

Inteligencia artificial aplicada al análisis de imágenes histológicas para apoyo al diagnóstico: una revisión sistemática de la evidencia 2025-2026

Artificial intelligence applied to histological image analysis for diagnostic support: a systematic review of the evidence 2025-2026

Inteligência artificial aplicada à análise de imagens histológicas para apoio ao diagnóstico: uma revisão sistemática da evidência 2025-2026

Adán Joel Villanueva Sosa, <https://orcid.org/0000-0001-5784-6929>

Universidad César Vallejo, Chiclayo-Doctorado en Gestión Pública y Gobernabilidad

Autor para correspondencia: avillanuevaso@ucvvirtual.edu.pe

RESUMEN

La inteligencia artificial (IA) aplicada a imágenes histológicas digitalizadas ha evolucionado rápidamente hacia tareas diagnósticas complejas. El objetivo de este trabajo es sintetizar la evidencia publicada entre 2025 y 2026 sobre aplicaciones de IA en histopatología diagnóstica. Como parte de la metodología se utilizó la revisión sistemática según PRISMA 2020 en PubMed, Google Académico y ScienceDirect, incluyendo estudios originales con imágenes histológicas digitalizadas para diagnóstico, clasificación o estratificación clínica. Se incluyeron 22 estudios, con predominio de patología oncológica (tracto gastrointestinal, mama, sistema nervioso central). Las arquitecturas de IA evolucionaron desde redes neuronales convolucionales hacia aprendizaje multi-instancia, transformers y foundation models. Más de la mitad reportó validación externa. Las métricas principales fueron AUROC, accuracy e índices de concordancia. La IA en patología digital muestra creciente sofisticación, pero persiste heterogeneidad metodológica que limita su implementación clínica generalizada.

Palabras clave: Inteligencia artificial; histopatología; diagnóstico; aprendizaje profundo; revisión sistemática.

ABSTRACT

Artificial intelligence (AI) applied to digital histological images has rapidly evolved toward complex diagnostic tasks. To synthesize the evidence published between 2025 and 2026 on AI applications in diagnostic histopathology. Methodology: Systematic review following PRISMA 2020 in PubMed, Google Scholar, and ScienceDirect, including original human studies using digital histological images for diagnosis, classification, or clinical stratification. Twenty-two studies were included, predominantly in oncologic pathology (gastrointestinal tract, breast, central nervous system). AI architectures evolved from convolutional neural networks toward multi-instance learning, transformers, and foundation models. More than half reported external validation. Main metrics were AUROC, accuracy, and concordance indices. AI in digital pathology shows increasing sophistication, but methodological heterogeneity persists, limiting widespread clinical implementation.

Keywords: Artificial intelligence; histopathology; diagnosis; deep learning; systematic review.

RESUMO

A inteligência artificial (IA) aplicada a imagens histológicas digitalizadas evoluiu rapidamente para tarefas diagnósticas complexas. Sintetizar as evidências publicadas entre 2025 e 2026 sobre aplicações de IA em histopatologia diagnóstica. Revisão sistemática segundo PRISMA 2020 nas bases PubMed, Google Acadêmico e ScienceDirect, incluindo estudos originais em humanos com imagens histológicas digitalizadas para diagnóstico, classificação ou estratificação clínica. Foram incluídos 22 estudos, com predomínio de patologia oncológica (trato gastrointestinal, mama, sistema nervoso central). As arquiteturas de IA evoluíram de redes neurais convolucionais para aprendizado multi-instância, transformers e foundation models. Mais da metade

reportou validação externa. As principais métricas foram AUROC, acurácia e índices de concordância. A IA em patologia digital mostra crescente sofisticação, mas persiste heterogeneidade metodológica que limita sua implementação clínica generalizada.

Palavras-chave: Inteligência artificial; histopatologia; diagnóstico; aprendizado profundo; revisão sistemática.

Recibido: 25/4/2025 Aprobado: 7/5/2026

Introducción

La digitalización de la anatomía patológica mediante imágenes de lámina completa (whole-slide images, WSI) ha transformado el análisis histológico en la última década, permitiendo la aplicación sistemática de técnicas de inteligencia artificial (IA) al diagnóstico en patología humana. Al inicio, los enfoques de aprendizaje automático (machine learning) se basaban en la extracción manual de características morfológicas y texturales; sin embargo, el desarrollo de arquitecturas de aprendizaje profundo (deep learning) ha permitido la construcción de modelos capaces de aprender representaciones jerárquicas directamente desde imágenes histológicas digitalizadas, alcanzando desempeños comparables o superiores a la evaluación humana en tareas específicas (Treillard *et al.*, 2026).

En oncología, la IA aplicada a histopatología ha evolucionado desde tareas básicas de clasificación tisular hasta modelos capaces de predecir pronóstico, alteraciones genómicas y respuesta terapéutica a partir de imágenes teñidas con hematoxilina-eosina (H&E). Takamatsu (2025) indicó que la integración de grandes bases de datos como The Cancer Genome Atlas (TCGA), junto con modelos de aprendizaje profundo, ha permitido avanzar hacia una patología computacional con implicancias diagnósticas y terapéuticas, incluyendo la estratificación molecular basada exclusivamente en morfología digital. De manera complementaria, Marra *et al.* (2025) señalaron que la IA está transitando de un rol asistencial a uno potencialmente integrado en la toma de decisiones clínicas, particularmente en tumores sólidos.

Diversas revisiones sistemáticas por órgano han confirmado la rápida expansión de estas aplicaciones. En mama, Li *et al.* (2025) describieron avances en clasificación tumoral, predicción de receptores hormonales y estimación de riesgo pronóstico mediante modelos de aprendizaje profundo. En hígado, Grignaffini *et al.* (2024) reportaron aplicaciones orientadas tanto al diagnóstico como a la evaluación de fibrosis y esteatosis mediante análisis automatizado. En tiroides, Kussaibi y Alsafwani (2023) hallaron una tendencia hacia modelos de clasificación binaria y multiclase con métricas de desempeño crecientemente robustas.

El enfoque metodológico también se ha diversificado. La aplicación de aprendizaje federado ha emergido como una estrategia para entrenar modelos colaborativos sin compartir datos sensibles entre instituciones, abordando limitaciones regulatorias y de privacidad (Touhami *et al.*, 2026). Revisiones centradas en características biológicamente interpretables han planteado la necesidad de mayor explicabilidad y transparencia en modelos de IA, especialmente frente a marcos regulatorios emergentes (Treillard *et al.*, 2026). Estas preocupaciones se alinean con la discusión sobre validación externa, reproducibilidad y heterogeneidad interinstitucional, aspectos críticos para la implementación clínica.

En ese sentido, el objetivo del presente estudio fue revisar y sintetizar la evidencia científica reciente (2025-2026) sobre la aplicación de inteligencia artificial al análisis de imágenes histológicas para apoyo al diagnóstico en patología humana. Específicamente, se describieron las áreas clínicas y órganos estudiados, las tareas diagnósticas abordadas, los tipos de imágenes y arquitecturas de modelos empleadas, así como el tamaño muestral, la presencia de validación externa y las principales métricas de desempeño reportadas, con el fin de identificar tendencias metodológicas, niveles de madurez y limitaciones relevantes para su potencial implementación clínica.

Metodología

Se realizó una revisión sistemática conforme a la guía PRISMA 2020 (Page *et al.*, 2021), con el objetivo de identificar estudios recientes que aplicaran inteligencia artificial al análisis de imágenes histológicas digitalizadas con finalidad diagnóstica en patología humana. La búsqueda bibliográfica se efectuó en PubMed, Google Académico y ScienceDirect, restringiendo los resultados a publicaciones de los años 2025 y 2026.

En PubMed (búsqueda realizada en febrero de 2026) se utilizó la siguiente ecuación: ((“Deep Learning”[MeSH] OR “Artificial Intelligence”[MeSH] OR “machine learning”[Title/Abstract])) AND (“Histopathology”[Title/Abstract] OR “histopathological image”[Title/Abstract] OR “histological image”[Title/Abstract] OR “whole slide image*”[Title/Abstract] OR “digital pathology”[Title/Abstract]) AND (diagnosis[Title/Abstract] OR

detection[Title/Abstract] OR classification[Title/Abstract]) NOT (radiology OR MRI OR CT OR ultrasound OR cytology). Esta estrategia identificó 184 registros; tras aplicar el filtro por año (2025-2026) y realizar la evaluación por título y resumen, se seleccionaron 16 estudios que fueron confirmados tras revisión a texto completo.

En Google Académico se empleó la búsqueda “whole slide image diagnosis benign malignant 2025-review-segmentation-grading-arxiv”, que arrojó 30 registros iniciales; cinco cumplieron criterios preliminares y, luego de la revisión completa, se incluyó un estudio. En ScienceDirect se utilizó la estrategia “whole slide image AND differential diagnosis AND deep learning”, aplicando filtros por año (2025-2026) y tipo de documento (“research articles”); de 95 registros iniciales, 36 cumplieron los filtros y, tras la evaluación por elegibilidad, cinco estudios fueron incluidos.

En total se identificaron 309 registros. Luego de la eliminación de duplicados y de la exclusión de estudios que no cumplieran criterios clínicos estrictos (infraestructura sin aplicación diagnóstica directa, benchmarking sin cohorte clínica independiente o ausencia de validación adecuada), 22 estudios conformaron la síntesis cualitativa final (16 de PubMed, 1 de Google Académico y 5 de ScienceDirect, con depuración posterior por duplicidad y elegibilidad final conforme al diagrama PRISMA).

Se incluyeron estudios originales en población humana que utilizaran imágenes histológicas digitalizadas, principalmente whole-slide images, y aplicaran técnicas de aprendizaje automático o profundo con finalidad diagnóstica, diagnóstico diferencial, detección o estratificación con implicancia clínica directa. Se excluyeron estudios centrados exclusivamente en segmentación sin componente diagnóstico, revisiones, encuestas, preprints no indexados y trabajos basados en modalidades de imagen no histológicas. La extracción de datos se realizó de manera estructurada considerando año de publicación, órgano o sistema estudiado, tipo de imagen, arquitectura del modelo, tamaño muestral, validación externa y métricas principales de desempeño. Los hallazgos se sintetizaron de forma cualitativa enfatizando la aplicabilidad diagnóstica, la solidez metodológica y la presencia de validación externa.

RESULTADOS Y DISCUSIÓN

Se incluyeron 22 estudios originales publicados entre 2025 y 2026 que evaluaron la aplicación de inteligencia artificial al análisis de imágenes histológicas digitalizadas con finalidad diagnóstica, pronóstica o de estratificación clínica. La mayoría de los estudios se desarrolló en Asia y Europa, con predominio de China como país del autor correspondiente, seguido de Alemania, Corea del Sur, Reino Unido, Estados Unidos y consorcios multicéntricos europeos, evidenciándose una concentración geográfica de la producción científica en centros con infraestructura consolidada en patología digital.

Distribución por país (n=22): China (8), Alemania (3), Corea del Sur (2), Reino Unido (2), Estados Unidos (2), consorcios europeos (2), otros (3). (Nota: esta distribución se basa en los autores correspondientes de los estudios incluidos).

Los órganos y sistemas más frecuentemente estudiados fueron el tracto gastrointestinal (cáncer gástrico, colorrectal y esofágico), el cáncer de mama, las neoplasias uroteliales y los tumores del sistema nervioso central. Un número menor de investigaciones abordó patologías específicas o de menor prevalencia, como tumores odontogénicos, enfermedades inflamatorias hepáticas autoinmunes, dermatosis inflamatorias, biopsias de arteria temporal y enfermedad de Hirschsprung. En términos generales, predominó la patología oncológica, reflejando el interés por aplicaciones con alto impacto diagnóstico, pronóstico y terapéutico.

Distribución por órgano o sistema (n=22): Tracto gastrointestinal (6), mama (4), sistema nervioso central (3), urotelio (2), odontogénico (2), hígado (1), piel (1), útero (1), próstata (1), pleura (1).

Todos los estudios utilizaron imágenes histológicas digitalizadas, principalmente whole-slide images (WSI) teñidas con hematoxilina-eosina. Varios trabajos incorporaron análisis basados en patches, tiles o regiones de interés derivados de las WSI, y algunos incluyeron citología digital o bloques celulares como modalidades complementarias. Las tareas de inteligencia artificial abarcaron clasificación diagnóstica, diagnóstico diferencial, detección de lesiones específicas, grading histológico, estratificación pronóstica, predicción de biomarcadores moleculares y respuesta a terapia neoadyuvante, evidenciando una expansión del rol de la patología digital más allá del diagnóstico morfológico tradicional.

Desde el punto de vista metodológico, las arquitecturas más empleadas fueron redes neuronales convolucionales (CNN), aprendizaje multi-instancia (MIL) y, de manera creciente, transformers y foundation models entrenados mediante aprendizaje auto-supervisado. El enfoque MIL fue particularmente frecuente en escenarios con anotación limitada a nivel de slide, mientras que los foundation models se asociaron a estudios con grandes volúmenes de datos y objetivos de generalización multicéntrica. Algunos trabajos incorporaron

estrategias multimodales o modelos vision-language, especialmente en tareas de pronóstico y cuantificación de biomarcadores.

El tamaño muestral mostró una marcada heterogeneidad, con estudios que incluyeron desde decenas de WSI hasta decenas de miles de imágenes en fases de preentrenamiento. Más de la mitad de los estudios reportó validación externa mediante cohortes independientes, evaluaciones multicéntricas o datasets internacionales, mientras que el resto se limitó a validaciones internas o particiones temporales. Los estudios con mayor tamaño muestral y validación externa se concentraron principalmente en cáncer gastrointestinal, mama, urotelio y tumores del sistema nervioso central.

Las métricas de desempeño variaron según la tarea clínica abordada, reportándose principalmente AUROC, accuracy, F1-score, balanced accuracy, índices de concordancia y hazard ratios. En general, los modelos mostraron desempeños elevados en validaciones internas, con una reducción moderada al evaluarse en cohortes externas; sin embargo, varios estudios demostraron estabilidad del rendimiento en escenarios multicéntricos. En los trabajos de naturaleza pronóstica o de estratificación de riesgo, los modelos basados en aprendizaje profundo mostraron asociaciones estadísticamente significativas con desenlaces clínicos relevantes, sugiriendo un potencial valor añadido frente a enfoques convencionales.

Tabla 1. Características metodológicas y clínicas de los estudios incluidos sobre inteligencia artificial aplicada al análisis de imágenes histológicas con finalidad diagnóstica (2025-2026)

Autores	Órgano/Patología	Tipo de Imagen	Tarea de IA	Tipo de Modelo	N pacientes/WSI	Validación Externa	Métrica Principal
Wang et al. (2025a)	Cáncer gastrointestinal	WSI (H&E), patches	Pronóstico / estratificación de riesgo	Foundation model (autosupervisado)	104.876 WSI (pre); 4.213 pacientes (val)	Sí (7 cohortes)	C-index 0,714-0,797
Gustav et al. (2025)	Cáncer colorrectal	WSI (H&E)	Clasificación molecular	Transformer	1.376 (entren) + 536 (val ext)	Sí	AUROC 0,78-0,93
Markowski et al. (2025)	Cáncer próstata metastásico	WSI (H&E)	Pronóstico / estratificación	IA multimodal	456 pacientes	No	HR, supervivencia
Lou et al. (2025)	Cáncer colorrectal	WSI (H&E)	Pronóstico / beneficio quimioterapia	MIL	883 WSI (2 cohortes)	Sí	C-index, HR
Zhang et al. (2025a)	Cáncer de mama	WSI (H&E), ROI	Clasificación molecular / subtipado	MIL (BBMIL)	No especificado (multicéntrico)	Sí (multicéntrico)	AUROC
Peng et al. (2026)	Carcinoma urotelial	WSI (H&E), tiles	Pronóstico / biomarcadores	CNN + Vision-Language (CONCH)	805 pacientes	Sí	C-index 0,672-0,795
Tan et al. (2025)	Esófago (Barrett)	Citología digital	Estratificación riesgo / displasia	Machine learning (no especificado)	910 pacientes	Sí	PPV, NPV, RR
Schirris et al. (2025)	Cáncer de mama	WSI (H&E)	Cuantificación TIL / pronóstico	Foundation model + regresión DL	2.340 pacientes	Sí	AUROC 0,80-0,94, HR 0,85
Zhang et al. (2025b)	Cáncer gástrico y colorrectal	WSI (H&E) + features nucleares	Clasificación molecular (MSI, TMB)	MIL (CLAM) + CNN (Hover-Net)	TCGA: 350+376 y 400+387	Sí (dataset externo)	AUC hasta 0,81
Zhou et al. (2025)	Cáncer de mama	WSI (H&E)	Predicción respuesta a terapia neoadyuvante	Foundation extractor (UNI) + MIL	826 pacientes	Sí	AUC 0,819 externa
Wang et al. (2026)	Útero (tumores músculo liso)	WSI	Diagnóstico (benigno vs maligno)	MIL (weakly supervised)	728 pacientes (1.797 WSI) + 117 ext	Sí	AUC 0,9976 int; 0,9889 ext
Giraldo-Roldán et al. (2026)	Tumores odontogénicos	WSI (H&E) + 455.107 patches	Clasificación multicategoría	CNN (DenseNet, EfficientNet, etc.)	64 WSI	No	AUC hasta 0,91; accuracy 0,79
Bharadwaj et al. (2025)	Lesiones odontogénicas (OKC vs DC)	WSI H&E 20x; 684 tiles	Diagnóstico diferencial binario	CNN propio OdonGAP	110 WSI	No	Accuracy 73,72%
Doeleman et al. (2025)	Piel (MF vs dermatosis benignas)	WSI H&E	Diagnóstico diferencial	Weakly supervised DL (MIL/atención)	924 WSI; 233 MF + 353 benignos	No (test temporal)	AUC 0,827; balanced acc 76,2%
Lee et al. (2026)	Gástrico (GIST/ leiomioma/ schwannoma)	WSI H&E	Clasificación diagnóstica (3 clases)	CNN + CutMix + Confident Instance Voting	No indicado	Sí	Accuracy 0,9262 int; mejora F1 ext

Lalchungnunga et al. (2025)	SNC (52 tipos tumorales)	WSI	Clasificación diagnóstica (top-1/top-2)	Deep learning (Neuropath-AI)	Entren: 5.835; Test: 5.516	Sí (multi-institucional)	Top-1 acc 80% (bal 66%); Top-2 acc 86%
Wang et al. (2025b)	Derrame pleural (MPE vs BPE)	WSI citología (smears + cell blocks)	Diagnóstico (maligno vs benigno)	Self-supervised + atención multi-escala (IMA-SSL; MIL)	194 smears + 188 cell block	No indicado	p<0,001 vs SOTA
Bourgade et al. (2025)	Biopsia arteria temporal (GCA)	WSI (H&E-saffron)	Diagnóstico (GCA vs control)	CTransPath + attention-MIL	366 entren; 58 test ext	Sí	AUROC 0,987 (CV); 0,994 (ext)
Gerussi et al. (2025)	Hígado (AIH vs PBC)	WSI H&E	Diagnóstico diferencial	Transformer-based ALNE	354 entren; 92 ext	Sí	AUC 0,81 externa
Azam et al. (2025)	SNC (meningioma vs SFT)	WSI	Diagnóstico diferencial	Conical transformers + voting	92 pacientes	No (CV)	Accuracy 92,27%; CV 94,68%; F1 95,07%
Steimetz et al. (2025)	Colon (adenomas LGD vs HGD)	WSI + patches	Grading / clasificación	CNN ResNet34 + agregación patches	200 slides (160 entren, 40 test)	No	Accuracy 95,0%; AUC 0,981; F1 0,923
Demir et al. (2025)	Hirschsprung (células ganglionares)	WSI H&E + patches	Detección diagnóstica	CNN ResNet-50 + Grad-CAM	668 slides (164 pacientes, 3 centros)	Sí	Accuracy 91,3%/92,8%/90,1%; mejora lectores 77%→85,8%

Nota: WSI: whole-slide image; H&E: hematoxilina-eosina; MIL: aprendizaje multi-instancia; CNN: red neuronal convolucional; DL: aprendizaje profundo; ROI: región de interés; TIL: linfocitos infiltrantes de tumor; MSI: inestabilidad de microsatélites; TMB: carga mutacional tumoral; GIST: tumor estromal gastrointestinal; SNC: sistema nervioso central; MPE: derrame pleural maligno; BPE: derrame pleural benigno; GCA: arteritis de células gigantes; AIH: hepatitis autoinmune; PBC: colangitis biliar primaria; SFT: tumor fibroso solitario; LGD: displasia de bajo grado; HGD: displasia de alto grado; CV: validación cruzada; SOTA: state of the art.

En el presente estudio se sintetizó la evidencia publicada entre 2025 y 2026 sobre la aplicación de inteligencia artificial al análisis de imágenes histológicas con fines diagnósticos. Los hallazgos principales indican un predominio de la patología oncológica, especialmente en tracto gastrointestinal, mama y sistema nervioso central, así como una expansión hacia tareas de mayor complejidad clínica, tales como la predicción de biomarcadores moleculares, la estratificación pronóstica y la estimación de respuesta terapéutica. Este patrón coincide con lo descrito en revisiones recientes que señalan al cáncer como el principal campo de aplicación de la patología computacional (Marra *et al.*, 2025; Takamatsu, 2025;).

En relación con la evolución metodológica, se encontró una progresión clara desde arquitecturas convolucionales clásicas hacia enfoques más sofisticados basados en aprendizaje multi-instancia (MIL), transformers y foundation models entrenados mediante aprendizaje autossupervisado. El uso de MIL fue particularmente frecuente en escenarios con anotación débil a nivel de lámina, permitiendo entrenar modelos sin segmentaciones exhaustivas (Lou *et al.*, 2025; Wang *et al.*, 2026). Los foundation models y extractores universales de características se asociaron a estudios con mayor tamaño muestral y validación multicéntrica, lo que podría interpretarse como un marcador de madurez tecnológica (Schirris *et al.*, 2025; Wang *et al.*, 2025a). Asimismo, algunos trabajos incorporaron estrategias multimodales y modelos visión-lenguaje, ampliando el paradigma hacia modelos integrativos (Peng *et al.*, 2026).

No obstante, el tamaño muestral y la validación externa mostraron una heterogeneidad considerable. Mientras ciertos estudios incluyeron miles de pacientes y múltiples cohortes independientes, otros se limitaron a series retrospectivas pequeñas sin validación externa formal. La reducción moderada del desempeño al evaluarse en cohortes externas, cuando fue reportada, refuerza la importancia de validar modelos fuera del entorno institucional de entrenamiento. La presencia de validación externa en más de la mitad de los estudios incluidos constituye un avance respecto a generaciones previas de investigaciones, pero aún no puede considerarse un estándar universal.

En cuanto a las métricas de desempeño, predominó el uso de AUROC, accuracy, F1-score e índices de concordancia. En estudios pronósticos se reportaron hazard ratios estadísticamente significativos, sugiriendo asociación independiente con desfechos clínicos (Lou *et al.*, 2025; Markowski *et al.*, 2025). Sin embargo, la variabilidad en métricas y definiciones dificulta la comparación directa entre estudios y limita la posibilidad de realizar síntesis cuantitativas robustas. La ausencia de reportes estandarizados sobre calibración, análisis de decisión clínica o impacto incremental sobre modelos convencionales constituye una limitación metodológica

relevante, tal como lo han señalado revisiones previas (Treillard *et al.*, 2026).

Un hallazgo destacado es el desplazamiento hacia tareas de mayor complejidad clínica más allá de la clasificación diagnóstica binaria tradicional. Por ejemplo, varios estudios demostraron la capacidad de la IA para predecir biomarcadores moleculares como MSI y TMB directamente a partir de WSI de rutina (Gustav *et al.*, 2025; Zhang *et al.*, 2025b), lo que sugiere que la IA puede explorar correlaciones morfo-moleculares con potencial valor añadido. Del mismo modo, la cuantificación automatizada de linfocitos infiltrantes de tumor (TIL) mediante foundation models mostró una fuerte asociación pronóstica en cáncer de mama (Schirris et al., 2025), ofreciendo una alternativa reproducible y estandarizada a la evaluación visual subjetiva.

En el ámbito del diagnóstico diferencial, modelos basados en transformers y atención-MIL alcanzaron desempeños muy elevados en condiciones clínicamente desafiantes, como la distinción entre hepatitis autoinmune y colangitis biliar primaria (AUC 0,81) (Gerussi *et al.*, 2025) o el diagnóstico de arteritis de células gigantes en biopsias de arteria temporal (AUROC 0,994) (Bourgade *et al.*, 2025). Estos resultados sugieren que la IA podría actuar como una herramienta de apoyo de segundo nivel en patologías de baja prevalencia o diagnóstico morfológico difícil.

Un aspecto crítico es la distancia existente entre el desempeño experimental y la implementación clínica real. Solo una minoría de estudios incluyó validaciones prospectivas o integración directa en flujos clínicos. La mayoría continúa en fase de evaluación retrospectiva, lo que sugiere que, pese al alto rendimiento reportado, la adopción rutinaria aún requiere estudios de impacto clínico, análisis de costo-efectividad y evaluación regulatoria formal. La creciente complejidad de transformers y foundation models resalta la necesidad de contar con estrategias interpretables, como lo subrayan revisiones centradas en características biológicas explicables (Treillard *et al.*, 2026). Además, propuestas como el aprendizaje federado emergen como alternativas para mitigar limitaciones regulatorias y de privacidad (Touhami *et al.*, 2026), aunque su adopción en patología digital todavía es incipiente.

La heterogeneidad en diseño, órganos estudiados, arquitecturas y métricas impidió la realización de un metaanálisis cuantitativo, por lo que la síntesis se orientó a un análisis cualitativo estructurado, coherente con el objetivo de identificar tendencias metodológicas y niveles de madurez más que estimar un efecto combinado. Asimismo, no se aplicó una herramienta formal de evaluación del riesgo de sesgo, dado el carácter descriptivo y comparativo de la revisión; sin embargo, se consideraron de manera sistemática elementos clave como tamaño muestral, validación externa y tipo de modelo para contextualizar la solidez de cada estudio. Los resultados deben interpretarse dentro de este marco analítico.

A pesar de estas limitaciones, la evidencia revisada sugiere que la inteligencia artificial aplicada a imágenes histológicas se encuentra en una fase de consolidación metodológica con creciente sofisticación tecnológica y expansión hacia tareas clínicamente relevantes. Las direcciones futuras deberían incluir: (1) la estandarización de métricas de desempeño y protocolos de validación externa; (2) el desarrollo de modelos más interpretables y transparentes; (3) la realización de estudios prospectivos que evalúen el impacto clínico real; (4) la inclusión de poblaciones diversas para garantizar la generalización; y (5) la integración de enfoques multimodales que combinen datos histológicos, genómicos y clínicos.

CONCLUSIONES

Se encontró un predominio de la patología oncológica, especialmente en tracto gastrointestinal, mama y sistema nervioso central, así como una expansión hacia tareas de mayor complejidad clínica, incluyendo predicción de biomarcadores moleculares, estratificación pronóstica y respuesta terapéutica. Metodológicamente, se observó una transición desde redes neuronales convolucionales convencionales hacia aprendizaje multi-instancia, transformers y foundation models, particularmente en estudios con mayor tamaño muestral y validación multicéntrica. Aunque más de la mitad de los estudios reportó validación externa, persiste heterogeneidad en métricas, tamaños muestrales y estrategias de evaluación, lo que indica un nivel creciente de madurez metodológica, pero aún con limitaciones para su implementación clínica generalizada.

Referencias bibliográficas

Azam, M. T., Balaha, H. M., Mistry, A., Ali, K. M., Mobley, B. C., Leelatian, N., ... & (2025). A novel rotation and scale-invariant deep learning framework leveraging conical transformers for precise differentiation between meningioma and solitary fibrous tumor. *Journal of Pathology Informatics*, *17*, Article 100422. <https://doi.org/10.1016/j.jpi.2025.100422>

Bharadwaj, A. P., Shivanna, D. B., Rao, R. S., & Astekar, M. (2025). Deep learning-based differential diagnosis of odontogenic keratocyst and dentigerous cyst in haematoxylin and eosin-stained whole slide images. *Digital Dentistry*

Journal, *2*(2), Article 100028. <https://doi.org/10.1016/j.ddj.2025.100028>

Bourgade, R., Elhannani, M., Loussouarn, D., Guédon, A., Péteri, R., Allix-Béguet, C., ... & (2025). Deep learning for giant cell arteritis diagnosis on temporal artery biopsy. *Computers in Biology and Medicine*, *196*, Article 110707. <https://doi.org/10.1016/j.combiomed.2025.110707>

Demir, D., Ozyoruk, K. B., Durusoy, Y., Cinar, E., Serin, G., Basak, K., ... & (2025). The future of surgical diagnostics: Artificial intelligence-enhanced detection of ganglion cells for Hirschsprung disease. *Laboratory Investigation*, *105*(2), Article 102189. <https://doi.org/10.1016/j.labinv.2024.102189>

Doeleman, T., Brussee, S., Hondelink, L. M., Westerbeek, D. W. F., Sequeira, A. M., Valkema, P. A., ... & (2025). Deep learning-based classification of early-stage mycosis fungoides and benign inflammatory dermatoses on H&E-stained whole-slide images: A retrospective, proof-of-concept study. *Journal of Investigative Dermatology*, *145*(5), 1127-1134. e8. <https://doi.org/10.1016/j.jid.2024.07.036>

Gerussi, A., Saldanha, O. L., Cazzaniga, G., Verda, D., Carrero, Z. I., Engel, B., ... & (2025). Deep learning helps discriminate between autoimmune hepatitis and primary biliary cholangitis. *JHEP Reports*, *7*(2), Article 101198. <https://doi.org/10.1016/j.jhepr.2024.101198>

Giraldo-Roldán, D., Nakamura, T. C. R., Claret, A. F., Dos Santos, G. C., Pulido-Díaz, K., Gerber-Mora, R., ... & (2026). Impact of transfer learning on convolutional neural networks for odontogenic tumor diagnosis. *Head and Neck Pathology*, *20*(1), Article 24. <https://doi.org/10.1007/s12105-025-01875-y>

Grignaffini, F., Barbuto, F., Troiano, M., Piazza, L., Simeoni, P., Mangini, F., ... & (2024). The use of artificial intelligence in the liver histopathology field: A systematic review. *Diagnostics*, *14*(4), Article 388. <https://doi.org/10.3390/diagnostics14040388>

Gustav, M., van Treeck, M., Reitsam, N. G., Carrero, Z. I., Loeffler, C. M. L., Meneghetti, A. R. (2025). Assessing genotype-phenotype correlations in colorectal cancer with deep learning: a multicentre cohort study. *The Lancet Digital Health*, *7*(8). <https://doi.org/10.1016/j.landig.2025.100891>

Kussaibi, H., & Alsafwani, N. (2023). Trends in AI-powered classification of thyroid neoplasms based on histopathology images- a systematic review. *Acta Informatica Medica*, *31*(4), 280-286. <https://doi.org/10.5455/aim.2023.31.280-286>

Lalchungnunga, H. (2025). Neuropath-AI: Deep learning for molecular inference and classification of 52 central nervous system tumor types from whole-slide images. *Nature Medicine*, *31*, 1120-1130.

Lee, H., Lee, Y. J., Kim, E., Lee, J., Ahn, S., & Lee, S. H. (2026). Subtype classification of gastric spindle cell tumors in whole slide images. *Computers in Biology and Medicine*, *201*, Article 111410. <https://doi.org/10.1016/j.combiomed.2025.111410>

Li, W., Ye, S., Jin, Z., Chen, L., Chao, Y., Wei, G. (2025). Artificial intelligence in digital pathology of breast cancer, new era of practice? *International Journal of Surgery*, *111*(11), 8270-8283. <https://doi.org/10.1097/JS9.0000000000002953>

Lou, S., Huang, Y., Du, F., Xue, J., Mo, G., Li, H. (2025). Development and validation of a deep learning-based pathomics signature for prognosis and chemotherapy benefits in colorectal cancer: a retrospective multicenter cohort study. *Frontiers in Immunology*, *16*, Article 1602909. <https://doi.org/10.3389/fimmu.2025.1602909>

Markowski, M. C., Ren, Y., Tierney, M., Royce, T. J., Yamashita, R., Croucher, D., ... & (2025). Digital pathology-based artificial intelligence biomarker validation in metastatic prostate cancer. *European Urology Oncology*, *8*(3), 755-762. <https://doi.org/10.1016/j.euo.2024.11.009>

Marra, A., Morganti, S., Pareja, F., Campanella, G., Bibeau, F., Fuchs, T., ... & (2025). Artificial intelligence entering the pathology arena in oncology: current applications and future perspectives. *Annals of Oncology*, *36*(7), 712-725. <https://doi.org/10.1016/j.annonc.2025.03.006>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, 10(1), 89. <https://doi.org/10.1186/s13643-021-01626-4>

Peng, X., Tan, H., Xiao, B., Tan, Y., Yue, X., Cao, Y., ... & (2026). Deep learning for prognostic stratification and biomarker exploration in upper tract urothelial carcinoma: a multicenter retrospective cohort study. *International Journal of Surgery*, *112*(1), Article 1402. <https://doi.org/10.1097/JS9.0000000000003581>

Schirris, Y., Voorthuis, R., Opdam, M., Liefwaard, M., Sonke, G. S., Dackus, G., ... & (2025). Label-efficient computational tumour infiltrating lymphocyte assessment in breast cancer (ECTIL): multicentre validation in 2340 patients with breast cancer. *The Lancet Digital Health*, *7*(11). <https://doi.org/10.1016/j.landig.2025.100921>

Steimetz, E., Simsek, Z. C., Saha, A., Xia, R., & Gupta, R. (2025). Deep learning model for detecting high-grade dysplasia in colorectal adenomas. *Journal of Pathology Informatics*, *17*, Article 100441. <https://doi.org/10.1016/j.jpi.2025.100441>

Takamatsu, M. (2025). Transforming histologic assessment: artificial intelligence in cancer diagnosis and personalized treatment. *British Journal of Cancer*, *133*(12), 1765-1775. <https://doi.org/10.1038/s41416-025-03206-y>

Tan, W. K., Ross-Innes, C. S., Somerset, T., Markert, G., Markowitz, F., O'Donovan, M., ... & (2025). Biomarker risk stratification with capsule sponge in the surveillance of Barrett's oesophagus: prospective evaluation of UK real-world implementation. *The Lancet*, *406*(10500), 271-282. [https://doi.org/10.1016/S0140-6736\(25\)01021-9](https://doi.org/10.1016/S0140-6736(25)01021-9)

Touhami, M., Ahmad Fauzi, M. F., Ur Rehman, Z., & Mansor, S. (2026). Federated learning for histopathology image classification: A systematic review. *Diagnostics*, *16*(1), Article 137. <https://doi.org/10.3390/diagnostics16010137>

Treillard, S., Schwob, R., Mouysset, S., Brousset, P., Cussat-Blanc, S., & Franchet, C. (2026). Biological feature-based machine learning in histopathological images: a systematic review. *Journal of Pathology Informatics*, *20*, Article 100539. <https://doi.org/10.1016/j.jpi.2025.100539>

Wang, C. W., Muzakky, H., Chung, Y. P., Lai, P. J., & Chao, T. K. (2025b). Interpretable multi-scale deep learning to detect malignancy in cell blocks and cytological smears of pleural effusion and identify aggressive endometrial cancer. *Medical Image Analysis*, *106*, Article 103742. <https://doi.org/10.1016/j.media.2025.103742>

Wang, X., Jiang, Y., Yang, S., Wang, F., Zhang, X., Wang, W. (2025a). Foundation model for predicting prognosis and adjuvant therapy benefit from digital pathology in GI cancers. *Journal of Clinical Oncology*, *43*(32), 3468-3481. <https://doi.org/10.1200/JCO-24-01501>

Wang, X., Shen, X., Yang, M., Yang, L., He, Y., Ren, Y. (2026). An interpretable model based on weakly supervised learning for uterine smooth muscle tumor diagnosis: A multi-center study. *Pathology – Research and Practice*, *278*, Article 156337. <https://doi.org/10.1016/j.prp.2025.156337>

Zhang, X., Chen, Y., Cai, C., Wang, Y., Tan, J., Fang, Z. (2025a). Artificial intelligence predicts multiclass molecular signatures and subtypes directly from breast cancer histology: a multicenter retrospective study. *International Journal of Surgery*, *111*(4), Article 3109. <https://doi.org/10.1097/JS9.0000000000002220>

Zhang, Y., Han, J., Chen, H., Hu, F., Huang, Y., Tian, G. (2025b). Deep learning-based fusion of nuclear segmentation features for microsatellite instability and tumor mutational burden prediction in digestive tract cancers: a multicenter validation study. *Briefings in Bioinformatics*, *26*(6), Article bbaf580. <https://doi.org/10.1093/bib/bbaf580>

Zhou, Y., Shu, X., Wang, F., Xu, H., Tang, H. Q., Fang, H. (2025). Prediction of neoadjuvant therapy response in breast cancer based on interpretable artificial intelligence. *International Journal of Surgery*, *112*(1), 1066-1080. <https://doi.org/10.1097/JS9.0000000000003326>

Declaración de conflicto de intereses: El autor no presenta ningún conflicto de interés.

Declaración de contribución de los autores/as utilizando la Taxonomía CRediT:

El investigador trabajó en la totalidad de investigación, redacción y metodología del artículo.

Declaración de aprobación por el Comité de Ética: El autor declara que la investigación fue aprobada por el Comité de Ética de la institución responsable, en tanto la misma implicó a seres humanos.

Declaración de originalidad del manuscrito: El autor confirma que este texto no ha sido publicado con anterioridad, ni ha sido enviado a otra revista para su publicación.