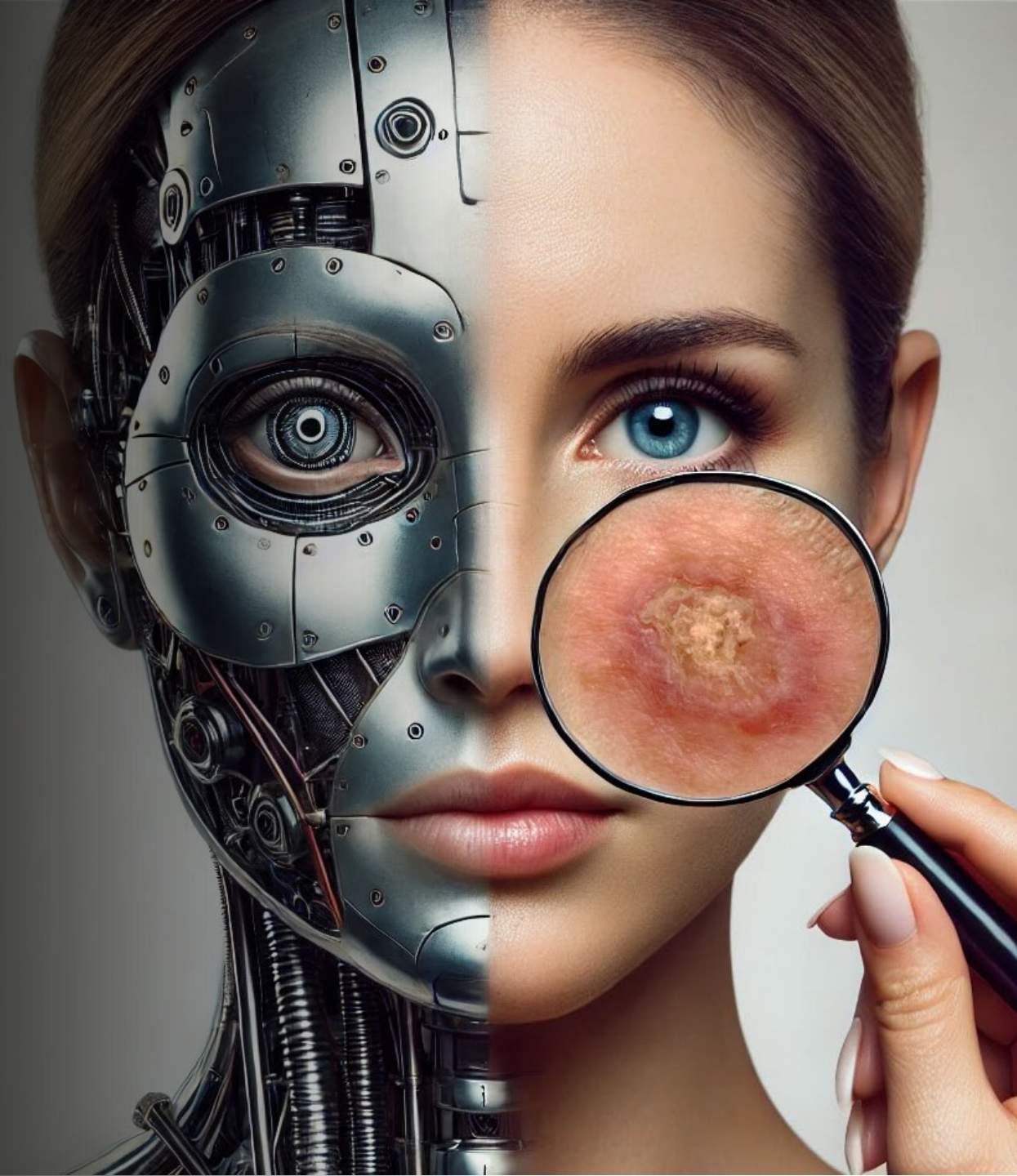


Proyecto de Inteligencia Artificial para la Clasificación de Problemas de Piel

Creado Por:

Kevin Gomez, Eduardo Lezama, Irina M.
Sinning, Bryan Ruiz.

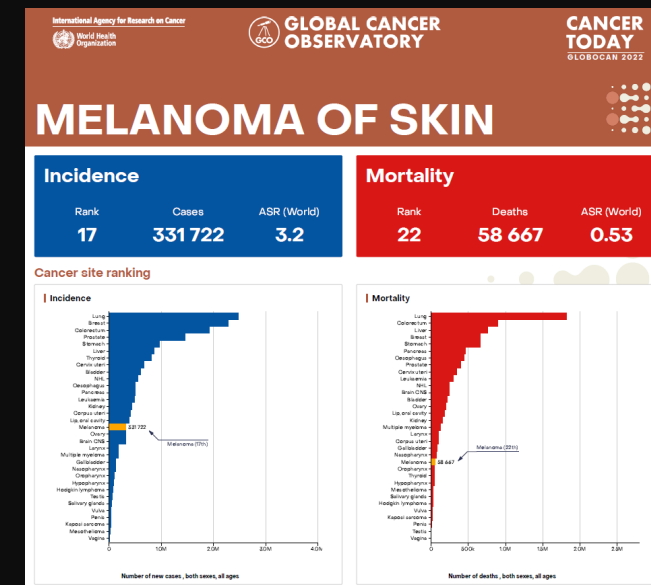
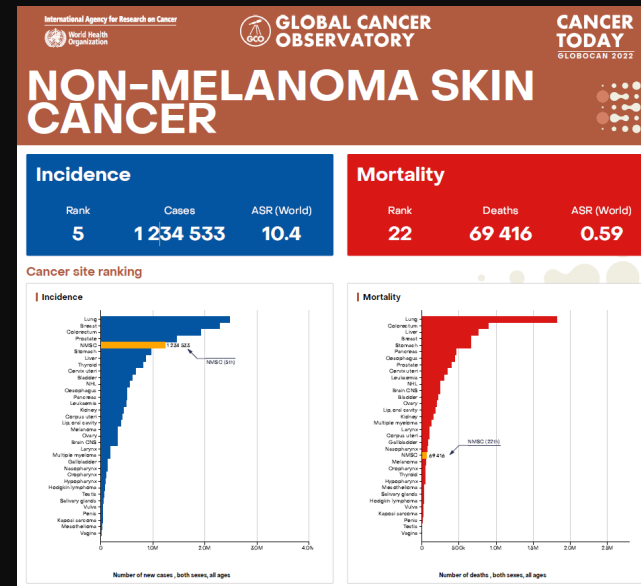


Planteamiento del problema

Más de 1.5 millones de nuevos casos reportados a nivel mundial en el año 2022.

El melanoma, tipo de cáncer más agresivo, es responsable de la mayoría de las muertes por su tendencia a metastásis (invasión de otro órganos)

Diagnóstico temprano es clave para una posible cura.



Contexto del Problema

La red pública de salud enfrenta demoras en citas con dermatólogos.

Una solución basada en IA puede ayudar a médicos generales a:

Clasificar y priorizar lesiones sospechosas.

Acelerar el triaje y la derivación de casos urgentes

El diagnóstico temprano, especialmente en melanoma, es clave para reducir la mortalidad.



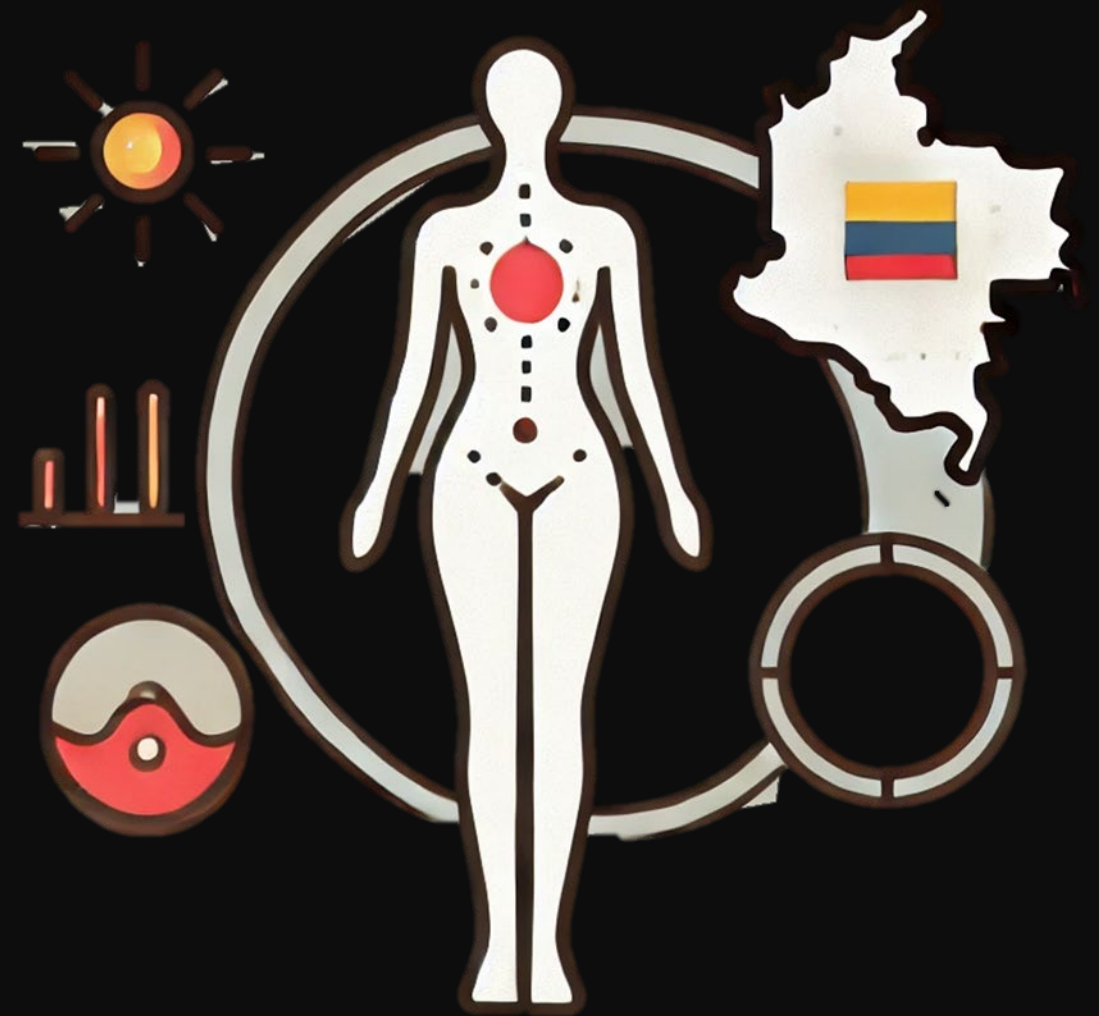
Estadísticas

Incidencia global (2022)

- Más de 1,500,000 casos nuevos de cáncer de piel no melanoma.
- 324,635 casos de melanoma maligno.
- 57,043 muertes por melanoma maligno.

Incidencia nacional (2022)

- 1062 casos nuevos de cáncer de piel reportados por el Instituto Nacional de Cancerología (INC).
- Regiones más afectadas: Bogotá, D.C., seguida de los departamentos de Cundinamarca, Boyacá, Tolima y Meta.



Objetivo Principal

Desarrollar un modelo de CNN para clasificar imágenes de piel en 9 clases dermatológicas mediante augmentación de datos, balanceo de clases y transfer learning, mejorando así la precisión del modelo en esta tarea.

Objetivos específicos

1. Aplicar transfer learning para adaptar modelos preentrenados a la clasificación dermatológica.
2. Implementar augmentación y balanceo de datos para una mejor representatividad de cada clase.
3. Evaluar el rendimiento del modelo en un conjunto de datos de prueba.

Dataset

Skin Cancer 9 Classes (ISIC)

Descripción: Conjunto de imágenes de lesiones cutáneas clasificadas en 9 categorías dermatológicas.

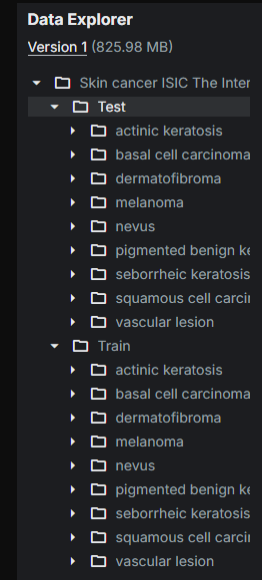
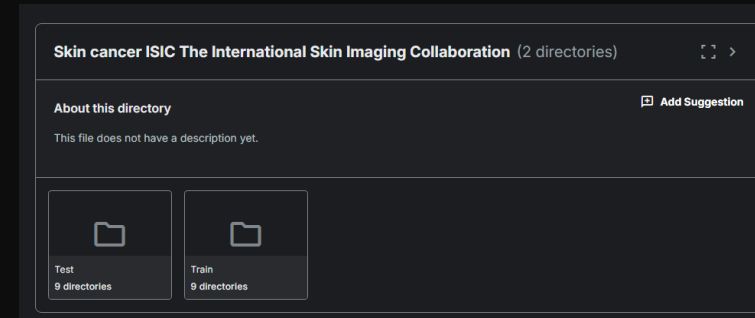
Clases: Incluye tipos de condiciones de la piel como melanoma, carcinoma, y otras lesiones comunes.

Cantidad de Imágenes: Aproximadamente 23,000 imágenes en total.

Objetivo: Proporcionar datos para la investigación y entrenamiento de modelos de IA en la clasificación de cáncer de piel.

Desafíos:

- **Desbalanceo de Clases:** Algunas categorías tienen más imágenes que otras.
- **Variabilidad de las Imágenes:** Diferentes tipos de iluminación, tamaño y calidad.



Metodología

Preparación del Dataset:

- División del conjunto de datos en entrenamiento, validación y prueba.
- Creación de un flujo de datos a partir de las carpetas de imágenes.

Entrenamiento Inicial:

- División del conjunto de datos en entrenamiento, validación y prueba.
- Creación de un flujo de datos a partir de las carpetas de imágenes.

Modelo Base:

- Uso de modelo preentrenado como ResNet50 mediante transfer learning.
- Ajuste fino de las capas superiores del modelo para adaptarse al dataset específico de cáncer de piel.

Preprocesamiento:

- Redimensionamiento de las imágenes a 224x224 píxeles.
- Augmentación de datos (rotación, zoom, y volteo) para aumentar la variabilidad y reducir el sobreajuste.
- Entrenamiento con Augmentación de datos.

Análisis de Desbalance de Clases:

- Identificación de desbalance en las clases y creación de datos sintéticos mediante la librería Augmentor para equilibrar las categorías minoritarias.

Entrenamiento y Evaluación:

- Entrenamiento con optimización Adam y función de pérdida de entropía cruzada categórica.
- Validación en cada época para monitorear el rendimiento y detectar posibles signos de sobreajuste.

Mejora y Ajuste del Modelo:

- Uso de dropout y augmentación adicional para reducir el sobreajuste.
- Almacenamiento del modelo final y evaluación en datos de prueba para medir precisión y pérdida final.

Estructura de la CNN

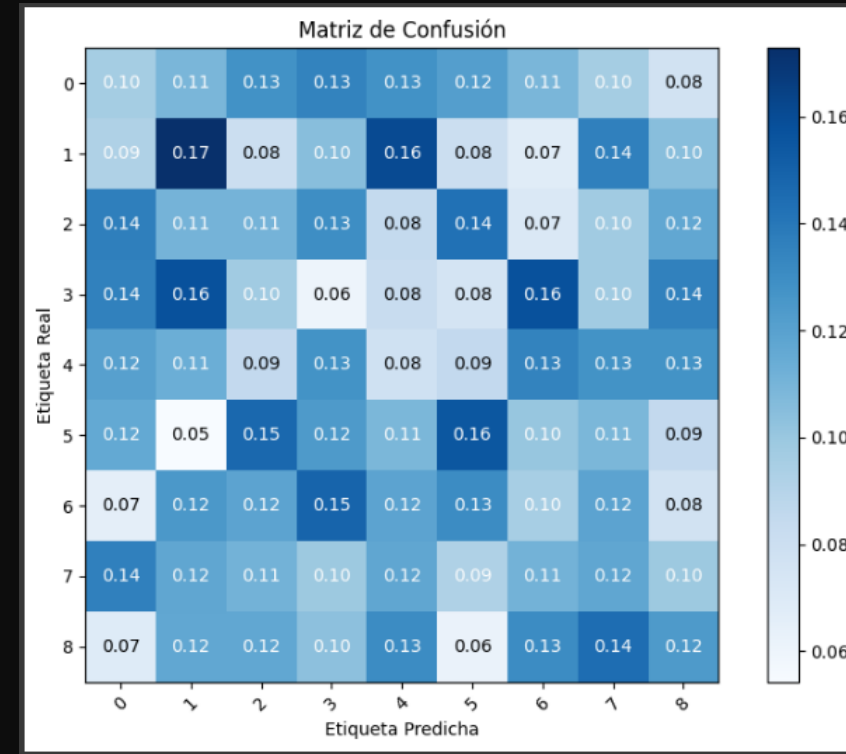
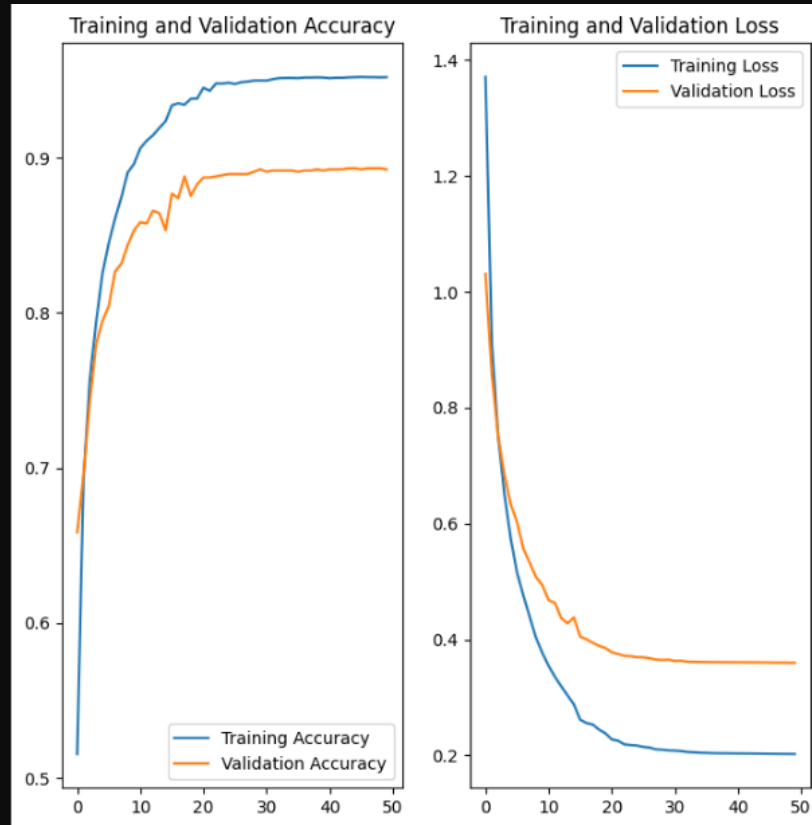
Layer (type)	Output Shape	Param #
sequential_1 (Sequential)	(None, 224, 224, 3)	0
resnet50 (Functional)	(None, 7, 7, 2048)	23,587,712
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 2048)	0
dense_2 (Dense)	(None, 9)	18,441

Total params: 23,606,153 (90.05 MB)

Trainable params: 18,441 (72.04 KB)

Non-trainable params: 23,587,712 (89.98 MB)

Gráficas



Referencias

- American Cancer Society. (n.d.). Skin cancer risk factors. American Cancer Society. Accedido 18-10-2024, from <https://www.cancer.org/cancer/skin-cancer/prevention-and-early-detection/risk-factors.html>
- Barrera-Valencia, C., & Perea-Flórez, E. X. (2024). Comparison of costs in teledermatology using PC and camera versus smartphone. Telemedicine and e-Health, 30(7), e2087-e2095. <https://doi.org/10.1089/tmj.2023.0369>
- Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F. (2024). Cancer fact sheets. Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer. Disponible en: <https://gco.iarc.fr/today/en/fact-sheets-cancers>, accedido 18-10-2024.
- Anuario estadístico 2022. Instituto Nacional de Cancerología (INC). Bogotá, D. C.: INC; 2023.
- Skin Cancer ISIC [Dataset]. Kaggle. Disponible en: <https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic/>, accedido 18-10-2024.
- Valenzuela, J. I., Arguello, A., Cendales, J. G., & Rizo, C. A. (2007). Web-based asynchronous teleconsulting for consumers in Colombia: A case study. Journal of Medical Internet Research, 9(4), e33. <https://doi.org/10.2196/jmir.9.4.e33>