VISIÓN ARTIFICIAL PARA PATOLOGÍAS OCULARES



DANIEL MAURICIO ORDOÑEZ SAAVEDRA

PONTIFICIA UNIVERSIDAD JAVERIANA

TUTORA: ANDREA DEL PILAR RUEDA OLARTE

FACULTAD DE CIENCIAS

CARRERA DE CIENCIA DE DATOS

BOGOTÁ D.C

Agradecimientos:

Quiero expresar mi más sincero agradecimiento a todas las personas que me apoyaron a lo largo de este proceso, en especial a mis amigos, quienes fueron un pilar fundamental desde el inicio de mi formación académica.

Asimismo, extiendo un profundo agradecimiento a mi tutora de trabajo de grado, la docente Andrea Rueda Olarte, por su invaluable guía, dedicación y compromiso constantes, los cuales fueron esenciales para la correcta culminación de cada etapa de este proyecto.

Tabla de contenidos

1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA 8

2. OBJETIVOS DEL PROBLEMA 9

2.1. Objetivo general 9

2.2. Objetivos especificos 9

3. MARCO TEÓRICO 10

3.1. Evaluación del avance en técnicas y rendimientos de modelos para la detección de diferentes retinopatías. 10

3.1.1. Conclusión del Estudio 12

3.1.2. Limitaciones de los Algoritmos 12

3.2. Análisis automatizado de imagines de fondo de ojo para el diagnóstico de enfermedades retinianas: una revisión. 13

3.2.1. Glaucoma 13

3.2.2. Retinopatia diabética 14

3.2.3. Degeneración macular senil – *DMS* 15

3.2.4. Cataratas 17

3.3. Mobile-HR: un Sistema de clasificación oftalmológico para el diagnóstico de la retinopatía hipertensiva usando la arquitectura MobileNet optimizada 18

3.3.1. Convolución en profundidad y convolución puntual 19

3.3.2. Bloques densos: 20

3.3.3. MobileNet: 21

3.3.4. Recolección de datos 22

3.3.5. Pre procesamiento y aumento de datos 22

3.3.6. Implementación del modelo 22

3.3.7. Evaluación del modelo 22

3.3.8. Resultados y recomendaciones 23

3.4. Predicción retiniana a través de la segmentación y clasificación de vasos sanguíneos usando aproximación basadas en ensamble y aprendizaje profundo. 23

3.4.1. Arquitecturas usadas en el ensamble final 24

3.5. Estudios secundarios relacionados 27

4. METODOLOGÍA PROPUESTA 31

4.1. Protocolo de recolección de imágenes y pre procesamiento de imágenes 32

4.2. Protocolo de pos procesamiento 41

4.2.1. Segmentación de disco y copa óptica 41

4.2.2. Segmentación de micro aneurismas: 49
4.2.3. Segmentación de drusas: 53
4.3. Construcción del modelo discriminativo final 56
4.3.1. Concatenación de mapas de segmentación 56
4.3.2. Construcción de arquitectura de clasificación compuesta 57
4.4. Diseño de sistema de diagnóstico final 60
4.4.1. Mapeo de características con Grad-CAM 60
4.4.2. Grad-CAM ++ 61
4.4.3. Sintomatología relacionada con retinopatía diabética y degeneración macular senil 63
4.5. Flujo final ilustrado 64
5. RESULTADOS 64
5.1. Insumos a nivel de software utilizados 64
5.2. Filtrado de documentos y selección de muestras 65
5.3. Preprocesamiento para segmentación de disco, y copa óptica 71
5.4. Preprocesamiento de drusas 73
5.5. Preprocesamiento para segmentación de micro aneurismas 73
5.6. Métricas de evaluación para el preprocesamiento 74
5.7. Resultados obtenidos después de la aplicación de las respectivas técnicas de preprocesamiento: 77
5.8. Evaluación de resultados en el post procesamiento 77
5.9. Segmentación de disco y copa óptica 81
5.10. Segmentación de posibles Drusas 83
5.11. Segmentación de posibles micro aneurismas 85
5.12. Evaluación de modelo de clasificación de ensamble 86
5.13. Aplicación de métodos de explicabilidad 88
5.13.1. Mapa de activación para glaucoma 89
5.13.2. Mapa de activación para retinopatía diabética 90
5.13.3. Mapa de activación para degeneración macular senil 91
5.14. Escenarios multietiqueta 91
5.15. Sensibilidad del modelo 93
6. DISCUSIÓN Y CONCLUSIONES 95
7. REFERENCIAS 98

Índice de figuras

Figura 1-1 Retinopatía hipertensiva Figura 1-2 Retinopatía diabética 9
Figura 3-1 modificación de la resolución de la imagen y segmentación de la imagen. Fecha de acceso: 4 de noviembre de 2024. Tomado de: http://arxiv.org/abs/1905.02544 12
Figura 3-2 Diferencias entre el diámetro del disco y la copa óptica para un ojo en empeoramiento progresivo. Tomado de: [17] 14
Figura 3-3 Área de interés para la detección de glaucoma. Tomado de: [17] 14
Figura 3-4 Lesiones más importantes detectadas en ojos con diabetes. Tomado de: [17] 15
Figura 3-5 Evolución de DMS. Tomado de: [17] 17
Figura 3-6 Evolución de cataratas. Tomado de: [17] 18
Figura 3-7 Características de una retinografía hipertensiva. Tomado de: [18] 19
Figura 3-8 Reducción de dimensionalidad generada por depthwise convolution. Fecha de acceso: 4 de noviembre de 2024.Tomado de: https://arxiv.org/abs/1704.04861. 20
Figura 3-9 Ilustración de la red final. Tomado de: [18] 21
Figura 3-10 Flujo final de la operación investigativa. Tomado de: [18] 22
Figura 3-11 Imagen antes y después del pre procesamiento. Tomado de: [19] 25
Figura 3-12 Curvas de aprendizaje del modelo propuesto. Tomado de: [19] 26
Figura 3-13 Segmentaciones obtenidas para las diferentes clases de enfermedades. Tomado de: [19] 26
Figura 3-14 Flujo del proceso de investigación desde la adquisición de los datos. Tomado de: [19] 27
Figura 4-1 Filtro de mediana para imagen comparado con filtro promedio para imagen. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www.youtube.com/watch?v=TYNbTdzm-5w 38
Figura 4-2 Comparación del histograma usando AHE vs CLAHE. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://en.wikipedia.org/wiki/Adaptive_histogram_equalization 39
Figura 4-3 Comparación de imagen procesada usando AHE vs CLAHE. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www-geeksforgeeks-org.translate.goog/clahe-histogram-eqalization-opencv/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc 39
Figura 4-4 Imagen de fondo de ojo recortada con su respectiva mascara para la identificación de micro aneurismas. Elaboración propia 41
Figura 4-5 Representación de recorte por parches con solapamiento. Fecha de acceso. 21 de abril de 2025. Tomado de [72] 41
Figura 4-6 Bloque de bottleneck de mobileNetv2. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://blog.paperspace.com/transfer-learning-explained-2/ 43
Figura 4-7 Comparativa entre tipos de convolución de imagen. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://arxiv.org/abs/1801.04381 43
Figura 4-8 Ilustración la adición de un bloque squeeze and excite a la arquitectura. Tomado de [70] 44
Figura 5-3 Representación de encuadre generado por MobileNetV2 con SSD. Fecha de acceso: 4 de noviembre de 24 Tomado de: https://github.com/saunack/MobileNetv2-SSD/blob/master/model.ipynb 45
Figura 4-9 Representación gráfica de la arquitectura U-NET. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://www.youtube.com/watch?v=NhdzGfB1q74 47
Figura 4-10 Arquitectura convolucional propuesta por Kou et al. Fecha de acceso: 4 de noviembre de 2024. Tomado de: [66] 50

Figura 4-11 Ilustración de bloque RRCNN con capa RCL. Fecha de acceso. 21 de abril de 2025. Tomado de: [66]. 51
Figura 4-12 Ilustración del proceso para la segmentación de drusas. Tomado de [65] 53
Figura 4-13 Arquitectura deeplabv3. Fecha de acceso: 8 de abril de 25 Tomado de: https://learnopencv.com/deeplabv3-ultimate-guide/. 54
Figura 4-14 Arquitectura deeplabv3+ con la adición del módulo decoder. Fecha de acceso: 21 de abril de 2025. Tomado de: [75]. 55
Figura 4-15 Imagen de fondo de ojo concatenada. Elaboración propia 56
Figura 4-16 Arquitectura VGG 19 ilustrada. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www.researchgate.net/figure/llustration-of-the-network-architecture-of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356 58
Figura 4-17 Ilustración de un bloque convolucional para redes residuales. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://es.wikipedia.org/wiki/Red_neuronal_residual 58
Figura 4-18 Metodología de escalado ilustrada. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://huggingface.co/google/efficientnet-b0 59
Figura 4-19 Mapeo de clases con Grad Cam Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://towardsdatascience.com/understand-your-algorithm-with-grad-cam-d3b62fce353 61
Figura 4-20 Comparación de un mapa de características Grad-CAM con respecto a un Grad-CAM++. Tomado de [74] 63
Figura 4-21 Flujo final ilustrado: Elaboración propia 64
Figura 5-1 conducto de pre procesamiento inicial para las imágenes en escala de grises. Elaboración propia. 72
Figura 5-2 Parche de fondo de ojo con drusas y parche de mascara con área de interés resaltada. Elaboración propia 73
Figura 5-4 Ejemplo de encuadre del modelo sobre el área de interés. Elaboración propia 82
Figura 5-5 Ejemplo de imagen encuadrada y segmentada para disco y copa óptica. Elaboración propia. 83
Figura 5-6 Ejemplo de parche de imagen con las drusas segmentadas. Elaboración propia. 85
Figura 5-7 Ejemplo de parche de imagen con micro aneurismas segmentados.: Elaboración propia 86
Figura 5-8 Mapa de calor sobre imagen de fondo de ojo con glaucoma. Elaboración propia 89
Figura 5-9 Mapa de calor sobre imagen de fondo de ojo con retinopatía diabética. Elaboración propia 90
Figura 5-10 Mapas de activación con gradCAM++ para una imagen de fondo de ojo con degeneración macular senil. Elaboración propia. 91
Figura 5-11 Imagen ensamble de fondo de ojo con presencia de glaucoma y degeneración macula. Elaboración propia. 92
Figura 5-12 Imagen ensamble de fondo de ojo con presencia de glaucoma y retinopatía diabética. Elaboración propia. 92
Figura 5-13 Imagen ensamble de fondo de ojo con presencia de retinopatía diabética y degeneración macular. Elaboración propia. 93
Figura 5-14 Ojo con glaucoma erróneamente clasificado como normal. Elaboración propia. 94
Figura 5-15 Ojo normal erróneamente clasificado como afectado por glaucoma. Elaboración propia. 94

Índice de Tablas

Tabla 1 Estudios secundarios referenciados 31
Tabla 2 Datasets revisados 37
Tabla 3 Descripción de modificaciones a los dataset 70
Tabla 4 Rendimiento de las técnicas de preprocesamiento aplicadas 77
Tabla 5 Valores promedio en las métricas de encuadre de disco y copa óptica para los conjuntos de entrenamiento y validación 81
Tabla 6 Resultados de Validación por folds para modelo de encuadre de disco y copa óptica en el conjunto de datos de validación 81
Tabla 7 Resultados de Validación por folds para modelo de encuadre de disco y copa óptica en el conjunto de datos de prueba 82
Tabla 8 Valores promedio en las métricas para el modelo de segmentación de disco y copa optica en diferentes conjuntos de datos 83
Tabla 9 Resultados de Validación por folds para modelo de segmentación de disco y copa óptica en el conjunto de datos de validación 83
Tabla 10 Resultados de Validación por folds para modelo de segmentación de disco y copa óptica en el conjunto de datos de prueba 83
Tabla 11 Valores promedio en las métricas para el modelo de segmentación de drusas en diferentes conjuntos de datos 84
Tabla 12 Resultados de Validación por folds para modelo de segmentación de drusas en el conjunto de datos de validación 84
Tabla 13 Resultados de Validación por folds para modelo de segmentación de drusas en el conjunto de datos de prueba 84
Tabla 14 Valores promedio en las métricas para el modelo de segmentación de microaneurismas en diferentes conjuntos de datos 85
Tabla 15 Resultados de Validación por folds para modelo de segmentación de microaneurismas en el conjunto de datos de validación 85
Tabla 16 Resultados de Validación por folds para modelo de segmentación de microaneurismas en el conjunto de datos de prueba 86
Tabla 17 Valores promedio en las métricas para el modelo de clasificación de ensamble en el conjunto de entrenamiento 87
Tabla 18 Valores promedio en las métricas para el modelo de clasificación de ensamble en el conjunto de prueba 88
Tabla 19 Resultados de Validación por folds para modelo de clasificación de ensamble en el conjunto de datos de prueba 88

1. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

El diagnóstico temprano de patologías oculares es una de las tareas más importantes en el mundo contemporáneo. Según el último informe mundial sobre la visión de la OMS [4] al menos 2.200 millones de personas padecen enfermedades oculares. Estas patologías abarcan una amplia gama de condiciones que afectan no solo al ojo (Cómo es el caso de las retinopatías hipertensivas y diabéticas). En este informe, la OMS estima que hay 2.600 millones de personas con miopía, 146 millones con retinopatía diabética, 196 millones con degeneración macular senil y 76 millones con glaucoma.

Si bien algunas de estas afecciones son más fáciles de detectar sin la ayuda de equipo médico, ciertas patologías como la diabetes y la hipertensión pueden mostrar diferentes y mejores resultados en los escáneres de retinografía dependiendo del análisis de un especialista [3]. Esto se debe a que la retina es uno de los órganos más comprometidos por estas patologías, lo que otorga una mayor importancia a los resultados de estos escáneres. Por estos motivos, a los pacientes con estas enfermedades se les recomienda visitar al especialista de forma anual para revisar el estado de su enfermedad.

Dada la relación entre diversas patologías oculares que afectan mayormente a la retina y a otras partes del ojo [22], es posible comenzar a detectar patrones dentro de los diagnósticos. No obstante, hay ocasiones en que no es tan clara la identificación de estos patrones, ya que, en los estados tempranos, la sintomatología puede no mostrar indicios notorios de una enfermedad que realmente está presente en el paciente. Para superar estos desafíos diagnósticos, se emplean diferentes tipos de escáneres oculares: Tomografía de Coherencia Óptica (OCT), retinografía, ecografía ocular, tomografía computarizada (CT), etc... Entre todos estos métodos, las retinografías juegan un papel especialmente importante. Estas proporcionan imágenes detalladas y de alta resolución de la retina a bajo costo, lo que facilita la detección temprana de patologías antes de que los síntomas se vuelvan evidentes, esto sumado a las faltas de disponibilidad que usualmente se ve sobre las herramientas que utilizan los demás métodos [25], las retinografias marcan la diferencia a la hora de recibir un tratamiento de forma eficaz.

De acuerdo con las proyecciones de la OMS [4], para 2030 se estima que el número de personas con 60 años o más en todo el mundo será de 1.400 millones. En cuanto a patologías, se prevé que la cantidad de personas con glaucoma senil aumentará 1.3 veces entre 2020 y 2030, alcanzando los 95.4 millones de afectados. Además, con las altas tasas de urbanización y los cambios en los estilos de vida laborales, se predice un aumento general en las enfermedades que afectan directamente a la retina y otros órganos. Asimismo, como se mencionó anteriormente, el incremento en la población mayor de 60 años provocará un aumento significativo en los casos de enfermedades degenerativas oculares.



Figura 1-1 Retinopatía hipertensiva

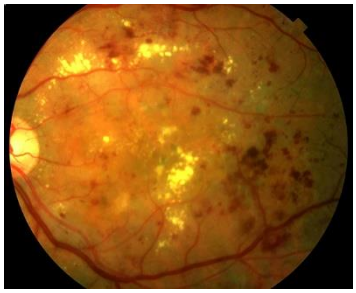


Figura 1-2 Retinopatía diabética

Gracias a los avances en técnicas de procesamiento para predecir estados y patologías en pacientes [17 y 12], parece imprescindible proveer sistemas automatizados que faciliten el proceso de diagnóstico, permitiendo determinar con cierta probabilidad posibles patologías presentes y el estado en que se están desarrollando. Esta tarea se justifica también en la relación ya probada entre la miopía alta y la propensión a ciertas enfermedades como las cataratas [9], generando una necesidad a nivel de monitoreo del paciente.

Para la investigación en curso se hizo énfasis en 3 patologías para ayudar con el proceso de identificación y diagnóstico, estas son: Retinopatía diabética, Degeneración macular senil, y glaucoma. Esto a través del análisis de parámetros en estructuras tal como se realiza en el entorno medico cotidiano.

2. OBJETIVOS DEL PROBLEMA

2.1. Objetivo general

Proponer una metodología de procesamiento de retinografías para facilitar la clasificación de diferentes patologías oculares.

2.2. Objetivos especificos

- Recopilar y caracterizar datos de retinografías que permitan identificar un número considerable de patologías oculares en diferentes estados.
- Explorar posibles algoritmos y técnicas que hayan sido previamente utilizados en el análisis de retinografías.
- Proponer una nueva metodología de preprocesamiento que integre un proceso de extracción de características únicas de cada patología.
- Identificar los efectos de incorporar el nuevo método de preprocesamiento en la tarea de clasificación de retinografías utilizando algoritmos de aprendizaje de máquina.
- Evaluar los resultados del método propuesto en comparación con otros enfoques existentes.

3. MARCO TEÓRICO

Aunque métodos como la Tomografía de Coherencia Óptica (TCO) se consideran actualmente más efectivos para la detección de enfermedades oculares, gracias al análisis microscópico de relieves que ofrecen, enfrentan limitaciones significativas en términos de accesibilidad debido al alto costo de las cámaras y de los propios exámenes [26 y 27]. Este desafío, junto con la creciente literatura sobre la aplicación de técnicas de machine learning y deep learning en la clasificación de imágenes retinianas, sugiere que una investigación exhaustiva en estas metodologías podría abrir nuevas vías de aproximación al problema. Estas aproximaciones podrían facilitar el desarrollo de enfoques innovadores para la detección de patologías oculares en contextos multiclase, desde el monitoreo de los diferentes estados de una patología hasta la detección de múltiples patologías mediante arquitecturas de menor costo. Esto promovería soluciones que mejorarían la accesibilidad para la creación de sistemas de apoyo para los especialistas en el área médica. La implementación de tales soluciones podría reflejarse en futuros censos mundiales, reduciendo la prevalencia de patologías oculares que podrían haberse prevenido.

El propósito final del marco teórico es tomar todos aquellos elementos de otros informes (aun cuando sean otros tipos de examen) para revisar que se puede tomar de la literatura existente desde el punto de vista técnico para su adaptación al contexto actual, estos informes estarán disponibles en la tabla de resultados expuesta al final de la sección.

3.1. Evaluación del avance en técnicas y rendimientos de modelos para la detección de diferentes retinopatías.

Desde la Universidad de París de Gustave Eiffel [12], se publicó un artículo que aborda la problemática de la clasificación multiclase de patologías oculares, considerando que un mismo ojo puede presentar una o más afecciones. En este artículo se realiza una revisión exhaustiva de las diferentes técnicas de preprocesamiento utilizadas en investigaciones previas, y se evalúan según varios factores: patología detectada (una o más), modelo de deep learning aplicado, arquitectura de deep learning usada, conjunto de datos trabajado, rendimiento de la detección y costo computacional (software y hardware).

El estudio revela que, aunque los métodos de preprocesamiento varían según el contexto del proyecto, se identificaron técnicas dominantes en la mayoría de los trabajos. Estas son:

- **Ecualización del histograma:**

  Este método busca mejorar el contraste de una imagen al redistribuir los niveles de intensidad de los pixeles de manera uniforme. Al expandir el rango dinámico de los valores de intensidad, esto hace que las áreas oscuras y claras de

la imagen sean más visibles, lo que es útil dado que, dentro de las imágenes recolectadas, puede que haya zonas en las que el contraste es tan bajo que no permita hacer la correcta identificación de una lesión especifica.

- **Eliminación de ruido en la imagen:**

  Este conjunto de técnicas puede variar dependiendo de la enfermedad que se esté detectando, la aplicación de filtros y modificación del contraste permiten resaltar o eliminar la notoriedad de características dentro de la imagen, asimismo es posible utilizar filtros específicos que, desde la distribución de los pixeles, buscan eliminar variaciones no deseadas que distorsionan la imagen**.**

- **Normalización de balance, color e iluminación local:**

  En este proceso se ajustan los colores y la iluminación de una imagen para que sean uniformes y consistentes. La normalización del balance de blancos corrige tonos que pueden no ser naturales dentro de la imagen, mientras que el manejo de la iluminación permite identificar áreas que pueden estar subexpuestas o sobre expuestas.

- **Aumento de Datos**:

  Se modifican algunas características de los datos existentes (tonalidades, luminancia, tamaño) con el propósito de contar con un conjunto de datos más diverso.

- **Rotación de Imágenes**:

  Se aplica varias veces con ángulos aleatorios para aumentar el tamaño del conjunto de datos.

- **Reflejo o Espejo de Imágenes**:

  Se eligen orientaciones aleatorias o se emplean volteos horizontales y verticales.

En este artículo, se discuten dos métodos principales para procesar imágenes antes de su introducción en la arquitectura de una red neuronal. Uno de estos métodos es el redimensionamiento de la imagen, y el otro es el particionamiento de la retinografía, en donde

la imagen se divide en parches que son luego procesados por la red neuronal.

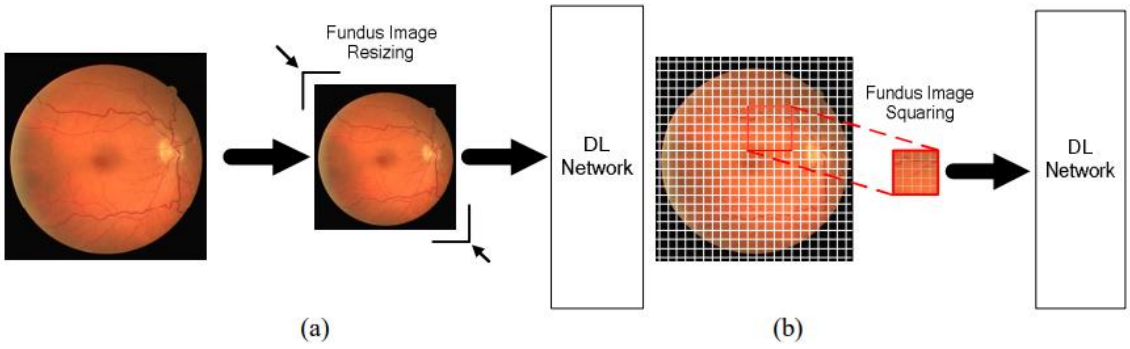


Figura 3-1 modificación de la resolución de la imagen y segmentación de la imagen. Fecha de acceso: 4 de noviembre de 2024. Tomado de: http://arxiv.org/abs/1905.02544

3.1.1. Conclusión del Estudio

El artículo concluye que, aunque los métodos de preprocesamiento son variados, las técnicas de mejora de la imagen y aumento de datos son importantes para mejorar la robustez de los modelos predictivos, aunque son técnicas que se deben manejar con cuidado dada la propensión al sobreajuste y la posibilidad de tener un modelo que no generalice frente al ingreso de nuevos datos.

3.1.2. Limitaciones de los Algoritmos

El estudio también examina varias limitaciones de los algoritmos de deep learning:

- **Pérdida de Características**:

  La reducción de la resolución de las imágenes puede llevar a la pérdida de detalles importantes.

- **Variabilidad en el Tamaño de Entrada**:

  Existe una gran divergencia en los tamaños de las imágenes de entrada, lo que puede afectar la consistencia del rendimiento del modelo.

- **Costo Computacional**:

  Los métodos de deep learning requieren considerable poder computacional y recursos de hardware, lo que puede ser una barrera para su implementación en entornos clínicos con recursos limitados.

3.2. Análisis automatizado de imagines de fondo de ojo para el diagnóstico de enfermedades retinianas: una revisión.

La Sociedad de Ingeniería Biomédica de Brasil [17] llevó a cabo un estudio exhaustivo cuyo objetivo fue presentar el estado del arte hasta el año 2023 en relación con los métodos empleados para el análisis de diversas retinopatías. Este estudio integró perspectivas desde los parámetros médicos, el uso de inteligencia artificial y la aplicación de técnicas de aprendizaje profundo (Deep Learning). El análisis se centró en comprender los desafíos inherentes a estos enfoques en un marco temporal que abarcó dos décadas, desde 2002 hasta 2022.

La investigación incluyó una revisión detallada de las metodologías específicas aplicadas a la clasificación de enfermedades oculares con alta prevalencia mundial. Asimismo, se abordaron los conjuntos de datos utilizados para cada patología, junto con aspectos anatómicos especializados del área de estudio correspondiente a cada enfermedad.

A continuación, se ofrece un resumen de los principales hallazgos obtenidos por el equipo de investigadores en relación con el glaucoma:

3.2.1. Glaucoma

Los dos tipos más comunes son el glaucoma primario de ángulo abierto (GPAA), y el glaucoma primario de ángulo cerrado (GPAC). Mientras que el GPAA implica un drenaje deficiente del humor acuoso a través de la malla trabecular con un ángulo abierto, el GPAC presenta un ángulo cerrado que dificulta aún más el drenaje, aumentando la presión sobre el nervio óptico.

Entre los principales parámetros clínicos se encontró la relación copa-disco (RCD), que evalúa la proporción entre la copa y el disco óptico, siendo un valor superior a 0.3 indicativo de posibles anomalías. Otro parámetro importante es la escala de probabilidad de daño del disco (DDLS por sus siglas en inglés), que cuantifica la progresión del daño en el nervio óptico, dividiendo la enfermedad en 10 etapas. También se utilizan técnicas como la tomografía de coherencia óptica (TCO) y la fotografía del fondo de ojo para detectar cambios patológicos en la capa de fibras nerviosas retinianas (CFNR por sus siglas en inglés), aunque actualmente se tiene una mayor fiabilidad sobre las TCO gracias a su dinamismo sobre la comprensión microscópica del ojo

Los principales métodos de procesamiento de imagen para esta patología se han basado en la extracción del disco óptico como punto de análisis y se ha pasado por la eliminación de los vasos sanguíneos como fuente de ruido, hasta su reincorporación

en el análisis visto como punto de referencia para la ubicación del disco, también se ha explorado la variación en el color para definir los límites de contorno del disco y la copa.

Desde el punto de vista de Deep learning, ha habido diferentes metodologías que buscan un ahorro computacional y una mejora en la efectividad del modelo, dentro de estas destacan:

- Recorte de la imagen para estimar solamente el radio entre el disco y la copa óptica
- Red neuronal convolucional profunda de 18 capas para la extracción de características robustas
- Utilización de diferentes redes para la detección de la copa y disco óptico inicialmente de forma independiente, y después de forma conjunta.

	DDLS stage	Narrowest rim width (rim/disk ratio) (average disk size: 1.50-2.00 mm)	Examples
At risk	1	0.4 or more	
	2	0.3 to 0.39	
	3	0.2 to 0.29	
	4	0.1 to 0.19	
Glaucoma damage	5	Less than 0.1	
	6	0 (Extension: Less than 45°)	
	7	0 (Extension: 46° to 90°)	
Glaucoma disability	8	0 (Extension: 91° to 180°)	
	9	0 (Extension: 181° to 270°)	
	10	0 (Extension: More than 270°)	

Figura 3-2 Diferencias entre el diámetro del disco y la copa óptica para un ojo en empeoramiento progresivo. Tomado de: [17]

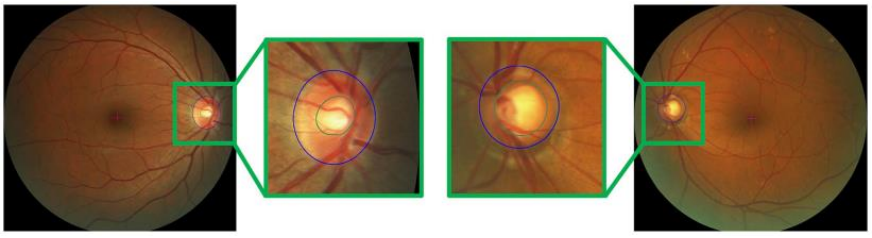


Figura 3-3 Área de interés para la detección de glaucoma. Tomado de: [17]

3.2.2. Retinopatia diabética

Los parámetros clínicos de la Retinopatía diabética (DR por sus siglas en inglés) se basan en la observación de imágenes de fondo de ojo, un proceso que depende de la experiencia del oftalmólogo. Las lesiones clave utilizadas para clasificar la gravedad de la DR incluyen microaneurismas, exudados duros y blandos, y hemorragias. Los microaneurismas son signos tempranos de la enfermedad, mientras que las hemorragias intrarretinianas ocurren cuando estos se rompen. La clasificación de la DR se basa en la presencia y extensión de estas lesiones, con el protocolo estándar "Early Treatment Diabetic Retinopathy Study" (ETDRS) utilizado históricamente, aunque su complejidad ha llevado a la adopción de la Escala Internacional de Gravedad de la Retinopatía Diabética, que clasifica la enfermedad en cinco niveles de severidad.

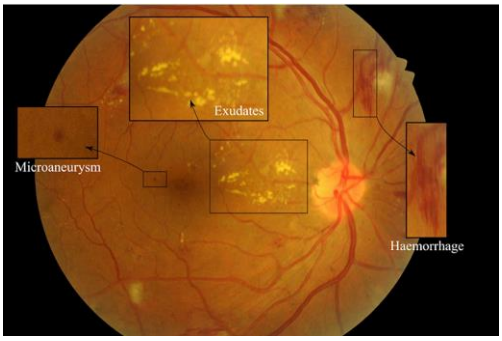


Figura 3-4 Lesiones más importantes detectadas en ojos con diabetes. Tomado de: [17]

3.2.3. Degeneración macular senil – *DMS*

En las etapas iniciales, la enfermedad es asintomática, y dado que el deterioro visual es gradual y afecta ambos ojos, muchas personas no se percatan hasta que la visión está gravemente afectada. Existen dos formas principales de DMS la "seca" (atrófica) y la "húmeda" (neovascular), siendo la forma seca la más común (85-90% de los casos), mientras que la forma húmeda es más agresiva y causa pérdida de visión más rápida.

La DMS seca se caracteriza por el adelgazamiento de la mácula y la acumulación de drusas, pequeñas acumulaciones de proteínas bajo la retina que causan atrofia y cicatrización. En cambio, la forma húmeda implica el crecimiento anormal de vasos sanguíneos debajo de la retina, que pueden filtrar líquidos y causar cicatrices. Aunque no existe cura para la DMS, los tratamientos con inyecciones intravítreas de anti-VEGF pueden ralentizar la progresión de la enfermedad inhibiendo el factor de crecimiento de los vasos anómalos, ayudando así a reducir la filtración y la cicatrización. Sin embargo, estos tratamientos no revierten el daño existente y requieren inyecciones frecuentes, lo que puede resultar doloroso y representar una

carga considerable para los pacientes. Sin embargo, estos tratamientos no revierten el daño y pueden ser dolorosos, lo que representa una carga significativa para los pacientes.

El análisis de drusas, anormalidades pigmentarias, atrofia geográfica y neovascularización coroidea en imágenes de fondo de ojo son las principales medidas para evaluar la gravedad de la DMS. Las escalas de clasificación comúnmente evalúan tanto las lesiones retinianas individuales como su ubicación en la mácula, siendo crucial analizar el tamaño, área y número de estas lesiones para un diagnóstico preciso.

Estudios han demostrado una efectividad del 100% en la detección de drusas como principal indicador para la clasificación de esta patología. Sin embargo, estos modelos aún presentan limitaciones al identificar drusas "suaves", las cuales tienen una morfología que dificulta la detección de bordes y el mapeo preciso del área afectada.

Desde 2010, se han logrado avances importantes mediante el análisis de texturas y la segmentación basada en clusters. Estas metodologías incluyen la adición de un paso en el pre procesamiento, donde se aplican filtros que resaltan la intensidad de las texturas superficiales. Como resultado, los modelos alcanzan una precisión (accuracy) del 96% para identificar drusas o la presencia de la enfermedad, mientras que el diagnóstico directo alcanza un accuracy del 85%.

Actualmente, existe un modelo que funciona como ensamble analizando la imagen en su totalidad, así como a través de procesos por parches. Este modelo ha logrado una precisión del 99%, con un puntaje de dato del 62%. Los clasificadores más eficaces hasta la fecha han sido los basados en máquinas de soporte vectorial (SVM por sus siglas en inglés), que han conseguido una accuracy del 92% en la identificación de estados avanzados de la enfermedad. La utilización de redes neuronales convolucionales (CNN por sus siglas en inglés) también ha demostrado ser eficaz para clasificar DMS húmeda y DMS seca. Estas redes neuronales buscan descomponer la imagen y detectar diferentes estructuras en cada capa, con el objetivo final de identificar patrones que permitan clasificar determinadas características o estructuras relacionadas con la enfermedad. Por otro lado, las máquinas de soporte vectorial, las máquinas de soporte vectorial se han destacado en la clasificación precisa de estados avanzados de la enfermedad debido a su capacidad para encontrar el hiperplano óptimo que separa las diferentes clases de datos. Las SVM son especialmente eficaces en conjuntos de datos con un alto grado de diferenciación entre clases, ya que maximizan el margen entre las clases posicionadas sobre un

espacio de características numérico, lo que mejora su capacidad de generalización y reduce el riesgo de sobreajuste.

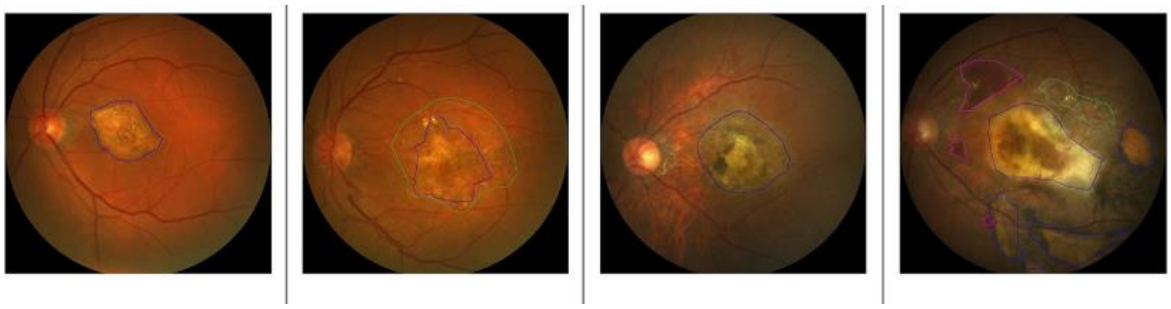


Figura 3-5 Evolución de DMS. Tomado de: [17]

3.2.4. Cataratas

La evaluación de las cataratas en imágenes de fondo de ojo se basa en la claridad de la imagen y los detalles visibles del ojo. En una imagen sin cataratas, el disco óptico y los vasos sanguíneos se ven claramente. Con cataratas leves, se observan menos detalles de los vasos sanguíneos, mientras que en cataratas moderadas solo se pueden distinguir los vasos sanguíneos grandes y el disco óptico. En cataratas severas, la visión en la imagen es casi nula. La progresión de las cataratas puede variar según la causa: aquellas relacionadas con el envejecimiento suelen avanzar lentamente, mientras que las cataratas inducidas por enfermedades, lesiones o factores genéticos o ambientales pueden progresar más rápidamente.

Los primeros métodos para la detección de cataratas se basaron en la transformación de la imagen mediante el uso de transformadas de Fourier, lo que permitió demostrar que el espectro de las imágenes cae en frecuencias más bajas con mayor rapidez en los casos de cataratas, en comparación con las imágenes de ojos sanos. Para las primeras aproximaciones, también se utilizaron discriminantes lineales para la segmentación, alcanzando una accuracy del 95%.

Clasificadores como Maquinas de soporte vectorial *(SVM en inglés)* han predominado en las aproximaciones que integran metodologías de aprendizaje automático. Estos algoritmos buscan la descomposición de la imagen en valores numéricos, con el objetivo de hallar el hiperplano que mejor separa las clases entre imágenes (en este caso para ojos sanos y con cataratas). El mejor rendimiento se alcanzó aplicando un procesamiento en lotes, con especial atención al disco óptico, dado que es un punto crucial para la detección. A través de la selección de características, el clasificador alcanzó una precisión del 95.33%, mejorando así la clasificación por grados.

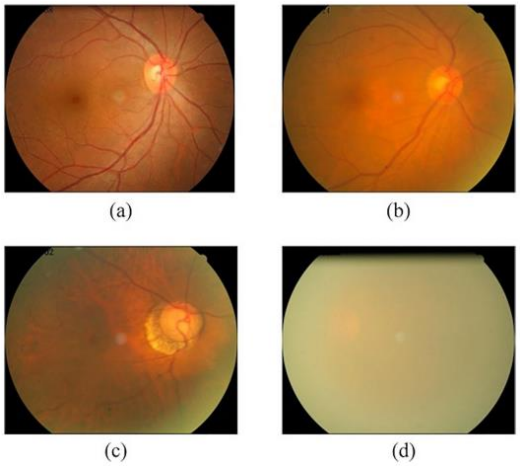


Figura 3-6 Evolución de cataratas. Tomado de: [17]

Aunque cada enfermedad oftalmológica tiene sus propios desafíos en cuanto a la detección, los estudios han puesto énfasis en la extracción de características clave y en la reducción de la dimensionalidad mediante técnicas como el análisis de componentes principales. A pesar de estos avances, todavía existen limitaciones significativas en cuanto a la subjetividad de los datasets utilizados y la especialización requerida para la detección de patologías oculares. Un ojo puede sufrir múltiples lesiones, lo que puede complicar el diagnóstico de una única patología. Finalmente, se resalta la importancia de desarrollar algoritmos multicapa que mejoren la identificación de ruido en las imágenes, así como la comprensión detallada de cada metodología específica.

3.3. Mobile-HR: un Sistema de clasificación oftalmológico para el diagnóstico de la retinopatía hipertensiva usando la arquitectura MobileNet optimizada

Un grupo de investigadores de diferentes países [18], realizó un proyecto enfocado en la detección de retinopatías hipertensivas con el objetivo de desarrollar una arquitectura de bajo costo que permitiera distinguir entre la presencia y ausencia de la enfermedad, independientemente del estado de avance en que se encontrara. La particularidad de la metodología utilizada radica en la extracción de características desde el punto de vista profesional médico, con el propósito de utilizarlas en el proceso de entrenamiento del modelo.

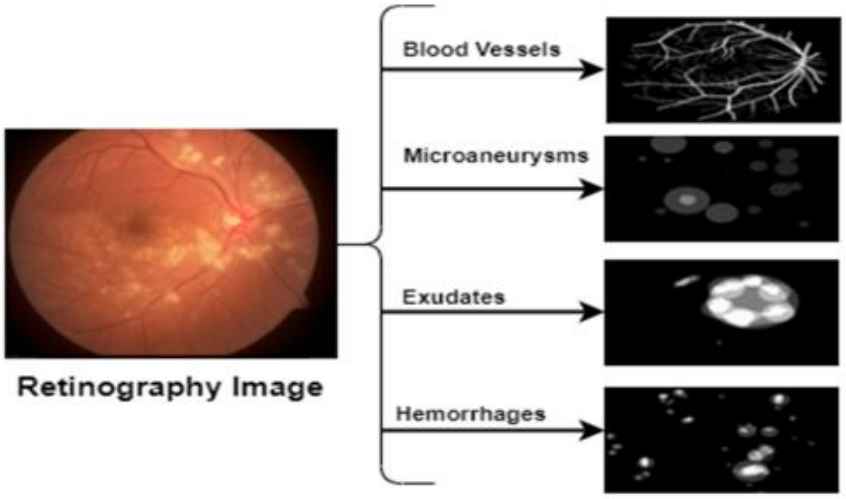


Figura 3-7 Características de una retinografía hipertensiva. Tomado de: [18]

El bajo costo de esta implementación se logró gracias a un enfoque nuevo en redes neuronales convolucionales, que integra características médicas especializadas directamente en el modelo. Esto permitió reducir tanto la necesidad de grandes volúmenes de datos de entrenamiento como la complejidad computacional. Este enfoque se complementa con el uso de la red neuronal *MobileNet*, que disminuye la complejidad del modelo mediante un proceso de convolución más compacto. Además, la inclusión de *dense blocks* mejora la interconexión entre las diferentes capas, permitiendo relacionar de manera efectiva todas las características definidas con el apoyo de expertos.

3.3.1. Convolución en profundidad y convolución puntual

Este proceso utiliza una metodología, en la que en lugar de aplicar un filtro a una imagen con diferentes valores (por ejemplo, RGB), reduce la dimensión al ingresar cada entrada de forma independiente, estas entradas se pasan por una capa convolucional adicional que combina la información obtenida desde los diferentes canales.

Una aproximación matemática de la reducción dimensional que ofrece este método puede ser vista de la siguiente forma:

Una convolución convencional contiene las dimensiones $\boldsymbol{K} * \boldsymbol{K} * \boldsymbol{M}$, haciendo referencia a las dimensiones bidimensionales del kernel, y el **M** hace referencia al rango de valores que puede contener ese filtro aplicado. Una convolución en profundidad, elimina una dimensión y convierte la matriz en un espacio bidimensional en el que aplica cada entrada por separado, quedando la dimensión final como $\boldsymbol{K} * \boldsymbol{K}$ para cada convolución [41].

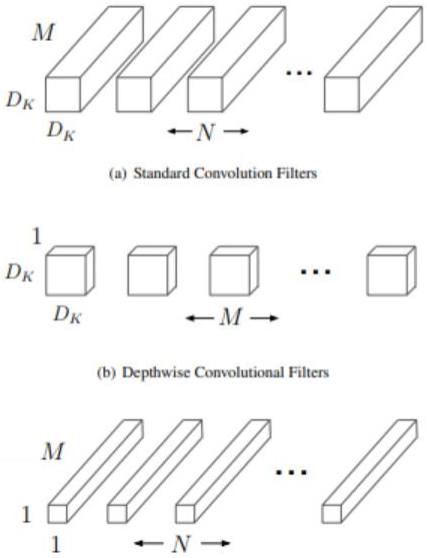


Figura 3-8 Reducción de dimensionalidad generada por depthwise convolution. Fecha de acceso: 4 de noviembre de 2024.Tomado de: https://arxiv.org/abs/1704.04861.

Por otro lado, las convoluciones puntuales tienen como objetivo combinar las características extraídas por las convoluciones en profundidad, reuniendo los canales de la imagen que han sido procesados de forma independiente. Estas convoluciones se aplican mediante filtros de tamaño 1×1, lo que permite que cada pixel de la salida influya en todos los canales de la entrada.

Las convoluciones puntuales pueden definirse como:

$$y = x_d * W_{pointwise} + b_{pointwise}$$

En donde:

x_d es la salida de la convolución en profundidad

$W_{pointwise}$ son los pesos a aplicar a cada canal

$b_{pointwise}$ es el sesgo asociado a cada canal de la salida anterior

Y la salida final termina siendo una imagen reconstruida con las características segmentadas.

3.3.2. Bloques densos:

Las redes neuronales convolucionales basadas en bloques densos buscan mejorar la eficiencia y precisión de modelos a través de la reutilización de características aprendidas en capas anteriores. En redes convolucionales convencionales, se tiene la siguiente formulación:

$$f_l = \sigma(W_l * f_{l-1} + b_l)$$

En la que $\boldsymbol{\sigma}$ es la función de activación para la capa actual, $\boldsymbol{W_l}$ son los filtros aplicados previamente, $\boldsymbol{b_l}$ es el sesgo de la capa, y $\boldsymbol{f_{l-1}(x)}$ es la función de la capa anterior.

Como es posible notar, la capa actual está siendo influida mayormente por la capa inmediatamente anterior, mientras que en las redes que utilizan bloques densos, se tiene la siguiente formulación:

$$f_l = \sigma(W_l * [f_{l-1}(x), f_{l-2}(x), \dots, f_0(x)] + b_l)$$

En la cual, es posible notar la influencia de las capas previas a la capa actual, lo que ayuda a la propagación del gradiente que estima los valores finales de cada capa convolucional, y la capacidad de la red para posicionarse en la capa actual a través de la influencia de capas anteriores.

3.3.3. MobileNet:

Mobilenet es una arquitectura que busca ser eficiente en términos computacionales para la implementación de modelos de clasificación en dispositivos con recursos limitados (dispositivos móviles y tablets), este tipo de redes utiliza convoluciones depthwise y pointwise cuya aplicabilidad busca adaptarse a redes con profundidad considerable las cuales se "comprimen" para ser accesibles a un menor costo computacional [41].

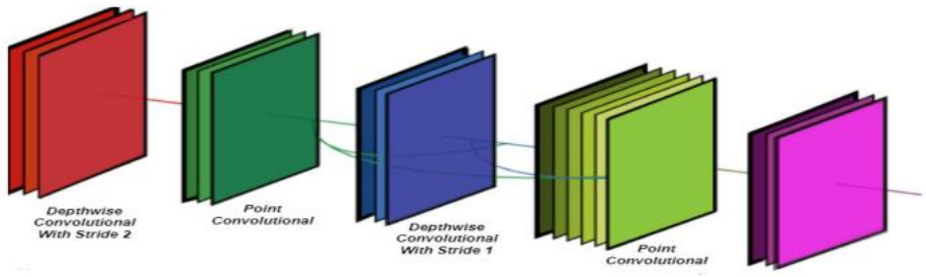


Figura 3-9 Ilustración de la red final. Tomado de: [18]

El flujo final del proyecto puede ser descrito de la siguiente manera:

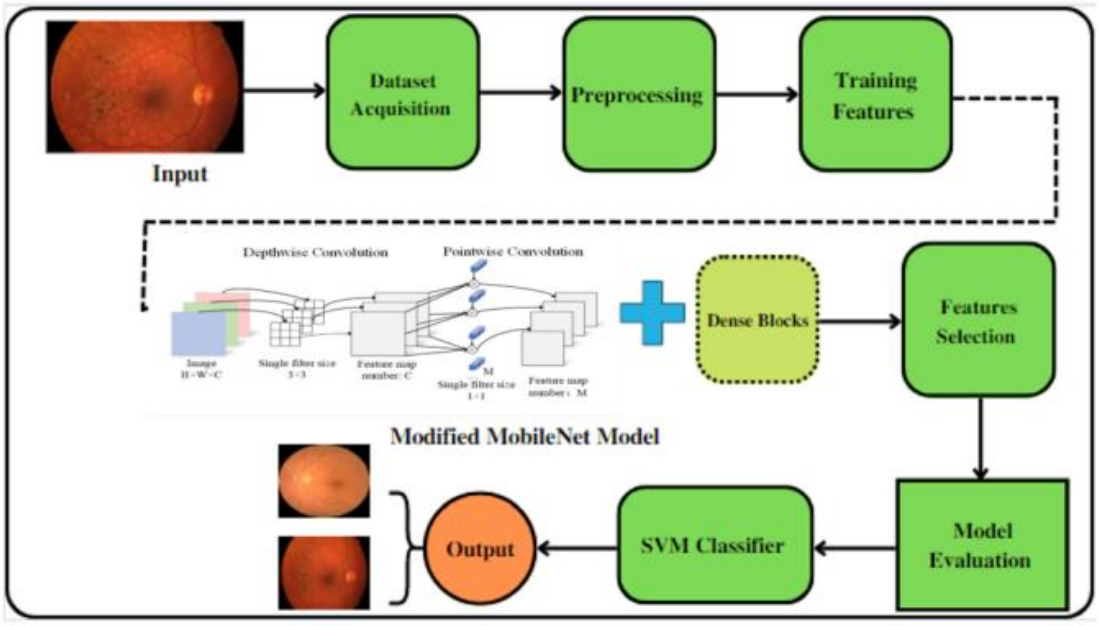


Figura 3-10 Flujo final de la operación investigativa. Tomado de: [18]

3.3.4. Recolección de datos

Se recopilaron datos de hospitales en Pakistán, con un total de 5,590 muestras retínales. De estas, 2,100 pertenecían a pacientes con retinopatía hipertensiva (HR). Adicionalmente, se incluyeron 3,170 muestras extraídas de un conjunto separado, que contenía 1,130 imágenes de pacientes con HR.

3.3.5. Pre procesamiento y aumento de datos

Para balancear las clases, se aplicó un aumento de datos utilizando diversas técnicas, siguiendo la metodología flip flop (rotación y reflejo de imagenes). Esto permitió corregir el desequilibrio de clases y mejorar la calidad del entrenamiento.

3.3.6. Implementación del modelo

Se utilizó el modelo *MobileNet* adaptado para la transferencia de aprendizaje en la detección de HR. La red neuronal fue entrenada utilizando un enfoque de validación cruzada de 10 pliegues, con parámetros estandarizados y el uso de *Stochastic Gradient Descent*. Los modelos se entrenaron en lotes de 64 imágenes debido a limitaciones de memoria GPU, con una tasa de aprendizaje inicial de 0.001.

3.3.7. Evaluación del modelo

Se evaluó la precisión del modelo en la clasificación mediante medidas estadísticas como precisión, recall, especificidad, y F1-score. La precisión alcanzó un 99% y la tasa de valores acertados fue del 100% para ambas clases. Se aplicó la regularización para evitar el sobreajuste (*overfitting*) y se usaron optimizadores como el optimizador adaptativo por estimación de momentos (ADAM por sus siglas en ingles) el cual busca acelerar la convergencia de aprendizaje de la red al ajustar dinámicamente las tasas de aprendizaje de cada parámetro teniendo en cuenta gradientes pasados para cada capa de la red.

Además del modelo *MobileNet*, se utilizó un clasificador de *Support vector machines* para identificar retinopatía hipertensiva en imágenes retinográficas. Este enfoque complementó la capacidad del modelo para manejar el ruido en las imágenes.

3.3.8. Resultados y recomendaciones

El sistema propuesto *Mobile-HR*, que integra bloques densos con la arquitectura *MobileNet*, ha demostrado una mejora significativa en el rendimiento de clasificación de imágenes de retinopatía hipertensiva (*HR*), alcanzando una precisión del 99%. Sin embargo, el estudio presenta limitaciones como la falta de evaluación de diversas clases de HR y el tamaño insuficiente del conjunto de datos utilizado. Para superar estas limitaciones, los autores recomiendan aumentar el tamaño del dataset y considerar el análisis de un espectro más amplio de clases de HR. Además, aunque el modelo es eficiente y ligero, se considera que un buen punto de evaluación está en la aplicación a dispositivos compactos para verificación de costo y adaptabilidad. Estas medidas permitirán mejorar la precisión del modelo y su aplicabilidad en escenarios prácticos.

3.4. Predicción retiniana a través de la segmentación y clasificación de vasos sanguíneos usando aproximación basadas en ensamble y aprendizaje profundo.

Una investigación llevada a cabo en el Instituto Nacional de Tecnología de Hamirpur [19] se centró en el desarrollo de un modelo de predicción de retinopatías sin basarse en la utilización de un criterio experto para la selección de características. El proceso comenzó con el preprocesamiento de las imágenes. En esta etapa se utilizó un análisis de componentes principales de la imagen en formato RGB para determinar la mejor forma de transformar la imagen a escala de grises. Posteriormente, se utilizó Ecualización de Histograma Limitada por Contraste (CLAHE) y operador de Contraste Toggle para mejorar la calidad en términos de contraste para las imágenes en escala de grises, esto debido a que, en un primer momento, no era posible diferenciar los vasos sanguíneos del fondo de la foto. Se aplicó un enfoque de filtro adaptativo basado en la Función de Densidad de Probabilidad (PDF) de un núcleo Gaussiano, lo que generó una imagen en alta resolución que resalta los vasos sanguíneos de la imagen de fondo de ojo.

El paso siguiente involucró la segmentación de la imagen, utilizando umbralización por entropía, filtrado por longitud y enmascaramiento de la imagen, con el fin de resaltar los vasos sanguíneos a través de un proceso más exhaustivo, analizando cambios entre pixeles adyacentes, y eliminando pequeños bloques de pixeles que separan los vasos sanguíneos del fondo de la foto. Estas imágenes segmentadas fueron luego clasificadas mediante un modelo de aprendizaje profundo en el que se combinaron tres arquitecturas: *VGG16, Efficient Net B0 y ResNet-152*. Estas arquitecturas buscaban la extracción de un conjunto de características de alta calidad basado en las imágenes con los vasos sanguíneos ya segmentados, las cuales fueron fusionadas mediante un enfoque de ensamble para mejorar la precisión de la clasificación.

3.4.1. Arquitecturas usadas en el ensamble final

- **ResNet:**

  Este tipo de redes introduce una metodología que busca mejorar el rendimiento de redes neuronales profundas, reduciendo la degradación del modelo con el avance y adición de capas que extraen características cada vez más complejas. El principal aporte de estas redes son las "conexiones residuales", que aflojan la rigurosidad del modelo, permitiendo que el flujo de información se salte ciertas capas que pueden degradar la optimización en la búsqueda del gradiente, este enfoque se basa en la idea de que la salida de una capa no solo depende de su capa anterior, sino también de la suma de su entrada original [39].

- **ResNet 152:**

  Es un modelo de la familia *ResNet* que fue preentrenado con 1.2 millones de imágenes y que busca tener un mayor poder representativo en la calidad y variedad de imágenes clasificadas.

- **EfficientNet:**

  Los modelos de tipo *efficientNet*, buscan estandarizar los componentes que conforman una red a través de la modificación de las capas vistas en su dimensionalidad total (Profundidad, ancho, y resolución), esta modificación está basada en la elección de recursos computacionales a utilizar. Por ejemplo: si tenemos un modelo que queremos utilizar a con 2^N Recursos, es posible escalar la red para que las dimensiones crezcan al mismo factor, la profundidad crece en α^N, el ancho crece en β^N, y el tamaño de imagen crece en γ^N, estos factores se determinan a través de un algoritmo de búsqueda en el modelo original [38].

- **EfficientNet B0:**

Es un modelo especifico de la familia EfficientNet, entrenado con más de un millón de imagen para la clasificación de más de 100 categorías.

- **VGG:**

  Las redes de tipo VGG, se caracterizan por la aplicación de un filtro 3X3 para cada capa, siguiendo la metodología usual en redes neuronales convolucionales a través de la cual se tiene una capa convolucional seguida de una capa de pooling para la reducción espacial sin pérdida del espacio de características, estos modelos buscan mejorar la precisión a través de la complejización en términos de adición de capas convolucionales a la red [40].

- **VGG16:**

  VGG16 es un modelo de la familia VGG que cuenta con 16 capas convolucionales (de las cuales 3 están conectadas completamente) que utiliza activación Relu, y que es capaz de identificar 1000 imágenes de 1000 categorías distintas.

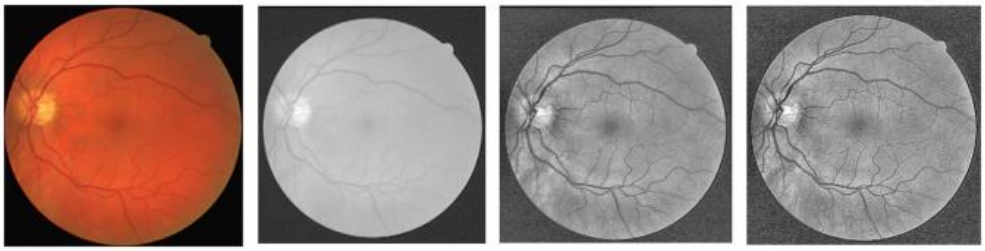


Figura 3-11 Imagen antes y después del pre procesamiento. Tomado de: [19]

La salida de estas tres arquitecturas, en conjunto con un clasificador SoftMax, permitió que el modelo lograra una alta precisión en la identificación de once clases diferentes de enfermedades retinianas. Entre las características clave de estas arquitecturas se destacan las capas convolucionales y de pooling de VGG16, que permitieron reducir la complejidad computacional y evitar el sobreajuste, así como la arquitectura residual de ResNet-152, que ayudo al aprendizaje de las características evitando el desvanecimiento del gradiente en el proceso.

El modelo propuesto demostró una mejora significativa con respecto a la literatura revisada, en términos de precisión de clasificación de enfermedades retinianas, logrando una tasa de precisión de 99.71%, una precisión de 98.63%, un recall de 98.25%, y un Fscore de 99.22%, para un total de 6 patologías, algunas en más de 3 estados de evolución. La investigación concluye que el uso de técnicas avanzadas de aprendizaje profundo y el enfoque de ensamble proporcionan una solución efectiva para la predicción precisa de enfermedades retinianas a partir de

imágenes de retinografías, aunque se declara la necesidad imperante de utilizar datos más significativos que permitan una mejor generalización.

La veracidad de los resultados fue evaluada a través de la utilización de curvas de aprendizaje a lo largo de un conjunto de épocas, para así determinar los puntos de minimización (en el caso de la función de perdida) y maximización (para la medición de métricas como el accuracy) este método busca verificar las capacidades de generalización del modelo sobre los datos de prueba desde la evolución del mismo.

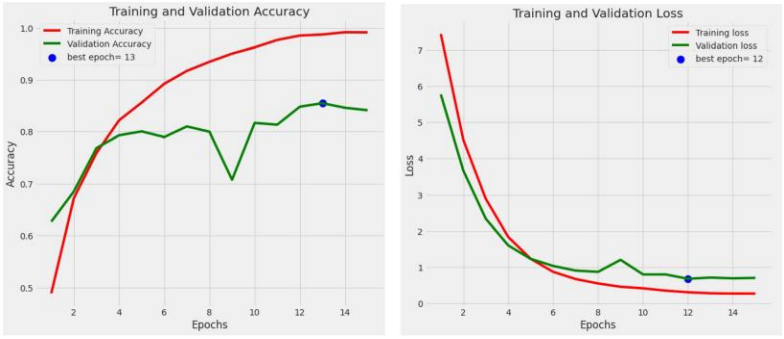


Figura 3-12 Curvas de aprendizaje del modelo propuesto. Tomado de: [19]

Finalmente, el conjunto de datos utilizado para entrenar y probar el modelo consistió en 20.077 imágenes para el conjunto de entrenamiento y 4.214 imágenes para el conjunto de prueba (partición 80-20), lo que resalta la robustez y eficacia del modelo propuesto en la clasificación de enfermedades retinianas en un conjunto de datos sustancialmente grande.

Table 5 Segmentation outcomes

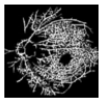
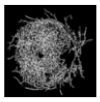
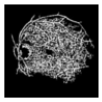
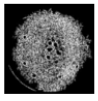
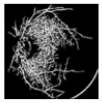
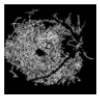
Segmented sample images	Class	Segmented sample images	Class
	Dry AMD		Hypertensive Retinopathy
	Moderate AMD		Cataract
	Pathological Myopia		Glaucoma
	Proliferate DR		Normal fundus
	Severe DR		Pathological Myopia

Figura 3-13 Segmentaciones obtenidas para las diferentes clases de enfermedades. Tomado de: [19]

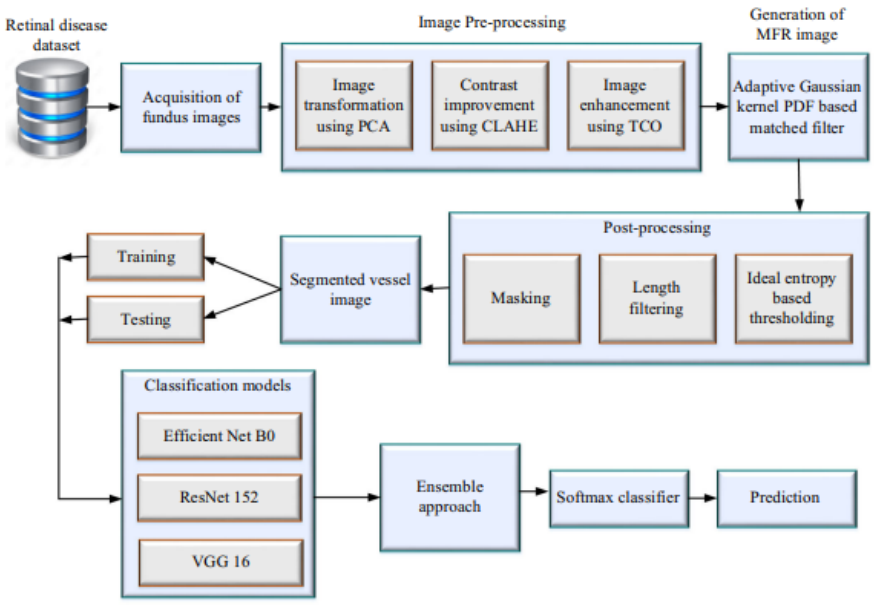


Figura 3-14 Flujo del proceso de investigación desde la adquisición de los datos. Tomado de: [19]

3.5. Estudios secundarios relacionados

Otros proyectos relacionados con este campo de investigación también buscan innovar mediante la implementación de técnicas específicas para mejorar el rendimiento y la extensión de los modelos. Por ejemplo, Sunija et al. [14] propuso un modelo basado en seis bloques convolucionales para identificar tres tipos de retinopatías en tomografías de coherencia óptica (OCT). El equipo de Google DeepMind [15] desarrolló una solución que aplica dos redes neuronales para predecir de manera temprana la degeneración macular relacionada con la edad. Una de las redes procesa la imagen OCT, mientras que la otra analiza la segmentación de tejidos de esa misma imagen para realizar la predicción. De manera similar, Subramanian et al. [16] destacan la importancia del transfer learning en las OCT y del aumento de imágenes en la etapa de preprocesamiento, con el objetivo de mejorar la generalización de los resultados.

Desde la perspectiva actual, se busca realizar una comparación basada en métricas como la exactitud y el puntaje F1. Además, se espera poder comparar los errores de predicción mediante una matriz multiclase que permita identificar la presencia o ausencia de enfermedades, basándose en características evaluadas por expertos, esta comparación se hará durante todo el desarrollo del proyecto, con el objetivo de tomar los mejores avances de cada proyecto de la literatura revisada, en esto se incluye: La búsqueda de un ensamble algorítmico de bajo costo computacional que permita procesar un número extenso de registros fotográficos, la utilización de preprocesamiento de

imágenes a partir de la extracción de características de forma automática y previa a la aplicación del modelo.

A continuación, se presenta una tabla con los resultados obtenidos hasta ahora vistos desde la literatura examinada.

Estudio	[18]	[19]	[14]	[15]	[16]	[29]
Enfermedades detectadas	Retinopatía hipertensiva	Degeneración macular senil (2 estados) Retinopatía diabética (4 estados) Cataratas Glaucoma Retinopatía hipertensiva Miopía patológica	Degeneración macular senil (Drusas) Neovascularización Coroidea	Degeneración macular	AMD CNV DME DRUSEN	Diabetes (4 estados)
Dataset usado	DRIVE DiaRetD80 Pak Eyes Hospitals (Dataset privado) Imam-HR (Dataset privado)	Messidor-2 dataset STARE ARIA HRF ODIR dataset DRIVE	EyePACS APTOS 2019 MESSIDOR2 MESSIDOR IDRiD DIARETDB0 DIARETDB1 DRIVE STARE ROC DDR DDR	Dataset privado	Dataset de kaggle y dataset con aumento de datos artificiales Dataset final: “ Retinal OCT–C8”	EyePACS APTOS 2019 MESSIDOR2 MESSIDOR IDRiD DIARETDB0 DIARETDB1 DRIVE STARE ROC DDR DDR
Hardware	Procesador HP-i7 octacore, 16 gb RAM, 2GB NVIDIA GPU	No menciona	PC con Nvidia GeForce, RTX2080 11 GB GPU	No menciona	No menciona	32GB RAM, i7processor 10th generation CPU along with GPU NVIDIA Quadro RTX 5000 with 16 GB RAM
Rendimiento	Tope alcanzado en 10 epochs con 2.4 segundos	Tope alcanzado en 14 epochs	Tope alcanzado con un ratio de aprendizaje de 0.003 con regularización L2	No menciona	75 epochs con parade temprana	0.169 a 199 segundos de procesamiento para hacer clasificación de imagen Tope alcanzado en los 100 epochs
Objetivo	Definición de features importantes previa a la aplicación del modelo para mejorar el rendimiento	Enfoque en el preprocesamiento para segmentar los vasos sanguíneos	Busqueda de reducción en coste omputacional con la utilización de una red con bloques	Utilización de segmentación de tejidos de un oct y de escáner digital para dos modelos DL que predigan el estado futuro de la enfermedad	Arquitectura basada en optimización bayesiana	Tomar arquitectura preentrenadas y añadir (en algunas) capas a la red para mejorar el mapeo de caracteristicas
Técnicas usadas	Uso de bloques densos + clasificador SVM	Ensamble compuesto por 3 CNNS pre entrenadas	Red convolucional en 6 bloques con clasificador softmax	Modelo de segmentación en 3D y modelo de clasificación	VGG16, DenseNet201, InceptionV3, y Xception	Arquitecturas: • MobileNetV2 • MobileNetV3 • NASNetMobileMobile • EfficientNetV2B0 • InceptionV3 Clasificador multiclase para los posibles estados de la enfermedad
Métricas	ACC: 99.71% Recall:98.25% F1: 98.22%	ACC: 99.71% Recall:97.85% F1: 99.21%	ACC: 99.69% Recall:99.69% F1: 99.68	ACC: 94.5% AUC: 88.6% Recall: 80%	ACC: 99.71% Recall:96.65% F1: 96.33%	ACC: 98.4% Recall: 95.6% F1: 98.4%

[33]	[30]	[31]	[34]	[35]
Glaucoma	Diabetes (4 estados)	Diabetes (4 estados)	Papiledema	Retinopatía diabetica
Conjuntos de datos propios extraídos del hospital de Beijing	APTOS 2019	APTOS 2019	Datos de estudio externo FOTO-ED [38]	Datos locales privados
No menciona	Intel Core i7-CPU 4.5 GHz	No menciona	No menciona	No menciona
No menciona	Tope alcanzado con 50 epochs	Tope alcanzado con 30 epochs	No menciona	No menciona
Estimación del grosor de la capa del nervio retinal para la detección de glaucoma	Utilizar un mapeo amplio de caracteristicas para mejorar el rendimiento	Utilizar una red distribuida que se entrene con los mismo datos y finalice con una toma de decisión conjunta de ambas arquitecturas preentrenadas	Utilización de la data de un estudio previo con una nueva metodología de Deep learning para comparar resultados	Utilización de un esquema de aprendizaje por transferencia sobre OCT
• Arquitectura ResNet 50 clasificador SVM	Red neuronal profunda con clasificador softmax	Doble arquitectura con funcionamiento en paralelo y merge para clasificación • ResNet50 EfficientNetB0	Utilización de la arquitectura BONSAI especializada en la detección de características del disco optico	Arquitecturas: • VGG19 DenseNet
AUC:86% RECALL:84.94%	ACC: 99% Recall: 99% F1: 99%	ACC: 89.60% Recall:98.60% F1: 89.15	ACC: 98.4% Recall: 84% AUC:97%	DenseNet: ACC: 89.7% Recall: 93.1%

[36]	Glaucoma	Drishti-GS, RIM-ONE-R3, ACRIMA, DRIVE	Intel (R) Core (tm) i9-10900 k cpu@3.7 GHz, 64.0 GB of ram, NVIDIA GEFORCE GTX 3080	50 Epochs	Enfoque en la detección del disco optico, eliminando ruido generado por los vasos sanguineos.	Arquitectura: • Resnet 50 Bloque de atención residual (feature extracción)	ACRIMA dataset: ACC: 97.87% AUC: 97.89%
[37]	Diabetes Hipertensión AMD Neovascularización Glaucoma	DRIVE, STARE, CHASE DB1, HRF, DRISHTI-GS	No menciona	No menciona	Utilización de arquitectura con optimizador y parámetros específicos para verificar el rendimiento multiclase	Arquitectura • Red neuronal con gradiente activo Optimización de araña roja	ACC: 97.65% Recall: 84.62% AUC: 98.54%

Tabla 1 Estudios secundarios referenciados

4. METODOLOGÍA PROPUESTA

Según la literatura revisada, se identifica una limitación central en los modelos de detección de patologías oculares: la ausencia de explicabilidad diagnóstica. La identificación de retinopatías (como la diabética y la degeneración macular senil) requiere metodologías alineadas con el criterio experto, fundamentadas en parámetros clínicos que validen cada diagnóstico.

Este proyecto propone un avance inicial para abordar dicha problemática, desarrollando un modelo que combine:

- Capacidad de generalización mediante el uso de datos diversos, libres de sesgos geográficos, tecnológicos o demográficos
- Mecanismos explicativos que vinculen las decisiones del algoritmo con hallazgos localizados en las imágenes, como alteraciones morfológicas en estructuras oculares o lesiones patognomónicas.

La mayoría de los modelos existentes presentan limitaciones críticas: entrenamiento con conjuntos de datos restringidos (geográfica o tecnológicamente), lo que limita su aplicabilidad en entornos clínicos diversos. Esta rigidez operativa constituye

un obstáculo en escenarios donde la confiabilidad del sistema y su transparencia diagnóstica resultan tan cruciales como la precisión técnica.

La falta de explicabilidad de los modelos de IA en medicina reviste especial gravedad, al omitir los criterios determinantes detrás de cada diagnóstico. La integración de técnicas como Grad-CAM (para visualizar regiones anatómicas determinantes en las predicciones) no solo optimiza la interpretabilidad, sino que valida los presupuestos metodológicos mediante correlación clínico-computacional.

La implementación se estructurará en cuatro fases, garantizando una transición fluida hacia un sistema operativo listo para producción.

1. La primera fase consistirá en la recopilación y preprocesamiento de un conjunto de datos diverso.
2. La segunda fase incluirá el desarrollo de las metodologías de segmentación.
3. La tercera fase se centrará en obtener el mapeo final de características junto con la construcción del modelo de clasificación de retinopatías
4. Finalmente. La cuarta etapa consistirá en la aplicación de técnicas de explicabilidad, asegurando que los resultados sean interpretables y confiables desde las métricas a evaluar, así como desde un punto de vista más humano.

4.1. Protocolo de recolección de imágenes y pre procesamiento de imágenes

Jordan et al. [46], realizaron una revisión para todos los repositorios de fondo de ojo disponibles públicamente con fecha de 2017, para este estudio se tuvieron en cuenta características de los conjuntos cómo:

- Especificaciones de la instrumentación utilizada
- Numero de imágenes
- Resolución espacial en pixeles
- Campo de visión
- Propósito del conjunto de datos

De la misma manera, Tomasz Krzywicki et al [47]. Realizaron la revisión de 127 repositorios de datos, teniendo en cuenta la misma información del estudio realizado por Jordan, no obstante. Añaden los países de origen para una mejor trazabilidad de las bases de datos.

Dado el alcance actual del proyecto, se buscará recolectar información relacionada con: Glaucoma, Degeneración macular senil (DMS) y Retinopatía diabética. Esto con el propósito de lograr unificar las metodologías principalmente usadas hasta la fecha para la identificación de lesiones y estructuras (segmentación basada en análisis de copa óptica y disco óptico, junto con segmentación de microaneurismas y drusas). También se filtrarán los conjuntos de datos a partir de las regiones de orígenes que permitan

alcanzar una mejor representatividad de acuerdo con los objetivos planteados. No se cuenta con un criterio específico de selección más allá de la distribución de los datos a lo largo de diferentes regiones, junto con las características relacionadas a la misma foto (instrumentación, resolución, definición, etc.), esto debido a la variabilidad aportada por las mismas diferencias en los métodos de recolección de las imágenes con el propósito de velar por la adaptabilidad del modelo.

A continuación, se presentan los repositorios escogidos para el desarrollo del proyecto, junto con sus especificaciones contextuales.

Nombre	Derbi_Hackathon_Retinal_Fundus_Image_Dataset [50]	ACRFAODAWDCMFERLRLASMAGD Versión 4 [51]	Drishti-GS1 [62]	ODIR [52]	E opththa [60]
Numero de fotos	100	2560	101	5,000	463
Resolución	Diferentes: 3000 × 3152 px; 1728 × 2592 px; . . .	Diferentes: 1934 × 2032 px; 576 × 760 px; . . .	Different: 2048 × 1760 px; 2049 × 1749; . . .	No especifica	Diferentes f 1440 × 960 to 2544 × 1696 pixels.
Tipo de recolección	Grados de patología	Grados de patología	Anotaciones en la imagen	Grados de patología	Anotaciones
Enfermedades registradas	Retinopatía diabética, Retinopatía hipertensiva	Degeneración macular senil, Glaucoma	Glaucoma	Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age related Macular Degeneration (A), Hypertension (H),	Retinopatía diabetica

Retinal Fundus Multi-disease Image Dataset 2.0 [61]	ACRIMA [53]	MESSIDOR-2 [54]	DRIVE [55]	Retinal Fundus Multi-disease Image Dataset [56]	REFUGE2 [57]	INSPIRE-AVR [59]	DR HAGIS [58]
860	705	1058	40	3200	1200	65	40
Diferentes	Diferentes: 349 × 349 px; 871 × 871 px; . . .	Diferentes	565 X 584	Diferentes	Difrentes: 2124 × 2056 1634 × 1634 1940 × 1940 1848 × 1848	2392 × 2048 px	Diferentes: 4752x3168 pixels, 3456x2304 pixels, 3126x2136 pixels, 2896x1944 pixels or 2816x1880
Grados de patología	Grados de patología	Grados de patología	Grados de patología	Grados de patología	Grados de patología	Grados de patología	Grados de patología
Glaucoma, AMD	Glaucoma	Retinopatía diabética	Retinopatía diabética	Retinopatía diabética, degeneración macular senil, glaucoma	Glaucoma	Retinopatía hipertensiva, Glaucoma, Enfermedades vasculares de la retina	Glaucoma, Hipertensión, diabetes, AMD

EYE PACS[67]	OCT-AND-EYE-FUNDUS-DATASET [68]	Dataset from fundus images for the study of diabetic retinopathy [69]	IDRiD [71]	drusen_segmentation [74]	ORIGA [78]
20	1548	1437	80	140	650
No especifica	1000x1000 pixels	No especifica	No especifica	No especifica	No especifica
Grados de patología	Grados	Anotaciones en imagen	Anotaciones en imagen	Anotaciones	Anotaciones
Diabetes	AMD	Diabetes	Diabetes	Drusas	Glaucoma

Glaucoma	Anotacoines	No especifica	1020	G1020[78]

Tabla 2 Datasets revisados

Una vez con los datos en crudo, se pasan los escáneres a través de un conducto estándar de pre procesamiento para consolidar las imágenes finales. El filtro inicial consiste en clasificar las imágenes, excluyendo aquellas que no estén dentro de las enfermedades ya definidas anteriormente.

- **Técnicas de mejoramiento de la calidad de la imagen:**

  El proceso de mejoramiento de imagen se dividirá en tres partes:

  1. **Preprocesamiento de disco, copa óptica, y drusas**

     - **Descomposición de la imagen en canales**

       Muchas de las técnicas actuales para consolidación de imágenes, están basadas en la modificación de la imagen en escala de grises, es por este motivo que Sonali et al. [48], Proponen una metodología en la cual se separan los canales de la imagen (Rojo, verde, y azul) para posteriormente aplicar los filtros que se consideren necesario sobre estos tres canales sin tener que estar cerrado a buscar una técnica para pasar la imagen a escala de grises teniendo en cuenta los 3 canales esta de forma simultánea.

     - **Filtros de reducción de ruido**

       En el trabajo de Sonali et al. [48], realizaron una revisión exhaustiva sobre el impacto de diferentes métodos de reducción de ruido en imágenes de fondo de ojo. En su estudio, se evaluaron dos tipos predominantes de ruido: el ruido gaussiano, que se presenta como fluctuaciones aleatorias en la intensidad de los píxeles, y el ruido de sal y pimienta, que se manifiesta como píxeles brillantes y oscuros esparcidos por la imagen de forma aleatoria. Dentro de su estudio, descubrieron que el filtro de mediana ponderada presenta los mejores resultados para un proceso de des y recomposición de imagen cuando se quiere utilizar un enfoque mixto basado en una subsecuente corrección de contraste utilizando CLAHE (Técnica utilizada también más adelante).

       Por otro lado, Vijayakumar et al**.** [49] también investigaron este tema y destacaron la eficacia del filtro por mediana adaptativo. Este filtro ajusta

dinámicamente el tamaño de la ventana de mediana en función de las características locales de la imagen a nivel de intensidades en escala de grises.

- **Filtro de mediana**

  El filtro de mediana es una técnica de reducción de ruido que se basa en el cálculo de la mediana de un conjunto de píxeles dentro de una ventana. Esta técnica ha demostrado ser eficaz para eliminar diferentes clases de ruido [48 y 49], como el ruido de "sal y pimienta" caracterizado por la aparición de píxeles aleatorios que son significativamente más claros o más oscuros que sus vecinos. Gracias a la robustez de este filtro, es posible conservar la información de bordes y texturas, lo que resulta útil en el análisis de exudados, micro aneurismas, y drusas en las imágenes de fondo de ojo dada la tendencia a manifestarse como estructuras pequeñas sin distribución definida.

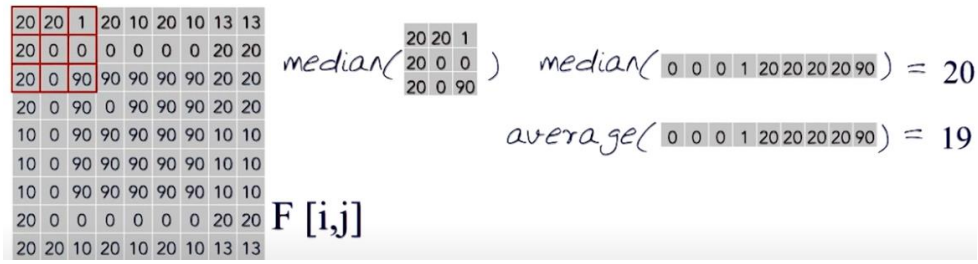


Figura 4-1 Filtro de mediana para imagen comparado con filtro promedio para imagen. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www.youtube.com/watch?v=TYNbTdzm-5w

  - **Ventana o kernel:** Para cada píxel en la imagen, se define una ventana de píxeles alrededor de él. En esta ventana, se calcula la mediana de los valores de intensidad para hacer el reemplazo sobre el valor original de los pixeles.

- **CLAHE (Contrast Limited Adaptive Histogram Equalization)**

  Una vez se tenga la imagen procesada con el primer filtro, se va a buscar mejorar esta en términos de calidad de contraste e iluminación para destacar incluso aún más los detalles concernientes a las apariciones específicas de enfermedades en las imágenes.

  Las metodologías basadas en Ecualización adaptativa del histograma de una imagen, buscan dividir imágenes en secciones para la aplicación de la redistribución de los valores de intensidad. En un principio, se crearon los modelos "AHE" que buscaban dividir la imagen y aplicar ecualización del histograma de forma local, teniendo en cuenta los pixeles vecinos para un área determinada, debido a este acercamiento, el algoritmo tiende a amplificar el ruido causado por la influencia de

vecinos con una iluminación homogénea, generando así que la distribución del histograma alcance picos que pueden sobrexponer la imagen en su totalidad. Frente a estas limitaciones, el algoritmo "*CLAHE"* redistribuye los valores de intensidad de los píxeles para mejorar el contraste, pero con un límite en el número de píxeles que se pueden redistribuir, lo que evita el sobre contraste en regiones homogéneas visto desde la distribución de la misma función de acumulada de los pixeles en la imagen.

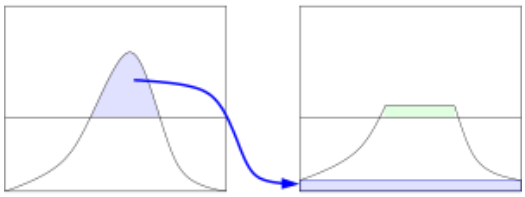


Figura 4-2 Comparación del histograma usando AHE vs CLAHE. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://en.wikipedia.org/wiki/Adaptive_histogram_equalization

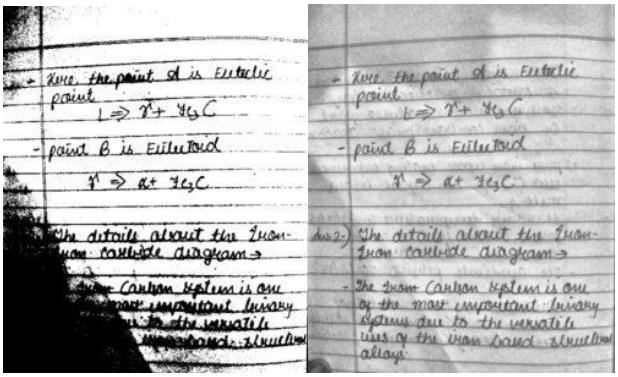


Figura 4-3 Comparación de imagen procesada usando AHE vs CLAHE. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www-geeksforgeeks-org.translate.goog/clahe-histogram-eqalization-opencv/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc

En el presente proyecto se va a realizar la aplicación del *CLAHE* sobre el canal de luminancia de la imagen, para de esta forma mitigar los posibles picos generados por sobresaturación o falta de iluminación dentro de la misma.

2. Preprocesamiento para segmentación de drusas

- **Red de imagen completa**

  Para las drusas se utilizará una red neuronal previa de imagen completa (descrita en la sección 4.2.3), la cual va a generar un mapa probabilístico para zonas candidatas de presencia de drusas en cada imagen.

- **Recorte de las imágenes en parches**

  Para estructuras pequeñas y que representan una pequeña proporción de la imagen (como lo son las drusas) se añadirá un paso consistente en el recorte de la imagen en parches que contengan la estructura de interés para así manejar el desbalance de clases causado por la naturaleza de la afección.

- **Concatenación de mapa probabilistico con parche de imagen**

  De la red aplicada a la imagen completa, se seleccionarán aquellos parches del mapa probabilístico que coincidan con las máscaras de las drusas. De este modo, se concatenará la imagen con el mapa probabilístico para obtener tensores de dimensiones (4×128×128).

3. **Preprocesamiento para segmentación de microaneurismas**

- **Aislamiento y mejora del canal verde de la imagen**

  Dado que los microaneurismas presentan un mayor contraste en el canal verde de la imagen [64], se utilizará este para hacer la extracción de características y la segmentación de las estructuras de interés. Para mejorar la diferenciación de los microaneurismas, se aplicarán técnicas de realce de contraste, como la ecualización de histograma limitada "*CLAHE*" o métodos basados en transformaciones adaptativas, con el objetivo de resaltar estas pequeñas lesiones con respecto el fondo de la imagen.

- **Recorte en parches de las imágenes con las estructuras de interés**

  Los microaneurismas son lesiones de pequeño tamaño que pueden aparecer en diversas áreas del fondo de ojo que, en términos de proporción, ocupan menos del 0.01% de la imagen completa. Esto genera un fuerte desbalance de clases al separar el fondo de la imagen de los microaneurismas, lo que puede afectar el rendimiento de los modelos de segmentación. Para mitigar este problema, se implementará un proceso de recorte en parches bajo un factor constante, enfocándose en regiones con alta probabilidad de contener microaneurismas. Esta estrategia permitirá aumentar la representatividad de la clase minoritaria y mejorar la capacidad del modelo para detectar las estructuras.

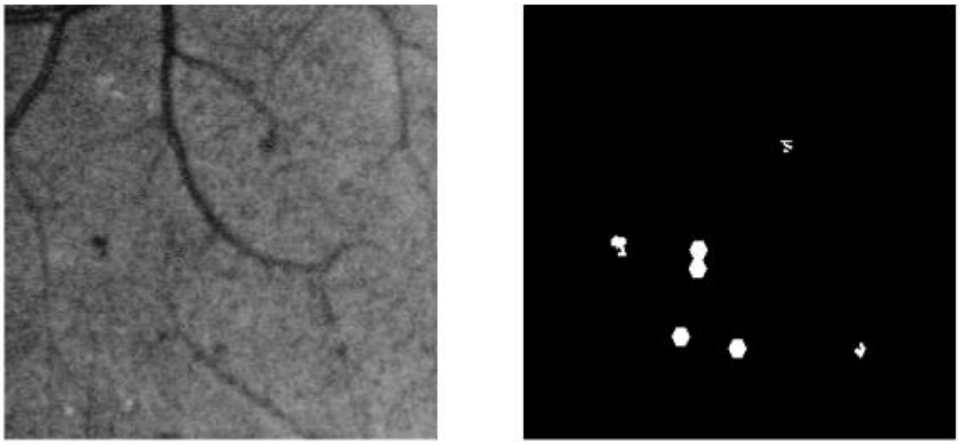


Figura 4-4 Imagen de fondo de ojo recortada con su respectiva mascara para la identificación de micro aneurismas. Elaboración propia

Para este proceso, las imágenes serán divididas en parches de 48x48 píxeles con el objetivo de mitigar problemas como la escasez de muestras para las etapas de entrenamiento y evaluación, así como el desbalance entre clases. Adicionalmente, se empleará un parámetro de solapamiento que determinará la velocidad del desplazamiento del kernel sobre la imagen, permitiendo así generar múltiples vistas de la región de interés y aumentar la representación de las lesiones a segmentar.

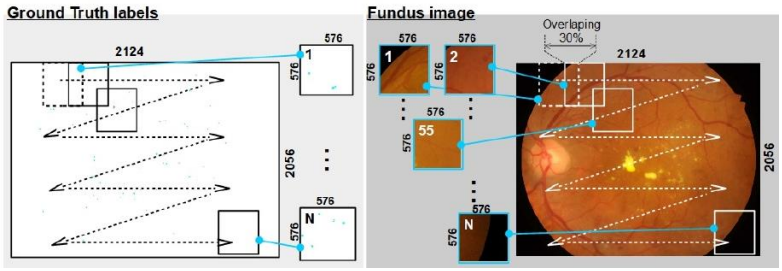


Figura 4-5 Representación de recorte por parches con solapamiento. Fecha de acceso. 21 de abril de 2025. Tomado de [72]

4.2. Protocolo de pos procesamiento

Esta sección pasará a explicar el funcionamiento de los algoritmos dedicados a la segmentación inicial de estructuras concernientes a las enfermedades seleccionadas para el desarrollo del proyecto.

4.2.1. Segmentación de disco y copa óptica

- **MobilenetV2:**

diseñada específicamente para dispositivos con limitados recursos de procesamiento, como teléfonos móviles y dispositivos embebidos. Esta

arquitectura se basa en una optimización llamada convolución separable en profundidad (ya explicada en la sección 3.3.1), que reduce el número de parámetros y operaciones, manteniendo un alto rendimiento en tareas de detección y clasificación de objetos. La innovación detrás de esta arquitectura se basa en un proceso compuesto el cual implica:

1. **Bloque invertido residual (bottlenecks)**

   Un bloque invertido residual va a contener 3 pasos antes de obtener la salida para una capa:

   1. **Expansión**:

      En esta fase se busca aumentar el número de canales de una pasada por una convolución inicial en dimensión 2X2, la salida de esta capa es entonces procesada para obtener más canales a través de una capa con función de activación RELU que se puede definir como:

      $$y_1 = \text{ReLU6}(x * W_1 + b_1)$$

      En el que:

      - x = tensor de entrada que proviene de la capa anterior
      - $w_1 =$ peso utilizado en la convolución definido como parámetros que se hallarán haciendo backpropagation, y que se definen como aleatorios inicialmente
      - $* =$ Operación de convolución
      - $b_1 =$ Sesgo agregado a la capa del modelo
      - ReLU6 es una función de activación para los valores calculados anteriormente.

   2. **Convolución en profundidad:**

      Remítase a la sección (3.3.1) para la explicación de las convoluciones en profundidad, el proceso es esencialmente el mismo.

   3. **Convolución puntual:**

      Remítase a las pagina (3.3.1) para la explicación de las covoluciones puntuales, el proceso es esencialmente el mismo.

   4. **Adición de conexión residual:**

      Finalmente, se adiciona una conexión residual, a la salida de la capa, esto hará que se adicione la entrada original para así mantener características que no se pierdan por el desvanecimiento del gradiente en la red (la lógica es la misma utilizada en la sección 4.2.2 para las redes residuales "ResNet").

MobileNetV2 introduce el concepto de **bloque de convolución inversa residual** con el objetivo de permitir que las características importantes de la entrada se preserven utilizando conexiones de salto.

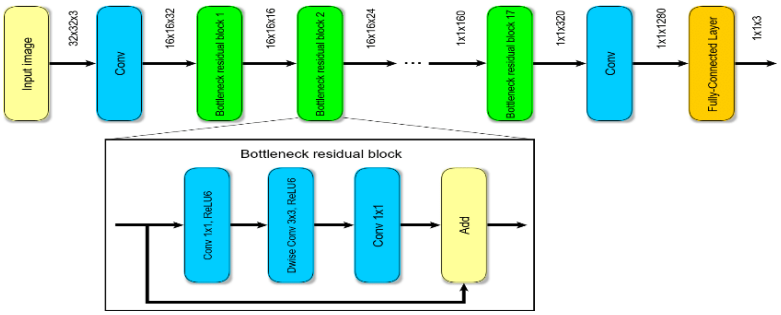


Figura 4-6 Bloque de bottleneck de mobileNetv2. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://blog.paperspace.com/transfer-learning-explained-2/

Básicamente el algoritmo toma la entrada x que se llevaba en ese momento, realiza una descomposición de la imagen en el bloque, y adiciona la entrada a las características extraídas para la función de activación $F(x)$ con esto no se pierde información y se genera una mejora paulatina del modelo a mayor profundidad.

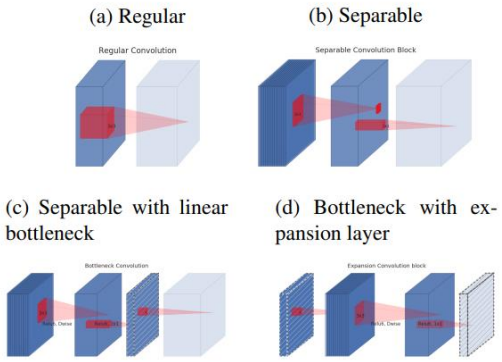


Figura 4-7 Comparativa entre tipos de convolución de imagen. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://arxiv.org/abs/1801.04381

- **MobilenetV3:**

En comparación a su predecesora (Mobilenetv2) esta nueva generación de red neuronal introduce ciertos cambios que permiten una mejora en la velocidad y calidad de los resultados. Estas mejoras incluyen:

- **Cambio en la función de activación**

En lugar de utilizar funciones RELU6 en la nueva arquitectura se utilizan funciones de activación "hard swish", las cuales buscan resolver problemas como la propagación en caso de encontrar pesos negativos manteniendo un costo computacional bajo. Esta función se define como:

$$h - swish(x) = x\frac{RELU6(x+3)}{6}$$

Esta función de activación no lineal, se integra en la primera capa de la red, así como en una mayoría de capas "bottlenecks" intermedias, para usarse de forma conjunta con funciones RELU.

- **Modificación en los cuellos de botella**

Dentro de las capas de cuellos de botella de la red, se añaden bloques de 'Squeeze-and-Excitation', los cuales se basan en la idea de asignar parámetros a cada canal dentro de una imagen, permitiendo que la red ajuste de forma adaptativa los pesos de cada mapa de características resultante. Esto implica la descomposición de los mapas de características en vectores unidimensionales, que pasarán por una serie de capas paralelas, las cuales generarán una serie de pesos utilizables en los mapas de características iniciales.

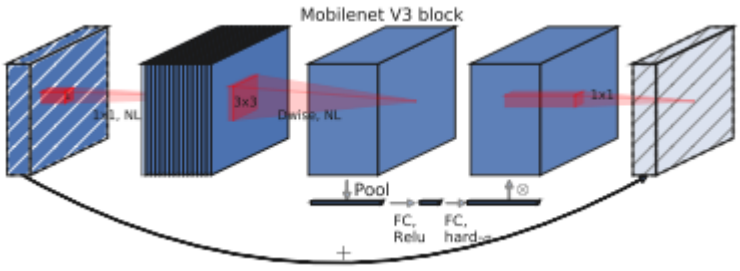


Figura 4-8 Ilustración la adición de un bloque squeeze and excite a la arquitectura. Tomado de [70]

El motivo detrás de la utilización de este modelo, es descomponer las imágenes para la obtención de mapas características que posteriormente permitan encuadrar la imagen de fondo de ojo de forma automática dentro de un área de interés que contenga el disco y la copa óptica, esto con el propósito de complementar la metodología utilizando de forma secuencial la imagen encuadrada, y la arquitectura *U-net* para la estimación semántica por pixeles

del disco y la copa óptica. Este procedimiento tiene como justificación el trabajo realizado en [63] en el cual se destaca la mejora en resultados a través de esta metodología compuesta usando MobilenetV2.

- **Detector multicaja de disparo único**

El detector multicaja de disparo único (SSD por sus siglas en inglés) hace una única pasada sobre la imagen de salida final que genera la red neuronal, la idea es definir una serie de cajas limitantes de área optimizando el costo computacional de la predicción final, esto lo hace a través de una asignación de pertenencias de clase para categorías definida como:

$$P(c|z) = \frac{e^{z_c}}{\sum_{j=1}^{c} e^{z_j}}$$

Donde:

- z es la salida de la capa final del modelo
- c es la categoría o categorías presentes.

La idea es buscar la minimización de una función de perdida definida como:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(P(c|z_i))$$

Donde:

- y_{ic} es un indicador de si la caja i pertenece a la categoría c
- N es el número total de cajas propuestas

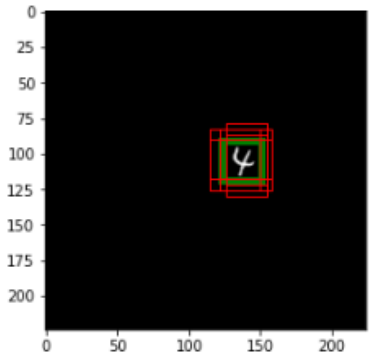


Figura 4-9 Representación de encuadre generado por MobileNetV2 con SSD. Fecha de acceso: 4 de noviembre de 24 Tomado de: https://github.com/saunack/MobileNetv2-SSD/blob/master/model.ipynb

Esta arquitectura está concatenada al final de la arquitectura MobileNetv3 con el objetivo de hacer las predicciones de encuadre sobre una imagen.

- **U-NET:**

La arquitectura *U-Net* es una red neuronal convolucional diseñada inicialmente para tareas de segmentación de imágenes biomédicas y enmascaramiento de características. Su nombre proviene de su estructura en forma de "U", que refleja el proceso de codificación y decodificación simétrico similar al usado en los autoencoders variacionales.

U-Net sigue un enfoque de *encoder-decoder*, en el cual la primera mitad (encoder) se encarga de descomponer la imagen mediante la extracción progresiva de características a través de capas convolucionales y operaciones de *max pooling* para reducción del espacio de características, reduciendo la resolución de la imagen. La segunda mitad (decoder) se enfoca en la reconstrucción de la imagen, incrementando gradualmente la resolución usando capas de *upsampling* y *convoluciones transpuestas*, mientras adiciona las características extraídas en las capas correspondientes del encoder. Esto permite a la red combinar la información detallada de las capas inferiores con la información contextual de las capas superiores.

Matemáticamente, cada capa convolucional en la fase de codificación aplica un filtro W sobre la entrada X para generar un conjunto de mapas de características F, esta operación está definida más arriba (página 15, formula 1) , y es la definición convencional para el cálculo del mapa de características. La diferencia radica en la fase de decodificación de la imagen, en la que se utilizan filtros transpuestos guardados en la codificación de la imagen, esta operación se puede ver como:

$$X = W^T * F + b$$

Donde "*" es la operación de convolución, y W^T es la matriz de filtros transpuesta.

A nivel gráfico, es posible entender que los features extraídos por los decoders, contienen información espacial sobre la ubicación del objeto, mientras los encoders mantienen información sobre los pixeles exactos que hacen referencia a las características segmentadas.

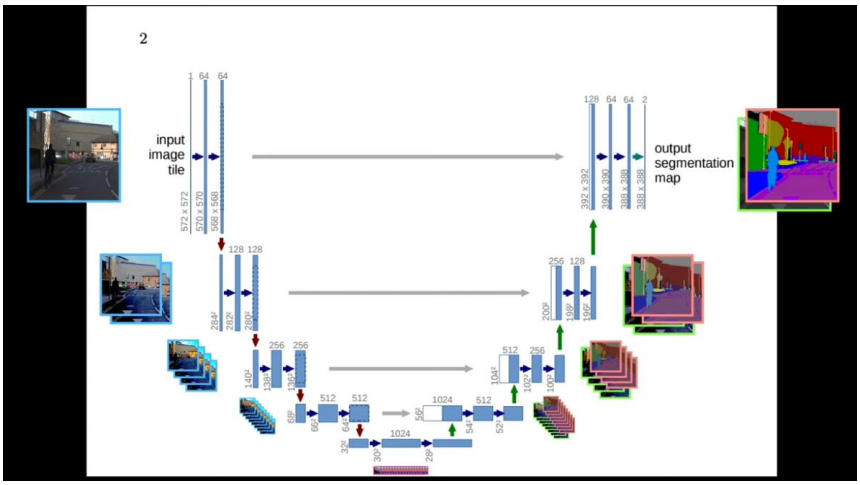


Figura 4-10 Representación gráfica de la arquitectura U-NET. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://www.youtube.com/watch?v=NhdzGfB1q74

Gracias a las capacidades con las que cuenta *U-Net*, se va a utilizar este algoritmo como principal punto de apoyo para la estimación de áreas de interés como lo son el disco y la copa óptica.

- **Detalles de implementación:**

Para el modelo de detección de disco y copa óptica se empleó el optimizador Adam, en combinación con la función de pérdida típica de redes con detector de disparo único multicaja (Single Shot MultiBox Detector, SSD). Esta función combina dos componentes: la pérdida de localización y la pérdida de confianza:

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}(x, c) + \alpha L_{loc}(x, l, g))$$

Donde:

- N es el número de cajas predeterminadas que coinciden con objetos de la clase de interés
- α es un parámetro de ponderación que equilibra la contribución de la perdida de localización (importancia de la localización de la caja por sobre la confianza del modelo sobre esta)

La pérdida de localización mide qué tan bien se ajustan las cajas predichas a las cajas reales, en términos de centro, ancho y alto. Se calcula utilizando la función *Smooth L1*, menos sensible a valores extremos que el error cuadrático medio:

$$L_{loc} = \sum_{i \in Pos}^{N} \sum_{m \in \{cx,cy,w,h\}} x_{ij}^{k} \, smooth_{L1}(l_i^m - \hat{g}_j^m)$$

Donde:

- x_{ij}^k es un indicador binario cuyo valor es 1 si la caja predicha i se asigna a la caja real j de la clase k, y 0 en caso contrario. Usualmente este valor se determina midiendo el índice de jaccard (definido en esta misma sección) para ambas cajas.
- l_i^mson las dimensiones predichas para la caja i.
- $\hat{g}_j^m$ son las coordenadas de la caja real j, ajustadas en relación con la caja predeterminada i, según:

$$\hat{g}_j^{cx} = \frac{\left(g_j^{cx} - d_i^{cx}\right)}{d_i^w}$$

$$\hat{g}_j^{cy} = \frac{\left(g_j^{cy} - d_i^{cy}\right)}{d_i^h}$$

$$\hat{g}_j^w = \log\left(\frac{g_j^w}{d_i^w}\right)$$

$$\hat{g}_j^h = \log\left(\frac{g_j^h}{d_i^h}\right)$$

En donde g_j representa la caja real, y d_i representa una caja guía para el ajuste de la caja real con respecto a desviaciones entre los valores de las bounding boxes para distintas muestras.

La pérdida de confianza evalúa si el modelo asigna correctamente la clase (incluyendo la clase "fondo" para cajas sin objeto) a cada caja. Se basa en la entropía cruzada softmax:

$$L_{conf}(x, c) = -\sum_{i \in Pos}^{N} x_{ij}^p \log\left(\hat{c}_i^p\right) - \sum_{i \in Neg} \log(\hat{c}_i^0)$$

Este componente:

- Penaliza cajas positivas clasificadas incorrectamente
- Penaliza cajas negativas (fondo) no se predicen como tal

La probabilidad predicha $\hat{c}_i^p$ para cada clase p se calcula mediante la aplicación de la función softmax:

$$\hat{c}_i^p = \frac{\exp(c_i^P)}{\sum_p exp(c_i^p)}$$

Donde c_i^p es el logit (salida en crudo) predicho por el modelo para una clase p en una caja i.

Como es posible notar, esta función de perdida se caracteriza por evaluar al mismo tiempo la calidad de la clasificación en términos tanto de alineación con la caja real, así como correspondencia entre clases predichas en una misma imagen.

Para la segmentación de disco y copa óptica se utilizó el optimizador Adam, en conjunto con la función de pérdida de Jaccard. Esta función puede definirse matemáticamente como:

$$Jaccard\ Loss = 1 - \frac{|Y \cap \hat{y}|}{|Y \cup \hat{y}|}$$

Donde Y representa la máscara real y $\hat{\boldsymbol{y}}$ la máscara predicha. Esta pérdida se basa en el índice de Jaccard o IoU (Intersection over Union) en dónde.

- $|Y \cap \hat{y}|$: Es la intersección de los valores verdaderos de la máscara con respecto los predichos
- $|Y \cup \hat{y}|$: Es la unión de los valores verdaderos de la máscara con respecto los predichos

El motivo principal detrás de su elección radica en que este tipo de pérdida favorece una segmentación precisa de la forma y contorno de las regiones de interés. Es particularmente útil en escenarios donde las estructuras a segmentar, como el disco y la copa óptica, presentan una morfología específica que el modelo debe aprender a replicar fielmente.

A diferencia de otras funciones como el coeficiente de Dice, la Jaccard Loss es más estricta, ya que no introduce un factor de suavizado que pueda favorecer una clase sobre otra. Lo que la convierte en una opción más robusta cuando se busca una correspondencia exacta entre la predicción y la máscara real.

4.2.2. Segmentación de micro aneurismas:

Dado que los microaneurismas son estructuras caracterizadas por tener tamaños reducidos, se decide abordar el problema a través de la aplicación directa de algoritmos de segmentación semántica en particiones de una misma imagen. Esto se hace con el objetivo de mitigar el desbalance de clases entre las estructuras a

identificar y reducir el margen de error que podría presentarse si se adoptara una metodología de encuadre y subsecuente segmentación. A continuación, se presenta el método propuesto para la segmentación de estas lesione:

- **Segmentación con el uso de U-net recurrente residual**

Kou et al. [66] presentan una metodología completa para la detección de microaneurismas en imágenes de retina, cuya etapa de segmentación se basa en una variante de U-Net (denominada **R2U-Net** [74]) que sustituye los bloques de convolución convencionales por bloques recurrentes residuales (RRCNN). La principal innovación de esta arquitectura reside en la dotación a cada nivel de codificación y decodificación de una "subred" recurrente-residual, capaz de refinar iterativamente las representaciones espaciales sin incrementar de manera drástica el número de parámetros a entrenar.

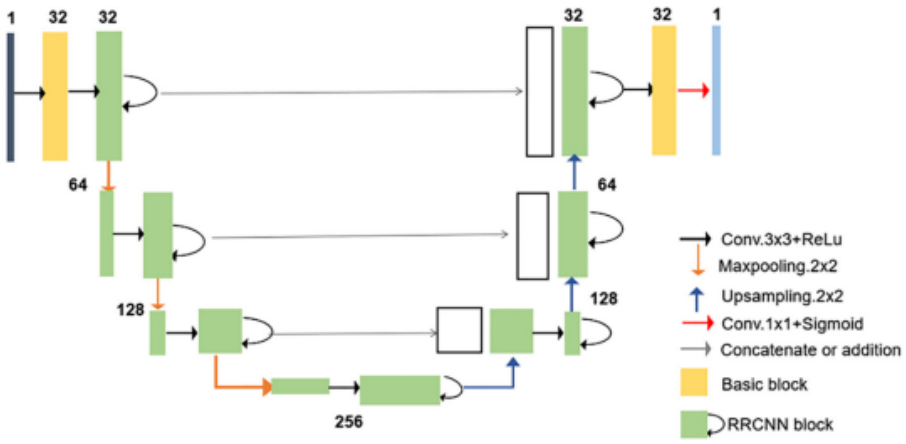


Figura 4-11 Arquitectura convolucional propuesta por Kou et al. Fecha de acceso: 4 de noviembre de 2024. Tomado de: [66]

Cada bloque RRCNN integra una capa de convolución recurrente (RCL) que, para un píxel (i,j) y el mapa de características "k" en la capa "l", calcula en el paso de tiempo "t" la salida intermedia

$$z_{ijlk} = (w_k^f)^T * u^{f(i,j)}(t) + (w_k^r)^T * v^{r(i,j)}(t-1) + b_k$$

En donde $u^{f(i,j)}(t)$ y $v^{r(i,j)}(t-1)$ son el input de la capa anterior, y la salida recurrente en el paso t-1 Respectivamente. Con w_k^f y w_k^r siendo los valores de los pesos de una capa de convolución estándar y de la capa RCL del k-esimo mapa de características, y con b_k como el sesgo adicionado a la operación.

A continuación, se aplica una función de activación ReLU para obtener la salida del bloque para el tiempo t:

$$F(x_l, w_l)(t) = f\left(z_{ijk}(t)\right) = \max\left(0, z_{ijk}(t)\right)$$

Este proceso recurrente se realiza de forma iterativa "T" veces para la refinación progresiva de las características sin la adición de capas convolucionales adicionales.

Al final del bloque se añade un "atajo" residual, sumando el input original x^l a la salida recurrente final $F(x_l, w_l)(T)$:

$$x_{l+1} = x_l + F(x_l, w_l)(t)$$

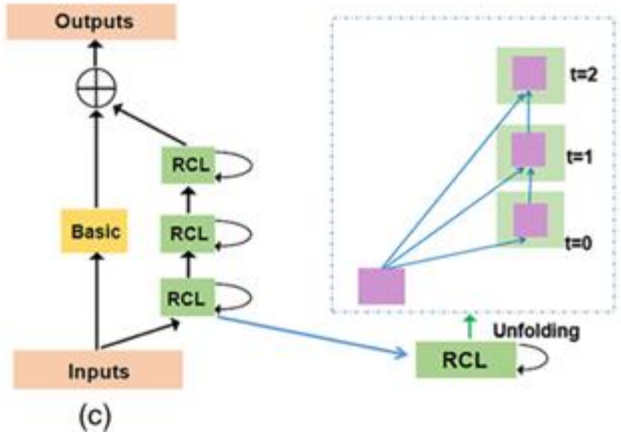


Figura 4-12 Ilustración de bloque RRCNN con capa RCL. Fecha de acceso. 21 de abril de 2025. Tomado de: [66].

- **Detalles de implementación**

  Para el presente problema de segmentación, se empleó el optimizador Adam, combinando dos funciones de pérdida en distintas etapas del entrenamiento. Inicialmente, se utilizó una función de pérdida Focal Tversky con el objetivo de priorizar la detección precisa de microaneurismas, dada su escasa representación espacial en las imágenes y la alta probabilidad de ser ignorados por el modelo durante el aprendizaje.

  Una vez alcanzada una detección adecuada de las regiones de interés, se sustituyó esta función por la pérdida basada en el coeficiente de Dice, con el propósito de refinar las predicciones morfológicas de las lesiones segmentadas. Esta estrategia secuencial permitió dirigir el aprendizaje hacia los casos más difíciles y minimizar errores críticos (como los falsos negativos), y al mismo tiempo, mejorar la coherencia morfológica de las regiones segmentadas, favoreciendo la precisión en los contornos y la forma de las lesiones.

- **Función de perdida Focal Tversky**

La función de pérdida Focal Tversky fue diseñada específicamente para abordar problemas de segmentación con clases altamente desbalanceadas, donde las regiones de interés (por ejemplo, lesiones médicas) representan una fracción muy pequeña del total de la imagen. Su propósito es doble: enfocar el aprendizaje en las muestras difíciles (donde el modelo comete mayores errores), y ajustar la sensibilidad hacia los falsos positivos y falsos negativos mediante una ponderación explícita.

La función introduce tres hiperparámetros que permiten personalizar la penalización hacia los errores en función del dominio del problema y del tipo de clase prioritaria. La Focal Tversky Loss (FTL) se define como:

$$FTL = (1 - TI)^{\gamma}$$

Donde TI es el índice de tversky, definido como:

$$TI = \frac{TP}{TP + \alpha FN + \beta FP}$$

Siendo

- TP: los valores predichos correctamente para la clase positiva
- FN: los valores predichos de forma errónea para la clase negativa
- FP: Los valores predichos de forma errónea para la clase positiva

Y los hiperparametros de la función:

- α: Coeficiente de penalización para los falsos negativos. Un valor alto prioriza la reducción de este tipo de error.
- β: Coeficiente de penalización para los falsos positivos. Un valor alto fuerza al modelo a ser más conservador al predecir positivos.
- γ: Parámetro de focalización. A medida que γ aumenta, se incrementa el enfoque del modelo en los casos más difíciles (en donde 1-TI es alto), reduciendo el impacto de los ejemplos bien clasificados en la función de pérdida. Esto permite reforzar el aprendizaje en regiones donde el modelo está fallando más.

- **Función de perdida Dice**

La pérdida de Dice está basada en el coeficiente de similitud de Dice, una métrica comúnmente utilizada en tareas de segmentación para cuantificar la superposición entre una predicción y la verdad de

terreno. Esta función penaliza explícitamente las discrepancias entre las áreas segmentadas por el modelo y las regiones reales, lo que la hace adecuada para refinar la precisión morfológica en la segmentación final.

Dada una máscara con los valores reales de una imagen Y y una predicción $\hat{y}$, la función de perdida de dice se define como:

$$Perdida\ Dice = 1 - \frac{2 * |Y \cap \hat{y}|}{|Y| + |\hat{y}|}$$

Donde:

- $|Y \cap \hat{y}|$: Representa la intersección (valores coincidentes) entre la máscara real y los valores predichos.
- $|Y| + |\hat{y}|$: Representa La suma total de las áreas segmentadas y reales.

4.2.3. Segmentación de drusas:

Para la segmentación de drusas se va a utilizar un procedimiento de segmentación en dos pasos [65] que va a consistir en el procesamiento de la imagen completa, y el posterior refinamiento de los mapas probabilísticos obtenidos para la segmentación de las áreas de interés en parches de determinado tamaño.

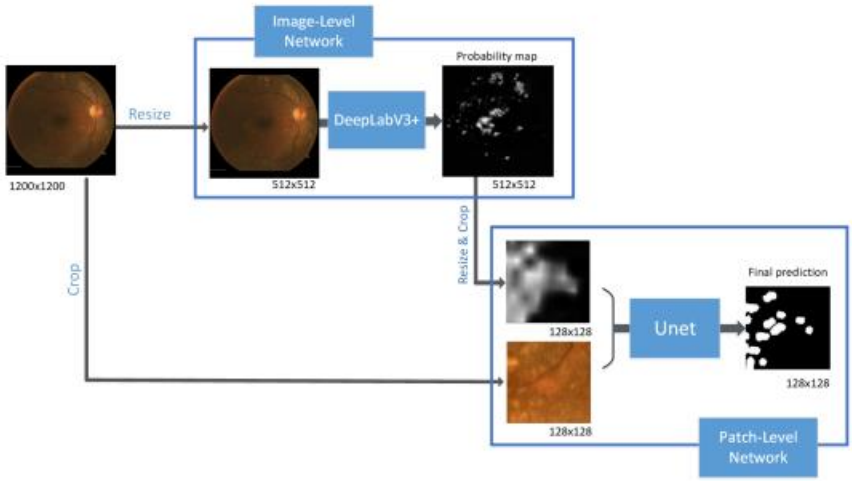


Figura 4-13 Ilustración del proceso para la segmentación de drusas. Tomado de [65]

- **Red a nivel de imagen:**

Para esta etapa se va a utilizar una red neuronal DeepLabV3+, esta arquitectura diseñada para tareas de segmentación semántica analiza de mejor forma información espacial de distintas dimensiones a través del uso de

convoluciones dilatadas para el manejo del campo receptivo y la resolución del mapa de características sin incrementar el número total de parámetros a entrenar.

El objetivo final de esta etapa es generar un mapa probabilístico que se ingerirá después por una red U-net la cual estará encargada de refinar las predicciones obtenidas por el modelo inicial.

- **Convoluciones "Atrous"**

  Suponga que se tiene un tensor unidimensional, para cada posición "i" de la salida "y" con filtro "w", una convolución "atrous" es aplicada sobre el mapa de características inicial "x" de la forma:

$$y[i] = \sum_{k} x[i + r \cdot k]w[k]$$

  En donde "r" define la dilatación del kernel que se pasa sobre el mapa de características de entrada.

  Esta definición busca separar el tamaño del kernel con respecto a la información espacial que este extrae, sin modificar la cantidad de parámetros utilizados en cada convolución.

- **Atrous spatial pyramid pooling (ASPP)**

  Los atrous spatial pyramid pooling (ASPP) son bloques convolucionales que aplican de forma paralela convoluciones "atrous" variando el parámetro de dilatación "r" junto con el tamaño del kernel. Esto con el propósito de extraer características en diferentes escalas. El proceso termina con la concatenación ponderada de los mapas de características obtenidos por cada filtro.

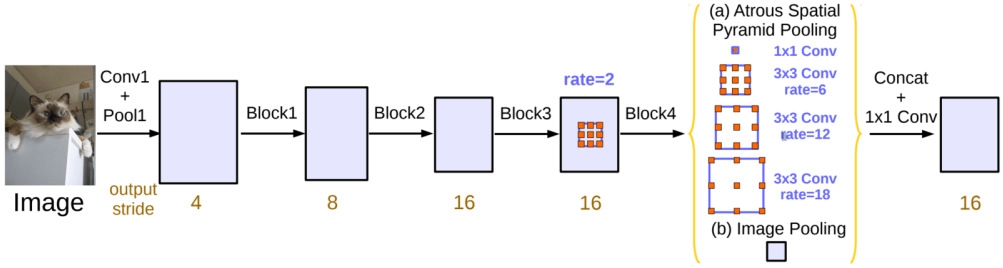


Figura 4-14 Arquitectura deeplabv3. Fecha de acceso: 8 de abril de 25 Tomado de: https://learnopencv.com/deeplabv3-ultimate-guide/.

- **Adición de modulo decoder**

  DeepLabv3+ añade un módulo decoder a la arquitectura original, el cual utiliza conexiones de tipo *skip* entre el encoder (basado en DeepLabv3) y el decoder. Esto permite combinar

características de alta resolución (extraídas en capas superficiales del backbone) con las características semánticas más profundas y contextuales generadas por el módulo ASPP, concatenándolas con las salidas de este último. De esta forma, se mejora significativamente la precisión en los bordes de los objetos segmentados y se obtiene una segmentación más detallada, especialmente útil en estructuras pequeñas o con contornos suaves.

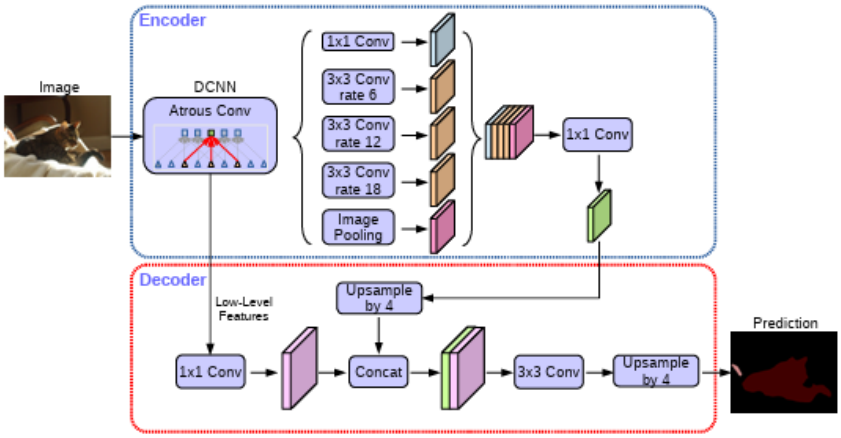


Figura 4-15 Arquitectura deeplabv3+ con la adición del módulo decoder. Fecha de acceso: 21 de abril de 2025. Tomado de: [75].

- **Detalles de implementación**

Para la red de imagen completa se utilizó un optimizador Adam junto con una función de pérdida Binary Cross-Entropy Ponderada (weighted BCE), con el objetivo de solventar el problema del desbalance de clases inherente a la detección de drusas en imágenes de fondo de ojo.

Esta función de perdida se define como:

$$L_{wBCE} = -\frac{1}{N}\sum_{i=0}^{N}[wy_i \log(x_i) + (1-y_i)\log(1-x_i)]$$

Donde:

- N: es el número total de pixeles en la imagen
- $y_i \in \{0,1\}$ representa la etiqueta binaria verdadera para un pixel i
- $x_i \in (0,1)$ representa la probabilidad predicha por el modelo para un pixel i

- $w > 1$ es un hiperparametro que amplifica el impacto de los falsos negativos (drusas)

Aunque otras funciones como Dice loss, o focal los pueden ser utilizados como alternativas, de acuerdo con el estudio [65], estas funciones pueden favorecer que los mapas generados no estén lo suficientemente poblados, dado que regiones ambiguas (como lo pueden ser zonas con drusas suaves) son más propensas a clasificarse como positivas.

- **Consolidación a detalle de áreas con probabilidad de drusas.**

Para esta etapa se empleó una red neuronal U-Net para estimar las áreas correspondientes a las drusas. En la fase inicial, se recortarán las imágenes en parches de 128x128 de aquellas zonas en las que el modelo de imagen completa indicó la posible presencia de drusas. Posteriormente, se concatenará el mapa probabilístico con el tensor de la imagen recortada del fondo de ojo, lo que proporcionará una guía para orientar al modelo hacia las áreas de interés. El proceso consiste en extraer áreas específicas durante la etapa de codificación y, en la fase final de concatenación y upsampling de la red de segmentación, reensamblar el mapa de características.

- **Detalles de implementación de red por parches**

Para la red por parches de la imagen se utilizó un optimizador ADAM junto con una función de perdida Focal tversky, definida en la sección (4.2.2).

4.3. Construcción del modelo discriminativo final

La siguiente sección detalla el proceso de construcción del modelo final de clasificación, abarcando desde la definición de las entradas esperadas hasta el diseño arquitectónico y la implementación de técnicas específicas orientadas a la obtención de los diagnósticos finales.

4.3.1. Concatenación de mapas de segmentación

Una vez generados los mapas de segmentación correspondientes a cada estructura anatómica y patológica, se implementa un **protocolo de concatenación horizontal** que combina las representaciones gráficas con la imagen original del fondo de ojo. Este proceso permite una comparación visual integrada, como se ejemplifica en la Figura 4.14.



Figura 4-16 Imagen de fondo de ojo concatenada. Elaboración propia

Para este ejemplo es posible ver de izquierda a derecha la concatenación de las siguientes características:

5. Fondo de ojo original
6. Mapa de segmentación de disco y copa óptica
7. Mapa de segmentación de drusas
8. Mapa de segmentación de microaneurismas

El orden secuencial se mantiene incluso ante la ausencia de elementos segmentados (ej: casos sin drusas detectables o con obstrucción del disco óptico). Esta aproximación se utiliza para tener un formato estandarizado y legible para la interpretación de múltiples muestras con patologías variadas partiendo de la imagen original como referencia para los subsecuentes mapas de segmentación.

4.3.2. Construcción de arquitectura de clasificación compuesta

La idea de esta sección es utilizar un enfoque de ensamble como clasificador final que nos dé las probabilidades más altas de que una imagen de fondo de ojo presente determinada enfermedad o enfermedades, para este modelo, se planean utilizar las siguientes arquitecturas:

- **VGG 19**

  VGG 19 es un modelo de la familia VGG que cuenta con 19 capas convolucionales (de las cuales 4 están conectadas completamente) que utiliza activación Relu, y que es capaz de identificar 1000 imágenes de 1000 categorías distintas.

  La serie de redes neuronales VGG 19 está caracterizada por aplicar operaciones de convolución sencilla para un kernel fijo 3X3. La red extrae las características y reduce la dimensión de este mapeo, manteniendo información detallada en cada bloque.

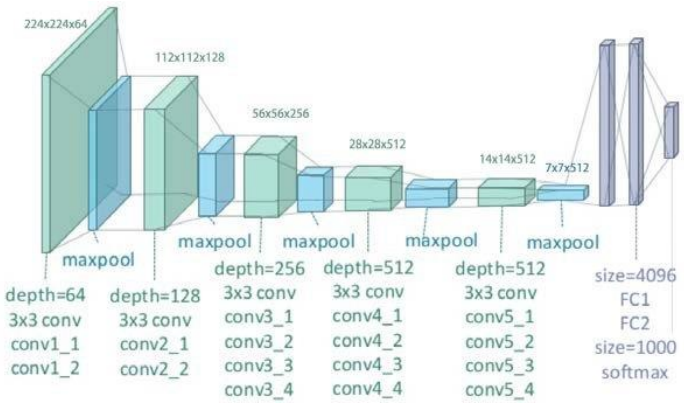


Figura 4-17 Arquitectura VGG 19 ilustrada. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://www.researchgate.net/figure/Illustration-of-the-network-architecture-of-VGG-19-model-conv-means-convolution-FC-means_fig2_325137356

- **ResNet 152**

ResNet-152 pertenece a la familia de redes neuronales residuales y cuenta con 152 capas profundas. Utiliza bloques residuales, que permiten un flujo continuo de gradientes a través de la red, resolviendo el problema del desvanecimiento que suele presentarse en redes muy profundas. Este modelo es altamente eficaz para la clasificación de imágenes complejas, como las imágenes de fondo de ojo, ya que puede captar características profundas sin perder información relevante a través de las capas. ResNet-152 es especialmente útil en tareas de mejora de imágenes y en la extracción de características complejas dentro de las mismas.

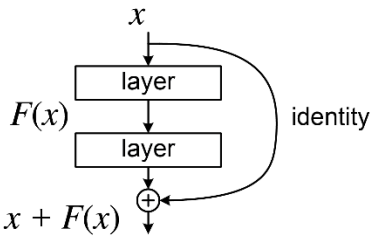


Figura 4-18 Ilustración de un bloque convolucional para redes residuales. Fecha de acceso: 4 de noviembre de 2024. Tomado de: https://es.wikipedia.org/wiki/Red_neuronal_residual

Esta arquitectura, propuesta en 2015, supera el límite de las redes que se estancan al intentar extraer características muy específicas de imágenes complejas. En la ilustración anterior, la entrada x se procesa mediante filtros convolucionales que extraen características. Una vez se llega a la función de activación $F(x)$, se suma la entrada original a la función de activación con el objetivo de mantener la importancia del gradiente en las capas más profundas.

- **EfficientNet V2**

La familia de modelos EfficientNet consiste en redes neuronales convolucionales optimizadas mediante una técnica de escalado compuesto, la cual ajusta de forma conjunta el ancho, la profundidad y la resolución de la red en proporción a un factor definido. Esta arquitectura se destaca por su eficiencia computacional, ya que logra un alto rendimiento en conjuntos de datos como "imagenet" con menos parámetros en comparación con modelos más grandes, como VGG o ResNet, los cuales no garantizan mejoras en la capacidad de clasificación a través de la adición progresiva de capas. El enfoque de escalado de EfficientNet permite un uso óptimo de los recursos sin sacrificar la precisión, lo que lo convierte en una opción ideal para la detección de patologías en imágenes médicas.

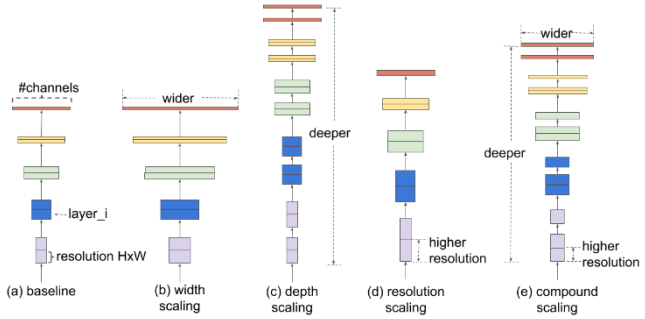


Figura 4-19 Metodología de escalado ilustrada. Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://huggingface.co/google/efficientnet-b0

EfficientNetV2 introduce importantes mejoras respecto al modelo original propuesto en el paper de EfficientNet. En esta nueva versión se implementa un escalado no uniforme de la red, ajustando de manera diferenciada la profundidad y el ancho, lo que optimiza el uso de los recursos computacionales. Además, incorpora técnicas de regularización avanzadas, como RandAugment y MixUp, que refuerzan la generalización del modelo y ayudan a evitar el sobreajuste durante el entrenamiento, esta nueva generación cuenta con 3 versiones. Small, medium, y large.

- **Capa softmax**

Después de aplicar los modelos base (VGG 19, ResNet 152 y EfficientNet V2-Small), los resultados de clasificación de cada uno se pasan a una capa softmax en el modelo de ensamble. La capa softmax se utiliza para convertir las salidas de los modelos en probabilidades unitarias. Para cada muestra se clasificará la imagen en base a la clase a la que se atribuya la mayor probabilidad.

- **Detalles de implementación**

Para el modelo de clasificación se utilizó el optimizador Adam junto con una función de pérdida de entropía categórica cruzada, la cual mide las diferencias entre la distribución de probabilidad predicha por el modelo y la distribución real para cada muestra.

Esta función se define como:

$$L(y, \hat{y}) = \sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

Donde:

- y_i es el valor verdadero de la clase para la muestra
- $\hat{y}_i$ es la probabilidad predicha para la clase i
- C es el número de clases

Dado que esta función de pérdida trabaja con probabilidades en lugar de logits (salidas en crudo del modelo), es necesario aplicar previamente una función softmax (definida en la sección 4.2.1). Esta transformación convierte los logits en una distribución de probabilidad sobre las clases, lo que permite calcular correctamente la entropía cruzada en función de la clase verdadera.

4.4. Diseño de sistema de diagnóstico final

Una vez validadas las capacidades de clasificación del modelo final, se procederá con la obtención de mapas diagnósticos que permitan entender qué secciones de la imagen están teniendo mayor influencia sobre la clasificación hecha, y de esta manera corroborar los beneficios aportados por los pasos previamente implementados.

4.4.1. Mapeo de características con Grad-CAM

Grad-CAM (Gradient-weighted Class Activation Mapping) es una técnica de visualización que ayuda a entender cómo funcionan los modelos de redes neuronales convolucionales al resaltar las áreas de la imagen que tienen mayor influencia en las decisiones del modelo. En el sistema actual, se utilizarían los gradientes de las capas convolucionales finales para crear un mapa de calor que se superpone a la imagen del fondo de ojo. Esto permite identificar las zonas en las que el modelo se concentra para detectar patologías.

- **Metodología:**

  Grad-CAM adopta un enfoque que permite identificar las regiones de una imagen que influyen de manera positiva en la decisión de clasificación de una red neuronal. Para ello, se procesa la imagen a través de la red y se calculan los gradientes que indican la contribución de cada píxel a la puntuación de una determinada clase. sea Y^c el puntaje que tiene determinada

clase “c” y sea A_{ij}^{k} el mapa de activación del canal “k” en la última capa de una red convolucional. El gradiente se calcula como:

$$\frac{\partial Y^{c}}{\partial A_{ij}^{k}}$$

Posteriormente se hace un promedio local (global average pooling) el cual genera una serie de valores escalares que funcionaran como pesos para cada canal “k” con respecto a una categoría “c”:

$$w_{k}^{c} = \frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial Y^{c}}{\partial A_{ij}^{k}}$$

Finalmente, se efectúa una combinación lineal de los mapas de activación ponderados por estos pesos, y se aplica la función ReLU para conservar únicamente las contribuciones positivas en el mapa de saliencia resultante:

$$L_{Grad-CAM}^{c} = ReLU(\sum_{k} \alpha_{k}^{c} A^{k})$$

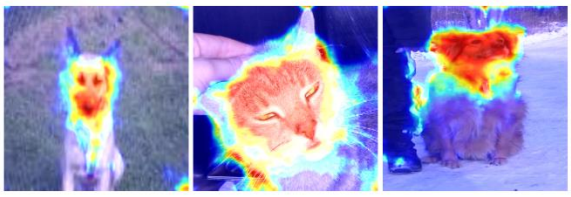


Figura 4-20 Mapeo de clases con Grad Cam Fecha de acceso: 4 de noviembre de 2024 Tomado de: https://towardsdatascience.com/understand-your-algorithm-with-grad-cam-d3b62fce353

4.4.2. Grad-CAM ++

Grad-CAM++ es una mejora respecto a su predecesora, ya que esta nueva versión busca optimizar la visualización del mapa de activaciones final mediante el siguiente proceso:

Como se recordará (sección 4.4.1), en Grad-CAM el peso asignado a un mapa de saliencia se calcula como el promedio de los gradientes sobre el número de píxeles del mapa de características “Z”. Sin embargo, esta definición puede estar sesgada y “desvanecer” áreas de activación en aquellos mapas de características donde las regiones con mayor activación sean pequeñas en comparación con el resto del mapa, lo que implica perder información espacial importante en la representación.

Para solventar este problema, en [74] se propone una reformulación en la obtención de los pesos para la combinación lineal final mediante la siguiente metodología:

Se parte de la idea de que, en ciertos puntos del mapa, el gradiente $\frac{\partial y^c}{\partial A_{ij}^k}$ alcanza un valor característico (por ejemplo, igual a 1 cuando A_{ij}^kcumple cierto criterio de activación). A partir de esta base, se redefine el peso del canal k para una categoría c de la siguiente manera:

$$w_k^c = \sum_i \sum_j \alpha_{ij}^{kc} ReLU(\frac{\partial Y^c}{\partial A_{ij}^k})$$

Donde los coeficientes α_{ij}^{kc} se calculan de modo que ponderen la contribución individual de cada píxel, tomando en cuenta derivadas de orden superior (por ejemplo, segundos y terceros) de respecto A_{ij}^k. Esta formulación permite que incluso las áreas con activación relativamente pequeña, pero significativas para la clasificación final, tengan la influencia adecuada en la generación del mapa de saliencia. Esta formulación se representa a través de:

$$\begin{cases} \alpha_{ij}^{kc} = \dfrac{1}{\sum_{i,m} \dfrac{\partial y^c}{\partial A_{ij}^k}} \; Si \; \dfrac{\partial y^c}{\partial A_{ij}^k} = 1 \\ 0 \; E.O.C \end{cases}$$

Como se puede ver, esta reformulación busca dar mayor peso a la contribución individual de cada píxel, evitando que se pierda información espacial relevante en mapas de activación con zonas pequeñas pero importantes.

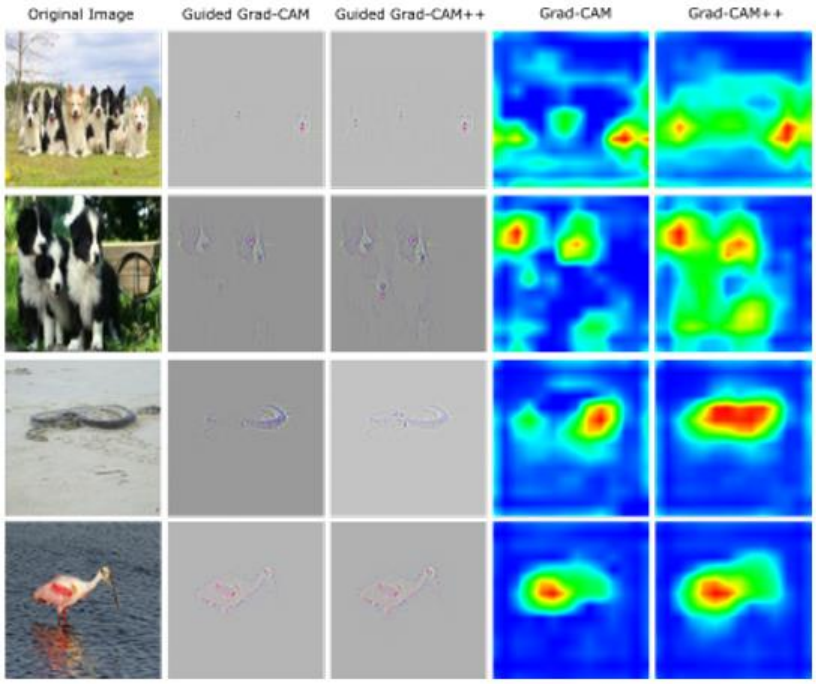


Figura 4-21 Comparación de un mapa de características Grad-CAM con respecto a un Grad-CAM++. Tomado de [74]

4.4.3. Sintomatología relacionada con retinopatía diabética y degeneración macular senil

Para el análisis de patologías como la retinopatía diabética y la degeneración macular senil, se emplearán los mapas de segmentación generados en etapas previas. En el caso de la retinopatía diabética, se aplicarán los métodos de interpretabilidad ya explicados para verificar que la inclusión de estos mapas influye significativamente en las decisiones del modelo. Respecto a la degeneración macular senil, el enfoque estará en la detección de drusas (acumulaciones de depósitos lipídicos en la retina). Estos mapas permitirán resaltar las zonas afectadas y optimizarán la evaluación médica al ofrecer una visualización precisa de las lesiones oculares de interés.

El objetivo final de esta sección es implementar un mapeo diagnóstico que permita comprender, a nivel de píxel, cómo se activan las funciones en cada capa de la red neuronal durante el proceso de interpretación. Esto facilitará la trazabilidad de las decisiones del modelo y brindará una perspectiva detallada de los patrones asociados a cada patología.

4.5. Flujo final ilustrado

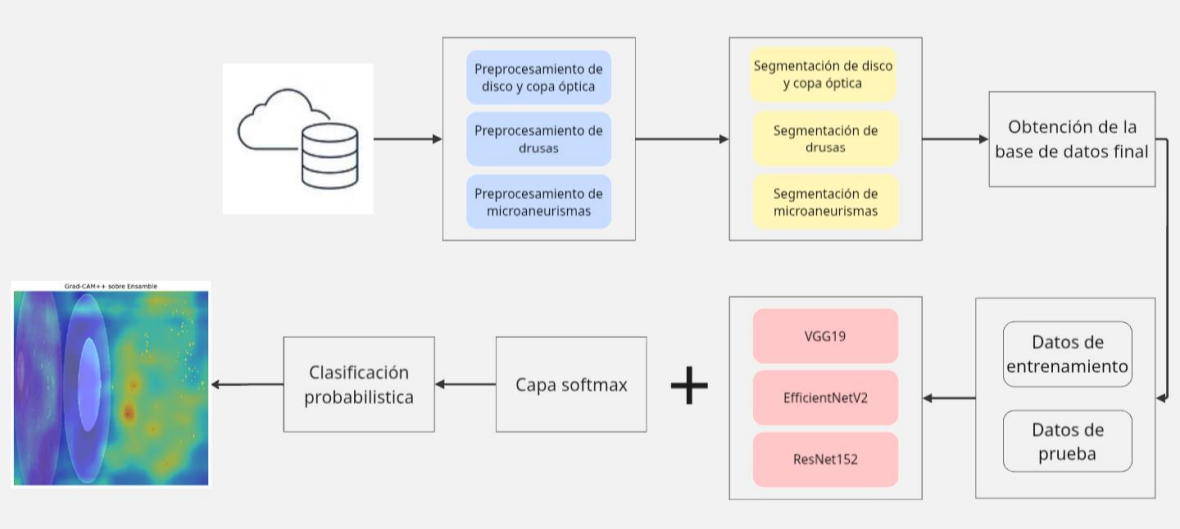


Figura 4-22 Flujo final ilustrado: Elaboración propia

5. RESULTADOS

Para los resultados preliminares, se decidió crear las funciones correspondientes a cada etapa para un dataset inicial, con el fin de definir el flujo por el cual pasarán los datos antes de ser procesados en cada modelo. Esto permite anticipar el costo computacional que requeriría aplicar cada etapa sobre el total de las imágenes.

5.1. Insumos a nivel de software utilizados

Todas las operaciones realizadas a continuación, se construyeron con la herramienta de software Python. Las principales librerías a utilizar incluyen:

- **Pandas:**

  Pandas se utilizó para manipular los conjuntos de datos, lo cual incluyó la aplicación de filtros y la asignación de dataframes para los conjuntos de datos extraídos.

- **Os:**

  La librería os se utilizó para explorar directorios a través de rutas desde el entorno de programación, los directorios principales se crearon utilizando esta librería, así como la asignación de la información de cada dataset.

- **Numpy:**

  Se utilizó NumPy para transformar las imágenes originalmente en crudo, en matrices con valores numéricos de intensidad, sobre los cuales se pueden realizar operaciones matemáticas.

- **Opencv:**

Se utilizó OpenCV para la manipulación de imágenes en la etapa de preprocesamiento los procedimientos incluyen: la aplicación de filtros, la ecualización del histograma y la medición de la calidad de las imágenes resultantes con los tratamientos aplicados.

- **Pytorch:**

  Se utilizó pytorch con el objetivo de tener un mayor control sobre las redes a utilizar, así como para la implementación de determinadas fases del pos procesamiento, y la construcción del modelo final.

- **Keras:**

  La librería keras se utilizó con el objetivo de construir algunos de los modelos de segmentación para lesiones siguiendo buenas prácticas como aumento de datos y regularización de los modelos.

5.2. Filtrado de documentos y selección de muestras

En esta sección inicial, se aplica un filtro a los documentos que contienen las imágenes, basado en las enfermedades determinadas anteriormente. Es importante recordar que el criterio de selección de las enfermedades (retinopatía diabética, glaucoma y degeneración macular senil) se debe a que la manifestación de cada una de estas patologías es única, lo que facilita identificar los factores clave para un diagnóstico preciso.

A continuación, se documentan los pasos específicos para la construcción de los directorios finales:

- **Filtrado Individual de las Bases de Datos**

  Este paso implica la filtración de datos correspondientes a las patologías seleccionadas, así como la inclusión de una muestra de control que contenga imágenes de ojos sanos.

- **Establecimiento de Directorios por tipos de imágenes recolectadas**

  Dado que en el posprocesamiento se aplicarán metodologías de estimación de áreas para las imágenes extraídas, es importante contar con datasets que incluyan imágenes anotadas por especialistas en relación con las áreas a estimar. Esto permitirá definir un margen de error sobre el cual validar los resultados de las técnicas aplicadas. Las anotaciones pueden presentarse en distintos formatos, como imágenes superpuestas con el área de interés resaltada o escalas de intensidad para los píxeles que componen la zona. Por esta razón, se definieron dos directorios para clasificar las imágenes con y sin anotaciones. A continuación, se construye una tabla con los filtros aplicados a cada dataset:

Nombre	Derbi_Hackathon_Retinal_Fundus_Image_ Dataset [50]	ACRFAODAWDCMFERLRLASMAGD Versión 4 [51]
Filtrado	El dataset contiene una serie de carpetas que almacenan las patologías específicas que afectan al ojo, por lo que se filtraron aquellas relacionadas con cada enfermedad: Afecciones específicas de diabetes: ['0.3.DR1', '1.0.DR2', '1.1.DR3', '20.Massive hard exudates', '22.Cotton-wool spots', '27.Laser Spots', '29.0.Blur fundus without PDR', '29.1.Blur fundus with suspected PDR'] Afecciones específicas de glaucoma: ['Large optic cup', 'Possible glaucoma', 'Optic atrophy'] Afecciones específicas de AMD: ['Maculopathy']	Este conjunto de datos contiene información de escáneres de fondo de ojo y tomografías de coherencia óptica de varios grupos de personas. El conjunto está dividido en directorios de la siguiente forma: 1. OD (Optic Disc): Este directorio contiene información de ojos sanos y ojos con glaucoma. Cada registro de persona incluye anotaciones relacionadas con la estimación del disco y la copa óptica. Se eliminaron los escaneos realizados con tomografía de coherencia óptica y se almacenaron en el directorio de conjuntos de datos con anotaciones. 2. Mácula: Este directorio contiene 6 carpetas correspondientes a 3 enfermedades diferentes que afectan la mácula. Se seleccionaron y almacenaron los directorios relacionados con la degeneración macular senil, y se eliminaron los fotogramas de tomografía de coherencia óptica de las imágenes. Estos fueron enviados al directorio de imágenes con anotaciones.

Drishti-GS1 [62]	ODIR [52]	E optha [60]
Este conjunto de datos cuenta con una jerarquía de directorios organizada de la siguiente manera: 1. Directorio con imágenes. 2. Directorio con especificaciones del área de interés para la enfermedad (disco y copa óptica). Dentro de los archivos con especificaciones hay: 2.1. Intensidad de píxeles en los 3 canales de la imagen. 2.2. Mapeo del disco y la copa óptica. Se conservaron los archivos y se enviaron junto con los dataframes que contienen la información anotada sobre las patologías.	Este conjunto de datos también incluye un dataframe secundario que contiene información de las personas cuyos ojos fueron evaluados para consolidar los registros. Esta tabla cuenta con una serie de banderas que identifican la patología que afecta a cada paciente. En algunos casos, un paciente puede presentar dos afecciones en un mismo ojo, por lo que, para las pruebas iniciales, solo se tomaron registros con una única aparición en alguna de las banderas. Para pruebas posteriores, se utilizará el dataset multiclase. ocurrir la situación en que un mismo paciente presente dos afecciones en un mismo ojo, por lo que, para las pruebas iniciales, se tomaron solamente aquellos registros que tuvieran una aparición única en alguna de las banderas. Para pruebas posteriores, se hará utilización del dataset multiclase.	Este dataset contiene información sobre retinopatía diabética y cuenta con afecciones segmentadas manualmente. Se almacena sin realizar un filtro más exhaustivo.

Retinal Fundus Multi-disease Image Dataset 2.0 [61]	ACRIMA [53]	MESSIDOR-2 [54]	DRIVE [55]	REFUGE2 [57]
Para este conjunto de datos, se pasa por una serie de tablas que contienen información sobre la patología presente en cada ojo. Se realiza un filtro para seleccionar los registros que tienen únicamente retinopatía diabética o degeneración macular senil, y se almacenan los datos para las pruebas preliminares. Al igual que en ODIR, estos datos pasarán por un segundo filtro en etapas posteriores, con el objetivo de tratarlos desde un enfoque multiclase.	Este dataset incluye imágenes recortadas alrededor de la copa y el disco óptico. El filtro aplicado solo separa el conjunto de datos en ojos con y sin patologías, definiendo un espacio específico separado del resto, ya que su utilidad no puede extenderse más allá del pos procesamiento.	Para este conjunto de datos no se realiza ningún filtro, excepto su colocación en la carpeta de datos para pruebas sin anotaciones. El conjunto de datos incluye registros con y sin retinopatía diabética.	En este dataset no se aplican filtros adicionales, y la información se pasa directamente al directorio de datasets con anotaciones.	Este dataset contiene información sobre ojos sanos y afectados por glaucoma, con el propósito de probar algoritmos de clasificación del área de interés y contrastar los resultados con las imágenes anotadas. No se aplicó ningún filtro adicional, salvo su almacenamiento junto con los datasets anotados.

INSPIRE-AVR [59]	DR HAGIS [58]	EYE PACS [67]	OCT-AND-EYE-FUNDUS-DATASET [68]
En este dataset tampoco se realizó ningún filtro; fue almacenado en el directorio que contiene datos con anotaciones.	El dataset contaba con la información separada de la siguiente manera: • Imágenes 1 a 10: Glaucoma • Imágenes 11 a 20: Hipertensión • Imágenes 21 a 30: Retinopatía diabética • Imágenes 31 a 40: AMD Se aplicó el filtro para traer solo las patologías relacionadas con la investigación, y se eliminó el resto, se re direccionaron los archivos al directorio con datasets anotados.	Este conjunto de datos contiene información de un concurso de Kaggle cuyo objetivo era evaluar la capacidad de modelos de clasificación para detectar retinopatía diabética. La información se almacenó directamente	Para este conjunto de datos se almacena información de tomografías de coherencia óptica e imágenes de fondo de ojo de varios pacientes. Se filtraron las imágenes para conservar solo las relacionadas con retinopatía diabética.

Dataset from fundus images for the study of diabetic retinopathy [69]	drusen_segmentation [74]	ORIGA [78]	G1020 [78]
Para este conjunto de datos no se realizó ningún filtro adicional, salvo la separación entre las imágenes con retinopatía diabética y las de ojos sin afección alguna.	Este conjunto de datos contiene ojos con áreas afectadas por drusas. No se aplica ningún filtro dado que todas las muestras son de interés.	Este conjunto de datos contiene ojos con áreas afectadas glaucoma. No se aplica ningún filtro dado que todas las muestras son de interés.	Este conjunto de datos contiene ojos con áreas afectadas por glaucoma. No se aplica ningún filtro dado que todas las muestras son de interés.

Tabla 3 Descripción de modificaciones a los dataset

- **Muestras resultantes**

  Para el primer directorio con imágenes sin anotaciones, se obtuvieron: 1.769 imágenes para retinopatía diabética, 1.915 imágenes para glaucoma, y 487 para degeneración macular senil.

Para el segundo directorio con imágenes con anotaciones, se obtuvieron: 90 imágenes con retinopatía diabética, 87 para glaucoma, y 160 para Degeneración macular senil, y 1.200 imágenes para segmentación de disco y copa óptica.

- **Limitaciones en la carga de los datos al entorno de trabajo**

  Durante la obtención de las imágenes para el procesamiento se contó con problemas de descarga de varios de los repositorios, esto debido a limitaciones en la capacidad de almacenamiento del equipo usado, aunque varios de los datasets contaban con repositorios disponibles en *kaggle*, permitiendo así su descarga sin afectar el almacenamiento del equipo, hay otros que tuvieron que ser descargados en forma de archivos comprimidos.

5.3. Preprocesamiento para segmentación de disco, y copa óptica

- **Filtro de mediana**

  Como se mencionó en la sección 4.1.1, el filtro de mediana aplica una ventana móvil que recorre toda la imagen. Para cada posición de la ventana, se toman los valores de intensidad de los píxeles en una escala única (en este caso, escala de grises) y se ordenan de menor a mayor. El valor que se encuentra en la posición central de esta lista ordenada, (es decir, la mediana), reemplaza el valor del píxel central de la ventana. Este proceso se repite para cada píxel de la imagen, ayudando a reducir el ruido sin difuminar significativamente los bordes.

  Ejemplo:

  Dada una venta de 3 x 3 píxeles centrada en un píxel (i,j), los valores de los píxeles se denotan como:

$$\begin{matrix} p_{i-1,j-1}, & p_{i-1,j} & p_{i-1,j+1} \\ p_{i,j-1} & p_{i,j} & p_{i,j+1} \\ p_{i+1,j-1} & p_{i+1,j} & p_{i+1,j+1} \end{matrix}$$

  Estos valores se ordenan de acuerdo a su intensidad, y se toma el valor del medio para hacer el reemplazo, esto puede ser visto como

$$P'_{i,j} = \text{Mediana}(\text{Arreglo ordenado de pixeles})$$

- **CLAHE**

  Una vez aplicados los filtros para hacer la remoción de ruido de la imagen, se pasa el Optimizador de contraste (CLAHE por sus siglas en inglés), para así obtener una imagen con mayor detalle sobre las lesiones existentes para cada enfermedad, puede encontrar una descripción más detallada del funcionamiento de este algoritmo en la sección 4.1.1 CLAHE (Contrast Limited Adaptive Histogram Equalization).

Formalmente, se puede definir la ecualización de histograma por CLAHE, como la reasignación de valores de pixeles a través de una serie de pasos descritos a continuación:

1. **División de la imagen en fragmentos:**

   La imagen se divide en bloques de tamaño m x n para hacer la construcción y ecualización del histograma.

2. **Limitación del histograma:**

   Se establece un valor límite (clip limit) que determina el número máximo de píxeles que pueden contribuir a cualquier valor de intensidad en el histograma. Cuando el número de píxeles de un valor de intensidad supera este límite, se redistribuyen uniformemente los excedentes entre los otros valores de intensidad en el histograma. Esto ayuda a prevenir la amplificación excesiva del ruido en áreas con alta concentración de píxeles. La fórmula para hallar el límite de histograma se puede definir como:

$$\beta = \frac{MN}{L}\left(1 + \frac{\alpha}{100}(s_{max} - 1)\right)$$

   En donde:

   - β Es el límite del histograma.
   - M N es el número de pixeles en cada región
   - L es el valor máximo en escala de grises (256)
   - α es un factor de recorte entre 0 y 100 (representa que tan exigente debe ser con respecto al recorte de valores en el histograma)
   - S_{max} es la pendiente máxima que va a tener el histograma redistribuido.

   La salida final es una imagen reconstruida con información recuperada del archivo en crudo inicial.

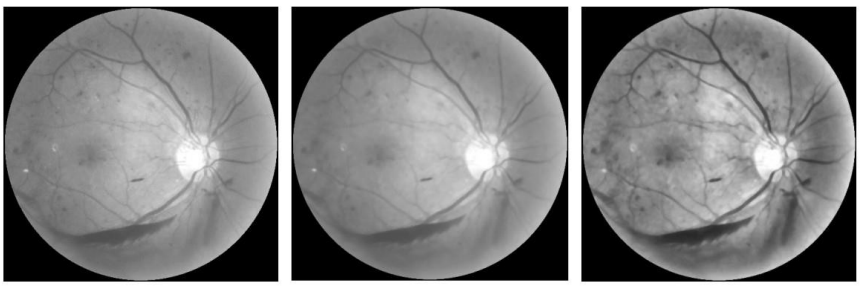


Figura 5-1 conducto de pre procesamiento inicial para las imágenes en escala de grises. Elaboración propia.

5.4. Preprocesamiento de drusas

Para las drusas se realiza un proceso de recorte de la imagen dado que, a diferencia del proceso de segmentación para disco y copa óptica, segmentar drusas es un problema desbalanceado el cual se busca mitigar a través de la siguiente serie de pasos:

1. **Obtención de mapa probabilístico para la imagen completa de las drusas**

   Se utiliza una red neuronal para generar un mapa probabilístico sobre la imagen completa donde se podrían encontrar las respectivas áreas de interés.

2. **Recorte de las imágenes en parches**

   Las imágenes se recortan en una resolución definida para así encuadrar áreas de interés que no serían notorias si se utilizara la imagen completa.

3. **Obtención de parches con zonas de interés dentro de la imagen**

   Teniendo en cuenta las zonas marcadas como áreas de interés dentro de las máscaras, se hace un filtrado para obtener únicamente los parches que contengan estas.

4. **Concatenación del mapa probabilístico con la imagen original**

   Se convierte la imagen a tensor y se concatena con el respectivo parche del mapa de probabilidades para su posterior refinamiento con una red neuronal a nivel de parches.

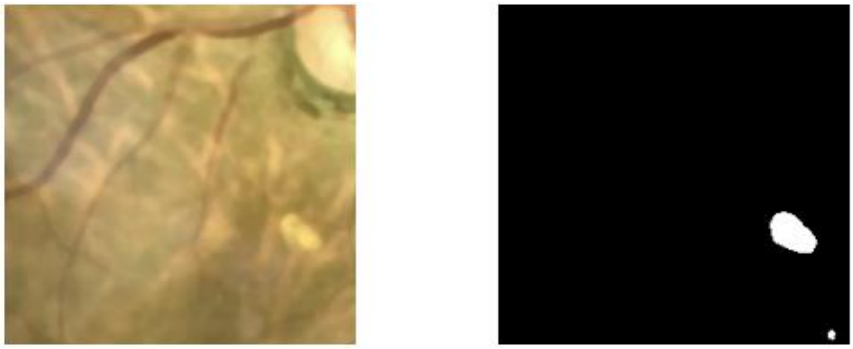


Figura 5-2 Parche de fondo de ojo con drusas y parche de mascara con área de interés resaltada. Elaboración propia

5.5. Preprocesamiento para segmentación de micro aneurismas

- **Aislamiento del canal verde de la imagen**

  Como se menciona en la sección 4.1. Para los micro aneurismas se va a hacer el aislamiento del canal verde para obtener una mejor visión sobre el área de interés (micro aneurismas). Dado que estas presentan un mayor contraste en este canal.

- **Recorte en parches de las imágenes con la estructura de interés**

  Cómo se menciona en la sección 4.1. Se van a realizar recortes de las imágenes en parches para así solventar el problema del desbalance entre las clases "fondo-micro aneurisma"

- **Superposición para las áreas de interés**

  Dado que para la segmentación de microaneurismas no se cuenta con una muestra representativa, se decide hacer superposición de los parches con una ventana deslizante cuyo objetivo es encuadrar la clase de interés en diferentes posiciones para así aumentar el número de imágenes con las que entrenar y probar el modelo.

  Considerando que a esta región de interés no se le aplica ninguna modificación que altere la estructura de la imagen en términos de iluminación o valores de píxeles, no se evalúan las métricas de preprocesamiento descritas al final de la sección.

5.6. Métricas de evaluación para el preprocesamiento

Con las imágenes preprocesadas, se hace la evaluación de la calidad de la calidad del resultado con el propósito de verificar la eficacia de los métodos utilizados para la eliminación de posible ruido en los escáneres, así como la preservación de estructuras en la imagen original con respecto a la imagen modificada. Estas métricas son:

- **Peak signal to noise Ratio**

  El peak signal to noise ratio (*PSNR)* es una métrica que busca evaluar la calidad en la reconstrucción de imágenes pre procesadas con respecto a su original, el *PSNR* busca cuantificar la diferencia entre las imágenes a través del poder máximo posible de la imagen, sin procesar con respecto a la procesada.

Se expresa como:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{(\mathrm{L}-1)^2}{\mathrm{MSE}}\right)$$

En el que "L" es la intensidad máxima que puede tener la imagen (en escala de grises es 255).

El "MSE" se define como:

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(O(i,j) - D(i,j))^2$$

- m = Son la fila de pixeles
- n = Son las columnas de pixeles

"O" representa la matriz con los valores de pixel original de la imagen. "D" representa la matriz de la imagen procesada, haciendo así un cálculo que busca contener las diferencias medibles de la imagen a nivel de pixel.

- **Structural similarity index**

Similar al *PSNR*, el structural similarity index (*SSIM*) es un índice que busca medir la diferencia entre imágenes procesadas desde un punto de vista más robusto, incluyendo contraste y una concepción de similitud más relacionada con la forma en que el ojo percibiría las diferencias. El SSIM entre dos imágenes se calcula mediante una combinación de términos

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2+\sigma_y^2 + c_2}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{c_3}$$

Para así obtener:

$$\mathrm{SSIM}(x,y) = [l(x,y)^{\propto} * c(x,y)^{\beta} * s(x,y)^{\gamma}]$$

Y cuando los pesos α, β, γ son iguales a 1. Tenemos:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2+\sigma_y^2 + c_2)}$$

Donde:

- $\mu_x, \mu_y =$ media de los valores de los pixeles de la imagen x e y respectivamente
- ${\sigma_x}^2, {\sigma_y}^2 =$ son las varianzas de los pixeles de la imagen x e y respectivamente.
- σ_{xy} es la covarianza entre los valores de pixeles de la imagen x e y.

- $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ Son "estabilizadores del valor para los casos en que el denominador sea un valor cercano a 0.
- L es el cociente entre el valor más grande y más chico con respecto al total de pixeles de la imagen (típicamente se calcula como
- $k_1 = 0.01$ y $k2 = 0.03$ por defecto
- $c_3 = \frac{c_2}{2}$

Esta métrica nos va a ser útil para comparar los métodos de eliminación de ruido para cada una de las imágenes, esto dado que las imágenes que tengan un mayor índice, van a guardar una mayor similaridad con la imagen original una vez aplicados los filtros para remoción de ruido.

- **Normalized mutual information**

El *Normalized Mutual Information (NMI)* es una métrica utilizada para evaluar la similitud de la distribución de intensidades entre dos imágenes a partir del uso de la entropía. Se calcula de la siguiente forma:

$$NMI = \frac{H(X) + H(Y)}{H(X,Y)}$$

Donde:

- H (X, Y) = es la entropía conjunta para las imágenes X, Y.
- H(X) y H(Y) son las entropías de las imágenes original y procesada, respectivamente.

El rango de valores que toma esta métrica va de 1 a 2. En donde 1 representa imágenes no correlacionadas, mientras que 2 equivale a imágenes perfectamente correlacionadas.

A nivel interpretativo, el *NMI* puede verse como un indicador de preservación de la información estructural posterior al procesamiento. Cuanto más similar sea la distribución de intensidades entre la imagen procesada y la original, mayor será el *NMI*. En cambio, valores bajos indican una mayor pérdida o alteración de la información original.

- **Contrast improvement index**

El *contrast improvement index (CII por sus siglas en inglés)* es una métrica que evalúa la mejora en el contraste de una imagen a partir del cálculo del coeficiente entre el contraste local promedio de la imagen procesada con respecto a la original.

$$CII = \frac{Contraste\ local\ promedio\ de\ la\ imagen\ procesada}{Contraste\ local\ promedio\ de\ la\ imagen\ original}$$

En donde:

- $Contraste\ local\ promedio = \frac{1}{N}\sum_{i,j}(I_{max}(i,j) - I_{\min(i,j)})$

Siendo:

- $I_{max\ (i,j)} =$ valor máximo de intensidad en una vecindad de pixeles i,j
- $I_{min(i,j)} =$ valor mínimo de intensidad en una vecindad de pixeles i,j
- $N =$ número total de pixeles en la imagen

El rango de valores que toma esta métrica es:

- CII < 1: Equivale a una reducción en el contraste local en la imagen procesada con respecto a la original, esto puede ser un indicio de pérdida del detalle.
- CII = 1: Representa que no hubo modificaciones dentro de la imagen modificada con respecto a la original.
- CII > 1: Representa un mayor contraste local en la imagen procesada con respecto a la original, lo que implica una aplicación de mejora de contraste exitosa.

5.7. Resultados obtenidos después de la aplicación de las respectivas técnicas de preprocesamiento:

Dado que el preprocesamiento de drusas y microaneurismas busca priorizar mejoras en el contraste de la imagen, se utilizarán métricas que evalúen la similaridad para estos componentes específicos. En este caso, *SSIM* mide la similitud estructural considerando factores como el contraste, mientras que el *NMI* busca medir la similaridad entre imágenes a partir de la teoría de la información. El *CII* por su parte busca una mejora en el contraste local de la imagen procesada con respecto a la original, evaluando el incremento en la diferencia de intensidades dentro de una agrupación local píxeles.

Etapa del preprocesamiento	Datasets evaluados	Métricas de evaluación:
Preprocesamiento para segmentación de disco y Copa óptica	ODIR-5K RFMID MESSIDOR2	PSNR: 32.38 SSIM: 0.856 NMI: 1.380
Preprocesamiento para segmentación de micro aneurismas	ODIR-5K RFMID E-OPTHA	SSIM: 0.878 NMI: 1.777 CII: 1.605

Tabla 4 Rendimiento de las técnicas de preprocesamiento aplicadas

5.8. Evaluación de resultados en el post procesamiento

Para el post procesamiento se utilizarán diferentes métricas de evaluación dependiendo de la fase etapa que se esté evaluando, no obstante, esta evaluación se limitará a hacer la validación con el uso de las siguientes métricas como mucho:

- **Índice de Jaccard (IoU)**

  El índice de Jaccard (También conocido como Intersection over Union) es una métrica que evalúa la calidad de una predicción partir de la similitud entre un área predicha y un área de referencia (ground truth) en tareas de clasificación. Matemáticamente es posible denotar esta métrica como:

$$IoU = \frac{Area\ de\ sobreposición}{Area\ de\ unión}$$

  En donde:

  - Área de sobreposición: conjunto de píxeles en los que la predicción y la máscara de referencia coinciden
  - Área de unión: Es el área total ocupada por la predicción y la máscara de referencia.

  Otra forma de representación de esta métrica es:

$$IoU = \frac{TP}{TP + FP + FN}$$

  En donde:

  - TP: son los valores correctamente clasificados para determinada categoría
  - FP: son los valores "falsos positivos" o valores que fueron predichos como pertenecientes a la categoría positiva, que en realidad no pertenecen a esta.
  - FN: son los valores clasificados erróneamente dentro de la categoría complementaria a la evaluada.

  El índice de jaccard toma un rango de valores de 0 a 1. En donde 1 representa una sobreposición perfecta.

- **Coeficiente Dice:**

  El coeficiente Dice es una métrica que busca evaluar la calidad de las predicciones la cual busca dar más peso a los verdaderos positivos, favoreciendo resultados que minimizan los falsos negativos. Matemáticamente está representado por:

$$2 * \frac{Área\ de\ sobreposición}{Área\ total}$$

En donde:

- Área de sobreposición: es el rango de pixeles en los que la predicción coincide con los verdaderos asociados a una imagen.
- Área de total: área que incluye la predicción y el margen de verdad para la muestra.

Otra forma de representar esta métrica es con:

$$Dice\ Coefficient = \frac{2TP}{2TP + FP + FN}$$

En donde:

- TP: son los valores correctamente clasificados para determinada categoría
- FP: son los valores "falsos positivos" o valores que fueron predichos como pertenecientes a la categoría positiva, que en realidad no pertenecen a esta.
- FN: son los valores clasificados erróneamente dentro de la categoría complementaria a la evaluada.
- El coeficiente dice toma un rango de valores de 0 a 1. En donde 1 representa una sobreposición perfecta.

- **Exactitud:**

La exactitud es una métrica que toma todas las predicciones hechas correctamente por el modelo, y las divide sobre las predicciones totales, esta métrica tiene en cuenta una visión más "general" con respecto al comportamiento del modelo sobre todas las clases, se puede expresar como:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

En donde:

- TP: son los valores correctamente clasificados para determinada categoría
- FP: son los valores "falsos positivos" o valores que fueron predichos como pertenecientes a la categoría positiva, que en realidad no pertenecen a esta.
- FN: son los valores clasificados erróneamente dentro de la categoría complementaria a la evaluada.

- TN: son los valores correctamente clasificados dentro de la categoría complementaria a la evaluada.

- **Precisión**

  Esta métrica mida la tasa de valores predichos como pertenecientes al área que fueron correctos, con respecto a las predicciones positivas totales realizadas por el modelo (correctas y erróneas).

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

- **Recall**

  Mide la calidad del modelo para las predicciones positivas con respecto al total de valores totales positivos dados por el modelo y las anotaciones.

$$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

  Se decide hacer uso de estas métricas con el objetivo de poder interpretar correctamente los resultados teniendo en cuenta los posibles desafíos presentes dentro de cada proceso de segmentación.

- **Tasa de detección:**

  Para el modelo de detección y segmentación de disco y copa óptica, los scores generados por el modelo representan la confianza para una predicción especifica. Estos scores pueden interpretarse como una probabilidad en el rango [0,1]**,** indicando qué tan seguro está el modelo de que una determinada área contiene el disco y la copa óptica. Esta puede ser definida como:

$$Tasa\ de\ detección = \frac{\#\ de\ registros\ con\ el\ área\ de\ interés\ detectada}{\#\ de\ registros\ total}$$

  En donde un registro se considera "detectado" si el modelo predice correctamente la presencia del disco y la copa óptica con una probabilidad de más del 70%. Esta métrica será evaluada únicamente en conjuntos de datos en los cuales todos los registros presenten el área de interés (Disco y copa óptica).

- **Validación de resultados**
  - En primera instancia, los datos se particionarán en conjuntos de entrenamiento, prueba y validación, o únicamente en entrenamiento y prueba, dependiendo de la fase del problema que se esté desarrollando. El propósito de esta partición es disponer de imágenes con composiciones distintas a las utilizadas durante el entrenamiento, lo que permitirá una evaluación más objetiva del rendimiento del modelo.
  - Se empleará una estrategia de validación cruzada de cinco particiones (5-fold cross-validation). En cada iteración, se registrarán las métricas

del modelo para diferentes grupos de datos provenientes de los conjuntos de prueba y validación, la idea detrás de este método, es validar el poder predictivo del modelo frente a diferentes conjuntos de datos.

- Con el propósito de robustecer los modelos frente a la presencia de datos en condiciones adversas, se introducen muestras con modificaciones aleatorias (cambios de orientación, escalado aleatorio, introducción de ruido gaussiano, modificación aleatoria del contraste, etc.), de manera que el modelo se entrene con estos datos y desarrolle una mayor capacidad de generalización.

Los procedimientos de validación descritos serán aplicados a todas las secciones posteriores.

5.9. Segmentación de disco y copa óptica

Cómo ya se explicó anteriormente (sección 4.2.1). La arquitectura MobilenetV3 es una mejora frente a su predecesora (MobilenetV2) con respecto a la velocidad y calidad en las respuestas. A esta se le hace la adición de una sección final que utiliza la lógica de los detectores de caja de disparo único para entrenar y hacer predicciones sobre el modelo.

Conjunto de datos	Porcentaje de la población representada	Tasa de detección	Índice de Jaccard (IoU)
Entrenamiento	80%	0.8906	0.8149
Prueba	10%	0.8667	0.8727
Validación	10%	0.8750	0.8202

Tabla 5 Valores promedio en las métricas de encuadre de disco y copa óptica para los conjuntos de entrenamiento y validación

Iteración del Fold	Tasa de detección	Índice de Jaccard
1	0.8750	0.8285
2	0.9167	0.8357
3	0.8750	0.8220
4	0.8750	0.7991
5	0.8333	0.7977

Tabla 6 Resultados de Validación por folds para modelo de encuadre de disco y copa óptica en el conjunto de datos de validación

Iteración del Fold	Tasa de detección	Índice de Jaccard
1	0.9167	0.8734
2	0.9583	0.8738
3	0.9167	0.8669
4	0.7917	0.8820
5	0.7500	0.8673

Tabla 7 Resultados de Validación por folds para modelo de encuadre de disco y copa óptica en el conjunto de datos de prueba



Figura 5-3 Ejemplo de encuadre del modelo sobre el área de interés. Elaboración propia

- **Segmentación semántica del disco y la copa óptica:**

Una vez con la salida de la etapa anterior, se utiliza la imagen para hacer la estimación de la copa y disco óptico de manera individual a través del proceso con descomposición y recomposición que utiliza *U-net* (sección 4.2.1). Los resultados son expuestos a continuación.

Conjunto de datos	Porcentaje de la población representada	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
Entrenamiento	80%	0.8790	0.7891	0.8790	0.8873	0.8918
Prueba	10%	0.8843	0.7764	0.8698	0.8787	0.8857
Validación	10%	0.8874	0.7846	0.8753	0.8887	0.8854

Tabla 8 Valores promedio en las métricas para el modelo de segmentación de disco y copa optica en diferentes conjuntos de datos

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.8897	0.7921	0.8798	0.8894	0.8928
2	0.9009	0.8040	0.8892	0.9042	0.8898
3	0.8700	0.7532	0.8538	0.8678	0.8744
4	0.8851	0.7798	0.8721	0.8780	0.8891
5	0.8914	0.7938	0.8816	0.9042	0.8810

Tabla 9 Resultados de Validación por folds para modelo de segmentación de disco y copa óptica en el conjunto de datos de validación

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.8973	0.7998	0.8863	0.8962	0.8912
2	0.8790	0.7699	0.8654	0.8724	0.8858
3	0.8692	0.7533	0.8542	0.8661	0.8721
4	0.8820	0.7726	0.8674	0.8788	0.8832
5	0.8939	0.7863	0.8757	0.8789	0.8861

Tabla 10 Resultados de Validación por folds para modelo de segmentación de disco y copa óptica en el conjunto de datos de prueba

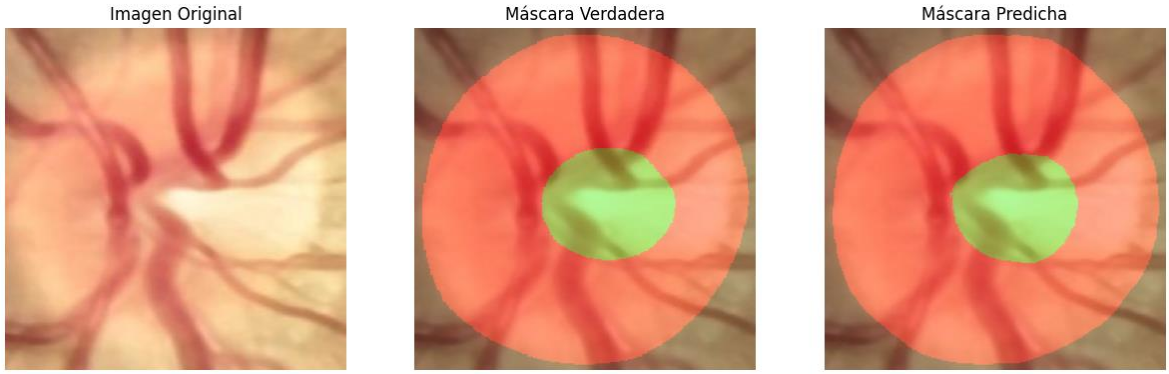


Figura 5-4 Ejemplo de imagen encuadrada y segmentada para disco y copa óptica. Elaboración propia.

5.10. Segmentación de posibles Drusas

Para la segmentación de drusas, se utilizarán los mapas de probabilidad generados por la red de imagen completa, y se concatenarán con las imágenes

recortadas como se explica en la sección 4.2.3. En esta etapa del proceso, se emplea la red de segmentación U-Net para la refinación de las posibles drusas obtenidas.

Conjunto de datos	Porcentaje de la población representada	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
Entrenamiento	80%	0.9974	0.3676	0.5178	0.4236	0.7356
Prueba	10%	0.9978	0.3200	0.4646	0.3968	0.6653
Validación	10%	0.9971	0.3692	0.5126	0.4111	0.7985

Tabla 11 Valores promedio en las métricas para el modelo de segmentación de drusas en diferentes conjuntos de datos

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.9984	0.4067	0.5680	0.4585	0.8584
2	0.9971	0.3905	0.5245	0.5607	0.3970
3	0.9953	0.3617	0.5056	0.5410	0.7863
4	0.9960	0.4038	0.5262	0.6268	0.7607
5	0.9958	0.2881	0.3815	0.5896	0.4812

Tabla 12 Resultados de Validación por folds para modelo de segmentación de drusas en el conjunto de datos de validación

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.9978	0.3096	0.4516	0.3326	0.7793
2	0.9984	0.2562	0.3974	0.9984	0.6699
3	0.9972	0.2525	0.3712	0.2782	0.7313
4	0.9966	0.2665	0.4093	0.3724	0.4996
5	0.9981	0.4409	0.6052	0.5139	0.7513

Tabla 13 Resultados de Validación por folds para modelo de segmentación de drusas en el conjunto de datos de prueba

Aunque el modelo presenta limitaciones en la detección general de las lesiones de interés, destaca por su buen desempeño en recall al identificar correctamente los casos positivos de la clase de drusas. Esto refleja la capacidad de la red para reconocer patrones asociados a esta lesión específica, lo cual es fundamental en aplicaciones en donde minimizar falsos negativos es lo más

importante.

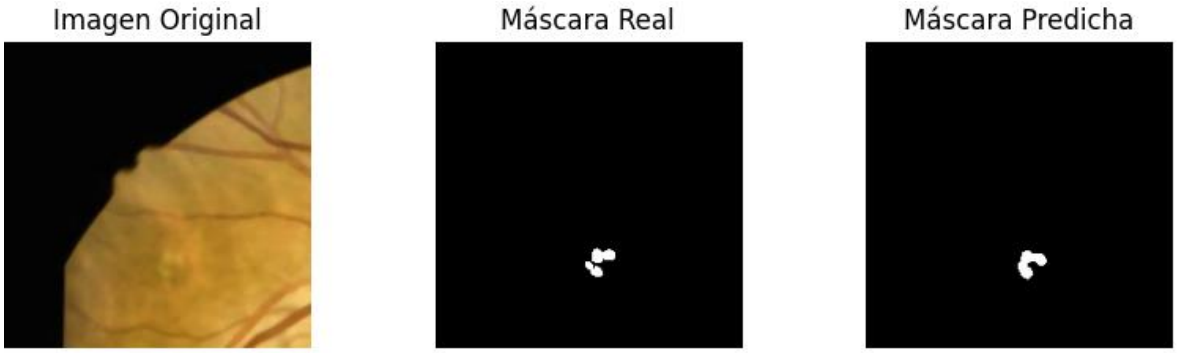


Figura 5-5 Ejemplo de parche de imagen con las drusas segmentadas. Elaboración propia.

5.11. Segmentación de posibles micro aneurismas

De la misma forma que con las drusas (sección 5.8). Se utiliza el modelo de segmentación directamente sobre los parches preprocesados del fondo de ojo para compensar el desbalance de clases. En este caso, se utilizó una red U-Net Recurrente Residual para la extracción de los mapas de segmentación.

Conjunto de datos	Porcentaje de la población representada	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
Entrenamiento	80%	0.9431	0.4780	0.6881	0.6007	0.8053
Prueba	10%	0.9440	0.5046	0.6808	0.5046	0.7996
Validación	10%	0.9421	0.5122	0.6831	0.5915	0.8084

Tabla 14 Valores promedio en las métricas para el modelo de segmentación de microaneurismas en diferentes conjuntos de datos

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.9428	0.5455	0.7059	0.6167	0.8253
2	0.9451	0.5408	0.7020	0.6129	0.8215
3	0.9419	0.5073	0.6731	0.5833	0.7957
4	0.9403	0.4951	0.6623	0.5733	0.7839
5	0.9406	0.5031	0.6694	0.5691	0.8121

Tabla 15 Resultados de Validación por folds para modelo de segmentación de microaneurismas en el conjunto de datos de validación

Iteración del fold	Accuracy	Índice de Jaccard (IoU)	Dice Score	Precisión	Recall
1	0.9425	0.5254	0.6889	0.6037	0.8022
2	0.9411	0.5056	0.6717	0.5810	0.7959
3	0.9470	0.5183	0.6827	0.6127	0.7709

4	0.9454	0.5064	0.6723	0.5723	0.8147
5	0.9443	0.5242	0.6879	0.5949	0.8153

Tabla 16 Resultados de Validación por folds para modelo de segmentación de microaneurismas en el conjunto de datos de prueba

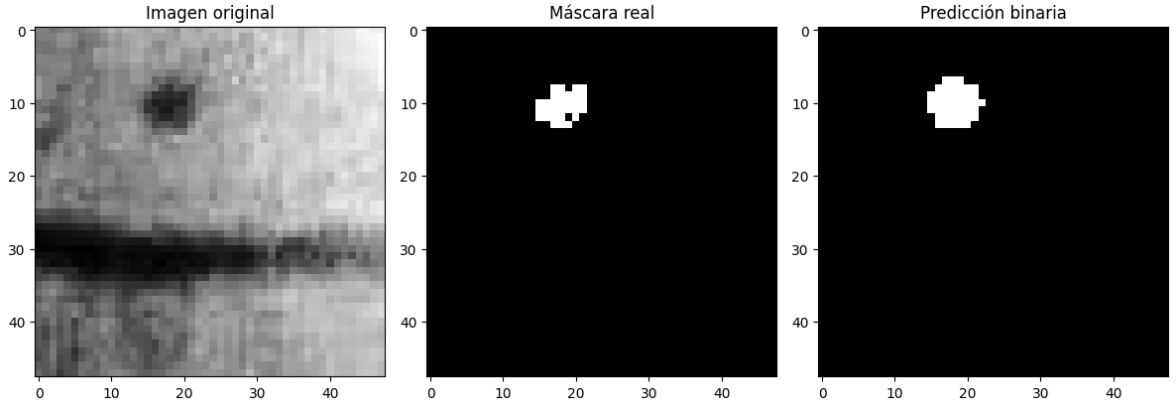


Figura 5-6 Ejemplo de parche de imagen con micro aneurismas segmentados.: Elaboración propia

Según los resultados obtenidos, al finalizar el proceso se dispone de un modelo con la capacidad de segmentar de forma precisa microaneurismas en parches extraídos de imágenes de fondo de ojo. Cabe resaltar que métricas como la exactitud (accuracy) y la precisión (precision) pueden no ser del todo representativas en este tipo de problemas, dado que son susceptibles al sesgo inducido por el desbalance entre clases. Esto es particularmente importante en el caso de los microaneurismas, cuya presencia puede representar menos del 1% del área total de la imagen, lo que reduce la sensibilidad de estas métricas frente a verdaderos positivos en regiones pequeñas, pero clínicamente relevantes.

5.12. Evaluación de modelo de clasificación de ensamble

Para el siguiente modelo se hizo una partición de los datos en un conjunto de entrenamiento y prueba para evaluar los resultados de clasificación. Las métricas de evaluación para el ensamble final buscaran cuantificar la capacidad del modelo para generalizar sobre las clases objetivo evitando el sobreajuste en los datos de entrenamiento, y contando con una capacidad de clasificación equilibrada para todos los tipos de enfermedad a clasificar. Para esto se reutilizarán las métricas utilizadas de clasificación de la sección 5.7. (recall, y precisión). Junto con las siguientes métricas:

- **AUC-ROC**

  El área bajo la curva (AUC-ROC) es una métrica que evalúa el desempeño de un modelo de clasificación binaria o multiclase, a través del cálculo de la sensibilidad y la tasa de falsos positivos modificando los

thresholds de confianza para así determinar una serie de puntos los cuales van a definir qué tan sensible es el modelo al hacer estos cambios con respecto a la calidad de las predicciones para cada clase. Para esto se define:

- Sensibilidad = $\frac{TP}{TP+FN}$
- Tasa de falso positivos = $\frac{FP}{FP+TN}$

En donde:

- TP son los valores de la clase predichos correctamente
- FN son los valores de la clase complementaria predichos de manera errónea
- FP son los valores de la clase predichos de forma errónea
- TN son los valores de la clase complementaria predichos correctamente

Aunque el AUC-ROC se usa comúnmente en tareas de clasificación binaria, también puede aplicarse en problemas multiclase mediante una estrategia "uno contra todos" (one-vs-rest). En este enfoque, cada clase se evalúa como positiva frente a todas las demás clases consideradas negativas. Así, se calcula un AUC por clase, lo que permite medir qué tan bien el modelo distingue esa clase en particular sin confundirse con las demás.

- **F1-Score**

El F1-Score es la media armónica entre la precisión y el recall. Esta métrica mejora de forma proporcional a ambas, evitando que una predomine sobre la otra, lo que permite obtener una evaluación más equilibrada del desempeño del modelo en tareas de clasificación. Esta métrica se puede definir como:

$$F1 = \frac{2 * TP}{2 * TP + FP + FN}$$

Patología	**Precisión**	**Recall**	**AUC-ROC**	**F1-score**
Glaucoma	0.87	0.70	0.9453	0.77
Diabetes	0.99	1	0.9995	0.99
DMS (degeneración macular senil)	1	0.99	0.9998	0.99
Ojo sano	0.75	0.90	0.9441	0.81

Tabla 17 Valores promedio en las métricas para el modelo de clasificación de ensamble en el conjunto de entrenamiento

Patología	Precisión	Recall	AUC-ROC	F1-score
Glaucoma	0.82	0.65	0.9225	0.72
Diabetes	0.99	0.98	0.9997	0.99
DMS (degeneración macular senil)	0.99	0.99	0.9998	0.99
Ojo sano	0.65	0.87	0.9180	0.75

Tabla 18 Valores promedio en las métricas para el modelo de clasificación de ensamble en el conjunto de prueba

Iteración del fold	Patología	Precisión	Recall	AUC-ROC	F1-score
1.	**Glaucoma**	0.93	0.68	0.9676	0.79
	Diabetes	1	1	1	1
	DMS	1	1	1	1
	Ojo sano	0.71	0.94	0.9591	0.81
2.	**Glaucoma**	0.80	0.50	0.9284	0.62
	Diabetes	0.94	1	1	0.97
	DMS	1	0.95	1	0.97
	Ojo sano	0.67	0.89	0.9340	0.76
3.	**Glaucoma**	0.79	0.58	0.9383	0.67
	Diabetes	1	1	1	1
	DMS	1	1	1	1
	Ojo sano	0.60	0.80	0.8627	0.69
4.	**Glaucoma**	0.73	0.40	0.8627	0.52
	Diabetes	1	0.92	0.9964	0.96
	DMS	1	1	1	1
	Ojo sano	0.61	0.86	0.8636	0.72
5.	**Glaucoma**	0.81	0.62	0.9105	0.70
	Diabetes	1	1	1	1
	DMS	1	1	1	1
	Ojo sano	0.68	0.85	0.9127	0.76

Tabla 19 Resultados de Validación por folds para modelo de clasificación de ensamble en el conjunto de datos de prueba

5.13. Aplicación de métodos de explicabilidad

Para la evaluación de los métodos de explicabilidad, se optó por una aproximación visual que permitiera validar los resultados obtenidos. En particular, los métodos

basados en la contribución del gradiente, como Grad-CAM++, ofrecen una perspectiva más clara para corroborar supuestos relacionados con el preprocesamiento de las imágenes. En este caso, se busca confirmar que cada enfermedad esté asociada a una "zona de enfoque" específica que la red utiliza para su clasificación. Un ejemplo de ello es la activación de ciertas regiones ante la presencia de microaneurismas, lo que permite determinar si una imagen corresponde a un cuadro de retinopatía diabética.

5.13.1. Mapa de activación para glaucoma

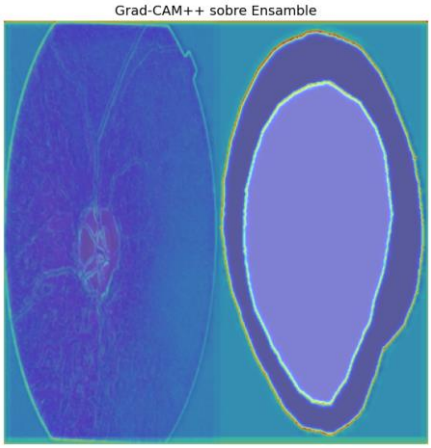


Figura 5-7 Mapa de calor sobre imagen de fondo de ojo con glaucoma. Elaboración propia

En este ejemplo, el modelo clasificó la imagen como correspondiente a un caso de glaucoma. De izquierda a derecha se presentan:

- la imagen original del fondo de ojo
- el mapa de segmentación del disco y la copa óptica.

En el mapa de activación, se observa que la respuesta se concentra principalmente en los bordes del disco y la copa óptica, lo que indica que, en las capas seleccionadas, el modelo se guía mayormente de la morfología de estas estructuras para la generación de la clasificación final.

5.13.2. Mapa de activación para retinopatía diabética

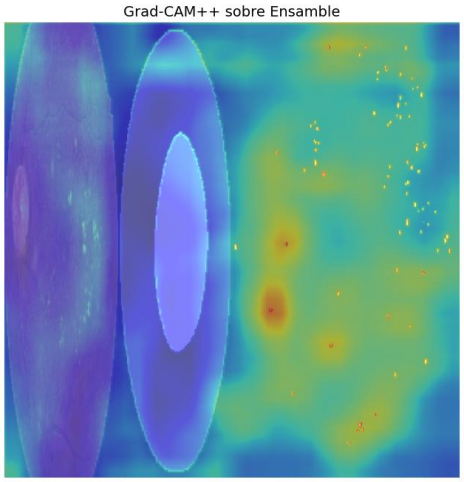


Figura 5-8 Mapa de calor sobre imagen de fondo de ojo con retinopatía diabética. Elaboración propia

En este caso, el modelo identificó correctamente un cuadro de retinopatía diabética. La imagen muestra, de izquierda a derecha:

- La fotografía original del fondo de ojo
- El mapa de segmentación de disco y copa óptica
- Los microaneurismas detectados

Se observa que las áreas más intensas (en tonos rojos) del mapa de atención coinciden con la región donde se detectaron los microaneurismas. Esto indica que el modelo basó su diagnóstico principalmente en esta anomalía, demostrando que su "foco visual" se alinea con los signos clínicos clave de la enfermedad desde etapas tempranas.

5.13.3. Mapa de activación para degeneración macular senil

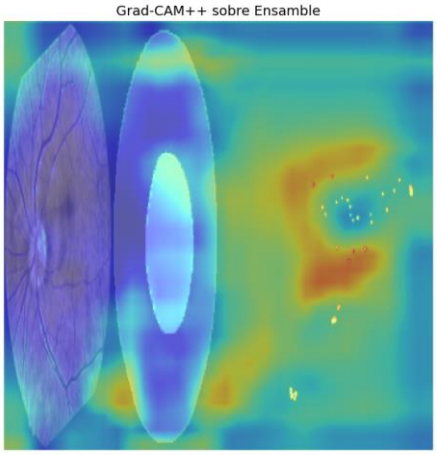


Figura 5-9 Mapas de activación con gradCAM++ para una imagen de fondo de ojo con degeneración macular senil. Elaboración propia.

En este caso, el modelo identificó correctamente un caso de degeneración macular. La imagen muestra, de izquierda a derecha:

- La fotografía original del fondo de ojo
- El mapa de segmentación de disco y copa óptica
- Las drusas detectadas

Para este ejemplo, el foco visual se sitúa sobre el área de la imagen que contiene las drusas, indicando su importancia sobre la decisión del modelo en la clasificación final.

5.14. Escenarios multietiqueta

En escenarios clínicos reales, es frecuente que un mismo ojo manifieste múltiples patologías. Ante este escenario de clasificación multietiqueta, se requiere un enfoque que permita estimar probabilidades individualizadas e independientes para cada condición, preservando la capacidad diagnóstica del modelo.

Para resolver este desafío, se sustituyó la función de activación softmax por una sigmoide en la capa de salida. Esta adaptación:

- Elimina la restricción de suma unitaria inherente a softmax.
- Permite estimaciones probabilísticas independientes, reflejando mejor la coexistencia real de enfermedades.
- Facilita la interpretación clínica al desacoplar las predicciones para cada retinopatía.

A continuación, se presentan algunos resultados de esta aproximación (para ca da una de las imágenes de fondo de ojo, se van a mostrar las probabilidades concernientes a cada combinación de patologías estudiada).

- **Ojos con degeneración macular senil y glaucoma:**

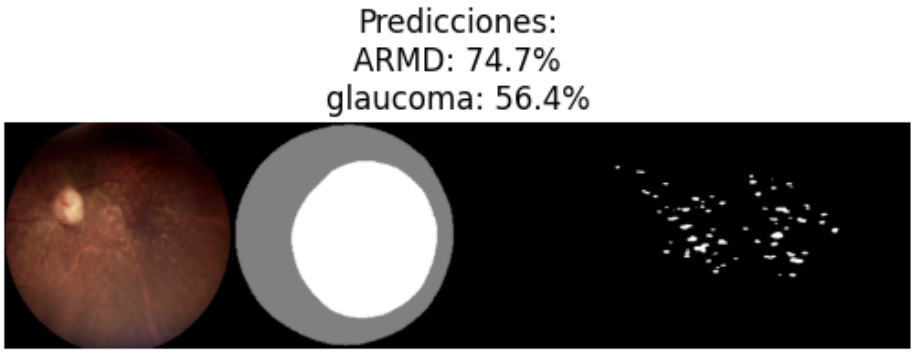


Figura 5-10 Imagen ensamble de fondo de ojo con presencia de glaucoma y degeneración macula. Elaboración propia.

En este caso, y en consonancia con lo observado mediante los métodos de explicabilidad, se puede apreciar que los mapas de segmentación muestran alteraciones morfológicas en estructuras como el disco y la copa óptica. Además, la detección de drusas refuerza la decisión del modelo al identificar la presencia de degeneración macular senil y glaucoma.

- **Ojo con glaucoma y retinopatía diabética**

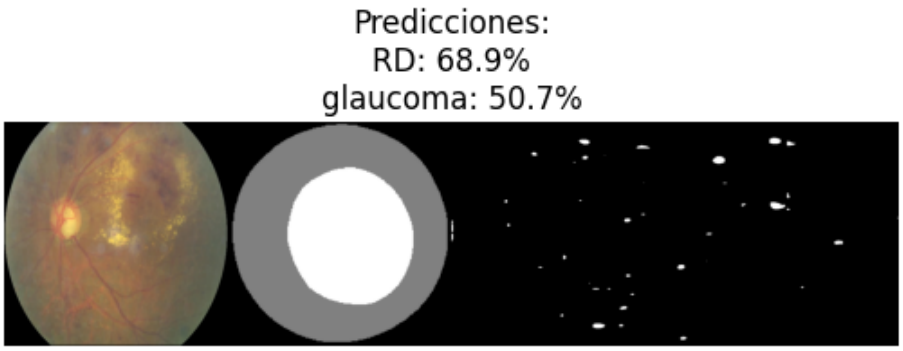


Figura 5-11 Imagen ensamble de fondo de ojo con presencia de glaucoma y retinopatía diabética. Elaboración propia.

Para esta muestra, aunque el modelo no presente una alta confianza en el diagnóstico de glaucoma, se evidencia su capacidad para identificar en el segundo mapa de segmentación la distribución correspondiente a los microaneurismas, así como la anomalía correspondiente a la forma del disco y la copa óptica.

- **Ojo con Retinopatía diabética y degeneración macular**

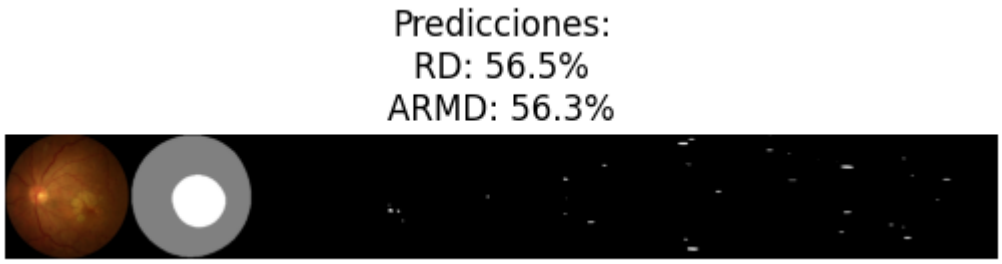


Figura 5-12 Imagen ensamble de fondo de ojo con presencia de retinopatía diabética y degeneración macular. Elaboración propia.

En casos donde un mismo ojo presenta retinopatía diabética y degeneración macular senil, se observa que, a pesar de que las predicciones más elevadas se asocian con las patologías correspondientes, el margen de confianza es limitado. Esto podría indicar que el modelo encuentra dificultades para diferenciar de forma precisa entre la distribución de drusas y la de microaneurismas segmentadas en una misma imagen en este tipo de escenarios.

Si bien el modelo muestra potencial para detectar múltiples enfermedades oculares simultáneas, es importante destacar que no fue entrenado específicamente con casos complejos donde coexisten varias patologías. Esta limitación surge principalmente por la dificultad de acceder a conjuntos de datos públicos que incluyan estas combinaciones reales, ya que la mayoría de bancos de imágenes de fondo de ojo con estas características son de acceso restringido o requieren licencias.

Este panorama genera dos desafíos clave:

1. Los modelos actuales podrían subestimar la frecuencia real de casos con múltiples enfermedades coexistentes debido a la limitada disponibilidad de conjuntos de datos públicos que representen estas combinaciones de manera adecuada.
2. Aunque la sustitución de la activación softmax por funciones sigmoides permite estimaciones probabilísticas independientes, dichas probabilidades podrían no estar correctamente calibradas, afectando la fiabilidad de las decisiones basadas en las predicciones del modelo.

5.15. Sensibilidad del modelo

Aunque el modelo muestra una alta capacidad discriminatoria para la detección de casos con degeneración macular senil y retinopatía diabética, este puede presentar ciertas fallas al diferenciar casos de glaucoma de ojos sin afección. A continuación, se exponen algunos ejemplos de clasificación y se ilustra cómo el modelo procesa la

información antes de emitir un diagnóstico, utilizando las metodologías visuales de la sección 5.12.

- **Ojos erróneamente clasificados como normales**

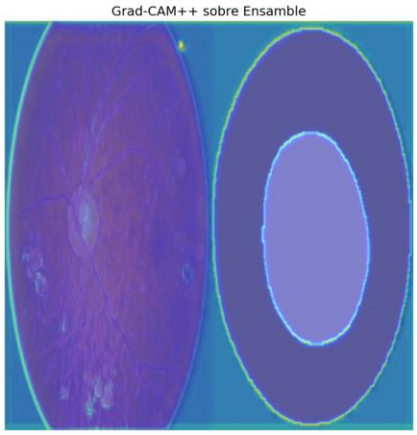


Figura 5-13 Ojo con glaucoma erróneamente clasificado como normal. Elaboración propia.

En este caso, cuando el modelo clasifica erróneamente un ojo como "normal" a pesar de presentar síntomas de glaucoma, se evidencia la influencia de la morfología de las áreas de interés. El problema radica en el modelo de segmentación del disco y la copa óptica, que para este ejemplo estima incorrectamente la forma de la copa óptica. La estructura, en la imagen original, exhibe una morfología considerablemente más compleja que la simplificada por el modelo, lo que conduce a la clasificación errónea.

- **Ojos erróneamente clasificados como afectados por glaucoma**

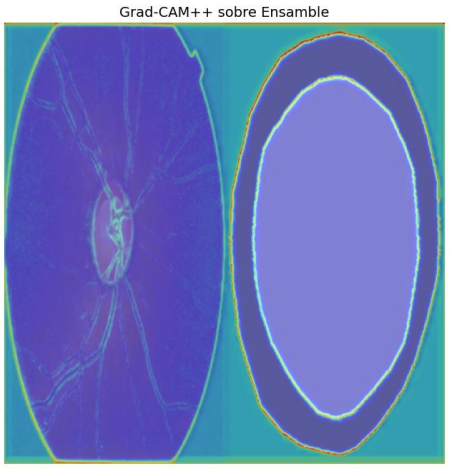


Figura 5-14 Ojo normal erróneamente clasificado como afectado por glaucoma. Elaboración propia.

Como se observa en este ejemplo, las capas seleccionadas para la visualización del gradiente reflejan una marcada influencia de la morfología

del disco y la copa óptica. Sin embargo, surge un problema en la etapa de clasificación, ya que el modelo se ha condicionado a recibir imágenes con estas características que han sido clasificados como afectadas por glaucoma, pese a las imprecisiones en la segmentación de las áreas de interés. La inexactitud en la estimación de dichas áreas introduce un sesgo en el diagnóstico del modelo.

Para reducir los errores de clasificación observados en estos escenarios, se plantean dos estrategias principales:

1. **Robustecimiento del modelo de segmentación de disco y copa óptica:**

   Se propone el fortalecimiento del modelo de segmentación para que sea capaz de manejar morfologías atípicas y variaciones extremas, disminuyendo así las representaciones simplificadas que inducen a error en la clasificación.

2. **Ampliar la diversidad de los datos de entrenamiento para el modelo de clasificación**

   Se esperaría que, a través de la incorporación de más muestras de ojos con diferencias morfológicas sutiles entre condiciones normales y glaucomatosas, permitiera que el modelo mejore su capacidad de generalización y pueda distinguir patrones más complejos, reduciendo la confusión entre ambas clases.

6. DISCUSIÓN Y CONCLUSIONES

En este trabajo se implementó un sistema para la identificación de patologías en imágenes de fondo de ojo mediante una metodología diagnóstica basada en la detección de zonas de interés con distribuciones específicas, simulando los procedimientos empleados en la práctica médica especializada. Esta aproximación busca no solo facilitar la clasificación de anomalías en las imágenes, sino también potenciar la capacidad de soporte diagnóstico ofrecida por los modelos de clasificación.

Se tomaron como referencia proyectos anteriores de segmentación y clasificación de enfermedades retinográficas [17,19,77,63,65] para optimizar las etapas de preprocesamiento y generación de los mapas de segmentación utilizados en el modelo de ensamble final. A partir de esta revisión, se profundizó en las limitaciones de los enfoques actuales, especialmente en cuanto a la explicabilidad de los modelos de aprendizaje profundo. Para abordar esta carencia, se integraron técnicas de visualización de activaciones como Grad-CAM++, con el fin de resaltar las regiones de mayor influencia en las predicciones de cada patología. Asimismo, se reforzó la capacidad de generalización del modelo para resultados en datos no vistos mediante estrategias de

aumento y balance de datos en los datos de entrenamiento, además de implementar validación cruzada para evaluar la consistencia del modelo sobre muestras no vistas.

Durante el desarrollo surgieron diversos desafíos, entre ellos la obtención de datos clínicos anotados para la generación y validación de los mapas de segmentación, así como la escasez de conjuntos multietiqueta que permitieran utilizar una metodología multiclase en casos de coexistencia de múltiples patologías retinianas. Para mitigar el desbalance de clases en estructuras de baja presencia (como microaneurismas y drusas) se ajustaron hiperparámetros en las funciones de pérdida y se aplicaron estrategias de recorte focalizado en las regiones de interés.

Aunque el modelo de ensamble propuesto demuestra viabilidad para la clasificación de las retinopatías seleccionadas, el flujo de datos (desde la generación de mapas hasta la clasificación final) puede resultar computacionalmente costoso debido a sus múltiples etapas. En [19] se sugiere una aproximación alternativa basada en la segmentación de vasos sanguíneos para simplificar el proceso de clasificación, reduciendo la complejidad de preprocesamiento y análisis. No obstante, dicha estrategia puede limitar la riqueza de los patrones clínicos obtenidos necesarios para un diagnóstico basado en parámetros específicos, tal como se propone en este estudio.

Para el desarrollo del modelo de clasificación y de los módulos de segmentación de estructuras, se comprobó que emplear redes con pesos preentrenados superaba en rendimiento en comparación a entrenar desde cero con pesos inicializados de forma aleatoria. Estudios como [17, 18] destacan el uso de arquitecturas pre entrenadas para tareas de detección de retinopatías, señalando mejoras en precisión y velocidad de convergencia.

Como posibles líneas de mejora, se podrían explorar distintas estrategias de fusión de mapas de segmentación más allá de la concatenación espacial horizontal empleada en este trabajo. En particular, la concatenación a lo largo de la dimensión de canales (tensorial) en lugar de en las dimensiones de alto y ancho permitiría preservar íntegramente la información espacial de cada mapa de segmentación (disco y copa óptica, microaneurismas y drusas), al tiempo que agiliza el procesamiento de los datos dentro de la red neuronal. Asimismo, podría evaluarse el uso de redes de segmentación con bloques de atención que permitan extraer información semántica de estructuras con distribución determinada de manera más eficiente, gracias a los attention gates, los cuales podrían facilitar la convergencia del modelo sobre regiones específicas como disco y copa óptica. Por último, se podría aprovechar el potencial de las redes generativas adversariales para robustecer la segmentación de estructuras finas, como microaneurismas y drusas, mediante un enfoque de aprendizaje mutuo entre el generador y el discriminador [76,77].

Este estudio constituye un primer paso hacia la generación de modelos interpretables y explicables que faciliten la evaluación diagnóstica por parte de profesionales de la salud en el ámbito optométrico. Se prevé que, con el continuo desarrollo de nuevos modelos y la disponibilidad de conjuntos de datos curados profesionalmente, se optimice tanto la calidad y la variedad de las imágenes de entrenamiento como la eficiencia computacional, permitiendo una integración efectiva en entornos con recursos limitados.

7. REFERENCIAS

1. Fernández-Vega, I. O. (2017, octubre 20). *Principales causas de enfermedad de la retina*. Fernandez-vega.com. https://fernandez-vega.com/blog/enfermedades-de-la-retina/
2. *Detectar enfermedades por los ojos*. (2023, enero 16). COP - Clinica Oftalmológica del Penedes ES. https://copenedes.com/es/detectar-enfermedades-por-los-ojos/
3. Manresa, J. M., Forés, R., Vázquez, X., Alzamora, M. T., Heras, A., Delgado, P., & Torán, P. (2020). *Fiabilidad de la retinografía para la detección de retinopatía hipertensiva en Atención Primaria*. Atencion primaria, 52(6), 410–417. https://doi.org/10.1016/j.aprim.2019.06.005
4. *Informe mundial sobre la visión*. (8 de octubre de 2019). Who.int; World Health Organization. https://www.who.int/es/publications/i/item/9789241516570
5. What Are the Risks of High Myopia? (s.f) MiSight® 1 day. Retrieved September 14, 2024, from https://misight.com/blog/what-are-risks-high-myopia
6. Larxel. (2020). Ocular Disease Recognition [Data set].
7. Bazan, J. (2020, mayo 29). *The link between myopia, cataracts, & glaucoma. Parkslopeeye.com*; Park Slope Eye. https://parkslopeeye.com/the-link-between-myopia-cataracts-and-glaucoma/
8. Tham, Y.-C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., & Cheng, C.-Y. (2014). *Global prevalence of glaucoma and projections of glaucoma burden through 2040*. Ophthalmology, 121(11), 2081–2090.
9. Holden, B. A., Fricke, T. R., Wilson, D. A., Jong, M., Naidoo, K. S., Sankaridurg, P., Wong, T. Y., Naduvilath, T. J., & Resnikoff, S. (2016). *Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050.* Ophthalmology, 123(5), 1036–1042.
10. El manejo optométrico del paciente con Hipertensión Arterial. (s.f). Optometristas.org. Recuperado el 25 de julio de 2024,
11. Tan, T.-E., & Wong, T. Y. (2023). *Diabetic retinopathy: Looking forward to 2030. Frontiers in endocrinology, 13.*
12. Elloumi, Y., Akil, M., & Boudegga, H. (2019). *Ocular Diseases Diagnosis in Fundus Images using a Deep Learning: Approaches, tools and Performance evaluation*. En arXiv [eess.IV]. http://arxiv.org/abs/1905.02544
13. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. ArXiv.org. http://arxiv.org/abs/1704.04861
14. Sunija, Kar, S., Gayathri, Gopi, V. P., & Palanisamy, P. (2021). OctNET*: A lightweight CNN for retinal disease classification from optical coherence tomography images. Computer Methods and Programs in Biomedicine*, 200(105877), 105877. https://doi.org/10.1016/j.cmpb.2020.105877
15. Yim, J., Chopra, R., Spitz, T., Winkens, J., Obika, A., Kelly, C., Askham, H., Lukic, M., Huemer, J., Fasler, K., Moraes, G., Meyer, C., Wilson, M., Dixon, J., Hughes, C.,

Rees, G., Khaw, P. T., Karthikesalingam, A., King, D., … De Fauw, J. (2020). *Predicting conversion to wet age-related macular degeneration using deep learning. Nature Medicine,* 26(6), 892–899. https://doi.org/10.1038/s41591-020-0867-7

16. Subramanian M, Kumar MS, Sathishkumar VE, Prabhu J, Karthick A, Ganesh SS, Meem MA (2022) *Diagnosis of retinal diseases based on Bayesian optimization deep learning network using optical coherence tomography images*. Comput Intell Neurosci. Vol. 2022
17. Berto, A., Scarpa, F., Tsiknakis, N., Manikis, G., Fotiadis, D. I., Marias, K., & Scarpa, A. (2024). *Automated analysis of fundus images for the diagnosis of retinal diseases: a review.* Research on Biomedical Engineering, 40(1), 225–251. https://doi.org/10.1007/s42600-023-00320-9
18. Sajid, M. Z., Qureshi, I., Abbas, Q., Albathan, M., Shaheed, K., Youssef, A., Ferdous, S., & Hussain, A. (2023). Mobile-HR: *An ophthalmologic-based classification system for diagnosis of hypertensive retinopathy using optimized MobileNet architecture.* Diagnostics (Basel, Switzerland), 13(8), 1439. https://doi.org/10.3390/diagnostics13081439
19. Kumar, K. S., & Singh, N. P. (2023). *Retinal disease prediction through blood vessel segmentation and classification using ensemble-based deep learning approaches*. Neural Computing & Applications, 35(17), 12495–12511. https://doi.org/10.1007/s00521-023-08402-6
20. Lim, G., Bellemo, V., Xie, Y., Lee, X. Q., Yip, M. Y. T., & Ting, D. S. W. (2020). *Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review.* Eye and Vision (London, England), 7(1). https://doi.org/10.1186/s40662-020-00182-7
21. Mitra, A., Banerjee, P. S., Roy, S., Roy, S., & Setua, S. K. (2018). *The region of interest localization for glaucoma analysis from retinal fundus image using deep learning. Computer Methods and Programs in Biomedicine, 165, 25–35.* https://doi.org/10.1016/j.cmpb.2018.08.003
22. Myopia and glaucoma: Sorting out the diagnosis. (2013, noviembre 1). *American Academy of Ophthalmology.* https://www.aao.org/eyenet/article/myopia-glaucoma-sorting-out-diagnosis
23. Coronado, I., Pachade, S., Trucco, E., Abdelkhaleq, R., Yan, J., Salazar-Marioni, S., Jagolino-Cole, A., Bahrainian, M., Channa, R., Sheth, S. A., & Giancardo, L. (2023). *Synthetic OCT-A blood vessel maps using fundus images and generative adversarial networks. Scientific Reports*, 13(1), 1–12. https://doi.org/10.1038/s41598-023-42062-9
24. *Optical coherence tomography category - ophthalmicmart.* (s.f). Recuperado el 6 de septiembre de 2024, de https://www.ophthalmicmart.com/product-category/oct/
25. *OCT Examination VS Fundus Photography: What to Choose*. (s.f.). Retrieved September 14, 2024, from https://www.altris.ai/article/oct-examination-vs-fundus-photography-which-method-to-choose/

26. Li, Z., He, Y., Keel, S., Meng, W., Chang, R. T., & He, M. (2018). *Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs*. Ophthalmology, 125(8), 1199–1206. https://doi.org/10.1016/j.ophtha.2018.01.023
27. Asaoka, R., Murata, H., Hirasawa, K., Fujino, Y., Matsuura, M., Miki, A., Kanamoto, T., Ikeda, Y., Mori, K., Iwase, A., Shoji, N., Inoue, K., Yamagami, J., & Araie, M. (2019). *Using Deep Learning and Transfer Learning to Accurately Diagnose Early-Onset Glaucoma From Macular Optical Coherence Tomography Images. American Journal of Ophthalmology, 198, 136–145.* https://doi.org/10.1016/J.AJO.2018.10.007
28. Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., … Zhang, K. (2018). *Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell*, *172*(5), 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010
29. Kazmi, M., Hafeez, B., Fatima, D., Qamar, M., Ahmed Qazi, S., Rehman Siddiqui, M. A., & Saad Ahmed Qazi, B. (s.f). *Multimedia Tools and Applications Deep learning based diabetic retinopathy screening for resource constraint applications*. https://doi.org/10.1007/s11042-023-18036-4
30. Durairaj, S., Subramanian, P., & Swamy, C. S. M. (2024). *Classification of diabetic retinopathy severity level using deep learning. International Journal of Diabetes in Developing Countries*. https://doi.org/10.1007/s13410-024-01329-1
31. Guefrachi, S., Echtioui, A., & Hamam, H. (2024*). Diabetic Retinopathy Detection Using Deep Learning Multistage Training Method. Arabian Journal for Science and Engineering*, 1–18. https://link-springer-com-443.webvpn.synu.edu.cn/article/10.1007/s13369-024-09137-9
32. Rizvana, M., & Narayanan, S. (2024). *Deep learning of fundus images and optical coherence tomography images for ocular disease detection – a review.* Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-18938-x
33. Xu, Y., Liu, H., Sun, R., Wang, H., Huo, Y., Wang, N., & Hu, M. (2024*). Deep learning for predicting circular retinal nerve fiber layer thickness from fundus photographs and diagnosing glaucoma*. Heliyon, 10(13). https://doi.org/10.1016/j.heliyon.2024.e33813
34. Biousse, V., Najjar, R. P., Tang, Z., Lin, M. Y., Wright, D. W., Keadey, M. T., Wong, T. Y., Bruce, B. B., Milea, D., Newman, N. J., Fraser, C. L., Micieli, J. A., Costello, F., Bénard-Séguin, É., Yang, H., Chan, C. K. M., Cheung, C. Y., Chan, N. C., Hamann, S., … Kilangalanga, J. N. (2024). *Application of a Deep Learning System to Detect Papilledema on Nonmydriatic Ocular Fundus Photographs in an Emergency Department*. American Journal of Ophthalmology, 261, 199–207. https://doi.org/10.1016/J.AJO.2023.10.025

35. Wang, L., Li, B., Pan, J., Zhang, C., & Wang, T. (2024). *Optical coherence tomography image recognition of diabetic retinopathy based on deep transfer learning. Journal of Radiation Research and Applied Sciences, 17(3), 101026.* https://doi.org/10.1016/J.JRRAS.2024.101026
36. Yi, S., & Lingxiang Zhou, ·. (2024). *Multi-step framework for glaucoma diagnosis in retinal fundus images using deep learning.* https://doi.org/10.1007/s11517-024-03172-2
37. Subramaniam, K., & Naganathan, A. (2024). *Enhancing retinal fundus image classification through Active Gradient Deep Convolutional Neural Network and Red Spider Optimization. Neural Computing and Applications.* https://doi.org/10.1007/s00521-024-09989-0
38. EfficientNet. Paperswithcode.com. Retrieved September 16, 2024, from https://paperswithcode.com/method/efficientnet
39. Papers with code - ResNet explained. (s.f.). Paperswithcode.com. Retrieved September 16, 2024, from https://paperswithcode.com/method/resnet
40. Boesch, G. (2021, October 6). *Very deep convolutional networks (VGG) essential guide. Viso.Ai.* https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/
41. Arxiv.org. Retrieved September 16, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications 2024, from http://arxiv.org/abs/1704.04861
42. *Dense Block Explained | Papers With Code*. Retrieved September 15, 2024, from https://paperswithcode.com/method/dense-block
43. Sengar, N., Joshi, R. C., Dutta, M. K., & Burget, R. (2023). EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network. *Neural Computing and Applications*, *35*(14), 10551–10571. https://link.springer.com/article/10.1007/s00521-023-08249-x
44. A. Nawaz, T. Ali, G. Mustafa, M. Babar and B. Qureshi, "Multi-Class Retinal Diseases Detection Using Deep CNN With Minimal Memory Consumption," in IEEE Access, vol. 11, pp. 56170-56180, 2023, https://ieeexplore.ieee.org/document/10141609
45. Gour, N., & Khanna, P. (2021). Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomedical Signal Processing and Control*, *66*, 102329. https://doi.org/10.1016/J.BSPC.2020.102329
46. Jordan, K. C., Menolotto, M., Bolster, N. M., Livingstone, I. A. T., & Giardini, M. E. (2017). *A review of feature-based retinal image analysis. Expert Review of Ophthalmology, 12(3), 207–220*. https://doi.org/10.1080/17469899.2017.1307105
47. Krzywicki, T., Brona, P., Zbrzezny, A. M., & Grzybowski, A. E. (2023). *A Global Review of Publicly Available Datasets Containing Fundus Images: Characteristics, Barriers to Access, Usability, and Generalizability. Journal of Clinical Medicine, 12*(10). https://doi.org/10.3390/jcm12103587

48. Sonali, Sahu, S., Singh, A. K., Ghrera, S. P., & Elhoseny, M. (2019). An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Optics & Laser Technology*, *110*, 87–98. https://doi.org/10.1016/J.OPTLASTEC.2018.06.061
49. Vijaya, H. S., Jayaram, M. A., Gowda, A., & P.T., B. (2016). A Comparative Study on Filters with Special Reference to Retinal Images. *International Journal of Computer Applications*, *138*(5), 36–41. https://doi.org/10.5120/IJCA2016908834
50. *1000 Fundus images with 39 categories*. (s.f.). Retrieved October 4, 2024, from https://www.kaggle.com/datasets/linchundan/fundusimage1000
51. Hassan, T., Akram, M. U., & Nazir, M. N. (2021). *A Composite Retinal Fundus and OCT Dataset with Detailed Clinical Markings of Retinal Layers and Retinal Lesions to Grade Macular and Glaucomatous Disorders*. *4*. https://doi.org/10.17632/TRGHS22FPG.4
52. *Ocular Disease Recognition*. (s.f.). Retrieved October 4, 2024, from https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k
53. *ACRIMA Dataset*. (s.f.). Retrieved October 4, 2024, from https://www.kaggle.com/datasets/toaharahmanratul/acrima-dataset
54. *Messidor-2 - ADCIS*. (s.f.). Retrieved October 4, 2024, from https://www.adcis.net/en/third-party/messidor2/
55. *DRIVE Digital Retinal Images for Vessel Extraction*. (s.f.). Retrieved October 4, 2024, from https://www.kaggle.com/datasets/andrewmvd/drive-digital-retinal-images-for-vessel-extraction
56. *Retinal Fundus Multi-disease Image Dataset (RFMiD) | IEEE DataPort*. (n.d.). Retrieved October 4, 2024, from https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid
57. *REFUGE2*. (s.f.). Retrieved October 4, 2024, from https://www.kaggle.com/datasets/victorlemosml/refuge2
58. *DR HAGIS Dataset | Papers With Code*. (s.f.). Retrieved October 4, 2024, from https://paperswithcode.com/dataset/dr-hagis
59. *Inspire Datasets | Department of Ophthalmology and Visual Sciences*. (s.f.). Retrieved October 4, 2024, from https://medicine.uiowa.edu/eye/inspire-datasets
60. Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J. C., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M., Danno, R., Elie, D., Massin, P., Viktor, Z., Erginay, A., Laÿ, B., & Chabouis, A. (2013). TeleOphta: Machine learning and image processing methods for teleophthalmology. *IRBM*, *34*(2), 196–203. https://doi.org/10.1016/J.IRBM.2013.01.010
61. *Inspire Datasets | Department of Ophthalmology and Visual Sciences*. (s.f.). Retrieved October 4, 2024, from https://medicine.uiowa.edu/eye/inspire-datasets
62. *Drishti-GS Dataset Webpage*. (s.f.). Retrieved October 4, 2024, from https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php

63. Septiarini, A., Hamdani, H., Setyaningsih, E., Junirianto, E., & Utaminingrum, F. (2023). Automatic Method for Optic Disc Segmentation Using Deep Learning on Retinal Fundus Images. *Healthcare Informatics Research*, *29*(2), 145. https://doi.org/10.4258/HIR.2023.29.2.145
64. Akram, M. U., Khalid, S., & Khan, S. A. (2013). Identification and classification of microaneurysms for early detection of diabetic retinopathy. *Pattern Recognition*, *46*(1), 107–116. https://doi.org/10.1016/J.PATCOG.2012.07.002
65. Pham, Q. T. M., Ahn, S., Song, S. J., & Shin, J. (2020). Automatic Drusen Segmentation for Age-Related Macular Degeneration in Fundus Images Using Deep Learning. *Electronics 2020, Vol. 9, Page 1617*, *9*(10), 1617. https://doi.org/10.3390/ELECTRONICS9101617
66. Kou, C., Li, W., Liang, W., Yu, Z., & Hao, J. (2019). Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network. *Journal of Medical Imaging*, *6*(2), 1. https://doi.org/10.1117/1.JMI.6.2.025008
67. *Data analysis*. (2018, diciembre 3). EyePACS; #creator. https://www.eyepacs.com/data-analysis
68. OCT-AND-EYE-FUNDUS-DATASET: Dataset of eye fundus and OCT images for the study of diabetic macular edema and diabetic retinophaty. (s.f.). https://github.com/Traslational-Visual-Health-Laboratory/OCT-AND-EYE-FUNDUS-DATASET
69. Benítez, V. E. C., Castro Matto, I., Román, J. C. M., Noguera, J. L. V., García-Torres, M., Ayala, J., Pinto-Roa, D. P., Gardel-Sotomayor, P. E., Facon, J., & Grillo, S. A. (2021). Dataset from fundus images for the study of diabetic retinopathy [Data set]. Zenodo. https://zenodo.org/records/4891308#.YtRwDS-plQI
70. Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. v, Adam, H., Ai, G., & Brain, G. (n.d.). *Searching for MobileNetV3*. https://arxiv.org/abs/1905.02244
71. Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian Diabetic Retinopathy Image Dataset (IDRiD). *IEEE Dataport*. https://doi.org/10.21227/H25W98
72. Raudonis, V.; Kairys, A.; Verkauskiene, R.; Sokolovska, J.; Petrovski, G.; Balciuniene, V.J.; Volke, V. Automatic Detection of Microaneurysms in Fundus Images Using an Ensemble-Based Segmentation Method. *Sensors* **2023**, *23*, 3431. https://doi.org/10.3390/s23073431
73. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11045 LNCS*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

74. Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). *Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation*. https://arxiv.org/abs/1802.06955v5
75. *[1802.02611] Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. (n.d.). Retrieved April 20, 2025, from https://arxiv.org/abs/1802.02611
76. Shyamalee, T., & Meedeniya, D. (2022). Attention U-Net for Glaucoma Identification Using Fundus Image Segmentation. *2022 International Conference on Decision Aid Sciences and Applications, DASA 2022*, 6–10. https://doi.org/10.1109/DASA54658.2022.9765303
77. Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S. L., Sanders, K. M., & Baker, S. A. (2021). RV-GAN: Segmenting Retinal Vascular Structure in Fundus Photographs Using a Novel Multi-scale Generative Adversarial Network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12908 LNCS*, 34–44. https://doi.org/10.1007/978-3-030-87237-3_4/TABLES/1
78. Bajwa, M. N., Singh, G. A. P., Neumeier, W., Malik, M. I., Dengel, A., & Ahmed, S. (2020). G1020: A Benchmark Retinal Fundus Image Dataset for Computer-Aided Glaucoma Detection. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN48605.2020.9207664