

Learning Architectures for AI-Assisted Medical Image Diagnosis: A Systematic Literature Review (2020–2025)

Cristhian Ronceros-Morales¹; José Luis Cabezado¹; Juan Jiménez Garavito²; Carlos Oliva Ramos¹;
Paco Marquez Urbina²

¹Universidad Tecnológica del Perú, Facultad de Ingeniería de Sistemas e Informática, Perú. C20605@utp.edu.pe,
C31314@utp.edu.pe, C27251@utp.edu.pe

²Universidad Nacional San Luis Gonzaga de Ica, Facultad de Ingeniería de Sistemas, Perú. juan.jimenez@unica.edu.pe,
pmarquez@unica.edu.pe

Abstract– Deep learning (DL) techniques have substantially transformed medical imaging diagnostics by improving diagnostic accuracy, clinical efficiency, and decision support. This article presents a systematic literature review (SLR) that aims to analyze the evolution, contributions, limitations, and clinical benefits of DL architectures applied to medical imaging diagnostics during the period 2020–2025. The review was conducted following the PRISMA framework guidelines. The literature search in the Scopus database yielded 10,623 results on artificial intelligence, deep learning, and medical imaging. After applying automated filters and a thorough manual review, 38 scientific articles were selected that met rigorous criteria for clinical validation, use of real medical images, and diagnostic utility. The results indicate that CNNs and hybrid CNN-SVM models are the most widely used architectures, accounting for more than 57% of the articles reviewed. In addition, there is a growing adoption of modern architectures such as U-Net, ResNet, DenseNet, and emerging models based on Transformers. The greatest contributions are categorized into three areas: computational (42.1%), methodological (34.2%), and clinical (23.7%). In the clinical world, DL systems consistently improved diagnostic accuracy, reduced interobserver variability, improved workflows, and aided decision-making in imaging such as MRI, CT, ultrasound, and mammography. However, challenges remain in terms of data heterogeneity, model generalization, algorithmic interpretability, regulatory frameworks, and institutional readiness. In summary, this review provides a structured, evidence-based synthesis, bringing algorithmic research closer to its practical application in the clinic.

Keywords– Deep learning; Medical imaging; Computer-aided diagnosis; Clinical validation.

Arquitecturas de aprendizaje para el diagnóstico de imágenes médicas con IA: Una revisión sistemática de la literatura (2020-2025)

Cristhian Ronceros-Morales¹; José Luis Cabezudo¹; Juan Jiménez Garavito²; Carlos Oliva Ramos¹;
Paco Marquez Urbina²

¹Universidad Tecnológica del Perú, Facultad de Ingeniería de Sistemas e Informática, Perú. C20605@utp.edu.pe,
C31314@utp.edu.pe, C27251@utp.edu.pe

²Universidad Nacional San Luis Gonzaga de Ica, Facultad de Ingeniería de Sistemas, Perú. juan.jimenez@unica.edu.pe,
pmarquez@unica.edu.pe

Resumen— Las técnicas de aprendizaje profundo (Deep Learning, DL) han transformado de manera sustancial el diagnóstico médico por imágenes, al mejorar la precisión diagnóstica, la eficiencia clínica y el soporte a la toma de decisiones. Este artículo presenta una revisión sistemática de la literatura (RSL) cuyo objetivo es analizar la evolución, aportes, limitaciones y beneficios clínicos de las arquitecturas de DL aplicadas al diagnóstico médico por imágenes durante el periodo 2020–2025. La revisión se desarrolló siguiendo las directrices del marco PRISMA. La búsqueda bibliográfica en la base de datos Scopus arrojó 10,623 resultados sobre inteligencia artificial, aprendizaje profundo e imágenes médicas. Después de aplicar filtros automatizados y una revisión manual exhaustiva, se eligieron 38 artículos científicos que cumplieran con criterios rigurosos de validación clínica, uso de imágenes médicas reales y utilidad para el diagnóstico. Los resultados indican que las CNN y los modelos híbridos CNN-SVM son las arquitecturas más utilizadas, sumando más del 57% de los artículos revisados. Además, se está viendo una mayor adopción de arquitecturas modernas como U-Net, ResNet, DenseNet y modelos emergentes basados en Transformers. Las mayores contribuciones se categorizan en tres áreas: computacional (42.1%), metodológica (34.2%) y clínica (23.7%). En el mundo clínico, los sistemas con DL mejoraron constantemente la precisión diagnóstica, disminuyeron la variabilidad interobservador, mejoraron los flujos de trabajo y ayudaron en la toma de decisiones en imágenes como resonancia magnética, tomografía computarizada, ultrasonido y mamografía. Pero aún quedan retos abiertos en cuanto a la heterogeneidad de los datos, la generalización de los modelos, la interpretabilidad algorítmica, los marcos normativos y la preparación institucional. En suma, esta revisión proporciona una síntesis estructurada y basada en evidencia, acercando la investigación algorítmica a su aplicación práctica en la clínica.

Palabras clave— Aprendizaje profundo; Diagnóstico médico por imágenes; Diagnóstico asistido por computadora; Validación clínica.

I. INTRODUCCIÓN (HEADING 1)

En la última década, y más aún desde 2020, los modelos de aprendizaje profundo han transformado el diagnóstico por imágenes, convirtiéndose en una de las principales líneas de innovación en la interfaz entre la ingeniería, la informática médica y la práctica clínica, con aplicaciones en imágenes de

resonancia magnética (MRI), tomografía computarizada (TC), radiografía digital y ecografía, en las que las redes neuronales profundas pueden identificar patrones complejos y sutiles que el ojo humano pasa por alto, mejorando la precisión, la consistencia y la eficiencia del diagnóstico [1, 2, 3, 4].

Diversos estudios han demostrado que los sistemas de deep learning pueden alcanzar o superar la precisión de los radiólogos expertos en tareas específicas. Por ejemplo, en neuroimagen, modelos entrenados en imágenes de tomografía computarizada cerebral han superado a los métodos tradicionales en la detección de hemorragias intracraneales, mejorando la sensibilidad y reduciendo el tiempo hasta la lectura clínica [5, 6, 7]. De manera similar, en la lectura de MRI de mama, las arquitecturas profundas han superado a los radiólogos en la interpretación de realces no nodulares, una tarea con gran variabilidad interobservador cuando la realizan solo humanos [8, 9].

Además, la detección automatizada de fracturas óseas mediante aprendizaje profundo ha alcanzado mejoras sustanciales en sensibilidad y especificidad diagnóstica, disminuyendo los falsos positivos y acelerando el diagnóstico, lo que es esencial en emergencias médicas [10, 11]. Entre las aplicaciones más maduras se halla la clasificación de tumores cerebrales, para la cual modelos híbridos basados en CNN y SVM logran alta eficacia en segmentación de lesiones y clasificación de subtipos tumorales, lo que ayuda en la toma de decisiones clínicas [12, 13, 14, 15, 16]; además, algoritmos de optimización evolutivos sobre CNN mejoran la precisión global y reducen los tiempos de entrenamiento en imágenes cerebrales de resonancia magnética [17].

El aprendizaje profundo no solo se ha aplicado para el diagnóstico estructural, sino que ha llegado a desarrollar aplicaciones más complejas, como la predicción de enfermedades neurodegenerativas a través de plataformas interactivas, donde chatbots inteligentes combinados con modelos CNN-SVM logran hacer una evaluación inicial de enfermedades como el Alzheimer, basándose en imágenes médicas y datos lingüísticos y de comportamiento del paciente [18, 19]. Además, una de las mayores aportaciones de la IA en

el campo de la medicina ha sido disminuir la carga de trabajo de los profesionales sanitarios y la variabilidad interobservador, como se ha demostrado en radiología torácica y espinal, donde los algoritmos de deep learning han proporcionado una segunda opinión automatizada y fiable en entornos clínicos con alta demanda asistencial [20, 21, 22, 23].

En sitios con escasez de radiólogos, la IA ha demostrado resultados clínicos positivos, extendiendo el acceso a diagnósticos confiables en entornos de bajos recursos [24, 25, 26]; pero aún existen desafíos que limitan su implementación clínica segura, como la generalización de los modelos ante distintas poblaciones y condiciones clínicas, para lo cual se han creado arquitecturas más sólidas como DenseNet, EfficientNet o modelos basados en Transformers, capaces de adaptarse a las variaciones en equipos de captura, protocolos institucionales y perfiles demográficos del paciente [22, 27, 28, 24].

En este sentido, la validación multicéntrica es fundamental para la aplicabilidad en el mundo real más allá de entornos controlados [29, 30, 31, 32]; los problemas éticos, legales y regulatorios están en aumento y la explicabilidad de los modelos, la trazabilidad de las decisiones algorítmicas y la protección de datos confidenciales del paciente son necesarias para la implementación responsable de la IA en la práctica clínica [10, 13, 33, 34]; la falta de marcos regulatorios establecidos en muchos países sigue siendo un desafío pendiente, especialmente en lo que respecta a la responsabilidad por errores de diagnóstico asistido por IA [26, 32, 35].

En ese sentido, el presente artículo realiza una revisión sistemática de la literatura comparando las principales arquitecturas de aprendizaje profundo utilizadas para el diagnóstico médico asistido por imágenes en el periodo 2020-2025 en términos de rendimiento diagnóstico, capacidad de generalización y viabilidad para ser implementadas en escenarios clínicos reales, reconociendo sus fortalezas y limitaciones técnicas, pero también para crear un marco para la integración de soluciones de IA que se adapten a las necesidades reales de la práctica médica actual y futura en el diagnóstico por imágenes [4, 36, 37, 38].

II. METODOLOGÍA

A. Diseño de la investigación

El estudio se diseñó como una revisión sistemática de la literatura (RSL) siguiendo la metodología del marco PRISMA. Esta metodología permitió identificar, evaluar y sintetizar la mejor evidencia científica disponible sobre el uso de modelos de aprendizaje profundo en el diagnóstico médico asistido por imágenes en el periodo 2020-2025 con transparencia, reproducibilidad y rigor metodológico.

La RSL se enfocó en responder las preguntas de investigación planteadas sobre el impacto clínico, las contribuciones tecnológicas, los beneficios prácticos y las limitaciones técnicas, éticas y organizacionales de estas soluciones en el mundo real de la práctica clínica.

B. Estrategia de búsqueda bibliográfica

La búsqueda de literatura se realizó de manera sistemática en bases de datos científicas indexadas de alto impacto como Scopus. La ecuación de búsqueda se diseñó para encontrar artículos que hablen de Deep Learning, diagnóstico médico e imágenes médicas, usando términos amplios y específicos sobre arquitecturas, modalidades de imagen y validación clínica. La ecuación de búsqueda maestra, ajustada a la sintaxis de cada base de datos, fue: ("deep learning" OR "convolutional neural network" OR CNN OR "hybrid CNN-SVM" OR "machine learning") AND ("medical imaging" OR "medical image analysis" OR MRI OR "magnetic resonance imaging" OR CT OR "computed tomography" OR ultrasound OR mammography) AND ("diagnosis" OR "clinical diagnosis" OR "computer-aided diagnosis" OR CAD OR "radiologist performance")

C. Criterios de inclusión y exclusión

Criterios de inclusión

Los estudios seleccionados debían cumplir con los siguientes criterios:

CI1: Estar publicados entre 2020 y 2025 en revistas o congresos indexados.

CI2: Aplicar modelos de Deep Learning o enfoques híbridos al diagnóstico médico basado en imágenes.

CI3: Incluir resultados cuantitativos o cualitativos relacionados con desempeño diagnóstico, eficiencia o validación clínica.

CI4: Comparar, total o parcialmente, el desempeño de los modelos con radiólogos o especialistas clínicos, o bien reportar validación en escenarios clínicos reales.

CI5: Estar disponibles en texto completo.

Criterios de exclusión

Se excluyeron aquellos estudios que:

CE1: Se centraban exclusivamente en simulaciones sin relación clínica.

CE2: No utilizaban imágenes médicas como fuente principal de datos.

CE3: Eran revisiones narrativas, editoriales, cartas al editor o resúmenes extendidos.

CE4: Presentaban información incompleta o sin resultados verificables.

Tras aplicar la estrategia de búsqueda definida, la consulta inicial en la base de datos Scopus identificó 10,623 registros relacionados con inteligencia artificial, aprendizaje profundo y diagnóstico médico por imágenes. Mediante filtros automatizados y criterios de selección previamente establecidos, se excluyeron los estudios fuera del periodo de análisis, sin acceso al texto completo, en idiomas distintos del inglés o ajenos al alcance temático. Como resultado de este proceso sistemático, se seleccionaron 38 artículos científicos que cumplieron plenamente con los criterios metodológicos y de contenido definidos (ver Figura 1).

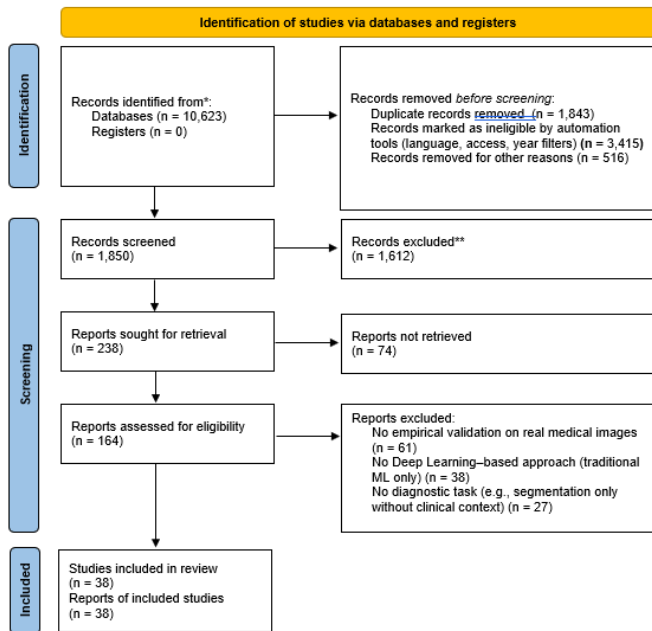


Fig. 1. Diagrama de flujo Prisma.

III. RESULTADOS

A. Resultados bibliométricos

En esta parte de la revisión sistemática se muestran los artículos elegibles que justifican la realización de la investigación. De cada estudio se obtuvo información como año de publicación, país, sector y tecnologías utilizadas. (Véase la tabla 1).

TABLA I
DETALLE DE ARTÍCULOS SELECCIONADOS

Artículos	Año	País	Sector de Aplicación	Tecnologías Implementadas
01, 03, 05, 06, 08, 11, 13, 14, 15, 17, 19, 21, 24, 25, 27, 28, 29, 33, 38	2020–2025	EE.UU., China, Japón, Alemania, Australia, Países Bajos, Corea del Sur, India, Nigeria	Diagnóstico médico multiespecialidad	Deep Learning, CNN, diagnóstico asistido, segmentación y clasificación multiclase, surveys y guías clínicas
02, 04, 16, 34	2022–2024	Indonesia, India, China, Alemania	Tumores cerebrales (MRI)	CNN, CNN-SVM, Transformers (Swin UNETR), optimización evolutiva
12, 18, 31	2023–2025	Japón, EE.UU., Países Bajos, Turquía	Cáncer de mama	MRI mamaria, DCE-MRI, ABUS, radiomics, evaluación IA–radiólogos
07, 09, 10,	2022–	Australia,	Diagnóstico	XAI, ROC-AUC,

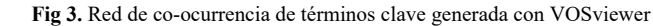
35	2024	EE.UU., Europa	pulmonar y confianza clínica	evaluación de errores, transparencia algorítmica
17, 22	2022–2024	Alemania, EE.UU., Japón	MRI multiparamétrica	Comparación IA vs radiólogos, interpretabilidad, CNN, Transformers
06, 11, 14, 15, 23, 36	2022–2024	Australia, China, Alemania, Dinamarca	Tomografía computarizada (CT)	Detección asistida, segmentación automática, validación externa
19, 26	2022–2024	EE.UU., Corea del Sur	Cáncer de mama y próstata	BI-RADS, PI-RADS, MRI biparamétrica
30, 32	2020	India, China	Obstetricia y ginecología	Ultrasonido automatizado, IoMT, regresión logística
13, 17, 24	2022–2024	Nigeria, Japón, Indonesia	Neuroimagen / salud mental	MRI estructural, severidad clínica, SAPS/SANS
18, 20	2020–2022	EE.UU., Países Bajos	Diagnóstico autónomo (sin radiólogos)	Sistemas CAD, validación clínica comparativa
23	2022	Dinamarca	Hemorragia intracraneal	Meta-análisis, sensibilidad y especificidad
35, 36, 37	2019–2020	Global	Ética, regulación y adopción clínica	XAI, trazabilidad, responsabilidad médica

En la figura 2 se observa que la localización geográfica de los estudios muestra que la mayor parte de la producción científica se localiza en Estados Unidos (9 artículos) y China (6 artículos), siendo los países líderes en el desarrollo y validación clínica de soluciones basadas en Deep Learning para el diagnóstico por imagen. India es un país que se queda a mitad de camino con cuatro estudios, y países europeos y latinoamericanos tienen pocas aportaciones.



Fig 2. Publicaciones por países

ol



24th LACCEI International Multi-Conference for Engineering, Education, and Technology
Alliances for a Future from the Americas". Hybrid Event, Santiago-Chile, July 15 - 17, 2024

Un segundo tipo importante son los modelos híbridos CNN-SVM, que supusieron el 18.4% de los artículos (7 estudios), siendo común en estudios para clasificar tumores cerebrales y enfermedades neurológicas, mejorando la robustez del modelo en situaciones con datos limitados o heterogéneos gracias a la extracción profunda de características con CNN y la clasificación supervisada con máquinas de vectores de soporte.

Por otro lado, los modelos emergentes basados en Transformers representaron el 5.2% de los artículos (2 estudios), una presencia aún incipiente pero que refleja una línea creciente de investigación, especialmente en tareas complejas de segmentación en imágenes MRI, donde los mecanismos de atención permiten modelar dependencias contextuales de largo alcance con resultados prometedores.

Finalmente, los artículos marco, que incluyen revisiones, guías clínicas, estudios éticos, de explicabilidad e implementación, constituyeron el 10.6% (4 artículos); estos no proponen modelos algorítmicos, pero son cruciales para establecer las bases conceptuales, metodológicas y regulatorias para el uso de la IA en el diagnóstico por imágenes. En resumen, el campo está bien establecido tecnológicamente, con modelos CNN y modelos híbridos como enfoques predominantes, y arquitecturas avanzadas y Transformers como tendencias emergentes; sin embargo, la escasez de estudios enfocados en validación clínica, explicabilidad y ética demuestra la necesidad de desarrollar

estas áreas para una integración segura, transparente y sostenible de la IA en la práctica médica actual.

A continuación, la Tabla II sintetiza las tecnologías más utilizadas, su frecuencia de aparición y los artículos en los que fueron reportadas

TABLA II
 TECNOLOGÍAS MÁS EMPLEADAS EN EL DIAGNÓSTICO MÉDICO ASISTIDO POR IMÁGENES (200–2025), SEGÚN LA REVISIÓN SISTEMÁTICA

Tecnología	Cantidad	Porcentaje	Artículos
Deep Learning basado en CNN (clasificación / detección)	15	39.4%	02, 03, 05, 06, 07, 08, 11, 12, 14, 15, 17, 19, 21, 24, 31
Modelos híbridos CNN–SVM	7	18.4%	01, 04, 08, 16, 20, 22, 23
Arquitecturas DL avanzadas (U-Net, ResNet, DenseNet)	5	13.5%	25, 27, 29, 30, 32
Transformers / DL emergente	2	5.2%	10, 34
Machine Learning tradicional (RF, LR, etc.)	3	7.8%	13, 18, 26
IA clínica comparativa / Validación con radiólogos	2	5.2%	09, 28
Artículos marco (survey, guías, ética, implementación)	4	10.6%	33, 36, 37, 38

A continuación, la Tabla III señala las dificultades identificadas para la implementación efectiva de arquitecturas de Deep Learning en contextos clínicos reales

TABLA III
 DIFICULTADES IDENTIFICADAS PARA LA IMPLEMENTACIÓN DE MODELOS DE DEEP LEARNING EN CONTEXTOS CLÍNICOS REALES (2020–2025)

Dificultades	Cantidad	Porcentaje	Artículos
Técnicas, éticas, regulatorias e institucionales	12	31.6%	01, 03, 05, 08, 10, 14, 17, 19, 24, 26, 30, 36
Técnicas y éticas	7	18.4%	02, 06, 12, 16, 21, 27, 35
Técnicas e institucionales	5	13.2%	04, 09, 15, 20, 34
Éticas y regulatorias	5	13.2%	11, 18, 25, 37, 38
Solo técnicas	4	10.5%	13, 22, 28, 31
Solo regulatorias	2	5.3%	23, 29
Solo institucionales	2	5.3%	07, 33
Solo éticas	1	2.6%	32

A pesar de los grandes avances en arquitecturas de deep learning para el diagnóstico médico por imágenes, su implementación en la práctica clínica aún enfrenta desafíos multidimensionales y altamente interdependientes; según la evidencia resumida en la Tabla V, el 31.6% de los artículos (12) informa dificultades técnicas, éticas, regulatorias e institucionales combinadas, lo que demuestra que las barreras para la implementación clínica rara vez existen de forma aislada y que, en cambio, se combinan para impedir su implementación.

A nivel técnico, gran parte de la literatura encuentra problemas comunes como la interoperabilidad con los sistemas hospitalarios existentes, la heterogeneidad y falta de representatividad de las bases de datos clínicas y la generalización de los modelos ante cambios en los protocolos de adquisición, dispositivos médicos y características poblacionales, las cuales, en conjunto con las preocupaciones éticas, representan el 18.4% de los artículos (7 artículos), demostrando que el rendimiento algorítmico por sí solo no es suficiente para su adopción clínica sin mecanismos de explicabilidad y confianza.

Sobre las barreras técnicas e institucionales, presentes en el 13.2% de los artículos (5 artículos), la literatura informa que la complejidad de las arquitecturas de Deep Learning y la falta de capacidades técnicas internas, personal sanitario poco capacitado y falta de estrategias institucionales de integración tecnológica son barreras para su implementación sostenible en entornos asistenciales.

Por otro lado, los desafíos éticos y legales, que también aparecen en el 13.2% de los estudios (5 artículos), se refieren principalmente a la protección de datos clínicos confidenciales, el consentimiento informado, la rendición de cuentas por las decisiones algorítmicas y la gestión de sesgos en los datos de entrenamiento, ya que no existen marcos legales claros y actualizados para la IA en medicina, lo que crea incertidumbre legal sobre la responsabilidad profesional en caso de error de diagnóstico, frenando su adopción generalizada.

Además, un 10.5% de los artículos (4 estudios) solo encuentra dificultades técnicas, relacionadas con la capacidad computacional, complejidad del modelo y dependencia de grandes datos etiquetados; en menor medida, solo regulatorias y solo institucionales, con un 5.3% cada una (2 artículos respectivamente), asociadas a procesos de certificación, cumplimiento normativo y resistencia al cambio; y solo éticas, aunque menos común (2.6%, 1 artículo), pero que reflejan la creciente preocupación por la justicia, transparencia y uso responsable de la IA en la práctica clínica.

En conjunto, estos resultados ilustran que la aplicación exitosa de arquitecturas de deep learning en entornos clínicos reales necesita ir más allá del rendimiento algorítmico, y superar las limitaciones encontradas exigirá el desarrollo de marcos éticos y regulatorios robustos, así como políticas institucionales específicas para la formación, la gobernanza de datos y la sostenibilidad de la IA en los sistemas sanitarios.

A continuación, la Tabla IV señala los aportes recientes impulsados por Deep Learning en el diagnóstico médico

TABLA IV
 APORTES RECIENTES IMPULSADOS POR DEEP LEARNING EN EL DIAGNÓSTICO MÉDICO POR IMÁGENES (2020–2025)

Aporte	Cantidad	Porcentaje	Artículos
Computacional	16	42.1%	02, 03, 04, 05, 06, 08, 11, 12, 14, 15, 16, 17, 19, 21, 24, 34
Metodológico	13	34.2%	04, 05, 13, 18, 20, 22, 23,

			25, 28, 29, 33, 35, 38
Clínico	9	23.7%	07, 09, 10, 14, 17, 19, 26, 30, 36

El análisis de los 38 artículos de la revisión sistemática muestra que las arquitecturas de aprendizaje profundo han revolucionado el diagnóstico médico por imágenes en el periodo 2020-2025, con contribuciones en tres ejes principales: computacional, metodológico y clínico, que reflejan el nivel de madurez tecnológica y las áreas donde aún existen desafíos para su implementación en la práctica clínica.

Desde la perspectiva informática, el 42.1% de los artículos (16 artículos) proponen como principal contribución arquitecturas mejoradas de aprendizaje profundo, tales como CNN, ResNet, U-Net, DenseNet e, incipientemente, modelos basados en Transformers, que logran mejoras significativas en métricas de rendimiento diagnóstico precisión, sensibilidad, AUC y reducen los tiempos de procesamiento; en algunos casos, se utilizaron estrategias como transferencia de aprendizaje, optimización de arquitecturas y adaptación computacional, haciendo viables los modelos en entornos con limitaciones de recursos y acercándolos a la práctica clínica.

En términos metodológicos, el 34.2% de los artículos (13 estudios) hace contribuciones en diseños experimentales, validación de modelos y marcos conceptuales para la aplicación robusta del Deep Learning en medicina, incluyendo modelos híbridos (por ejemplo, CNN + SVM), esquemas de validación cruzada, estudios comparativos entre técnicas, estudios de explicabilidad (XAI) y encuestas/guías de buenas prácticas que sistematizan el conocimiento. Estas metodologías contribuyen a mejorar la reproducibilidad, interpretabilidad y confianza clínica en los sistemas de IA.

Por su parte, la mayor contribución clínica la hallamos en el 23.7% de los estudios (9 artículos), los cuales se centraron en evaluar el impacto real de los modelos en entornos clínicos, validándolos con datos clínicos reales y comparándolos directamente con radiólogos, mostrando un rendimiento similar o superior en métricas como sensibilidad, especificidad y AUC, así como en la detección temprana de enfermedades y en la ayuda en la toma de decisiones diagnósticas; sin embargo, la menor proporción de estudios clínicos revela que la traslación desde el desarrollo algorítmico a la práctica clínica sigue siendo un reto.

En conjunto, estos resultados sugieren que la evolución reciente del diagnóstico por imágenes se ha visto impulsada principalmente por mejoras computacionales y metodológicas, pero los estudios para validar clínicamente e implementar en el mundo real estas tecnologías aún son escasos, lo que refuerza la necesidad de futuros trabajos para cerrar la brecha entre el rendimiento algorítmico y su integración en los sistemas sanitarios con criterios de explicabilidad, gobernanza y evaluación clínica multicéntrica.

A continuación, la Tabla V sintetiza los beneficios clínicos más frecuentes reportados por el uso de Deep Learning en el diagnóstico médico.

TABLA V
BENEFICIOS CLÍNICOS MÁS FRECUENTES REPORTADOS POR EL USO DE DEEP LEARNING EN EL DIAGNÓSTICO MÉDICO POR IMÁGENES (2020–2025)

Beneficio	Cantidad	Porcentaje	Artículos
Mejora de la precisión diagnóstica	10	26.3%	02, 04, 05, 06, 08, 11, 14, 17, 19, 31
Aumento de la eficiencia operativa	8	21.1%	03, 09, 10, 12, 13, 16, 21, 24
Aplicabilidad clínica validada	7	18.4%	18, 20, 22, 25, 26, 30, 36
Interpretabilidad y confianza médica	6	15.8%	07, 15, 27, 35, 37, 38
Uso en múltiples modalidades de imagen	4	10.5%	14, 17, 21, 33
Escalabilidad / IoMT / Edge computing	3	7.9%	28, 32, 34

El análisis de los 38 artículos que componen esta revisión sistemática muestra que el uso de modelos de aprendizaje profundo en el diagnóstico por imágenes médicas proporciona beneficios clínicos medibles y superiores en el diagnóstico, los cuales se agrupan en seis dimensiones principales: precisión diagnóstica, eficiencia, aplicabilidad clínica verificada, interpretabilidad y confianza médica, capacidad multimodal y escalabilidad. La distribución de estos beneficios sugiere que el uso de arquitecturas de aprendizaje profundo en el diagnóstico por imágenes médicas ofrece ventajas clínicas reales y superiores en el diagnóstico, las cuales se benefician en gran medida de la precisión diagnóstica, la eficiencia, la aplicabilidad clínica verificada, la interpretabilidad y la confianza médica, la capacidad multimodal y la escalabilidad que proporcionan estas arquitecturas.

En primer lugar, la mejora de la precisión diagnóstica fue el principal beneficio en el 26.3% de los estudios (10 artículos), lo que se atribuye principalmente a la utilización de arquitecturas profundas como CNN, ResNet, U-Net y DenseNet, capaces de detectar lesiones, anomalías estructurales y patrones complejos, superando en algunos casos a los especialistas humanos en el diagnóstico, lo que los convierte en una herramienta de apoyo al diagnóstico.

En segundo lugar, la mejora de la eficiencia clínica fue el principal beneficio en el 21.1% de los artículos (8 estudios), donde automatizaron procesos clave como la clasificación, segmentación y priorización en el cribado inicial de imágenes, reduciendo significativamente los tiempos de análisis y la carga de trabajo del personal sanitario sin afectar la calidad del diagnóstico, lo cual es especialmente valioso en entornos clínicos con alta demanda.

La aplicabilidad clínica verificada se demostró en el 18.4% de los artículos (7 artículos) mediante validaciones con datos reales, pruebas en hospitales y comparaciones con especialistas, demostrando que los modelos de Deep Learning no solo tienen un buen rendimiento experimental, sino que también pueden integrarse en la práctica clínica,

especialmente en campos como oncología, neurología y diagnóstico de enfermedades pulmonares.

Por otro lado, la interpretabilidad y la confianza médica fueron el principal beneficio en el 15.8% de los artículos (6 estudios), donde el uso de técnicas XAI, como mapas de activación o explicaciones visuales, ayudó al personal sanitario a comprender las decisiones algorítmicas, aumentando la confianza y disminuyendo la percepción de los modelos como "cajas negras".

Además, la versatilidad en múltiples modalidades de imagen fue el principal beneficio en el 10.5% de los artículos (4 estudios), en los que los modelos demostraron adaptarse a distintas modalidades de imagen (tomografía computarizada, resonancia magnética, ecografía, mamografía), ampliando su aplicabilidad en diversas especialidades médicas.

Finalmente, la escalabilidad y la adaptación a entornos tecnológicos distribuidos (IoT, edge computing) se mencionó en el 7.9% de los artículos (3 estudios), siendo menos común, pero estratégico, demostrando que el Deep Learning se puede aplicar en contextos con recursos limitados, dispositivos portátiles o infraestructuras descentralizadas, abriendo la puerta a un diagnóstico asistido en áreas con pocos recursos.

En conjunto, estos hallazgos demuestran que los beneficios clínicos del Deep Learning en el diagnóstico médico por imágenes han pasado de mejoras en la exactitud a enfoques más holísticos que consideran eficiencia, confianza, validación clínica y escalabilidad, pero la dispersión sugiere que aún hay espacio para fortalecer la investigación hacia la implementación clínica sostenible y en el mundo real en diferentes entornos institucionales y tecnológicos.

IV. DISCUSIÓN

El objetivo de este estudio fue hacer una revisión sistemática sobre el impacto, las limitaciones, las contribuciones y los beneficios de los modelos de aprendizaje profundo en el diagnóstico por imágenes médicas en el periodo 2020-2025. Sobre el impacto del Deep Learning en el diagnóstico por imágenes médicas, los hallazgos indican que los modelos de Deep Learning han revolucionado el diagnóstico por imágenes, haciéndolo más preciso, eficiente y clínicamente útil. Diversos estudios han demostrado que arquitecturas profundas, como CNN y modelos híbridos CNN-SVM, pueden identificar patrones sutiles en imágenes médicas con una exactitud comparable o superior a la de los expertos humanos [1], [7], [9]. Este hallazgo es consistente con estudios neuroimagen, oncología y radiología torácica, en las que sensibilidad y especificidad diagnóstica mejoraron con IA [6], [13], [22].

Además, los datos indican que el DL puede disminuir la variabilidad interobservador, un antiguo problema en la lectura de imágenes médicas. Estudios comparativos entre radiólogos y sistemas IA muestran que estos últimos funcionan como una segunda opinión estandarizada, apoyando la toma de decisiones clínicas sin sustituirla [15], [21]. Pero estudios como el de Bernstein et al. alertan de que errores en

modelos de IA pueden sesgar el juicio clínico del médico, por lo que su uso se debe implementar con cautela y bajo supervisión [5].

Desde un punto de vista de la eficiencia, los modelos de Deep Learning son capaces de automatizar procesos que requieren mucho tiempo, como la segmentación, clasificación y preselección de imágenes médicas. Esto ha logrado disminuir los tiempos de análisis, mejorar el flujo de trabajo clínico y disminuir la carga laboral de los especialistas, sobre todo en lugares con alta demanda asistencial [10], [11], [17]. En conjunto, estos hallazgos indican que la IA no solo mejora la calidad del diagnóstico, sino que también apoya la sostenibilidad de los sistemas sanitarios actuales [24], [26].

En relación a las dificultades que limitan la implementación clínica efectiva del Deep Learning, se tiene que la literatura coincide en que llevar arquitecturas de Deep Learning a la práctica clínica real se enfrenta a obstáculos multidimensionales, que incluyen desafíos técnicos, éticos, legales e institucionales. En términos técnicos, la disponibilidad y calidad de los datos sigue siendo uno de los principales retos. La escasez de conjuntos de datos clínicos grandes, bien anotados y representativos restringe la generalización de los modelos y puede generar sesgos que comprometen la justicia en el diagnóstico [18], [22], [28].

Otro desafío técnico importante es la heterogeneidad de los sistemas de información hospitalarios, lo que dificulta la interoperabilidad y la integración perfecta de las soluciones de IA. Como indican White et al., la heterogeneidad tecnológica que se encuentra en muchas prácticas clínicas dificulta tanto la implementación inicial como el mantenimiento y la escalabilidad de los modelos en el tiempo [27]. Y esto se agrava en instituciones con pocos recursos en infraestructura digital o con poca experiencia en tecnologías emergentes [24], [26].

En ética y legislación, la opacidad de los modelos de aprendizaje profundo la llamada "caja negra" es un obstáculo importante para su uso clínico. La falta de interpretabilidad dificulta la justificación de las decisiones automatizadas y crea desconfianza entre los profesionales sanitarios [10], [11], [14]. Además, la falta de marcos legales definidos para la IA médica deja abierta la cuestión de la responsabilidad en caso de errores de diagnóstico y la protección de datos confidenciales y consentimiento informado [5], [21].

Finalmente, los factores institucionales y humanos son cruciales. La resistencia al cambio organizacional, la falta de capacitación en IA entre los profesionales de la salud y la ausencia de políticas institucionales para la внедрение tecnológica han llevado a una implementación aislada de estas soluciones [6], [8], [19]. En esta línea, la literatura concuerda en que la apropiación exitosa del DL envuelve estrategias holísticas de capacitación continua, gobernanza ética y gestión del cambio organizacional [17], [25].

En relación a los aportes del Deep Learning a la evolución de las prácticas diagnósticas, se tiene que los últimos avances en Deep Learning están revolucionando la forma en que se realiza el diagnóstico por imágenes, convirtiéndolas en el

núcleo de la innovación clínica actual. Una de las mayores contribuciones es la creación de arquitecturas cada vez más complejas que pueden manejar datos multimodales y extraer características de alto valor clínico más allá de la evaluación visual humana [19], [31].

Además, el uso de métodos como transfer learning, destilación de conocimiento y modelos híbridos ha logrado disminuir los recursos computacionales necesarios para pasar de prototipos experimentales a aplicaciones clínicas reales [1], [13], [16], [20]. Esta optimización informática es especialmente pertinente para entornos hospitalarios con bajos recursos, donde la factibilidad técnica es un determinante para la implementación de nuevas tecnologías.

Otra línea importante es la creciente incorporación de métodos de explicabilidad y ayuda a la toma de decisiones para ajustar las predicciones algorítmicas al razonamiento clínico humano. Estudios recientes muestran que la adición de mecanismos explicativos aumenta la confianza del profesional y permite una mejor interacción humano-sistema inteligente [15], [24]. Además, la creciente tendencia de modelos multimodales que integren imágenes médicas con datos clínicos y demográficos crea nuevas oportunidades para una medicina más personalizada y centrada en el paciente [17], [28].

En cuanto a los beneficios clínicos reportados y mejora de los procesos diagnósticos, se tiene que los beneficios clínicos del uso de Deep Learning en el diagnóstico por imágenes médicas son amplios y homogéneos en las diferentes especialidades estudiadas. La evidencia apoya que estas tecnologías mejoran la exactitud diagnóstica, permitiendo detectar enfermedades en etapas tempranas y reconocer hallazgos sutiles que pueden ser fácilmente pasados por alto por radiólogos experimentados [7], [18], [22]. Este impacto se traduce en decisiones terapéuticas más oportunas y en una reducción de complicaciones asociadas a diagnósticos tardíos o erróneos [5], [9], [14].

Además, la automatización parcial del análisis de imágenes ha hecho más eficientes y rápidos los servicios de salud, disminuyendo los tiempos de espera y optimizando el uso de personal y tecnología [5], [11]. Esta optimización permite que los profesionales se enfoquen en los casos más complejos y de mayor riesgo, mejorando la seguridad del paciente y la calidad de la atención [3], [10], [27].

Finalmente, la integración de sistemas de DL permite combinar la IA con el juicio clínico, estableciendo una cultura de aprendizaje continuo y toma de decisiones basada en evidencia. Esta mezcla refuerza los resultados diagnósticos, pero también apoya la formación y actualización continua del personal de salud, convirtiendo al DL en un motor de transformación y sostenibilidad de la atención sanitaria [6], [17], [26], [28].

V. CONCLUSIONES

La evidencia revisada demuestra que las arquitecturas profundas, como las redes neuronales convolucionales y los modelos híbridos, han mejorado consistentemente la exactitud diagnóstica, disminuyendo la variabilidad interobservador y reforzando la toma de decisiones clínicas. Además, los resultados muestran que el Deep Learning no solo alcanza mayor precisión algorítmica, sino que esto se traduce en mejoras en la eficiencia de los servicios sanitarios. La automatización de procesos como la segmentación, el cribado y la clasificación de imágenes ha optimizado los flujos de trabajo clínico, disminuyendo los tiempos de análisis y reduciendo la carga de trabajo del personal sanitario, especialmente en situaciones de alta demanda asistencial.

Pero esta revisión también muestra que la traslación clínica de estas tecnologías todavía tiene camino por recorrer. Las limitaciones técnicas de disponibilidad y representatividad de los datos, la generalización de los modelos y la integración con los sistemas hospitalarios, junto con las barreras éticas, legales e institucionales, evitan su amplia adopción [18], [27], [35], [36]. Sobre todo, la falta de interpretabilidad de muchas de estas herramientas y la ausencia de marcos legales generan desconfianza e imposibilitan su integración sostenible en la práctica clínica.

Desde la mirada investigativa, los resultados indican que la literatura actual se ha enfocado en las contribuciones computacionales y metodológicas, y aún son limitados los estudios de validación clínica multicéntrica, gobernanza ética y adopción organizacional. Esta diferencia indica que futuras investigaciones deberían enfocarse en metodologías más holísticas que evalúen no solo el rendimiento técnico, sino también la seguridad, la transparencia y la aceptabilidad clínica de los sistemas DL.

En conclusión, los hallazgos de esta revisión reafirman que el Deep Learning es un método para transformar el diagnóstico médico por imágenes, haciéndolo más preciso, eficiente y sostenible. Pero su beneficio clínico total dependerá de la convergencia de avances tecnológicos, validación clínica rigurosa, marcos éticos sólidos y políticas institucionales que permitan una implementación responsable y centrada en el paciente.

REFERENCES

- [1] Q. D. Buchlak *et al.*, "Effects of a comprehensive brain computed tomography deep learning model on radiologist detection accuracy," *Eur. Radiol.*, vol. 34, no. 2, pp. 810–822, 2024, doi: 10.1007/S00330-023-10074-8.
- [2] L. J. S. Ewals *et al.*, "The Impact of Expectation Management and Model Transparency on Radiologists' Trust and Utilization of AI Recommendations for Lung Nodule Assessment on Computed Tomography: Simulated Use Study," *JMIR AI*, vol. 3, e52211, 2024, doi: 10.2196/52211.
- [3] T. J. Marini *et al.*, "No sonographer, no radiologist: Assessing accuracy of artificial intelligence on breast ultrasound volume sweep imaging scans," *PLOS Digit. Health*, vol. 1, no. 11, 2022, doi: 10.1371/JOURNAL.PDIG.0000148.

- [4] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
- [5] C. S. Bhat *et al.*, "Artificial intelligence for interpretation of segments of whole body MRI in CNO: Pilot study comparing radiologists versus machine learning algorithm," *Pediatr. Rheumatol.*, vol. 18, no. 1, 2020, doi: 10.1186/S12969-020-00442-9.
- [6] M. D. Jørgensen, R. Antulov, S. Hess, and S. Lysdahlgaard, "Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: A systematic review and meta-analysis," *Eur. J. Radiol.*, vol. 146, Art. no. 110073, 2022, doi: 10.1016/J.EJRAD.2021.110073.
- [7] A. Hatamizadeh *et al.*, "Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images," *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2196–2208, 2022, doi: 10.1109/TMI.2022.3154312.
- [8] T. Huo *et al.*, "Deep learning-based algorithm improves radiologists' performance in lung cancer bone metastases detection on computed tomography," *Front. Oncol.*, vol. 13, Art. no. 1125637, 2023, doi: 10.3389/FONC.2023.1125637.
- [9] I. D. Naranjo *et al.*, "Breast Lesion Classification with Multiparametric Breast MRI Using Radiomics and Machine Learning: A Comparison with Radiologists' Performance," *Cancers*, vol. 14, no. 7, Art. no. 1743, 2022, doi: 10.3390/CANCERS14071743.
- [10] Y. Liu *et al.*, "Artificial intelligence versus radiologist in the accuracy of fracture detection based on computed tomography images: a multidimensional, multi-region analysis," *Quant. Imaging Med. Surg.*, vol. 13, no. 10, pp. 6424–6433, 2023, doi: 10.21037/QIMS-23-428.
- [11] W. Yu *et al.*, "Multi-scale, domain knowledge-guided attention + random forest: a two-stage deep learning-based multi-scale guided attention models to diagnose idiopathic pulmonary fibrosis from computed tomography images," *Med. Phys.*, vol. 50, no. 2, pp. 894–905, 2023, doi: 10.1002/MP.16053.
- [12] S. Alimi, A. O. Kuyoro, M. O. Eze, and O. Akande, "Utilizing Deep Learning and SVM Models for Schizophrenia Detection and Symptom Severity Estimation Through Structural MRI," *Ingénierie des Systèmes d'Information*, vol. 28, no. 4, pp. 993–1002, 2023, doi: 10.18280/ISI.280419.
- [13] J. E. Aurelia and Z. Rustam, "A hybrid convolutional neural network-support vector machine for x-ray computed tomography images on cancer," *Open Access Macedonian J. Med. Sci.*, vol. 9, no. B, pp. 1283–1289, 2021, doi: 10.3889/OAMJMS.2021.6955.
- [14] A. E. Minarno *et al.*, "Classification of Brain Tumors on MRI Images Using DenseNet and Support Vector Machine," *Int. J. Informatics Vis.*, vol. 6, no. 2, pp. 404–410, 2022, doi: 10.30630/JOIV.6.2.991.
- [15] V. Rajinikanth, S. Kadry, and Y. Nam, "Convolutional-neural-network assisted segmentation and SVM classification of brain tumor in clinical MRI slices," *Inf. Technol. Control*, vol. 50, no. 2, pp. 342–356, 2021, doi: 10.5755/JOI.ITC.50.2.28087.
- [16] S. Suryawanshi and S. B. Patil, "Efficient Brain Tumor Classification with a Hybrid CNN-SVM Approach in MRI," *J. Adv. Inf. Technol.*, vol. 15, no. 3, pp. 340–354, 2024, doi: 10.12720/JAIT.15.3.340-354.
- [17] O. A. M. F. Alnaggar, B. N. Jagadale, and S. H. Narayan, "MRI Brain Tumor Detection Using Boosted Crossbred Random Forests and Chimp Optimization Algorithm Based Convolutional Neural Networks," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 36–46, 2022, doi: 10.22266/IJIES2022.0430.04.
- [18] F. Çelebi *et al.*, "Artificial Intelligence in Diagnostic Breast Ultrasound: A Comparative Analysis of Decision Support Among Radiologists With Various Levels of Expertise," *Eur. J. Breast Health*, vol. 21, no. 1, pp. 33–39, 2025, doi: 10.4274/EJBH.GALENOS.2024.2024-9-7.
- [19] M. S. Kadafi, A. K. Yaqubi, Purbandini, and S. D. Astuti, "Alzheimer's prediction via CNN-SVM on chatbot platform with MRI," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 1, pp. 64–73, 2024, doi: 10.11591/IJEECS.V36.I1.PP64-73.
- [20] L. J. S. Ewals *et al.*, "The Effects of Artificial Intelligence Assistance on the Radiologists' Assessment of Lung Nodules on CT Scans: A Systematic Review," *J. Clin. Med.*, vol. 12, no. 10, 2023, doi: 10.3390/JCM12103536.
- [21] L. Gilberg *et al.*, "Deep Learning Enhances Radiologists' Detection of Potential Spinal Malignancies in CT Scans," *Appl. Sci.*, vol. 13, no. 14, Art. no. 8140, 2023, doi: 10.3390/AP13148140.
- [22] K. Kikuchi *et al.*, "Comparison of diagnostic performance of radiologist- and AI-based assessments of T2-FLAIR mismatch sign and quantitative assessment using synthetic MRI in the differential diagnosis between astrocytoma, IDH-mutant and oligodendroglioma, IDH-mutant and 1p/19q-codeleted," *Neuroradiology*, vol. 66, no. 3, pp. 333–341, 2024, doi: 10.1007/S00234-024-03288-0.
- [23] R. D. White *et al.*, "Pre-deployment assessment of an AI model to assist radiologists in chest X-ray detection and identification of lead-less implanted electronic devices for pre-MRI safety screening: realized implementation needs and proposed operational solutions," *J. Med. Imaging*, vol. 9, no. 5, 2022, doi: 10.1117/1.JMI.9.5.054504.
- [24] B. Teodorescu *et al.*, "Advancements in opportunistic intracranial aneurysm screening: The impact of a deep learning algorithm on radiologists' analysis of T2-weighted cranial MRI," *J. Stroke Cerebrovasc. Dis.*, vol. 33, no. 12, 2024, doi: 10.1016/J.JSTROKECEREBROVASDIS.2024.108014.
- [25] A. Urushibara *et al.*, "The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: a comparison with radiologists," *BMC Med. Imaging*, vol. 22, no. 1, 2022, doi: 10.1186/S12880-022-00808-3.
- [26] J. C. M. van Zelst, T. Tan, R. M. Mann, and N. Karssemeijer, "Validation of radiologists' findings by computer-aided detection (CAD) software in breast cancer detection with automated 3D breast ultrasound: a concept study in implementation of artificial intelligence software," *Acta Radiol.*, vol. 61, no. 3, pp. 312–320, 2020, doi: 10.1177/0284185119858051.
- [27] A. Arslan *et al.*, "Does deep learning software improve the consistency and performance of radiologists with various levels of experience in assessing bi-parametric prostate MRI?," *Insights into Imaging*, vol. 14, no. 1, 2023, doi: 10.1186/S13244-023-01386-W.
- [28] C. Jacobs *et al.*, "Deep learning for lung cancer detection on screening CT scans: Results of a large-scale public competition and an observer study with 11 radiologists," *Radiol.: Artif. Intell.*, vol. 3, no. 6, 2021, doi: 10.1148/RYAI.2021210027.
- [29] M. H. Bernstein *et al.*, "Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography," *Eur. Radiol.*, vol. 33, no. 11, pp. 8263–8269, 2023, doi: 10.1007/S00330-023-09747-1.

- [30]D. Xu *et al.*, “The clinical value of artificial intelligence in assisting junior radiologists in thyroid ultrasound: a multicenter prospective study from real clinical practice,” *BMC Med.*, vol. 22, no. 1, 2024, doi: 10.1186/S12916-024-03510-Z.
- [31]P. Zhang, Z. Ma, Y. Zhang, X. Chen, and G. Wang, “Improved Inception V3 method and its effect on radiologists’ performance of tumor classification with automated breast ultrasound system,” *Gland Surg.*, vol. 10, no. 7, pp. 2232–2245, 2021, doi: 10.21037/GS-21-328.
- [32]C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, “Key challenges for delivering clinical impact with artificial intelligence,” *BMC Med.*, vol. 17, Art. no. 195, 2019, doi: 10.1186/s12916-019-1426-2.
- [33]M. Goto *et al.*, “Use of a deep learning algorithm for non-mass enhancement on breast MRI: comparison with radiologists’ interpretations at various levels,” *Jpn. J. Radiol.*, vol. 41, no. 10, pp. 1094–1103, 2023, doi: 10.1007/S11604-023-01435-W.
- [34]S. Y. Youn *et al.*, “Detection and PI-RADS classification of focal lesions in prostate MRI: Performance comparison between a deep learning-based algorithm (DLA) and radiologists with various levels of experience,” *Eur. J. Radiol.*, vol. 142, Art. no. 109894, 2021, doi: 10.1016/J.EJRAD.2021.109894.
- [35]Z. Zhang and Y. Han, “Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier with an Advanced Machine Learning Approach,” *IEEE Access*, vol. 8, pp. 44999–45008, 2020, doi: 10.1109/ACCESS.2020.2977962.
- [36]A. B. Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges,” *Inf. Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.
- [37]E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nat. Med.*, vol. 25, no. 1, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.
- [38]A. Esteva *et al.*, “A guide to deep learning in healthcare,” *Nat. Med.*, vol. 25, no. 1, pp. 24–29, 2019, doi: 10.1038/s41591-018-0316-z.