

Universidad Valparaíso
Escuela de Kinesiología
Facultad de Medicina



**IMPLICANCIA DE LA INTELIGENCIA ARTIFICIAL EN LA DISMINUCIÓN
DEL ERROR DE DIAGNÓSTICO RADIOLÓGICO.
UN METANÁLISIS.**



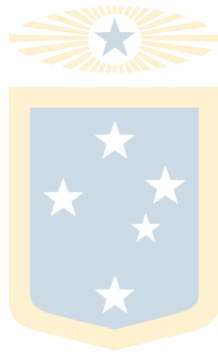
**SEMINARIO DE INVESTIGACIÓN PARA OPTAR AL GRADO DE
LICENCIADO EN KINESIOLOGÍA**

**AUTORES : AMANKAY LIN KAY RAY ROMERO
CLARA VILLEGAS CAVAL**

**PROFESOR GUÍA : KLGO. JUAN CRISTIAN ROJAS MONTERO
PROFESOR CORREFERENTE : SR. ANGELO BARSCH**

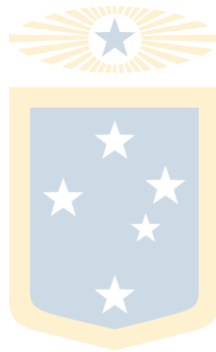
Valparaíso - Chile
2019

Fija tu rumbo a una estrella y podrás navegar a través de cualquier tormenta ...Leonardo da Vinci



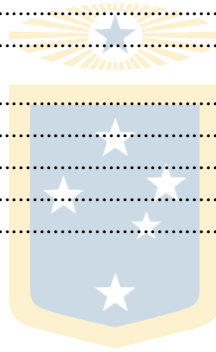
AGRADECIMIENTOS

Deseamos profundamente agradecerle a todos nuestros seres queridos que nos acompañaron y apoyaron durante éste largo proceso. Agradecemos a todo el cuerpo docente de la Escuela de Kinesiología de la Universidad de Valparaíso, especialmente a nuestro tutor y cotutor, Juan Cristian Rojas Montero y Angelo Bartsch, quienes de manera indudable nos cooperaron y facilitaron de su ayuda.



Índice de Figuras

1.1. Esquema de clasificación del error médico	3
2.1. National Academies of Sciences, Engineering, and Medicine. (2015). Improving diagnosis in health care. National Academies Press.....	5
2.2. Imagen de un equipo de rayos X y sus partes	6
3.1. Tabla MeSH	12
3.2. Tabla de criterios de inclusión	16
4.1. Esquema de resultados	18
4.2. Tabla resumen	19
4.3. Tabla de resultados paper n°1	20
4.4. Tabla de resultados paper n°3	21
4.5. Tabla de resultados paper n°5	23



1 | Introducción

1.1. Planteamiento del problema

Cada año la World Health Organization emite un reporte sobre las enfermedades más prevalentes, incidentes, incapacitantes y mortales a nivel mundial, sin embargo, el error médico o iatrogenia médica contribuye una condición que al no encontrarse directamente en los certificados de defunción World Health Organization(2018)ni en los registros médicos,(Martín-Vegue et al., 2002) suele ser ignorada por todos los reportes, siendo un óbice para estudios cuantitativos y comparativos (Landrigan et al., 2010).

Uno de los primeros informes realizados a modo de cuantificar la tasa de error médico fue emitido en 1984 por la unidad de Harvard donde se evidencia que de 2,671,863 pacientes atendidos, 98,609 presentaron algún tipo de efecto adverso en el proceso de tratamiento, solo un 56,6 % mostró lesiones leves, un 13,7 % presentó lesiones de moderado efecto discapacitante, 2,6 % de los pacientes presentó lesiones permanentes o incapacitantes y en el 13,6 % implicó la defunción. Se concluyó que un 3,7 % de los pacientes obtuvo perjuicio al manejo médico, y que los reportes médicos no fomentan la visualización del problema (Brennan et al., 1991).

En 1994, se manifiesta un reporte insinuando que su tasa de incidencia era significativamente más altas, refutando el 3,7 % descrito anteriormente siendo éstas un 1,5 % más alto que las encontradas por los investigadores de Harvard, es decir, un porcentaje de afección del 5,2 % de los pacientes (Leape et al., 1993).

Se define el error médico como: "la falla en una acción planeada para ser llevada a cabo, o el uso de un plan incorrecto para lograr un objetivo", en base a ésto se evidenció una cifra 9800-44000 de muertes anuales en EEUU, por lo que se posiciona como la octava causa de muerte en América, teniendo un costo total estimado entre los 17 y 29 billones (Kohn et al., 2008) . Este reporte marcó un antes y un después sobre el

pensamiento general respecto al actuar médico, desplegando un nuevo debate sobre la preeminencia de los profesionales de la salud que ejercen la medicina (Hayward y Hofer, 2001).

Años más tarde un análisis, promedia el número de muertes estipuladas en los distintos estudios emitidos hasta el año 2015, el cual asoció un número de 251.000 muertes anuales. Dado los siguientes resultados se corona al error médico como la tercera causa de muerte en América, lo que conlleva a cavilar que más número de personas mueren al año a causa de una iatrogenia médica que del, sida, accidentes automovilísticos, suicidios, o cáncer de seno (Makary y Daniel, 2016).

Institute for Healthcare Improvement dice que de todos los errores médicos solo el 45 % de ellos son reportados siendo el más prevaeciente el error diagnóstico (Sheets et al., 2017).

Henty Garland radiólogo Irlandés, quien dedicó 41 años de su vida a evaluar la exactitud de los procesos diagnósticos (Berlin, 2007) mostró en dos de sus estudios, que el porcentaje de error inter-evaluador corresponde al 16.2 %, el de intra-evaluador al 8 % y el error total a un 20 %, este estudio fue realizado en 1949, por 3 especialistas broncopulmonares y 2 radiólogos, conclusión que se determinó tras el estudio de 53 radiografías de tórax (Garland, 1949). Años más tarde, en 1952 se evaluó la capacidad de 6 radiólogos expertos en clasificar un total de 150 muestras divididas en, la progresión del paciente en cuestión, habría mejorado, empeorado o se mantuvo en la misma situación, se concluyó que el error de diagnóstico contribuye al 29.1 % el error inter-evaluador correspondía al 21.5 % y el error intra-evaluador corresponde al 30.1 % (Garland et al., 2014). Estos estudios nos permiten apreciar que los porcentajes de error se mantienen bastante altos sin importar la muestra.

Con el pasar del tiempo más investigadores presentaron estudios sobre el error radiológico, lo que permitió clasificar los criterios de error diagnóstico radiológicos más comunes y sus porcentajes correspondientes, se coronó que las causas más preponderantes eran la satisfacción de búsqueda y la deficiencia de identificación, lo que se desprende de esta investigación que el diagnóstico en radiología depende principalmente del evaluador, específicamente de su sensibilidad y especificidad a la hora de diagnosticar una imagen (Bruno et al., 2015).

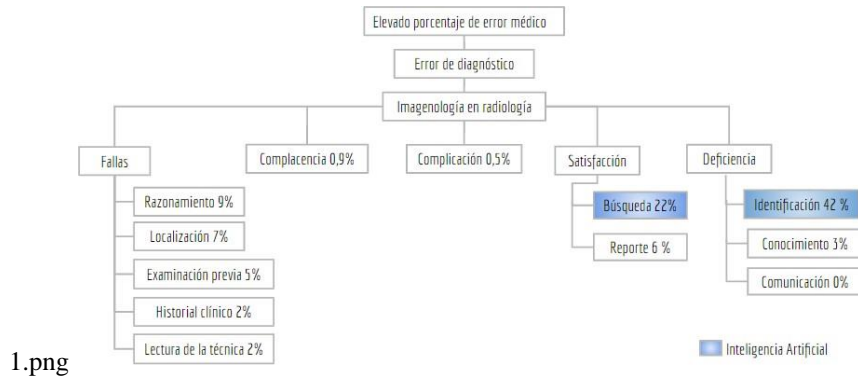


Figura 1.1: Esquema de clasificación del error médico.

1.2. Justificación del problema

En 1956 la conferencia de Darmouth dio inicio a la La inteligencia artificial (Benítez, 2014) para unos años después lograr que esta sea una de las ciencias que ha presentado más investigación con respecto al diagnóstico de imágenes presentando más de 3000 investigaciones en los años comprendidos entre el 2013 y el 2016 (Jiang et al., 2017) , esta no solo cuenta con sustento teórico ya que su funcionalidad también ha sido probada en revistas tan importantes como Nature quien el año 2017 publicó una investigación sobre la creación de una aplicación móvil capaz de diagnosticar el cáncer de piel mejor que un radiólogo (Esteva et al., 2017) , éste estudio no se encuentra aislado ya que cada año la universidad de stanford emite el ImageNet challenge el cual incita a los investigadores para que creen algoritmos que permitan clasificar imágenes. Este reto logró igualar el nivel del ojo humano en condiciones óptimas y desde este año en adelante ya ha superado a ser humano no tanto en especificidad y sensibilidad a la hora de reconocer imágenes, sino que también en el tiempo empleado para analizar grandes grupos de imágenes (Von Zitzewitz, 2017)

Ahora bien la inteligencia artificial no solo ha demostrado que puede “ver” mejor que los humanos, esta también puede tomar decisiones informadas en bases a los datos entregados por el usuario, por lo que podemos apreciar que la máquina podría “pensar” de manera bastante similar a la de los humanos (Kowalski, 2011) . En conjunto con esto si tenemos en cuenta que el uso de inteligencia artificial corresponde a un menor costo del punto de vista del sistema de salud que el costo asociado al error médico (Kohn et al., 2008) .

1.3. Formulación del problema

Planteamos la siguiente pregunta de investigación ¿Puede la Inteligencia Artificial (IA) disminuir el porcentaje de error de diagnóstico médico en imagenología radiológica?

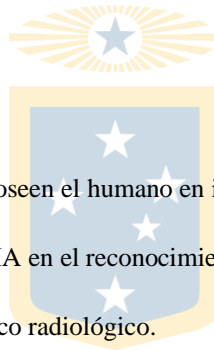
1.4. Objetivos

1.4.1. Objetivo general

Determinar el rango de error que presenta la IA en el reconocimiento de hallazgos radiológicos en comparación al médico.

1.4.2. Objetivos específicos

- Determinar el rango de error que poseen el humano en identificar hallazgos radiológicos.
- Determinar el rango de error de la IA en el reconocimiento de hallazgos radiológicos.
- Determinar el proceso de diagnóstico radiológico.
- Comparar rangos de error de la IA y el ser humano.



2 | Marco teórico

2.1. Proceso de diagnóstico radiológico

El proceso diagnóstico es caracterizado por ser una acción colaborativa, compleja, iterativa e integrativa, que busca la mayor precisión al momento de etiquetar una enfermedad. Su desarrollo, consta de situaciones de retroalimentación, con el fin de disminuir las probabilidades de error diagnóstico, en efecto, aumentar la sensibilidad y especificidad de manera más exacta. Éste, de manera constante está recabando, integrando e interpretando información, describiendo a este proceso como reiterativo y cíclico. El que aplique un sistema de trabajo multidireccional, facilita el espacio de aprendizaje y por ende a su fin (Bockor et al.,2017)

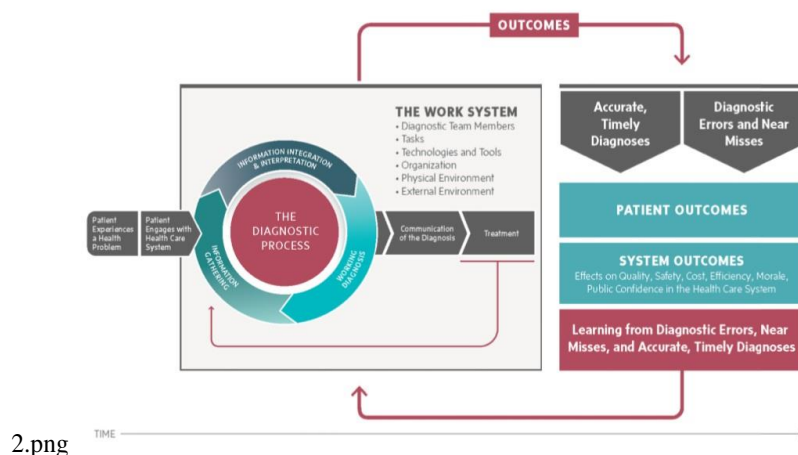


Figura 2.1: National Academies of Sciences, Engineering, and Medicine. (2015). Improving diagnosis in health care. National Academies Press.

Se define como radiografía a aquel procedimiento que es utilizado para hacer fotografías del interior de un cuerpo por medio de rayos X (Real academia Española [RAE], 2016), ésta herramienta tecnológica contribuye

en el proceso diagnóstico que realiza el médico, siendo un apoyo a la toma de decisión diagnóstica.

El proceso de captura radiográfica emite haces de radiación, que pueden ser rayos X, gamma o de algún otro tipo similar, cierta parte de ésta radiación es absorbida por el sujeto expuesto y otra es depositada en una placa de celulosa que puede ser de una radiografía convencional o de un detector electrónico de ser digital (Hernando et al., 2003).

Se pueden cuantificar cuatro etapas en el proceso diagnóstico de la radiografía: la adquisición de la imagen, el procesamiento, la proyección de la imagen y la interpretación de la radiografía por miembros del personal de salud (Doi et al., 2006)

Dentro de la primera etapa se encuentra, la captura de la imagen, este proceso depende de dos elementos, el óptimo funcionamiento del equipo y correcto posicionamiento del paciente (Bansal et al., 2006) El posicionamiento del paciente depende de la patología que esté cursando y las posibles disfunciones asociadas a ella, sin embargo generalmente el posicionamiento se encuentra estandarizado en guisa clínicas (Ahmad, 2008)

El equipo puede ser de dos tipos: convencional o digital, la diferencia entre ellos fundamentalmente radica en que el equipo convencional realiza la captura en una placa celulosa llamada Chasis, mientras que el equipo digital lo realiza sobre una proyección computarizada de la imagen captada (Mattoon et al., 2004). Los equipos cuentan con 3 partes principales: panel de control, panel receptor de imagen y el cabezal, compuesto por un tubo de rayos X, encargado de emitir la energía ionizante mediante el desplazamiento de un fotón desde el cátodo hacia el ánodo del mismo (Hambali et al., 2009). Figura 2.2

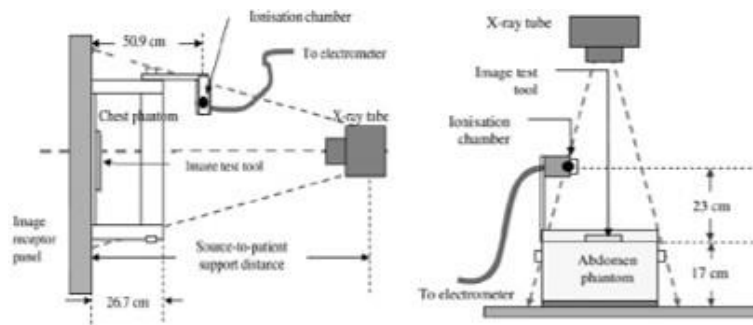


Figura 2.2: Imagen de un equipo de rayos X y sus partes.

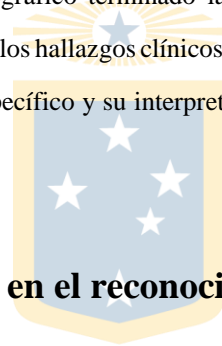
Tras emitir las partículas de radiación son expulsadas de manera dispersa, logrando que las distintas

estructuras del cuerpo sean penetradas de forma desigual dependiendo de la densidad del tejido (Raju et al., 2014).

La energía será captada por el receptor, de quedar en la placa deberán pasar por un proceso de revelado, no obstante, en el caso de ser una radiografía digital el procesamiento de la imagen es más expedito, pasando por un transductor y siendo revelada momentaneamente tras su captura (Alkhalifah et al., 2004)

Posterior a la exposición de la placa a la radiación ionizante comienza el proceso de revelado, en donde se crea una imagen latente del Chasis, una vez finalizado el proceso se logra apreciar la imagen de color negro/plateado metalizado. En éste punto se debe lavar la placa para eliminar el exceso de revelado, luego para la fijación y el endurecimiento de ésta se disuelven los cristales de haluro de plata no expuestos. Ulteriormente se vuelve a enjuagar con el secado de la imagen para eliminar el exceso de agua (Beutel et al., 2000).

Finalmente y con el proceso radiográfico terminado la proyección debiera ser evaluada por algún profesional autorizado el cual dará valor a los hallazgos clínicos encontrados (Bail et al., 2015). Cabe destacar, que éste examen es más sensible que específico y su interpretación depende en gran medida del evaluador (Degnan et al., 2018)



2.2. Inteligencia Artificial en el reconocimiento de imágenes:

Inteligencia Artificial

El termino fue acuñado en 1956 por John McCarthy, jefe del departamento de computación de la Universidad de Standford. Define la IA como ciencia e ingeniería de hacer maquinas inteligentes". Este concepto fue cambiando con los años según los diferentes autores, asociándose cada vez a replicar o mejorar las funciones cognitivas humanas a través del uso de maquinas (Grewal et al., 2014)

Machine learning

Rama de la inteligencia artificial destinada a la realización de una tarea o función, mediante el uso de maquinas. En base a la tarea que se esté buscando, la máquina debe adquirir una habilidad, la cual deber ser desarrollada a través de un proceso de aprendizaje en base a datos entregados por un programador (Domingos cols., 2012).

Artificial Neural Networks

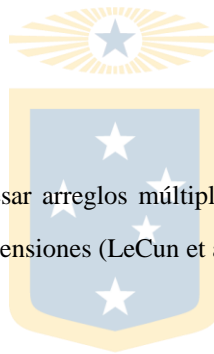
Su sistema de aprendizaje de máquinas es creado con el fin de imitar la configuración del sistema nervioso humano, esto lo realiza a través de una capa de ingreso (input layer), una capa oculta(hidden layer) y una de salida(output layer). Dicho entramado de redes posee la función de tomar decisiones, realizar predicciones o clasificar información (Kurita et al., 2019).

Deep learning

Una variante de las ANNs indica que la cantidad de capas ocultas siempre es mayor a una, permitiendo hacer predicciones más precisas en bases a datos analizados, con un mayor entramado estructural(Schidhuber et al., 2015)

Convolutional neural networks

Tipo de ANNs utilizada para procesar arreglos múltiples de datos, tales como las imágenes. Éstas presentan una configuración en 2 o 3 dimensiones (LeCun et al., 2015)



2.2.1. Análisis de imágenes

Sensing

Es aquel proceso de capacitación de imagen generado por medio de un transductor, convirtiendo las tonalidades propias de la realidad en arreglos bi o tridimensionales de pixeles.

Preprocessing

Se define como la modificación primaria de la imagen, donde se busca la eliminación del ruido de los posibles artefactos, además realiza ajustes de luz y sombras en la imagen antes de procesarla (Boe et al., 2002)

Segmentation

Consiste en aislar el segmento de una totalidad de imagen a analizar, diferenciando un objeto en específico de un grupo de objetos con características similares.

Feature extraction

Segregación de rasgos característicos o sobresalientes del objeto a analizar, para posteriormente ser comparados con los mismos rasgos, pero de otros objetos (Guyon et al., 2008)

Classification

Comparación y estratificación de los rasgos estudiados. Es menester para el modelamiento, permite realizar un análisis previo a los datos y escoger el modelo más representativo para la muestra en cuestión.

Modeling

Es la creación del algoritmo, sustancial para el inicio del proceso de aprendizaje del programa o máquina. Éste es el paso que presenta mayor variabilidad, por ende, debe ser descrito para cada grupo de datos en particular, según las órdenes correspondientes del programador.

Training

Método por el cual se lleva a cabo el aprendizaje, puede o no ser supervisado y al igual que el modelamiento, la elección del método para realizar el entrenamiento depende directamente de la finalidad que se le desee otorgar al uso de los datos (Steels et al., 2018)

Cost function

Función por el cual se pretende optimizar el porcentaje de error del programa tras una primera instancia de entrenamiento. Se describe la existencia de dos tipos, convencional o cuadrática, sus usos dependen de la fiabilidad que se le quiera imprimir al programa (Chen et al., 2016)

Post processing

Corresponde al proceso posterior a la programación del algoritmo, donde se afinan los parámetros que pudiesen haberse alterado o comprometido a lo largo del proceso . Esta etapa consta de la asignación del valor real a la propuesta, relacionando los valores predictivos obtenidos con la realidad observable (Steels et al., 2018)



3 | Metodología

3.1. Tipo de investigación

El estudio realizado es de tipo explicativo observacional, de colección retrospectiva de datos. Revisión bibliográfica de estudios clínicos experimentales. La presente investigación se encuadra en el marco de una serie de revisiones sistemática, las cuales pretende establecer una relación numérica objetiva entre el error asociado al médico radiólogo, en contraste con el mecanismo de inteligencia artificial.

3.2. Estrategias de búsqueda

El estudio fue llevado a cabo durante los meses comprendidos entre Marzo y Diciembre del presente año, siendo realizado por dos investigadores de forma independientes, los cuales establecieron en una primera instancia los parámetros generales de la investigación según el cronograma (Ver anexo cronograma), para luego determinar los aspectos específicos de la búsqueda bibliográfica como se indican a continuación.

El diseño de investigación que utilizó en todas las revisiones sistemáticas, fue la utilización de buscadores online especializados en el uso de artículos científicos, tales como: PubMed, Scencedirect, Wiley y SpringerLink, en donde se incluyeron a todos aquellos que cumplieran los criterios temáticos, de calidad y metodológicos establecidos por los mismo, para finalmente ser revisados, analizados, resumidos y comparados.

Para la estrategia de búsqueda, se definieron en primera instancias las palabras claves y los operadores booleanos.

3.3. Términos de búsqueda

Las palabras clave o MeSH (Medical Subject Headings) fueron seleccionadas en relación con los objetivos establecidos previamente determinando 4 focos principales dentro de la investigación que se corresponden al error médico, la inteligencia artificial, el profesional radiólogo y las placas radiológicas. Estos son: “error range”; “error rate”; “incidence of error”; “accuracy”; “sesibility”; “specificity”; “artificial intelligence”; “radiography”; “X-rays”; “radiologist”; “specialist”. Los operadores booleanos se ordenan según su pertenencia, “OR” si pertenece al mismo pilar de investigación, o “AND” si pertenecen a distintos pilares, estos fueron unidos por medio de los operadores.

MeSH

PILARES	PALABRAS CLAVES
Error médico o de IA	<ul style="list-style-type: none"> ▪ Error range ▪ Error rate ▪ Incidence of error ▪ Accuracy ▪ Sensibility ▪ Specificity
Radiólogo o Especialista	<ul style="list-style-type: none"> ▪ Radiologist ▪ Specialist
Placa radiográfica	<ul style="list-style-type: none"> ▪ Radiography ▪ X-rays
Inteligencia artificial	<ul style="list-style-type: none"> ▪ Artificial Intelligence

Figura 3.1: Tabla MeSH

Para establecer los términos de búsqueda, se agruparon los pilares de error médico- radiografía con inteligencia artificial, de forma paralela error médico-radiografía con radiólogo y finalmente todos los pilares en conjunto. Lo que concluyó en tres revisiones sistemáticas con los siguientes términos de búsqueda.

- Revisión n°1: [“error range” OR “error rate” OR “incidence of error” OR “accuracy” OR “sesibility” OR “specificity”] AND [“radiography” OR “X-rays”] AND [“artificial intelligence”].
- Revisión n°2: [“error range” OR “error rate” OR “incidence of error” OR “accuracy” OR “sesibility”

OR “specificity”]AND[“radiography” OR “X-rays”]AND[“radiologist” OR “specialist”].

- Revisión n°3:[“error range” OR “error rate” OR “incidence of error” OR “accuracy” OR “sesibility” OR “specificity”]AND[“radiography” OR “X-rays”]AND[“artificial intelligence”]AND[“radiologist” OR “specialist”].

3.3.1. Elección de buscadores

Los buscadores se instauraron por medio de una lista de buscadores pertinentes a la temática entregada por el docente del área de investigación de la Universidad de Valparaíso de la carrera de Kinesiología, Juan Cristian Rojas Montero, el cual considero las siguientes plataformas: Pubmed, Scienedirect, SpringerLink, Nature, Wiley Online Library, Scopus, Karger, Jstor, EBSCO y Scielo. De estos buscadores se seleccionaron cuatro en base al número de papers presentados sobre la temática. Estos fueron: Wiley Online Library, ScienceDirect, Spriger y Pubmed. A cada uno de estos buscadores se les ingresaron los términos de búsqueda establecidos para cada una de las revisiones.

3.4. Criterios de búsqueda de artículos científicos

Los criterios para la búsqueda bibliográfica fueron divididos en tres categorías:

3.4.0.1. Criterios propios de la publicación

Dentro de ellos se encuentran todos aquellos aspectos de la publicación que pudiesen ser filtrados por el mismo buscador dentro de las diferentes plataformas y corresponden a la fecha de publicación, idioma, tipo de investigación y la especie en la que se realizado el estudio.

3.4.0.2. Criterios temáticos

Son determinados como los principales criterios de pertenencia para la investigación. Estos se establecen en relación con las líneas investigativas presentadas, analizando su presencia o ausencia dentro de las distintas revisiones. Se espera que para cada revisión se presente la totalidad de los pilares incluidos en ella, junto con una explícita relación entre los mismos.

3.4.0.3. Criterios de calidad

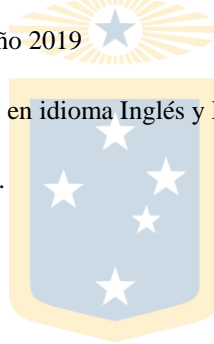
Su función es denotar el carácter cualitativo de cada una de las investigaciones ingresadas en la revisión. Se instituyen dos mecanismos de valoración: Impact factor (IF) y SRJ.

3.5. Criterios de inclusión

Serán aceptados solo aquellos estudios que presenten las siguientes características:

3.5.0.1. Criterios de la publicación

- a.- Fecha de publicación: Estudios con un máximo de cinco años de antigüedad, comprendiendo este lapso entre enero del año 2014 a julio del año 2019
- b.- Idioma: Solo serán aceptados estudios en idioma Inglés y Español.
- c.- Tipo de investigación: Estudio clínico.
- d.- Especie: Humanos.



3.5.0.2. Criterios temáticos

- a.- Error: El estudio debe mostrar explícitamente un valor numérico, ya sea a modo de rango, porcentaje o número de casos del error establecido, tanto para la inteligencia artificial como para el desempeño del especialista o radiólogo.
- b.- Radiólogo: Debe explicitar la capacidad de todo aquellos sujetos que participe dentro del estudio para evaluar una placa radiográfica.
- c.- Inteligencia artificial: Se aceptará el uso de cualquier tipo de inteligencia artificial siempre y cuando se establezca el uso de ésta directamente en el análisis de la placa, no solamente de los datos obtenidos o de otras aplicaciones.
- d.- Placa radiológica: Se debe especificar el uso de algún tipo de placa radiológica, ya sea virtual o física.

e.- Relación: Se incluirán todos aquellos estudios que presenten algún tipo de comparación o correlación entre los pilares.

3.5.0.3. Criterios de calidad

a.- IF: Todos los estudios deben poseer un IF > 1,5

b.- SJR: Se aceptarán aquellas publicaciones con un puntaje >1,5

3.6. Criterios de exclusión

Serán excluidos de la investigación todos aquellos estudios que no presenten alguno de los criterios previamente planteados.

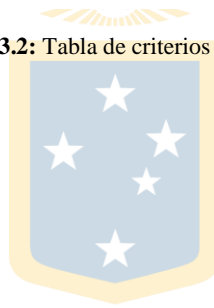


3.7. Documentos incluidos

Cada investigador deberá revisar todos los archivos encontrados por los buscadores y determinará según la tabla de criterios de inclusión, la pertenencia de cada documento en el estudio. En caso de encontrar discordancia entre ambos evaluadores sobre la pertenencia de algún estudio en particular, ésta disputa será resuelta por el docente a cargo de la investigación el cual determinará si este debiese ser incluido o no.

Criterios	Categorías	Paper 1	Paper 2
Criterios propios de la publicación	Idioma	Si	Si
	Fecha	Si	Si
	Especie	Si	Si
	Tipo	Si	Si
Criterios temáticos	Error médico	Si	Si
	Radiólogo	Si	Si
	IA	Si	Si
	Placa radiológica	Si	Si
	Relación	Si	Si
Criterios de calidad	IF >1,5	No	Si
	SJR >1,5	Si	Si
Resultado final	Aceptación	Rechazado	Aceptado

Figura 3.2: Tabla de criterios de inclusión



3.8. Materiales

- Ordenadores (Notebooks)
- Red de conexión a internet
- Microsoft Office Word
- Microsoft Office Excel
- Google Drive
- Mendeley Desktop
- Texmaker

4 | Resultados

Dentro del metaanálisis realizado, se obtuvieron un total de 30.256 papers encontrados en los distintos buscadores determinados previamente, de los cuales en la primera revisión sistemática se obtuvieron un total de 14.487 papers distribuidos de la siguiente forma: Wiley Online Library con un total de 3.967, ScienceDirect con 5.307, Springerlink con 5.069 y finalmente Pubmed con 144 papers. Tras la aplicación del primer criterio, propios de la publicación, se obtuvieron un total 402 papers, quedando 23 papers al agregar el criterio temático, sin embargo con el último criterio, criterio de calidad, el resultado fue de tres papers.

En la segunda revisión dio un total de 12.095 papers distribuidos en los buscadores: Wiley Online Library con un total 6.498, ScienceDirect con 3.483, Springerlink con 2.109 y finalmente Pubmed con 5 papers. Al emplear el primer filtro se obtuvo un total de 641 papers, al segundo filtro 47 papers, mientras que al aplicar el último criterio se obtuvo como resultado final 5 papers.

En la tercera y última revisión se encontraron un total de 3.674 papers obteniendo distintos resultados en los buscadores, en Wiley Online Library un total 1.573, en ScienceDirect con 706, Springerlink con 1.375 y finalmente Pubmed con 20 papers. Dicha revisión fue la que mayor diferencia obtuvo al momento de la aplicación de los criterios, propios de la publicación, temáticos y de calidad, obteniendo como resultados 211, 2 y 1 paper respectivamente.

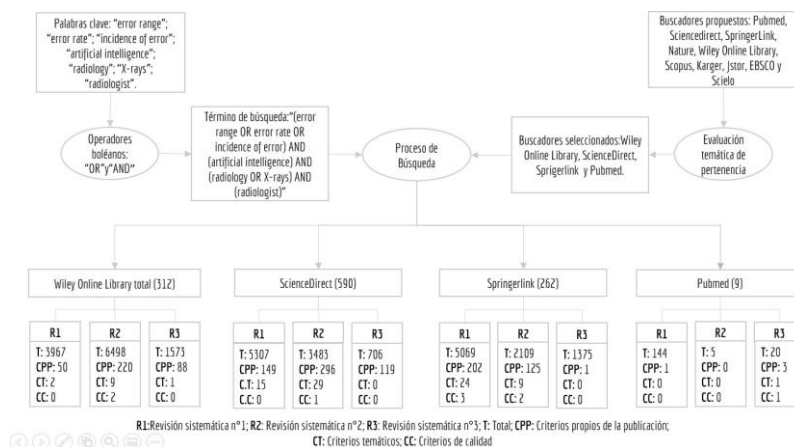
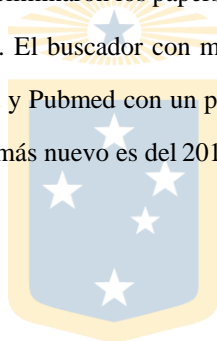


Figura 4.1: Esquema de resultados

Del total de papers seleccionados se eliminaron los papers repetidos entre las revisiones quedando ocho papers como se observa en la tabla final. El buscador con más papers fue Springerlink con ocho papers, seguido por Wiley con dos, ScienceDirect y Pubmed con un paper cada uno de ellos. El paper más antiguo corresponde al año 2016 mientras que el más nuevo es del 2019.



Revisión	Buscador	Año	Autor	Título
Revisión 1	Springerlink	2018	Choi, J. A., Kim, Y. C., Min, S. J., & Khil, E. K.	A simple method for bone age assessment: the capitohamate planimetry
Revisión 1	Springerlink	2019	Fan, Y., Chen, Q., Chen, T., Wang, H., Zhu, X., Fang, Z., & Lu, Y.	Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays
Revisión 2	Science Direct	2017	Crabtree, N. J., Chapman, S., Höglér, W., Hodgson, K., Chapman, D., Bebbington, N., & Shaw, N. J.	Vertebral fractures assessment in children: Evaluation of DXA imaging versus conventional spine radiography
Revisión 2	Springerlink	2019	Cheng, C. T., Ho, T. Y., Lee, T. Y., Chang, C. C., Chou, C. C., Chen, C. C., ... & Liao, C. H.	Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs
Revisión 2	Springerlink	2018	Walton, T. R., & Layton, D. M.	Intra- and inter-examiner agreement when assessing radiographic implant bone levels: Differences related to brightness, accuracy, participant demographics and implant characteristics
Revisión 2	Wiley	2016	Henostroza, G., Harris, J. B., Kancheya, N., Nhandu, V., Besa, S., Musopole, R., ... & Reid, S. E.	Chest radiograph reading and recording system: evaluation in frontline clinicians in Zambia
Revisión 2	Wiley	2017	Eliyas, S., Briggs, P. F. A., Harris, I. R., Newton, J. T., & Gallagher, J. E.	Development of quality measurement instruments for root canal treatment
Revisión 3	Pubmed	2018	Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Patel, B. N.	Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists.

Figura 4.2: Tabla resumen

A continuación se realizará una breve descripción de los objetivos, metodología y resultados de todos aquellos estudios que finalmente fueron seleccionados para realizar el metaanálisis:

4.1. Deep learning for chest radiograph diagnosis: A retrospective of the ChEXNetXt algorithm to practicing radiologists

El estudio plantea el desarrollo de una red CNN para evaluar 14 patologías pulmonares llamada CheX-NeXt, para posteriormente comparar el rendimiento del programa con el de un panel de expertos (9 radiólogos) en materias de rendimiento y tiempo. Lo que busca es, medir el rendimiento del programa y del panel de expertos al evaluar un set de 420 radiografía de torax con las 14 patologías planteadas. A ambos se les evalúa área bajo la curva(AUC),para curva de sensibilidad/especificidad y el tiempo de evaluación. Los resultados

arrojados fueron:

Resultados	
Tiempo promedio (420 Rx): <ul style="list-style-type: none"> • Radiólogos 240 min • IA 1,5 min AUC Promedio <ul style="list-style-type: none"> • Radiólogo: 0.80 • IA: 0.86 Atelectasia: <ul style="list-style-type: none"> • Radiólogo: 0.80 • IA: 0.86 Cardiomegalia: <ul style="list-style-type: none"> • Radiólogo: 0.88 • IA: 0.83 Consolidación: <ul style="list-style-type: none"> • Radiólogo: 0.84 • IA: 0.89 Edema: <ul style="list-style-type: none"> • Radiólogo: 0.91 • IA: 0.92 Derrame: <ul style="list-style-type: none"> • Radiólogo: 0.90 • IA: 0.90 Enfisema: <ul style="list-style-type: none"> • Radiólogo: 0.91 • IA: 0.70 	Fibrosis: <ul style="list-style-type: none"> • Radiólogo: 0.89 • IA: 0.80 Hernia: <ul style="list-style-type: none"> • Radiólogo: 0.98 • IA: 0.85 Infiltración: <ul style="list-style-type: none"> • Radiólogo: 0.73 • IA: 0.72 Masa: <ul style="list-style-type: none"> • Radiólogo: 0.88 • IA: 0.90 Nódulo: <ul style="list-style-type: none"> • Radiólogo: 0.89 • IA: 0.89 Fibrosis pleural <ul style="list-style-type: none"> • Radiólogo: 0.77 • IA: 0.79 Neumonía: <ul style="list-style-type: none"> • Radiólogo: 0.82 • IA: 0.85 Neumotórax <ul style="list-style-type: none"> • Radiólogo: 0.94 • IA: 0.94

Figura 4.3: Tabla de resultados paper n°1

4.2. Development of quality measurement instruments for root canal treatment

Pretenden diseñar un instrumento de medición para evaluar la calidad del tratamiento del conducto radicular, para evaluar la capacitación y el resultado de los odontólogos generales que trabajan en un entorno de área pública. La metodología utilizada, consistió en establecer un sistema de puntuación con calidad de tratamiento de 4 dominios,.

- 1) Calidad de proceso de tratamiento clínico (proceso)
- 2) Calidad de relleno de la raíz, como se ve radiográficamente (proceso).
- 3) Curación, como se ve clínicamente (resultado).

- 4) Curación , como se ve radiográficamente (resultado)
- Complejidad de los dientes a tratar (estructura)

Los resultados obtenidos fueron de fiabilidad intra-evaluador: 22-100 %, fiabilidad interevaluador: 18-99 %

4.3. Chest radiograph reading and recording system:evaluation in frontline clinicians in Zambia

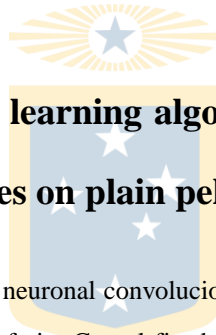
El siguiente paper tiene como objetivo evaluar la sensibilidad, especificidad y fiabilidad inter-evaluador de un grupo de 4 médicos en Zambia ,por medio de la detección anómala de 571 placas radiográficas de tórax asociadas a Tuberculosis en prisioneros con y sin VIH.

Resultados
Fiabilidad inter-evaluador: <ul style="list-style-type: none"> • Pacientes con VIH: 76-96% • Pacientes sin VIH: 65-93%
Sensibilidad: <ul style="list-style-type: none"> • Pacientes con VIH: 50-77% • Pacientes sin VIH: 55-70%
Especificidad: <ul style="list-style-type: none"> • Pacientes con VIH:35-59% • Pacientes sin VIH: 33-60%

Figura 4.4: Tabla de resultados paper n°3

4.4. Intra and inter examiner agreement when assessing radiographic implant bone levels: Differences related to brightness, accuracy, participant demographics and implant characteristics

Su objetivo es evaluar la concordancia intra e interexaminador de la evaluación radiográfica del nivel del hueso marginal (MBL) alrededor de los implantes individuales Brånemark discriminando el nivel de precisión en base al brillo de la radiografía de torax, ésto a través del uso de 100 radiografías evaluadas por 74 profesionales, las cuales fueron evaluadas 2 veces con distintos niveles de brillo y se calculó el promedio de fiabilidad intra e inter-evaluador. Los resultados obtenidos arrojaron un promedio de porcentaje de fiabilidad intra-evaluador: 76 %, promedio de porcentaje de fiabilidad inter-evaluador: 72 %.



4.5. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs

El estudio pretende evaluar una red neuronal convolucionar profunda que detecta fracturas de cadera por medio de radiografía del miembro inferior. Con el fin de determinar según la base de datos PXR del año 2017, la sensibilidad, especificidad, rango de falso-negativo y el AUC del programa, junto con la sensibilidad y especificidad de un grupo de expertos conformado por médicos fisiatras, radiólogos y cirujanos ortopédicos.

Resultados
CNN: <ul style="list-style-type: none">• Sensibilidad: 98%• Especificidad: 91%• Rango de falso-negativo 2%• AUC: 0.98
Médicos fisiatras: <ul style="list-style-type: none">• Sensibilidad: 84-100%• Especificidad: 46-94%
Radiólogos más cirujanos: <ul style="list-style-type: none">• Sensibilidad: 98-100%• Especificidad: 81-93%



Figura 4.5: Tabla de resultados paper n°5

4.6. A simple method for bone age assessment: the capitohamate planimetry

A modo de resumen el paper busca evaluar la planimetría entre el hueso ganchoso y grande para determinar la edad ósea y poder compararlo con el método Greulich-Pyle. Se evalúan 391 radiografías de ambas manos de niños midiendo su planimetría por medio de 4 radiólogos, obteniendo como resultado una precisión interobservador de un 96 %.

4.7. Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays

Medición del ángulo de coob, en base a radiografías de tórax, por medio de un método de inteligencia artificial, para ser comparado con medición manual realizada por expertos. Para esto se requirió de dos radiólogos que de manera individual, realizaron una medición manual del ángulo de coob en 248 radiografías. Tras esta medición se evaluó la fiabilidad intra en interobservador de ambos y se comparó con la medición realizada por la IA para evaluar la reproductibilidad, sensibilidad y especificidad. Los resultados concluyeron en, fiabilidad intraobservador: 94 %, fiabilidad interobservador:88 %, reproductibilidad:85 %, sensibilidad:89,5 %, especificidad.79.3 %.



4.8. Vertebral fractures assessment in children: Evaluation of DXA imaging versus conventional spine radiography

Éste estudio pretende evaluar la presencia de fracturas vertebrales en niños por medio de densitometría ósea y radiografías. Se analizaron a 120 niños con radiografías de distintos grados de severidad de fractura, las cuales fueron estudiadas por un radiólogo pediátrico y dos pediatras, con el fin de determinar el nivel de sensibilidad, especificidad y la fiabilidad inter-evaluador. Los resultados entregados fueron, fiabilidad interevaluador: 13-68 %, sensibilidad:81.3-99.3 %, especificidad:62.5-99.2 %.

Bibliografía

- Ahmad, Naveed (2008). X-Ray Patient Positioning Manual. *Auntminnie*, (pp. 152). 2.1
- American Journal of Roentgenology*, 188(5), 1173–1178. 1.1
- Al Khalifah, K., & Brindhavan, A. (2004). Comparison between conventional radiography and digital radiography for various kVp and mAs settings using a pelvic phantom. *Radiography*, 10(2), 119-125.
- Bansal, G. J. (2006). Digital radiography. A comparison with modern conventional imaging. *Postgraduate medical journal*, 82(969), 425-428.
- Benítez, Antonio (2014). Inteligencia Artificial : la intrahistoria. (pp. 1–19). 1.2
- Berlin, Leonard (2007). Accuracy of Diagnostic Procedures: Has It Improved Over the Past Five Decades?
- Bow, S. T. (Ed.). (2002). *Pattern recognition and image preprocessing*. CRC press.
- Brennan, Troyen A; Leape, Lucian L; Laird, Nan a N M; Ph, D; Hebert, Liesi; Sc, D; Localio, A Russell; Lawthers, Ann G N N G; Newhouse, Joseph P; Weiler, Paul C; Ll, M; y Hiatt, Howard H (1991). Incidence of Adverse Events and Negligence in Hospitalized Patients. Results of the Harvard Medical Practice Study I. *N Engl J Med*, 324(Appendix I), 370–376. 1.1
- Bruno, Michael A.; Walker, Eric A.; y Abujudeh, Hani H. (2015). Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction. *RadioGraphics*, 35(6), 1668–1676. 1.1
- Beutel, J., Kundel, H. L., & Van Metter, R. L. (2000). *Handbook of medical imaging* (Vol. 1). Spie Press.
- Cheng, C. T., Ho, T. Y., Lee, T. Y., Chang, C. C., Chou, C. C., Chen, C. C., ... & Liao, C. H. (2019). Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *European radiology*, 1-9
- Chen, F., Chen, M., Li, Q., Meng, K., Zheng, Y., Guerrero, J. M., & Abbott, D. (2016). Cost-based droop schemes for economic dispatch in islanded microgrids. *IEEE Transactions on Smart Grid*, 8(1), 63-74.
- Chen, L., Qu, H., Zhao, J., Chen, B., & Principe, J. C. (2016). Efficient and robust deep learning with correntropy-induced loss function. *Neural Computing and Applications*, 27(4), 1019-1031.
- Choi, J. A., Kim, Y. C., Min, S. J., & Khil, E. K. (2018). A simple method for bone age assessment: the capitohamate planimetry. *European radiology*, 28(6), 2299-2307.
- Crabtree, N. J., Chapman, S., Höglér, W., Hodgson, K., Chapman, D., Bebbington, N., & Shaw, N. J.

- (2017). Vertebral fractures assessment in children: evaluation of DXA imaging versus conventional spine radiography. *Bone*, 97, 168-174
- Degnan, Andrew J.; Ghobadi, Emily H.; Hardy, Peter; Krupinski, Elizabeth; Scali, Elena P.; Stratchko, Lindsay; Ulano, Adam; Walker, Eric; Wasnik, Ashish P.; y Auffermann, William F. (2018). Perceptual and Interpretive Error in Diagnostic Radiology—Causes and Potential Solutions. *Academic Radiology*, (pp. 1–13). 2.1
- Doi, K. (2006). Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Physics in Medicine & Biology*, 51(13), R5.
- Domingos, P., & Webb, W. A. (2012, July). A tractable first-order probabilistic logic. In Twenty-Sixth AAAI Conference on Artificial Intelligence.
- Eliyas, S., Briggs, P. F. A., Harris, I. R., Newton, J. T., & Gallagher, J. E. (2017). Development of quality measurement instruments for root canal treatment. *International endodontic journal*, 50(7), 652-666
- Esteva, Andre; Kuprel, Brett; Novoa, Roberto A; Ko, Justin; Swetter, Susan M; Blau, Helen M; y Thrun, Sebastian (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115. 1.2
- Garland, L Henry (1949). On the scientific evaluation of diagnostic procedures: presidential address thirty-fourth annual meeting of the radiological society of north america. *Radiology*, 52(3), 309–328. 1.1
- Garland, L. H.; Miller, E. R.; Zwerling, H. B.; Harkness, J. T.; Hinshaw, H. C.; Shipman, S. J.; y Yerushalmy, J. (2014). Studies on the Value of Serial Films in Estimating the Progress of Pulmonary Disease. *Radiology*, 58(2), 161–177. 1.1
- Grewal, R., Lilien, GL y Mallapragada, G. (2006). Ubicación, ubicación, ubicación: cómo la integración de la red afecta el éxito del proyecto en sistemas de código abierto. *Ciencias de la gestión*, 52 (7), 1043-1056.
- Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (Eds.). (2008). *Feature extraction: foundations and applications* (Vol. 207). Springer.
- Hambali, A., Sapuan, S. M., Ismail, N., & Nukman, Y. (2009). Application of analytical hierarchy process in the design concept selection of automotive composite bumper beam during the conceptual design stage. *Scientific Research and Essays*, 4(4), 198-211.
- Hayward, Rodney A y Hofer, Timothy P (2001). Due to Medical Errors Preventability Is in the Eye of the Reviewer. *Jama*, 286(4), 415–420. 1.1
- Henostroza, G., Harris, J. B., Kancheya, N., Nhandu, V., Besa, S., Musopole, R., ... & Reid, S. E. (2016). Chest radiograph reading and recording system: evaluation in frontline clinicians in Zambia. *BMC infectious diseases*, 16(1), 136.
- Jiang, Fei; Jiang, Yong; Zhi, Hui; Dong, Yi; Li, Hao; Ma, Sufeng; Wang, Yilong; Dong, Qiang; Shen, Haipeng; y Wang, Yongjun (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243. 1.2
- Kohn, Linda T; Corrigan, Janet M; y Molla, S (2008). *Rapporteur's Report Session I: Origin of the problem: Malcolm Ross*, volume 52. 1.1, 1.2

- Kowalski, Robert (2011). Artificial intelligence and human thinking. *IJCAI International Joint Conference on Artificial Intelligence*, (pp. 11–16). 1.2
- Kurita, K., Vyas, N., Pareek, A., Black, AW y Tsvetkov, Y. (2019). Medición de sesgo en representaciones de palabras contextualizadas. *preimpresión de arXiv arXiv: 1906.07337*.
- Landrigan, Christopher P.; Parry, Gareth J.; Bones, Catherine B.; Hackbarth, Andrew D.; Goldmann, Donald A.; y Sharek, Paul J. (2010). Temporal Trends in Rates of Patient Harm Resulting from Medical Care. *New England Journal of Medicine*, 363(22), 2124–2134. 1.1
- Leape, L L; Lawthers, A G; Brennan, T A; y Johnson, W G (1993). Preventing medical injury. *QRB. Quality review bulletin*, 19(5), 144–9. 1.1
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Pan, Y., Chen, Q., Chen, T., Wang, H., Zhu, X., Fang, Z., & Lu, Y. (2019). Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. *European Spine Journal*, 28(12), 3035-3043.
- Raju, G. S. (2014). Endoscopic clip closure of gastrointestinal perforations, fistulae, and leaks. *Digestive Endoscopy*, 26, 95-104.
- Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Patel, B. N. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS medicine*, 15(11), e1002686.
- Makary, Martin A. y Daniel, Michael (2016). Medical error-the third leading cause of death in the US. *BMJ (Online)*, 353(May), 1–5. 1.1
- Martín-Vegue, AR; Vázquez-Barquero, JL; y Castanedo, S Herrera (2002). Cie-10 (i): Introducción, historia y estructura general. *Papeles Méd*, 11(1), 24–35. 1.1
- Mattoon, J. S., & Smith, C. (2004). Breakthroughs in radiography: computed radiography. *Compendium*.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- Sheets, Fact; September, Released; Err, To; y States, United (2017). Americans' Experiences with Medical Errors and Views on Patient Safety How many Americans report experiencing medical errors? 1.1
- Steels, L., & Brooks, R. (2018). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Routledge.
- Von Zitzewitz, Gustav (2017). Survey of neural networks in autonomous driving. 1.2
- Walton, T. R., & Layton, D. M. (2018). Intra-and inter-examiner agreement when assessing radiographic implant bone levels: Differences related to brightness, accuracy, participant demographics and implant characteristics. *Clinical oral implants research*, 29(7), 756-771.

