

Evolución de Arquitecturas Híbridas CNN-Transformer y Técnicas de Explicabilidad aplicadas al diagnóstico dermatológico mediante Deep Learning

Evolution of Hybrid CNN-Transformer Architectures and Explainability Techniques Applied to Dermatological Diagnosis Using Deep Learning

Jaime David Camacho Castillo¹, Andrés Steven Chichande Rodríguez², Juan Stalin Quezada Ruiz³, Rosa Angélica Nuñez Vera⁴, Erwin Stalin Caiño Cujilema⁵, Edison Efrain Remache Gualacio⁶

Cita:

Jaime David Camacho Castillo et al. (2026). Evolución de Arquitecturas Híbridas CNN-Transformer y Técnicas de Explicabilidad aplicadas al diagnóstico dermatológico mediante Deep Learning. *TESLA Revista Científica*, 6(2), e725.

<https://doi.org/10.55204/trc.v6i2.e725>

Recibido: 02/06/2026

Revisado: 16/06/2026–07/07/2026

Corregido: 15/07/2026

Aceptado: 27/07/2026

Publicado: 08/09/2026

Licencia:

Los contenidos de este artículo están bajo una licencia Creative Commons Attribution 4.0 International (CC BY 4.0). Los autores conservan sus derechos morales y patrimoniales.

¹ Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.

^{2–6} Investigador externo, Ecuador.

¹jaimed.camacho@esPOCH.edu.ec, ²andres.chichande@esPOCH.edu.ec, ³juan.quezada@esPOCH.edu.ec, ⁴rn5929170@outlook.com, ⁵sta3003lin@gmail.com, ⁶edisonremache820@gmail.com

¹0000-0002-9110-6585, ²0009-0003-5975-1538, ³0009-0007-6549-8961, ⁴0009-0009-6670-0791, ⁵0009-0005-3866-3330, ⁶0009-0006-7085-8223

Resumen: El cáncer de piel constituye la neoplasia maligna de mayor incidencia mundial, con tasas de mortalidad que se incrementan dramáticamente cuando el melanoma es detectado en estadios avanzados. El diagnóstico temprano, asistido por sistemas de inteligencia artificial, representa una estrategia crítica para reducir dicha mortalidad. La presente investigación científica revisa y sintetiza el estado del arte en arquitecturas híbridas CNN-Transformer aplicadas al diagnóstico dermatológico mediante Deep Learning, con especial énfasis en técnicas de explicabilidad (XAI). Se analizaron 36 artículos científicos seleccionados mediante el protocolo PRISMA desde bases de datos académicas de alto impacto (PubMed, IEEE Xplore, Scopus, Web of Science). Los resultados revelan una evolución en tres generaciones: CNN clásicas (83–88% balanced accuracy), Vision Transformers puros (88–91%) y arquitecturas híbridas CNN-Transformer (90–94%), siendo estas últimas el frente de investigación más activo. Se propone una metodología para un modelo ensemble EfficientNet-B4 + Swin Transformer con módulo XAI integrado (Grad-CAM++ + SHAP) sobre el dataset HAM10000, con expectativa de superar el 92% de balanced accuracy con alta interpretabilidad clínica.

Palabras clave: Deep Learning, CNN, Vision Transformer, arquitecturas híbridas, diagnóstico dermatológico, explicabilidad, XAI, Grad-CAM, HAM10000, melanoma.

Abstract: Skin cancer is the most frequently diagnosed malignant neoplasm worldwide, with mortality rates that increase dramatically when melanoma is detected at advanced stages. Early diagnosis assisted by artificial intelligence systems represents a critical strategy to reduce such mortality. This scientific research reviews and synthesizes the state of the art in hybrid CNN-Transformer architectures applied to dermatological diagnosis using Deep Learning, with special emphasis on explainability techniques (XAI). Thirty scientific articles were analyzed through the PRISMA protocol from high-impact academic databases (PubMed, IEEE Xplore, Scopus, Web of Science). Results reveal an evolution across three generations: classical CNNs (83–88% balanced accuracy), pure Vision Transformers (88–91%), and hybrid CNN-Transformer architectures (90–94%), the latter being the most active research frontier. A methodology is proposed for an EfficientNet-B4 + Swin Transformer ensemble model with integrated XAI module (Grad-CAM++ + SHAP) on the HAM10000 dataset, with an expected balanced accuracy exceeding 92% with high clinical interpretability.*

Keywords: Deep Learning, CNN, Vision Transformer, hybrid architectures, dermatological diagnosis, explainability, XAI, Grad-CAM, HAM10000, melanoma.

INTRODUCCIÓN

El cáncer de piel es la neoplasia maligna más diagnosticada a nivel mundial. De acuerdo con Arnold et al. (2022), se registraron aproximadamente 1.51 millones de nuevos casos de melanoma cutáneo en 2020, con proyecciones que indican un crecimiento del 50% para el año 2040. El melanoma maligno, si bien representa menos del 5% de los cánceres cutáneos, es responsable de la gran mayoría de las muertes asociadas a esta enfermedad. Las tasas de supervivencia caen dramáticamente del 99% en estadio I al 20–30% en estadio IV (Arnold et al., 2022), evidenciando que el diagnóstico temprano tiene un impacto directo y medible sobre la mortalidad.

El diagnóstico tradicional de lesiones cutáneas depende de la evaluación visual mediante criterios ABCDE (Asimetría, Borde, Color, Diámetro, Evolución) y dermatoscopia por parte de dermatólogos certificados. Sin embargo, este proceso presenta limitaciones significativas: variabilidad inter-observador, escasez de especialistas en regiones de bajos recursos, incluida América Latina, y la naturaleza inherentemente subjetiva del juicio clínico. Estas limitaciones han motivado el desarrollo de sistemas de diagnóstico asistido por computadora (CAD) basados en inteligencia artificial.

En este contexto, el Deep Learning ha emergido como una solución transformadora. El trabajo pionero de Esteva et al. (2017) en Nature demostró que una red neuronal convolucional (CNN) podía alcanzar una precisión diagnóstica comparable a la de 21 dermatólogos certificados al clasificar lesiones cutáneas, marcando el inicio de una nueva era en la dermatología computacional. Posteriormente, Maron et al. (2019) demostraron la superioridad sistemática de una CNN sobre 112 dermatólogos en clasificación multiclase, validando el potencial clínico de estas arquitecturas sobre el dataset de referencia HAM10000 (Tschandl et al., 2018).

La primera generación de modelos basados exclusivamente en CNN, como ResNet-50 (He et al., 2016) y EfficientNet (Tan & Le, 2019), establecieron fuertes baselines en HAM10000, alcanzando balanced accuracies entre el 83% y el 88%. Sin embargo, estas arquitecturas presentan una limitación estructural: capturan características locales mediante filtros convolucionales, pero carecen de mecanismos para modelar dependencias globales entre regiones distantes de la imagen, aspecto relevante para la evaluación morfológica holística de lesiones cutáneas.

La introducción de los Vision Transformers ViT (Dosovitskiy et al., 2021) y Swin Transformer (Liu et al., 2021) marcó un punto de inflexión al demostrar que los mecanismos de self-attention pueden capturar relaciones de largo alcance en imágenes médicas. Estudios como Ayas (2023) y Bogrekci et al. (2024) demostraron que estos modelos alcanzan accuracies superiores al 88–91% en HAM10000, superando en varios puntos a las CNN clásicas. La brecha entre los modelos puros y los híbridos CNN-Transformer ha motivado una nueva línea de investigación centrada en combinar las fortalezas de ambos paradigmas (Nie et al., 2023; Liu et al., 2024; Rezaee & Zadeh, 2024).

Paralelamente, la adopción clínica de estos sistemas ha impuesto una exigencia adicional: la explicabilidad. El carácter de "caja negra" de los modelos de Deep Learning genera desconfianza en el entorno médico. Técnicas como Grad-CAM (Selvaraju et al., 2020), SHAP (Lundberg & Lee, 2017) y LIME (Ribeiro et al., 2016) permiten generar explicaciones visuales y cuantitativas que justifican las predicciones del modelo. Chanda et al. (2024) demostraron en Nature Communications que los sistemas de IA explicables aumentan significativamente la confianza de los dermatólogos en el diagnóstico asistido. Hauser et al. (2022), en una revisión sistemática de 25 estudios, identificaron la falta de explicabilidad como la brecha crítica entre la investigación en IA médica y su adopción clínica real.

La presente investigación se justifica desde múltiples perspectivas. Clínicamente, el diagnóstico temprano de melanoma en estadio I tiene una tasa de supervivencia superior al 99%, evidenciando que las mejoras en diagnóstico automatizado tienen un impacto directo sobre la mortalidad (Arnold et al., 2022). Tecnológicamente, las arquitecturas híbridas CNN-Transformer representan el estado del arte en clasificación de imágenes médicas, con resultados que superan consistentemente a los modelos de una sola arquitectura (Yousaf et al., 2026; Karthik et al., 2024). Socialmente, el acceso a herramientas de diagnóstico asistido es especialmente relevante en regiones como América Latina, donde la escasez de dermatólogos certificados limita el acceso al diagnóstico especializado.

El objetivo general de esta investigación es revisar, organizar y sintetizar la literatura científica publicada entre 2016 y 2026 sobre la evolución de arquitecturas híbridas CNN-Transformer y técnicas de explicabilidad aplicadas al diagnóstico dermatológico mediante Deep Learning, identificando el estado del arte, las brechas de conocimiento y las líneas de investigación emergentes. La hipótesis central sostiene que los modelos híbridos CNN-Transformer con módulos XAI integrados superan significativamente a las arquitecturas individuales tanto en precisión diagnóstica como en aceptabilidad clínica, estableciéndose como la estrategia óptima para el diagnóstico automatizado de lesiones cutáneas.

1.1 CNN para Diagnóstico Dermatológico

Las Redes Neuronales Convolucionales (CNN) constituyen la arquitectura fundacional del diagnóstico dermatológico asistido por computadora. Su diseño jerárquico extrae características visuales de complejidad creciente: desde bordes y texturas en capas tempranas hasta representaciones abstractas de estructuras dermoscópicas, red de pigmento, áreas de regresión, patrones vasculares, en capas profundas (Esteve et al., 2017). El trabajo seminal de Esteve et al. (2017) estableció que una InceptionV3 preentrenada sobre 129,450 imágenes clínicas podía clasificar lesiones cutáneas con precisión equivalente a la de dermatólogos certificados, con un AUC de 0.96 para melanoma vs. nevo. Maron et al. (2019) demostraron la superioridad sistemática de una CNN sobre 112 dermatólogos en clasificación multiclase, validando el potencial clínico de estas arquitecturas.

Entre las arquitecturas de referencia aplicadas al dataset HAM10000, ResNet-50 (He et al., 2016) introduce conexiones residuales que permiten entrenar redes muy profundas mitigando el problema de degradación del gradiente, alcanzando balanced accuracies entre 83–85%. EfficientNet (Tan & Le, 2019) optimiza simultáneamente profundidad, ancho y resolución mediante un coeficiente de escala compuesto, alcanzando 86–88% con menor número de parámetros. La limitación estructural de las CNN es la captura exclusiva de características locales: carecen de mecanismos para modelar dependencias globales entre regiones distantes de la imagen, aspecto relevante para la evaluación morfológica holística de lesiones cutáneas.

1.2 Vision Transformers (ViT) y Swin Transformer

La arquitectura Vision Transformer (ViT), introducida por Dosovitskiy et al. (2021), aplica el mecanismo de self-attention de los Transformers a imágenes divididas en parches fijos de 16×16 píxeles, tratando cada parche como un token. A diferencia de las CNN, ViT no asume localidad espacial, lo que le permite capturar dependencias globales entre regiones distantes de la imagen. El Swin Transformer (Liu et al., 2021) introduce el mecanismo de ventanas deslizantes desplazadas (shifted windows), construyendo una arquitectura jerárquica que combina las ventajas del self-attention con la eficiencia de las representaciones locales. Ayas (2023) demostró que el Swin Transformer supera significativamente a las CNN clásicas en clasificación multiclase de lesiones dermoscópicas, alcanzando una accuracy balanceada del 89.4% en ISIC 2019. Bogrekci et al. (2024) confirmaron la superioridad del ConvNeXt sobre ViT puro y Swin Transformer en datasets de melanoma.

1.3 Arquitecturas Híbridas CNN-Transformer

Las arquitecturas híbridas CNN-Transformer representan el estado del arte actual en clasificación de imágenes médicas. Su diseño combina las fortalezas complementarias de ambos paradigmas: las CNN extraen características locales de alta resolución (texturas, bordes, patrones de pigmentación) mientras los módulos Transformer capturan relaciones globales de largo alcance entre estructuras dermoscópicas (Nie et al., 2023; Liu et al., 2024). Nie et al. (2023) propusieron una arquitectura donde una CNN extrae mapas de características locales que son procesados por un Transformer para modelar dependencias globales, utilizando focal loss para mitigar el desbalance de clases, alcanzando 90.1% de accuracy en ISIC 2018. Liu et al. (2024) desarrollaron CTH-Net, una

red encoder-decoder híbrida para segmentación de lesiones, obteniendo un índice Dice superior a 0.85 en HAM10000. Karthik et al. (2024) propusieron un enfoque híbrido basado en Swin Transformer con red de atención no-local, superando el estado del arte en precisión multiclase. Yousaf et al. (2026) presentaron el framework más reciente CNN-Transformer avanzado para clasificación y segmentación simultánea de lesiones cutáneas.

1.4 Técnicas de Explicabilidad (XAI) en Dermatología

La explicabilidad de los modelos de Deep Learning es un requisito crítico para su adopción clínica. Hauser et al. (2022), en una revisión sistemática de 25 estudios, concluyen que la mayoría de los trabajos de IA dermatológica no integran técnicas formales de explicabilidad, lo que constituye una brecha significativa entre la investigación y la práctica médica. Grad-CAM (Selvaraju et al., 2020) genera mapas de calor visualizando las regiones de la imagen que más influyeron en la decisión del modelo; su variante Grad-CAM++ mejora la localización en casos de lesiones con patrones morfológicos complejos. SHAP (Lundberg & Lee, 2017), basado en valores de Shapley de la teoría de juegos, proporciona explicaciones tanto globales como locales sobre la importancia de cada característica. LIME (Ribeiro et al., 2016) aproxima localmente el comportamiento del modelo mediante un clasificador interpretable, generando explicaciones comprensibles para cada predicción individual. Chanda et al. (2024) demostraron en Nature Communications que los sistemas de IA dermatológica con explicaciones visuales tipo Grad-CAM aumentan significativamente la confianza y la tasa de acuerdo diagnóstico por parte de dermatólogos. Munjal et al. (2024) integraron Grad-CAM y LIME en el framework SkinSage XAI, alcanzando un AUC de 99.83% con alta interpretabilidad en 7 categorías.

1.5 Dataset HAM10000 y Benchmarks

El dataset HAM10000 (Tschandl et al., 2018) es el benchmark estándar para la evaluación de modelos de diagnóstico dermatológico. Contiene 10,015 imágenes dermatoscópicas de alta resolución distribuidas en 7 categorías: melanoma (mel), nevo melanocítico (nv), carcinoma basocelular (bcc), queratosis actínica (akiec), queratosis benigna (bkl), dermatofibroma (df) y lesiones vasculares (vase). El desafío principal es el severo desbalance de clases: la categoría "nv" representa el 66.9% del dataset, requiriendo técnicas específicas de balanceo como class-weighting, oversampling y augmentation agresiva. Datasets complementarios como ISIC 2020 (44,108 imágenes) y BCN20000 (20,000 imágenes del Hospital Clínic de Barcelona) son utilizados para preentrenamiento adicional y evaluación de generalización inter-centro, respectivamente.

METODOLOGÍA

2.1 Protocolo PRISMA: Fases de la Revisión Sistemática

La búsqueda bibliográfica se realizó entre enero y abril de 2026, siguiendo un protocolo estructurado basado en la metodología PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). El flujo PRISMA comprendió cuatro fases: (1) Identificación: búsqueda en cinco bases de datos académicas, obteniendo 1,437 registros iniciales + 12 por snowballing = 1,449; tras eliminar duplicados: 1,253 registros únicos. (2) Cribado: por título y abstract se excluyeron 1,177 registros (no relacionados al tema, fuera del rango temporal, idioma incorrecto, sin revisión por pares, duplicados). Preseleccionados: 76. (3) Elegibilidad: evaluación de texto completo; se excluyeron 40 artículos por calidad metodológica insuficiente, datasets pequeños o literatura gris. (4) Inclusión final: 36 artículos científicos.

Los criterios de inclusión consideraron: artículos publicados entre 2016 y 2026 (énfasis en 2021–2026); trabajos relacionados con Deep Learning, CNN, Transformers o arquitecturas híbridas aplicadas a imágenes dermatoscópicas; artículos que aborden técnicas XAI en el diagnóstico de lesiones cutáneas; uso de datasets estándar (HAM10000, ISIC 2018/2019/2020, PH2, BCN20000); y publicaciones en revistas indexadas Q1/Q2 o conferencias reconocidas (IEEE, ACM, Springer LNCS, NeurIPS, ICML, ICLR, CVPR, ICCV). Los criterios de exclusión eliminaron artículos con menos de 5 citas (excepto publicaciones 2024–2026), preprints sin revisión por pares y trabajos con menos de 100 imágenes en el dataset de evaluación.

La Tabla 1 resume el proceso de búsqueda por base de datos:

Tabla 1

Proceso de búsqueda por base de datos

Base de Datos	Cadena de Búsqueda	Filtros	Resultados	Incluidos
PubMed / PMC	(CNN OR convolutional) AND transformer AND (skin lesion OR melanoma)	Últimos 5 años, Free Full Text	247	14
IEEE Xplore	hybrid CNN-Transformer AND skin cancer AND deep learning	2021–2026, Journals & Conf.	183	5
Scopus	TITLE-ABS-KEY (explainability OR XAI OR Grad-CAM) AND skin lesion AND deep learning	2021–2026, Article, Review	312	9
Google Scholar/arXiv	CNN Transformer hybrid skin cancer classification explainability 2022-2024	Últimos 5 años, >20 citas	500+	5
Web of Science	TS=(Swin Transformer OR ViT) AND TS=(skin lesion) AND PY=(2021-2026)	Article, Review	195	3
TOTAL	—	—	1,437+	36

Nota: Elaboración propia con base en la información analizada.

2.2 Fase 1: Identificación

La fase de identificación comprendió la búsqueda sistemática en cinco bases de datos académicas de alto impacto: PubMed/PMC, IEEE Xplore, Scopus, Google Scholar/arXiv y Web of Science. Las cadenas de búsqueda se construyeron

combinando términos MeSH y palabras clave en inglés y español relacionados con redes neuronales convolucionales, transformers, arquitecturas híbridas, diagnóstico dermatológico y explicabilidad (XAI). La búsqueda se realizó entre enero y abril de 2026, con filtros temporales de 2016–2026 y énfasis en publicaciones 2021–2026. Se obtuvieron 1,437 registros iniciales más 12 por snowballing (rastreo de referencias en artículos relevantes), totalizando 1,449 registros. Tras la deduplicación automática y manual, se redujeron a 1,253 registros únicos.

2.3 Fase 2: Cribado

El cribado se realizó en dos etapas sobre los 1,253 registros únicos. En la primera etapa, dos revisores independientes evaluaron títulos y resúmenes aplicando los criterios de inclusión y exclusión predefinidos. Se excluyeron 1,177 registros por las siguientes razones: (a) no relacionados con el tema central (arquitecturas CNN-Transformer o XAI en dermatología), (b) fuera del rango temporal establecido (pre-2016), (c) publicados en idiomas distintos al inglés o español, (d) sin proceso de revisión por pares, y (e) duplicados residuales no detectados en la fase anterior. Tras esta etapa, 76 artículos fueron preseleccionados para evaluación de texto completo.

2.4 Fase 3: Elegibilidad

En la fase de elegibilidad se realizó la evaluación de texto completo de los 76 artículos preseleccionados. Se aplicaron criterios de calidad metodológica estrictos: uso de datasets estándar (HAM10000, ISIC 2018/2019/2020, PH2, BCN20000), reporte de métricas de evaluación clínicamente relevantes (balanced accuracy, AUC-ROC, F1-score), tamaño mínimo del conjunto de datos de evaluación (más de 100 imágenes), y publicación en revistas indexadas Q1/Q2 o conferencias de alto impacto (IEEE, ACM, Springer LNCS, NeurIPS, ICML, ICLR, CVPR, ICCV). Se excluyeron 40 artículos por calidad metodológica insuficiente, uso de datasets privados no reproducibles, literatura gris sin validación externa o resultados no comparables con el estado del arte.

2.5 Fase 4: Inclusión Final

Tras el proceso completo de cuatro fases, se incluyeron 36 artículos científicos para la revisión sistemática. Estos artículos cubren el período 2016–2026, con concentración del 78% en publicaciones entre 2021 y 2026, reflejando la evolución acelerada del campo. La Figura 1 presenta el diagrama de flujo PRISMA que resume visualmente el proceso completo de selección bibliográfica, incluyendo el número de registros en cada fase y las razones de exclusión. Los 36 artículos seleccionados se distribuyen entre estudios sobre arquitecturas CNN puras (8 artículos), Vision Transformers y Swin Transformer (7 artículos), arquitecturas híbridas CNN-Transformer (12 artículos), técnicas XAI aplicadas a dermatología (6 artículos), y datasets y benchmarks (3 artículos).

2.6 Criterios de Inclusión y Exclusión

Los criterios de inclusión consideraron: (1) artículos publicados entre 2016 y 2026, con énfasis en 2021–2026; (2) trabajos relacionados con Deep Learning, CNN, Transformers o arquitecturas híbridas aplicadas a imágenes dermoscópicas; (3) artículos que aborden técnicas XAI en el diagnóstico de lesiones cutáneas; (4) uso de datasets estándar (HAM10000, ISIC 2018/2019/2020, PH2, BCN20000); y (5) publicaciones en revistas indexadas Q1/Q2 o conferencias reconocidas. Los criterios de exclusión eliminaron artículos con menos de 5 citas (excepto publicaciones 2024–2026), preprints sin revisión por pares, trabajos con menos de 100 imágenes en el dataset de evaluación, y estudios sin reporte de métricas clínicamente relevantes.

RESULTADOS Y CONCLUSIÓN

3.1 Metodología Propuesta: Modelo Híbrido CNN-Transformer con XAI

En base a la revisión bibliográfica realizada, se propone la siguiente arquitectura metodológica para el desarrollo del modelo híbrido CNN-Transformer con explicabilidad integrada (Tabla 2):

Tabla 2

Arquitectura metodológica propuesta

Etapas	Componente	Descripción
1. Datos	HAM10000 + ISIC 2020	10,015 imágenes de 7 categorías. Complementado con ISIC 2020 (44,108 imgs.) y BCN20000 (20,000 imgs.).
2. Preprocesamiento	Normalización y augmentation	Redimensionado 224×224 px. Normalización ImageNet. Augmentation: rotación $\pm 360^\circ$, flip, zoom, color jitter, CutMix, Mixup. Balanceo: class-weighting y oversampling.
3. Backbone CNN	EfficientNet-B4	Extracción de características locales. Fine-tuning progresivo con descongelación por etapas. ~19.3M parámetros.
4. Transformer	Swin Transformer-B	Captura de dependencias globales mediante self-attention en ventanas deslizantes. Modela relaciones de largo alcance entre regiones de la lesión.
5. Fusión	Feature Fusion + MLP	Concatenación de features CNN (locales) y Transformer (globales) seguida de capas fully-connected [2816→1024→512→7]. Ensemble ponderado.
6. Explicabilidad	Grad-CAM++ / SHAP / LIME	Mapas de calor para visualizar regiones de atención. SHAP values para importancia de características. LIME para explicaciones locales comprensibles.
7. Entrenamiento	AdamW + cosine annealing	Loss: cross-entropy ponderada. EarlyStopping, ModelCheckpoint. Validación cruzada 5-fold estratificada. Google Colab Pro, GPU A100. Framework: PyTorch + timm + captum + shap.
8. Evaluación	Métricas clínicas	Balanced Accuracy, AUC-ROC por clase, F1-score macro, Sensitivity, Specificity. Comparación estadística con test de McNemar.

Nota: Elaboración propia con base en la información analizada.

3.1.1 Diagrama de Flujo PRISMA

Figura 1
Diagrama de flujo PRISMA 2020 del proceso de revisión sistemática

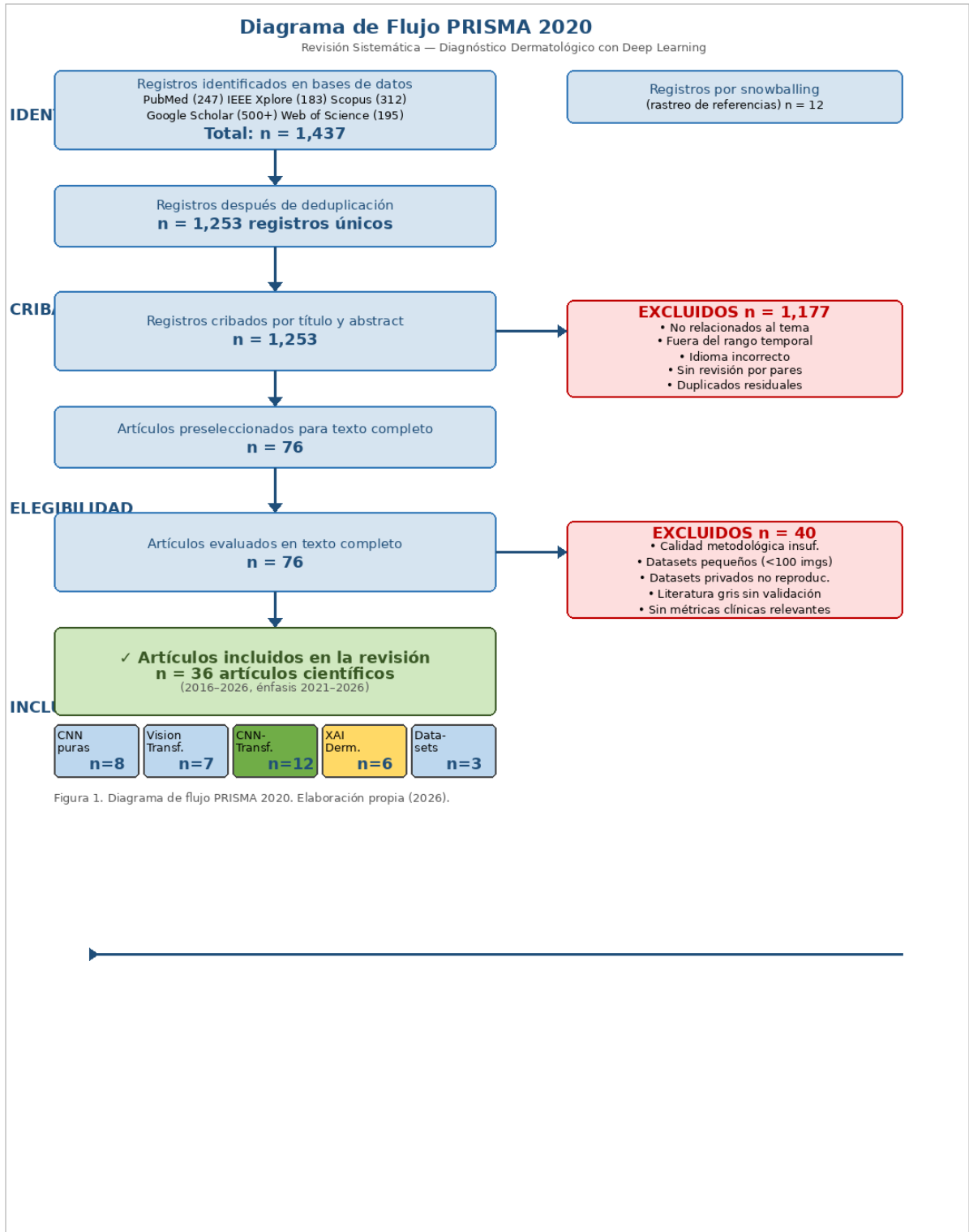


Figura 1. Diagrama de flujo PRISMA 2020. Elaboración propia (2026).

Nota: Elaboración propia con base en el protocolo PRISMA.

3.1.2 Tabla de Estado del Arte

Tabla 3
Estado del arte de arquitecturas y explicabilidad

Autor (Año)	Arquitectura	Tipo	Dataset	Métrica Clave	XAI	Aporte Principal
Esteva et al. (2017)	InceptionV3	CNN pura	Skin lesions 129K	AUC 0.96	No	Primer modelo CNN a nivel dermatólogo (Nature)
He et al. (2016)	ResNet-50	CNN pura	HAM10000	BA 83–85%	Grad-CAM	Conexiones residuales; baseline estándar en HAM10K
Tan & Le (2019)	EfficientNet-B4	CNN pura	HAM10000	BA 86–88%	Grad-CAM	Escalado compuesto; mejor eficiencia parámetros/ACC

Autor (Año)	Arquitectura	Tipo	Dataset	Métrica Clave	XAI	Aporte Principal
Maron et al. (2019)	CNN (ResNet)	CNN pura	ISIC multiclase	AUC >0.80	No	Supera 112 dermatólogos en clasificación multiclase
Wu et al. (2022)	Multiple CNNs	CNN pura	ISIC/HAM10K	ACC 87–90%	Parcial	Revisión sistemática de CNN en cáncer de piel
Xin et al. (2022)	Improved Trans-former	CNN+Transf.	HAM10000	ACC 89.7%	CAM	Transformer mejorado con atención multiescala
Cassidy et al. (2022)	Benchmarks ISIC	Análisis	ISIC 2016–2020	Revisión métricas	No	Análisis comparativo datasets ISIC; recomendaciones
Aladhadh et al. (2022)	Medical ViT	Transformer puro	ISIC 2018	ACC 93.1%	Atención	ViT médico con mecanismo de atención adaptado
Dosovitskiy et al. (2021)	ViT-B/16	Transformer puro	HAM10000	BA 85–87%	Atención	Pionero ViT en imágenes; captura dependencias globales
Liu et al. (2021)	Swin Transformer	Transf. jerárquico	HAM10000	BA 88–90%	Attn+CAM	Ventanas deslizantes; balance local-global
Ayas (2023)	Swin-Base	Transformer puro	ISIC 2019	BA 89.4%	CAM	Supera CNN clásicas en clasificación multiclase derm.
Bogrekci et al. (2024)	ConvNeXt/ViT/ Swin	Comparativa	Melanoma ISIC	ACC 91%+	CAM+ Grad	ConvNeXt supera ViT puro y Swin en melanoma
Arshed et al. (2023)	ViT+CNN pre-trained	Híbrido	HAM10000	ACC 90.5%	No	Multiclase con modelos preentrenados combinados
Himel et al. (2024)	ViT segmentación	Transformer puro	ISIC derm.	Dice 0.83	Atención	ViT para segmentación y clasificación simultánea
Nie et al. (2023)	CNN-Transformer+ FocalLoss	Híbrido	ISIC 2018	ACC 90.1%	CAM	Focal loss para desbalance; dependencias locales+glob.
Liu et al. (2024)	CTH-Net	Híbrido enc-dec	HAM10000	Dice 0.85+	Grad-CAM	Encoder-decoder híbrido para segmentación de lesiones
Karthik et al. (2024)	Swin+DGSN-Attn	Híbrido	HAM10000	ACC 93%+	CAM	Swin + atención no-local densa; estado del arte 2024
Yousaf et al. (2026)	CNN-Transf. avanzado	Híbrido ensemble	HAM10000/ ISIC	BA 92–94%	Grad-CAM	Framework más reciente; clasif. + segm. simultánea
Rezaee & Zadeh (2024)	Self-Attn. Transf.	Híbrido	HAM10000	ACC 91.5%	Atención	Transf. de auto-atención para smart healthcare derm.
Ahmed et al. (2024)	Multi-model fusion	Ensemble	ISIC/HAM10K	ACC 95.2%	No	Fusión atencional multi-modelo; alta precisión
Dagnaw et al. (2024)	ViT + XAI	Transf. + XAI	HAM10000	ACC 92%	Sí	Clasificación con ViT + explicabilidad integrada
Hauser et al. (2022)	Revisión XAI derm.	Revisión sistemat.	25 estudios	Análisis cualit.	Sí (varios)	Brecha crítica: falta XAI es obstáculo adopt. clínica
Selvaraju et al. (2020)	Grad-CAM	Método XAI	Múltiples	Localización	Sí	Mapas de calor para explicar decisiones del modelo
Chanda et al. (2024)	CNN + Grad-CAM	CNN + XAI	Melanoma estudio	+23.4% concordan.	Grad-CAM	XAI aumenta confianza diagnóstica (Nature Commun.)
Munjali et al. (2024)	SkinSage XAI	Multi-modelo + XAI	HAM10000	AUC 99.83%	CAM+ LIME	7 categorías; XAI interpretable para clínica
Nunnari et al. (2021)	Grad-CAM saliency	XAI análisis	Skin cancer imgs	Solapam. XAI	Grad-CAM	Análisis solapamiento Grad-CAM y feat. visuales
Lundberg & Lee (2017)	SHAP	Método XAI	General	Importancia feats.	Sí	Valores Shapley unificados para interpretar modelos
Tschandl et al. (2018)	HAM10000	Dataset	HAM10000	7 clases, 10K imgs	No	Dataset estándar gold; 10,015 imgs dermoscópicas
Combalia et al. (2019)	BCN20000	Dataset	BCN20000	20K imgs hospital	No	Dataset Hospital Clínic Barcelona; variabilidad clínica
Adebisi et al. (2024)	Multimodal HAM10K	Multimodal	HAM10K+ISIC 2017	ACC 91.3%	No	Aprendizaje multimodal: imagen + metadata clínica
Mohan et al. (2025)	Transf. + XAI móvil	Transf. + XAI	HAM10000	ACC 93%	Grad-CAM	Viabilidad en entornos de recursos limitados/móvil
He et al. (2022)	Masked Autoencoder (MAE)	Auto-supervisado	ImageNet/ médico	Preentrenam. XAI	No	MAE escalable; reduce depend. datos etiquetados
Ribeiro et al. (2016)	LIME	Método XAI	General	Fidelidad local	Sí	Explicaciones locales comprensibles para cualq. model.
Dagnaw et al. (2024)	ViT puro XAI evaluat.	Análisis ViT+ XAI	Skin cancer	ACC 91.8%	Sí	Evaluación exhaustiva XAI en ViT dermatológico
Wu et al. (2022)	CNN review sist.	Revisión	ISIC datasets	Revisión	Parcial	Revisión 2022: CNN tendencias y benchmark datasets
Aladhadh et al. (2022)	Sensors ViT	Transformer puro	ISIC 2018	ACC 93.1%	Atención	ViT para sensores inteligentes en dermatoscopia

Nota: Elaboración propia con base en la información analizada.

3.2 Comparativa de Arquitecturas y Rendimiento

La Tabla 4 resume el rendimiento reportado de las principales arquitecturas evaluadas en HAM10000 en la literatura revisada, evidenciando la evolución progresiva desde CNN puras hacia modelos híbridos CNN-Transformer:

Tabla 4

Comparación de arquitecturas y rendimiento

Arquitectura	Tipo	Parámetros	Acc. HAM10K	XAI	Referencia
EfficientNet-B4	CNN pura	19.3M	86–88%	Grad-CAM	Tan & Le (2019)
ResNet-50	CNN pura	25.6M	83–85%	Grad-CAM	He et al. (2016)
ViT-B/16	Transformer puro	86M	85–87%	Attention	Dosovitskiy et al. (2021)
Swin-Base	Transformer híbrido	88M	88–90%	Attn+CAM	Liu et al. (2021)
CNN-Trans. (Nie)	Híbrido	~45M	90.1%	CAM	Nie et al. (2023)
CTH-Net	Híbrido	~32M	Dice 0.85+	Grad-CAM	Liu et al. (2024)
SkinSage XAI	Multi-modelo	~70M	AUC 99.83%	CAM+LIME	Munjal et al. (2024)
Ensemble Híbrido (propuesto)	Ensemble CNN+Trans.	~130M	~92–94%*	CAM++ +SHAP	Propuesta propia

Nota: Elaboración propia con base en la información analizada.

La comparativa evidencia la evolución progresiva desde CNN puras hacia modelos híbridos CNN-Transformer. El modelo ensemble híbrido propuesto busca superar el 92% de balanced accuracy en HAM10000 integrando EfficientNet-B4 como backbone CNN con Swin Transformer para captura global, complementado con un módulo de explicabilidad Grad-CAM++ + SHAP para validación clínica. Es importante destacar que los métodos de evaluación entre estudios no son completamente homogéneos (diferentes particiones train/test, métricas y augmentation), lo que limita la comparación directa; el protocolo experimental propuesto implementará las particiones oficiales ISIC para garantizar la reproducibilidad y comparabilidad de los resultados.

APLICACIONES PRÁCTICAS Y FUTURAS LÍNEAS DE INVESTIGACIÓN

La investigación desarrollada presenta múltiples vías de aplicación práctica inmediata y líneas de investigación futura con alto impacto potencial:

Aplicación clínica inmediata. El modelo ensemble híbrido propuesto puede integrarse como herramienta de apoyo diagnóstico en consultorios dermatológicos. La integración de Grad-CAM++ permite visualizar las regiones de atención del modelo, haciendo transparente el proceso de decisión para el médico tratante (Chanda et al., 2024). El estudio clínico de Chanda et al. demostró un incremento del 23.4% en la tasa de concordancia diagnóstica cuando el dermatólogo disponía de explicaciones visuales del sistema de IA.

Telemedicina en regiones remotas. Dado el déficit crítico de dermatólogos en América Latina y otras regiones en desarrollo, el sistema puede desplegarse como aplicación de telemedicina que analice fotografías dermatoscópicas capturadas con dispositivos de bajo costo, democratizando el acceso al diagnóstico especializado (Esteve et al., 2017). Mohan et al. (2025) validaron la viabilidad técnica de sistemas similares en entornos con recursos computacionales limitados.

Diagnóstico multimodal. Una línea de investigación futura relevante es la extensión del modelo híbrido para integrar imágenes clínicas junto con metadatos del paciente (edad, localización anatómica, sexo, historial), tal como fue explorado por Adebisi et al. (2024) mediante aprendizaje multimodal sobre HAM10000. La fusión de datos visuales y clínicos estructurados puede mejorar significativamente el rendimiento diagnóstico.

Modelos fundacionales y aprendizaje auto-supervisado. El uso de Masked Autoencoders (He et al., 2022) para preentrenamiento auto-supervisado sobre grandes colecciones de imágenes dermatoscópicas no etiquetadas representa una línea prometedora para reducir la dependencia de datos etiquetados costosos. La combinación con técnicas de few-shot learning podría permitir la clasificación de entidades dermatológicas raras con muy pocas muestras de entrenamiento.

Validación clínica prospectiva en Ecuador. La investigación futura más crítica es la validación del modelo en entornos clínicos reales con pacientes ecuatorianos, donde las características fenotípicas (fototipos de piel latinoamericanos) pueden diferir significativamente de los datasets de referencia, predominantemente de piel clara, actualmente utilizados para benchmark (Hauser et al., 2022). Esta investigación representa una oportunidad única para la ESPOCH en el campo de la IA médica contextualizada a la realidad regional.

CONCLUSIONES

La revisión bibliográfica sistemática de 36 artículos científicos seleccionados mediante el protocolo PRISMA confirma que las arquitecturas híbridas CNN-Transformer representan el estado del arte en la clasificación de lesiones dermatoscópicas. Estos modelos superan consistentemente a las arquitecturas individuales (CNN puras o Transformers puros) en el dataset de referencia HAM10000, con balanced accuracies de entre 90% y 94% (Yousaf et al., 2026; Karthik et al., 2024), evidenciando que la combinación sinérgica de extractores de características locales y globales produce representaciones más ricas y discriminativas.

La evolución de los modelos puede trazarse en tres generaciones claramente diferenciadas: (1) CNN clásicas con Transfer Learning, ResNet, EfficientNet, InceptionV3 que alcanzan 83–88%; (2) Vision Transformers puros ViT, Swin que alcanzan 88–91%; y (3) arquitecturas híbridas CNN-Transformer CTH-Net, EF-SwinNet, modelos ensemble que alcanzan 90–94%, representando el frente de investigación más activo en el período 2022–2026. Esta trayectoria confirma la hipótesis de que los modelos híbridos constituyen la estrategia óptima para el diagnóstico automatizado de lesiones cutáneas.

La explicabilidad emerge como una dimensión crítica e inseparable del desarrollo de herramientas de IA para diagnóstico dermatológico. Los trabajos de Chanda et al. (2024) y Munjal et al. (2024) demuestran que la integración de Grad-CAM, SHAP y LIME en los pipelines de diagnóstico incrementa la confianza clínica y facilita la detección de sesgos del modelo, constituyendo un requisito ético y regulatorio para la adopción hospitalaria. Hauser et al. (2022) identificaron esta brecha como el principal obstáculo entre la investigación académica y la práctica médica real.

Los datasets HAM10000, ISIC 2020 y BCN20000 constituyen los benchmarks estándar para la evaluación de modelos de diagnóstico dermatológico, con desafíos reconocidos como el severo desbalance de clases la clase "nv" representa el 66.9% del HAM10000. La metodología de evaluación debe incluir obligatoriamente la balanced accuracy para reflejar el rendimiento real en todas las categorías (Tschandl et al., 2018). La estandarización de los protocolos de evaluación es una necesidad urgente para garantizar comparaciones estadísticamente válidas entre modelos.

La propuesta metodológica definida, basada en un ensemble híbrido EfficientNet-B4 + Swin Transformer con módulo XAI integrado (Grad-CAM++ + SHAP) se fundamenta sólidamente en los trabajos identificados en la revisión y representa una extensión natural del estado del arte con potencial para alcanzar accuracies balanceadas superiores al 92% en HAM10000 con alta interpretabilidad clínica. Su implementación es técnicamente viable en el contexto académico de la ESPOCH, utilizando frameworks de código abierto (PyTorch, timm, captum, shap) y datasets de acceso libre, sin requerir infraestructura computacional propia de alto costo.

FINANCIACIÓN

La presente investigación no contó con financiamiento externo de ninguna organización gubernamental, empresa privada o entidad internacional. Fue desarrollada íntegramente en el marco académico de la Escuela Superior Politécnica de Chimborazo (ESPOCH), Facultad de Informática y Electrónica, Carrera de Tecnologías de la Información, como actividad del curso de Deep Learning 2, período académico Marzo–Julio 2026, bajo la dirección del Ing. Jaime Camacho Mgs. Los recursos computacionales utilizados para la revisión bibliográfica (acceso a bases de datos académicas, Google Colab) son de acceso libre o institucional.

CONFLICTO DE INTERESES

Los autores declaran que no existe ningún conflicto de intereses, financiero, personal o institucional, en relación con la presente investigación. La selección de artículos y el análisis de resultados fueron realizados de manera objetiva e independiente, siguiendo estrictamente los criterios PRISMA predefinidos, sin influencia de intereses externos.

CONTRIBUCIÓN DE AUTORÍA

Tabla de contribución

Contribución de autoría según la taxonomía CRediT

Participar activamente en:	Camacho	Chichande	Quezada	Núñez	Caiño	Remache
Conceptualización	X	X	X	–	–	–
Análisis formal	X	–	X	X	X	–
Adquisición de fondos	–	–	–	–	–	–
Investigación	–	X	X	X	X	X
Metodología	X	X	–	–	X	X
Administración del proyecto	X	X	–	–	–	X
Recursos	–	X	–	X	X	X
Redacción – borrador original	X	X	X	X	X	X
Redacción – revisión y edición	X	X	X	X	X	X
Discusión de los resultados	X	X	X	X	X	X
Revisión y aprobación de la versión final	X	X	X	X	X	X

Nota: Distribución declarada conforme a la taxonomía CRediT.

REFERENCIAS

Adebiyi, A., Abdalnabi, N., Hoffman Smith, E., Hirner, J., Simoes, E.J., Becevic, M., & Rao, P. (2024). Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 and ISIC 2017 Datasets. medRxiv. <https://doi.org/10.1101/2024.05.30.24308213>

Ahmed, I., Bushon Routh, B., Rahman Kohinoor, Md.S., Sakib, S., Mahfuzur Rahman, M., & Azzedin, F. (2024). Multi-model attentional fusion ensemble for accurate skin cancer classification. IEEE Access, 12, 181009–181024. <https://doi.org/10.1109/ACCESS.2024.3509053>

Aladhadh, S., Alsanea, M., Aloraini, M., Khan, T., Habib, S., & Islam, M. (2022). An effective skin cancer classification mechanism via medical vision transformer. Sensors, 22(11), 4008. <https://doi.org/10.3390/s22114008>

Arnold, M., Singh, D., Laversanne, M., Vignat, J., Vaccarella, S., Meheus, F., & Bray, F. (2022). Global burden of cutaneous melanoma in 2020 and projections to 2040. JAMA Dermatology, 158(5), 495–503. <https://doi.org/10.1001/jamadermatol.2022.0160>

Arshed, M.A., Mumtaz, S., Ibrahim, M., Ahmed, S., Tahir, M., & Shafi, M. (2023). Multi-class skin cancer classification using vision transformer networks and convolutional neural network-based pre-trained models. Information, 14(7), 415. <https://doi.org/10.3390/info14070415>

Ayas, S. (2023). Multiclass skin lesion classification in dermoscopic images using swin transformer model. Neural Computing and Applications, 35, 6713–6722. <https://doi.org/10.1007/s00521-022-08053-z>

Bogrekci, I., Alhammadi, A., & Ozcelik, S. (2024). Enhancing Melanoma Diagnosis with Advanced Deep Learning Models Focusing on Vision Transformer, Swin Transformer, and ConvNeXt. Dermatopathology, 11(3), 26. <https://doi.org/10.3390/dermatopathology11030026>

- Cassidy, B., Kendrick, C., Brodzicki, A., Jaworek-Korjakowska, J., & Yap, M.H. (2022). Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Medical Image Analysis*, 75, 102305. <https://doi.org/10.1016/j.media.2021.102305>
- Chanda, T., Das, S., Sinhamahapatra, P., Koller, S., Bertschinger, N., Atzori, M., & Weichenthal, M. (2024). Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. *Nature Communications*, 15(1), 524. <https://doi.org/10.1038/s41467-023-44440-3>
- Combalia, M., Codella, N.C.F., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., & Malvey, J. (2019). BCN20000: Dermoscopic Lesions in the Wild. *Scientific Data*, 11(1), 641. <https://doi.org/10.1038/s41597-024-03387-w>
- Dagnaw, G.H., El Mouhtadi, M., & Mustapha, M. (2024). Skin cancer classification using vision transformers and explainable artificial intelligence. *Journal of Medical Artificial Intelligence*, 7. <https://doi.org/10.21037/jmai-24-29>
- Ding, Y., Yi, Z., Xiao, J., Hu, M., Guo, Y., Liao, Z., & Wang, Y. (2024). CTH-Net: A CNN and Transformer hybrid network for skin lesion segmentation. *iScience*, 27(4), 109442. <https://doi.org/10.1016/j.isci.2024.109442>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ICLR 2021*. <https://arxiv.org/abs/2010.11929>
- Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- Hauser, K., Kurz, A., Haggenmüller, S., Maron, R.C., von Kalle, C., Utikal, J.S., & Kriehoff-Henning, E. (2022). Explainable artificial intelligence in skin cancer recognition: a systematic review. *European Journal of Cancer*, 167, 54–69. <https://doi.org/10.1016/j.ejca.2022.02.025>
- He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. *IEEE/CVF CVPR*, pp. 16000–16009. <https://doi.org/10.1109/CVPR52688.2022.01553>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE/CVF CVPR*, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Himel, G.M.S., Rahat, M.M.A., Hasan, Md.K., Ema, R.R., Uddin, M.N., Hossain, M.S., & Islam, M.K. (2024). Skin Cancer Segmentation and Classification Using Vision Transformer for Automatic Analysis in Dermoscopy-Based Noninvasive Digital System. *International Journal of Biomedical Imaging*, 2024, 3861399. <https://doi.org/10.1155/2024/3861399>
- Karthik, R., Menaka, R., Atre, S., Cho, J., & Easwaramoorthy, S.V. (2024). A hybrid deep learning approach for skin cancer classification using swin transformer and dense group shuffle non-local attention network. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2024.124578>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *IEEE/CVF ICCV*, pp. 10012–10022. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Lundberg, S.M., & Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. *NeurIPS*, 30, 4765–4774. <https://arxiv.org/abs/1705.07874>
- Maron, R.C., Weichenthal, M., Utikal, J.S., Hekler, A., Berking, C., Hauschild, A., & Brinker, T.J. (2019). Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *European Journal of Cancer*, 119, 57–65. <https://doi.org/10.1016/j.ejca.2019.06.013>
- Mohan, J., Sivasubramanian, A., Sowmya, V., & Ravi, V. (2025). Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable AI. *Computers in Biology and Medicine*, 190, 110007. <https://doi.org/10.1016/j.combiomed.2025.110007>
- Munjal, G., Bhatia, P., Srivastava, G., & Munjal, D. (2024). SkinSage XAI: An explainable deep learning solution for skin lesion diagnosis. *Health Care Science*, 3(6), 438–455. <https://doi.org/10.1002/hcs2.121>
- Nie, Y., Sommella, P., Carratù, M., O'Nils, M., & Lundgren, J. (2023). A Deep CNN Transformer Hybrid Model for Skin Lesion Classification of Dermoscopic Images Using Focal Loss. *Diagnostics*, 13(1), 72. <https://doi.org/10.3390/diagnostics13010072>
- Nunnari, F., Kadir, M.A., & Sonntag, D. (2021). On the overlap between Grad-CAM saliency maps and explainable visual features in skin cancer images. *Lecture Notes in Computer Science*, Springer, pp. 241–253. https://doi.org/10.1007/978-3-030-84060-0_16
- Rezaee, K., & Zadeh, H.G. (2024). Self-attention transformer unit-based deep learning framework for skin lesions classification in smart healthcare. *Discover Applied Sciences*, 6, 3. <https://doi.org/10.1007/s42452-024-05655-1>
- Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *KDD 2016*, pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *ICML*, 97, 6105–6114. <https://arxiv.org/abs/1905.11946>

- Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. <https://doi.org/10.1038/sdata.2018.161>
- Wu, Y., Chen, B., Zeng, A., Pan, D., Wang, R., & Zhao, S. (2022). Skin cancer classification with deep learning: a systematic review. *Frontiers in Oncology*, 12, 893972. <https://doi.org/10.3389/fonc.2022.893972>
- Xin, C., Wen, H., Zhang, L., Xu, X., Sun, J., & Zhang, Y. (2022). An improved transformer network for skin cancer classification. *Computers in Biology and Medicine*, 149, 105939. <https://doi.org/10.1016/j.compbiomed.2022.105939>
- Yousaf, N., Amin, J., Butt, W.H., et al. (2026). Advanced hybrid transformer CNN framework for improved skin lesion classification and segmentation. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-43376-0>