

Tamizaje asistido por inteligencia artificial en mamografías: clasificación preliminar y priorización automatizada

Afiliación: Laboratorio de Procesamiento de Imágenes Digitales, Universidad Nacional de Costa Rica

Abstract

Breast cancer remains a leading cause of mortality among women worldwide. Despite the impact of screening programs, mammography interpretation is time-consuming and subject to inter-reader variability, motivating decision-support tools. This study presents a proof-of-concept pipeline that retrains a convolutional neural network to classify mammograms as malignant or benign and then produces a priority ranking of all cases according to malignancy probability, enabling triage-oriented reading order. Using the MIAS dataset restricted to labeled images ($n = 322$), we performed metadata cleaning, class-balanced augmentation, transfer learning, model evaluation, and large-scale inference to export probabilistic scores and a ranked CSV. The model predicted 121 images as malignant and 201 as benign (threshold 0.5), with a mean malignancy probability of 0.39 ± 0.40 . While performance optimization and diverse data remain necessary, these results illustrate the feasibility of an AI-assisted prioritization strategy to support clinical workflows, with future work focusing on calibration, external validation, and integration into PACS-based environments.

Resumen

Este artículo presenta un estudio preliminar de un sistema de tamizaje asistido por inteligencia artificial (IA) en mamografías. Se implementó un procesamiento modular basado en redes neuronales convolucionales (CNNs) para clasificar imágenes como malignas o benignas, y se añadió un mecanismo de priorización automática que ordena estudios según la probabilidad de malignidad. Para la evaluación inicial se utilizó el dataset MIAS, restringido a las imágenes con clasificación disponible ($n = 322$). El modelo alcanzó una proporción de 121 casos clasificados como malignos y 201 como benignos, con una probabilidad media de malignidad de 0.39 ± 0.40 . A pesar de las limitaciones del conjunto de datos, los resultados sugieren que esta herramienta puede constituir un apoyo al diagnóstico médico, permitiendo priorizar casos sospechosos y optimizar el flujo de trabajo clínico.

Introducción

El cáncer de mama continúa siendo una causa principal de mortalidad en mujeres; la mamografía, núcleo del tamizaje poblacional, mitiga esa carga cuando detecta lesiones en etapas tempranas. Sin embargo, la lectura es demandante y variable entre observadores, por lo que se han propuesto sistemas de apoyo basados en aprendizaje profundo que complementan —no reemplazan— el juicio clínico. En los últimos años, modelos de IA han mostrado rendimientos comparables o superiores a lectores humanos en evaluaciones retrospectivas y multicéntricas, y se han publicado estudios prospectivos que evalúan su impacto real en el cribado, abriendo la puerta a flujos como lectura asistida, segunda lectura y triage (priorización de casos).

Este estudio fue desarrollado en el Laboratorio de Procesamiento de Imágenes Digitales de la Universidad Nacional de Costa Rica, en el marco de iniciativas que buscan explorar aplicaciones de inteligencia artificial adaptadas a contextos regionales y a necesidades de los sistemas de salud en América Latina.

Materiales y Métodos

El proceso experimental se estructuró como un procesamiento modular diseñado para vincular clasificación y priorización de mamografías. La primera etapa consistió en la preparación del dataset MIAS, restringido a las imágenes con clasificación disponible. Se eliminaron metadatos incrustados en los archivos y se reorganizó el conjunto en particiones de entrenamiento, validación y prueba con proporciones de 80/10/10, preservando la distribución de clases.

Para mitigar la escasez de muestras, se aplicaron técnicas de aumentación controlada, entre ellas rotaciones de hasta $\pm 15^\circ$, traslaciones horizontales y verticales (10%), zoom aleatorio (10%), reflejos horizontales y ajustes en brillo/contraste. Estas operaciones buscaban incrementar la diversidad del conjunto manteniendo la fidelidad de los patrones radiológicos.

El modelo principal se fundamentó en la arquitectura MobileNetV2, seleccionada por su balance entre precisión y eficiencia computacional, aspecto crítico en escenarios con recursos limitados. Se utilizaron pesos preentrenados en ImageNet como punto de partida. La arquitectura se extendió con un bloque superior compuesto por:

- Global Average Pooling sobre la salida convolucional.
- Una capa densa de 256 neuronas con activación ReLU, que actúa como clasificador intermedio.
- Dos capas de dropout ($p=0.5$ y $p=0.3$) intercaladas para reducir sobreajuste.

- Una capa final densa de una neurona con activación sigmoide, que produce una probabilidad en el rango $[0,1]$ asociada al diagnóstico de malignidad.

El entrenamiento se llevó a cabo en dos fases. En la primera fase, las capas convolucionales del modelo base se congelaron, entrenando únicamente la cabeza añadida con el optimizador Adam (tasa de aprendizaje $1e-4$) y función de pérdida binary crossentropy. En la segunda fase, se habilitó el ajuste fino (*fine-tuning*) liberando las últimas 50 capas de MobileNetV2, entrenadas con una tasa de aprendizaje reducida ($1e-5$) para estabilizar la convergencia.

Se trabajó con un tamaño de lote (batch size) de 32 imágenes, un máximo de 500 épocas para la fase inicial y 200 para el fine-tuning, ambos con mecanismos de early stopping (paciencia de 100 épocas) y ModelCheckpoint para conservar el mejor modelo según la exactitud de validación. La normalización de entradas se realizó dividiendo valores de píxel por 255, escalando las intensidades al rango $[0,1]$.

Estas configuraciones, combinadas con la estrategia de aumentación y la estructura ligera de MobileNetV2, permitieron entrenar un clasificador binario robusto en un dataset reducido, reduciendo riesgos de sobreajuste y maximizando la capacidad de generalización.

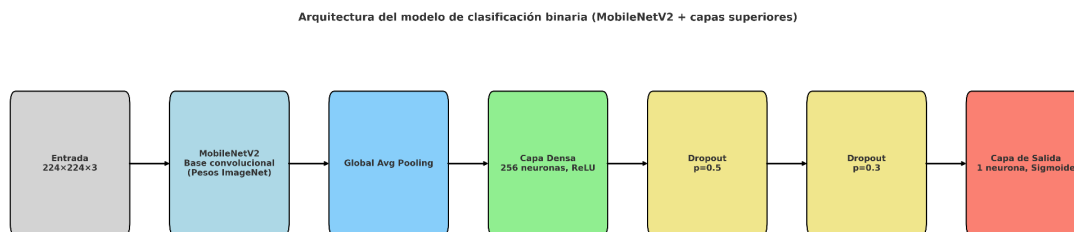


Figura 0. Representación esquemática de la arquitectura empleada basada en MobileNetV2 con capas adicionales para clasificación binaria.

Resultados

El procesamiento modular abarcó 322 mamografías. Bajo un umbral de decisión de 0.5, 121 imágenes fueron clasificadas como malignas y 201 como benignas. La probabilidad media de malignidad se situó en 0.39 ± 0.40 , mientras que la de benignidad fue 0.61 ± 0.40 . Estas cifras reflejan la capacidad del modelo para discriminar casos dentro de un dataset reducido, aunque con variabilidad significativa entre ejemplos individuales.

El ranking de priorización permitió identificar de manera inmediata los casos más sospechosos: el top-10 de estudios alcanzó probabilidades cercanas a 1.0 en la clase maligna, lo cual ilustra el potencial de la herramienta para apoyar la revisión temprana de

hallazgos críticos. Adicionalmente, el análisis de la distribución de probabilidades mostró un espectro bimodal, con concentraciones en valores bajos y altos, lo que sugiere que el modelo tiende a generar predicciones confiadas, aunque se requieren técnicas de calibración para garantizar su uso en escenarios clínicos.

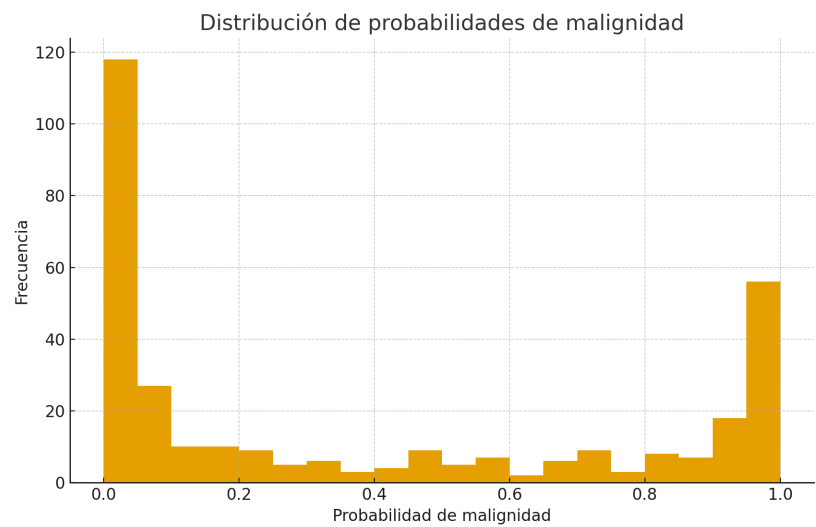


Figura 1. Distribución de probabilidades de malignidad.

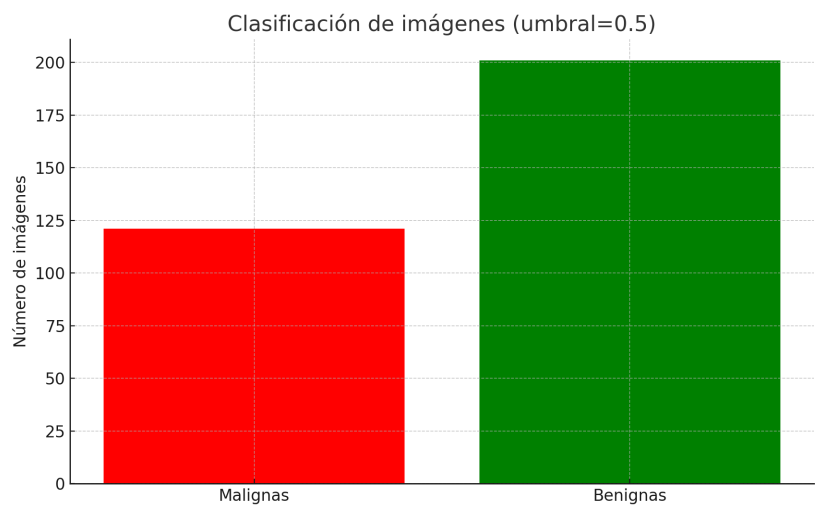


Figura 2. Número de imágenes clasificadas como malignas vs benignas.

Tabla 1. Top-10 imágenes con mayor probabilidad de malignidad.

Imagen	Probabilidad de malignidad
--------	----------------------------

mdb238.pgm	1.00
mdb216.pgm	1.00
mdb287.pgm	1.00
mdb213.pgm	1.00
mdb027.pgm	1.00
mdb233.pgm	1.00
mdb158.pgm	1.00
mdb170.pgm	1.00
mdb256.pgm	1.00
mdb124.pgm	1.00

Discusión

Los resultados confirman que un procesamiento modular compacto puede articular clasificación y priorización de mamografías, proponiendo un marco operativo de triage asistido por IA. Este planteamiento se alinea con trabajos que, en entornos simulados, demostraron que descartar de forma segura fracciones de estudios o reordenar colas de lectura puede aumentar especificidad y eficiencia sin deteriorar sensibilidad, y con evaluaciones stand-alone que encuentran rendimientos comparables al promedio de radiólogos en escenarios retrospectivos. En paralelo, experiencias prospectivas —incluidas implementaciones a gran escala— muestran que la IA puede incrementar la detección y reducir carga de lectura, si bien los efectos en tasas de recall y flujos de trabajo dependen del diseño local y del rol de la IA. Nuestro estudio se sitúa como paso previo: demuestra el mecanismo de priorización basado en probabilidad de malignidad dentro de un procesamiento modular reproducible, con la intención explícita de integrarlo en flujos clínicos y medir impacto real.

Desde el punto de vista metodológico, dos aspectos emergen como palancas de mejora. Primero, la calibración probabilística y la incertidumbre: transformar puntajes en probabilidades bien calibradas y cuantificar la confianza habilita umbrales clínicamente interpretables y reglas de triage más seguras. Segundo, la generalización: MIAS es un conjunto clásico y pequeño; por ello, la robustez ante diversidad fenotípica, densidad mamaria y variaciones de hardware exige validaciones externas multicéntricas (CBIS-DDSM, INbreast) y, preferentemente, evaluaciones prospectivas que midan indicadores operativos junto con métricas diagnósticas estándar.

Trabajo Futuro

Con base en lo anterior, el siguiente ciclo de investigación se articulará en torno a tres ejes que se retroalimentan. En primer lugar, la optimización del modelo mediante fine-tuning de arquitecturas modernas e incorporación de calibración pos-entrenamiento y estimación de incertidumbre, con el fin de definir umbrales operativos y estrategias de priorización segura. En segundo lugar, la validación externa y prospectiva, trasladando el procesamiento modular a conjuntos multicéntricos y posteriormente a un entorno clínico real con integración en PACS o visor, donde se mida no sólo el rendimiento diagnóstico sino también el impacto en el flujo de trabajo. Un aspecto clave será extender el estudio hacia la representación de biotipos y fenotipos característicos de poblaciones de Centroamérica, considerando variaciones en densidad mamaria, equipamiento radiológico y perfiles epidemiológicos regionales. Esta ampliación permitirá evaluar la robustez del modelo frente a la diversidad propia de la región y aportar evidencia para su eventual incorporación en programas de tamizaje locales. En tercer lugar, la explicabilidad y la aceptación clínica, complementando la priorización con mapas de atención y protocolos de interacción que documenten cuándo y por qué el sistema sugiere priorizar un estudio, favoreciendo transparencia, auditoría y adopción.

Conclusiones

Este trabajo presenta un procesamiento modular de clasificación y priorización de mamografías sobre MIAS que transforma probabilidades de malignidad en un orden de lectura orientado al riesgo. La contribución central no reside en alcanzar un techo de rendimiento en un conjunto pequeño, sino en demostrar un mecanismo reproducible para materializar el triage asistido por IA: desde la limpieza y aumentación, pasando por el reentrenamiento con transferencia, hasta la inferencia masiva y la exportación de una lista priorizada interpretable por el clínico. A la luz de la evidencia disponible, que reporta rendimientos competitivos de IA respecto a radiólogos y beneficios potenciales en escenarios prospectivos, la ruta razonable es avanzar hacia validaciones externas multicéntricas y estudios prospectivos que cuantifiquen beneficios operativos y seguridad clínica. Si tales estudios confirman que la priorización mejora los tiempos de lectura sin penalizar la calidad diagnóstica, la integración de esta herramienta como soporte —nunca sustituto— de la práctica radiológica podría contribuir a un tamizaje más eficiente y equitativo.

Referencias

- [1] J. Suckling et al., 'The Mammographic Image Analysis Society Digital Mammogram Database,' *Excerpta Medica, International Congress Series*, vol. 1069, pp. 375–378, 1994.
- [2] Google Dataset Search, 'The Mammographic Image Analysis Society database of digital mammograms (v1.21), 322 images (PGM). Accessed Sep. 15, 2025.'

- [3] S. M. McKinney et al., 'International evaluation of an AI system for breast cancer screening,' *Nature*, vol. 577, pp. 89–94, 2020.
- [4] A. Yala, T. Schuster, R. Miles, R. Barzilay, and C. Lehman, 'A Deep Learning Model to Triage Screening Mammograms: A Simulation Study,' *Radiology*, vol. 293, no. 1, pp. 38–46, 2019.
- [5] A. Rodríguez-Ruiz et al., 'Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101 Radiologists,' *J. Natl. Cancer Inst.*, vol. 111, no. 9, pp. 916–922, 2019.
- [6] K. Dembrower et al., 'Artificial intelligence for breast cancer detection in screening mammography: prospective evidence,' *Lancet Digital Health*, 2023.
- [7] A. Y. Ng et al., 'Prospective implementation of AI-assisted screen reading for breast cancer detection,' *Nat. Med.*, 2023.
- [8] L. Wang et al., 'Mammography with deep learning for breast cancer detection: recent achievements,' *BMC Medical Imaging*, 2024.
- [12] Radiological Society of North America (RSNA), 'Deep Learning Predicts Interval and Screening-Detected Breast Cancer,' *RadioGraphics*, 2021.
- [18] O. Díaz et al., 'Artificial Intelligence for breast cancer detection,' 2024.
- [24] T. J. A. van Nijnatten et al., 'Overview of trials on AI algorithms in screening mammography,' 2023.