Análisis documental del riesgo asociado a ataques adversariales en el flujo de trabajo radiológico: consecuencias en la validez del diagnóstico por (IA) y retos para la protección de datos sensibles bajo estándares internacionales

Arley Llanos Jamioy

Carlos Julio Nova Ballesteros

Helbert Hugo Nova Rodríguez

Marcela Ocampo Guerra

Yoselen Cecilia Triana Galeano

Asesor

Edna Rocío Jamaica Guio

Universidad Nacional Abierta y a Distancia - UNAD

Escuela de Ciencias de la Salud - ECISA

Tecnología en Radiología e Imágenes Diagnósticas

2026

Resumen

La presente investigación analiza el impacto de los ataques adversariales en los sistemas de inteligencia artificial aplicados a la radiología, con énfasis en sus efectos sobre la validez diagnóstica y la seguridad de los datos clínicos. El estudio se desarrolla bajo un enfoque cualitativo, con un diseño documental basado en la revisión de literatura científica relacionada con inteligencia artificial, ciberseguridad y diagnóstico por imágenes.

A partir del análisis de fuentes académicas, se identifican las principales vulnerabilidades de los sistemas de IA frente a la manipulación de imágenes médicas, evidenciando riesgos asociados a diagnósticos erróneos, alteración de la información clínica y posibles afectaciones a la seguridad del paciente. Asimismo, se examinan las implicaciones técnicas, éticas y clínicas derivadas de estos ataques en entornos hospitalarios digitalizados.

De igual forma, se destacan estrategias de mitigación basadas en la literatura, orientadas al fortalecimiento de la ciberseguridad, la protección de datos sensibles y la confiabilidad de los sistemas diagnósticos asistidos por inteligencia artificial.

El estudio resalta la necesidad de fortalecer los marcos de gobernanza de datos y la supervisión humana en los sistemas de inteligencia artificial, considerando que la automatización en radiología no elimina la responsabilidad clínica del profesional. En este sentido, se plantea que la implementación segura de estas tecnologías requiere una integración equilibrada entre innovación tecnológica, regulación ética y capacitación especializada del personal, con el fin de garantizar diagnósticos confiables y minimizar riesgos asociados a la manipulación adversarial en entornos clínicos digitalizados.

Palabras clave: inteligencia artificial, radiología, ataques adversariales, ciberseguridad, diagnóstico médico.

Abstract

This study analyzes the impact of adversarial attacks on artificial intelligence systems applied to radiology, with a focus on their effects on diagnostic validity and clinical data security. The research adopts a qualitative approach, using a documentary design based on the review of scientific literature related to artificial intelligence, cybersecurity, and medical imaging.

Through the analysis of academic sources, the main vulnerabilities of AI systems to the manipulation of medical images are identified, highlighting risks such as misdiagnosis, alteration of clinical information, and potential threats to patient safety. Additionally, the study examines the technical, ethical, and clinical implications of these attacks within digitalized hospital environments.

Furthermore, mitigation strategies found in the literature are discussed, aimed at strengthening cybersecurity measures, protecting sensitive data, and improving the reliability of AI-assisted diagnostic systems.

The study highlights the need to strengthen data governance frameworks and human oversight in artificial intelligence systems, recognizing that automation in radiology does not replace clinical responsibility. In this context, it is argued that the safe implementation of these technologies requires a balanced integration of technological innovation, ethical regulation, and specialized staff training, in order to ensure reliable diagnoses and minimize risks associated with adversarial manipulation in digitalized clinical environments

Keywords: artificial intelligence, radiology, adversarial attacks, cybersecurity, medical diagnosis.

.

Tabla de Contenido

Lista de Tablas

Lista de Figuras

Introducción

En los últimos años, la inteligencia artificial ha tomado un papel cada vez más importante dentro del sector salud, especialmente en el área de la radiología, donde su uso ha permitido mejorar la rapidez y precisión en la interpretación de imágenes diagnósticas como radiografías, tomografías y resonancias magnéticas. Gracias a estas herramientas tecnológicas, los profesionales pueden contar con un apoyo adicional al momento de identificar anomalías, priorizar estudios urgentes y optimizar los tiempos de respuesta en la atención de los pacientes. Esto ha convertido a la inteligencia artificial en un recurso de gran valor dentro de los servicios de diagnóstico por imágenes.

Sin embargo, así como la implementación de estas tecnologías representa múltiples beneficios, también trae consigo nuevos desafíos que no pueden ser ignorados. Uno de los más importantes son los llamados ataques adversariales, que consisten en la alteración intencional de imágenes médicas o datos clínicos para engañar a los sistemas de inteligencia artificial y modificar sus resultados. Aunque estas alteraciones pueden ser mínimas e incluso imperceptibles para el ojo humano, tienen la capacidad de generar diagnósticos incorrectos que pueden afectar directamente la salud del paciente y la toma de decisiones médicas.

Además del riesgo clínico, esta problemática también involucra la seguridad de la información. Actualmente, los hospitales trabajan con sistemas digitales interconectados donde se almacenan grandes cantidades de datos sensibles de los pacientes, como historias clínicas, estudios radiológicos e información personal. Cuando estos sistemas presentan fallas de ciberseguridad o son vulnerables a ataques externos, no solo se compromete la calidad del diagnóstico, sino también la confidencialidad y la integridad de la información médica, generando consecuencias éticas, legales e institucionales.

Por esta razón, resulta necesario analizar de qué manera los ataques adversariales afectan la validez del diagnóstico radiológico basado en inteligencia artificial y qué estrategias pueden implementarse para reducir estos riesgos. Esta investigación busca comprender las principales vulnerabilidades presentes en estos sistemas, así como las posibles soluciones orientadas a fortalecer la ciberseguridad, la supervisión profesional y la protección de los datos clínicos. De esta manera, se pretende aportar una visión más clara sobre la importancia de implementar la inteligencia artificial de forma segura, responsable y controlada dentro de los entornos hospitalarios.

Planteamiento del Problema

En el mundo de la medicina ha surgido la necesidad de obtener diagnósticos más rápidos y precisos. En este contexto, el uso de sistemas de inteligencia artificial (IA) ha transformado la manera de generar dichos diagnósticos. Sin embargo, estos sistemas también presentan vulnerabilidades frente a ataques adversariales, los cuales consisten en la manipulación intencional de imágenes y datos con el fin de alterar los resultados diagnósticos.

El objetivo de estos ataques es comprometer tanto la validez diagnóstica como la seguridad de los datos clínicos. Por un lado, pueden generar interpretaciones erróneas, que conduzcan a decisiones medicas inadecuadas, afectando directamente al paciente. Por otro lado, ponen en riesgo la confidencialidad y la integridad de la información clínica compartida en los sistemas intrahospitalarios, lo que puede disminuir la confianza en las tecnologías basadas en IA.

El problema se centra en la limitada existencia de mecanismos de defensa suficientemente robustos y estandarizados para proteger los sistemas de inteligencia artificial aplicados a la radiología frente a ataques adversariales. En el contexto actual, donde múltiples instituciones de salud han incorporado herramienta de IA para apoyar el diagnóstico clínico, la vulnerabilidad de estos sistemas representa un riesgo significativo para la validez diagnostica y la seguridad del paciente.

La manipulación intencional de imágenes médicas puede alterar los resultados generados por los algoritmos, comprometiendo la confiabilidad de las decisiones clínicas. En consecuencia, resulta necesario analizar el impacto de los ataques adversariales en el flujo de trabajo radiológico y determinar que estrategias puedan implementarse para fortalecer la protección de los datos sensibles y la integridad diagnóstica.

En este contexto se hace necesario formular la siguiente pregunta de investigación: ¿Cómo afectan los ataques adversariales la validez del diagnóstico radiológico basado en la inteligencia artificial y que mecanismos de defensa pueden implementarse para fortalecer la seguridad y protección de los datos sensibles en entornos hospitalarios?

Justificación

La transformación digital del sector salud y la incorporación de sistemas de inteligencia artificial (IA) en los procesos diagnósticos basados en imágenes médicas han generado avances significativos en términos de rapidez, precisión y apoyo a la toma de decisiones clínicas. No obstante, estas tecnologías también presentan vulnerabilidades frente a ataques adversariales, los cuales consisten en la manipulación intencional de imágenes y datos clínicos con el propósito de alterar los resultados emitidos por lo algoritmos.

Las imágenes diagnósticas, como radiografías, tomografías computarizadas y resonancias magnéticas, constituyen insumos fundamentales para la toma de decisiones médicas. Cuando estas son modificadas de manera maliciosa para engañar los modelos de IA, pueden producirse diagnósticos erróneos, lo que representa un riesgo directo para la seguridad del paciente. En este sentido, la problemática no solo involucra un desafío técnico, sino también ético y clínico.

Asimismo, la manipulación de datos médicos compromete la integridad, confidencialidad y disponibilidad de la información clínica, principios fundamentales de la ciberseguridad en entornos hospitalarios. En sistemas altamente interconectados un ataque exitoso puede afectar no solo a un paciente individual, sino a toda una institución sanitaria disminuyendo la confianza en las tecnologías implementadas.

Por lo tanto, resulta necesario fortalecer los mecanismos de defensa y desarrollar estrategias de mitigación que permitan identificar, prevenir y contrarrestar las vulnerabilidades de los sistemas de IA radiológica frente a ataques adversariales, garantizando así la seguridad del paciente y la confiabilidad de los procesos diagnósticos.

Objetivos

Objetivo General

Determinar el impacto de los ataques adversariales en los sistemas de inteligencia artificial aplicados a la radiología a partir de la revisión documental, considerando su efecto en la validez diagnóstica y la seguridad de los datos clínicos.

Objetivos Específicos

Identificar las principales vulnerabilidades de los sistemas de inteligencia artificial en radiología frente a ataques adversariales, con base en la literatura científica.

Describir el funcionamiento de la inteligencia artificial en el flujo de trabajo radiológico y su relación con el proceso diagnóstico.

Analizar las consecuencias clínicas, técnicas y éticas derivadas de la manipulación de imágenes médicas y datos sensibles.

Examinar las estrategias de ciberseguridad y mecanismos de protección de datos reportados en la literatura para mitigar estos riesgos.

Marco Teórico

Transformación Digital de la Radiología y Consolidación de la Inteligencia Artificial

En las últimas décadas, la radiología ha experimentado una transformación estructural impulsada por la digitalización de los sistemas de adquisición, almacenamiento y transmisión de imágenes médicas. Este proceso no solo ha optimizado la gestión del flujo de trabajo, sino que ha abierto el camino para la incorporación de herramientas basadas en inteligencia artificial (IA) como apoyo al diagnóstico clínico. La transición desde modelos analógicos hacia entornos digitales fue acompañada por lineamientos técnicos orientados a la gestión de dosis, calidad de imagen y seguridad del paciente, como lo estableció la International Commission on Radiological Protection (2003/2014) al enfatizar la necesidad de equilibrio entre optimización diagnóstica y protección radiológica.

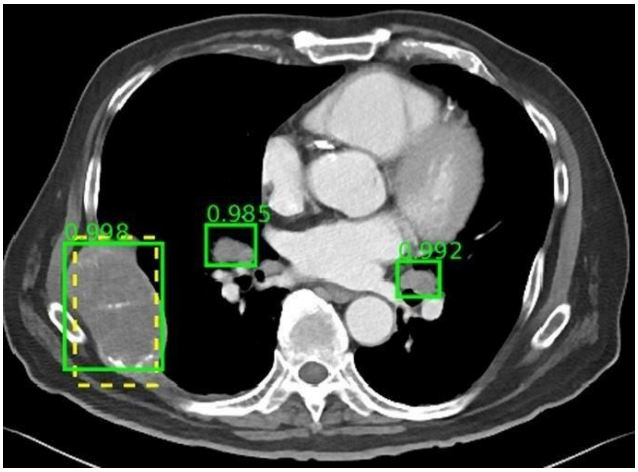
En este escenario de modernización, la IA ha sido descrita como uno de los avances más disruptivos en imagenología médica (Puentes et al., 2021; Gálvez, 2017). Su incorporación responde a la creciente demanda de diagnósticos oportunos, particularmente en servicios de urgencias, donde la rapidez en la interpretación puede incidir directamente en la toma de decisiones críticas (López Izquierdo, 2025). Katzman et al. (2023) sostienen que los algoritmos de aprendizaje profundo han demostrado utilidad en la priorización automática de estudios, la detección de hallazgos incidentales y la reducción de tiempos de respuesta en radiología de emergencia.

Desde una perspectiva latinoamericana, Rodríguez et al. (2023) señalan que la integración de nuevas tecnologías en radiología no solo modifica los procesos asistenciales, sino también redefine las competencias profesionales del radiólogo. Esta postura coincide con

Marangoni (2018), quien plantea que la IA no debe entenderse como una amenaza de sustitución, sino como un desafío adaptativo que exige actualización permanente y pensamiento crítico.

Figura 1

TC Tórax Corte Axial



Nota. Tomografía computarizada de tórax en la que un sistema de inteligencia artificial identifica y resalta automáticamente posibles anomalías pulmonares mediante recuadros y probabilidades de detección, apoyando al radiológico en el análisis de la imagen. Fuente. Inteligencia Artificial en Radiología, *por De La Cámara Egea, 2019*, Radiología Club.

Fundamentos Conceptuales de la Inteligencia Artificial en Imagen Médica

La inteligencia artificial aplicada a la radiología se fundamenta principalmente en técnicas de machine learning y deep learning, capaces de identificar patrones complejos en grandes volúmenes de datos (Aguirre et al., 2021; Machacado Rojas & Aparicio Pico, 2021). Estas tecnologías permiten entrenar modelos predictivos mediante bases de datos etiquetadas, generando sistemas capaces de clasificar, segmentary detectar anomalías con altos niveles de sensibilidad.

Acosta-Jiménez et al. (2023) destacan que las aplicaciones actuales abarcan desde la detección temprana de lesiones pulmonares hasta el análisis automatizado de estructuras

neuroanatómicas. En el campo específico de la tomografía computarizada, Higuera Mosquera et al. (2025) evidencian mejoras en la precisión diagnóstica cuando la IA se emplea como herramienta complementaria al criterio clínico. Asimismo, Nallino et al. (2024) documentan avances significativos en neuroimágenes, particularmente en la identificación precoz de alteraciones estructurales sutiles.

En términos de validación clínica, Silva Afonso et al. (2025) analizaron la fiabilidad de un sistema asistido por IA en radiografías de tórax y óseas en un servicio de urgencias hospitalario, concluyendo que, aunque los resultados fueron prometedores, la supervisión humana continúa siendo indispensable para garantizar la validez diagnóstica. Esta idea es reforzada por Castillo López et al. (2026), quienes sostienen que la IA debe concebirse como complemento y no como reemplazo del especialista.

Figura 2

MRI del Cerebro, Corte axial, Coronal y Sagital

Nota. La imagen presenta líneas de referencia y mediciones que son utilizadas por sistemas de inteligencia artificial para analizar estructuras cerebrales, detectar anomalías y apoyar el diagnóstico radiológico. Fuente. Diagnóstico por imágenes con inteligencia artificial: cómo la inteligencia artificial está transformando la atención médica, *por Quibim, 2025*, Quibim Website.

IA y Optimización del Flujo de Trabajo Radiológico

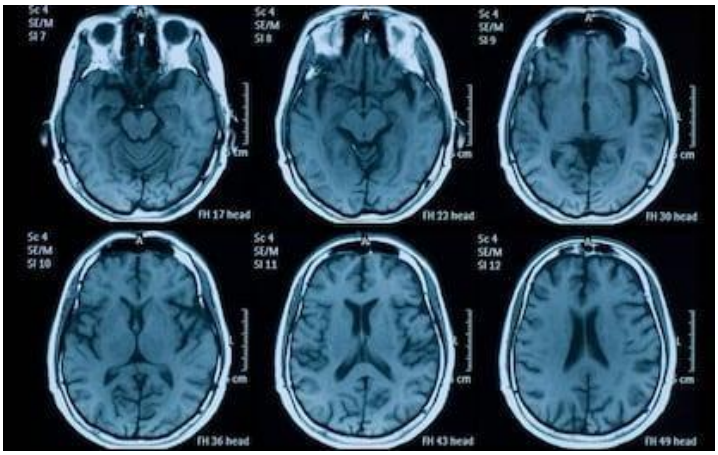
El concepto de "radiología 2.0" alude a un entorno digital interconectado donde la IA interviene en múltiples etapas del proceso diagnóstico (Parada et al., 2023). Pierre et al. (2023) describen la "radiology roundtrip" como un ciclo integral que incluye programación del estudio, adquisición de imagen, procesamiento, interpretación, informe y almacenamiento. En cada una de estas fases, los algoritmos pueden intervenir para optimizar tiempos y reducir errores humanos.

Narváez Pereira et al. (2024) resaltan el papel de la IA en el control de calidad y detección automática de artefactos, contribuyendo a mejorar la confiabilidad técnica de las imágenes. Díaz Ramírez et al. (2025) agregan que los algoritmos pueden ajustar parámetros técnicos de adquisición en radiografías de tórax, favoreciendo una mayor consistencia en los resultados.

No obstante, Jiménez-Rodríguez et al. (2024) advierten que la calidad y seguridad en los servicios de diagnóstico por imagen dependen de procesos sistemáticos de evaluación y estandarización, lo cual implica que la implementación tecnológica debe acompañarse de protocolos de validación rigurosos. Trapero y López (2017) complementan esta visión al subrayar la importancia de la gestión del ciclo de vida tecnológico, aspecto crucial para evitar obsolescencia y vulnerabilidades operativas.

Figura 3

MRI del Cerebro



Nota. La imagen presenta un derrame detectado por un sistema de IA en el tomógrafo. Fuente. Diagnóstico por imágenes con inteligencia artificial: cómo la inteligencia artificial está transformando la atención médica, *por Quibim, 2025*, Quibim Website.

Riesgos Asociados a la Implementación de IA en Radiología

Si bien el uso de la inteligencia artificial para mejorar la precisión diagnóstica mediante imágenes médicas radiológicas ofrece numerosas ventajas, también presenta algunos aspectos negativos potenciales. Estos costos pueden analizarse desde tres perspectivas distintas: técnica, ética y profesional. En primer lugar, la mayoría de estos costos se deben a la calidad de los datos obtenidos para el desarrollo del algoritmo, a los diversos parámetros involucrados en la creación de un modelo predictivo mediante aprendizaje automático y a la eficacia de las decisiones tomadas en conjunto con dichos modelos predictivos para lograr un diagnóstico más preciso en casos futuros.

El sesgo algorítmico es uno de los principales riesgos citados en la literatura, ya que los sistemas de IA se entrenan con bases de datos que no reflejan la verdadera diversidad de las poblaciones del mundo real. Cuando hay muchos más individuos de ciertos grupos demográficos

en un conjunto de datos de entrenamiento, o cuando hay más individuos con muchas patologías y afecciones clínicas representadas en el conjunto de datos de entrenamiento que con esos grupos o patologías, no se aprenderá lo suficiente sobre esos individuos para producir algoritmos precisos que los asignen a cada categoría.

La baja representatividad en el conjunto de datos de entrenamiento da como resultado que los algoritmos produzcan interpretaciones inexactas de las manifestaciones clínicas en esos grupos demográficos, lo que dificulta el diagnóstico equitativo de las poblaciones no representadas. Este problema en la IA es especialmente relevante para las aplicaciones médicas. Galarza Medina et al. (2023) enfatizan la importancia de la representatividad de los datos para evitar sesgos en las aplicaciones médicas de la IA.

Otro riesgo importante reside en la opacidad de los algoritmos, especialmente en el aprendizaje profundo. Muchos sistemas de IA funcionan como redes neuronales interconectadas cada vez más complejas, que ayudan a identificar patrones en los datos de imágenes médicas. Sin embargo, el proceso interno de toma de decisiones, o cómo el sistema llegó a una conclusión diagnóstica específica, no siempre es completamente comprensible para los profesionales sanitarios. El problema de la caja negra, tal como se describe en la literatura, impide que los profesionales comprendan realmente por qué el sistema proporcionó un diagnóstico específico.

Esta falta de transparencia puede mermar la confianza del profesional y, por lo tanto, limitar la validación clínica de los resultados del sistema. También existe un riesgo potencial asociado a una dependencia excesiva de la automatización. En particular, cuando la tecnología se integra demasiado en el proceso de diagnóstico, surge la incapacidad (o reticencia) de algunos profesionales para analizar críticamente o delegar en exceso la interpretación inicial de imágenes a un producto automatizado.

Iglesias López (2023) indica que, ante la falta de capacitación adecuada sobre la aplicación y la comprensión de estas herramientas, los profesionales pueden aceptar las recomendaciones del algoritmo como correctas, independientemente de cualquier evaluación realizada, lo que aumenta el riesgo de diagnósticos erróneos en situaciones donde el algoritmo presenta limitaciones. Existen diversos riesgos que podrían derivar en una brecha de seguridad de datos o en la confidencialidad de los pacientes.

Los sistemas de IA deben entrenarse y operar con grandes cantidades de datos médicos, por ejemplo, imágenes diagnósticas, historiales clínicos y otros tipos de datos sensibles. Cuando un sistema de IA no cuenta con los protocolos adecuados para proteger o anonimizar los datos que utiliza, existen vulnerabilidades que podrían provocar una brecha en la confidencialidad de los datos médicos. Esto es especialmente importante en hospitales, donde existen múltiples plataformas digitales interconectadas. Los errores derivados de la interferencia de la inteligencia artificial en su funcionamiento a nivel clínico pueden provocar falsos positivos o falsos negativos en la detección de patologías.

Los falsos positivos pueden conllevar pruebas excesivas e innecesarias, un aumento de los costes generales de atención y ansiedad en la persona examinada. Por otro lado, los falsos negativos pueden impedir el diagnóstico de una afección existente, lo que podría retrasar el inicio del tratamiento y, por consiguiente, afectar al pronóstico del paciente. Así pues, según Sotelo Ramírez (2023), independientemente del grado de automatización del sistema utilizado para el diagnóstico, la responsabilidad final del mismo recae en el médico.

En consecuencia, la adopción de la IA en radiología debe considerarse una herramienta que facilite la toma de decisiones informadas, en lugar de un sustituto del juicio clínico del profesional sanitario. Según Álvarez Córdova (2025), el rol del radiólogo está evolucionando

progresivamente hacia una función de supervisión, validando los resultados derivados de algoritmos informáticos y proporcionando una interpretación contextual de los mismos. Esto exige que los especialistas no solo sean capaces de interpretar imágenes médicas, sino que también comprendan las limitaciones, el alcance y los posibles riesgos asociados al uso de estas herramientas.

Por lo tanto, la incorporación responsable de la inteligencia artificial en radiología requiere estrategias que incluyan la validación clínica continua de los algoritmos, el desarrollo de bases de datos representativas, la capacitación del personal médico y la implementación de marcos éticos y regulatorios que garanticen la seguridad del paciente y la calidad del diagnóstico.

Ciberseguridad, Ataques Adversariales y Vulnerabilidad del Diagnóstico

Gracias a los avances tecnológicos, los radiólogos ahora pueden adquirir y almacenar imágenes digitales de los pacientes y utilizarlas para realizar diagnósticos. La digitalización aumentó la eficiencia de los médicos al permitirles trabajar con redes interconectadas de servidores de sistemas de archivo y comunicación de imágenes (PACS), redes internas del hospital, algoritmos de inteligencia artificial (IA) y plataformas de computación en la nube. Sin embargo, esta mayor integración también incrementó el número de posibles vectores de ataque. Por lo tanto, la ciberseguridad es fundamental para mantener la continuidad de las operaciones clínicas y salvaguardar los datos de los pacientes.

Los ciberataques dirigidos a los sistemas hospitalarios que comprometen tanto la disponibilidad como la integridad de los sistemas de información médica constituyen uno de los riesgos asociados a la infraestructura digital en radiología. Nguyen (2025) informa de varios casos de ciberataques que interrumpieron el funcionamiento de plataformas de diagnóstico o bloquearon el acceso a bases de datos clínicas a través de los servicios de radiología. Muchos de

estos incidentes involucraron malware o ransomware, lo que provocó la imposibilidad de los profesionales para acceder a los estudios radiológicos, retrasó las decisiones clínicas y afectó la atención al paciente. Además, estos ciberataques pueden provocar la exposición o filtración de datos sensibles y, por lo tanto, representan un importante riesgo para la ciberseguridad debido a su impacto en la confidencialidad de la información sanitaria.

Las instituciones radiológicas deben implementar medidas de protección tecnológica y desarrollar las competencias de su personal en materia de ciberseguridad a nivel institucional. Según las recomendaciones del OIEA, Strohal (2023) afirma que la formación del personal en seguridad digital es un elemento esencial de la gestión integral de riesgos para las instalaciones radiológicas y nucleares. La capacitación del personal en el manejo seguro de la información, la detección de amenazas y los protocolos de respuesta ante incidentes de ciberseguridad contribuye a mitigar las vulnerabilidades que podrían ser explotadas por amenazas externas.

Así como existen riesgos para las redes y sistemas informáticos utilizados con fines de seguridad, también existen riesgos relacionados con el uso de la inteligencia artificial en radiología, específicamente en torno a las vulnerabilidades asociadas con el funcionamiento de las herramientas de análisis de imágenes. Una de estas vulnerabilidades se denomina «ataque adversario» e implica la manipulación deliberada de los datos de entrada que se introducen en los modelos de inteligencia artificial para modificar su resultado.

En el caso de la radiología, estas manipulaciones pueden consistir en ajustes mínimos a las propias imágenes médicas, que resultan indetectables para el ojo humano, pero suficientes para afectar la forma en que los algoritmos las interpretan. Al realizar estas modificaciones en una imagen, es posible que un sistema automatizado la clasifique erróneamente, lo que genera graves problemas en cuanto a la fiabilidad de los sistemas automatizados utilizados para este fin.

Los ataques adversarios son relevantes porque explotan las propiedades internas del aprendizaje automático. Los modelos de IA (inteligencia artificial) pueden aprender a identificar patrones complejos en grandes cantidades de datos; sin embargo, también pueden ser fácilmente engañados por pequeñas perturbaciones realizadas deliberadamente para confundirlos. Según Berrones (2024), si bien los modelos de análisis de imágenes han demostrado una excelente capacidad para proporcionar resultados precisos en condiciones de laboratorio, su implementación en entornos clínicos reales requerirá considerar formas de fortalecer su resistencia a la corrupción o los errores derivados de la manipulación de sus datos de entrada.

Alterar una imagen diagnóstica, ya sea accidental o intencionadamente, añade un riesgo y una vulnerabilidad importantes al proceso diagnóstico. En radiología, los médicos utilizan la interpretación precisa de los hallazgos visuales en las imágenes médicas para tomar decisiones relacionadas con la atención del paciente. Si un algoritmo, entrenado para interpretar incorrectamente una imagen modificada de forma malintencionada, se pueden producir diagnósticos erróneos.

Por otra parte, la dimensión de la protección de datos médicos constituye un aspecto central en el debate sobre la seguridad digital en radiología. Las imágenes diagnósticas y los registros clínicos asociados contienen información altamente sensible que debe ser protegida conforme a principios de confidencialidad y privacidad reconocidos en normativas internacionales de protección de datos sanitarios. La alteración, acceso no autorizado o filtración de estos datos no solo implica riesgos éticos y legales, sino que también puede afectar la confianza de los pacientes en los sistemas de salud digitales.

En este escenario, la implementación segura de inteligencia artificial en radiología requiere una aproximación integral que combine medidas tecnológicas, protocolos de seguridad

informática y supervisión profesional. La adopción de sistemas robustos frente a ataques adversariales, junto con estrategias de ciberseguridad institucional y capacitación especializada del personal, constituye un elemento clave para reducir la vulnerabilidad del diagnóstico asistido por inteligencia artificial y garantizar que el uso de estas tecnologías contribuya efectivamente a mejorar la calidad y seguridad de la atención médica.

Protección de Datos Sensibles y Estándares Internacionales

La gestión de información radiológica implica el tratamiento de datos altamente sensibles. La ICRP (2003/2014) ya advertía que la digitalización requiere sistemas de control y trazabilidad para preservar la integridad de los registros. En la actualidad, la discusión se amplía hacia la necesidad de marcos normativos que armonicen innovación tecnológica con salvaguarda de derechos fundamentales.

Domínguez (s.f.) sostiene que la implementación de IA debe acompañarse de políticas claras sobre gobernanza de datos, anonimización y control de accesos. Asimismo, Rosales Chaves y Ramírez Morales (2024) destacan que la radiología personalizada y molecular incrementa la complejidad del manejo de información, reforzando la necesidad de estándares internacionales robustos.

La convergencia entre IA, ciberseguridad y ética clínica configura un nuevo paradigma donde la protección de datos no es un elemento accesorio, sino un requisito estructural del sistema.

Síntesis Conceptual del Riesgo Adversarial en el Flujo Radiológico

La literatura revisada coincide en que la IA ha transformado profundamente la radiología, mejorando eficiencia, precisión y capacidad predictiva (Martí-Bonmatí, 2024; Revista Ocronos, 2023; Gallego Piña et al., 2025). Sin embargo, también evidencia que la incorporación de

algoritmos en el flujo de trabajo amplía la superficie de ataque digital y plantea interrogantes sobre la resiliencia de los sistemas frente a manipulaciones intencionadas.

El riesgo asociado a ataques adversariales puede entenderse como la probabilidad de que una alteración maliciosa en los datos de entrada comprometa la integridad diagnóstica y la seguridad informacional. Este riesgo se ve amplificado en entornos hospitalarios interconectados donde convergen múltiples dispositivos y bases de datos.

En consecuencia, el análisis del impacto de estos ataques requiere integrar tres dimensiones teóricas:

La dimensión tecnológica, relacionada con la arquitectura y robustez de los algoritmos.

La dimensión clínica, vinculada a la validez diagnóstica y seguridad del paciente.

La dimensión ética y normativa, asociada a la protección de datos y cumplimiento de estándares internacionales.

Desde esta perspectiva, el marco teórico permite comprender que la innovación tecnológica en radiología, aunque prometedora, exige una evaluación crítica de sus vulnerabilidades. La consolidación de mecanismos de defensa robustos, protocolos de ciberseguridad y formación especializada constituye un reto ineludible para garantizar que la inteligencia artificial continúe siendo una herramienta de apoyo confiable y segura en el diagnóstico médico.

|Teoría de la Robustez Algorítmica y Resiliencia de Sistemas Inteligentes

El desarrollo de sistemas de inteligencia artificial aplicados a la radiología se fundamenta en modelos matemáticos capaces de aprender representaciones complejas a partir de grandes volúmenes de datos clínicos (Aguirre et al., 2021; Machacado Rojas & Aparicio Pico, 2021). Sin embargo, la teoría contemporánea de la robustez algorítmica sostiene que todo modelo predictivo

es vulnerable cuando pequeñas perturbaciones en los datos de entrada generan cambios significativos en la salida del sistema.

En el ámbito radiológico, esta vulnerabilidad adquiere especial relevancia debido a que los algoritmos operan sobre imágenes médicas que pueden ser modificadas de manera imperceptible al ojo humano, pero suficientemente significativas para alterar la clasificación automática. Desde la perspectiva técnica, la robustez se define como la capacidad del sistema para mantener su desempeño frente a variaciones intencionales o no intencionales en los datos (Berrones Berrones, 2024).

Pierre et al. (2023) señalan que la automatización del "radiology roundtrip" incrementa la eficiencia operativa, pero también amplía la superficie de exposición a interferencias digitales. En ese sentido, la resiliencia algorítmica implica no solo mejorar el rendimiento diagnóstico, sino diseñar modelos capaces de detectar anomalías, identificar inconsistencias y resistir manipulaciones adversariales.

Silva Afonso et al. (2025) evidencian que incluso sistemas con altos índices de sensibilidad y especificidad requieren validación continua en entornos clínicos reales. Esta observación se vincula con la teoría de la confiabilidad tecnológica, según la cual ningún sistema automatizado debe considerarse infalible sin mecanismos de auditoría y supervisión humana (Castillo López et al., 2026).

Asimismo, Galarza Medina et al. (2023) advierten que los modelos entrenados con bases de datos limitadas o sesgadas pueden presentar comportamientos impredecibles ante escenarios no contemplados durante el entrenamiento. Esta condición incrementa la vulnerabilidad frente a ataques adversariales, dado que el algoritmo podría amplificar errores derivados de alteraciones mínimas en la imagen.

Desde un enfoque sistémico, la resiliencia tecnológica implica incorporar capas de protección, tales como validaciones cruzadas, monitoreo continuo del rendimiento y protocolos de actualización segura (Trapero & López, 2017). La robustez no se limita al diseño inicial del modelo, sino que requiere una gestión integral del ciclo de vida del sistema.

En consecuencia, la teoría de la robustez algoritmica aporta un fundamento conceptual clave para comprender cómo los ataques adversariales pueden comprometer la validez diagnóstica y por qué resulta imprescindible fortalecer mecanismos preventivos y correctivos dentro del flujo radiológico digital.

Teoría Sociotécnica del Riesgo en Entornos Clínicos Digitalizados

La integración de inteligencia artificial en radiología no constituye únicamente un fenómeno tecnológico, sino un proceso sociotécnico en el que interactúan personas, dispositivos, protocolos y estructuras organizacionales. La teoría sociotécnica del riesgo plantea que los eventos adversos emergen de la interacción compleja entre componentes humanos y tecnológicos, más que de fallas aisladas (Puentes et al., 2021).

Desde esta perspectiva, la seguridad diagnóstica depende tanto del desempeño algoritmico como de la cultura organizacional, la capacitación del personal y los protocolos institucionales. Strohal (2023) enfatiza que la formación en ciberseguridad es un elemento estructural en instalaciones radiológicas, dado que la vulnerabilidad no reside únicamente en el software, sino también en prácticas operativas inadecuadas.

Nguyen et al. (2025) documentan que los ciberataques en radiología suelen explotar debilidades humanas, como contraseñas inseguras o falta de actualización de sistemas, evidenciando que el riesgo digital es multidimensional. Esta visión coincide con Jiménez-

Rodríguez et al. (2024), quienes subrayan que la calidad y seguridad en los servicios de diagnóstico por imagen requieren procesos integrales de gestión.

Álvarez Córdova (2025) argumenta que el rol del radiólogo se redefine en un entorno donde la IA participa activamente en la interpretación de estudios. La teoría sociotécnica sostiene que la confianza en la tecnología debe construirse mediante interacción crítica y supervisión constante, evitando tanto la dependencia excesiva como la resistencia injustificada (Sotelo Ramírez, 2023).

Además, la digitalización masiva incrementa la interdependencia entre sistemas hospitalarios. La International Commission on Radiological Protection (2003/2014) ya advertía que la gestión digital exige controles estrictos para preservar la integridad de los registros clínicos. En el contexto actual, dicha advertencia adquiere mayor relevancia debido a la complejidad de los entornos interconectados.

La teoría sociotécnica permite comprender que un ataque adversarial no solo afecta un algoritmo, sino que puede desencadenar consecuencias organizacionales, clínicas y éticas. Una alteración en la imagen diagnóstica puede repercutir en decisiones terapéuticas, generar responsabilidad legal y erosionar la confianza institucional.

|Teoría de la Gobernanza de Datos y Responsabilidad Algorítmica

Otro contexto complementario es la teoría de la gobernanza de datos, la cual sostiene que la gestión de información sensible debe basarse en principios de transparencia, trazabilidad y rendición de cuentas. En radiología, los datos incluyen imágenes, informes y metadatos clínicos que requieren protección reforzada (Domínguez, s.f.).

Martí-Bonmatí (2024) indica que la expansión de la IA en imagen médica obliga a establecer marcos regulatorios claros que delimiten responsabilidades y garanticen el uso ético

de la información. Rosales Chaves y Ramírez Morales (2024) agregan que la personalización diagnóstica, potenciada por algoritmos, incrementa el volumen y sensibilidad de los datos tratados.

La responsabilidad algorítmica implica que las instituciones deben ser capaces de explicar cómo se generan los resultados diagnósticos y cómo se protegen frente a manipulaciones externas (Iglesias López, 2023). Desde esta óptica, la gobernanza no se limita a la protección contra filtraciones, sino que abarca la prevención de alteraciones maliciosas que comprometan la integridad de la imagen.

En este sentido, la teoría de gobernanza de datos se articula directamente con el problema de investigación planteado, ya que los ataques adversariales no solo representan una amenaza técnica, sino una vulneración potencial de principios éticos y normativos internacionales relacionados con la confidencialidad y seguridad del paciente.

Marco Metodológico

Tipo, Diseño y Enfoque de la Investigación

Tipo de Investigación

La presente investigación se enmarca en un tipo descriptivo-analítico, ya que busca identificar y caracterizar las vulnerabilidades asociadas a los ataques adversariales en sistemas de inteligencia artificial aplicados a la radiología, así como analizar sus implicaciones en la validez diagnóstica y la seguridad de los datos clínicos.

El enfoque descriptivo permite detallar las características del fenómeno estudiado, mientras que el componente analítico facilita comprender las relaciones entre las variables involucradas, particularmente entre la manipulación de imágenes médicas y sus efectos en los resultados diagnósticos (Hernández Sampieri et al., 2014).

Diseño de Investigación

El estudio se desarrolla bajo un diseño documental o de revisión bibliográfica, dado que se basa en la recopilación, análisis e interpretación de información proveniente de fuentes científicas, artículos académicos y documentos especializados en inteligencia artificial, radiología y ciberseguridad.

Este diseño es pertinente debido a que el fenómeno de los ataques adversariales en entornos clínicos digitales ha sido ampliamente abordado en la literatura científica, permitiendo construir un análisis fundamentado sin necesidad de intervención experimental directa (Arias, 2012).

Además, el diseño documental permite identificar tendencias, vacíos de conocimiento y estrategias propuestas en estudios previos, lo cual es clave para sustentar las conclusiones del trabajo.

Enfoque de la Investigación

La investigación adopta un enfoque cualitativo, ya que se centra en la interpretación y análisis de información teórica y conceptual relacionada con el uso de inteligencia artificial en radiología y sus vulnerabilidades frente a ataques adversariales.

Este enfoque permite comprender el fenómeno desde una perspectiva integral, considerando no solo aspectos técnicos, sino también implicaciones clínicas, éticas y de seguridad de la información.

Según Creswell (2014), el enfoque cualitativo es adecuado cuando se busca interpretar fenómenos complejos en su contexto natural, lo cual resulta pertinente en el análisis de tecnologías emergentes como la inteligencia artificial en salud.

Fases del Diseño Metodológico

La investigación se desarrolló en las siguientes fases:

Fase 1: Búsqueda y recolección de Información

Se realizó una búsqueda sistemática de literatura científica en bases de datos académicas, seleccionando estudios relacionados con inteligencia artificial en radiología, ataques adversariales y ciberseguridad en salud.

Fase 2: Selección de Fuentes

Se filtraron los documentos encontrados mediante criterios de inclusión y exclusión previamente definidos, garantizando la pertinencia y calidad de la información.

Fase 3: Análisis de la Información

Se realizó una lectura crítica de los documentos seleccionados, identificando categorías como vulnerabilidades de la IA, impacto en el diagnóstico, riesgos en seguridad de datos, estrategias de mitigación.

Fase 4: Síntesis e Interpretación

Se integraron los hallazgos en un análisis estructurado que permitió responder la pregunta de investigación y sustentar los resultados del estudio.

Fuentes de Información y Bases de Datos Consultadas

La recolección de información se realizó mediante la consulta de bases de datos académicas y científicas reconocidas a nivel nacional e internacional, con el propósito de garantizar la calidad y confiabilidad de las fuentes utilizadas. Entre las principales bases de datos consultadas se encuentran PubMed, Scopus, ScienceDirect y Google Scholar, así como repositorios institucionales de universidades y organismos académicos. Estas plataformas permiten el acceso a artículos científicos indexados, revisiones sistemáticas y documentos especializados en áreas como inteligencia artificial, radiología y ciberseguridad. La selección de estas fuentes responde a su rigor científico y a la pertinencia temática frente al objeto de estudio.

Estrategia de Búsqueda

La estrategia de búsqueda se estructuró a partir de la definición de palabras clave o descriptores relacionados con el tema de investigación, tanto en español como en inglés, con el fin de ampliar el alcance de los resultados. Se emplearon términos como “inteligencia artificial”, “radiología”, “ataques adversariales”, “ciberseguridad” y “diagnóstico médico”, junto con sus equivalentes en inglés. Estos descriptores fueron combinados mediante operadores booleanos (AND, OR), permitiendo construir ecuaciones de búsqueda más precisas y específicas. Este proceso facilitó la identificación de literatura relevante, asegurando la inclusión de estudios directamente relacionados con el problema de investigación.

Tipo de Artículos Seleccionados

Para el desarrollo de la investigación se priorizó la selección de artículos científicos provenientes de revistas indexadas, revisiones sistemáticas, estudios de investigación aplicada y documentos académicos relevantes. Este tipo de publicaciones garantiza un alto nivel de rigurosidad metodológica y validez científica. Asimismo, se incluyeron trabajos que abordan la implementación de inteligencia artificial en radiología, los riesgos asociados a ataques adversariales y las estrategias de ciberseguridad en entornos clínicos. La elección de estos documentos permitió construir un análisis fundamentado y actualizado del fenómeno estudiado.

Criterios de Inclusión y Exclusión

Los criterios de inclusión se establecieron con el fin de garantizar la pertinencia y calidad de la información recopilada. Se consideraron artículos publicados en los últimos diez años, en idioma español e inglés, que abordaran temas relacionados con inteligencia artificial, radiología, ciberseguridad y ataques adversariales. Asimismo, se priorizaron documentos con respaldo académico, publicados en revistas indexadas o provenientes de fuentes institucionales confiables. Por otro lado, se excluyeron aquellos documentos que no contaran con revisión por pares, que provinieran de fuentes no académicas o que no guardaran relación directa con el objeto de estudio. También se descartaron publicaciones fuera del rango temporal establecido, salvo aquellas consideradas fundamentales desde el punto de vista teórico.

Técnica de Análisis de la Información

La información recopilada fue analizada mediante la técnica de análisis de contenido cualitativo, la cual permite interpretar y organizar la información de manera sistemática. A través de este proceso, se identificaron categorías relevantes como las vulnerabilidades de los sistemas de inteligencia artificial, el impacto en la validez diagnóstica, los riesgos en la seguridad de los

datos y las estrategias de mitigación propuestas en la literatura. Este análisis permitió establecer relaciones entre los diferentes hallazgos, facilitando una comprensión integral del fenómeno estudiado y aportando sustento teórico a las conclusiones de la investigación.

Resultados

A partir del análisis documental realizado, se evidenció que la inteligencia artificial aplicada a la radiología representa una herramienta de apoyo importante para mejorar la rapidez, precisión y eficiencia del diagnóstico por imágenes. Sin embargo, los autores consultados coinciden en que su implementación también introduce riesgos relevantes, especialmente cuando los sistemas son expuestos a ataques adversariales capaces de alterar imágenes médicas o datos clínicos de forma casi imperceptible para el ojo humano, pero suficiente para modificar la respuesta del algoritmo.

Uno de los principales resultados encontrados es que los ataques adversariales pueden afectar directamente la validez diagnóstica. La manipulación de una imagen radiológica puede generar falsos positivos o falsos negativos, lo cual representa un riesgo clínico significativo. En el caso de un falso positivo, el paciente podría ser sometido a exámenes innecesarios, tratamientos no requeridos o mayores costos en salud. En cambio, un falso negativo podría retrasar el diagnóstico real de una enfermedad, afectando el pronóstico y la oportunidad del tratamiento.

Tabla 1

Principales Riesgos Identificados

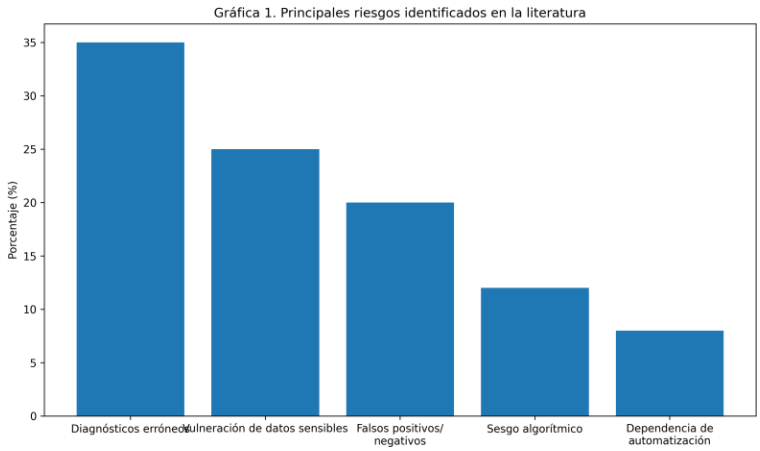
Autor	Riesgo identificado	Consecuencia principal	Impacto en el paciente o institución
Berrones Berrones (2024)	Manipulación adversarial de imágenes	Diagnósticos erróneos	Riesgo para la seguridad del paciente
Sotelo Ramírez (2023)	Falsos positivos	Exámenes o tratamientos innecesarios	Aumento de costos y ansiedad del paciente
Sotelo Ramírez (2023)	Falsos negativos	Retraso en el diagnóstico	Progresión de la enfermedad

Iglesias López (2023)	Sesgo algorítmico	Resultados poco confiables en ciertos grupos	Diagnóstico desigual o poco equitativo
Iglesias López (2023)	Falta de explicabilidad del algoritmo	Dificultad para comprender la decisión de la IA	Menor confianza del profesional
Strohal (2023)	Brechas de ciberseguridad	Exposición o alteración de datos sensibles	Riesgo legal, ético e institucional

Nota. Elaboración propia.

Figura 4

Principales Riesgos Identificados



Nota. Los diagnósticos erróneos representan un factor fundamental en los errores del uso de IA en la interpretación radiológica, favoreciendo tratamientos inadecuados y fallas graves en los servicios. Elaboración propia.

La grafica muestra los principales riesgos encontrados en la literatura científica sobre inteligencia artificial aplicada a la radiología. Se observa que los diagnósticos erróneos representan el mayor porcentaje, debido a que una alteración adversarial en la imagen puede llevar a falsos resultados clínicos que afectan directamente la seguridad del paciente.

En segundo lugar, aparece la vulneración de datos sensibles, lo cual demuestra que el problema no solo es clínico, sino también ético, legal e institucional. La filtración o alteración de información médica puede comprometer la confianza en los sistemas hospitalarios digitalizados.

Los falsos positivos y falsos negativos también representan una preocupación importante, ya que pueden generar tratamientos innecesarios o retrasar diagnósticos reales.

Otro resultado relevante es que la vulnerabilidad de la IA no depende únicamente del algoritmo, sino también del contexto hospitalario donde se utiliza. Los sistemas de radiología actuales están conectados a plataformas digitales, PACS, bases de datos clínicas, servidores y redes internas, lo cual amplía la superficie de ataque. Esto significa que un problema de ciberseguridad no afecta solo a una imagen aislada, sino que puede comprometer el flujo completo de trabajo radiológico, desde la adquisición de la imagen hasta el informe diagnóstico y el almacenamiento de la información.

También se encontró que la literatura revisada plantea una preocupación constante frente a la dependencia excesiva de la automatización. Aunque la IA puede apoyar al radiólogo, no debe reemplazar su criterio clínico. Los autores analizados resaltan que la supervisión humana sigue siendo indispensable, especialmente porque los sistemas automatizados pueden fallar ante imágenes alteradas, datos incompletos, sesgos en el entrenamiento o situaciones clínicas no contempladas durante el desarrollo del modelo.

Tabla 2

Beneficios y Desafíos de la IA

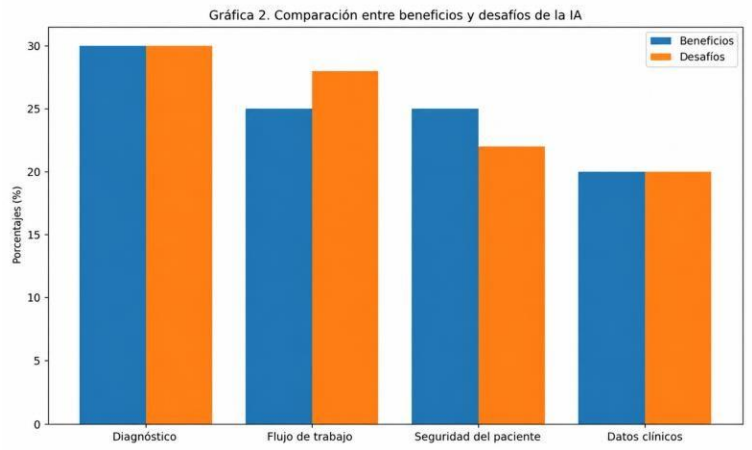
Autor	Aspecto	Beneficios	Desafíos
Katzman et al. (2023)	Diagnóstico	Mayor rapidez y apoyo en la detección de hallazgos	Posibilidad de errores por ataques adversariales

Trapero y López (2017)	Flujo de trabajo	Optimización de tiempos y priorización de estudios	Dependencia tecnológica y fallas operativas
Narváez Pereira et al. (2024)	Calidad de imagen	Detección de artefactos y apoyo en control de calidad	Necesidad de validación constante
Sotelo Ramírez (2023)	Seguridad del paciente	Apoyo a decisiones clínicas más oportunas	Riesgo de diagnósticos incorrectos
Rosales Chaves y Ramírez Morales (2024)	Datos clínicos	Mejor gestión digital de información	Riesgo de filtración, alteración o acceso no autorizado
Álvarez Córdova (2025)	Rol profesional	Fortalece el trabajo del radiólogo	Exige capacitación y supervisión continua

Nota. Elaboración propia

Figura 5

Comparación entre Beneficios y Desafíos de la IA



Nota. Si bien, los beneficios de la IA representan un gran avance en la atención en los servicios radiológicos, los desafíos siguen siendo un tema de abordaje crucial en los procesos de calidad. Elaboración propia

Esta gráfica permite analizar que la inteligencia artificial ofrece múltiples beneficios dentro del flujo de trabajo radiológico, especialmente en la rapidez diagnóstica, la optimización de procesos y la detección temprana de patologías.

Sin embargo, también se evidencian desafíos importantes, siendo los ataques adversariales y la ciberseguridad los principales factores de riesgo. Esto demuestra que la implementación de IA no debe centrarse únicamente en mejorar la eficiencia, sino también en garantizar la protección de los sistemas y la supervisión constante del proceso diagnóstico.

La comparación evidencia que los beneficios son altos, pero los riesgos requieren una gestión institucional sólida.

En cuanto a las posibles soluciones, los resultados muestran que no basta con mejorar el rendimiento de los algoritmos. Es necesario implementar una estrategia integral que combine ciberseguridad, gobernanza de datos, capacitación del personal y validación clínica continua. Entre las medidas más mencionadas se encuentran el monitoreo constante del desempeño de los modelos, la detección de anomalías en imágenes, el uso de bases de datos representativas, la actualización segura de los sistemas y la creación de protocolos institucionales frente a incidentes digitales.

Tabla 3

Estrategias de Mitigación de Errores de IA

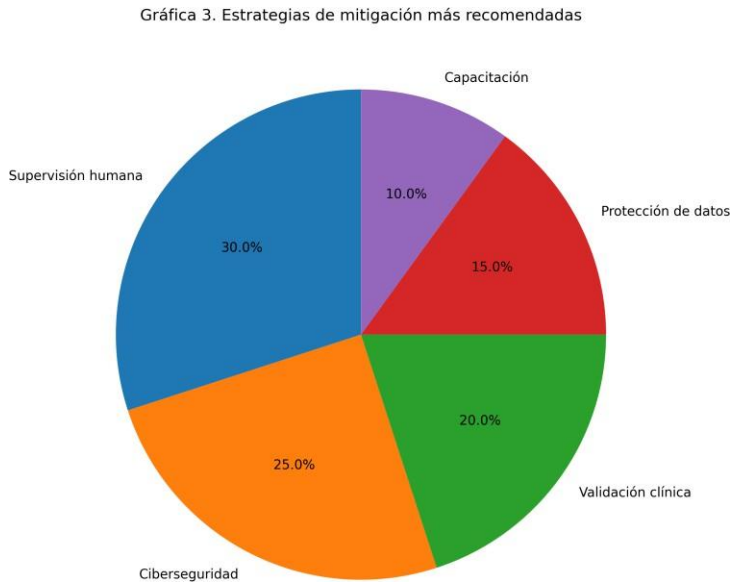
Autor	Solución propuesta	Propósito	Nivel de aplicación
Silva Afonso et al. (2025)	Validación clínica continua	Verificar que la IA mantenga resultados confiables	Clínico y técnico

Álvarez Córdova (2025)	Supervisión humana obligatoria	Evitar dependencia total del algoritmo	Profesional
Strohal (2023)	Capacitación en ciberseguridad	Reducir errores humanos y mejorar respuesta ante incidentes	Institucional
Rosales Chaves y Ramírez Morales (2024)	Protección de datos sensibles	Garantizar confidencialidad e integridad de la información	Ético y legal
Berrones Berrones (2024)	Modelos más robustos	Resistir alteraciones adversariales en imágenes	Tecnológico
Nguyen et al. (2025)	Protocolos de respuesta	Actuar rápidamente ante ataques o fallas	Organizacional
Trapero y López (2017)	Auditorías periódicas	Evaluar seguridad, trazabilidad y desempeño	Institucional

Nota. Elaboración propia

Figura 6

Estrategias de Mitigación de Errores en IA



Nota. Estrategias más recomendadas por los autores para reducir el impacto de los ataques adversariales en radiología. Elaboración propia

La supervisión humana ocupa el porcentaje más alto, lo que confirma que la inteligencia artificial debe funcionar como una herramienta de apoyo y no como reemplazo del profesional radiólogo. El criterio clínico sigue siendo indispensable para validar los resultados generados por los algoritmos.

El fortalecimiento de la ciberseguridad también ocupa un lugar prioritario, especialmente en instituciones con sistemas altamente digitalizados. La validación clínica continua y la protección de datos sensibles complementan este proceso, garantizando mayor confiabilidad y seguridad diagnóstica.

De manera general, los resultados permiten afirmar que el principal desafío no es solamente incorporar inteligencia artificial en radiología, sino hacerlo de manera segura, ética y controlada. La IA tiene un alto potencial para fortalecer el diagnóstico por imágenes, pero su uso sin mecanismos de defensa adecuados puede generar riesgos clínicos, técnicos y legales. Por esta razón, la protección de datos sensibles y la confiabilidad diagnóstica deben considerarse componentes centrales dentro de cualquier proceso de implementación tecnológica en servicios radiológicos.

El análisis de los autores e investigadores consultados permite establecer que los ataques adversariales representan una amenaza real para la validez del diagnóstico radiológico asistido por inteligencia artificial. Las soluciones más viables deben orientarse hacia un modelo de seguridad integral, donde el algoritmo, el profesional de salud, la institución y los protocolos de protección trabajen de manera articulada. Esto permitiría aprovechar los beneficios de la IA sin comprometer la seguridad del paciente ni la confianza en los sistemas digitales de salud.

Conclusiones

A partir del análisis documental realizado, se concluye que los ataques adversariales representan una amenaza importante para los sistemas de inteligencia artificial aplicados a la radiología, debido a que pueden alterar imágenes médicas o datos clínicos de manera casi imperceptible y modificar los resultados generados por los algoritmos. Esta situación compromete directamente la validez diagnóstica y puede generar falsos positivos o falsos negativos, afectando la seguridad del paciente y la toma de decisiones clínicas.

También se concluye que la inteligencia artificial aporta beneficios significativos al flujo de trabajo radiológico, especialmente en la rapidez del diagnóstico, la priorización de estudios, la detección de hallazgos y el apoyo al profesional de salud. Sin embargo, estos beneficios no eliminan los riesgos técnicos, éticos y clínicos asociados a su implementación. Por esta razón, la IA debe entenderse como una herramienta complementaria y no como un reemplazo del criterio profesional del radiólogo.

Otro hallazgo relevante es que la vulnerabilidad de los sistemas de IA no depende únicamente del algoritmo, sino también del contexto institucional donde se utilizan. Los hospitales y servicios radiológicos funcionan mediante redes digitales, sistemas PACS, bases de datos clínicas y plataformas interconectadas, lo que aumenta la superficie de exposición frente a ciberataques, filtración de información o alteración de datos sensibles.

De igual manera, se concluye que la protección de los datos clínicos debe ser un componente central en la implementación de inteligencia artificial en radiología. La confidencialidad, integridad y disponibilidad de la información médica son aspectos fundamentales para garantizar la confianza de los pacientes y la seguridad de los procesos

diagnósticos. Por ello, se requiere fortalecer la gobernanza de datos, los controles de acceso, la trazabilidad de la información y los protocolos de respuesta ante incidentes digitales.

Se establece que la implementación segura de la inteligencia artificial en radiología requiere un enfoque integral que combine validación clínica continua, supervisión humana, capacitación del personal, auditorías periódicas, modelos algorítmicos más robustos y estrategias institucionales de ciberseguridad. Solo mediante esta articulación entre tecnología, personal de salud y protocolos de protección será posible aprovechar los beneficios de la IA sin comprometer la seguridad del paciente ni la confiabilidad del diagnóstico radiológico.

Referencias Bibliográficas

Acosta-Jiménez, S., González-Chávez, S. A., Camarillo-Cisneros, J., Pacheco-Tena, C. F., & Ochoa-Albíztegui, R. E. (2023). Aplicaciones de la inteligencia artificial en la medicina y la imagenología médica. Anales de Radiología México, 22, 130–139. https://doi.org/10.24875/ARM.21000093

Aguirre, F., Carballo, L., González, X., & Gigirey, V. (2021). Inteligencia artificial aplicada a la imagen médica. Revisión de tema. Revista Imagenol, 24(2), 47–58. file:///C:/Users/USUARIO/Downloads/94-1-506-1-10-20211012.pdf

Álvarez Córdova, V. M. (2025). El rol del radiólogo con la implementación de la inteligencia artificial. Revista Científica Internacional Arandu UTIC, 12(3), 369–376. https://doi.org/10.69639/arandu.v12i3.1309

Arias, F. G. (2012). El proyecto de investigación. https://tauniversity.org/sites/default/files/libro_el_proyecto_de_investigacion_de_fidias_g_arias.pdf

Berrones Berrones, K. A. (2024). Integración de modelos de IA de análisis de imagen radiológica en el entorno clínico (Trabajo Fin de Grado). Escuela Politécnica Superior, Universidad de Alicante. https://rua.ua.es/server/api/core/bitstreams/8ba5059e-3c80-43f1-86db-c86c4a6f66c2/content

Caicedo Cobos, A. F., Caraballo Caldera, I. P., Rodríguez Gutiérrez, N. H., Barrios Parejo, R. D. J., & Mendoza Jiménez, Y. (2024). Inteligencia artificial en la interpretación de imágenes médicas. Universidad Nacional Abierta y a Distancia (UNAD), Escuela de Ciencias de la Salud ECISA.

https://repository.unad.edu.co/bitstream/handle/10596/64852/ipcaraballoc.pdf?isAllowed=y&sequence=1

Calva Sánchez, R. J., Jiménez Buri, K. F., Herrera Sarango, S. C., & Núñez Cabrera, C. M. (2023). Avances tecnológicos y científicos en radiología. RECIAMUC, 7(2), 457–465. https://doi.org/10.26820/reciamuc/7.(2).abril.2023.457-465

Castillo López, J. L., Puchi Lojano, M. D., Rivera Calle, D. F., Vásquez Sinchi, P. H., & Ordoñez, M. del C. (2026). El papel de la inteligencia artificial en el diagnóstico por imágenes: ¿Complemento o reemplazo? Ibero Ciencias, 5(1), 585–596. https://doi.org/10.63371/ic.v5.n1.a679

Creswell, J. (2014). Research design: Qualitative, quantitative and mixed methods approaches. https://www.ucg.ac.me/skladiste/blog_609332/objava_105202/fajlovi/Creswell.pdf

De La Cámara Egea, M. A. (2019). Inteligencia Artificial en Radiología. Radiología Club. https://radiologiaclub.com/2019/11/26/inteligencia-artificial-en-radiologia/

Díaz Ramírez, A. M., González Mancera, A., García Rojas, J. D., Liz Andela, L. G., & Villamil Castellanos, Y. P. (2025). Análisis de la implementación de inteligencia artificial (IA) mediante algoritmos en los parámetros técnicos de adquisición para la radiografía de tórax [Trabajo de investigación, Universidad Nacional Abierta y a Distancia]. Repositorio Institucional UNAD. https://repository.unad.edu.co/bitstream/handle/10596/78693/YPVILLAMILC.pdf?sequence=3&isAllowed=y

Domínguez, Y. (s.f.). El impacto del uso de la IA en la imagen médica. Grupo Oesía. https://grupooesia.com/wp-content/uploads/el-impacto-del-uso-de-la-inteligencia-artificial-en-la-imagen-medica.pdf

Estrada, D., Alvarado, O., & Carrillo, K. (2020). Inteligencia artificial en el área médica aplicada en la imagenología. file:///C:/Users/USUARIO/Downloads/admin,+73-Texto+del+art%C3%ADculo-275-1-15-20201228.pdf

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohan, I. S. (2019). Adversarial attacks on medical machine learning. Science, 363(6433), 1287–1289. https://doi.org/10.1126/science.aaw4399.

Galarza Medina, K. X., Maldonado Coronel, K., & Herrera Guanopatin, M. S. (2023). Beneficios y riesgos de la implementación de inteligencia artificial en los procesos de diagnóstico médico en el Ecuador. Ciencia Latina Revista Científica Multidisciplinar, 7(6), 7276–7299. https://doi.org/10.37811/cl_rcm.v7i6.9274

Gallego Piña, E. Y., Luna Martínez, J. J., Sierra Bedoya, J. A., Chiquillo Yepes, R., & Cogollo López, Y. J. (2025). La inteligencia artificial (IA) aplicada en la radiología para la detección temprana de patologías [Trabajo académico]. Universidad Nacional Abierta y a Distancia (UNAD). https://repository.unad.edu.co/jspui/bitstream/10596/77343/1/jjlunama.pdf

Gálvez, M. (2017). Inteligencia artificial en radiología: ¿Seremos reemplazados por las máquinas? Revista Chilena de Radiología, 23(3), 90. https://www.scielo.cl/scielo.php?pid=S0717-93082017000300001&script=sci_arttext

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.6572.

Hernández Sampieri, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación*. https://dialnet.unirioja.es/servlet/libro?codigo=775008

Higuera Mosquera, D. A., Pineda Ospino, J. A., & Velásquez Cruz, E. S. (2025). Análisis del uso y eficacia de la inteligencia artificial en tomografía computarizada (Trabajo de grado). Fundación Tecnológica Autónoma de Bogotá FABA. https://www.faba.edu.co/wp-content/uploads/2025/08/INTELIGENCIA-ARTIFICIAL-EN-TOMOGRAFIA-COMPUTARIZADA.pdf

Iglesias López, D. (2023). Impacto de la Inteligencia Artificial en la Radiología. Revista Cubana de Informática Médica, 15(1), e624. http://scielo.sld.cu/pdf/rcim/v15n1/1684-1859-rcim-15-01-e624.pdf

International Commission on Radiological Protection. (2003). Gestión de la dosis al paciente en radiología digital (Publicación ICRP 93). Annals of the ICRP. Traducción oficial al español publicada por la Sociedad Argentina de Radioprotección (2014). https://www.icrp.org/docs/p93_spanish.pdf

Jiménez-Rodríguez, L. A., Jiménez Ospina, J. S., & Agudelo Berrio, J. F. (2024). y de optimización hacia la calidad y seguridad en los servicios de diagnóstico por imagen. NOVA, 22(42). https://doi.org/10.22490/24629448.8182

Katzman, B. D., van der Pol, C. B., Soyer, P., & Patlas, M. N. (2023). Artificial intelligence in emergency radiology: A review of applications and possibilities. Diagnostic and Interventional Imaging, 104, 6–10. https://www.sciencedirect.com/science/article/pii/S2211568422001437

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

López Izquierdo, R. (2025). Papel de la inteligencia artificial en la interpretación de imágenes radiológicas en urgencias. Emergencias, 37, 403–404. https://revistaemergencias.org/wp-content/uploads/2025/10/403-404.pdf

Machacado Rojas, A. M., & Aparicio Pico, L. E. (2021). Técnicas de inteligencia artificial aplicadas al análisis de imágenes diagnóstico. https://pdfs.semanticscholar.org/c747/894f5cf20a88fccbdb2725708d2de5b12730.pdf

Marangoni, A. (2018). El arribo de la "inteligencia artificial" a la radiología – ¿Amenaza o desafío de adaptación? Revista Argentina de Radiología, 82, 55–56. https://doi.org/10.1055/s-0038-1656546

Martí-Bonmatí, L. (2024). Inteligencia artificial en imagen médica. Anales de la Real Academia Nacional de Medicina, 141(02), 111–118. https://doi.org/10.32440/ar.2024.141.02.rev02

Narváez Pereira, M., Herrera Rojas, D. A., & Ladino Gutiérrez, A. L. (2024). Impacto de la inteligencia artificial en el control de calidad de imágenes radiológicas y la detección de artefactos. Universidad Nacional Abierta y a Distancia (UNAD), Escuela de Ciencias de la Salud ECISA. https://repository.unad.edu.co/jspui/bitstream/10596/63440/3/alladinog.pdf

Nguyen, X. V., Petscavage-Thomas, J. M., Straus, C. M., & Ikuta, I. (2025). Cybersecurity in radiology: Cautionary tales, proactive prevention, and what to do when you get hacked. Current Problems in Diagnostic Radiology, 54, 245–250. https://www.sciencedirect.com/science/article/pii/S0363018824001221.

Nallino, M. B., Acevedo, P., Fernández, I., Fumagalli, A. I., & Ojeda, A. (2024). Inteligencia artificial en neuroimágenes. Anuario (Fundación Dr. J. R. Villavicencio), 31, e-1–e-7. https://villavicencio.org.ar/anuario/31/inteligencia-artificial-en.pdf

Parada, O., Quintero, J., Quintero, J., Cetina, J., & Carvajal, E. (2023). Radiología 2.0: El futuro de la inteligencia artificial en la identificación e interpretación de imágenes diagnósticas. https://herasmomeoz.gov.co/wp-content/uploads/2024/03/PROYECTO-17.pdf

Paschali, M., Conjeti, S., Navarro, F., Navab, N., & Wachinger, C. (2018). Generalizability vs. robustness: Investigating medical imaging networks using adversarial examples. Medical Image Analysis, 55, 192–207. https://doi.org/10.1016/j.media.2019.04.005.

Pierre, K., Haneberg, A. G., Kwak, S., Peters, K. R., Hochhegger, B., Sananmuang, T., Tunlayadechanont, P., Tighe, P. J., Mancuso, A., & Forghani, R. (2023). Applications of artificial intelligence in the radiology roundtrip: Process streamlining, workflow optimization, and beyond. Seminars in Roentgenology, 58, 158–169. https://doi.org/10.1053/j.ro.2023.02.003

Puentes, G., Salinas Miranda, E., & Triana, G. A. (2021). Inteligencia artificial y radiología: La disrupción tecnológica en la transformación de un paradigma. Med, 43(4), 594–605. https://doi.org/10.56050/01205498.1549

Quibim. (2025, 31 enero). Diagnóstico por imágenes con inteligencia artificial: ¿cómo está transformando la atención médica? - Quibim. Quibim Website. https://quibim.com/es/news/ai-diagnostic-imaging-medical-care/

Revista Ocronos. (2023). Aplicación de la inteligencia artificial en la interpretación de imágenes médicas: Panorama actual e impacto a futuro. Editorial Científico-Técnica Ocronos. https://doi.org/10.58842/OCRONOS

Rodríguez, A., Martínez, L., & Reyes Alvarado, S. (2023). Uso de nuevas tecnologías en radiología e imágenes diagnósticas y su relación con las competencias profesionales y/o perfil de egreso del Licenciado en Radiología de Panamá y Latinoamérica en los últimos

15 años. Ciencia Latina Revista Científica Multidisciplinar, 7(1), 6762–6788. https://doi.org/10.37811/cl_rcm.v7i1.4929

Rosales Chaves, M., & Ramírez Morales, L. (2024). Tendencias emergentes en radiología: La inteligencia artificial, la radiología molecular y la imagen personalizada. Revista Veritas de Difusión Científica, 5(2), 449–463. https://doi.org/10.61616/rvdc.v5i2.95

Salas Henriquez, A. (2025). El papel de la inteligencia artificial en la interpretación de imágenes radiológicas: avances recientes y desafíos éticos. Revista Electrónica de PortalesMedicos.com, Vol. XX, Nº 07, pp. 312

Silva Afonso, R. F., Gallardo-Rodríguez, P., Espinosa, B., Bautista, A., Serrano, J., Veguillas, M., Corell, M., Garrido Chamorro, R., Arenas Jiménez, J., Astor Rodríguez, C., Abellón Fernández, Á., Palazón Ruíz de Tremiño, Á., Garfias Baladrón, M. J., Marquina Arribas, V., Chico-Sánchez, P., Gras Valenti, P., Cabrer González, M., Martínez Riera, C., Moliner Mateu, D., Salinas Serrano, J. M., Vivancos Rubio, E., Valdivieso Martínez, B., Concepción-Aramendia, L., Sánchez-Payá, J., & Llorens, P. (2025). Fiabilidad y validez de un sistema asistido por inteligencia artificial para la detección de anomalías en las radiografías de tórax y óseas en un servicio de urgencias hospitalario. Emergencias, 37, 420–426. https://revistaemergencias.org/wp-content/uploads/2025/10/420-426.pdf

Sotelo Ramírez, S. E. (2023). Reflexiones sobre la inteligencia artificial en radiología. Revista Peruana de Radiología, 22, 24–27. https://socpr.org.pe/wp-content/uploads/2024/06/rev-vol-22-tema-de-opinion.pdf

Strohal, A. (2023). Innovación en la capacitación virtual en seguridad informática para instalaciones nucleares y radiológicas. Boletín del OIEA, junio 2023. https://www.iaea.org/sites/default/files/6421213es.pdf

Trapero, M. A., & López, I. (2017). Guía para la renovación y actualización tecnológica en radiología: Gestión de los ciclos de vida de la tecnología de diagnóstico por la imagen. Sociedad Española de Radiología Médica (SERAM). https://www.fundacionsigno.com/archivos/20240207082254.pdf

Wagner, A. (2026). ¿Cómo reforzar la ciberseguridad sin frenar la innovación en Radiología? RadiologiaLatam.