

**Modelo de detección de glaucoma en imágenes de fondo de ojo.**

Proyecto realizado por Joaquín Rodríguez, Técnico en Ciencias de Datos.

Este proyecto fue realizado como proyecto final de la Tecnicatura en Ciencias de Datos del Instituto Teclab.

[Linkedin](#) | [Github](#) | [Video explicativo en Youtube](#)

## Índice.

1. Resumen.
2. Introducción.
  - 2.1. Importancia clínica del glaucoma y necesidad de detección temprana
  - 2.2. Limitaciones de los métodos diagnósticos actuales.
  - 2.3. Interpretabilidad e integración clínica: uso de Grad-CAM.
  - 2.4. Selección de arquitectura eficiente para implementación práctica.
3. Materiales y métodos.
  - 3.1. Dataset.
  - 3.2. Arquitectura del modelo.
  - 3.3. Técnicas avanzadas implementada.
  - 3.4. Cálculo de “nivel de urgencia”.
4. Resultado.
  - 4.1. Métricas de desempeño (Accuracy, AUC, sensibilidad, especificidad, matriz de confusión).
  - 4.2. Análisis de interpretabilidad (mapas Grad-CAM, ejemplos).
5. Aplicación de escritorio.
  - 5.1. Arquitectura interna y funcionamiento.
  - 5.2. Paneles Principales.
  - 5.3. Proyección de uso clínico.
6. Discusión, conclusiones y trabajo futuro.
  - 6.1. Interpretación y conclusión sobre la Hipótesis.
  - 6.2. El rol del modelo como herramienta de apoyo clínico.
  - 6.3. Limitaciones del proyecto y líneas de trabajo futuro.
7. Referencia.

## 1. Resumen.

El glaucoma es una neuropatía óptica progresiva e irreversible que constituye la segunda causa principal de ceguera a nivel mundial, el aumento de la presión intraocular es el principal factor de riesgo, pero no el que causa el glaucoma, aunque su nivel determina el daño del nervio óptico.

La detección temprana del glaucoma es un reto de salud pública debido a su progresión asintomática en estadios iniciales y la elevada carga poblacional proyectada. Propongo un pipeline de detección automática basado en MobileNetV2 (transfer learning + fine-tuning parcial) entrenado sobre imágenes de fondo de ojo con aumentación agresiva. Para hacer la salida interpretable incorporamos Grad-CAM y definimos un nivel de urgencia calculado como el promedio entre la probabilidad del modelo y la proporción del área activada por Grad-CAM. En la validación el modelo alcanzó  $AUC \approx 0.94$  y produjo mapas de activación localizados en regiones compatibles con hallazgos clínicos (ej. región del disco óptico). Entregamos además una aplicación de escritorio que integra inferencia, visualización (heatmap puro y overlay), historial y exportación CSV para trazabilidad. El sistema está planteado como ayuda de cribado/triage, y requiere estudios prospectivos y validación externa para uso clínico.

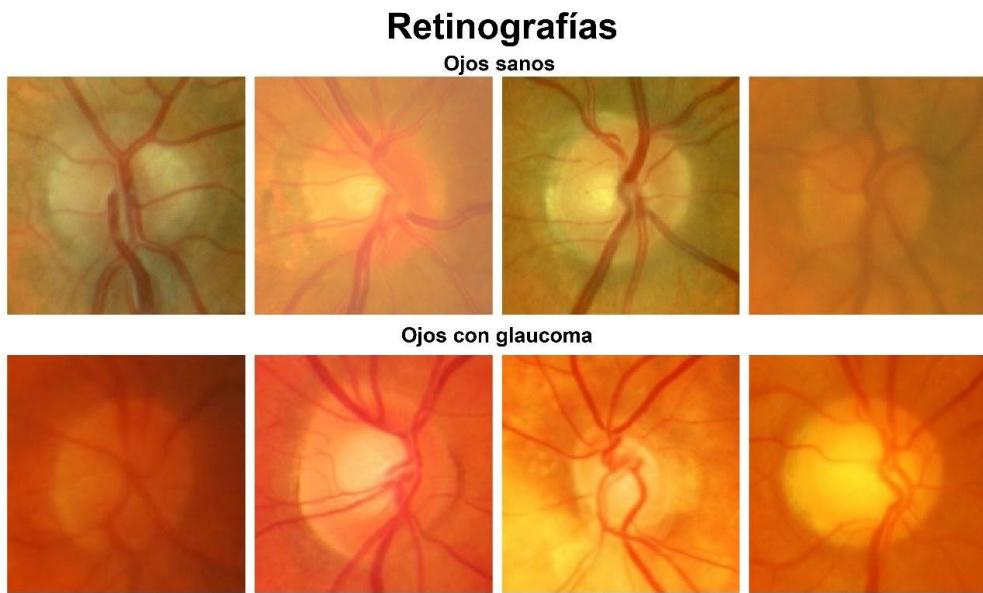
## 2. Introducción.

### 2.1. Importancia clínica del glaucoma y necesidad de detección temprana.

El glaucoma es una enfermedad crónica, progresiva e irreversible del nervio óptico que se caracteriza por la pérdida progresiva de las fibras nerviosas y cambios en su aspecto. En la mayoría de los casos, la enfermedad se asocia a un aumento patológico de la presión intraocular (PIO), causado por una dificultad en el drenaje del humor acuoso, lo que daña el nervio óptico. Es considerada la segunda causa principal de ceguera irreversible a nivel mundial.

El impacto global del glaucoma es masivo; se estima que 111.8 millones de personas padecerán la enfermedad para el año 2040. La prevalencia mundial se sitúa entre el 1.5% y el 2% en mayores de 40 años, y aumenta con la edad. El tipo más común, el glaucoma primario de ángulo abierto (GPAA), es de evolución lenta y generalmente es asintomático hasta etapas avanzadas de la enfermedad, cuando ya ha ocurrido la pérdida de visión. Esta característica le ha valido el apodo de "ladrón silencioso de la visión".

Debido a que la pérdida de visión causada por el glaucoma es permanente e irreversible, la detección temprana y la intervención son cruciales para prevenir el daño visual irreparable y preservar la calidad de vida del paciente. Sin embargo, se estima que entre el 50% y el 85% de la población afectada, incluso en países desarrollados, desconoce su diagnóstico. En Latinoamérica, este porcentaje de casos no diagnosticados puede ascender entre el 56% y el 75%



## 2.2. Limitaciones de los métodos diagnósticos actuales.

El diagnóstico clínico del glaucoma se basa en una exploración oftalmológica completa que incluye la medición de la PIO (tonometría), la evaluación del ángulo de drenaje (gonioscopia), el examen del nervio óptico y la valoración del campo visual (campimetría).

A pesar de la disponibilidad de pruebas, existen limitaciones significativas en los métodos diagnósticos actuales para el cribado a gran escala:

- Subjetividad y variabilidad humana: Los métodos diagnósticos tradicionales, como la evaluación subjetiva de la fundoscopia (retinografía), carecen de una alta reproducibilidad y fiabilidad interobservador. Los oftalmólogos pueden cometer errores comunes, como el subdiagnóstico en discos ópticos pequeños o el sobrediagnóstico en excavaciones fisiológicamente grandes.
- Insuficiencia del cribado masivo: El cribado masivo en la población general no se recomienda debido a la baja prevalencia, lo que resulta en un alto número de

falsos positivos. Además, la medición de la PIO por sí sola tiene una baja sensibilidad para detectar la enfermedad.

- Limitaciones tecnológicas: Si bien tecnologías como la Tomografía de Coherencia Óptica (OCT) son altamente precisas para cuantificar el daño estructural, son equipos caros y no fácilmente portátiles, lo que limita su viabilidad en programas de cribado masivo o en entornos de bajos recursos. Además, la evaluación estructural convencional con SDOCT depende de un proceso de segmentación automatizado que es imperfecto y propenso a errores (con tasas de error de 19.9% a 46.3%).

#### Aprendizaje profundo en oftalmología: oportunidades y desafíos

El Aprendizaje Profundo (Deep Learning o DL), una subrama de la Inteligencia Artificial (IA), se ha situado a la vanguardia debido a su capacidad para manejar grandes conjuntos de datos y superar a menudo el rendimiento humano en tareas de clasificación y reconocimiento de patrones, especialmente en imágenes. La oftalmología, por ser altamente dependiente de pruebas de imagen complementarias (como la retinografía), ha sido un área privilegiada para la aplicación de DL.

El DL utiliza el "aprendizaje de representación" para inferir características relevantes directamente de los datos brutos, obviando la necesidad de que expertos humanos diseñen manualmente las características de entrada, una limitación clave de los clasificadores tradicionales de machine learning. Las Redes Neuronales Convolucionales (CNN) son el principal tipo de DL utilizado en la visión por computadora y en el análisis de imágenes oftalmológicas.

### Oportunidades:

- Los algoritmos de DL pueden identificar de manera más rápida y precisa el daño glaucomatoso en pruebas diagnósticas.
- El enfoque Machine-to-Machine (M2M) ha demostrado que los algoritmos de DL pueden cuantificar el daño glaucomatoso a partir de fotografías de fondo de ojo con un rendimiento diagnóstico comparable al de la medición del grosor de la capa de fibras nerviosas de la retina (RNFL) con SDOCT (área bajo la curva ROC de 0.940 y 0.944, respectivamente).

### Desafíos:

- El principal desafío es que las características aprendidas automáticamente por los modelos de DL pueden no ser fáciles de entender o explicar, lo que genera la percepción de que son una "caja negra".
- Si los modelos se entranan para replicar las calificaciones humanas (que sirven como gold standard), pueden aprender y replicar errores comunes de los oftalmólogos, como el sobrediagnóstico en copas grandes o el subdiagnóstico en miopes altos.
- La validación de nuevos tests diagnósticos de IA debe basarse en una metodología rigurosa.

### **2.3. Interpretabilidad e integración clínica: uso de Grad-CAM.**

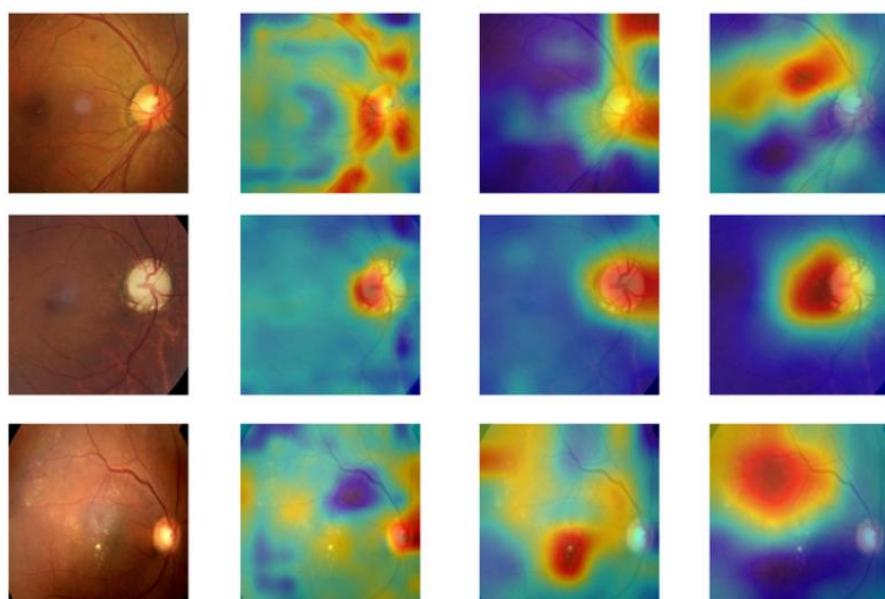
La aceptación y confianza clínica de los sistemas de DL dependen críticamente de su capacidad para demostrar la justificación de sus predicciones, mitigando el problema de la "caja negra".

Para abordar este desafío, este proyecto implementó el Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM es una herramienta de visualización que

genera mapas de calor (heatmaps) que resaltan qué regiones de la imagen son más relevantes para la clasificación del modelo.

La utilidad clínica de Grad-CAM ha sido confirmada en la literatura, mostrando que el algoritmo se enfoca en las estructuras anatómicas claves afectadas por la enfermedad. Por ejemplo, los mapas de activación de la clase (CAM) generados por modelos como el M2M mostraron que la red se enfocaba correctamente en el área del disco óptico y la Capa de Fibras Nerviosas de la Retina (RNFL) adyacente. Otros estudios que analizan redes neuronales para diagnosticar glaucoma mediante fundoscopia también observaron una tendencia marcada a focalizarse en la capa de fibras nerviosas al tomar sus decisiones.

La incorporación de Grad-CAM permite que el sistema no solo clasifique, sino que también proporcione explicaciones visuales para su interpretación clínica, lo que es fundamental para validar la confianza en el sistema antes de su integración en el flujo de trabajo médico.



([https://www.researchgate.net/figure/Grad-Cam-for-AlexNet-ResNet50-GoogLeNet-convolutional-neural-networks-for-the-eye\\_fig4\\_379551022](https://www.researchgate.net/figure/Grad-Cam-for-AlexNet-ResNet50-GoogLeNet-convolutional-neural-networks-for-the-eye_fig4_379551022))

## 2.4. Selección de arquitectura eficiente para implementación práctica.

Un objetivo primordial del proyecto fue la optimización para uso en entornos con recursos limitados, como los programas de telemedicina que buscan aumentar el acceso a servicios de salud en comunidades aisladas o vulnerables en Latinoamérica.

La arquitectura de red neuronal convolucional (CNN) seleccionada fue MobileNetV2. MobileNetV2 fue elegida por su eficiencia computacional y su reducido tamaño en comparación con arquitecturas más pesadas, lo que la hace ideal para su implementación en dispositivos móviles o sistemas embebidos.

La estrategia de utilizar Transfer Learning y Fine-Tuning sobre pesos preentrenados con ImageNet permitió entrenar una red de millones de parámetros (MobileNetV2 tiene 3.5M de parámetros) utilizando un conjunto de datos reducido de 1.707 imágenes de retinografías. La arquitectura del modelo final consistió en la base MobileNetV2, con la última capa totalmente conectada reemplazada por capas de average-pooling, batch normalization, dropout (20%) y una capadense final para la clasificación binaria.

El resultado de este enfoque es la APP\_DET\_GLAUCOMA, una aplicación de escritorio desarrollada en Python, PySide6 y TensorFlow/Keras, diseñada específicamente para la detección asistida de glaucoma en el screening masivo, proporcionando predicciones binarias y visualizaciones interpretables de Grad-CAM.

### 3. Materiales y métodos.

#### 3.1. Dataset.

El proyecto utilizó datasets clínicos de imágenes oculares. Las imágenes utilizadas para el entrenamiento contenían patrones compatibles con glaucoma y se clasificaron en dos categorías: Glaucoma y No Glaucoma. Se enfrentó la limitación de conseguir datos de calidad para el entrenamiento, dado que son datos sensibles y su acceso es difícil.

Partición, limpieza, preprocessamiento y aumentación: Las imágenes pasaron por un proceso de preprocessamiento.

- **Preprocessamiento y limpieza:** Las imágenes fueron redimensionadas a  $224 \times 224$  píxeles. Posteriormente, se realizó la normalización de los valores de píxeles en el rango.

- **Aumentación de Datos (Data Augmentation):** Se aplicó de manera agresiva para simular variaciones clínicas reales y mejorar la generalización. Las técnicas utilizadas incluyeron: rotaciones, giros horizontales/verticales, cambios de brillo y contraste, y zoom y recortes aleatorios.

- **Partición:** Los datos adquiridos fueron seleccionados para entrenamiento, validación y pruebas.

#### Balance del dataset y manejo del desbalance

El proceso metodológico incluyó la etapa de Balanceo de clases para verificar cuáles eran las clases más pesadas y asegurarse de que el modelo las tuviera en cuenta. El balanceo de clases se abordó mediante la aplicación de pesos de clase (class\_weight), que se aplicaron durante la configuración de los hiperparámetros.

### 3.2.Arquitectura del modelo

La construcción de un sistema de diagnóstico asistido por IA eficaz requiere una cuidadosa selección de la arquitectura de la red neuronal y un riguroso flujo de procesamiento de datos. Estos dos pilares son fundamentales para garantizar que el modelo sea no solo preciso, sino también eficiente y capaz de generalizar a partir de los datos de entrenamiento para evaluar nuevas imágenes clínicas.

#### Selección de la Arquitectura Base: MobileNetV2.

La elección de MobileNetV2 como arquitectura base fue una decisión estratégica motivada por un balance entre rendimiento y eficiencia. Las razones clave para esta selección incluyen:

- **Eficiencia Computacional:** MobileNetV2 está específicamente optimizado para su ejecución en dispositivos con recursos limitados, como plataformas móviles y sistemas embebidos, lo que facilita su despliegue en aplicaciones de escritorio sin requerir hardware especializado.
- **Transfer Learning:** Se aprovechan los pesos preentrenados en el vasto dataset de ImageNet. Esto permite que el modelo parte de un conocimiento general sobre la extracción de características visuales, acelerando el entrenamiento y mejorando el rendimiento en tareas específicas como la detección de patologías oculares.
- **Arquitectura de Residuales Invertidos:** Su diseño, basado en conexiones residuales invertidas y cuellos de botella lineales, permite construir redes profundas de manera muy eficiente en términos de parámetros y operaciones computacionales.
- **Tamaño Reducido:** El modelo posee significativamente menos parámetros en comparación con arquitecturas más pesadas como modelos como ResNet o EfficientNet, lo que se traduce en un menor consumo de memoria y tiempos de inferencia más rápidos.

### Arquitectura Detallada de la Red.

La arquitectura final del modelo combina el backbone de MobileNetV2 con un cabezal de clasificación personalizado, diseñado para la tarea específica de detección de glaucoma. El flujo de datos a través de la red es el siguiente:

1. **Entrada:** El modelo recibe imágenes de fondo de ojo preprocesadas con dimensiones de 224x224 píxeles y 3 canales de color (RGB).
2. **Backbone Convolucional:** Se utiliza el cuerpo principal de MobileNetV2 como extractor de características. Las capas iniciales se mantienen congeladas durante la primera fase del entrenamiento para preservar el conocimiento adquirido de ImageNet.
3. **Capa de Agrupación:** Una capa GLOBAL AVERAGE POOLING 2D se aplica a la salida del backbone, condensando los mapas de características en un vector de 1280 características por imagen.
4. **Cabezal de Clasificación Personalizado:** Se añade una secuencia de capas densas para realizar la clasificación final:
  - Una capa Dense de 256 neuronas con función de activación Swish y regularización L2 ( $\lambda=0.001$ ).
  - Una capa de Dropout con un ratio del 60% para mitigar el sobreajuste.
  - Una capa Dense de 128 neuronas con función de activación ReLU y regularización L2 ( $\lambda=0.001$ ).
  - Una capa de Dropout con un ratio del 50%.
  - Finalmente, una capa de salida Dense de 1 neurona con activación Sigmoid, que produce una probabilidad entre 0 y 1, indicando la confianza del modelo en la presencia de glaucoma.

### Flujo de Datos y Preprocesamiento.

El pipeline de preprocesamiento fue diseñado para estandarizar las imágenes y enriquecer el conjunto de datos. Primero, cada imagen se redimensiona a 224x224 píxeles y sus valores de píxeles se normalizan al rango [0, 1]. Posteriormente, se aplica un conjunto de técnicas de Data Augmentation de forma agresiva para simular la variabilidad clínica y mejorar la capacidad de generalización del modelo:

- Rotación: Aleatoria hasta  $\pm 45^\circ$ .
- Zoom: Rango de zoom aleatorio entre 0.7x y 1.3x.
- Desplazamiento: Horizontal y vertical de hasta  $\pm 15\%$ .
- Brillo: Ajuste aleatorio del brillo en un rango de 0.5x a 1.5x.
- Flip Horizontal: Aplicado con una probabilidad del 50%.
- Shear (Inclinación): Deformación aleatoria de hasta  $\pm 0.2$ .
- Channel Shift: Variación aleatoria en la intensidad de los canales de color de hasta  $\pm 50$ .

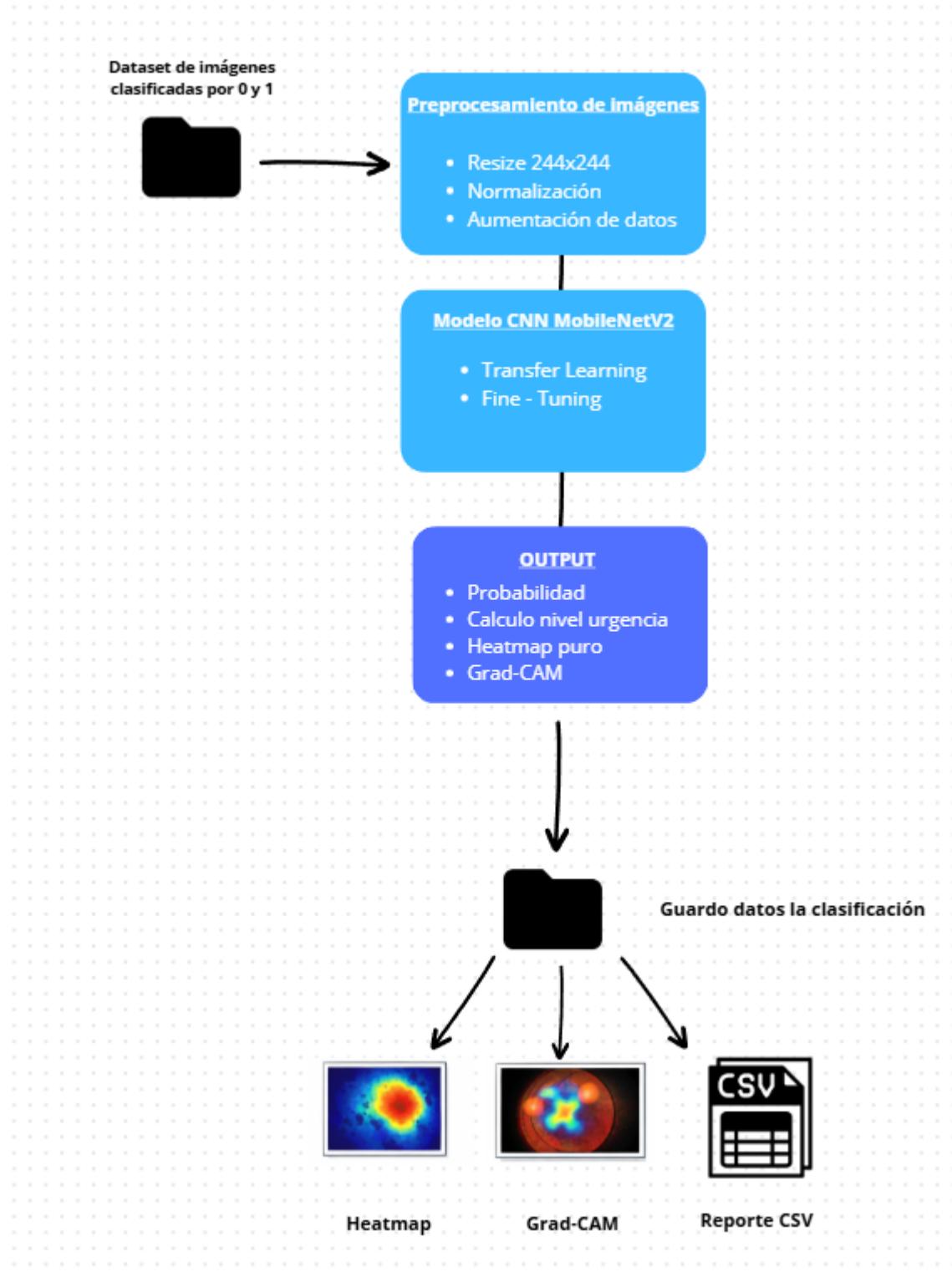
### Estrategia de Entrenamiento en Dos Fases.

Para maximizar el rendimiento, se implementó una estrategia de entrenamiento de dos fases que combina el Transfer Learning con el Fine-Tuning.

- **Fase 1:** Entrenamiento Inicial (Transfer Learning) En esta fase, las capas convolucionales del backbone de MobileNetV2 (específicamente, las primeras 60 capas) se mantuvieron congeladas, entrenando únicamente el cabezal de clasificación personalizado. Esto permite que el nuevo cabezal aprenda a interpretar las características extraídas por la red base sin alterar los pesos preentrenados. Se utilizó el optimizador Adam con una tasa de aprendizaje (learning rate) de 1e-3 y la función de pérdida Binary Crossentropy.

• **Fase 2:** Ajuste Fino (Fine-Tuning) Una vez que el cabezal de clasificación convergió, se procedió a descongelar las últimas 20 capas del modelo base. Se redujo la tasa de aprendizaje del optimizador Adam a 1e-4 para realizar ajustes sutiles y evitar la pérdida del conocimiento previamente adquirido. Además, se cambió la función de pérdida a Focal Loss ( $\gamma=2$ ,  $\alpha=0.25$ ) para enfocar el entrenamiento en los ejemplos más difíciles.

Esta metodología estructurada sienta las bases para la implementación de técnicas más específicas, las cuales se detallan a continuación, orientadas a optimizar aún más el rendimiento y la fiabilidad del modelo.



### 3.3.Técnicas avanzadas implementadas

Para superar desafíos comunes en el modelado de datos médicos, como el sobreajuste, el desbalance de clases y la necesidad de interpretabilidad, se implementaron un conjunto de técnicas avanzadas. Estas estrategias fueron cruciales para transformar un modelo base en un sistema robusto y clínicamente útil.

#### Mitigación de Desbalance de Clases

El desbalance entre el número de imágenes de pacientes sanos y pacientes con glaucoma es un problema recurrente. Para abordarlo, se combinaron dos estrategias complementarias:

**1. Pesos de Clase (class\_weight):** Durante la fase inicial de entrenamiento, se asignó un mayor peso a la clase minoritaria (glaucoma) para penalizar más fuertemente los errores en su clasificación. Los pesos calculados fueron:

- Clase 0 (Sin glaucoma): **0.814**
- Clase 1 (Con glaucoma): **1.295**

**2. Función de Pérdida Focal (Focal Loss):** En la fase de fine-tuning, se adoptó la función Focal Loss. Esta variante de la entropía cruzada binaria reduce la contribución a la pérdida de los ejemplos clasificados correctamente y con alta confianza, permitiendo que el modelo se enfoque en aprender de los casos más difíciles o ambiguos.

#### Estrategias de Regularización

Para prevenir el sobreajuste y mejorar la capacidad de generalización del modelo, se implementó un enfoque de regularización múltiple:

- **Regularización L2:** Se aplicó una penalización L2 con un factor lambda ( $\lambda$ ) de 0.001 a las capas densas del cabezal de clasificación, lo que ayuda a mantener los pesos del modelo pequeños y reduce la complejidad.
- **Dropout:** Se introdujeron dos capas de Dropout con ratios del 60% y 50% en el cabezal. Esta técnica desactiva aleatoriamente neuronas durante el entrenamiento, forzando a la red a aprender representaciones más robustas y menos dependientes de neuronas individuales.
- **Batch Normalization:** La arquitectura MobileNetV2 integra de forma nativa capas de Batch Normalization, que normalizan las activaciones de las capas intermedias, estabilizando y acelerando el proceso de entrenamiento.

#### Callbacks para la Optimización del Entrenamiento

Se utilizaron callbacks avanzados de Keras para monitorizar y controlar el proceso de entrenamiento de manera dinámica, asegurando que se obtuviera el mejor modelo posible sin intervención manual:

- **EarlyStopping:** Monitorizaba la métrica val\_auc (Área Bajo la Curva ROC en el conjunto de validación) con una paciencia de 10 épocas (patience=10) y en modo de maximización (mode='max'). Si esta métrica no mejoraba, el entrenamiento se detenía automáticamente, restaurando los pesos del mejor epoch para evitar el sobreajuste.
- **ModelCheckpoint:** Guardaba únicamente la mejor versión del modelo en disco, basándose también en el rendimiento de val\_auc. Esto garantiza que el artefacto final corresponda al punto óptimo del entrenamiento.
- **ReduceLROnPlateau:** Reducía dinámicamente la tasa de aprendizaje en un factor de 0.2 (factor=0.2) si la pérdida en el conjunto de validación (val\_loss) se estancaba por 7

épocas (patience=7). Esta técnica permite al modelo escapar de mínimos locales y converger de manera más precisa.

### Módulo de interpretabilidad (Grad-CAM)

Para proporcionar la interpretabilidad clínica necesaria, se implementó el algoritmo Grad-CAM (Gradient-weighted Class Activation Mapping). Este módulo fue diseñado para generar explicaciones visuales sobre la activación de la red en cada predicción.

- **Capa Objetivo:** La activación de la última capa convolucional, denominada Conv\_1, se utilizó para generar el mapa de calor.
- **Visualización:** El mapa de calor se generó con un colormap Jet y se superpuso a la imagen original mediante Alpha blending con un factor de 0.4.
- **Utilidad:** Este análisis permite la identificación de regiones anatómicas relevantes en la toma de decisiones del modelo.

### **3.4.Cálculo del “nivel de urgencia”**

#### Fórmula y justificación

El sistema calcula un Nivel de urgencia para reflejar no solo la probabilidad de glaucoma, sino qué tan extendida es la zona activa detectada en el ojo,. La fórmula utilizada es:

$$\text{Nivel de urgencia} = (\text{Probabilidad Glaucoma} + \text{Porcentaje de zona activa}) / 2$$

La Probabilidad de glaucoma es el valor entre 0 y 1 entregado por el modelo de deep learning, y el Porcentaje de zona activa es la proporción del área del ojo resaltada por Grad-CAM. La fórmula combina la certeza del modelo con el impacto anatómico.

#### Justificación y Limitaciones

- **Ventajas:** Es fácil de interpretar, mantiene el rango entre 0 y 1, y combina la certeza del modelo con el impacto anatómico.

- **Limitaciones:** La fórmula asigna un peso igual para la probabilidad y el área, lo cual no siempre es equivalente en la práctica clínica. Por ejemplo, un área pequeña localizada en la papila óptica podría ser clínicamente más importante. Además, el Grad-CAM es una aproximación y no considera la localización precisa de la zona activa (no es lo mismo en el borde periférico que en el disco óptico).

### Mejoras potenciales

Se sugiere la ponderación diferente (ejemplo:  $U=0.7\times P+0.3\times A$ ) o el uso de una función no lineal (ejemplo:  $U=P\times A$ ) que aumente la urgencia solo cuando tanto la probabilidad como el área activa son altas.

## 4. Resultados.

### 4.1. Métricas de desempeño (Accuracy, AUC, sensibilidad, especificidad, matriz de confusión)

El modelo MobileNetV2, optimizado mediante fine-tuning y el uso de Focal Loss ( $\gamma=2, \alpha=0.25$ ), alcanzó métricas de desempeño que son clínicamente aceptables para un sistema de detección temprana de glaucoma (superando el objetivo de >90% accuracy),.

Los resultados finales obtenidos por el modelo Fine-Tuned en las métricas clave de clasificación binaria fueron:

Métrica	Valor
Accuracy (Precisión Global)	93%
AUC (Area Under the Curve)	0.989
Recall (Sensibilidad)	96%
Precision	87%
F1-Score	91%
Especificidad	91%

El valor de AUC de 0.989 indica una excelente capacidad del modelo para la separación entre clases (Glaucoma/Normal). La métrica de Sensibilidad (Recall) del 96% es crucial, ya que se considera "Excelente para detectar casos positivos" y demuestra que el fine-tuning logró su objetivo de reducir significativamente los Falsos Negativos (FN).

La Matriz de Confusión del modelo final fue la siguiente, basada en el conjunto de datos de prueba:

	Predictión Normal	Predictión Glaucoma
Actual Normal	3226	313
Actual Glaucoma	91	2117

El análisis de la matriz revela dos puntos clave:

- **Una alta sensibilidad (Recall) del 96% ( $2117 / (2117 + 91)$ )**, lo que indica que el modelo identifica correctamente la gran mayoría de los casos de glaucoma.
- **Una especificidad del 91.1% ( $3226 / (3226 + 313)$ )**, demostrando una sólida capacidad para identificar correctamente a los pacientes sanos.

Clínicamente, el resultado más importante es el bajo número de falsos negativos (91 casos). Aunque el objetivo es minimizar este valor, el modelo ha logrado reducirlo significativamente, un factor crucial para una herramienta de tamizaje.

#### Impacto del Fine-Tuning en el Rendimiento

La estrategia de re-entrenamiento con fine-tuning demostró ser fundamental para pulir el rendimiento del modelo. Comparando las métricas antes y después de esta fase, se observaron mejoras sustanciales:

- **Precisión (clase 1 - Glaucoma):** Aumento del 3%, pasando de 0.84 a 0.87.
- **Recall:** Aumento del 2% para la clase 0 (Normal) y del 1% para la clase 1 (Glaucoma).

- **F1-Score:** Aumento del 2% para ambas clases, indicando un mejor equilibrio entre precisión y sensibilidad.

Estas mejoras demuestran que permitir que el modelo ajuste finamente sus capas convolucionales más profundas fue una estrategia efectiva para especializarlo en las características sutiles de las imágenes de fondo de ojo.

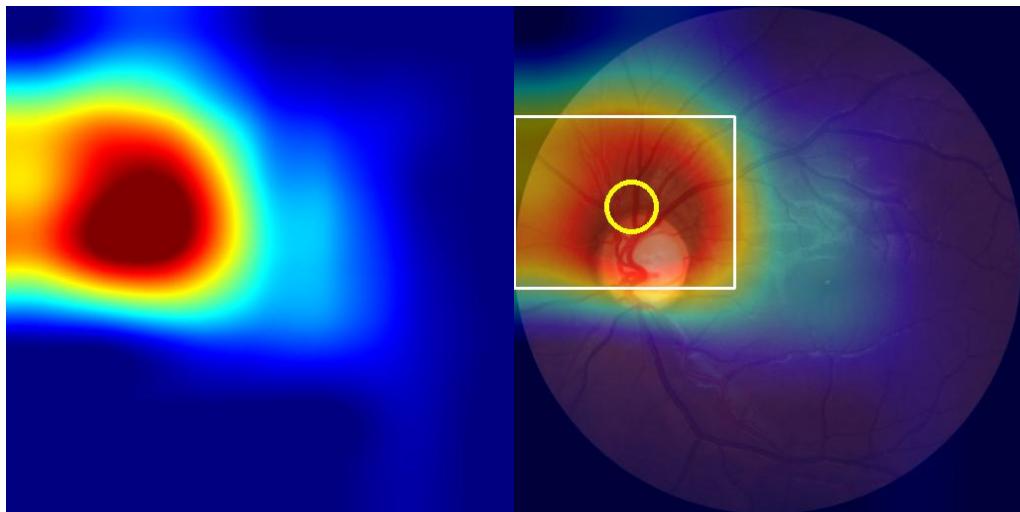
La solidez de estos resultados proporciona una base sólida para discutir las implicaciones del modelo en un contexto clínico real.

#### **4.2. Análisis de interpretabilidad (mapas Grad-CAM, ejemplos)**

Se implementó el algoritmo Grad-CAM (Gradient-weighted Class Activation Mapping) para generar explicaciones visuales de las predicciones del modelo, abordando la naturaleza de "caja negra" del Deep Learning. El sistema fue diseñado para que no solo clasifique, sino que también brinde explicaciones visuales para su interpretación clínica.

- **Proceso de Generación:** Los mapas de calor se generaron utilizando la última capa convolucional, denominada Conv\_1, como capa objetivo. La visualización se realizó superponiendo el colormap Jet sobre la imagen original mediante Alpha blending con un factor de 0.4.

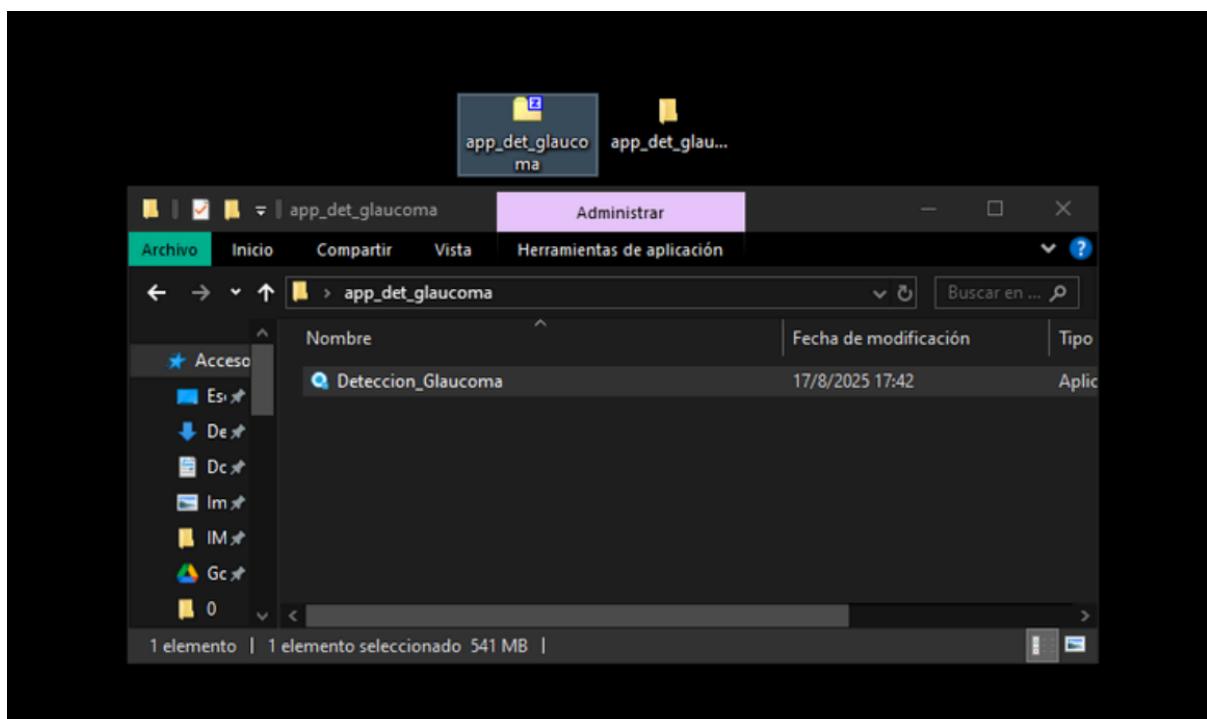
- **Interpretación Clínica:** El análisis de las activaciones (Grad-CAM) permitió la identificación de regiones anatómicas relevantes en la toma de decisiones de la red. La interpretación clínica de estos mapas ha demostrado que la red neuronal tiende a focalizarse en la capa de fibras nerviosas y el disco óptico al tomar su decisión, que son las estructuras anatómicas clave afectadas por el glaucoma.



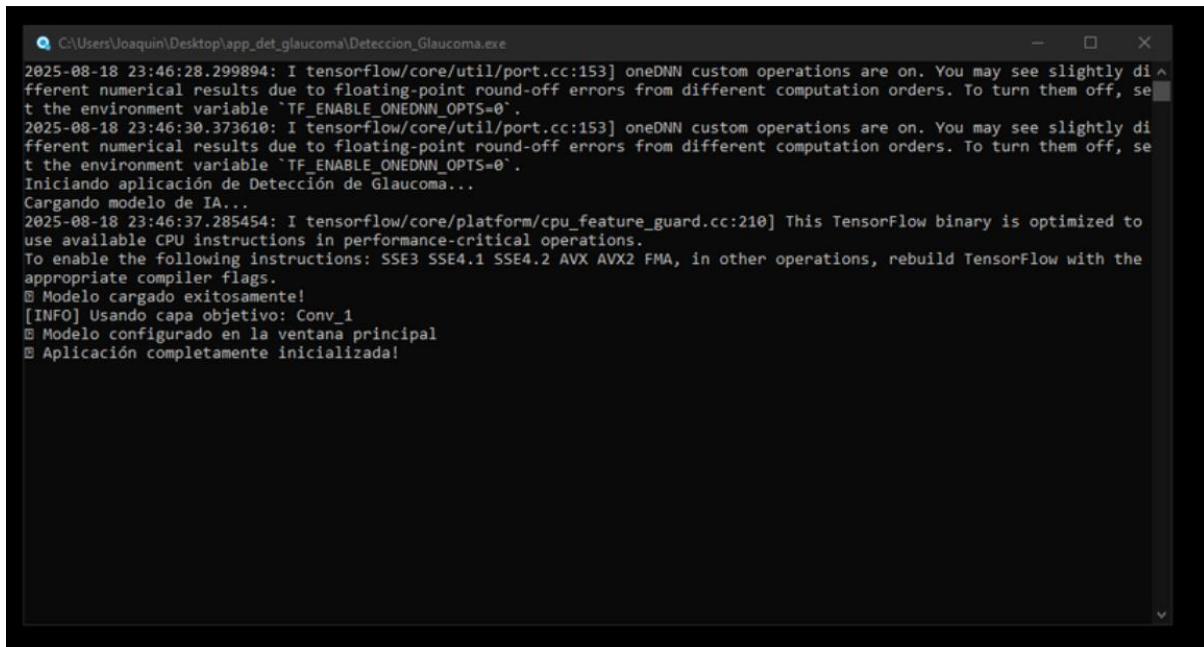
## 5. Aplicación de escritorio

### 5.1. Arquitectura interna y funcionamiento.

La APP\_DET\_GLAUCOMA es un software de escritorio desarrollado en Python, PySide6 y TensorFlow/Keras. El software es portable y no requiere instalación; se ejecuta mediante el archivo Deteccion\_Glaucoma.exe.



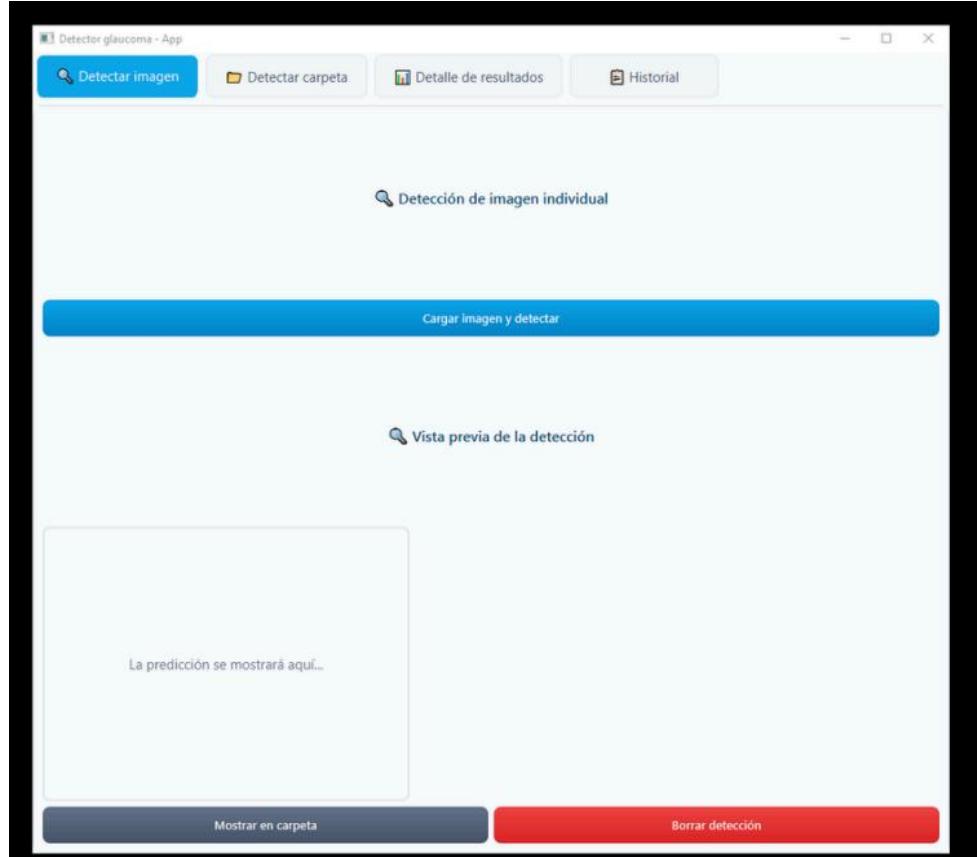
Al ejecutarse, una consola muestra el proceso de carga del modelo de IA, indicando que se usa la capa objetivo Conv\_1 para Grad-CAM.



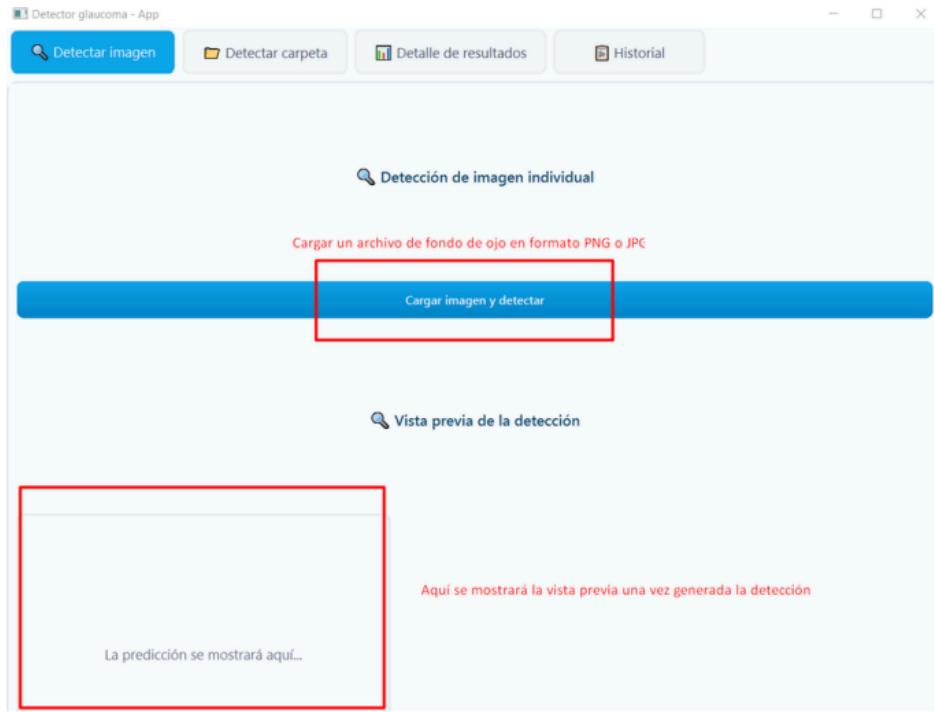
```
C:\Users\Joaquin\Desktop\app_det_glaucoma\Detencion_Glaucoma.exe
2025-08-18 23:46:28.299894: I tensorflow/core/util/port.cc:153] oneDNN custom operations are on. You may see slightly different numerical results due to floating-point round-off errors from different computation orders. To turn them off, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
2025-08-18 23:46:30.373610: I tensorflow/core/util/port.cc:153] oneDNN custom operations are on. You may see slightly different numerical results due to floating-point round-off errors from different computation orders. To turn them off, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
Iniciando aplicación de Detección de Glaucom...
Cargando modelo de IA...
2025-08-18 23:46:37.285454: I tensorflow/core/platform/cpu_feature_guard.cc:210] This TensorFlow binary is optimized to use available CPU instructions in performance-critical operations.
To enable the following instructions: SSE3 SSE4.1 SSE4.2 AVX AVX2 FMA, in other operations, rebuild TensorFlow with the appropriate compiler flags.
Modelo cargado exitosamente!
[INFO] Usando capa objetivo: Conv_1
Modelo configurado en la ventana principal
Aplicación completamente inicializada!
```

## 5.2. Paneles principales

La aplicación cuenta con una interfaz sencilla que ofrece tres paneles principales.

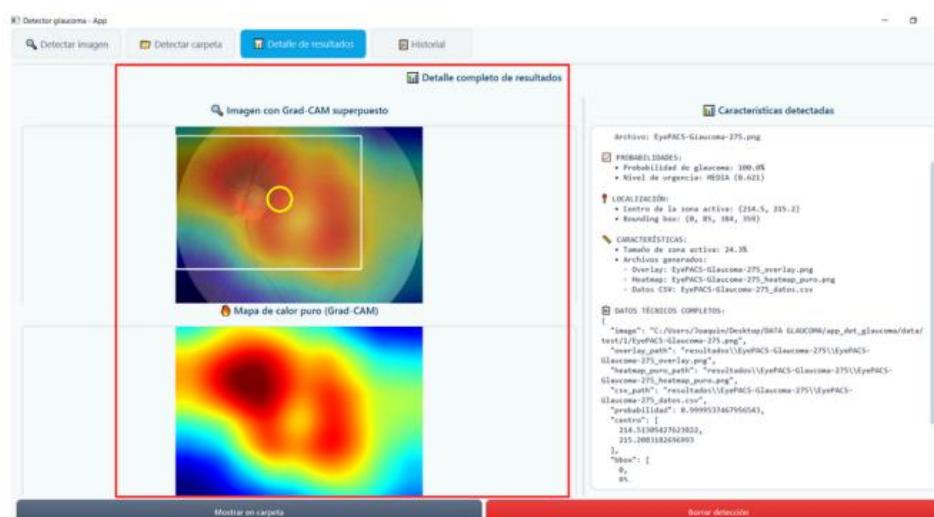


**1. Detectar Imagen:** Permite la carga de un archivo individual (PNG o JPG) y la visualización de una vista previa de la detección.

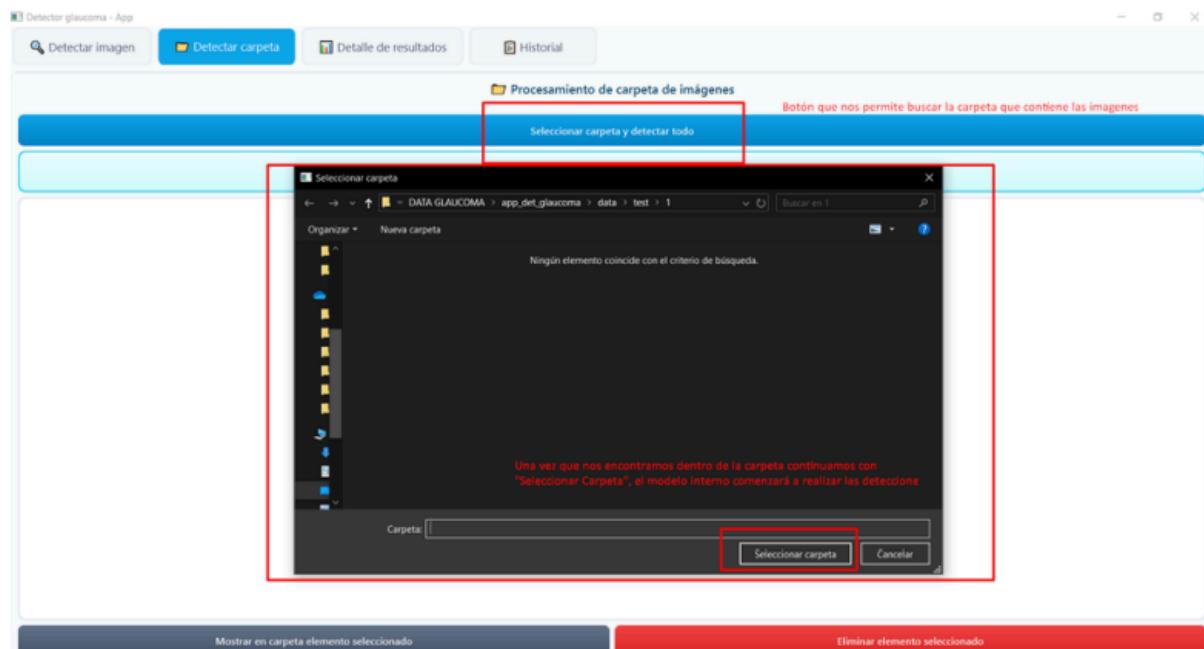


**2. Detalle de resultados:** Muestra la información completa de la detección:

- Visualización de Grad-CAM superpuesto y mapa de calor puro.
- Características detectadas: Nombre del archivo, Probabilidad de glaucoma, Nivel de urgencia, localización (centro y bounding box), tamaño de la zona activa y datos técnicos completos.



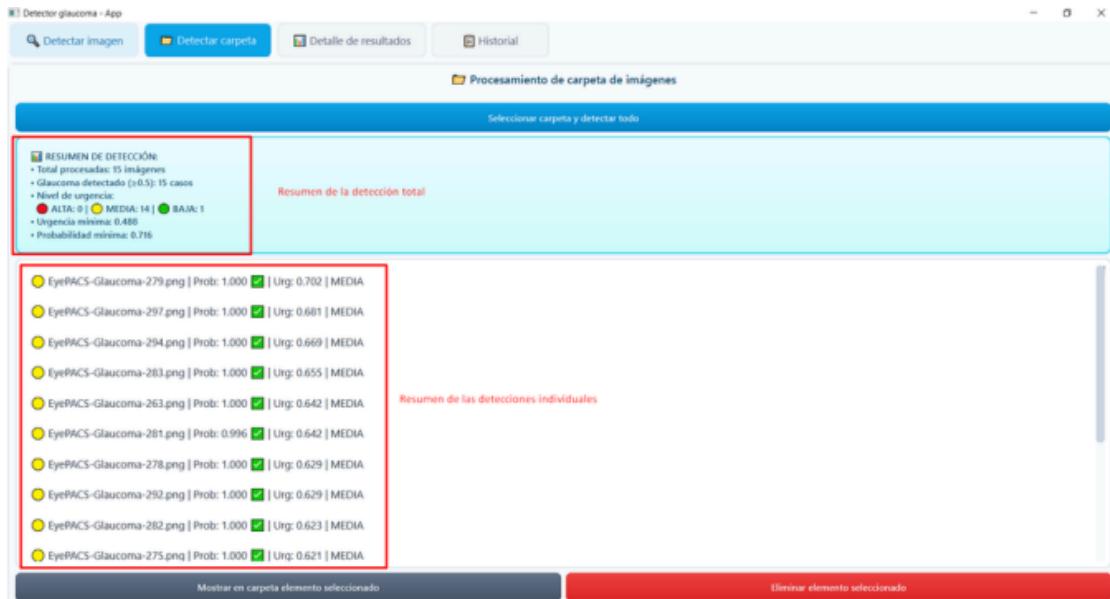
**3. Detectar Carpeta:** Permite el procesamiento por lotes de múltiples imágenes contenidas en una carpeta. Al finalizar, muestra un resumen de la detección total (imágenes



procesadas, casos de glaucoma detectados y el recuento de los niveles de urgencia ALTA, MEDIA y BAJA), y un resumen individual de cada imagen.

**4. Historial:** Permite visualizar todas las detecciones realizadas. Se puede ordenar la lista por fecha o nivel de urgencia, y filtrar por nivel de urgencia (Baja, Media, Alta).

Exportación de resultados: Los resultados de la detección (archivos de Overlay, Heatmap y CSV con los datos técnicos completos) se guardan automáticamente en la carpeta "resultados".



### 5.3. Proyección de uso clínico

El software tiene fines educativos y de investigación, y no reemplaza el diagnóstico de un especialista médico. Su rol es de herramienta complementaria y de apoyo a la detección asistida. La IA puede colaborar con el trabajo de un especialista, optimizando el tiempo de detección, indicando las zonas donde prestar más atención y listando los casos más urgentes, descartando la hipótesis de que la detección puede ser tercerizada al 100% a una IA.

## 6. Discusión, conclusiones y trabajo futuro

Esta sección final tiene como propósito interpretar el significado de los resultados obtenidos, evaluar la hipótesis inicial que motivó el proyecto y delinear las implicaciones prácticas del sistema, así como las futuras líneas de trabajo para su mejora continua.

## **6.1. Interpretación y Conclusión sobre la Hipótesis**

El proyecto comenzó con la hipótesis de que la detección de glaucoma podría ser tercerizada al 100% a una IA. Los hallazgos demuestran que el modelo de Deep Learning desarrollado es capaz de detectar glaucoma en imágenes de fondo de ojo con una excelente precisión general (93%) y un poder de discriminación sobresaliente (AUC 0.989). Sin embargo, la hipótesis inicial del proyecto, que planteaba la posibilidad de "tercerizar al 100% a una IA la detección de glaucoma", debe ser descartada.

La razón principal para esta conclusión radica en la existencia de un pequeño, pero clínicamente inaceptable, porcentaje de falsos negativos. En la práctica médica, omitir el diagnóstico de un paciente con glaucoma, aunque sea en un bajo número de casos, tiene consecuencias graves debido a la naturaleza irreversible de la enfermedad. Por lo tanto, el modelo no puede operar de forma totalmente autónoma como un sustituto del diagnóstico humano.

## **6.2. El rol del modelo como herramienta de apoyo clínico**

Aunque la automatización total no es viable, el sistema desarrollado se posiciona como una potente herramienta de colaboración y apoyo para el especialista. Su rol práctico no es reemplazar, sino aumentar las capacidades del oftalmólogo. Sus aplicaciones más directas incluyen:

- Optimizar el tiempo de detección en programas de screening masivo, permitiendo analizar grandes volúmenes de imágenes de manera rápida y eficiente.
- Indicar visualmente las zonas de la imagen que requieren mayor atención a través de los mapas de calor de Grad-CAM, guiando el ojo del clínico hacia posibles anomalías.

- Ayudar a priorizar los casos potencialmente más urgentes, permitiendo que los especialistas enfoquen sus recursos en los pacientes que presentan una mayor probabilidad de padecer la enfermedad.

### **6.3.Limitaciones del Proyecto y Líneas de Trabajo Futuro**

Durante el desarrollo se encontraron ciertas limitaciones. Las dos más significativas fueron la dificultad para obtener grandes volúmenes de datos clínicos de alta calidad debido a su naturaleza sensible, y la falta de hardware especializado (GPU), lo que ralentizó considerablemente los ciclos de entrenamiento e iteración en la búsqueda de hiperparámetros óptimos.

Con base en los logros y las limitaciones identificadas, se proponen los siguientes pasos para la evolución del proyecto:

- 1. Validación Clínica:** El siguiente paso crucial es realizar estudios formales en colaboración con oftalmólogos para validar el rendimiento del modelo en un entorno clínico real y compararlo con los estándares de diagnóstico actuales.
- 2. Optimización Continua:** Enfocar futuros esfuerzos de modelado en la reducción adicional del número de falsos negativos, explorando arquitecturas más avanzadas o técnicas de entrenamiento novedosas.
- 3. Despliegue:** El despliegue del modelo se ha validado exitosamente mediante su integración en una aplicación de escritorio funcional (APP\_DET\_GLAUCOMA), demostrando la viabilidad de su implementación práctica y sentando las bases para futuras adaptaciones a plataformas web o móviles.

## 7. Referencias

<https://github.com/joarodev/glaucoma-detection-dl/tree/main>

[https://github.com/joarodev/glaucoma-detection-dl/blob/main/Proyecto\\_detecci%C3%B3n\\_glaucoma.pdf](https://github.com/joarodev/glaucoma-detection-dl/blob/main/Proyecto_detecci%C3%B3n_glaucoma.pdf)

[https://github.com/joarodev/app\\_det\\_glaucoma](https://github.com/joarodev/app_det_glaucoma)

[https://github.com/joarodev/app\\_det\\_glaucoma/blob/main/manual\\_de\\_uso\\_app\\_det\\_glaucoma\\_V1.pdf](https://github.com/joarodev/app_det_glaucoma/blob/main/manual_de_uso_app_det_glaucoma_V1.pdf)