37 Congreso Nacional
CENTRO DE CONVENCIONES INTERNACIONALES
Barcelona 22/25 MAYO 2024
seram Sociedad Española de Radiología Médica
FERM FUNDACIÓN ESPAÑOLA DE RADIOLOGÍA MÉDICA
RC RADIÒLEGS DE CATALUNYA

Soy urgenciólogo; ¿Me fío de lo que me dice la Inteligencia Artificial sobre una radiografía de tórax?

Hospital Universitario quirónsalud Madrid

Julia López Alcolea[1], Alejandro Díaz Moreno[1], Ana Fernández Alfonso[1], Raquel Cano Alonso[1], David García Castellanos[1], Vicente Martínez de Vega[1], Ana Álvarez Vazquez[1], Cristina Andreu Vázquez[2] y Lucía Sanabria Greciano[1].

[1]Hospital Universitario QuironSalud Madrid, Pozuelo

[2]Universidad Europea de Madrid. Facultad de Ciencias Biomédicas y de la Salud, Villaviciosa de Odón (Madrid)

ABSTRACT

Objetivos:

1. Evaluar la sensibilidad (S) y especificidad (E) de la Inteligencia Artificial (IA) y del residente en la lectura de radiografías torácicas urgentes (RxU) respecto al Gold Standard (GS; radiólogo senior).
2. Evaluar la concordancia entre la IA y el residente.

Material y Métodos:

Estudio observacional, descriptivo, transversal, retrospectivo y doble ciego en una muestra de 784 RXU.
La IA evalúa cinco variables categóricas (nódulo pulmonar, opacidad pulmonar, derrame pleural, neumotórax y fractura) y proporciona una lectura positiva, negativa o dudosa. Comparamos el rendimiento diagnóstico de la IA y del residente frente al GS y describimos la frecuencia de casos dudosos y otras variables no evaluadas por la IA.

Resultados:

Se obtienen resultados buenos para fractura y neumotórax (S=100%), moderados para opacidad pulmonar (S=71-76%) y razonables para derrame pleural (S=60-67%), con VPN>95% y AUC>0,8. Para nódulo pulmonar, la S del residente es moderada (75%) y la de IA baja (33%), con VPN=0,99.
La IA duda más que el residente, siendo el porcentaje de diagnósticos positivos bajo.
La concordancia entre ambos es baja (coeficiente Kappa=0,3) para todas las variables salvo derrame pleural, que es moderada (0,5).
La prevalencia de otras variables es: 16% mediastino, 20% material quirúrgico y 20% otros hallazgos pulmonares.

Conclusiones:

Nuestro estudio evalúa el impacto de la IA en la práctica clínica, comparando su validez frente al radiólogo. Obtiene alta AUC y VPN para todas las variables salvo en nódulo pulmonar y destaca alta S para fractura y neumotórax.
La concordancia entre IA y residente es baja.

OBJETIVOS:

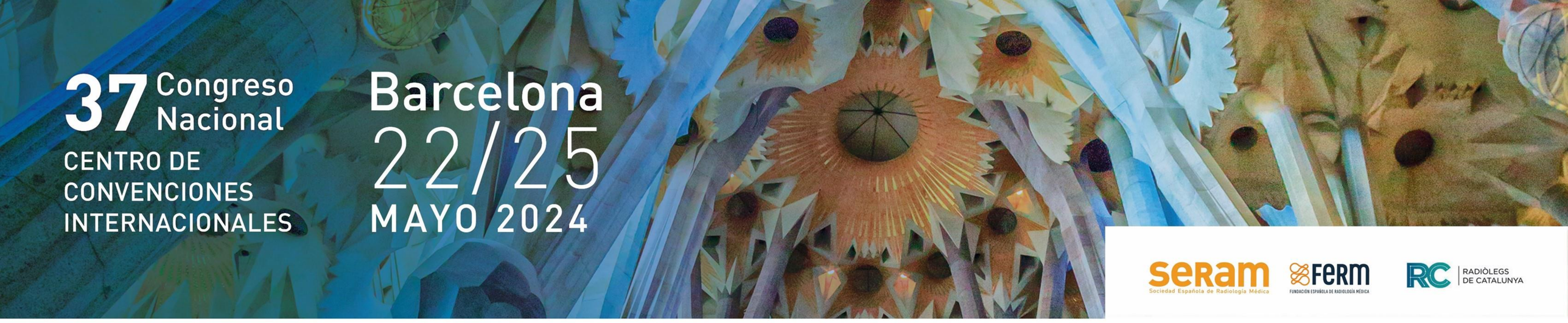
PRINCIPALES

- Evaluar la **sensibilidad** (S) y **especificidad** (E) de un software de **Inteligencia Artificial** (IA) y de un **residente** de radiología en la lectura de radiografías torácicas urgentes (RxU) frente a un **radiólogo senior**, considerado el Gold Standard (GS).
- Evaluar la **concordancia** entre la IA y el residente.

SECUNDARIOS

- Describir la frecuencia de **casos dudosos** en cada categoría y cuántos de ellos son considerados positivos por el GS.
- Evaluar **otras variables** que la IA no está entrenada para detectar con el objetivo de analizar sus debilidades y diagnósticos potenciales.

Figura 1. Diagrama del estudio.



CONTEXTO:

A lo largo de los últimos años ha aumentado significativamente el uso de la **Inteligencia Artificial** en el diagnóstico por imagen.

La **radiografía de tórax** constituye la prueba de imagen más común, particularmente en el servicio de **Urgencias**, dada su disponibilidad, bajo coste, baja dosis de radiación, y portabilidad. Es fundamental para el cribado, diagnóstico y manejo de múltiples patologías y el tiempo empleado en su lectura puede ser crucial [1-3]. Sin embargo, es una de las exploraciones radiológicas más difíciles de interpretar [4].

El empleo de la IA es especialmente **útil** en ambientes de trabajo en los que:

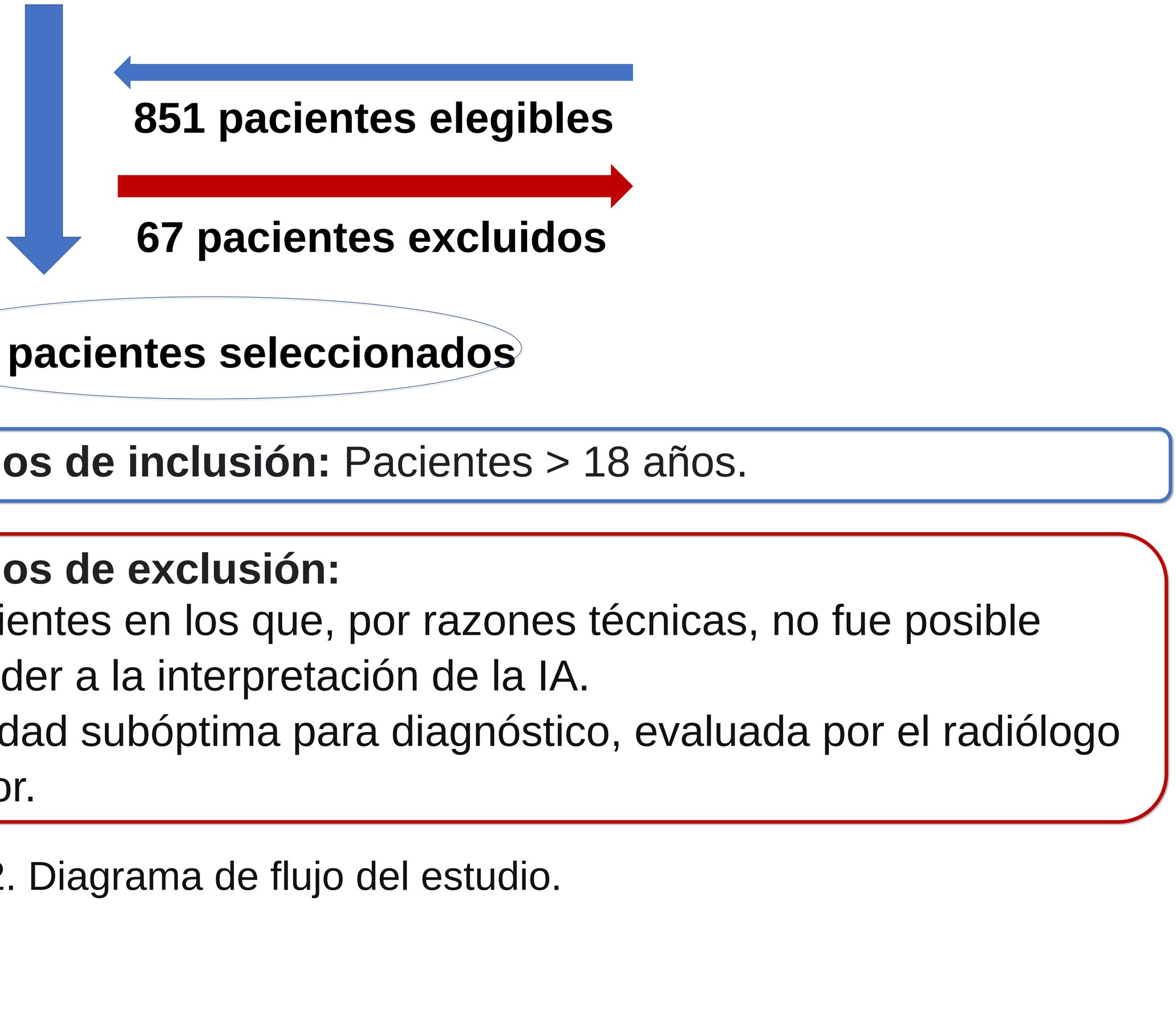
- Los radiólogos trabajan de forma remota.
- Hay una alta demanda de exploraciones radiológicas.
- Las radiografías son interpretadas directamente por los médicos de la Urgencia.

Ayudar a los médicos (tanto radiólogos como de Urgencias) en la evaluación de las radiografías torácicas con programas de IA podría contribuir a mejorar su precisión y el flujo de trabajo (priorizando el informe de estudios de pacientes con hallazgos urgentes y reduciendo el tiempo de lectura) y reduciendo el coste sanitario [5].

MATERIAL Y MÉTODOS:

Estudio observacional, descriptivo, transversal, retrospectivo y doble ciego en una muestra de 784 radiografías de tórax obtenidas con sistemas digitales de pacientes del Servicio de Urgencias del Hospital Universitario QuironSalud Madrid (HUQM) entre el 15 de Octubre y el 15 de Noviembre de 2022 (Figura 2).

Población diana: Pacientes del servicio de Urgencias del HUQM entre el 15 de Octubre y el 15 de Noviembre de 2022.



Criterios de inclusión: Pacientes > 18 años.

Criterios de exclusión:

- Pacientes en los que, por razones técnicas, no fue posible acceder a la interpretación de la IA.
- Calidad subóptima para diagnóstico, evaluada por el radiólogo senior.

Figura 2. Diagrama de flujo del estudio.

El algoritmo de **IA** consiste en una red neuronal convolucional profunda entrenada para detectar 5 variables cualitativas nominales:

1. **Nódulo pulmonar**
2. **Opacidad pulmonar**
3. **Derrame pleural**
4. **Neumotórax**
5. **Fractura**

Cataloga cada hallazgo como:

- "**positivo**" - (caja continua)
- "**negativo**"
- "**dudoso**" - (caja discontinua)

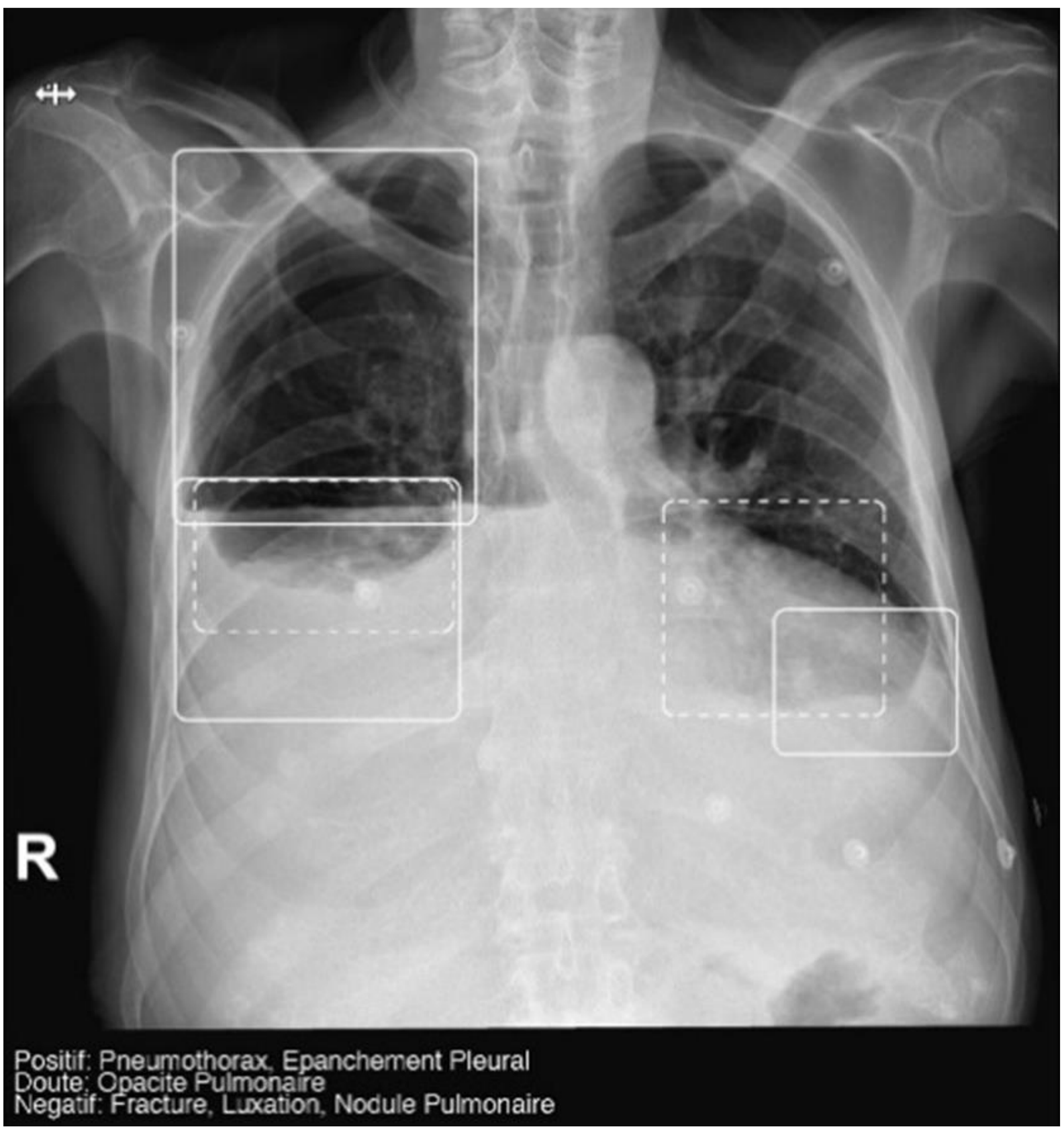


Figura 3. Rx de tórax PA analizada por la IA. Señala neumotórax derecho positivo, derrame pleural bilateral positivo (delimitados con cajas continuas) y opacidades bibasales dudosas (delimitadas con cajas discontinuas).

- Un **<u>residente de radiología</u>** (con 2 años de formación) leyó independientemente cada radiografía de forma cualitativa y clasificó los hallazgos como:
  - positivo
  - negativo
  - dudoso

- Un **<u>radiólogo senior (GS)</u>** (con más de 10 años de experiencia) también leyó independientemente cada estudio de forma cualitativa y clasificó los hallazgos como:
  - positivo
  - negativo

  En los casos en los que dudaba, consultó a un segundo radiólogo senior.

- Ninguno de los dos tuvo acceso a la información clínica y al resultado de la IA. Cada uno registró sus respectivos hallazgos en una tabla de datos y los hallazgos de la IA fueron registrados por un residente de primer año.

- Comparamos el **rendimiento diagnóstico** de la IA y del residente frente al GS y describimos **la frecuencia de hallazgos dudosos** y **de otras variables radiológicas no evaluadas por la IA** que podrían tener repercusión diagnóstica:
  - anomalías mediastínicas (cardiomegalia, hernia de hiato, ensanchamiento mediastínico)
  - material quirúrgico (grapas, válvulas metálicas, stents…)
  - otros hallazgos pulmonares (hiperinsuflación pulmonar)

Con respecto al **análisis estadístico**:

- Las variables cualitativas se describieron en frecuencias absolutas (n) y relativas (%).
- Para validar la capacidad diagnóstica de la IA y del residente, se usaron tablas de contingencia para calcular la sensibilidad, especificidad, valores predictivos, área bajo la curva (AUC) (en comparación con el GS) e intervalos de confianza (IC) al 95%.
- Para evaluar la concordancia interobservador entre el residente y la IA, se usó el coeficiente Kappa.

Variables cualitativas	Validación	Concordancia residente - IA
•Frecuencias absolutas (n) •Frecuencias relativas (%)	•Sensibilidad •Especificidad •VPP, VPN •AUC •IC 95%	•Coeficiente Kappa

Figura 4. Parámetros del análisis estadístico.

RESULTADOS:

CARACTERÍSTICAS DE LA MUESTRA

La muestra final incluyó 784 pacientes, de los cuales el 52% eran mujeres y el 48% eran hombres, con una mediana de edad de 58 años.

La mayoría de las radiografías (86,10%) incluyeron 2 proyecciones (posteroanterior y lateral) y fueron de calidad óptima (90,18%).

		Casos (n=784)
Edad del paciente mediana [Q1; Q3]		58 [44-74]
Sexo del paciente (n,%)		
	Hombre	378 (48,21)
	Mujer	406 (51, 79)
Número de proyecciones (n,%)		
	1	109 (13,9)
	2	675 (86,10)
Calidad de la radiografía (n,%)		
	Óptima	707 (90.18)
	Pobre	77 (9,82)
Q1: Primer cuartil = percentil 25. Q3: tercer cuartil = percentil 75		

Tabla 1. Características de la muestra.

PREVALENCIA DE LOS HALLAZGOS

La prevalencia está basada en el diagnóstico del GS y ajustada por intervalo de confianza, siendo por tanto más precisa.

El hallazgo más prevalente fue la opacidad pulmonar (14,29%), mientras que el menos prevalente fue el neumotórax (0,38%).

Curiosamente, la prevalencia de las variables no analizadas por la IA fue mayor que la de las analizadas por la misma: 16,33% mediastino, 20,15% material quirúrgico y 20,82% otros hallazgos.

	Casos
	(n=784)
Prevalencia* (n, % [CI 95%])	
Fractura/luxación	10 (1,28 [0,61-2,33])
Neumotórax	3 (0,38 [0,079-1,1])
Nódulo pulmonar	20 (2,55 [1,6-3,93])
Opacidad pulmonar	112 (14,29 [11-16,2])
Derrame pleural	71 (9.06 [7,1-11,3])
Mediastino	128 (16,33 [14-19,1])
Material quirúrgico	158 (20.15 [17-23,1])
Otros hallazgos	163 (20,82 [18-23,9])

*basado en el diagnóstico del GS

* La prevalencia ajustada por intervalo de confianza es más precisa

Tabla 2. Prevalencia de los hallazgos.

RENDIMIENTO DIAGNÓSTICO

En lo que respecta a la eficacia diagnóstica de la IA y del residente, los casos catalogados como dudosos fueron excluidos para calcular la S, E, VPP, VPN y AUC.

En términos generales, ambos obtuvieron resultados estadísticos:

- Buenos para la detección de fractura y neumotórax, con una sensibilidad del 100%.
- Moderados para la detección de opacidad pulmonar, con una sensibilidad entre 71-76%, VPN entre 95-96% y AUC de 0,85.
- Pobres para la detección de derrame pleural, con una sensibilidad entre 60-67% y VPN entre 96-97% y AUC entre 0,79-0,82.

En cuanto al nódulo pulmonar, la IA obtuvo una sensibilidad baja (33%) y el residente moderada (75%).

Inteligencia Artificial

La **sensibilidad** de la IA fue:

- alta (100%) para fractura y neumotórax
- moderada para opacidad pulmonar (75,6%)
- razonable para derrame pleural (59,7%)
- baja para nódulo pulmonar (33,3%).

De los casos catalogados como **dudosos** por la IA, fueron escasos los catalogados como positivos por el GS, salvo en la categoría de nódulo pulmonar.

		Total de casos	
		(n=784)	
		Rendimiento diagnóstico **RESIDENTE**	Rendimiento diagnóstico **IA**
FRACTURA (ratio diagnósticos dudosos / certeros)		0 / 784	19 / 765
Ratio diagnósticos dudosos /% positivos		NA	19/15,79
	Sensibilidad (%, IC 95%)	100 (69,2-100)	100 (59-100)
	Especificidad (%, IC 95%)	99,9 (99,3-100)	98,7 (97,6-99,4)
	VPP (%, IC 95%)	90,9 (58,7-99,8)	41,2 (184-67.1)
	VPN (%, IC 95%)	100 (99,5-100)	100 (99,5-100)
	AUC (IC 95%)	0,999 (0,998-1)	0,993 (0989-0,997)
NEUMOTÓRAX (ratio diagnósticos dudosos / certeros)		0 / 784	9 /775
Ratio diagnósticos dudosos /% positivos		NA	9/11,11
	Sensibilidad (%, IC 95%)	100 (29- 100)	100 (15,8-100)
	Especificidad (%, IC 95%))	100 (99,5-100)	100 (99,5-100)
	VPP (%, IC 95%)	100 (29,2-100)	100 (15,8-100)
	VPN (%, IC 95%)	100 (99,5-100)	100 (99,5-100)
	AUC (IC 95%)	1(1-1)	1 (1-1)
NÓDULO PULMONAR (ratio diagnósticos dudosos / certeros)		4/780	16/768
Ratio diagnósticos dudosos /% positivos		4/0	16/50
	Sensibilidad (%, IC 95%)	75 (50,9-91,3)	33,3 (9,92-65,1)
	Especificidad (%, IC 95%)	99,3 (98,5-99,8)	99,6 (98,8-99,9)
	VPP (%, IC 95%)	75 (50,9-91,3)	57,1 (18,4-90,1)
	VPN (%, IC 95%)	99,3 (98,5-99,8)	98,9 (97,6-99,5)
	AUC (IC 95%)	0,872 (0,774-0,969)	0,665 (0,525-0,804)
OPACIDAD PULMONAR (ratio diagnósticos dudosos / certeros)		16/768	170/614
Ratio diagnósticos dudosos /% positivos		16/50	170/17,65
	Sensibilidad (%, 95% CI)	71,2 (61,4-79,6)	75,6 (64,9 – 84,4)
	Especificidad (%, 95% CI)	99.1 (98-99,7)	95,3 (93,1-96,6)
	VPP (%, 95% CI)	92,5 (84,4-97,2)	71,3 (60,6-80,5)
	VPN (%, 95% CI)	95,6 (93,8-97)	96,2 (94,2-97,7)
	AUC (95% CI)	0,851 (0,807-0,895)	0,855 (0,807-0,902)
DERRAME PLEURAL (ratio diagnósticos dudosos / certeros)		12/772	21/763
Ratio diagnósticos dudosos /% positivos		12/8,33	21/42,86
	Sensibilidad (%, 95% CI)	67,1 (54,6-77,9)	59,7 (46,4-71,9)
	Especificidad (%, 95% CI)	97,4 (96-98,5)	98,9 (97,8-99,5)
	VPP (%, 95% CI)	72,3 (59,8-82,7)	82,2 (67,9-92)
	VPN (%, 95% CI)	96.7 (95,2-97,9)	96,5 (94,9-97,7)
	AUC (95% CI)	0,823 (0,767-0,879)	0,793 (0,731-0,854)

Tabla 3. Tabla resumen del rendimiento diagnóstico Residente/IA.

Desglose por hallazgos (IA)

FRACTURA (ratio diagnósticos dudosos / certeros)	19 / 765
Ratio diagnósticos dudosos /% positivos	19/15,79
Sensibilidad (%, IC 95%)	100 (59-100)
Especificidad (%, IC 95%)	98,7 (97,6-99,4)
VPP (%, IC 95%)	41,2 (184-67.1)
VPN (%, IC 95%)	100 (99,5-100)
AUC (IC 95%)	0,993 (0989-0,997)

La prevalencia fue baja: 10 casos (1,28% [0,61-2,33]).

De los 19 diagnósticos dudosos solo el 15,79% fueron verdaderos. Se obtuvieron excelentes resultados, con sensibilidad y VPN del 100% y AUC de 0,99.

Fue incapaz de diferenciar entre fracturas agudas y crónicas (callos de fractura).

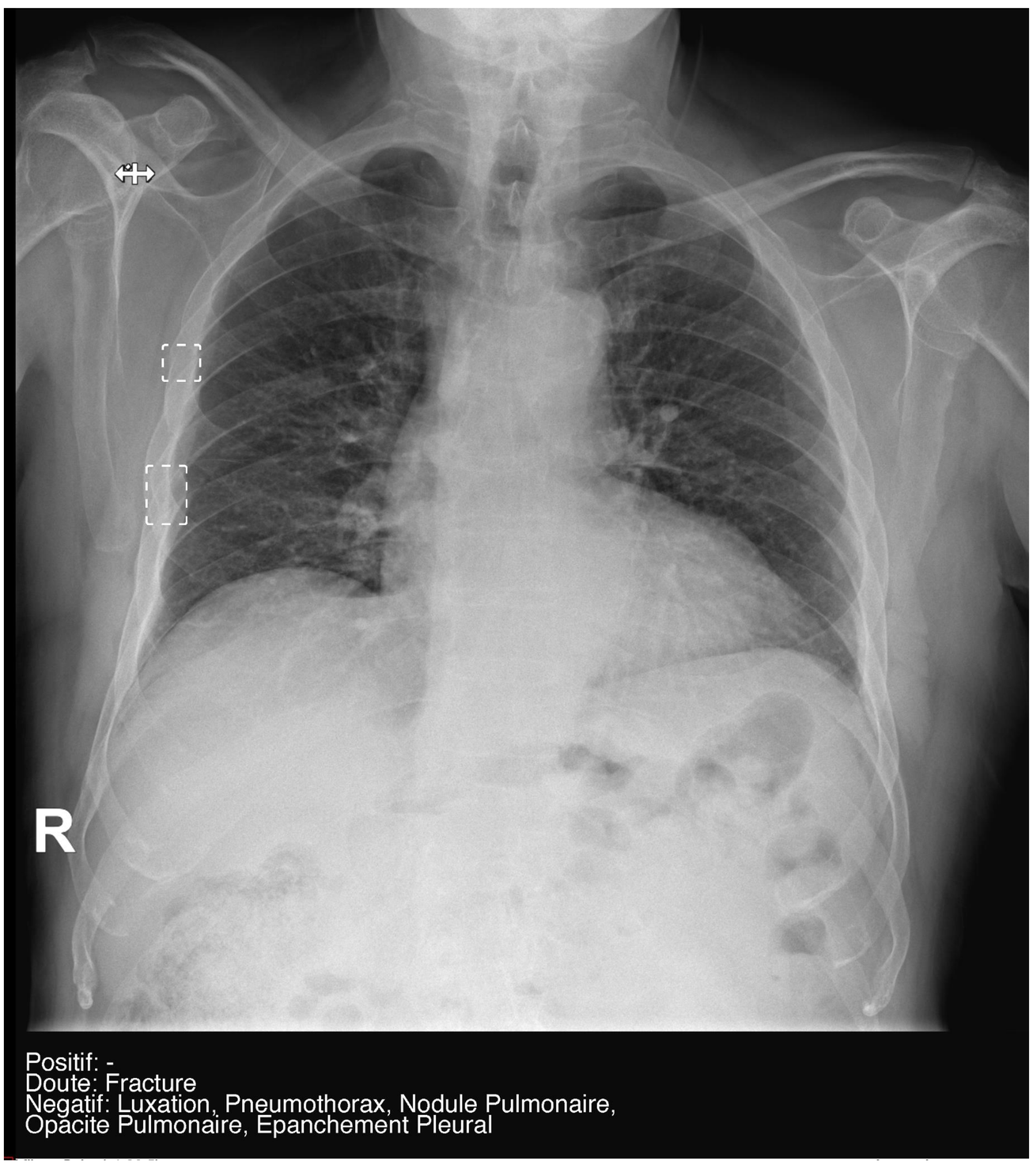


Figura 5. Radiografía de tórax PA analizada por la IA, que señala 2 fracturas costales derechas dudosas, que en realidad corresponden a callos de fractura.

NEUMOTÓRAX (ratio diagnósticos dudosos / certeros)	9 / 775
Ratio diagnósticos dudosos /% positivos	9 / 11,11
Sensibilidad (%, IC 95%)	100 (15,8-100)
Especificidad (%, IC 95%)	100 (99,5-100)
VPP (%, IC 95%)	100 (15,8-100)
VPN (%, IC 95%)	100 (99,5-100)
AUC (IC 95%)	1 (1-1)

La prevalencia también fue baja: 3 casos (0,38 % [0,079-1,1]).
La prevalencia baja puede afectar al rendimiento diagnóstico, especialmente al intervalo de confianza, que se vuelve más amplio y, por tanto, menos preciso. No obstante, la IA detectó correctamente los 3 casos.
Solo el 11% de los diagnósticos dudosos fueron verdaderos.
Todos los parámetros estadísticos obtuvieron excelentes resultados. La sensibilidad fue del 100% y el VPN del 100%.

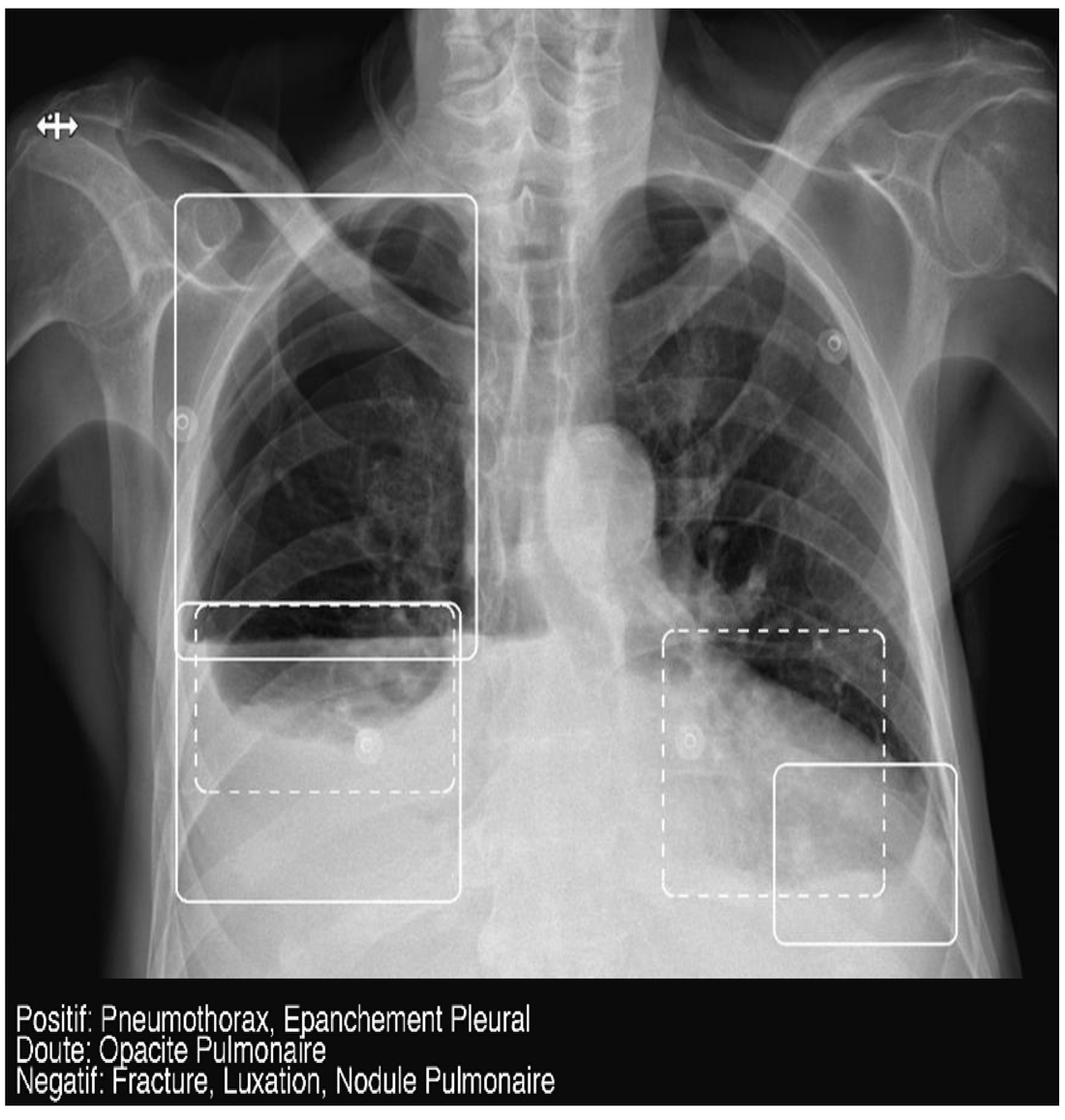


Figura 6. Rx de tórax PA analizada por la IA, que (entre otros) señala un neumotórax derecho positivo verdadero.

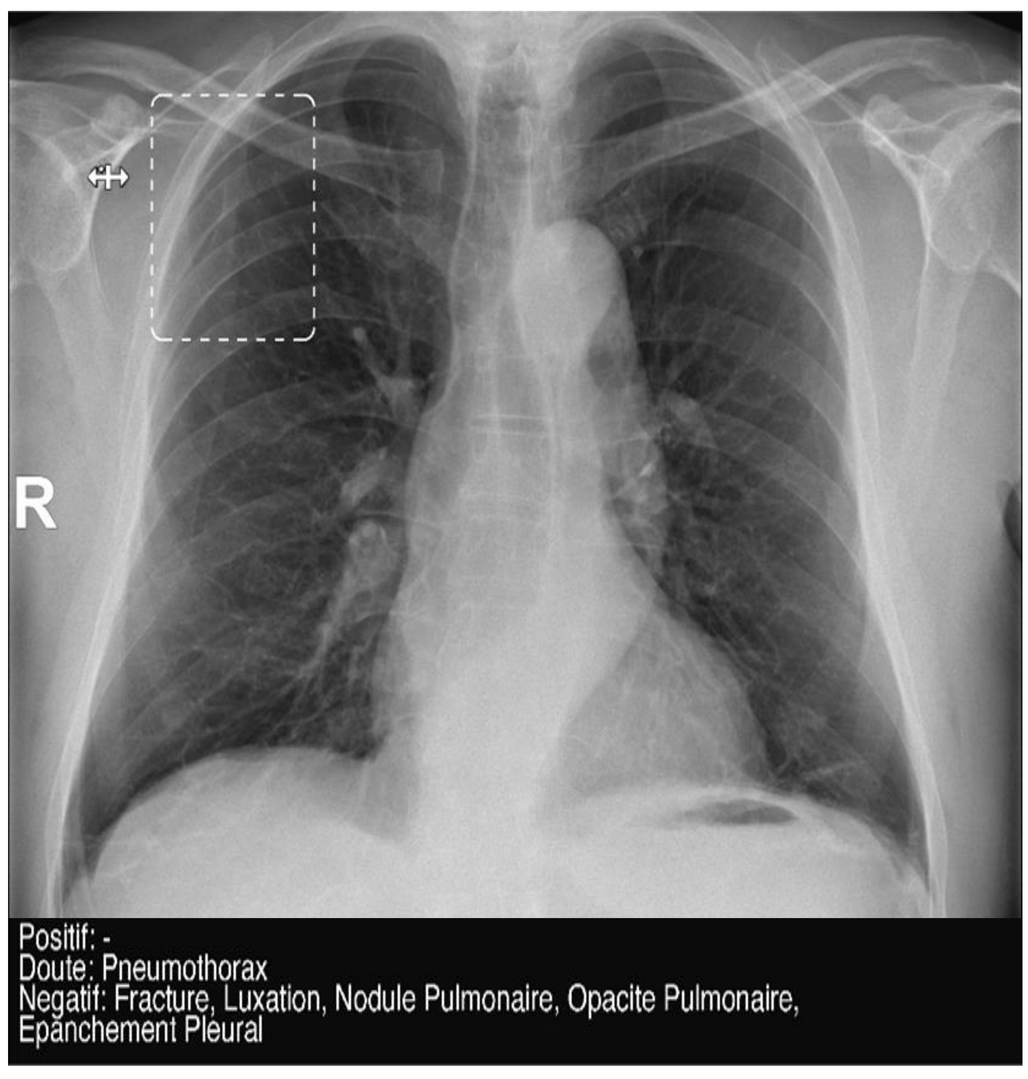


Figura 7. Rx de tórax PA analizada por la IA, que señala un neumotórax derecho dudoso erróneo.

NÓDULO PULMONAR (ratio diagnósticos dudosos / certeros)	16 / 768
Ratio diagnósticos dudosos /% positivos	16 / 50
Sensibilidad (%, IC 95%)	33,3 (9,92-65,1)
Especificidad (%, IC 95%)	99,6 (98,8-99,9)
VPP (%, IC 95%)	57,1 (18,4-90,1)
VPN (%, IC 95%)	98,9 (97,6-99,5)
AUC (IC 95%)	0,665 (0,525-0,804)

La prevalencia fue baja: 20 casos (2,55 %[1,6-3,93]).
Por tanto, los intervalos de confianza fueron amplios.
De los 16 diagnósticos dudosos, el 50% fueron verdaderos.
La sensibilidad fue del 33,3% (baja). Por otro lado, el VPN fue del 98,9 % (muy bueno).

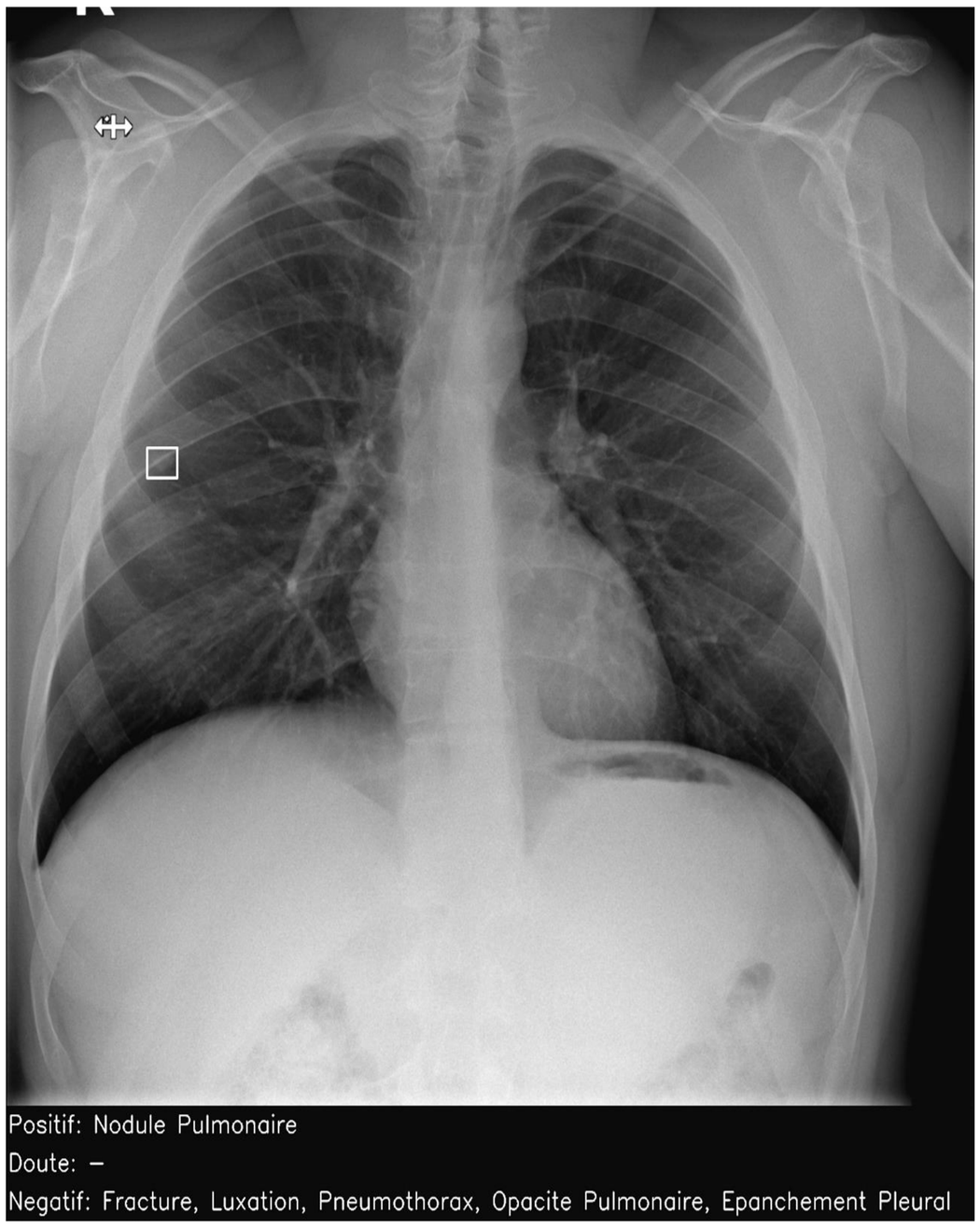


Figura 8. Rx de tórax PA analizada por la IA, que señala un nódulo pulmonar derecho positivo verdadero.

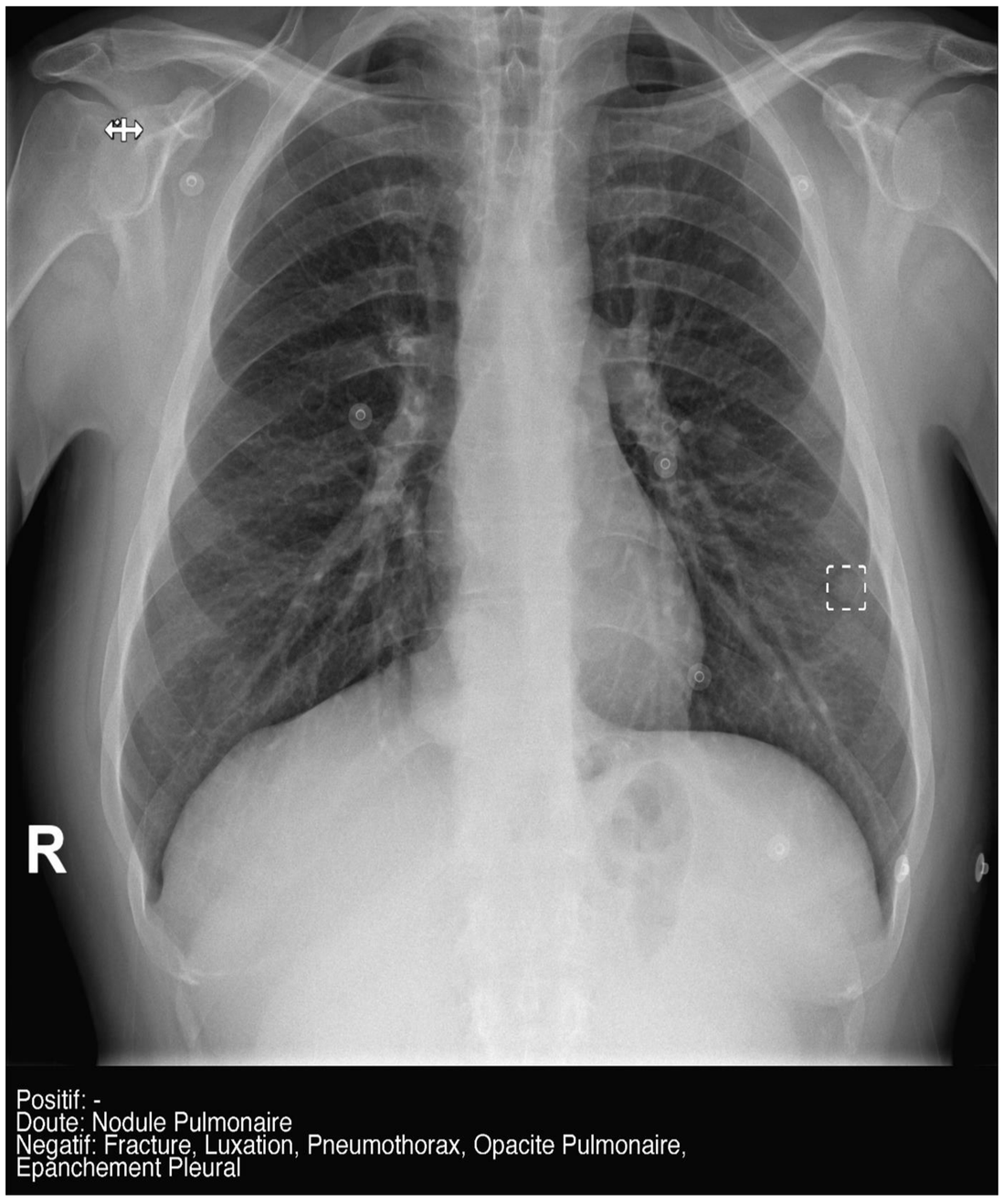


Figura 9. Rx de tórax PA analizada por la IA, que señala un nódulo pulmonar izquierdo dudoso verdadero.

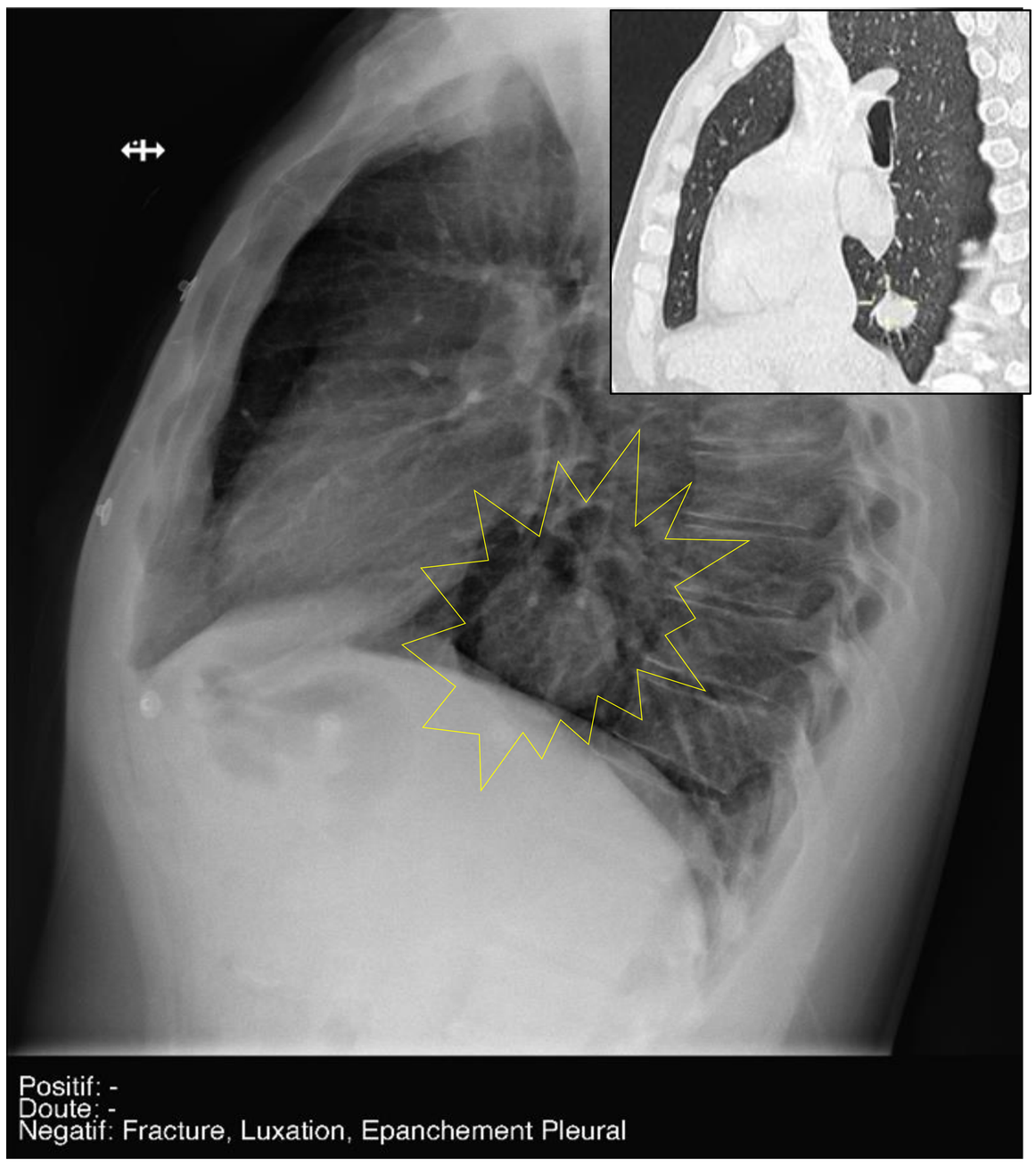


Figura 10. Rx de tórax sagital analizada por la IA, que no identificó un nódulo pulmonar en LII (negativo según su lectura), delimitado con bordes espiculados amarillos y corroborado con un TC torácico sagital complementario.

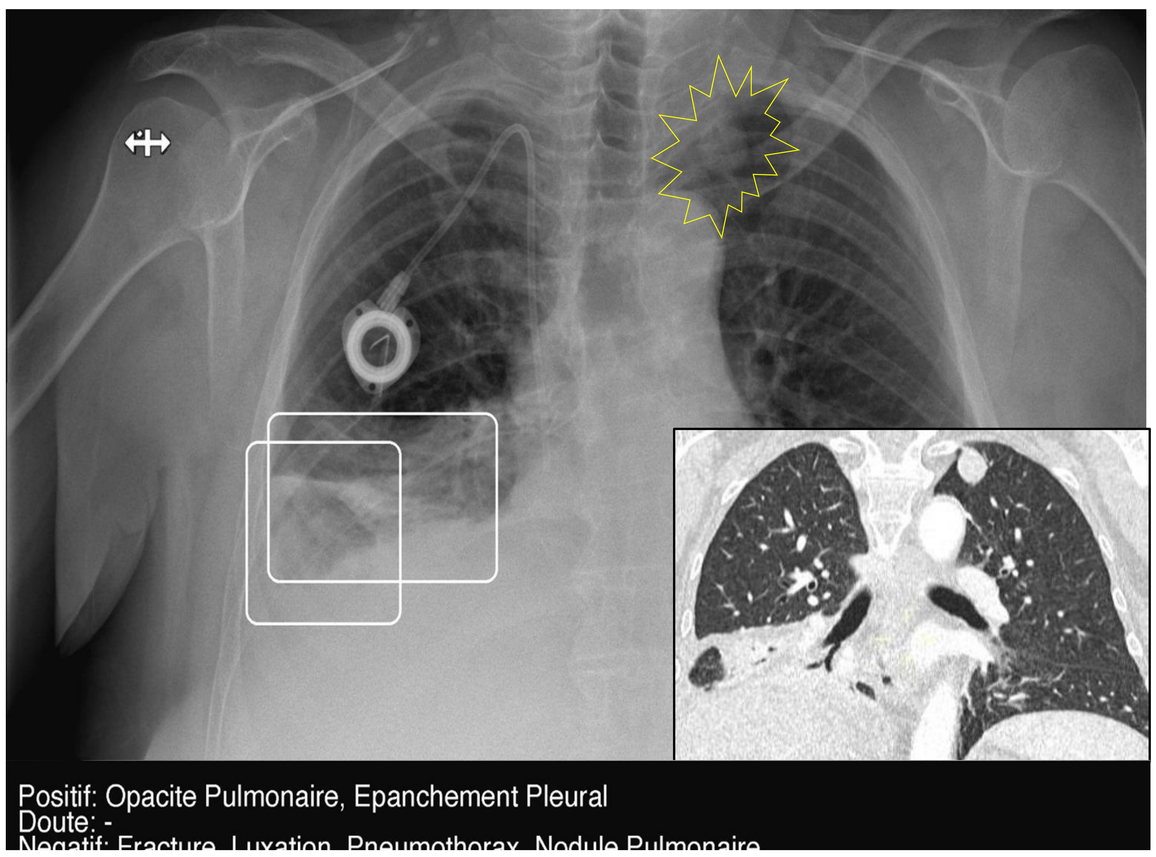


Figura 11. Rx de tórax PA analizada por la IA, que identificó correctamente derrame pleural y opacidad pulmonar derechos, pero no identificó un nódulo pulmonar en LSI (negativo según su lectura), delimitado con bordes amarillos y corroborado con un TC torácico coronal complementario.

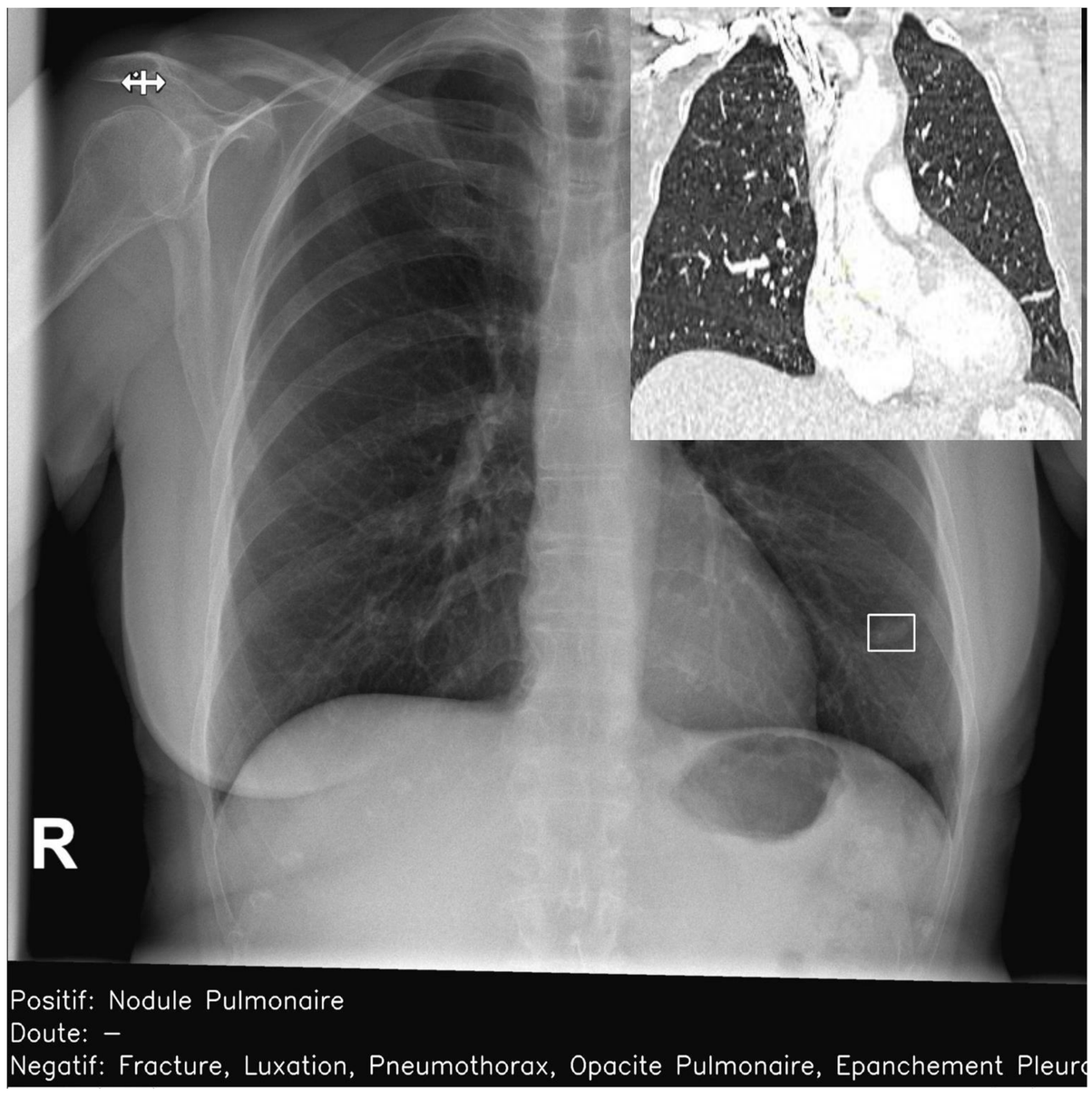


Figura 12. Rx de tórax PA analizada por la IA, que señala un nódulo pulmonar positivo falso, que corresponde a una atelectasia laminar en LII, como se muestra en e TC torácico coronal.

OPACIDAD PULMONAR (ratio diagnósticos dudosos / certeros)	170 / 614
Ratio diagnósticos dudosos /% positivos	170 / 17,65
Sensibilidad (%, IC 95%)	75,6 (64,9-84,4)
Especificidad (%, IC 95%)	95,3 (93,1-96,6)
VPP (%, IC 95%)	71,3 (60,6-80,5)
VPN (%, IC 95%)	96,2 (94,2-97,7)
AUC (IC 95%)	0,855 (0,807-0,902)

La prevalencia fue la más alta de todos los ítems: 112 casos (14,29 % [11-16,2]), por lo que los intervalos de confianza fueron más precisos.

La IA señaló muchos diagnósticos dudosos (170), de los cuales solo el 17% fueron verdaderos. La mayoría de los casos dudosos correspondían a la vascularización pulmonar normal del LM y LID.

La sensibilidad fue del 75,6% (moderada), mientras que el VPN fue del 96,2% (alto) y el AUC de 0,86 (alta).

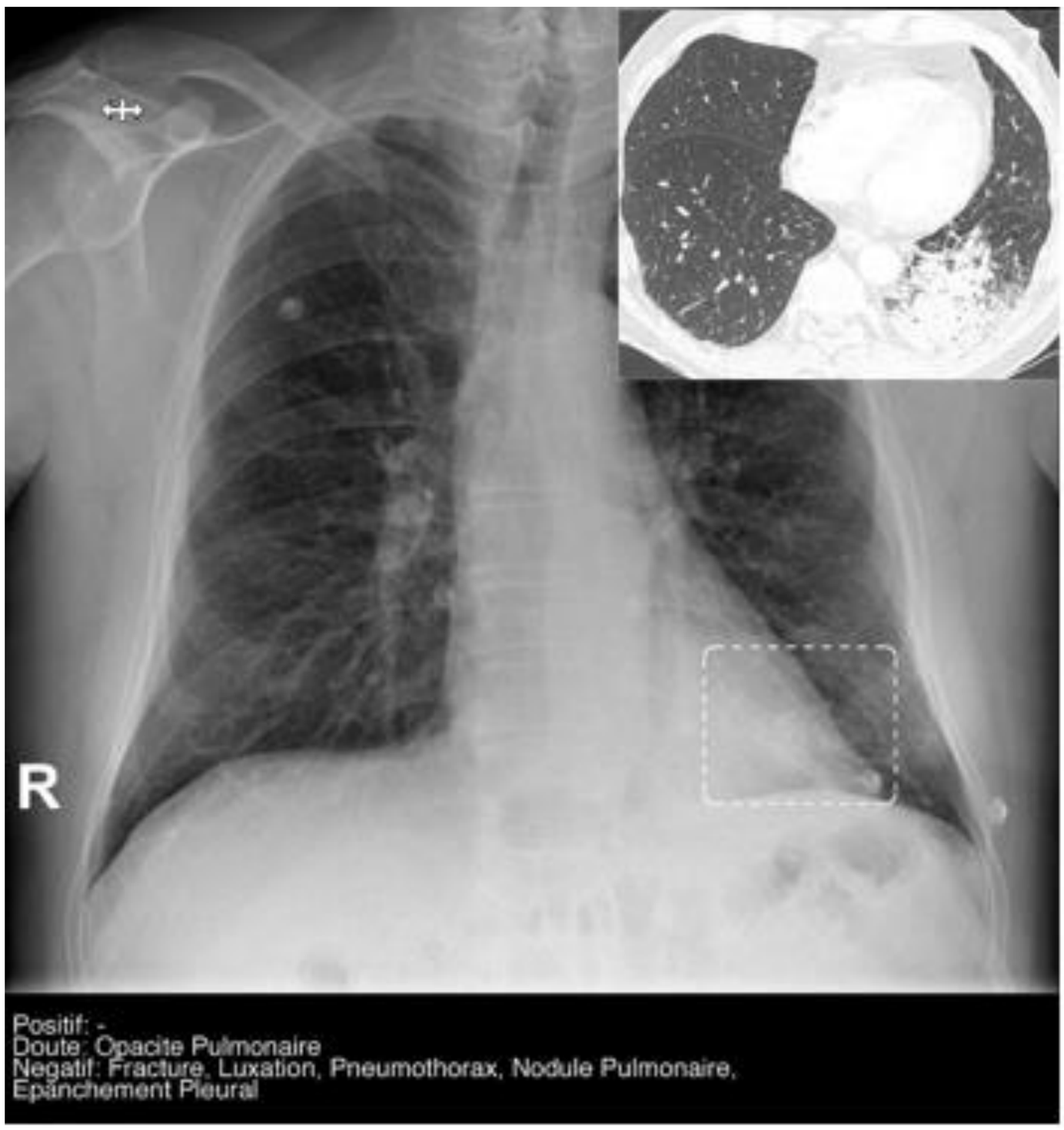
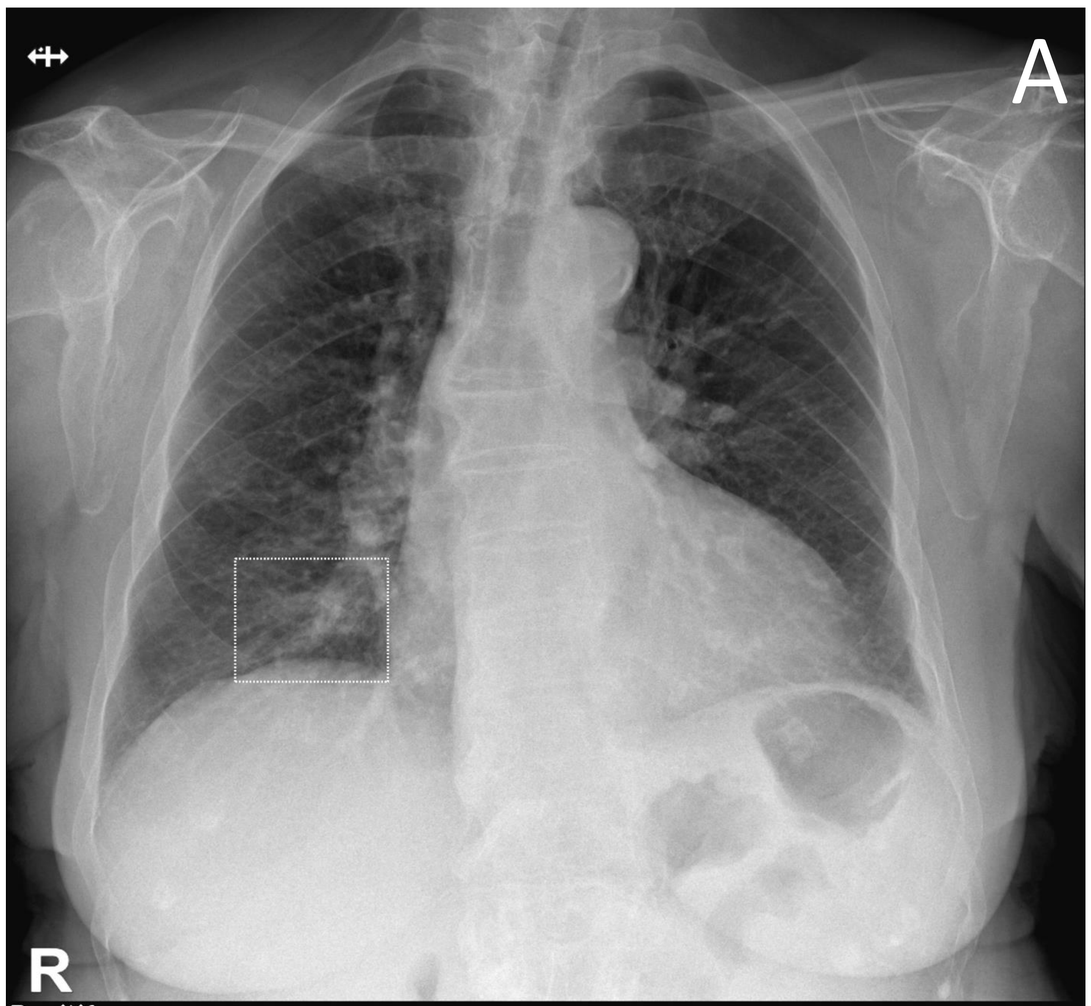
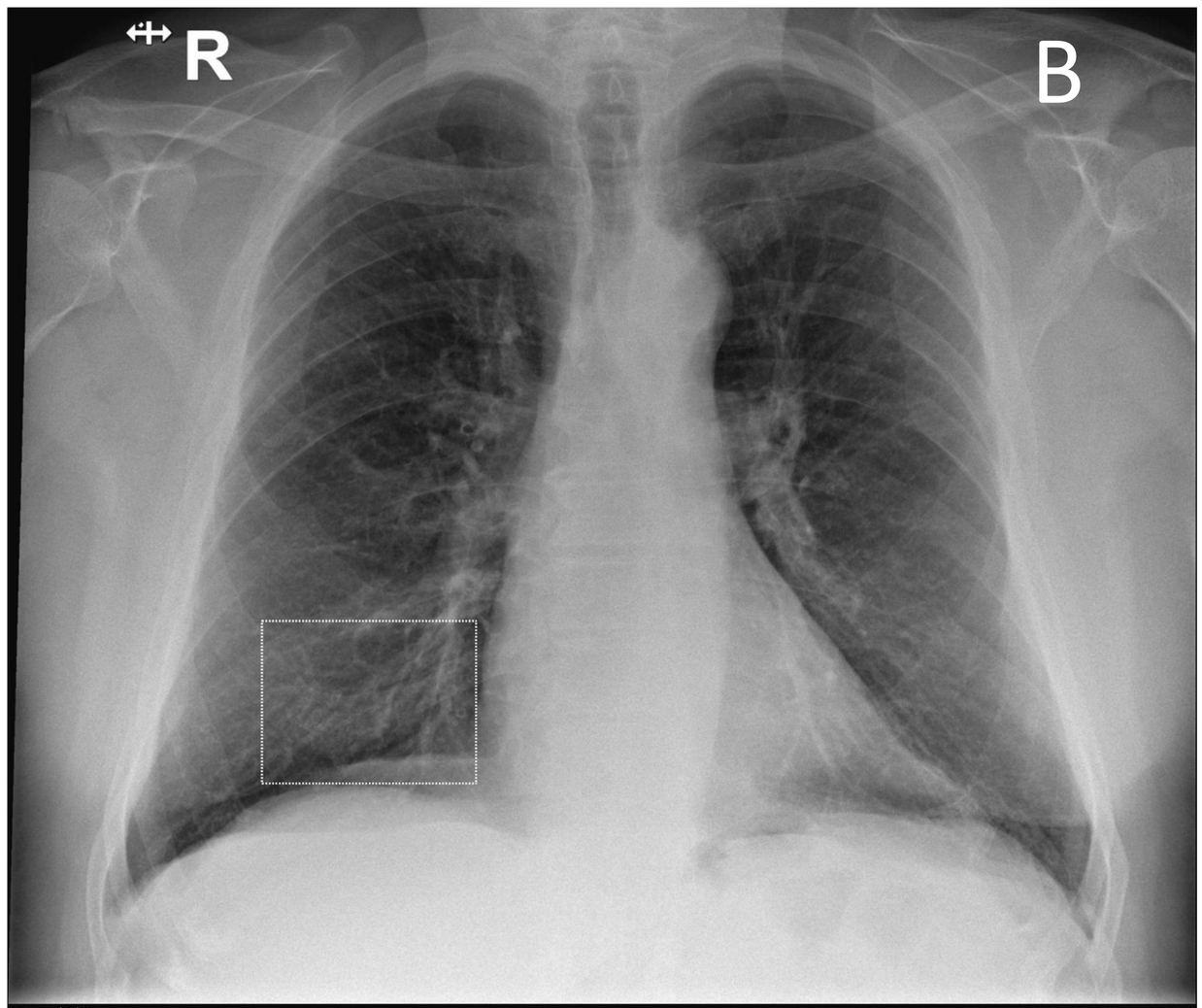


Figura 13. Rx de tórax PA analizada por la IA, que señala una opacidad pulmonar retrocardiaca dudosa que es verdadera, tal y como se objetiva en el TC torácico axial complementario.





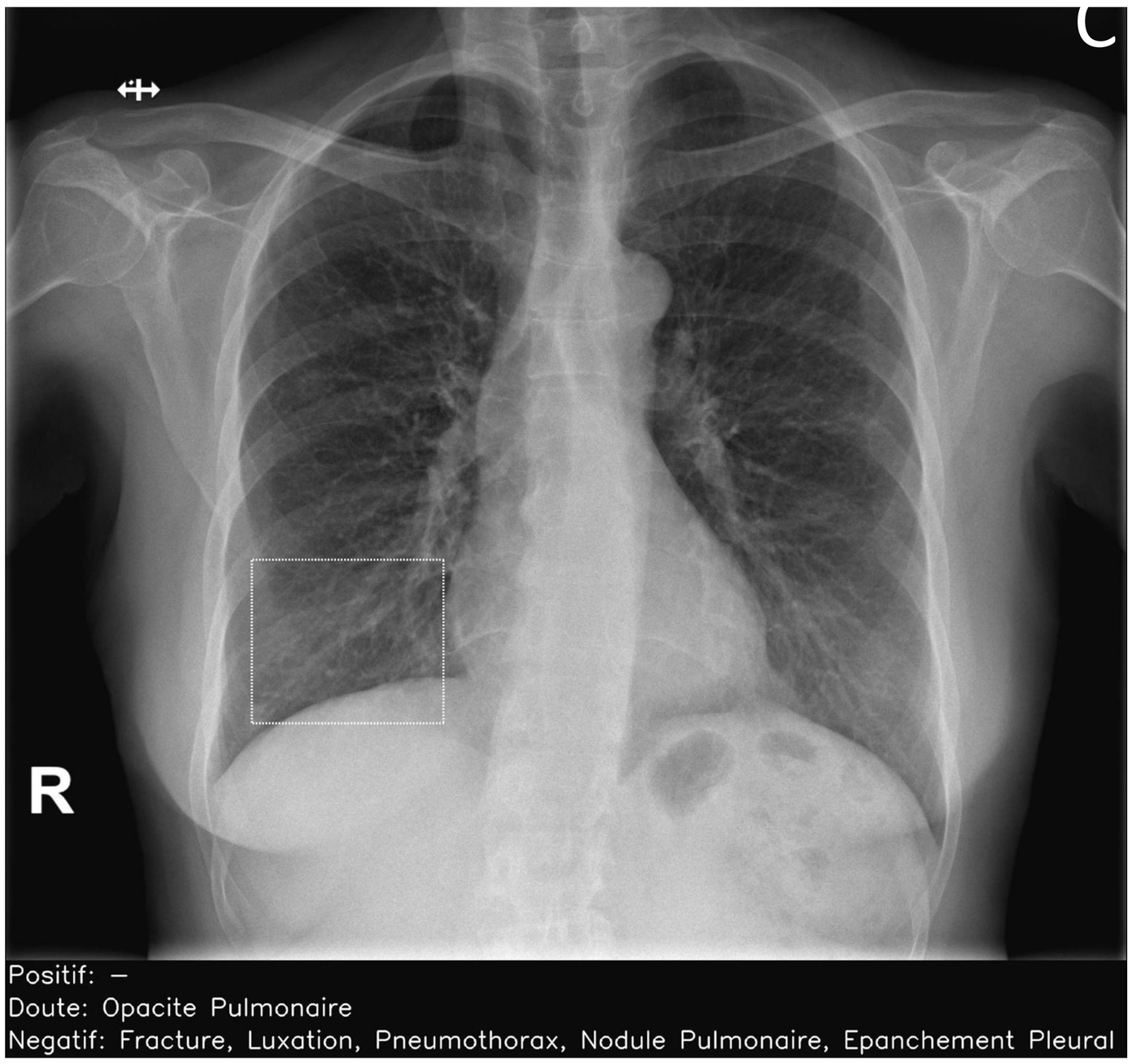


Figura 14. 3 rx de tórax PA (A, B C) analizadas por la IA, que señala opacidades pulmonares en LID dudosas falsas, que realmente corresponden a vasos pulmonares normales.

DERRAME PLEURAL (ratio diagnósticos dudosos / certeros)	21 / 763
Ratio diagnósticos dudosos /% positivos	21 / 42,86
Sensibilidad (%, IC 95%)	59,7 (46,4-71,9)
Especificidad (%, IC 95%)	98,9 (97,8-99,5)
VPP (%, IC 95%)	82,2 (67,9-92)
VPN (%, IC 95%)	96,5 (94,9-97,7)
AUC (IC 95%)	0,793 (0,731-0,854)

La prevalencia fue moderada: 71 (9.06 % [7,1-11,3]), la segunda más alta.
De los 21 diagnósticos dudosos, solo el 42% fueron verdaderos.
La sensibilidad fue del 59,7% (pobre). La IA detectó correctamente los derrames moderados y severos, pero tuvo dificultades a la hora de diferenciar entre derrames leves y hallazgos normales como aplanamiento diafragmático secundario a hiperinsuflación o simplemente como variante anatómica, en los que puede haber un leve borramiento del ángulo costofrénico.
Sin embargo, el VPN fue del 96,5% y el AUC de 0,79 (buenos).

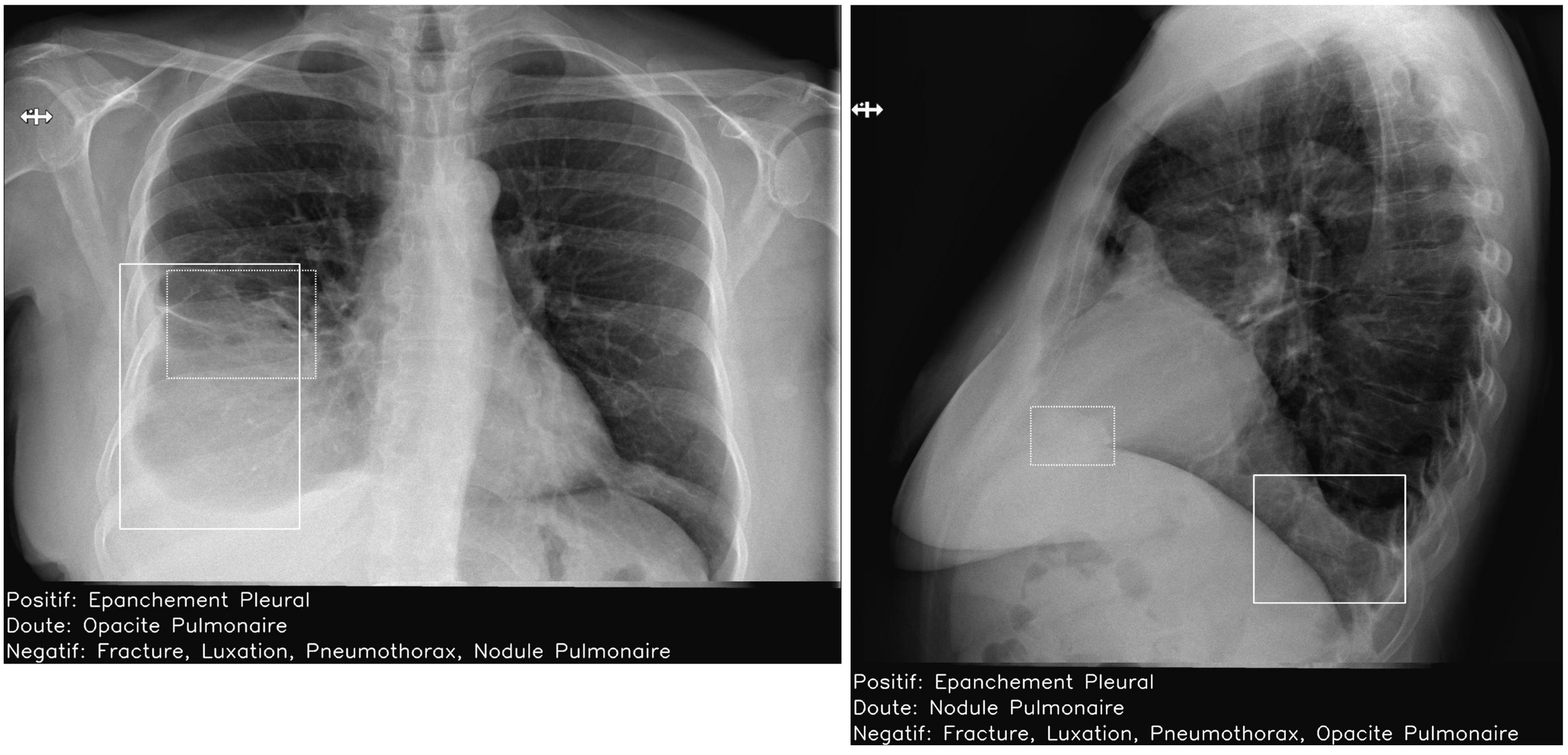


Figura 15. Rx de tórax PA y lateral del mismo paciente analizadas por la IA, que señala (entre otros) derrame pleural positivo verdadero (de cuantía moderada).

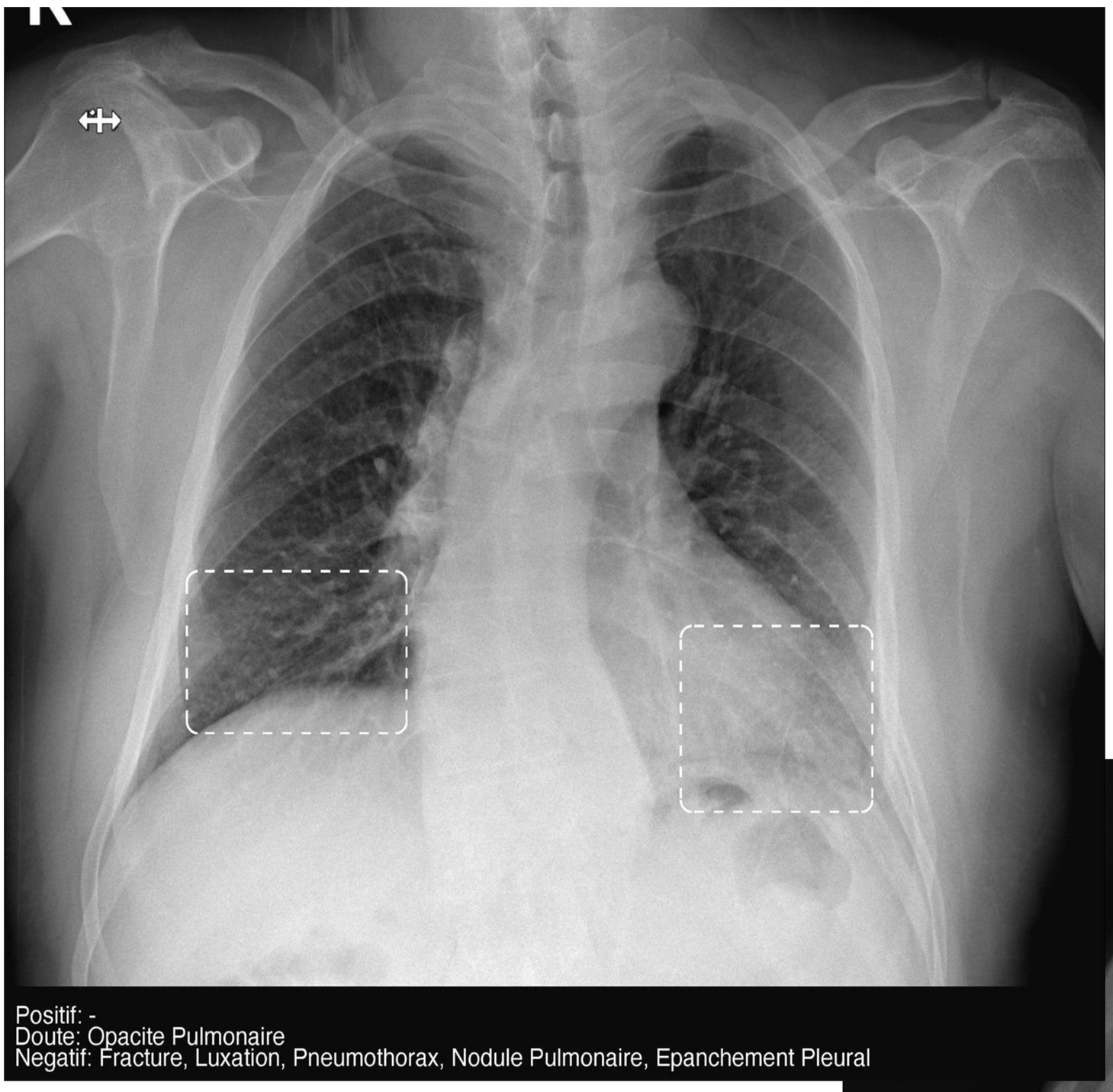


Figura 16. Rx de tórax PA y lateral del mismo paciente analizadas por la IA, que en la lateral señala derrame pleural positivo falso (corresponde a un sutil borramiento del ángulo costofrénico posterior izquierdo, como puede observarse en el TC torácico sagital complementario).

Residente

El residente dudó menos en todas las categorías y obtuvo la misma sensibilidad para fracturas y neumotórax (100%), moderadamente mayor para nódulo pulmonar (75%), discretamente mayor para derrame pleural (67,1%) y discretamente menor para opacidad pulmonar (71,2%).

OTRAS VARIABLES

La prevalencia de otras variables fue: 16,33% para mediastino, 20,15% material quirúrgico y 20,82% otros hallazgos pulmonares.

La cardiomegalia fue el hallazgo más común, (80,47% del total), mientras que la hiperinsuflación solo constituyó el 7,36%.

	Casos
	(n=784)
Prevalencia* (n, % [CI 95%])	
Mediastino	128 (16,33 [14-19,1])
Material quirúrgico	158 (20.15 [17-23,1])
Otros hallazgos	163 (20,82 [18-23,9])

*basada en el diagnóstico del GS

	Casos
	(n=128)
Proporción en mediastino (n, %)	
Cardiomegalia	103 (80.47)
Hernia de hiato	16 (12.50)
Ensanchamiento mediastínico superior	6 (4.69)

	Casos
	(n=163)
Proporción de otros hallazgos (n, %)	
Hiperinsuflación	12 (7.36)

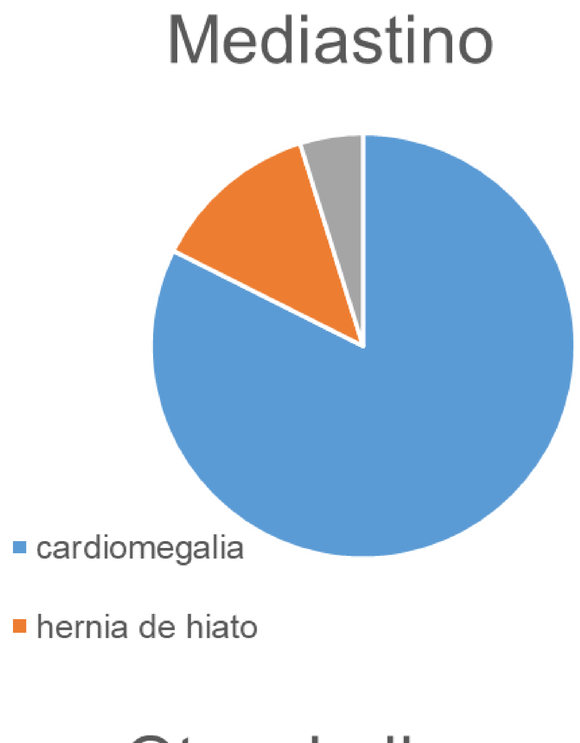


Tabla 4. Prevalencia de otros hallazgos.

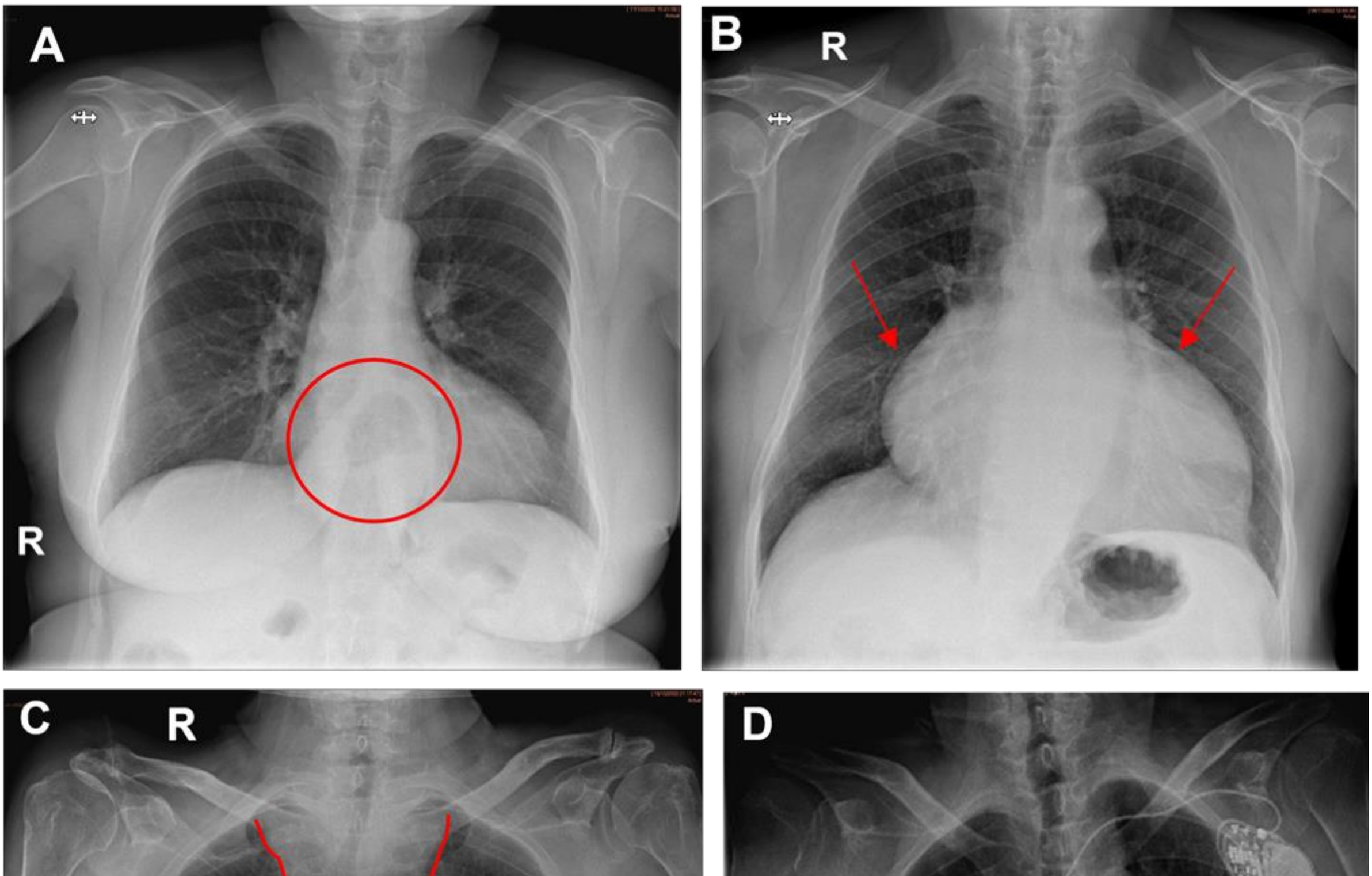
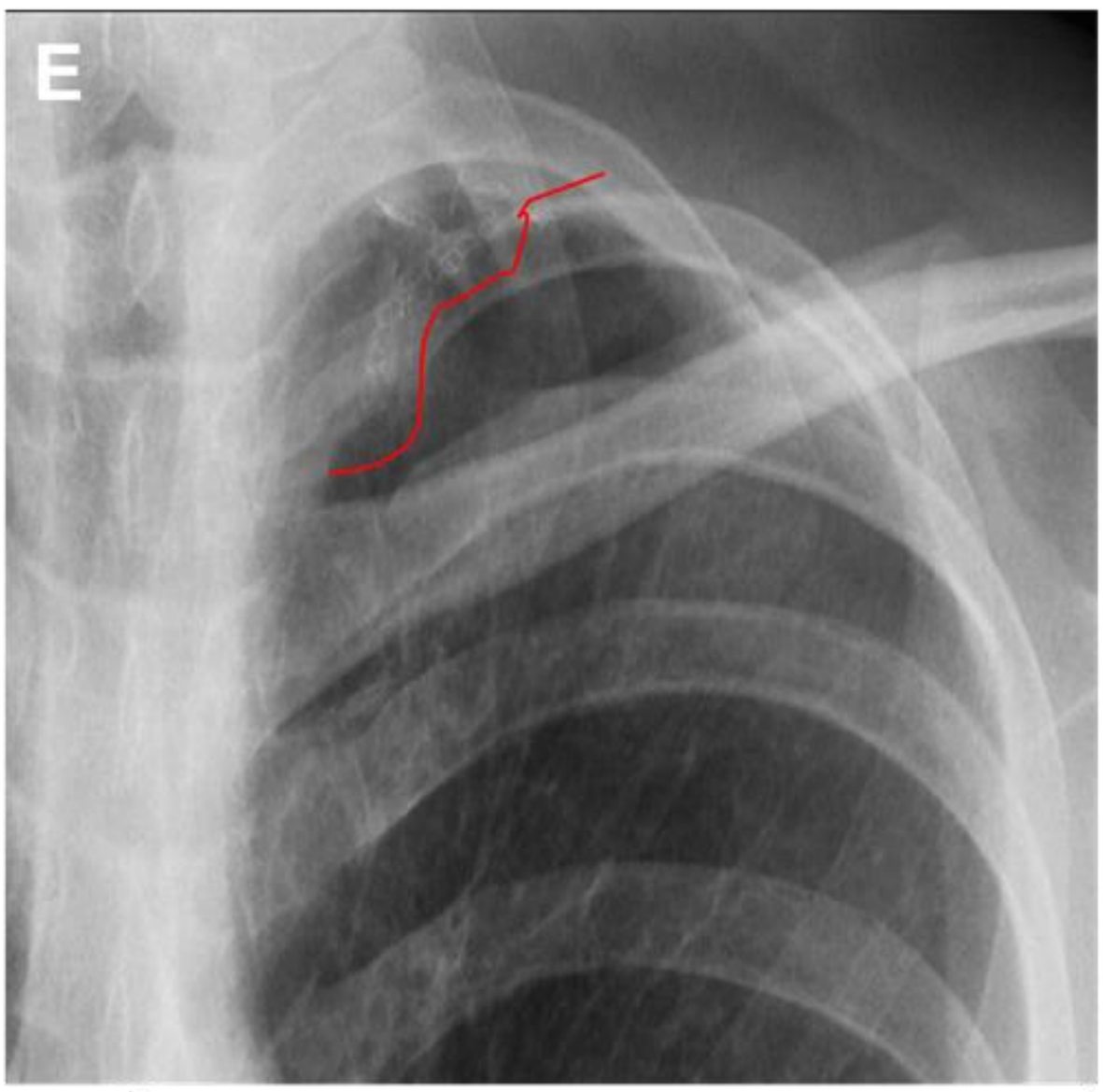
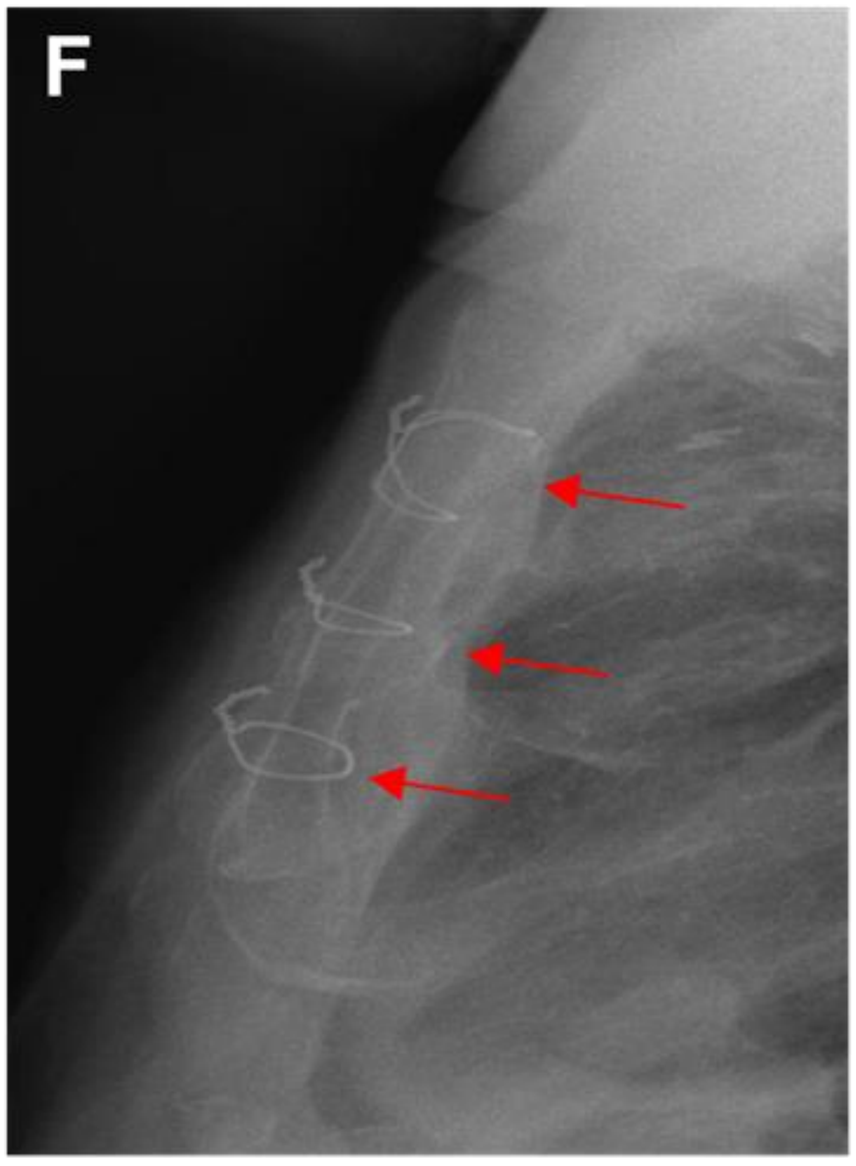
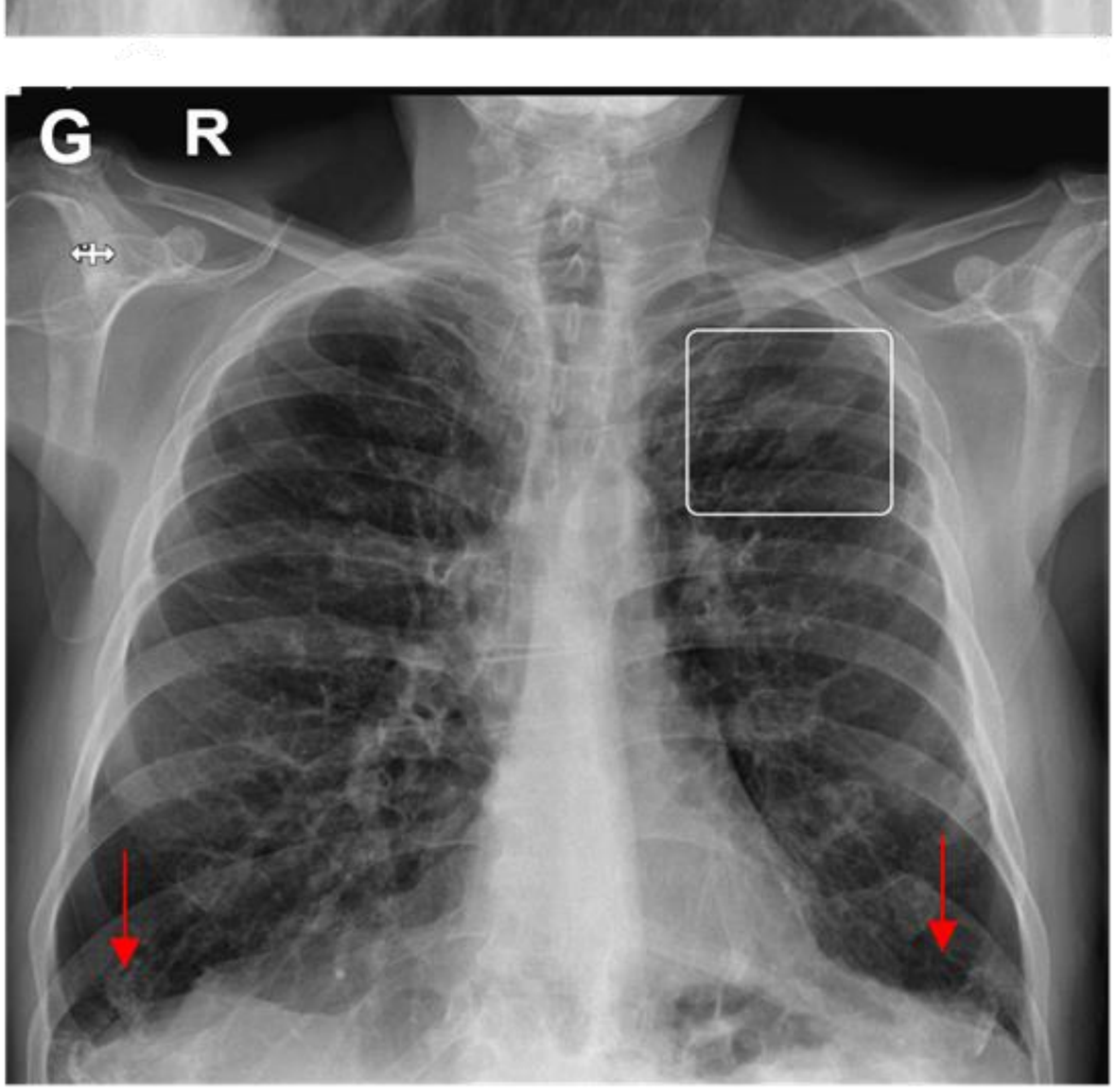


Figura 17. Radiografías de tórax PA con otros hallazgos. A) Hernia de hiato. B) Cardiomegalia. C) Ensanchamiento mediastínico superior. D) Marcapasos cardiaco.







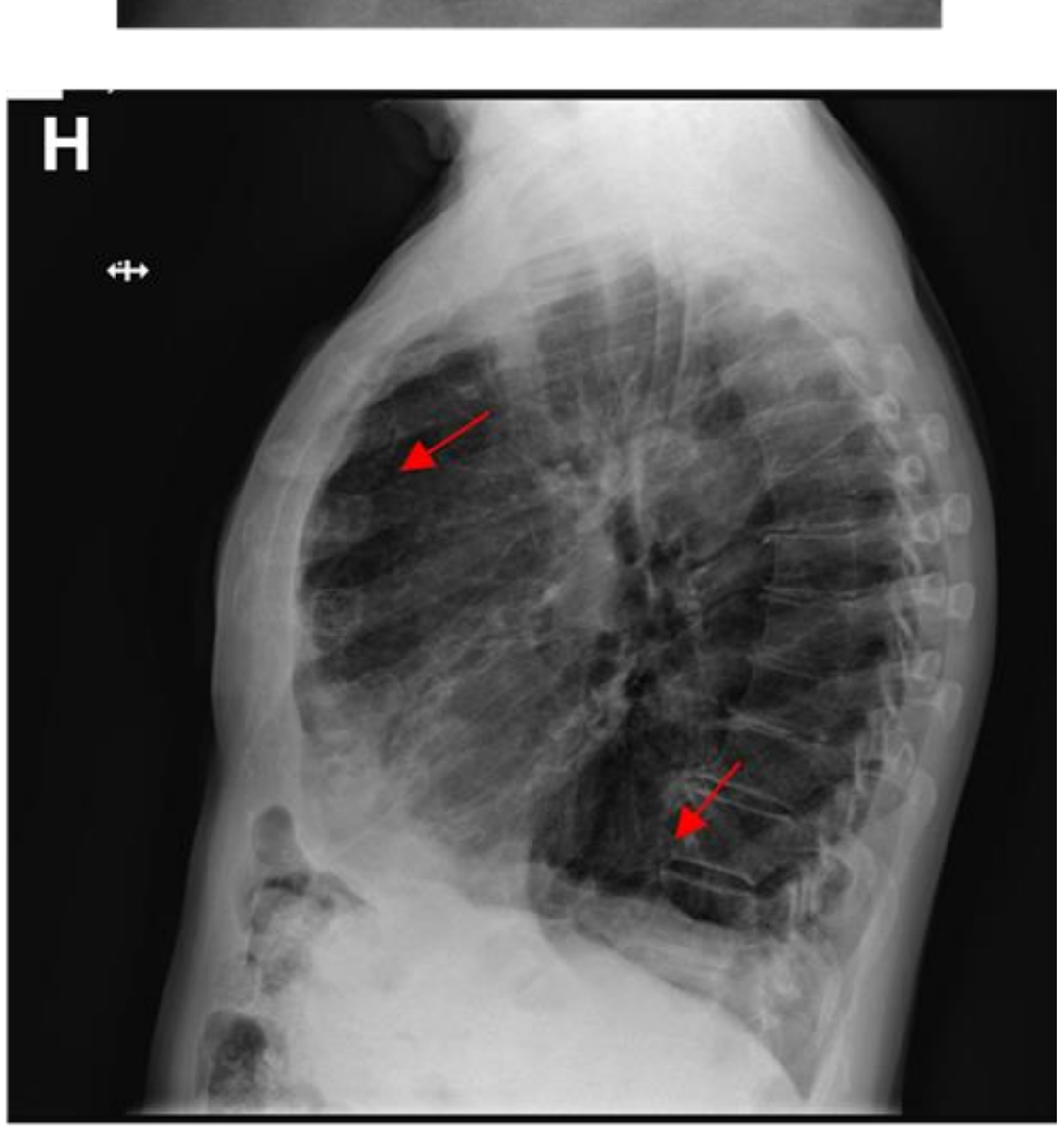


Figura 18. Radiografías de tórax PA con otros hallazgos. E) Material quirúrgico pulmonar. F) Material quirúrgico esternal. G y H) Hiperinsuflación pulmonar (aplanamiento diafragmático y ensanchamiento del espacio retroesternal y retrocardiaco).

CONCORDANCIA IA - RESIDENTE

La concordancia (valorada mediante el coeficiente Kappa de Cohen) entre el residente y la IA fue 0,3 (es decir, leve) para todas las variables salvo para derrame pleural, que fue 0,5 (moderada).

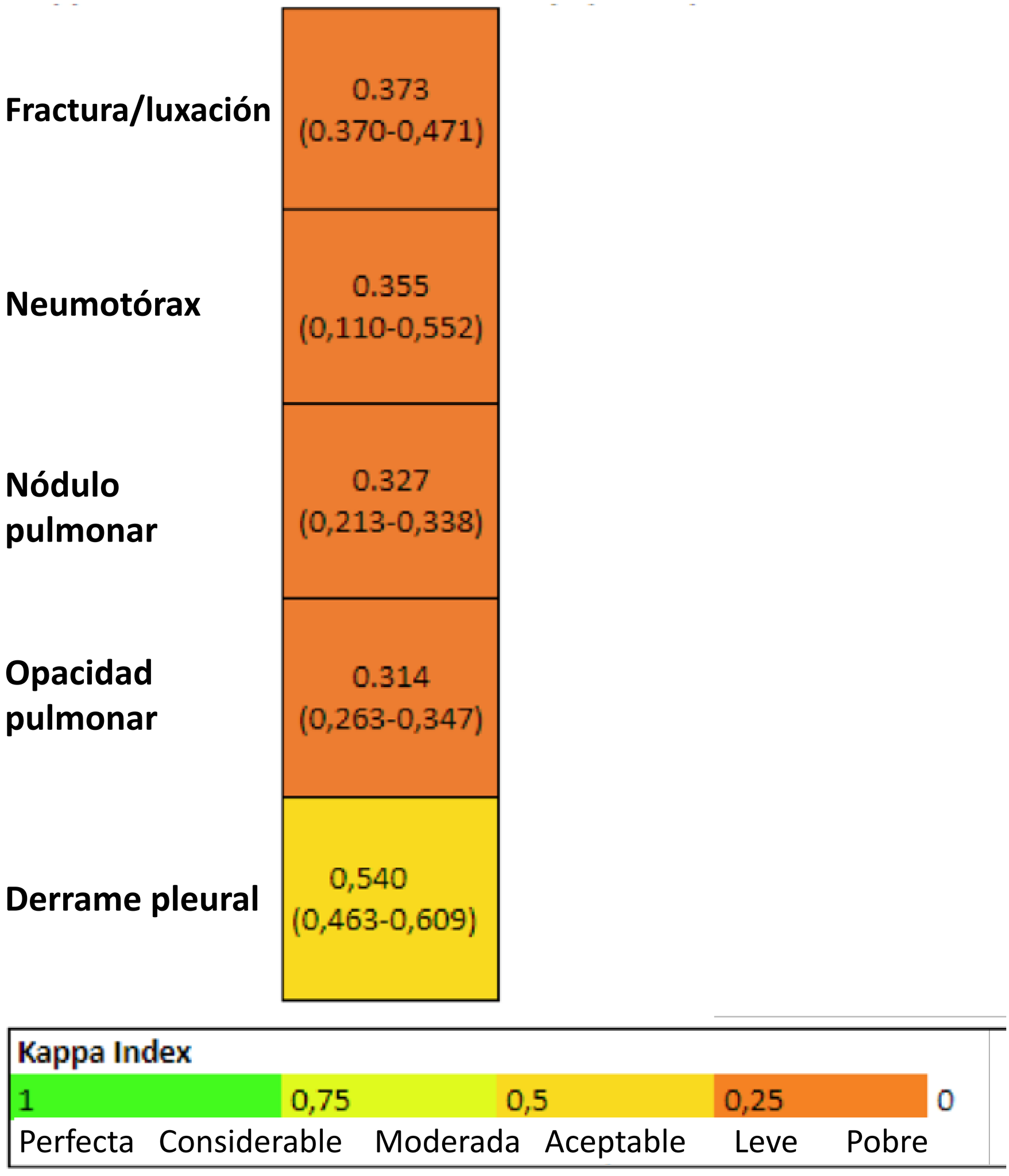


Figura 19. Concordancia IA . Residente (IC 95%)

DISCUSIÓN:

- La S de la IA y el residente para fractura y neumotórax fue alta (100%), moderada para opacidad pulmonar (IA 75,6%; residente 71,2%), razonable para derrame pleural (IA 59,7%; residente 67,1%) y baja para nódulo pulmonar por parte de la IA (33,3%) y moderada por parte del residente (75%).

- Nuestros resultados apoyan lo descrito previamente por otros autores en la literatura, como el estudio de Bennani et al (4), en el que analizan 500 radiografías y obtienen valores de S elevados para neumotórax (80%), consolidación (70,7%) y derrame pleural (78,8 %) y razonables para nódulo pulmonar (55,7%). Además, obtuvieron muy buenos resultados de AUC (0,94 para neumotórax, 0,96 derrame pleural, 0,93 consolidación y 0,74 nódulo pulmonar), coincidiendo así con los valores menos óptimos obtenidos en nuestro estudio para nódulo pulmonar.

- En el estudio de Kwang Nam et al (6), tras analizar 6006 radiografías en busca de nódulos/masas, consolidaciones y neumotórax obtuvieron una S global de 0.885

- La IA sería útil para detectar la mayoría de estos hallazgos, mientras que los resultados menos óptimos obtenidos en nuestro estudio (nódulo pulmonar) podrían deberse a las limitaciones del mismo. Sin embargo, la radiografía de tórax no constituye la técnica de cribado para la detección de nódulos pulmonares en pacientes con factores de riesgo, sino el TC de tórax de baja dosis.

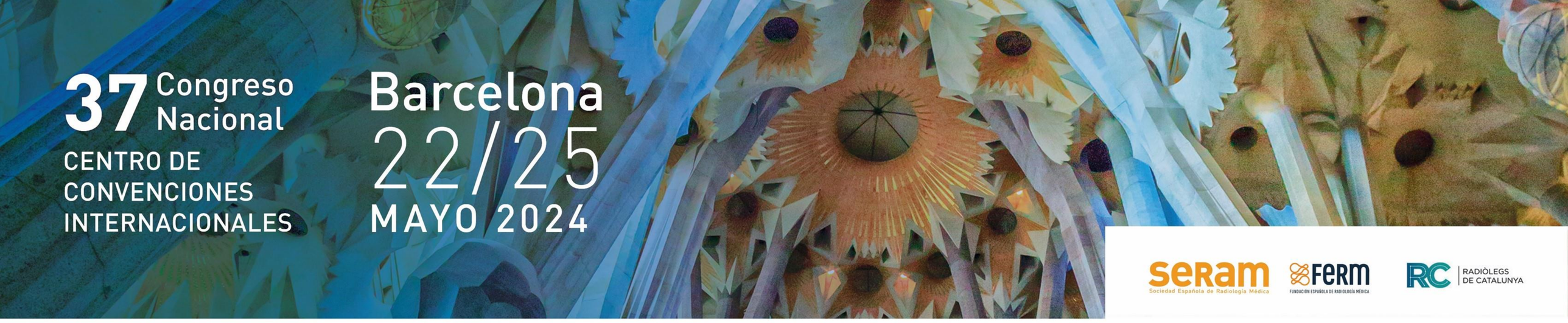
- El VPN tanto de la IA como del residente fue >95% y el AUC fue >0,8 (excepto para nódulo pulmonar), con un IC al 95% alto, lo que en términos estadísticos se traduce en que los resultados fueron robustos. De estos resultados se puede inferir que la IA sería una aplicación útil de screening en la Urgencia ya que clasificaría adecuadamente a los pacientes sin hallazgos patológicos, ahorrando tiempo al radiólogo y ayudando a los médicos de Urgencias.

- Dado que la mayoría de los casos catalogados como dudosos por la IA no resultaron ser positivos, sería interesante que los médicos de Urgencias consultaran al radiólogo los casos dudosos.

- Además, estos deberían tener en cuenta los hallazgos que la IA suele sobreestimar (fracturas –crónicas-), opacidades pulmonares en tercios medio e inferior pulmonares derechos y derrames pleurales leves y valorarlos con precaución.

LIMITACIONES

- Solo evaluamos una muestra pequeña de pacientes adultos durante un periodo de tiempo corto.
- Consideramos como Gold Standard a un radiólogo senior (que no deja de ser subjetivo) y no a una prueba de imagen confirmatoria más objetiva (por ejemplo TC torácico).
- Al habernos centrado únicamente en 5 hallazgos patológicos en vez de analizar las radiografías de forma completa en busca de cualquier hallazgo y al no disponer de la información clínica, el estudio no representa la práctica clínica diaria.
- Los intervalos de confianza para neumotórax y nódulo pulmonar fueron amplios (y menos precisos en términos estadísticos), puesto que dependen de la prevalencia y esta fue muy baja para estos ítems.
- El índice Kappa puede depender de la prevalencia de los hallazgos evaluados y de los casos dudosos excluidos para realizar el análisis estadístico, lo que explicaría sus valores tan bajos.



CONCLUSIONES:

- La IA sería útil para detectar la mayoría de los hallazgos estudiados (tal y como han concluido estudios previos), mientras que los resultados menos óptimos podrían deberse a las limitaciones del estudio referidas. Destaca su elevado VPN.

- Cuando la IA emite diagnósticos dudosos, en general suelen ser negativos por el GS, siendo especialmente destacable la valoración que hace de la vascularización normal de la base pulmonar derecha como dudosa opacidad pulmonar.

- Sería útil entrenar a la IA para detectar hallazgos adicionales en mediastino que podrían proporcionar información relevante en la práctica clínica.

- La IA ha venido para quedarse y bajo nuestro punto de vista, debemos verla como aliado, facilitando el trabajo diario y focalizando la valoración del radiólogo en los casos complejos.

REFERENCIAS:

1. Irmici G, Cè M, Caloro E, Khenkina N, Della Pepa G, Ascenti V, et al. Chest X-ray in Emergency Radiology: What Artificial Intelligence Applications Are Available? Diagnostics (Basel). 2023; 13(2): 216. doi:10.3390/diagnostics13020216

2. Li D, Pehrson LM, Lauridsen CA, Tøttrup L, Fraccaro M, Elliott D, Zając HD, Darkner S, Carlsen JF, Nielsen MB. The Added Effect of Artificial Intelligence on Physicians' Performance in Detecting Thoracic Pathologies on CT and Chest X-ray: A Systematic Review. . Diagnostics. 2021; 11(12):2206. doi:10.3390/diagnostics11122206

3. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and Validation of a Deep Learning–Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs. JAMA Netwwork open. 2019;2(3):e191095. doi:10.1001/jamanetworkopen.2019.1095

4. Bennani S, Regnard NE, Ventre J, Lassalle L, Nguyen T, Ducarouge A, et al. Using AI to Improve Radiologist Performance in Detection of Abnormalities on Chest Radiographs. Radiology. 2023 Dec;309(3):e230860. doi: 10.1148/radiol.230860. PMID: 38085079.

5. Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. Duron L, Ducarouge A, Gillibert A, Lainé J, Allouche C, Cherel N, et al. Radiology. 2021; 300(1): 120-129. doi: 10.1148/radiol.2021203886

6. Jin, K.N., Kim, E.Y., Kim, Y.J. et al. Diagnostic effect of artificial intelligence solution for referable thoracic abnormalities on chest radiography: a multicenter respiratory outpatient diagnostic cohort study. Eur Radiol 32, 3469–3479 (2022). https://doi.org/10.1007/s00330-021-08397-5