



**Universidad
Europea**

Máster Universitario en Bioinformática

**Recopilación de soluciones de inteligencia artificial con
Mercado CE en productos sanitarios de Clase IIa y IIb
bajo la Medical Device Regulation y evaluación del
asistente médico Sycai Medical**

Autor: Sergio García Dieste

Tutora: Júlia Rodríguez Comas

Curso 2022-23

RESUMEN

Objetivos: Los objetivos del presente trabajo son recopilar las herramientas de IA de clases IIa y IIb en el mercado que cumplen la regulación de dispositivos médicos MDR (Medical Device Regulation) y observar los tipos de ensayos a los que han sido sometidas; la posterior elaboración de una base de datos (BD) que contenga estas herramientas; y la comparación en términos de sensibilidad y especificidad con el asistente médico Sycal Medical.

Material y métodos: Se llevó a cabo una búsqueda de productos de IA para radiología diagnóstica utilizando la base de datos AI for Radiology de la plataforma Grand Challenge, aplicando filtros para dispositivos médicos de clase IIa o IIb. Además, se buscó incluir datos de ensayos clínicos que involucraran el uso de estas herramientas, utilizando ClinicalTrials.gov y EU Clinical Trials Register como bases de datos de referencia. También, se incluyeron pruebas de concepto con resultados de sensibilidad y especificidad.

Resultados: El resultado fue una base de datos que incluía soluciones de IA presentes en el mercado, su fabricante, el tipo de ensayo clínico, el diseño de intervención u observación o el responsable del estudio, entre otros. También se obtuvo una tabla con valores de especificidad y sensibilidad, y se compararon con el asistente médico Sycal Medical.

Conclusiones: Existen numerosos algoritmos de inteligencia artificial para la detección de patologías utilizando imágenes radiológicas, que han pasado por ensayos clínicos con diferentes diseños y que presentan valores de especificidad y sensibilidad muy diferentes. Aunque estas herramientas son una realidad en investigaciones clínicas, se necesitan más estudios y una mayor cantidad de datos de entrenamiento para mejorar la precisión y la eficacia de las herramientas de diagnóstico basadas en IA en la detección de patologías para su uso en la práctica clínica real.

Palabras clave: Inteligencia artificial, radiología, imagen médica, algoritmo, aprendizaje automático, deep learning, ensayos clínicos.

1. Introducción

La inteligencia artificial (IA) es un término amplio que abarca el uso de sistemas informáticos para imitar comportamientos inteligentes con poca intervención humana (Hamet & Tremblay, 2017). Los algoritmos de inteligencia artificial han mostrado en los últimos años un gran progreso en tareas relacionadas con el reconocimiento de imágenes. Basados en métodos como las redes neuronales convolucionales y los codificadores automáticos variacionales, estos algoritmos han demostrado ser de gran utilidad para múltiples aplicaciones en el campo de la medicina, impulsando su evolución de manera significativa.

Concretamente, los algoritmos de inteligencia artificial destacan en la identificación automática de patrones complejos en los datos de imágenes y ofrecen evaluaciones numéricas de las características radiográficas, en lugar de evaluaciones subjetivas. Es por ello por lo que han surgido numerosas herramientas basadas en IA con el objetivo de mejorar el diagnóstico a partir de imágenes radiológicas (Hosny, Parmar, Quackenbush, Schwartz, & Aerts, 2018).

La radiología es una de las ramas de la medicina que ha evolucionado más rápidamente en cuanto a la adopción de tecnología innovadora. Desde sus inicios, los radiólogos han sido pioneros en la implementación de nuevas herramientas y técnicas que les permiten obtener imágenes de alta calidad para el diagnóstico y tratamiento de enfermedades. Con el tiempo, la radiología ha experimentado un flujo constante de innovaciones tecnológicas, que se han ido incorporando a su práctica habitual en forma de nuevas modalidades de imágenes, como la tomografía por emisión de positrones (PET) y la tomografía por emisión de fotón único (SPECT), la resonancia magnética (RM), la radiografía digital, la tomografía computarizada (TC), los ultrasonidos y las imágenes ópticas. Estas herramientas han evolucionado en gran medida a partir de nuevos descubrimientos en física e ingeniería, así como de otros desarrollos técnicos (Gore, 2020).

El uso de estas nuevas modalidades de imágenes ha permitido a los radiólogos obtener imágenes más detalladas y precisas, lo que ha mejorado significativamente la capacidad de diagnóstico y tratamiento de enfermedades. En la actualidad, la radiología sigue siendo una de las ramas médicas más innovadoras en cuanto a la adopción de tecnología avanzada. La integración de la inteligencia artificial (IA) en la práctica radiológica se perfila como una nueva área de innovación, que promete mejorar aún más la precisión y eficacia del diagnóstico y tratamiento de enfermedades.

Esto se debe a que el uso de la IA permite procesar grandes cantidades de información y datos médicos, lo que les permite detectar patrones y anomalías en las imágenes de manera más rápida y precisa (Gore, 2020; Hosny et al., 2018).

De esta forma, la inteligencia artificial se está convirtiendo en un componente esencial en muchas aplicaciones dentro del cuidado de la salud. Entre ellas, se incluyen el descubrimiento de fármacos, el seguimiento remoto de pacientes, el diagnóstico médico y las imágenes, la gestión de riesgos, los wearables, los asistentes virtuales y la gestión hospitalaria. Además, la IA también se espera que beneficie el análisis en campos que manejan enormes cantidades de datos, como la secuenciación de ADN y ARN.

En particular, la radiología puede beneficiarse de la precisión y reproducibilidad mejoradas que se pueden lograr con la integración de la IA en el flujo de trabajo clínico como una herramienta para ayudar a los médicos. Dado que los datos de imágenes se recopilan durante la práctica clínica habitual, grandes conjuntos de datos están fácilmente disponibles, ofreciendo así un recurso increíblemente rico para el descubrimiento científico y médico (Gore, 2020).

Un ejemplo de algoritmo de IA aplicado al diagnóstico es Sycai Medical, un asistente médico desarrollado por Sycai Technologies SL que pretende mejorar la eficacia de los radiólogos identificando y clasificando lesiones quísticas pancreáticas (Duh et al., 2023).

En el presente trabajo, nos centraremos en los productos de IA para imagen médica con marcado CE (del francés Conformité Européenne) de clase IIa o IIb según el Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo de 5 de abril de 2017, disponibles en el mercado, así como en su evaluación frente al asistente Sycai Medical.

Quistes pancreáticos

Los quistes pancreáticos son muy comunes en la población, tanto que, en pacientes mayores de 70 años, las imágenes de resonancia magnética abdominal mostraron una incidencia del 40% de quistes pancreáticos incidentales. Los quistes pancreáticos suelen ser detectados en imágenes abdominales realizadas por otras razones, y su prevalencia en una población asintomática aumenta con la edad (Elta, Enestvedt, Sauer, & Lennon, 2018).

Existen varios tipos de quistes pancreáticos, entre ellos los pseudoquistes (PCYST), sin potencial maligno y que no requieren vigilancia; las neoplasias mucinosas papilares intraductales (IPMN), siendo las IPMN del conducto principal las menos comunes pero las que presentan mayor riesgo de malignidad, ya que el 38–68% de ellas presentan displasia de alto grado o cáncer de páncreas en muestras reseadas (Stark, Donahue, Reber, & Joe Hines, 2016); las neoplasias quísticas mucinosas (MCN), de las que alrededor del 10% contienen displasia o cáncer de páncreas; y los cistoadenomas serosos, con un riesgo de malignidad cercano a cero.

Actualmente, la única opción de tratamiento para los quistes pancreáticos es la cirugía, que tiene altos riesgos y mortalidad. Debido a esto, las recomendaciones sobre su manejo y seguimiento son contradictorias debido a la falta de estudios de alta calidad y el pequeño riesgo de transformación maligna (Elta et al., 2018).

Aunque el riesgo de transformación maligna es bajo, la tasa de conversión a cáncer invasivo es del 0,24% por año. Sin embargo, las tasas de cáncer informadas en estudios retrospectivos de quistes resecaos quirúrgicamente son más altas, con una proporción agrupada de quistes con adenocarcinoma invasivo del 15% (Scheiman, Hwang, & Moayyedi, 2015)

El coste del análisis y seguimiento de los quistes es elevado y el beneficio en términos de prevención del cáncer no está claro. No se han realizado análisis de coste-efectividad sobre el seguimiento de los quistes pancreáticos incidentales. Además, según la reciente revisión de Scheiman et al., la tasa de mortalidad por resección pancreática para quistes pancreáticos es del 2,1%, con una tasa de morbilidad del 30%.

En cuanto al diagnóstico, la resonancia magnética nuclear (RMN) o la colangiopancreatografía por resonancia magnética (CPRM) son las pruebas de elección debido a que no son invasivas, a la ausencia de radiación y a la gran precisión que ofrecen en la evaluación de la comunicación entre el conducto pancreático principal y el quiste (característica de las IPMNs de rama lateral). La tomografía computarizada (TC) o la ultrasonografía endoscópica (USE) son excelentes alternativas en pacientes que no pueden someterse a resonancia magnética (Elta et al., 2018).

2. Hipótesis y Objetivos

La hipótesis de la que se parte es que existen herramientas de IA en radiología que han llegado a un nivel de desempeño suficiente como para salir al mercado, con características y prestaciones que permiten su implementación en la práctica clínica.

El objetivo del trabajo es recopilar las herramientas de IA de clase II en el mercado que cumplen la regulación de dispositivos médicos MDR (Medical Device Regulation) y observar los tipos de ensayos a los que han sido sometidas; la posterior elaboración de una base de datos (BD) que contenga estas herramientas; y la comparación en términos de sensibilidad y especificidad con el asistente médico Sycal Medical.

3. Metodología

En primer lugar, se llevó a cabo una búsqueda de productos de IA para radiología diagnóstica de la clase II que cumplieran la regulación MDR, establecida en el *Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios*. Para ello, se utilizó la base de datos *AI for Radiology* de la plataforma Grand Challenge (“AI for radiology,” n.d.) al aplicar los filtros “CE Class = IIa” y “CE Class = IIb”, esto es, dispositivos médicos catalogados por la Comisión Europea como de clase IIa o IIb y que, por tanto, “requieren pruebas de producto, implementación del sistema de calidad (ISO 13485), preparación de expediente técnico, auditoría de organismos notificados y certificación” (“CE Certification - Class IIa Medical Device CE Marking,” n.d.; “CE Certification - Class IIb Medical Device CE Marking,” n.d.).

Asimismo, se pretendía incluir datos de ensayos clínicos (si los había) que involucraran el uso de estas herramientas. Para ello, se consultó la principal base de datos de

ensayos clínicos de Estados Unidos, *ClinicalTrials.gov* (“Home - ClinicalTrials.gov,” n.d.), así como su equivalente en la Unión Europea, *EU Clinical Trials Register* (“EU Clinical Trials Register” n.d.).

La búsqueda en la BD estadounidense consistió en buscar utilizando el nombre del producto (Trade_name) en la casilla “Other terms”. Si la búsqueda no arrojaba resultados, se repitió la búsqueda con el nombre de la empresa que desarrolló la IA (Manufacturer) para encontrar los ensayos en los que la herramienta aparecía con un nombre diferente al recogido en *AI for Radiology*. Este nombre tenía que referirse inequívocamente a una de las soluciones de AI recogidas en *AI for Radiology*.

En cambio, si los resultados eran muy numerosos (más de 10), se afinó la búsqueda incluyendo el nombre del producto y el nombre de la empresa, separados por la partícula “AND” (Trade_name AND Manufacturer). En algunos estudios, la herramienta de IA era un complemento a un diseño que podía incluir fármacos u otras intervenciones.

Una vez obtenida la pantalla de resultados, se aplicó el criterio de que el nombre del producto tenía que aparecer en la descripción del ensayo clínico para poder seleccionarlo. En la búsqueda, se descartaron algunos resultados, como el estudio “NCT05489471”, con la empresa Lunit Inc. como colaborador y presente en algunas entradas de la base de datos, pero donde no se pudo identificar qué IA de su catálogo se usó en el mismo; o el estudio “NCT05716958”, donde se utilizó la IA “Lunit INSIGHT DBT”, no registrada en la base de datos de la plataforma *Grand Challenge*. Del mismo modo, se repitió la búsqueda en la BD europea, que no arrojó ningún resultado.

Posteriormente, se generó una base de datos utilizando MySQL, un sistema de gestión de bases de datos relacional desarrollado por Oracle Corporation. En concreto, se hizo uso de la herramienta para escritorio *MySQL Workbench 8.0 CE*, donde se creó y editó la base de datos, que se nombró *AI_Medical_Imaging*.

Recogida y entrada de datos

La recogida y entrada de datos en la base de datos AI_Medical_Imaging fue manual. Se utilizaron las siguientes bases de datos:

1. *AI for Radiology*, Grand Challenge: <https://grand-challenge.org/aiforradiology/>
2. *ClinicalTrials.gov*, la principal base de datos pública de ensayos clínicos de Estados Unidos: <https://clinicaltrials.gov/>
3. *EU Clinical Trials Register*, la principal base de datos pública de ensayos clínicos de la Unión Europea: <https://www.clinicaltrialsregister.eu/ctr-search/search>

Criterios de inclusión y exclusión

Se incluyeron soluciones de inteligencia artificial de imagen médica recogidos en la base de datos AI for Radiology de la plataforma Grand Challenge catalogadas como CE Class IIa o IIb, y que cumplan el *Reglamento (UE) 2017/745 del Parlamento Europeo y del Consejo, de 5 de abril de 2017, sobre los productos sanitarios*.

Para seleccionar los algoritmos que cumplen con el Reglamento europeo sobre los productos sanitarios, se añadió el filtro “CE = MDR”. Así, se encontraron 30 resultados para la búsqueda “CE Class = IIa” y 14 resultados para “CE Class = IIb”.

A continuación, se muestra la selección de filtros utilizada en la base de datos *AI for Radiology*:

Products

Find the artificial intelligence based software for radiology that you are looking for. All products listed are available for the European market (CE marked).

Subspeciality: Modality: CE: ⁱ CE class: ⁱ FDA class: ⁱ Sort by:

All ▾ All ▾ MDR ▾ Class IIa ▾ All ▾ last modified ▾

Products

Find the artificial intelligence based software for radiology that you are looking for.
All products listed are available for the European market (CE marked).

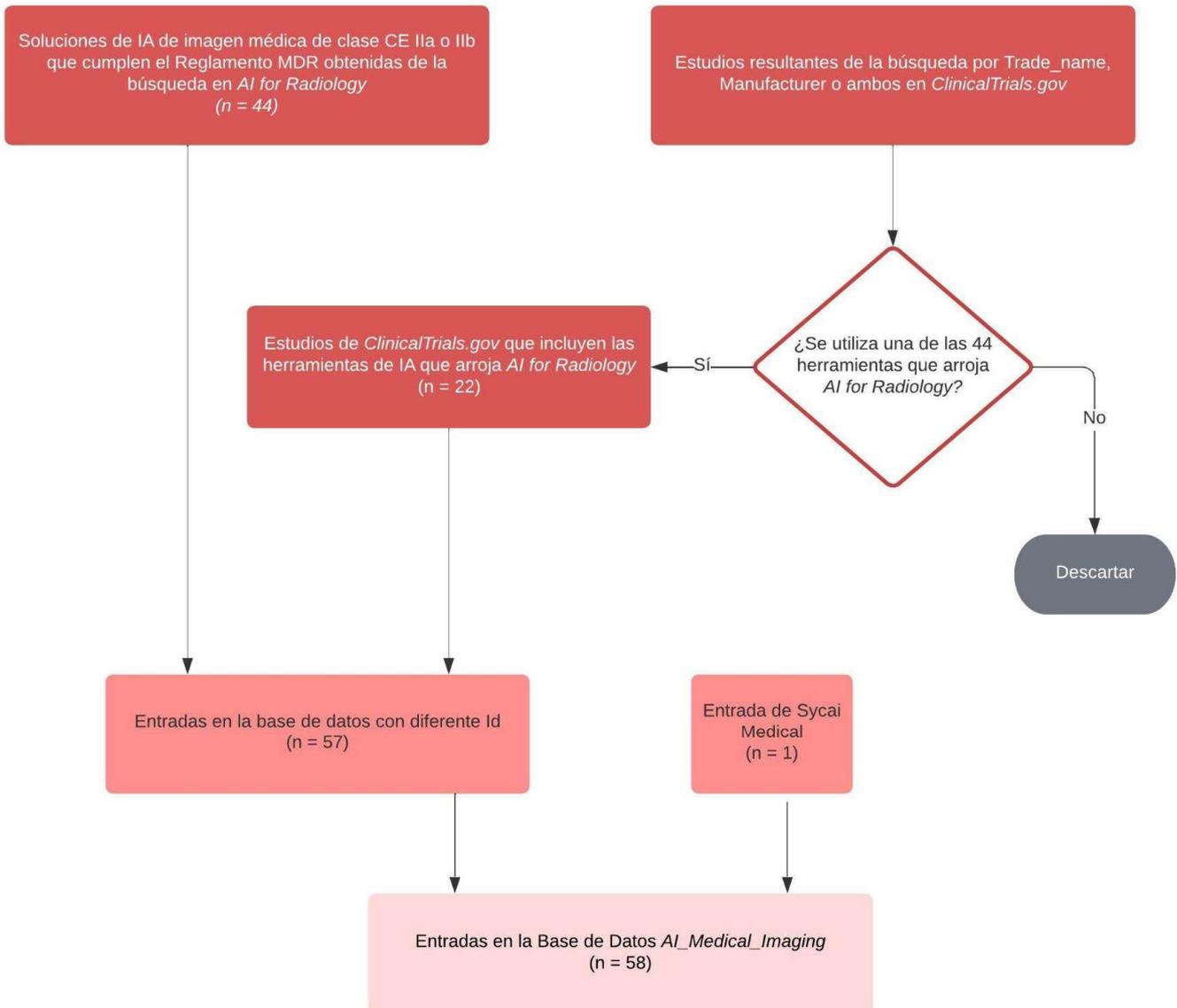
Subspeciality: Modality: CE: ⁱ CE class: ⁱ FDA class: ⁱ Sort by:

All ▾ All ▾ MDR ▾ Class IIb ▾ All ▾ last modified ▾

Por tanto, la base de datos cuenta con 45 soluciones de inteligencia artificial, las 44 recogidas en Grand Challenge más Sycal Medical.

Al realizar la búsqueda en *ClinicalTrials.gov* y *EU Clinical Trials Register*, y tras aplicar los criterios expuestos anteriormente, se extrajeron 22 estudios de la base de datos estadounidense y 0 estudios de la base de datos europea.

Diagrama de flujo para la inclusión de estudios y herramientas



Definición de variables

La base de datos cuenta con 23 variables, definidas como se explica a continuación:

Nombre de variable	Tipo y atributos	Null	Explicación	Fuente
Id	int AUTO_INCREMENT	No (Primary Key)	Número de identificación único	Propia
Trade_name	varchar(255)	Sí	Nombre comercial del producto	Grand Challenge
Manufacturer	varchar(255)	Sí	Empresa desarrolladora	Grand Challenge
FDA_Class	varchar(255)	Sí	Clasificación del Producto Sanitario según la Food and Drug Administration	Grand Challenge
Anatomic_site	varchar(255)	Sí	Sitio anatómico de las imágenes que analiza la herramienta	Grand Challenge
Modality	varchar(255)	Sí	Tipo de imagen médica que utiliza la herramienta	Grand Challenge
CE_Class	varchar(255)	Sí	Clasificación del Producto Sanitario según la Food and Drug Administration	Grand Challenge

Number_of_participants	int	Sí	Número de participantes del estudio	ClinicalTrials.gov
Male_percentage	float	Sí	Porcentaje de participantes hombres (0 cuando solo son las mujeres son elegibles)	ClinicalTrials.gov
Time_perspective	varchar(255)	Sí	Tipo de estudio según la perspectiva temporal (prospectivo o retrospectivo)	ClinicalTrials.gov
Center_setting	varchar(255)	Sí	Tipo de estudio según el nº de centros participantes en el estudio (unicéntrico o multicéntrico)	ClinicalTrials.gov
Sample_country	varchar(255)	Sí	País o países de la muestra participante	ClinicalTrials.gov
Ages_eligible	varchar(255)	Sí	Rango de edades elegibles para participar en el estudio	ClinicalTrials.gov
Indication	varchar(511)	Sí	Indicación o indicaciones que ayuda a diagnosticar la herramienta	Grand Challenge

Recruitment_status	varchar(255)	Sí	Estado de reclutamiento	ClinicalTrials.gov
Main_contact	varchar(255)	Sí	Contacto principal	ClinicalTrials.gov
Study_type	varchar(255)	Sí	Tipo de estudio según si existe intervención (intervencional) o no (observacional)	ClinicalTrials.gov
Observational_model	varchar(255)	Sí	Modelo de observación (sólo para estudios observacionales)	ClinicalTrials.gov
Intervention_model	varchar(255)	Sí	Modelo de intervención (sólo para estudios intervencionales)	ClinicalTrials.gov
Identifier	varchar(255)	Sí	Identificador de ClinicalTrials.gov (Formato NCT)	ClinicalTrials.gov
Design_country	varchar(255)	Sí	País de la institución donde trabaja el investigador principal o, en su defecto, país del responsable del estudio	ClinicalTrials.gov
Responsible_party	varchar(255)	Sí	Institución o persona responsable del estudio	ClinicalTrials.gov

Official_title	varchar(511)	Sí	Título oficial	ClinicalTrials.gov
----------------	--------------	----	----------------	--------------------

En cuanto al criterio para seleccionar los datos, las columnas *Trade_name*, *Manufacturer*, *FDA_Class*, *Anatomic_site*, *Modality*, *CE_Class* e *Indication* se completaron según la información recogida en *AI for Radiology*, con los cambios de nomenclatura que se muestran a continuación:

Nombre en <i>AI for Radiology</i>	Nombre en <i>AI_Medical_Imaging</i>
Product name	Trade_name
Company	Manufacturer
FDA	FDA_Class
Subspeciality*	Anatomic_site
Modality	Modality
CE	CE_Class
Disease targeted	Indication

*En la entrada de datos, se sustituyó Cardiac por Heart y Neuro por Brain, ya que todas las soluciones para el sistema cardiovascular leían imágenes del corazón, y para el sistema neurológico, imágenes del cerebro. Esta nomenclatura se ajusta mejor al nombre de la variable "*Anatomic_site*".

Por otra parte, las columnas *Number_of_participants*, *Male_percentage*, *Time_perspective*, *Center_setting*, *Sample_country*, *Ages_eligible*, *Recruitment_status*, *Main_contact*, *Study_type*, *Observational_model*, *Intervention_model*, *Identifier*, *Design_country*, *Responsible_party* y *Official_title* se completaron con datos de *ClinicalTrials.gov*, con cambios en el título de columna como se muestra debajo:

Nombre en <i>ClinicalTrials.gov</i>	Nombre en <i>AI_Medical_Imaging</i>
Estimated Enrollment / Actual Enrollment [†]	Number_of_participants
**	Male_percentage
Time Perspective ^{***}	Time_perspective
**	Center_setting
**	Sample_country
Ages Eligible for Study	Ages_eligible
Recruitment Status	Recruitment_status
Contact#	Main_contact
Study Type ^{***}	Study_type
Observational Model	Observational_model
Intervention Model	Intervention_model
ClinicalTrials.gov Identifier	Identifier
**	Design_country
Responsible Party	Responsible_party
Official Title	Official_title

[†]Se agrupó en una sola variable (Number_of_participants) las variables Estimated Enrollment y Actual Enrollment. De esta forma, para estudios con estado de reclutamiento “*completed*”, el número corresponde con los participantes reales, mientras que para cualquier otro estado, corresponde con el estimado de participantes que tendrá el estudio.

**Estos datos no aparecían explícitamente en la ficha, pero se pudieron deducir o establecer como se indica a continuación:

1. Male_percentage se estableció en 0 cuando el sexo elegible era únicamente femenino (indicado en el campo “*Sexes Eligible for Study*”), o se calculó dividiendo el número de participantes hombres entre el total y multiplicando por 100 en aquellos en los que aparecía el dato.
2. Center_setting no aparecía explícitamente, pero se buscó en el título, en la descripción o se infirió del número de localizaciones que reclutaban participantes.
3. Sample_country es una variable no existente en la ficha del estudio, y corresponde al país o países donde se situaban los centros reclutadores.
4. Design_country es una variable no existente en la ficha del estudio, y corresponde al país de la institución donde trabaja el investigador principal o, cuando no aparecía, al país del responsable del estudio (Responsible_party).

***En algunas ocasiones, no aparecían los valores para estas variables en la ficha del estudio y se buscaron en el título, en la descripción o se infirió al leer el diseño.

#Contact se seleccionó como el primer contacto en la sección “*Contacts*” y, cuando no aparecía, se seleccionó el investigador principal (Principal investigator).

Adicionalmente, se buscó recopilar los datos de sensibilidad y especificidad de las herramientas contenidas en la base de datos. Para ello, se recurrió al apartado “*Evidence*” en la ficha de cada software en *Grand Challenge*, donde aparecen los papers relacionados con los mismos. Asimismo, se realizaron búsquedas en la base de datos especializada en ciencias de la salud *PubMed* donde, en caso de obtenerse más de 5 resultados, se reformuló la búsqueda para refinarla, añadiendo, además del nombre de la herramienta, el término *artificial intelligence* o el nombre de la empresa fabricante.

Con estos datos de sensibilidad y especificidad, se realizó una tabla donde aparecía el nombre del software, la especificidad, la sensibilidad y la fuente (nombre del paper de donde se extrajeron los datos).

4. Resultados

El resultado obtenido es una base de datos con 58 filas y 23 columnas, lo que supone 1334 ítems, incluyendo celdas nulas.

Utilizando el comando `SHOW COLUMNS FROM AI_Medical_Imaging;` en MySQL Workbench, obtenemos la siguiente definición de las variables:

Field	Type	Null	Key	Default	Extra
Id	int	NO	PRI	NULL	
Trade_name	varchar(255)	YES		NULL	
Manufacturer	varchar(255)	YES		NULL	
FDA_Class	varchar(255)	YES		NULL	
Anatomic_site	varchar(255)	YES		NULL	
Modality	varchar(255)	YES		NULL	
CE_Class	varchar(255)	YES		NULL	
Number_of_participants	int	YES		NULL	
Male_percentage	float	YES		NULL	
Time_perspective	varchar(255)	YES		NULL	
Center_setting	varchar(255)	YES		NULL	
Sample_country	varchar(255)	YES		NULL	
Ages_eligible	varchar(255)	YES		NULL	
Indication	varchar(511)	YES		NULL	
Recruitment_status	varchar(255)	YES		NULL	
Main_contact	varchar(255)	YES		NULL	
Study_type	varchar(255)	YES		NULL	
Observational_model	varchar(255)	YES		NULL	
Intervention_model	varchar(255)	YES		NULL	
Identifier	varchar(255)	YES		NULL	
Design_country	varchar(255)	YES		NULL	
Responsible_party	varchar(255)	YES		NULL	
Official_title	varchar(511)	YES		NULL	

Como se puede observar, 18 variables admiten datos de tipo cadena de longitud variable (varchar) de hasta 256 caracteres; 2 admiten varchar de hasta 512 caracteres (así definidas por el tipo de datos que alberga); otras 2 admiten números enteros (int); y la variable Male_percentage admite datos de tipo float, es decir, números con parte decimal con una cierta precisión que se consideró suficiente.

De igual modo, se observa que la variable Id está definida como “primary key” o clave primaria, es decir, no admite dos valores iguales, por lo que identifica inequívocamente cada fila de una tabla. El valor por defecto para todas las variables es NULL.

En cuanto a los tipos de estudio, se incluyeron 11 estudios intervencionales y 12 estudios observacionales (incluyendo el de Sycai Medical). Para el conteo, se utilizó el siguiente comando (nótese que se eliminaron las entradas nulas, correspondientes a herramientas de IA sin estudio asociado):

```
SELECT Study_type, COUNT(*) FROM AI_Medical_Imaging WHERE Study_type IS NOT NULL GROUP BY Study_type;
```

Y se obtuvo el siguiente resultado:

Study_type	COUNT(*)
Interventional	11
Observational	12

En cuanto al número total de participantes, la suma de ellos asciende a 309046 participantes que participaron o que se estima que participarán en los 23 estudios recogidos en la base de datos. El input y output se muestran a continuación:

```
SELECT SUM(Number_of_participants) FROM AI_Medical_Imaging;
```

SUM(Number_of_participants)
309046

Un dato de mayor interés es calcular el número de estudios por estado de reclutamiento y, después, sumar los participantes de los ensayos con estatus “completo”:

```
SELECT Recruitment_status, COUNT(*) FROM AI_Medical_Imaging WHERE Recruitment_status IS NOT NULL GROUP BY Recruitment_status;
```

Recruitment_status	COUNT(*)
Recruiting	7
Unknown	5
Completed	9
Active, not recruiting	2

```
SELECT SUM(Number_of_participants) FROM AI_Medical_Imaging WHERE Recruitment_status = "Completed";
```

SUM(Number_of_participants)
20888

En total, son 20888 los participantes de estudios que han terminado la fase de reclutamiento.

Con relación al número de centros de reclutamiento, los resultados son los siguientes:

```
SELECT Center_setting, COUNT(*) FROM AI_Medical_Imaging WHERE Center_setting IS NOT NULL GROUP BY Center_setting;
```

Center_setting	COUNT(*)
Multicenter	9
Single center	14

Ahora, se agrupan las herramientas de IA por fabricante, obteniéndose los siguientes resultados:

```
SELECT Manufacturer, COUNT(*) FROM AI_Medical_Imaging GROUP BY Manufacturer;
```

Manufacturer	COUNT(*)
Koios Medical, Inc.	1
Quantib	3
Us2.ai	1
Siemens Healthineers	4
Quibim	1
annalise.ai	1
Aidence	1
Vara (MX Healthcare GmbH)	1
Thirona	10
Oxipit	1
Optellum	1
SenseTime	2
BrainScan Inc.	1
Lunit	5
contextflow	1
AZmed	1
Arterys	5
MRIguidance	1
AI4MedImaging	1
Radiobotics	2
Combinostics	1
Riverain Technologies	7
Ligence	1
BrainTale	1
ScreenPoint Medical	3
Sycal Technologies	1

De igual forma, al clasificar las herramientas por sitio anatómico, el resultado es el siguiente:

```
SELECT Anatomic_site, COUNT(*) FROM AI_Medical_Imaging WHERE Anatomic_site IS NOT NULL GROUP BY Anatomic_site;
```

Anatomic_site	COUNT(*)
Breast, Thyroid	1
Abdomen	4
Heart	4
Chest	29
Breast	8
Brain	7
MSK	4
MSK, Chest	1

MSK: sistema musculoesquelético

Por su parte, la clasificación según la perspectiva temporal del estudio es la siguiente:

```
SELECT Time_perspective, COUNT(*) FROM AI_Medical_Imaging WHERE Time_perspective IS NOT NULL GROUP BY Time_perspective;
```

Time_perspective	COUNT(*)
Prospective	18
Retrospective	5

Ahora, clasificando por modalidad de imagen médica, se obtiene lo siguiente:

```
SELECT Modality, COUNT(*) FROM AI_Medical_Imaging WHERE Modality IS NOT NULL GROUP BY Modality;
```

Modality	COUNT(*)
Ultrasound	3
MR	11
X-ray	23
CT	12
Mammography	8
CT, MR	1

En la siguiente tabla se recogen datos de la sensibilidad y especificidad de las diferentes herramientas:

Software	Especificidad	Sensibilidad	Fuente
Koios DS	50%	98%	Should We Ignore, Follow, or Biopsy? Impact of Artificial Intelligence Decision Support on Breast Ultrasound Lesion Assessment
Quantib Prostate	tumor volumes > 0.03 cc: 43% tumor volume > 0.1 cc images: 52% tumor volumes > 0.5 cc: 74%	tumor volumes > 0.03 cc: 82% tumor volume > 0.1 cc: 85% tumor volumes > 0.5 cc: 94%	Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI
Annalise Enterprise CXR	Any pneumothorax: 92.0% Tension pneumothorax: 95.3%	Any pneumothorax: 94.3% Tension pneumothorax: 94.5%	Evaluation of an Artificial Intelligence Model for Detection of Pneumothorax and Tension Pneumothorax in Chest Radiographs

Veye Lung Nodules	96%-100%	85-92%	Higher agreement between readers with deep learning CAD software for reporting pulmonary nodules on CT
Veye Lung Nodules	Determination of the composition of solid nodules: 68.4%.	Determination of the composition of solid nodules: 98.8%	Validation of a deep learning computer aided system for CT based lung nodule detection, classification, and growth rate estimation in a routine clinical population
Vara	89.5% on internal-test data 91.3% on external-test data	84.2% on internal-test data 84.6% on external-test data	Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis
CAD4TB	76.0%	91.5%	Comparing different versions of computer-aided detection products when reading chest X-rays for tuberculosis

CAD4TB	72.9%	90%	Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms
ChestLink	36.4 %	99.8 %	Using artificial intelligence to detect chest X-rays with no significant findings in a primary health care setting in Oulu, Finland
ChestLink	Normal posteroanterior chest radiographs: 28.0%	Abnormal radiographs: 99.1% Critical radiographs: 99.8%	Autonomous Chest Radiograph Reporting Using AI: Estimation of Clinical Impact
Lunit INSIGHT CXR	94.2%	35.3%	Multicentre external validation of a commercial artificial intelligence software to analyse chest radiographs in health screening environments with

			low disease prevalence
Lunit INSIGHT MMG	88.5%-90.2%	75.6%-76.1%	Applying Data-driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study
Rayvolve	Detection approach: 91.2% Enumeration and localization approaches: 88.8%	Detection approach: 95.7% Enumeration and localization approaches: 94.1%	External validation of a commercially available deep learning algorithm for fracture detection in children
ClearRead CT - Vessel Suppress	84.4%	72.5%	Computer-Aided Detection of Lung Nodules on CT With a Computerized Pulmonary Vessel Suppressed Function

Quantib ND	Depending on the disease: dementia, frontotemporal dementia, Alzheimer's disease: 60-90%	Depending on the disease: dementia, frontotemporal dementia, Alzheimer's disease): 50-97,5%	Comparing two artificial intelligence software packages for normative brain volumetry in memory clinic imaging
LungQ	66.7%	83.3%	Predicting Lung Volume Reduction after Endobronchial Valve Therapy Is Maximized Using a Combination of Diagnostic Tools
Transpara	97.6%	68.3%	Artificial intelligence in BreastScreen Norway: a retrospective analysis of a cancer-enriched sample including 1254 breast cancer cases
Sycal Medical	Detection of PCL: 81.8%	Detection of PCL: 93.1%	Deep Learning to Detect Pancreatic Cystic Lesions on Abdominal Computed Tomography Scans: Development and Validation Study

5. Discusión

De las 45 herramientas incluidas en la base de datos, solo 9 presentaban ensayos clínicos registrados en *ClinicalTrials.gov*, un total de 22 ensayos. La herramienta con más ensayos fue CAD4TB, con 9 ensayos que presentan diferentes estados de reclutamiento. El ensayo de Sycai se añadió a pesar de no estar registrado en las bases de datos seleccionadas para la búsqueda. Este ensayo fue retrospectivo, observacional y de un solo centro.

De los 23 ensayos en total, 11 fueron de intervención y 12 observacionales; 18 fueron prospectivos y 5 retrospectivos; y 14 fueron de centro único frente a 9 multicentro. Esto último evidencia que todavía se necesitan más ensayos que incluyan un mayor número de participantes y de centros de reclutamiento con el objetivo de evitar sesgos en los procedimientos.

En cuanto a la tabla, de valores de especificidad y sensibilidad, se puede observar que las herramientas presentan valores muy diversos: en el caso de la especificidad, la variación es del 28% al 100%, mientras que en términos de sensibilidad, varían del 35,3% al 99,8%.

La especificidad en el campo del diagnóstico se define como el número de verdaderos negativos (negativos detectados y que coinciden con el método de referencia) dividido entre el de todos los negativos (es decir, verdaderos negativos más falsos positivos, siendo estos últimos negativos según la referencia pero que positivos según nuestro método diagnóstico). La fórmula es la siguiente:

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

Por otra parte, cuando se habla de sensibilidad, se hace referencia al número de verdaderos positivos entre todos los positivos (los verdaderos positivos y los falsos negativos) (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008).

Por tanto, la fórmula de la sensibilidad es:

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

Así pues, los valores de sensibilidad y especificidad deben contextualizarse. Son difícilmente comparables ya que, probablemente, fueron tomados en distintos momentos del proceso de desarrollo de las herramientas (desde las primeras pruebas de concepto hasta su comercialización); detectan o clasifican tumores o patologías en órganos muy diferentes; y utilizan imágenes generadas por tecnologías distintas, entre ellas, tomografía computarizada, resonancia magnética nuclear, ultrasonidos o radiografía (mamografía).

Por su parte, el asistente médico Sycal Medical arroja, en el último paper publicado (Duh et al., 2023), unos datos de especificidad y sensibilidad del 81,8% y el 93,1% respectivamente. En él, se incluyeron 335 TACs, 297 de los cuales contenían una lesión quística pancreática. Por otra parte, los autores aclaran que cuenta con algunas limitaciones, como que el algoritmo, cuyo objetivo era detectar lesiones quísticas pancreáticas, fue entrenado con datos de pacientes provenientes de un solo hospital.

En esta línea, además, los datos de entrenamiento se dividieron en 2 grandes grupos: el primero incluía casos de los quistes más peligrosos y con mayor potencial de desarrollar malignidad (IPMN y MCN) y, el segundo, aquellos con un bajo potencial de generar cáncer (PCYST y SCA). Esto fue debido a la falta de datos para cada subtipo específico de quistes pancreáticos. Por ello, para lograr una clasificación de los

diferentes subtipos, se requiere un mayor número de casos en la base de datos de entrenamiento.

En el caso de las lesiones quísticas de páncreas, el órgano afectado es pequeño, está rodeado de muchos otros órganos y presenta una vaga delineación vaga, en parte, por los pseudópodos interdigitados. Por ello, la detección de lesiones sigue siendo una tarea difícil para los algoritmos de IA (Duh et al., 2023), especialmente si se quieren alcanzar valores especificidad y sensibilidad aceptables, y requiere una mayor variedad de casos de entrenamiento.

6. Conclusiones

Las inteligencias artificiales pueden ser entrenadas utilizando casos de estudios radiológicos, tales como resonancia magnética, tomografía computarizada, ultrasonidos y radiografías, con el objetivo de que aprendan a identificar o clasificar patologías muy diversas. En este estudio, se han recogido numerosas herramientas presentes en el mercado junto a los ensayos clínicos registrados en la base de datos *ClinicalTrials.gov*.

También, se recogen en una tabla los resultados de especificidad y sensibilidad que arrojan las herramientas, con valores muy diferentes pero difícilmente comparables por la diversidad de estadios de desarrollo, de patologías que se pretenden detectar y de tecnologías de radiología utilizadas.

En el caso del asistente médico Sycai Medical, se muestra una especificidad del 81,8% y una sensibilidad del 93,1% para la detección de lesiones quísticas pancreáticas. Sin embargo, los autores destacan que el algoritmo fue entrenado con datos de pacientes provenientes de un solo hospital y que se necesitan más casos para cada subtipo específico de quistes pancreáticos para lograr una clasificación más precisa.

A pesar del aumento de los estudios clínicos que utilizan inteligencia artificial (IA) en el campo de la medicina, la adopción de estas herramientas en la práctica clínica diaria sigue siendo limitada. Se han identificado varios factores que pueden contribuir a esta situación. En primer lugar, hay una falta de confianza en las herramientas de IA por parte de los médicos y los pacientes, ya que las decisiones clínicas basadas en la IA aún no son completamente comprensibles para la mayoría de las personas. En segundo lugar, la implementación de estas herramientas puede requerir una infraestructura tecnológica y un personal altamente capacitado para su uso. Además, puede ser difícil

obtener acceso a suficientes datos de entrenamiento para las herramientas de IA, lo que puede limitar su capacidad para detectar y clasificar patologías de manera efectiva.

En este sentido, es importante seguir investigando en el campo de la IA en la medicina para poder superar estos obstáculos y mejorar la precisión y eficacia de estas herramientas. Por ejemplo, se necesitan más estudios que evalúen la efectividad de las herramientas de IA en diferentes contextos clínicos y con diferentes grupos de pacientes. También se necesita una mayor colaboración entre los investigadores, los proveedores de atención médica y las compañías tecnológicas para asegurar que las herramientas de IA estén diseñadas para satisfacer las necesidades clínicas reales y que puedan ser fácilmente integradas en la práctica clínica diaria.

En resumen, aunque la IA es una herramienta prometedora en el campo de la medicina, aún hay una brecha entre los estudios clínicos y la adopción en la práctica clínica diaria. Se necesita seguir investigando y desarrollando estas herramientas para superar los desafíos actuales y permitir que la IA tenga un impacto más significativo en la mejora de la atención médica.

7. Referencias

- AI for radiology. (n.d.). Retrieved January 19, 2023, from <https://grand-challenge.org/aiforradiology/>
- CE Certification - Class IIa Medical Device CE Marking. (n.d.). Retrieved March 15, 2023, from <https://www.ce-certification.us/ce-marking-class-IIa-device.html>
- CE Certification - Class IIb Medical Device CE Marking. (n.d.). Retrieved March 15, 2023, from <https://www.ce-certification.us/ce-marking-class-IIb-device.html>
- Duh, M. M., Torra-Ferrer, N., Riera-Marín, M., Cumelles, D., Rodríguez-Comas, J., García López, J., & Fernández Planas, M. T. (2023). Deep Learning to Detect Pancreatic Cystic Lesions on Abdominal Computed Tomography Scans: Development and Validation Study. *JMIR AI*, 2, e40702. <https://doi.org/10.2196/40702>
- Elta, G. H., Enestvedt, B. K., Sauer, B. G., & Lennon, A. M. (2018). ACG Clinical Guideline: Diagnosis and Management of Pancreatic Cysts. *The American Journal of Gastroenterology*, 113(4), 464–479. <https://doi.org/10.1038/AJG.2018.14>
- EU Clinical Trials Register - Update. (n.d.). Retrieved April 16, 2023, from <https://www.clinicaltrialsregister.eu/>
- Gore, J. C. (2020). Artificial intelligence in medical imaging. *Magnetic Resonance Imaging*, 68, A1–A4. <https://doi.org/10.1016/J.MRI.2019.12.006>
- Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism: Clinical and Experimental*, 69S, S36–S40. <https://doi.org/10.1016/J.METABOL.2017.01.011>
- Home - ClinicalTrials.gov. (n.d.). Retrieved March 17, 2023, from <https://clinicaltrials.gov/>
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews. Cancer*, 18(8), 500–510. <https://doi.org/10.1038/S41568-018-0016-5>

Parikh, R., Mathai, A., Parikh, S., Sekhar, G. C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56(1), 45. <https://doi.org/10.4103/0301-4738.37595>

Scheiman, J. M., Hwang, J. H., & Moayyedi, P. (2015). American gastroenterological association technical review on the diagnosis and management of asymptomatic neoplastic pancreatic cysts. *Gastroenterology*, 148(4), 824-848.e22. <https://doi.org/10.1053/J.GASTRO.2015.01.014>

Stark, A., Donahue, T. R., Reber, H. A., & Joe Hines, O. (2016). Pancreatic Cyst Disease: A Review. *JAMA*, 315(17), 1882–1893. <https://doi.org/10.1001/JAMA.2016.4690>

8. Anexos

Posibles estados de reclutamiento (según ClinicalTrials.gov):

- **Not yet recruiting:** The study has not started recruiting participants.
- **Recruiting:** The study is currently recruiting participants.
- **Enrolling by invitation:** The study is selecting its participants from a population, or group of people, decided on by the researchers in advance. These studies are not open to everyone who meets the eligibility criteria but only to people in that particular population, who are specifically invited to participate.
- **Active, not recruiting:** The study is ongoing, and participants are receiving an intervention or being examined, but potential participants are not currently being recruited or enrolled.
- **Suspended:** The study has stopped early but may start again.
- **Terminated:** The study has stopped early and will not start again. Participants are no longer being examined or treated.
- **Completed:** The study has ended normally, and participants are no longer being examined or treated (that is, the last participant's last visit has occurred).
- **Withdrawn:** The study stopped early, before enrolling its first participant.
- **Unknown:** A study on ClinicalTrials.gov whose last known status was recruiting; not yet recruiting; or active, not recruiting but that has passed its completion date, and the status has not been last verified within the past 2 years.