



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias en Inteligencia Artificial

Modelo de inteligencia artificial para clasificación y segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo.

TESIS

Que como parte de los requisitos para obtener el grado de Maestro en Ciencias en Inteligencia Artificial

Presenta:

Ing. Javier Anguiano Almejo

Dirigida por:

Mtro. Gendry Alfonso Francia

SINODALES

Mtro. Gendry Alfonso Francia
Presidente
Dr. Saúl Tovar Arriaga
Secretario
Dr. Jesús Carlos Pedraza Ortega
Vocal
Dr. Marco Antonio Aceves
Fernández
Suplente
Dra. Mariana Badillo Fernández
Suplente

Centro Universitario
Querétaro, Qro.
Octubre de 2023
México



Dirección General de Bibliotecas y Servicios Digitales
de Información



Modelo de inteligencia artificial para clasificación y
segmentación de atrofia peripapilar Alfa y Beta en
imágenes de fondo de ojo

por

Javier Anguiano Almejo

se distribuye bajo una [Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0
Internacional](#).

Clave RI: IGMAC-309239

Dedicatoria

A mi querido abuelo, quien ha sido fuente de inspiración en mi vida, y a mis futuros hijos, a quienes dedico el tiempo y esfuerzo invertidos en este trabajo de investigación. Lo realicé con la firme esperanza de siempre contribuir al bienestar y apoyo de nuestra familia y de la sociedad.

Agradecimientos

Mis abuelos representan el corazón y el alma detrás de este proyecto. A mi abuela, siempre le estaré eternamente agradecido por su amor inquebrantable y la fe que depositó en mí. Mi abuelo, con su sabiduría y fortaleza, me brindó el aliento necesario para avanzar en cada fase de este proyecto. Aunque su partida fue transcurso de este proceso, su confianza en mí y sus valiosas enseñanzas se convirtieron en el faro que iluminó mi camino. Que en paz descanse.

Un agradecimiento especial para mis padres, Jaime Anguiano Solorzano y María de Jesús Almejo Ibarra, quienes desde mis primeros pasos me enseñaron el verdadero valor del esfuerzo y la perseverancia. Han sido el pilar fundamental en cada etapa de mi vida, y este logro es también suyo gracias a su amor y apoyo incondicional.

No puedo pasar por alto el apoyo fundamental de mi hermana y hermano, Carolina Anguiano Almejo y David Anguiano Almejo. Sin ellos, quizás no hubiera emprendido este proyecto. Su motivación, aliento y fe inquebrantable en mí han sido fuerzas motrices que me llevaron a culminar esta etapa.

A mis amigos y a mi pareja en la maestría, les debo inmensamente. Gracias por ser mi refugio tanto académico como emocional, por compartir conmigo cada risa, cada desafío y cada triunfo en esta travesía.

Mi reconocimiento y gratitud a la Universidad Autónoma de Querétaro y al Instituto Mexicano de Oftalmología por equiparme con las herramientas y brindarme las oportunidades y el entorno adecuado para mi desarrollo profesional. También agradezco a todos los profesionales que generosamente compartieron su tiempo y expertise para la consolidación de este proyecto.

Y, finalmente, mi profunda gratitud al Mtro. Gendry Alfonso Francia, mi director de tesis, por su guía y respaldo académico. Su paciencia y dedicación resultaron cruciales en la culminación de este proyecto.

Resumen

El Glaucoma es una patología oftálmica que requiere la identificación temprana para garantizar tratamientos adecuados y evitar la pérdida de visión. Un indicador clave para su detección es la presencia de atrofia peripapilar. En este trabajo Maestría, proponemos un modelo de inteligencia artificial para la clasificación y segmentación de atrofia peripapilar, enfocándonos especialmente en sus subclases: Atrofia Alfa y Beta. Esta propuesta busca marcar un precedente en la construcción de sistemas de asistencia para el diagnóstico más precisos y robustos.

La principal contribución de este estudio es el desarrollo de la primera base de datos para la segmentación de estas subclases de atrofia peripapilar, validada por especialistas en glaucoma del Instituto Mexicano de Oftalmología. Anteriores investigaciones abordaban la segmentación y clasificación de la atrofia de forma binaria, omitiendo la diferenciación crucial entre Alfa y Beta.

Además, hemos comparado metodologías de segmentación semántica y de instancias. Después de un análisis detallado de modelos de segmentación semántica como FCN, SegNet y Unet con variaciones en sus Backbones, y contrastándolos con el modelo MaskRCNN (que integra la segmentación de instancias), determinamos que los modelos de detección de objetos ofrecen una mejor precisión para nuestro objetivo. A esto le sumamos un estudio de ablación, que nos permitió determinar la combinación óptima de hiperparámetros para maximizar el rendimiento del modelo.

La segmentación de atrofia peripapilar presenta desafíos inherentes, como su estructura irregular y difusa, el desequilibrio entre clases y el tamaño reducido de la Atrofia Alfa. Aunque existen modelos con alto rendimiento en el área, ninguno diferencia entre las atrofas peripapilares Alfa y Beta, resaltando la originalidad y relevancia de nuestro trabajo.

Con nuestro enfoque y resultados, no solo destacamos en el ámbito de la oftalmología e inteligencia artificial, sino que también establecemos un sólido punto de partida para investigaciones futuras en esta dirección específica.

Abstract

Glaucoma is an ophthalmic pathology that requires early identification to ensure appropriate treatments and prevent vision loss. A key indicator for its detection is the presence of peripapillary atrophy. In this Master's thesis, we propose an artificial intelligence model for the classification and segmentation of peripapillary atrophy, focusing specifically on its subclasses: Alpha and Beta Atrophy. This proposal aims to set a precedent in the construction of more accurate and robust diagnostic assistance systems.

The main contribution of this study is the development of the first database for the segmentation of these subclasses of peripapillary atrophy, validated by glaucoma specialists from the Mexican Institute of Ophthalmology. Previous research approached the segmentation and classification of atrophy in a binary manner, overlooking the crucial differentiation between Alpha and Beta.

Furthermore, we compared semantic segmentation methodologies with instance-based ones. After a detailed analysis of semantic segmentation models like FCN, SegNet, and Unet with variations in their Backbones, and contrasting them with the MaskRCNN model (which integrates instance segmentation), we determined that object detection models offer better accuracy for our goal. Added to this, we conducted an ablation study, which allowed us to determine the optimal combination of hyperparameters to maximize the model's performance.

The segmentation of peripapillary atrophy presents inherent challenges, such as its irregular and diffuse structure, the class imbalance, and the smaller size of Alpha Atrophy. Although there are high-performing models in the area, none differentiate between Alpha and Beta peripapillary atrophies, highlighting the originality and relevance of our work.

With our approach and results, we not only stand out in the field of ophthalmology and artificial intelligence but also establish a solid starting point for future research in this specific direction.

ÍNDICE

Comentado [GAF1]: Agregar los sub epígrafes

DATOS GENERALES	10
I. INTRODUCCIÓN.....	11
1.1 GLAUCOMA	11
1.1.1 DIAGNÓSTICO.....	11
1.2 ATROFIA PERIPAPILAR.....	12
1.3 INTELIGENCIA ARTIFICIAL EN EL DIAGNÓSTICO MÉDICO	14
1.4 JUSTIFICACIÓN	15
1.5 DESCRIPCIÓN DEL PROBLEMA.....	16
1.6 HIPÓTESIS	17
1.7 OBJETIVOS.....	17
1.7.1 Objetivo General:.....	17
1.7.2 Objetivos Específicos:	17
II. ANTECEDENTES.....	18
III. MARCO TEÓRICO	23
3.1 Estructura Anatómica del Ojo.....	23
3.2 Imagen de fondo de ojo	24
3.3 Segmentación de atrofia peripapilar	24
3.4 Redes Neuronales	25
3.5 Aprendizaje Profundo.....	27
3.6 Modelos de Segmentación	29
3.6.1 Segmentación Semántica.....	29
3.6.1.1 Red Totalmente Convolutiva (FCN).....	30
3.6.1.2 Segmentation Network (SegNet)	30
3.6.1.3 Unet.....	31
3.6.1.4 Redes Codificadoras.....	32
3.6.1.4.1 ResNet.....	33
3.6.1.4.2 VGG.....	33
3.6.1.4.3 MobileNet	33
3.6.2 Segmentación de Instancias	33
3.6.2.1 Detección de Objetos	34
3.6.2.2 Mask RCNN.....	35

3.6.2.3 Cascade Mask RCNN	36
3.6.2.4 Mask Scoring RCNN	37
3.6.3 Métricas de Evaluación.....	38
3.6.3.1 Precisión.....	40
3.6.3.2 Sensibilidad (Recall).....	40
3.6.3.3 Puntuación F1 (F1 Score).....	40
3.6.3.4 Intersección sobre Unión (IoU).....	41
3.6.3.4 Average Precision (AP).....	41
3.6.3.4.1 AP@0.5 y AP@0.75	42
3.6.3.4.2 AP@[.5:.05:.95].....	42
3.6.3.4.3 APs, APm y API.....	42
IV. MATERIALES Y METODOLOGÍA	42
4.1 Google Colaboratoy.....	43
4.2 Python.....	43
4.3 MMDetection.....	43
4.4 PyTorch.....	44
4.5 Keras.....	45
4.6 Roboflow	45
4.7 Base de Datos	46
4.7.1 ORIGA.....	46
4.7.2 Retina	46
4.7.3 DRISHTI-GSI.....	47
4.8 Hardware.....	47
4.9 Metodología.....	48
4.10 Implementación	51
4.10.1 Propuesta de Investigación	51
4.10.2 Búsqueda de Base de Datos	52
4.10.3 Selección de línea de segmentación	52
4.10.3 Técnicas de Preprocesamiento.....	53
4.10.3.1 Extracción de región de interés (ROI).....	54
4.10.3.2 Aumento de datos volteo horizontal y vertical.....	54
4.10.3.3 Ajuste de contraste	55

4.10.3.4 Blur.....	55
4.10.3.5 Ajuste de exposición de imagen.....	56
4.10.4 Exploración de Hiperparámetros.....	57
4.10.4.1 Análisis Mask Loss Weight:	57
4.10.4.2 Análisis de función de pérdida: Cross Entrophy y Focal Loss.....	57
4.10.4.3 Análisis de Bbox Loss: L1 Loss, SmoothL1, DIoU, CIoU.....	59
4.10.4.5 Análisis de Optimizador: SGD y Adam.....	63
4.10.4.6 Análisis de Backbone: ResNet50 y ResNet101	65
4.10.4.7 Ajuste de cuadros de anclaje.....	66
4.10.5 Análisis de modelos: Cascade Mask RCNN y Mask Scoring	67
4.10.6 Entrenamiento General	67
V. RESULTADOS Y DISCUSIÓN	68
5.1 Adquisición de la base de datos.....	68
5.2 Resultados de Segmentación Semántica	69
5.3 Resultados de Segmentación de Instancias.....	74
5.4 Selección de línea de segmentación.....	76
5.5 Análisis de Errores.....	78
5.5.1 ROI.....	79
5.6 Análisis de Hiperparámetros.....	81
5.6.1 Análisis de peso de pérdida de máscara	81
5.6.2 Análisis de funciones de pérdida en regresión de Bbox.....	83
5.6.3 Análisis de Épocas.....	85
5.6.4 Análisis de Función de pérdida.....	87
5.6.5 Análisis de Optimizador	88
5.6.6 Análisis de Backbone.....	90
5.6.7 Análisis de Bbox Size.....	91
5.7 Análisis de técnicas de Preprocesamiento	96
5.7.1 Contraste	96
5.7.2 Aumento de Datos	97
5.8 Análisis de Modelos	98
5.9 Entrenamiento de los mejores modelos	100
CONCLUSIONES	103

REFERENCIAS..... 105
ANEXOS 112

DATOS GENERALES

Título del proyecto de Tesis: Modelo de inteligencia artificial para clasificación y segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo

Nombre del Alumno: Javier Anguiano Almejo

Número de expediente: 309239

Programa de Estudios para realizar: Maestría en Ciencias en Inteligencia Artificial

Director de Tesis: Mtro. Gendry Alfonso Francia

Secretario: Dr. Saúl Tovar Arriaga

Vocal: Dr. Jesús Carlos Pedraza Ortega

Lugar donde se realizará la investigación: Universidad Autónoma de Querétaro, Campus Aeropuerto.

Línea de investigación: Ingeniería Biomédica

Tipo de investigación: Aplicada

Horario de trabajo: 7:00 am a 3:00 pm

I. INTRODUCCIÓN

1.1 GLAUCOMA

El glaucoma es una neuropatía óptica progresiva caracterizada por alteraciones específicas del campo visual asociado a la muerte de las células ganglionares de la retina y cambios morfológicos específicos en el nervio óptico [1]. De acuerdo con la Organización Mundial de la Salud (OMS), el glaucoma es considerada como la segunda causa de ceguera en el mundo y como la primera causa de ceguera irreversible; posicionándose con una prevalencia del 1.5% de la población mundial. En México se estima que el 1.3% de la población lo padece, pero aproximadamente el 50% aún no se ha diagnosticado [2], [3].

El glaucoma se puede clasificar en cuatro diferentes tipos de acuerdo con su etiología: Glaucoma primario, glaucoma congénito, glaucoma secundario y glaucoma absoluto. El glaucoma primario posee mayor relevancia epidemiológica por tener una alta prevalencia y se puede dividir en: glaucoma primario de ángulo abierto (GPAA) y glaucoma primario de ángulo cerrado (GPAC) [1], [3].

El GPAA expresa una presencia del 80 al 85% en los casos de diagnóstico de glaucoma [3]. Este tipo de patología se presenta de forma asintomática durante las etapas iniciales de desarrollo; conforme aumenta el grado de la afección se va presentando una pérdida de la función visual irreversible, es por lo que en la mayoría de los casos se ve reflejada cuando se encuentra en un estado avanzado.

Comentado [GAF2]: Por lo que

1.1.1 DIAGNÓSTICO

El diagnóstico de glaucoma se realiza con base a estudios que evalúan el daño funcional y estructural en la retina, además de la evaluación de aspectos como el grosor corneal, presión intraocular y el ángulo iridocorneal [4], [5].

El daño glaucomatoso temprano puede ser difícil de identificar mediante la evaluación funcional de la visión, ya que esta se ve afectada cuando se encuentra en estados avanzados, por ello es por lo que se requiere de una observación minuciosa de las características morfológicas del nervio óptico.

Comentado [GAF3]: Por lo que

Existen diversos medios de diagnóstico que permiten la evaluación estructural del nervio óptico como lo son: la tomografía de coherencia óptica, tomografía de Heidelberg, análisis de fibras

nerviosas, pero la fotografía de fondo de ojo es una de las técnicas más utilizadas por los especialistas, ya que esta puede ser obtenida a través de los oftalmoscopios los cuales tiene una alta presencia en las clínicas debido a su portabilidad y costo accesible [6].

No obstante, la evaluación clínica de la imagen de fondo de ojo a través de un oftalmoscopio requiere de la interpretación experta de un oftalmólogo altamente capacitado que provea de una evaluación semi-cuantitativa [4].

1.2 ATROFIA PERIPAPILAR

Entre los factores estructurales que pueden ser observados dentro de las imágenes de fondo de ojo están: tamaño del disco óptico, tamaño y forma del nervio neuro retiniano, capas de fibras nerviosas retinianas, presencia de atrofia peripapilar y presencia de hemorragias retinianas o del disco óptico [5].

Se ha demostrado que la atrofia peripapilar (APP) presenta una fuerte correlación con el diagnóstico de glaucoma y proporcionar información valiosa sobre los estados normales y afectados por miopía [7].

Los exámenes histológicos revelan características distintivas de la zona Alfa, como la presencia de la Membrana de Bruch (BM) y el EPR, mostrando este último una estructura irregular como se muestra en la Figura 1. Curiosamente, la zona Alfa es frecuentemente el área más grande ubicada en el sector horizontal temporal, seguida por el área temporal inferior y la región temporal superior. Por el contrario, suele ser más pequeño y menos frecuente en la región nasal por la desaparición de fotorreceptores y finalmente el cierre de la coriocapilar. Funcionalmente, esto se correlaciona con un escotoma absoluto en la perimetría, lo que indica un punto ciego en el campo visual [8]

Comentado [GAF4]: Espacio con la referencia

Comentado [GAF5]: minúscula

Comentado [GAF6]: Trata de poner la figura lo más cerca posible de la referencia

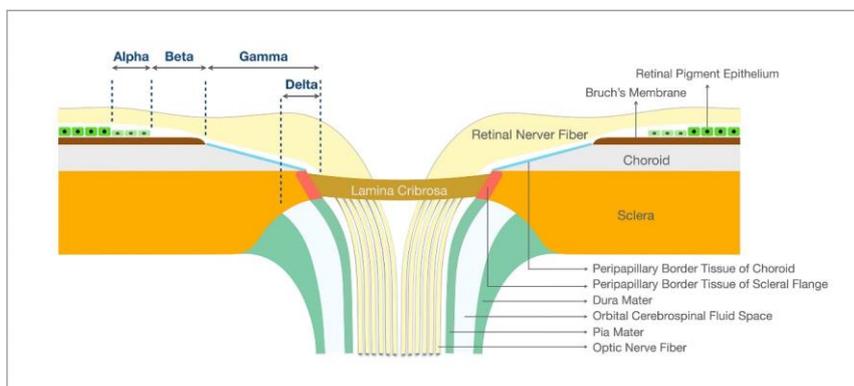


Figura 1. Figura esquemática de las zonas peri papilares[8].

La zona Beta se caracteriza por la presencia de la Membrana de Bruch y la ausencia de EPR. Esta región revela cambios estructurales, como el cierre de la coriocapilar y la ausencia de fotorreceptores retinales, especialmente cerca del borde del disco óptico. En contraste, la parte periférica de la zona Beta exhibe coriocapilares abiertos y la presencia de fotorreceptores, pero carece de EPR. Las implicaciones de estos hallazgos son cruciales. Sugieren que el desarrollo de la zona Beta involucra una secuencia de cambios degenerativos, comenzando con la pérdida de células EPR, seguida por la desaparición de fotorreceptores y finalmente el cierre de la coriocapilar. Funcionalmente, esto se correlaciona con un escotoma absoluto en la perimetría, lo que indica un punto ciego en el campo visual [8].

Varios estudios han explorado la relación entre la zona Beta y el glaucoma, revelando correlaciones entre su ubicación y la progresión del glaucoma en ojos miopes, la progresión del campo visual y los defectos de la capa de fibras nerviosas de la retina. A pesar de que su valor diagnóstico en el glaucoma varía entre los estudios, la zona Beta sigue siendo un elemento esencial en la evaluación morfológica del glaucoma [8]–[10]

En una visualización de la imagen de fondo de ojo, como se muestra en la Figura 2. La zona Beta tiene una ubicación adyacente al Disco Óptico (DO), mientras que la zona Alfa se encuentra posteriormente, bordeando la periferia de la zona Beta. En la Figura 2, se presenta una vista detallada del nervio óptico en una imagen 3D de fondo de ojo, donde se pueden distinguir

Comentado [GAF7]: Preferiría el párrafo a continuación de descripción de la atrofia Beta antes que este, para entrar en contexto de lo que es esa estructura.

Comentado [GAF8]: Dónde está definido DO?

Comentado [GAF9]: O siempre con mayúsculas o siempre con minúsculas

Comentado [GAF10]: Lo mismo que el comentario anterior

claramente las áreas correspondientes a la APP Alfa y Beta. Se ilustra la atrofia Beta, delineada por una región en color blanco, y el daño Alfa, representado por una región en color negro.

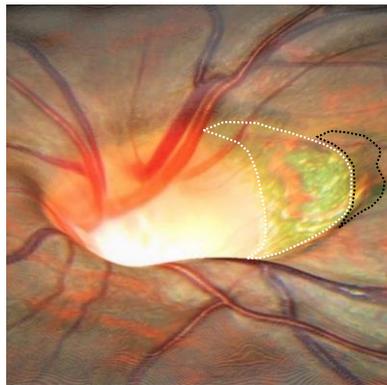


Figura 2. Acercamiento del nervio óptico de imagen 3D de fondo de ojo con atrofia peripapilar Alfa y Beta[11].

La evaluación de los cambios estructurales asociados a glaucoma por medio del análisis de fotografías de fondo de ojo permite realizar un registro detallado y sistemático de cada una de las características de la retina y nervio óptico que pueden ser utilizadas por los especialistas de la visión en el control y diagnóstico de glaucoma [9].

1.3 INTELIGENCIA ARTIFICIAL EN EL DIAGNÓSTICO MÉDICO

La Inteligencia Artificial (IA) es un campo de la informática que busca desarrollar sistemas capaces de realizar tareas que requieran de inteligencia humana, tales como el reconocimiento de patrones, la toma de decisiones, y el procesamiento del lenguaje natural, entre otros [12]. Con la masiva acumulación de datos y el desarrollo de algoritmos avanzados, la IA ha experimentado un rápido crecimiento en las últimas décadas, desbloqueando un potencial que abarca desde la robótica hasta el análisis financiero y, particularmente, el diagnóstico médico [13].

En el campo médico, la IA ha revolucionado el modo en que se analizan y procesan los datos clínicos, ofreciendo herramientas precisas que asisten en el diagnóstico, el pronóstico y el tratamiento de enfermedades. En el ámbito de la oftalmología, la adaptación de algoritmos de IA en la interpretación de imágenes ha mostrado ser especialmente prometedora [14].

La llegada del aprendizaje profundo, un subconjunto de la inteligencia artificial (IA) que utiliza redes neuronales artificiales, ha dado lugar a avances en las tareas de reconocimiento de imágenes.

Comentado [GAF11]: Espacios innecesarios, revisar en todo el documento.

Comentado [GAF12]: Pones toda una introducción de la parte médica pero nada de la parte técnica. Qué es la IA?, cómo ha evolucionado en el área médica, cómo se aplica en imágenes de retina. Son algunas ideas.

Esto ha generado un impacto positivo en el análisis de imágenes médicas, incluidas fotografías de fondo de ojo, tomografías de coherencia óptica (OCT) y angiogramas con fluoresceína, que se usan comúnmente en la práctica oftalmológica [15], [16].

La identificación y segmentación de estructuras anatómicas en imágenes retinianas sirven como componentes básicos para cualquier sistema de diagnóstico automatizado relacionado con enfermedades de la retina. Estas tareas, a pesar de su importancia crítica, siguen siendo desafíos sin resolver en el campo [17].

1.4 JUSTIFICACIÓN

El glaucoma alcanzó a nivel mundial un total de 60.5 millones de casos en el 2010, y de acuerdo con estudios de metaanálisis epidemiológicos que estimaron la prevalencia de glaucoma primario de ángulo abierto y cerrado, para los años 2020 y 2040 en pacientes entre los 40 y 80 años, obteniendo proyecciones donde revelaron que en el año 2013 vivían 64.3 millones de personas afectadas por glaucoma en el mundo, y que para el 2020 se incrementó a 76 millones, para finalmente alcanzar 111.8 millones en el 2040 lo que representa el 3.54% de la población. Los resultados de los estudios expresan un comportamiento de crecimiento desproporcionado. Dado que el glaucoma es una enfermedad que progresa lentamente con daño neuronal irreversible, el diagnóstico precoz y la monitorización sensible de la progresión son fundamentales para el tratamiento del glaucoma [18].

De acuerdo con las guías prácticas del IMSS, en México, el diagnóstico de glaucoma se realiza en hospitales de segundo nivel de atención médica mediante el estudio de perimetría automatizada, con el cual a su vez es posible determinar el grado de la patología; no obstante, este estudio va acompañado de otras pruebas que evalúan la calidad de la visión y rasgos morfológicos en el nervio óptico, como lo es la atrofia peripapilar Beta [18]. Es por ello por lo que es necesario desarrollar estrategias de salud pública para garantizar la detección oportuna y tratamiento temprano con el objetivo de retrasar la pérdida visual.

Con la adaptación de los principales avances en las tecnologías de diagnóstico que ofrecen una visión más amplia del estado de la retina y de las enfermedades oculares, los médicos y los proveedores de atención médica han adoptado nuevos métodos de medición de la actividad visual en un esfuerzo por mejorar la atención al paciente, acelerar el proceso de gestión y minimizar la cantidad de herramientas necesarias para cada evaluación.

Comentado [GAF13]: Por lo que

Con la finalidad de hacer uso de los más recientes avances de la tecnología entorno al procesamiento de imágenes médicas con la aplicación de técnicas de aprendizaje profundo, se desarrolla la presente investigación actuando como soporte al proyecto de “Detección temprana de Glaucoma asistido por redes neuronales artificiales” cuyo trabajo permitirá la clasificación y segmentación de APP Alfa y Beta en imágenes de fondo de ojo, para apoyar a brindar un diagnóstico certero y con una mejor gestión de datos para los oftalmólogos.

1.5 DESCRIPCIÓN DEL PROBLEMA

La mayoría de los estudios de segmentación de APP basados en aprendizaje profundo se llevan a cabo en condiciones experimentales con base de datos reducidas y de información limitada como se observa en la Figura 3. Esto puede proporcionar conocimiento inadecuado para el uso de aplicaciones de inteligencia artificial en entornos de atención médica heterogénea del mundo real [14].

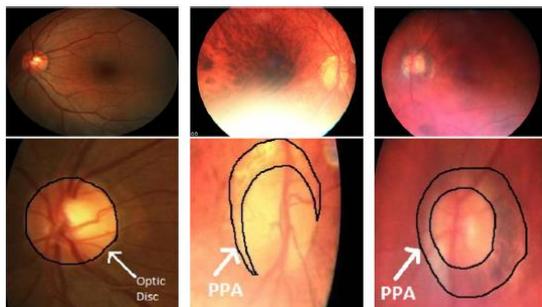


Figura 3. Comparación de imágenes de fondo de ojo con atrofia peripapilar en condiciones de iluminación diferentes [19].

La segmentación de la APP Alfa y Beta es una tarea compleja debido a una multitud de factores interrelacionados. En primer lugar, la variabilidad y la ambigüedad en la apariencia de la APP entre individuos y etapas de la enfermedad agregan una capa de complejidad al proceso de segmentación. En segundo lugar, la calidad de las imágenes de fondo de ojo puede diferir ampliamente, influenciada por elementos como el dispositivo de imagen, su configuración y las condiciones oculares del paciente. Esta variabilidad en la calidad de la imagen puede afectar directamente el rendimiento de los modelos de segmentación.

Además, el entrenamiento de modelos robustos de aprendizaje profundo depende de la disponibilidad de extensos conjuntos de datos de imágenes de fondo de ojo etiquetadas por

expertos. Sin embargo, la creación de dichos conjuntos de datos requiere de arduo trabajo y tiempo, lo que a menudo conduce a una escasez de datos etiquetados de calidad.

Otro desafío radica en las capacidades de generalización de estos modelos de IA. Los modelos entrenados en conjuntos de datos específicos pueden tener un rendimiento inferior cuando se prueban en datos de diferentes fuentes o poblaciones, lo que representa un obstáculo importante para la implementación en el mundo real de estas soluciones de IA.

Además, la presencia de APP en escenarios del mundo real rara vez está aislada. A menudo, la APP coexiste con otras patologías que afectan la estructura morfológica de la imagen del fondo de ojo. Estas condiciones simultáneas introducen una mayor variabilidad y complejidad en las imágenes, lo que complica aún más la tarea de segmentación. Por lo tanto, un modelo lo suficientemente robusto para manejar estas complejidades, capaz de distinguir las diferencias matizadas entre las zonas Alfa y Beta de la APP y dar cuenta de patologías adicionales, es el objetivo primordial en este campo.

1.6 HIPÓTESIS

Un modelo basado en técnicas de aprendizaje profundo para la detección y segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo, obtiene resultados comparables a los trabajos del estado del arte.

1.7 OBJETIVOS

1.7.1 Objetivo General:

Diseñar y desarrollar un algoritmo de clasificación y segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo mediante la integración de técnicas de inteligencia artificial.

1.7.2 Objetivos Específicos:

- Etiquetar y delimitar las regiones de atrofia Alfa y Beta de base de datos públicas.
- Diseñar e implementar una red de clasificación de atrofia Alfa y Beta.
- Diseñar e implementar una red de segmentación de atrofia Alfa y Beta.

Comentado [GAF14]: Poner hipótesis antes de objetivos

II. ANTECEDENTES

El avance de la Inteligencia Artificial (IA) ha desatado una transformación significativa en diversos sectores, particularmente en la salud. En el dominio de la oftalmología, la IA ha demostrado un potencial extraordinario, potenciando las capacidades diagnósticas y las estrategias de manejo de enfermedades.

En este enfoque, la importancia de una segmentación precisa es crucial, especialmente cuando se trata de la APP. La segmentación detallada de la APP, que comprende la zona Alfa externa y la zona Beta interna, con un nivel de atrofia más severo, es esencial para el seguimiento de la presencia y evolución del glaucoma.

La aplicación de la IA en este componente de la oftalmología no solo mejora la precisión del diagnóstico, sino que también aporta un entendimiento más profundo de la progresión de la enfermedad. Los modelos de segmentación impulsados por IA, en su constante evolución, prometen abrir nuevas vías para la detección temprana y el manejo del glaucoma y otras enfermedades relacionadas.

En la exploración de los últimos avances en la construcción de modelos de clasificación y segmentación de imágenes del fondo del ojo, se ha notado una tendencia hacia la integración de técnicas de IA para la detección de patrones. Esto permite obtener resultados altamente confiables en el procesamiento de imágenes del fondo del ojo. La Tabla 1 muestra trabajos recientes en la identificación de atrofia peripapilar en imágenes del fondo del ojo. Este avance en el campo de la oftalmología es un testimonio del potencial inexplorado de la IA en la medicina.

Tabla 1. Estado del Arte.

Autor	Artículo	Técnica	Base de datos	Métricas
Cheng-Kai Lu et al. 2010	Automatic Parapapillary Atrophy Shape Detection and Quantification in Colour Fundus Images	Para segmentar y cuantificar la OD y la PPA, se utiliza una combinación de varias técnicas, incluido el filtro de exploración, el umbral, el crecimiento de la región y el modelo Chan-Vese (CV) modificado con una restricción de forma.	Propia [40]	ACC: 91.3%
Chisako Muramatsu et al. 2011	Computerized Detection of Peripapillary Chorioretinal Atrophy by Texture Analysis	Análisis de textura para detectar las regiones de PPA.	Propia [171]	SEN: 73% ESP: 94%
Jun Cheng et al. 2012	Peripapillary Atrophy Detection by Sparse Biologically Inspired Feature Manifold	Esta técnica es una mejora de la técnica BIF e implica un aprendizaje de transferencia dispersa negativa para obtener resultados precisos.	SCORM [1584]	PREC: 90%
Hanxiang Li. 2018	Automatic segmentation of PPA in retinal images	Utiliza segmentos de líneas radiales orientados uniformemente para detectar los puntos límite de DO y DO aumentado (A-DO). Luego, se obtienen límites suaves y no discretos de A-DO y DO, se ajustan a elipses mediante el método de mínimos cuadrados. Se obtiene el límite de PPA, como el área que no se superpone de estos dos límites.	Propia H. Beijing Tongren [100]	F1: 67.1%
Anindita Septiaini et al. 2018	Automatic detection of peripapillary atrophy in retinal fundus images using statistical features	Red neuronal de retropropagación (BPNN)	D1 D2 RIM-ONE [155]	SEN: 93%, 100%, 100% ESP: 85%, 83%, 91% PREC: 95%, 96%, 96%
Anindita Septiaini, 2018	Peripapillary Atrophy Detection in Fundus Images Based on Sectors with Scan Lines Approach	Metodo de base de conocimientos para clasificación de las características obtenidas.	D1 D2 RIM-ONE [155]	SEN: 91%, 100%, 100% ESP: 72%, 64%, 71% PREC: 92%, 90%, 89%
Fakhira Zahra Zulfira et al. 2019	Multi-Class Peripapillary Atrophy for Detecting Glaucoma in Retinal Fundus Image	Maquina de vectores de soporte (SVM)	RIM-ONE KAGGLE [210]	SEN: 89%, 99% ESP: 86%, 87% PREC: 95% 94%
Yidong Chai et al. 2019	A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images	Red neuronal convolucional (CNN)	Multiples [1000]	PREC: 89.28%
Fakhira Zahra Zulfira et al. 2020	Detection of Multi-Class Glaucoma Using Active Contour Snakes and Support Vector Machine	Uso de serpiente de contorno activa para obtener el valor de OC y OD para medir la CDR, segmentación de Otsu para PPA y una SVM para la clasificación de clases de Glaucoma	RIMONE KAGGLE [210]	SEN: 90%, 97% ESP: 87%, 88% PREC: 95%
Ambika Sharma et al. 2020	Deep learning to diagnose Peripapillary Atrophy in retinal images along with statistical features	ResNet-50, Data augmentation, transfer learning	Drishti Refugee Rim Messidor Drive Drions AIIMS [600]	SEN: 90%, 97% PREC: 95%
Fakhira Zahra Zulfira et al. 2021	Segmentation technique and dynamic ensemble selection to enhance glaucoma severity detection	Uso de serpiente de contorno activa para obtener el valor de OC y OD para medir la CDR, segmentación de Otsu para PPA y extracción de características con matriz de co-ocurrencia de nivel de grises (GCLM) y uso del clasificador de selección de conjunto dinámico (DES) para clasificación de glaucoma.	RIM-ONE KAGGLE MESSIDOR [250]	SEN: 96% ESP: 88% PREC: 96%
Mengxuan Li et al. 2021	Peripapillary Atrophy Segmentation with Boundary Guidance	Propone un bloque de guía de límites junto con una función de pérdida de contorno para mejorar el rendimiento de la segmentación de PPA en los límites.	-	F1: 80.06% IoU: 67.29%
Abdullah Almansour et al. 2022	Peripapillary atrophy classification using CNN deep learning for glaucoma screening	El modelo se desarrolló en función de la localización de la región de interés (ROI) utilizando una máscara de redes neuronales convolucionales basadas en regiones R-CNN y una red de clasificación para la presencia de PPA utilizando algoritmos de aprendizaje profundo de CNN	Bin Rushed [195] Magrabi [94] HRF [45] Kaggle [495] ORIGA [49] EyePacs [487] KAIMRC [1178]	AUC: 0.83% LOCAL AUC: 0.89% PUB AUC: 0.87% COMB

Cheng-Kai Lu et al. (2010), realizó una extracción de la APP utilizando una combinación de métodos que incluyen el filtro de exploración, umbralización, crecimiento y el modelo Chan-Vese. Este método utiliza características para la segmentación de clases. Obteniendo una exactitud de 91.3% para una base de datos de 40 imágenes [20].

Chisako Muramatsu (2011), propuso una detección computarizada de la atrofia coriorretiniana peripapilar, que consiste en utilizar una técnica de análisis de textura en imágenes de fondo de ojo, las cuales fueron obtenidas en campo por los mismos desarrolladores del proyecto, clasificando las imágenes en: Atrofia moderada o severa. La sensibilidad y especificidad alcanzada fue de 73% y 94% respectivamente [21].

Jun Cheng (2012), introdujo un enfoque para la detección de APP basado en un colector de características dispersas de inspiración biológica. Esta técnica implica un aprendizaje de transferencia dispersa negativa para obtener resultados precisos en la identificación de PPA en imágenes de fondo de ojo. Estas fueron obtenidas de la base de datos SCORM cual cuenta con un total de 1584 imágenes. Se obtuvo una precisión del 90% en esta técnica [22].

Anindita Septiarini (2018), desarrolló una detección automática de APP en imágenes de fondo de ojo de la retina mediante funciones estadísticas haciendo uso de una red neuronal de retro propagación donde utilizó las imágenes provenientes de los conjuntos de datos D1, D2 y RIM-ONE, para establecer una clasificación de imágenes con y sin APP. La precisión, especificidad y sensibilidad media de esta investigación fue de: 95.66%, 86.33% y 97.66% respectivamente [23].

Anindita Septiarini (2018), también planteó la detección de APP en imágenes de fondo de ojo basada en sectores con enfoque de líneas de exploración utilizando un método de base de conocimientos para la clasificación de características obtenidas. De igual forma se trabajó con los conjuntos de datos D1, D2 y RIM-ONE, para la clasificación de imágenes con y sin APP. La precisión, especificidad y sensibilidad media de esta investigación fue de: 90.33%, 69% y 97% respectivamente [24].

Hanxiang Li. (2018), utilizó segmentos de líneas radiales orientados uniformemente para detectar los puntos límite candidatos de DO y DO aumentado (A-DO, que indica la región que combina APP y DO). Luego, para obtener límites suaves y no discretos de A-DO y DO, esos puntos se ajustan a elipses mediante el método de mínimos cuadrados. Finalmente, el límite de PPA se obtiene después de encontrar la parte que no se superpone de estos dos límites. Esta técnica provee de una puntuación F1 de 67.1 en una base de datos de 100 imágenes [25].

Fakhira Zulfira (2019) creó un método de identificación de clases múltiples de APP para la detección de glaucoma en imágenes de fondo de ojo de retina. Donde utilizó una máquina de vectores de soporte para establecer la clasificación: No APP, leve APP, severa APP. Los conjuntos de datos utilizados fueron RIM-ONE y KAGGLE, sus resultados en precisión, especificidad y sensibilidad fueron: 95%, 86%, 89% y 94%, 87%, 99% respectivamente para cada conjunto [26].

Yidong Chai (2019) introdujo un nuevo modelo de DL para la segmentación de área de APP en imágenes de fondo de ojo de la retina, utilizando una red neuronal convolucional para la clasificación de diversos conjuntos en imágenes con o sin APP. Obteniendo una precisión del 89.28% [27].

Fakhira Zulfira (2020), propuso una detección de múltiples clases de glaucoma usando una máquina de vectores de soporte, usando la relación copa disco obtenida a través de una serpiente de contorno activa y segmentación de Otsu para APP. Se desarrolló con base a los conjuntos de imágenes RIM-ONE y KAGGLE. Obtuvo una precisión, especificidad y sensibilidad media de: 95%, 87.5% y 93.5% respectivamente [28].

Ambika Sharma (2020), propuso el uso de técnicas de DL para el diagnóstico de atrofia peripapilar en imágenes retinianas haciendo uso de características estadísticas, utilizando un ResNet-50, aumento de datos y aprendizaje de transferencia. Haciendo uso de diversas bases de datos como: Drishti, Refugee, Rim, Messidor, Drive, Drions y AIIMS. La precisión y sensibilidad obtenida fue: 95.83% y 95.83 respectivamente [6].

Fakhira Zulfira (2021), desarrolló una técnica de segmentación y selección dinámica de conjuntos para mejorar la detección de la gravedad de glaucoma haciendo uso una de serpiente de contorno activa para obtener el valor de OC y OD para medir la CDR, segmentación de Otsu para APP y extracción de características con matriz de coocurrencia de nivel de grises (GCLM) y uso del clasificador de selección de conjunto dinámico (DES) para clasificación de glaucoma. Trabajó con los conjuntos de datos RIM-ONE, KAGGLE y MESSIDOR. Logrando una precisión, especificidad y sensibilidad de: 96%, 88% y 96% respectivamente [29].

Mengxuan Li et al. (2022), propone un bloque de guía de límites junto con una función de pérdida de contorno para mejorar el rendimiento de la segmentación de PPA en los límites. Generando un

mejor desempeño cualitativa y cuantitativamente, dando como puntuación F1 de 80.06% e IoU de 67.29% [30].

Abdullah Almansour et al. (2022), generó un modelo que se desarrolló en función de la localización de la región de interés (ROI) utilizando una máscara de redes neuronales convolucionales basadas en regiones R-CNN y una red de clasificación para la presencia de APP utilizando algoritmos de aprendizaje profundo de CNN. En conjuntos de datos de bases públicas y privadas 2543 imágenes. Obteniendo resultados de área bajo la curva (AUC) de 83%, 89% y 87% para la base de datos privada, publica y combinada respectivamente [31].

Desde la perspectiva de las ciencias computacionales se puede observar en la investigación previa, un enfoque que va dirigido a la integración de algoritmos de aprendizaje profundo para el desarrollo de modelos de clasificación y segmentación de imágenes de fondo de ojo con glaucoma en conjuntos de datos cada vez más grandes y homogéneos que permitan proveer de un modelo que se acerque más a los entornos de prueba reales.

III. MARCO TEÓRICO

3.1 Estructura Anatómica del Ojo

El ojo es un órgano sensorial con una estructura compleja capaz de percibir el mundo físico circundante a través de la recepción de la luz. Lo logra canalizando los rayos de luz a través de la córnea y el cristalino, que concentran los haces de luz en la retina para su posterior procesamiento cerebral. Este mecanismo nos permite discernir patrones, movimientos, colores y contrastes [32].

El fondo del ojo se refiere a la superficie interior del ojo, que incluye la retina, el nervio óptico, los vasos sanguíneos, la mácula y la fovea, y se sitúa frente al cristalino. Como parte integral del sistema nervioso central humano, el nervio óptico presenta una oportunidad única para su visualización y estudio directo. Curiosamente, es a través de la imagen del fondo del ojo que podemos lograr esta observación directa del sistema nervioso. Este acceso sin precedentes ofrece una valiosa vía para identificar posibles biomarcadores en imágenes oculares, específicamente dentro del fondo de ojo [33].

En el contexto más amplio de la investigación en inteligencia artificial, esto abre una nueva frontera para el desarrollo y la implementación de algoritmos de aprendizaje profundo. Dichos algoritmos pueden potencialmente analizar imágenes de fondo de ojo de manera escalable y eficiente, y detectar patrones sutiles que de otro modo podrían escapar al ojo humano[14].

La Figura 4 proporciona una ilustración detallada de la estructura anatómica del ojo humano. Comprender esta estructura es vital para el desarrollo de algoritmos capaces de interpretar imágenes de fondo de ojo con precisión y, posteriormente, la identificación exitosa de biomarcadores.

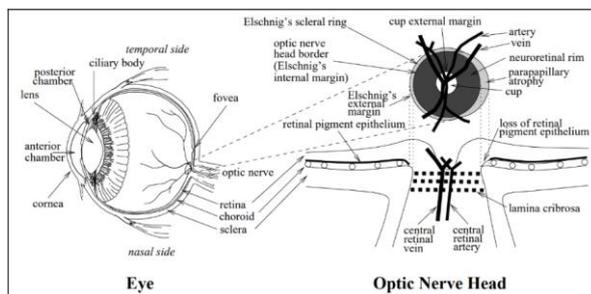


Figura 4. Anatomía de fondo de ojo [33]

3.2 Imagen de fondo de ojo

Sobre la base de los fundamentos de la anatomía del ojo humano, podemos profundizar en las estructuras primarias del fondo del ojo. Estas estructuras incluyen la retina, el disco óptico, la mácula, la fovea y los vasos sanguíneos, cada uno de los cuales desempeña un papel único en nuestra percepción visual. En la Figura 5 se muestra señaladas las principales estructuras morfológicas en la imagen de fondo de ojo.

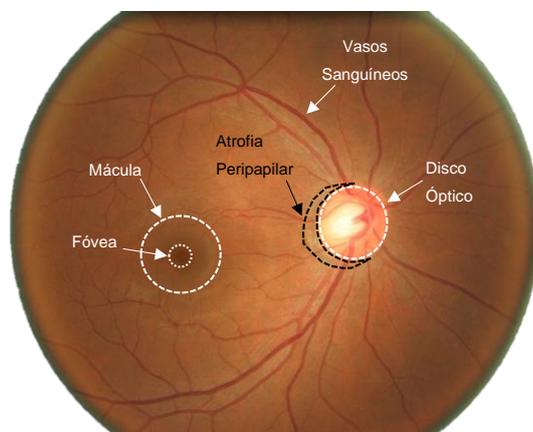


Figura 5. Estructuras principales de fondo de ojo [34]

Durante las últimas dos décadas, los oftalmólogos de todo el mundo han analizado una gran cantidad de imágenes de fondo de ojo utilizando técnicas de procesamiento de imágenes digitales, mejorando significativamente el diagnóstico de diversas enfermedades oculares como el glaucoma, la retinopatía diabética, la neovascularización y la degeneración macular relacionada con la edad [14].

Una de las áreas clave de interés en la obtención de imágenes del fondo de ojo es el disco óptico y los vasos sanguíneos de la retina donde establecen inserción con el nervio óptico. Es un área esencial para examinar, ya que los cambios en esta región pueden significar varias enfermedades oculares, incluido el glaucoma y las neuropatías ópticas[1].

3.3 Segmentación de atrofia peripapilar

Adyacente al disco óptico, encontramos la región de atrofia peripapilar (APP). La APP, caracterizada por cambios en el epitelio pigmentario de la retina y la coroides, se asocia comúnmente con glaucoma y miopía. Por lo general, se identifican dos tipos de APP: la APP de

Comentado [GAF15]: Ocupar espacios previos, no tienen que empezar las secciones en páginas nuevas.

la zona Beta, que es un área irregular hipopigmentada adyacente al disco óptico, y la zona Alfa, que es una región atrófica periférica de hiperpigmentación [35].

El análisis del área peripapilar es importante en la evaluación de diversas enfermedades del nervio óptico. En particular, el estudio de la atrofia peripapilar puede brindar información valiosa sobre la progresión y la gravedad de afecciones como el glaucoma. En consecuencia, el desarrollo de sistemas de IA capaces de analizar con precisión esta región en imágenes de fondo de ojo tiene un potencial sustancial para mejorar las estrategias de tratamiento y detección temprana.

El diagnóstico asistido por computadora (DAC) es una de las estrategias para automatizar la detección del glaucoma. Su objetivo es disminuir y estandarizar la evaluación de los cambios morfológicos y funcionales del ojo asociados al glaucoma. Este enfoque brinda a los profesionales médicos una opinión objetiva con información valiosa que permita establecer una detección temprana de la enfermedad [32].

Los sistemas DAC tienen el potencial de analizar imágenes provenientes de diferentes bases de datos, estableciendo técnicas de preprocesamiento y aprendizaje automático para la extracción y clasificación de características. En la búsqueda de avanzar en el diagnóstico de glaucoma, el enfoque de esta investigación gira en torno a la aplicación de metodologías de aprendizaje profundo para segmentar la APP en las subclases Alfa y Beta en imágenes de fondo de ojo. Este esfuerzo busca mejorar la eficiencia y precisión del análisis APP, enriqueciendo así el poder de diagnóstico de los sistemas DAC para detectar glaucoma y monitorear su progresión.

3.4 Redes Neuronales

En un contexto de las ciencias computacionales, las redes neuronales pueden verse como una forma de imitación digital de la funcionalidad del cerebro humano. Comprenden una colección de unidades de procesamiento, denominadas neuronas artificiales, que están interconectadas a través de pesos sinápticos similares a las conexiones neuronales en el cerebro biológico. La fuerza de estas redes radica en su capacidad para aprender y extraer patrones de los datos, lo que les permite llevar a cabo tareas complejas. Estas tareas van desde la clasificación, donde las redes clasifican los datos en diferentes grupos, hasta la regresión, donde predicen resultados continuos, e incluso la generación, donde crean nuevas instancias de datos en función de los patrones que han aprendido [36], [37].

Una red neuronal se compone de varios elementos clave, como se muestra en la Figura 6. La Capa de entrada recibe los datos sin procesar y los pasa a las capas posteriores para su posterior procesamiento. Las "Capas ocultas", a menudo múltiples, son donde tienen lugar el cálculo y el

aprendizaje, cada una de las cuales extrae progresivamente características de nivel superior de los datos de entrada. Finalmente, la 'Capa de salida' entrega el resultado de los cálculos de la red, como una etiqueta de clase en una tarea de clasificación o un valor numérico en una tarea de regresión[38].

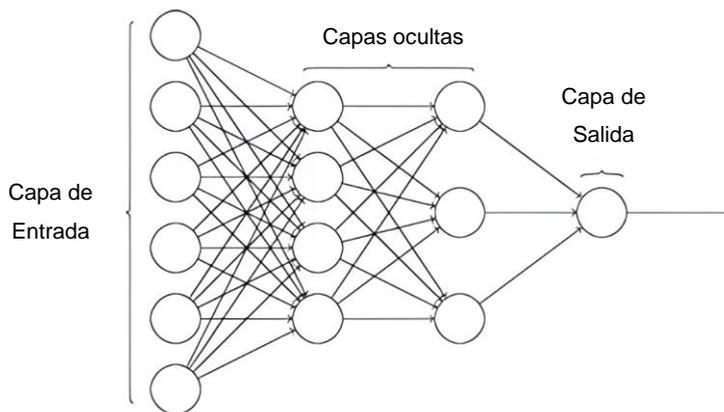


Figura 6. Arquitectura básica de red neuronal [38].

En esencia, las redes neuronales se basan en funciones matemáticas, a menudo no lineales, que les permiten aproximarse a una amplia gama de fenómenos. Cada neurona artificial en la red realiza una suma ponderada de su entrada, una combinación lineal de entradas y pesos, que luego pasa a través de una función de activación para producir una salida.

En términos matemáticos, el cálculo dentro de una neurona se puede expresar, dada una neurona con n entradas, cada entrada x_i tiene un peso asociado w_i . Además, la neurona tiene un término de sesgo (bias) b . La salida y de la neurona se calcula generalmente en dos pasos:

- **Combinación lineal:** Se calcula una suma ponderada de las entradas y el sesgo. Matemáticamente, esto se puede expresar como:

$$z = w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n + b \quad \text{Ec.1}$$

donde z es la suma ponderada de las entradas.

- **Activación:** Luego, la suma ponderada pasa por una función de activación f , que transforma z en la salida y de la neurona. La elección de la función de activación puede

variar, pero ejemplos comunes son la función sigmoide, tangente hiperbólica (tanh), ReLU (Rectified Linear Unit), entre otras. Así:

$$y = f(z) \quad \text{Ec.2}$$

En resumen, la operación dentro de una neurona de una red neuronal se puede expresar como $y = f(w_1 * x_1 + w_2 * x_2 + \dots + w_n * x_n + b)$. Este cálculo se realiza para cada neurona en la red, y los cálculos se propagan desde las capas de entrada hasta la capa de salida.

Es importante notar que los pesos y sesgos son los parámetros que la red aprende durante el entrenamiento. Estos parámetros se ajustan utilizando un algoritmo de optimización (como el descenso de gradiente) para minimizar una función de pérdida, que mide la discrepancia entre las predicciones de la red y los valores verdaderos[38].

3.5 Aprendizaje Profundo

En una extensión lógica del concepto de redes neuronales, se encuentra el aprendizaje profundo, o "deep learning". Este subcampo del aprendizaje automático se basa en redes neuronales con numerosas capas - redes 'profundas' - capaces de aprender características abstractas de alto nivel a partir de datos de entrada sin procesar [39]. El aprendizaje profundo ha demostrado ser particularmente poderoso en dominios como el reconocimiento de imágenes, el procesamiento del lenguaje natural y el reconocimiento de voz, entre otros [40].

El aprendizaje profundo se deriva de las redes neuronales en el sentido de que utiliza la misma arquitectura básica, pero aplica varias capas de estas neuronas para representar los datos. Cada capa de la red aprende una representación de nivel cada vez más alto de los datos. A medida que se avanza a través de las capas de la red, las representaciones se vuelven cada vez más abstractas, permitiendo a la red aprender patrones complejos y características de alto nivel de los datos.

Las redes neuronales profundas consisten en muchas capas de neuronas, cada una de las cuales se conecta a las capas adyacentes a través de conjuntos de pesos sinápticos, al igual que en una red neuronal regular. Aunque los detalles específicos pueden variar dependiendo del tipo de red, la idea central es la misma: las entradas a cada neurona se suman de manera ponderada (como en la ecuación presentada anteriormente), y luego se pasa a través de una función de activación para producir la salida. Al igual que con las redes neuronales más simples, la función de activación puede ser cualquier función matemática, aunque las funciones no lineales son las más comunes [38].

Un tipo específico de red neuronal profunda que merece especial atención son las Redes Neuronales Convolucionales (CNNs). Estas redes son especialmente útiles para el procesamiento de imágenes, ya que son eficientes para manejar datos con una topología de cuadrícula, como los píxeles en una imagen. Las CNNs introducen la operación de convolución en lugar de la multiplicación general de matrices en al menos una de sus capas[41].

La Figura 7 muestra la arquitectura típica de una red neuronal convolucional. Los componentes clave de una CNN son las capas convolucionales, las capas de agrupación (pooling) y las capas completamente conectadas.

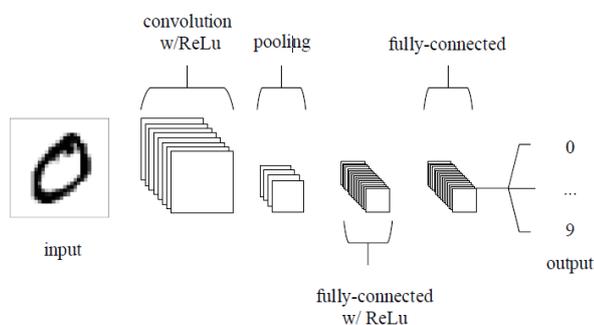


Figura 7. Arquitectura de Red Neuronal Convolutiva [41].

- **Capas convolucionales:** Realizan una operación de convolución en la entrada, pasando un filtro o kernel sobre la imagen de entrada y calculando el producto punto entre los pesos del filtro y la entrada. Esto permite a la red aprender características locales y proporciona la propiedad de invarianza a la traslación.
- **Capas de agrupación (Pooling):** Reducen la dimensión espacial (ancho y alto) de la entrada, lo que ayuda a disminuir la cantidad de parámetros y cálculos en la red, y también controla el sobreajuste.
- **Capas completamente conectadas:** Funcionan de la misma manera que en una red neuronal tradicional, toman las características extraídas de las capas anteriores y realizan la clasificación final.

Comentado [UdW16]: Y realizan.

El entrenamiento de una red neuronal convolucional implica ajustar los pesos de los filtros y las neuronas en la red para minimizar una función de pérdida, similar al entrenamiento de una red neuronal regular. Este proceso se realiza mediante un algoritmo de optimización, como el descenso de gradiente estocástico, y generalmente se requiere de grandes cantidades de datos y tiempo de cálculo.

3.6 Modelos de Segmentación

Las CNN no solo revolucionaron el campo del reconocimiento de imágenes, sino que también proporcionaron la base para el desarrollo de modelos de segmentación semántica y de instancias. La segmentación de imágenes es una tarea esencial en muchas aplicaciones de visión por computadora, que va más allá de la simple clasificación de imágenes para asignar una etiqueta de clase.

3.6.1 Segmentación Semántica

La segmentación semántica, área de gran interés en el campo de la visión artificial, busca asignar una etiqueta de clase a cada píxel de una imagen, delineando así diferentes objetos o regiones de interés con características compartidas. Como se observa en la Figura 8. Denotando en la imagen (b), la máscara de atrofia peripapilar Beta se encuentra representada de color rojo, la clase Alfa verde y el fondo de color negro. De esta forma cada píxel se encuentra clasificado a una de las clases de interés. Esta tarea ha experimentado avances significativos con la llegada de las redes neuronales convolucionales (CNN), principalmente debido a sus sólidas capacidades de extracción de características [42].

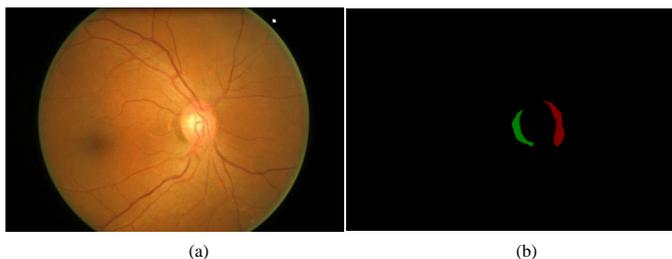


Figura 8. Imagen de Fondo de ojo (a) y máscara de segmentación de atrofia peripapilar (b) Alfa (verde) y Beta(rojo). Las CNN, con su capacidad para reconocer patrones locales, han demostrado ser excepcionalmente beneficiosas en tareas que requieren comprender la información espacial y contextual, como la segmentación semántica. Han allanado el camino para el desarrollo de modelos potentes como redes totalmente convolucionales (FCN), U-Net y SegNet, cada uno de los cuales aporta atributos distintivos y mejoras al dominio.

3.6.1.1 Red Totalmente Convolutiva (FCN)

Largo et al. introdujo un enfoque pionero para la segmentación semántica al proponer la arquitectura de red totalmente convolutiva (FCN). A diferencia de las CNN convencionales que emplean capas completamente conectadas hacia el final de la red, FCN las reemplaza con capas totalmente convolutivas, como se muestra en la Figura 9. Esta modificación crucial permite que la red genere predicciones densas en píxeles, que son esenciales para la tarea de análisis semántico [43].

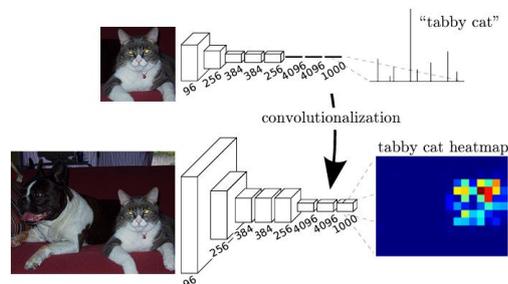


Figura 9. Arquitectura FCN [43].

La arquitectura FCN sentó las bases para los avances posteriores en el campo de la segmentación semántica. La combinación de sus innovaciones arquitectónicas y el poder de las CNN ha dado como resultado una mejora en precisión de segmentación.

3.6.1.2 Segmentation Network (SegNet)

SegNet, una arquitectura novedosa para la segmentación semántica por píxeles se adhiere a la estructura fundamental de las redes totalmente convolutivas (FCN), como se muestra en la Figura 10. Se caracteriza claramente por su diseño de tres partes, que comprende una red codificadora, una red decodificadora y una capa de clasificación por píxel [44].



Comentado [UdW17]: Espacio con la referencia. Revisar todas las demás.

Figura 10. Arquitectura Segnet [44].

La red del codificador refleja el diseño topológico de las trece capas convolucionales de la red VGG16, un modelo bien establecido y de alto rendimiento en el campo del aprendizaje profundo[45]. Esta red codificadora sirve para extraer características diversas e informativas de la imagen de entrada. Después del codificador está la red del decodificador, cuya función principal es restaurar las dimensiones espaciales de los mapas de características de resolución reducida para que coincidan con el tamaño de entrada original. Este paso es crucial para permitir la clasificación por píxeles, que se encuentra en el corazón de la segmentación semántica.

Lo que distingue a SegNet de otros modelos es su enfoque innovador para la ampliación de mapas de características de menor resolución en la red del decodificador. En lugar de depender de parámetros entrenables para el muestreo ascendente, el decodificador de SegNet emplea índices de agrupación calculados durante la operación de agrupación máxima en la etapa del codificador correspondiente. Esta técnica facilita el muestreo ascendente no lineal de los mapas de características, obviando así la necesidad de aprender a realizar el muestreo ascendente.

3.6.1.3 U-net

En 2015, Ronneberger et al. introdujo una arquitectura innovadora para la segmentación semántica, conocida como U-Net, que desde entonces ha ganado una tracción significativa en el campo de la segmentación de imágenes biomédicas. Esta red, basada en los principios de las redes totalmente convolucionales (FCN), es especialmente hábil para manejar los matices de las imágenes biomédicas, marcadas por una alta variabilidad y detalles intrincados [46].

La estructura de U-Net, como se muestra en la Figura 11, presenta una forma de "U" característica, que comprende una ruta de codificación (contracción) y una ruta de decodificación (expansión). La ruta de codificación, también conocida como codificador, es esencialmente una secuencia de capas de convolución y agrupación máxima. A través de este camino, U-Net aprende a capturar el contexto de la imagen mediante la extracción de características de alto nivel mientras reduce las dimensiones espaciales y la cantidad de parámetros, lo que aumenta la eficiencia computacional.

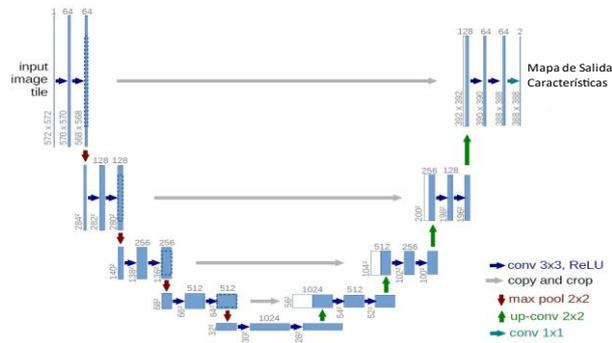


Figura 11. Arquitectura Unet [46].

La ruta de decodificación, o el decodificador, sirve para restaurar la resolución espacial perdida durante la codificación y permitir una localización precisa. Esto se logra a través de convoluciones transpuestas, que son esencialmente operaciones inversas a convoluciones y agrupación máxima, que aumentan la muestra de los mapas de características de baja resolución del codificador. Esta ruta de expansión simétrica aprovecha tanto la información contextual de alto nivel de la ruta de codificación como la información espacial detallada de la ruta de decodificación, lo que da como resultado una salida de segmentación más precisa y detallada [46].

3.6.1.4 Redes Codificadoras

Como se ha señalado, los modelos de segmentación semántica están diseñados en torno a una arquitectura fundamental, que normalmente incorpora una etapa de codificación que es fundamental en el proceso de extracción de características. Esta etapa, también conocida como codificador, es particularmente susceptible de modificaciones y mejoras destinadas a reforzar el rendimiento de los modelos.

En esta investigación, exploramos diferentes marcos de codificación de los modelos de segmentación semántica e introdujimos una serie de modificaciones en la etapa codificadora. Específicamente, integramos diversas redes neuronales convolucionales en las redes codificadoras de estos modelos. Este enfoque permite dotar de flexibilidad y adaptabilidad en la extracción de características, ya que facilita la integración de varios modelos preentrenados, cada uno de los cuales presta sus fortalezas únicas al codificador.

Al incorporar diversas arquitecturas como codificadores en nuestros modelos de segmentación semántica, buscamos aprovechar las fortalezas individuales de cada red. Esta estrategia optimiza

la capacidad de los modelos para aprender y adaptarse a varias tareas de segmentación, mejorando así el rendimiento general de los modelos.

3.6.1.4.1 ResNet

ResNet, abreviatura de Residual Network, es un modelo de CNN desarrollado por Microsoft que obtuvo la primera posición en ImageNet Large Scale Visual Recognition Challenge (ILSVRC) en 2015. ResNet presenta el concepto de "saltar conexiones" o "conexiones de acceso directo", que eluden una o más capas. Este enfoque novedoso mitiga de manera efectiva el problema del gradiente de fuga, un problema destacado que surge con las redes profundas, lo que lleva a la saturación y degradación de la precisión a medida que aumenta la profundidad de la red. La incorporación de estas conexiones de acceso directo permite que el modelo aprenda funciones residuales y admite el entrenamiento de redes más profundas, lo que aumenta su rendimiento [47].

3.6.1.4.2 VGG

VGG, otra red ampliamente adoptada, se caracteriza por su profundidad, lograda mediante el apilamiento de capas convolucionales. Las variantes de VGG, como VGG16 y VGG19, constan de varias capas completamente conectadas, cada una con 4096 canales, seguidas de otra capa completamente conectada con canales correspondientes a cada clase de predicción. La capa final totalmente conectada utiliza una función softmax para propósitos de clasificación. La fortaleza de esta arquitectura radica en su simplicidad y uniformidad, lo que la convierte en una opción sólida para la extracción de características en diversas tareas de visión artificial [45].

3.6.1.4.3 MobileNet

MobileNet adopta un enfoque diferente y ofrece una solución liviana y eficiente ideal para aplicaciones de visión integradas y móviles. La arquitectura MobileNet aprovecha las circunvoluciones separables en profundidad, que reducen significativamente el tamaño y la complejidad del modelo sin comprometer su rendimiento. Este diseño convierte a MobileNet en una excelente opción para aplicaciones que requieren procesamiento en tiempo real en dispositivos con recursos computacionales limitados [48].

3.6.2 Segmentación de Instancias

Mientras que la segmentación semántica tiene como objetivo clasificar cada píxel de una imagen en su clase correspondiente, la segmentación de instancias lleva esta tarea un paso más allá. La principal diferencia entre ambos modelos de segmentación es que la segmentación de instancias radica en un mayor nivel de detalle de la salida. Si bien ambos métodos tienen como objetivo

asignar una etiqueta de clase a cada píxel de una imagen, la segmentación de instancias va un paso más allá al diferenciar entre objetos separados de la misma clase [49].

En la Figura 11(a), se ilustra un ejemplo de segmentación semántica. En este caso, el modelo identifica correctamente todos los píxeles que pertenecen a las clases Alfa y Beta. Sin embargo, una limitación inherente de la segmentación semántica es que considera las diferentes atrofas identificadas dentro de una misma clase como una única entidad. En contraste, en la Figura 11(b), se muestra la segmentación de instancias donde cada atrofia de la clase Beta es identificada como una entidad separada. Esta particularidad de la segmentación de instancias la convierte en una herramienta esencial en ámbitos como la medicina, donde el reconocimiento diferenciado de objetos individuales resulta ser crucial.

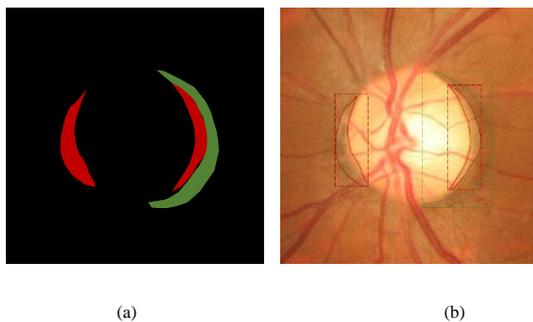


Figura 11. Comparación de modelos de Segmentación semántica (a) y de Instancias (b).

Como resultado, los modelos de segmentación de instancias tienden a ser más complejos y computacionalmente intensivos que sus contrapartes de segmentación semántica. Sin embargo, el detalle adicional proporcionado por la segmentación de instancias puede ser invaluable en aplicaciones donde el reconocimiento de objetos individuales es esencial.

3.6.2.1 Detección de Objetos

El aprendizaje profundo, un subcampo del aprendizaje automático, ha impulsado avances significativos en el dominio de la clasificación y detección de imágenes. Un elemento central de este progreso son los algoritmos de detección de objetos, que han demostrado velocidad y eficacia. El poder de estos algoritmos radica en su capacidad para realizar dos tareas al mismo tiempo: identificar diversas clases dentro de una imagen y ubicar sus posiciones si están presentes [50].

Los algoritmos iniciales empleaban un enfoque de ventana deslizante para el análisis de imágenes, aplicando CNN en diferentes regiones de una imagen. Este proceso fue iterativo, involucrando el escaneo sistemático de la imagen para identificar objetos de interés, seguido por la clasificación de estos objetos. Si bien fue efectivo, este enfoque fue computacionalmente intensivo, dada la necesidad de aplicar CNN a múltiples subregiones de imágenes [50].

Para mejorar la eficiencia, se introdujeron las redes neuronales convolucionales basadas en regiones (R-CNN). Este modelo proponía el uso de la búsqueda selectiva para identificar un número manejable de propuestas de regiones de objetos de cuadro delimitador. Luego, estas regiones fueron analizadas por una CNN, lo que redujo la carga computacional al enfocarse solo en las áreas de interés probables [51].

Fast R-CNN, una evolución del R-CNN original, mejoró aún más la eficiencia mediante la introducción de una técnica conocida como agrupación de regiones de interés (RoI), que permite que la red acepte imágenes de diferentes tamaños. La combinación de ROI simplificó el proceso al aplicar la CNN a toda la imagen solo una vez, en lugar de iterativamente para cada región propuesta [52].

Faster-R-CNN cambia la búsqueda selectiva de regiones, introduciendo una red de propuesta de región (RPN) que comparte características convolucionales de imagen completa con la red de detección, permitiendo usar este algoritmo en aplicaciones de tiempo real [53].

3.6.2.2 Mask RCNN

Mask R-CNN es una red que amplía a Faster R-CNN agregando una rama de predicción de máscaras que funciona en paralelo con la rama de reconocimiento de cuadro delimitador. Permite segmentar las clases de interés dentro de las imágenes [54].

La estrategia empleada por este algoritmo se puede representar en tres etapas como se muestra en la Figura 12, un Backbone, una red de propuestas regionales (RPN) y dos subredes paralelas para la detección de objetos y la predicción de máscaras, respectivamente.

La primera etapa es un Backbone que suele ser una red neuronal convolucional (CNN) previamente entrenada, como ResNet50 o ResNet101[47], que se utiliza para extraer características de la imagen de entrada. Luego, el RPN se usa para generar un conjunto de propuestas para la detección de objetos, que luego son refinadas por la subred de detección de objetos para obtener predicciones precisas del cuadro delimitador. Al mismo tiempo, la subred de máscaras procesa cada propuesta y genera una máscara binaria para cada objeto detectado.

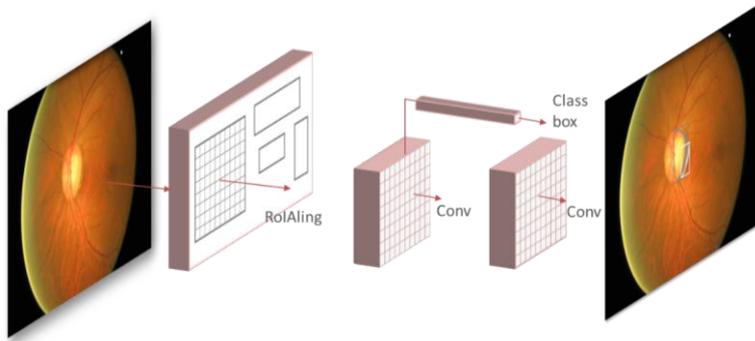


Figura 12. Arquitectura de Mask RCNN.

3.6.2.3 Cascade Mask RCNN

La exploración de la segmentación de instancias ha dado lugar a una serie de modelos sofisticados, cada uno iterando sobre el último para mejorar el rendimiento y la precisión. Sobre la base sólida del modelo Mask R-CNN, el modelo Cascade Mask R-CNN introduce modificaciones destinadas a mejorar la predicción de máscaras, ampliando así los límites de lo que se puede lograr en la segmentación de instancias.

Mask R-CNN, una extensión del modelo de detección de objetos Faster R-CNN, incorpora una rama adicional para predecir máscaras binarias, lo que permite la segmentación a nivel de píxeles. Cascade Mask R-CNN, a su vez, amplía la arquitectura de Mask R-CNN mediante la introducción de varias etapas en cascada, cada una de las cuales alberga una subred independiente para la predicción de máscaras.

Como se muestra en la Figura 13, cada subred recibe el mismo conjunto de Regiones de interés (RoI) como entrada, generada por la Red de propuesta de región (RPN). La característica distintiva del modelo Cascade Mask R-CNN radica en el mapa de características adicional introducido en cada etapa. Este mapa de características se obtiene recortando y redimensionando las máscaras previstas en la etapa anterior [54].

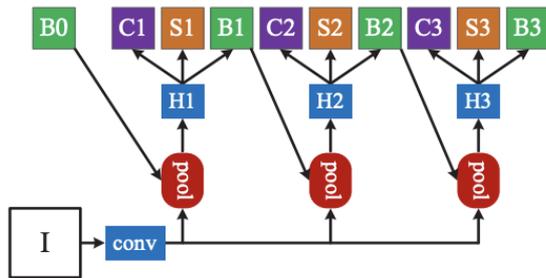


Figura 13. Arquitectura de Cascade Mask RCNN [54].

La subred luego utiliza esta información adicional para refinar las máscaras, generando así un nuevo conjunto de predicciones de máscaras. Este proceso recursivo de refinar máscaras a través de etapas sucesivas puede verse como una forma de optimización iterativa, destinada a mejorar progresivamente la precisión de la segmentación de instancias.

Al aprovechar la información espacial y contextual conservada en las máscaras refinadas de etapas anteriores, Cascade Mask R-CNN puede generar predicciones de máscara más precisas. Esto contribuye a mejorar el rendimiento en tareas de segmentación de instancias desafiantes, particularmente aquellas que involucran escenas complejas con múltiples objetos superpuestos.

3.6.2.4 Mask Scoring RCNN

El modelo Mask Scoring R-CNN se destaca como un refinamiento notable de la arquitectura Mask R-CNN, diseñado específicamente para mejorar la precisión de la segmentación de instancias. En el modelo tradicional Mask R-CNN, la segmentación de instancias se logra generando máscaras binarias para cada objeto detectado en una imagen. Si bien este enfoque proporciona resultados significativos, no evalúa la calidad de estas máscaras, un factor que podría mejorar potencialmente el resultado final de la segmentación.

Al abordar esta limitación, Mask Scoring RCNN presenta un nuevo módulo de puntuación de máscaras. En la Figura 14 se visualiza como el módulo inferior derecho. Este módulo innovador utiliza las características extraídas de la última capa convolucional de la subred de máscara, incorporando así información más compleja sobre la imagen en el proceso de segmentación [55].

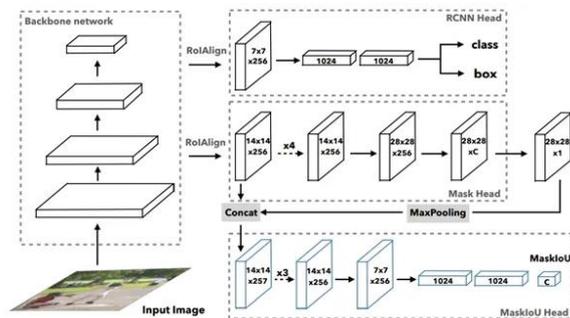


Figura 14. Arquitectura de Mask Scoring RCNN [55].

El módulo de puntuación de máscara tiene la tarea de generar una puntuación de calidad de máscara para cada objeto detectado. Esta puntuación es una evaluación cuantitativa de la adecuación de la máscara al objeto, considerando factores como la alineación entre la máscara y los límites del objeto, y la integridad de la representación del objeto dentro de la máscara.

Una vez que se generan estos puntajes de calidad de máscara, se utilizan para modular el resultado final del modelo. Específicamente, cada predicción de máscara original se multiplica por su puntaje de calidad de máscara correspondiente. Esta operación pesa efectivamente las predicciones de la máscara, acentuando las de mayor calidad y disminuyendo la influencia de las predicciones de menor calidad en el resultado final.

Al integrar este nivel adicional de refinamiento, Mask Scoring R-CNN mejora la precisión de la segmentación de instancias. No solo identifica objetos y genera máscaras respectivas, sino que también evalúa y ajusta estas máscaras en función de su calidad. Este enfoque sofisticado subraya el potencial de modificaciones y mejoras adicionales en los modelos de segmentación de instancias, lo que promete resultados aún más precisos en aplicaciones [55].

3.6.3 Métricas de Evaluación

El problema de la segmentación de imágenes puede ser visto como un problema de clasificación a nivel píxel y particularmente en el contexto de problemas de clasificación binaria, las medidas de desempeño son esenciales para evaluar la efectividad y precisión de los modelos. Cuatro términos fundamentales centrales para estas medidas son: Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN). Cada una de estas métricas ofrece información única sobre el rendimiento del modelo y se utilizan colectivamente para

evaluar la eficacia general. En la Figura 15 se encuentra representado un esquema donde se pueden identificar cada uno de los términos vistos en un problema de segmentación.

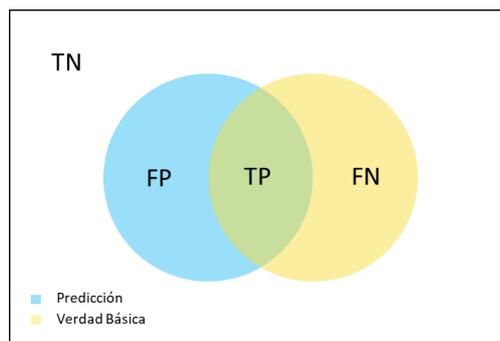


Figura 15. Esquema de evaluación de máscaras de segmentación.

Verdaderos positivos (TP): Los verdaderos positivos se refieren a las instancias en las que el modelo identifica correctamente una clase positiva. En el contexto de la segmentación de imágenes, un TP es una instancia en la que el modelo segmenta con precisión la región de interés. Por ejemplo, en la Figura 15 cada píxel correctamente identificado como perteneciente a la clase de interés, se muestra como en la intersección de ambas máscaras.

Falsos positivos (FP): Los falsos positivos representan las instancias en las que el modelo identifica incorrectamente una instancia negativa como positiva. En la segmentación de imágenes, esto significaría que el modelo identifica erróneamente una región como de interés. En nuestro ejemplo esta sección se encuentra representado por la región de color azul que no coincide con el ground truth.

Verdaderos negativos (TN): Los verdaderos negativos son las instancias en las que el modelo identifica correctamente una clase negativa. En términos de segmentación de imágenes, un TN es un píxel que el modelo identifica con precisión como no perteneciente a la región de interés. Denotando de color blanco esta connotación en la Figura 15.

Falsos negativos (FN): Los falsos negativos se refieren a las instancias en las que el modelo identifica incorrectamente una instancia positiva como negativa. En el contexto de la segmentación de imágenes, esto significaría que el modelo no logra identificar una región que sea de interés. Continuando con nuestro ejemplo, si el modelo no logra clasificar un píxel de la región de interés perteneciente a la verdad básica, sería un falso negativo.

Comentado [UdW18]: La?

Comentado [UdW19]: Revisar concordancia.

Comentado [JA20R19]: No entiendo a que e refiere

Mientras que TP y TN muestran las predicciones correctas del modelo, FP y FN brindan información sobre los tipos de errores que comete el modelo. Juntos, se utilizan para calcular métricas de rendimiento como la precisión, la sensibilidad, intersección sobre la unión y la puntuación F1.

3.6.3.1 Precisión

La precisión es una métrica que cuantifica la capacidad de un modelo para identificar con precisión instancias positivas. En el contexto de los modelos de segmentación, la precisión indica la proporción de píxeles positivos identificados correctamente (positivos verdaderos) con respecto al número total de píxeles que el modelo identificó como positivos (positivos verdaderos y positivos falsos) [56].

Matemáticamente, la precisión se puede calcular usando la siguiente fórmula:

$$\text{Precisión} = \frac{TP}{TP + FP} \quad \text{Ec.3}$$

3.6.3.2 Sensibilidad (Recall)

La recuperación (también conocida como sensibilidad o tasa de verdaderos positivos) mide la capacidad de un modelo para encontrar todas las instancias positivas. En la segmentación, es la proporción de píxeles positivos reales que el modelo identifica correctamente [56].

La fórmula para calcular Recall es:

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{Ec.4}$$

Los valores de Precisión y Recall varían de 0 a 1, y los valores más altos indican un mejor rendimiento. Sin embargo, a menudo hay una compensación entre estas dos métricas, ya que mejorar una puede hacer que la otra disminuya

3.6.3.3 Puntuación F1 (F1 Score)

La puntuación F1 es la media armónica de Precisión y Recall. Busca equilibrar la compensación entre Precisión y Recuperación, proporcionando una puntuación única que representa ambas métricas. Esto es especialmente importante en los casos en que una métrica puede ser alta a expensas de la otra [56].

La puntuación F1 se calcula de la siguiente manera:

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \quad \text{Ec.5}$$

La puntuación F1 es un valor entre 0 y 1, cuyo valor más cercano a 1 expresa la mayor similitud entre las muestras. En otro sentido la puntuación F1 es alta si tanto la precisión como la recuperación son altas, lo que la convierte en una medida más robusta que simplemente confiar en la Precisión o Recall.

3.6.3.4 Intersección sobre Unión (IoU)

La intersección sobre la unión es una métrica de evaluación que se utiliza para medir la superposición entre la predicción y la realidad del terreno. Calcula el área de superposición entre los dos segmentos dividida por el área de unión de los dos segmentos [56].

Se define como:

$$IoU = \frac{TP}{TP + FP + FN} \quad \text{Ec.6}$$

En esta ecuación, el numerador representa la intersección de la verdad fundamental y la predicción (es decir, los píxeles del objeto correctamente identificados), y el denominador representa la unión de la verdad fundamental y la predicción (es decir, todos los píxeles identificados como parte del objeto), ya sea en la verdad básica, la predicción o ambas). Por lo tanto, el IoU es una medida de la superposición entre la realidad básica y la predicción, en relación con su tamaño combinado.

3.6.3.4 Average Precision (AP)

AP es una métrica basada en el área bajo la curva Pr x Rc que ha sido preprocesada para eliminar un comportamiento Zig Zag. Para su cálculo se establecen los siguientes pasos:

- Se ordenan k diferentes valores de confianza $\tau(k)$ por el detector de objetos.

$$\tau(k) = k = 1, 2, \dots, k \quad \text{tal que } \tau(i) > \tau(j) \quad \text{para } i > j \quad \text{Ec.7}$$

- Se define un conjunto ordenado de valores de Recall $R_r(n)$.

$$R_r(n), \quad n = 1, 2, \dots, N \quad \text{tal que } R_r(m) > R_r(n) \quad \text{para } m > n \quad \text{Ec.8}$$

AP, se calcula utilizando los dos conjunto de datos previos, pero antes es necesario de (Pr x Rc) tienen que ser interpolados de tal manera que el resultado de (Pr x Rc) sea una curva monótona. El resultado de la interpolación de la curva es definida por una función continua, donde R es un valor real contenido en $[0,1]$, $Rc(\tau(k))$ es el valor de recall para la confianza dada en $\tau(k)$.

$$Pr_{interp}(R) = \max\{\Pr(\tau(k))\}, k | Rc(\tau(k)) \geq R \quad \text{Ec.9}$$

El valor de precisión interpolado en Recall R corresponde al valor máximo de precisión $Pr_{interp}(k)$ cuyo valor de Recall correspondiente es mayor o igual a R.

$$AP = \sum_{k=0}^k (R_r(k) - R_r(k+1)) \Pr_{interp}(R_r(k)) \quad \text{Ec. 10}$$

3.6.3.4.1 AP@0.5 y AP@0.75

En este método se realiza la interpolación de N=101 Recall puntos dado en la ecuación:

$$AP = \frac{1}{N} \sum_{n=1}^N \Pr_{interp}(R_r(n)) \quad \text{Ec. 11}$$

Los resultados se obtienen para cada clase y después son promediados, su diferencia radica en el valor de confianza brindado en IoU.

3.6.3.4.2 AP@[.5:.05:.95]

La métrica expande al método anterior empleando 10 umbrales ($t = [0.5, 0.55, \dots, 0.95]$) y tomando el promedio de todos los resultados.

3.6.3.4.3 APs, APm y API

Estas métricas son conocidas como AP across scales, aplican AP@[.5:.05:.95] teniendo en consideración el área del objeto de verdad fundamental.

- APs: (área < 32² pixeles).
- APm: (32² < área < 96² pixeles).
- API: (área > 96² pixeles).

El aporte científico del trabajo propuesto radica en la aplicación de algoritmos de segmentación para establecer una delimitación detallada de la APP Alfa y Beta, brindando un enfoque con mayor grado de granularidad que brinde de información de valor al seguimiento y diagnóstico de glaucoma.

IV. MATERIALES Y METODOLOGÍA

La eficacia de cualquier esfuerzo de investigación en el campo de la inteligencia artificial depende de las herramientas y metodologías adoptadas durante la investigación. Este capítulo proporciona

una descripción detallada de los materiales y métodos aprovechados en este estudio, centrándose en las herramientas computacionales y las bibliotecas utilizadas para ejecutar las diversas tareas, incluido el desarrollo de modelos, la capacitación, las pruebas y la evaluación.

4.1 Google Colaboratoy

Google Colaboratory, o Google Colab, es un entorno de programación de Python basado en la nube que ofrece una plataforma interactiva para el aprendizaje automático y el análisis de datos. Está construido sobre Jupyter Notebook y ofrece acceso gratuito a recursos informáticos sólidos, incluidas Central Processing Unit(CPU), Graphics Processing Unit (GPU) y Tensor Processing Unit (TPU), que son esenciales para los cálculos complejos necesarios en las tareas de aprendizaje profundo.

La utilización de Google Colab Pro+ en este estudio anuló la necesidad de una infraestructura de hardware de alto rendimiento, lo que permitió la ejecución de tareas computacionalmente intensivas, como la inferencia y el entrenamiento de modelos. Lo que la convirtió en una herramienta fundamental en esta investigación [57].

4.2 Python

Python es un lenguaje de programación interpretado, de alto nivel y de propósito general que ha tenido una amplia adopción en la comunidad científica y, en particular, en el campo de la investigación de inteligencia artificial. Creado por Guido van Rossum y lanzado por primera vez en 1991. Cuenta con un conjunto integral de bibliotecas para el aprendizaje automático, la manipulación de datos y la visualización de datos, lo que lo convierte en el lenguaje elegido para esta investigación. Su sintaxis fácil de entender y su uso generalizado en la comunidad científica han facilitado el desarrollo y la modificación de nuestros modelos [58].

Se utilizó Python en su versión 3.10.

4.3 MMDetection

MMDetection es una caja de herramientas integral, de alta calidad y de código abierto para tareas de detección de objetos, segmentación de instancias y segmentación semántica. Desarrollado por el Laboratorio Multimedia de la Universidad China de Hong Kong, fue diseñado para facilitar el diseño, entrenamiento y validación de varios modelos de aprendizaje profundo para visión artificial [59]na de las características que definen a MMDetection es su diseño modular. El marco segrega los diferentes componentes de un sistema de detección de objetos en módulos individuales, incluidas redes troncales, cuellos, cabezas densas y funciones de pérdida. Esto

Comentado [UdW21]: Deberías definir las siglas si son primera aparición.

permite a los investigadores ensamblar estas piezas de forma plug-and-play, lo que facilita el diseño y la experimentación de arquitecturas novedosas.

MMDetection es compatible con una amplia gama de modelos de segmentación y detección de objetos, incluidos, entre otros, detectores de dos etapas como Faster R-CNN, Mask R-CNN, Cascade R-CNN. Además, proporciona implementaciones de modelos de última generación propuestos en trabajos de investigación recientes, lo que permite a los investigadores utilizar estos modelos sin implementarlos desde cero.

Otra ventaja clave de MMDetection es su alta eficiencia. Utiliza el marco de aprendizaje profundo PyTorch para sus cálculos, beneficiándose del gráfico de cálculo dinámico y la interfaz intuitiva de PyTorch. Además, MMDetection proporciona compatibilidad con múltiples GPU, lo que permite el entrenamiento de modelos a gran escala en conjuntos de datos de tamaño considerable.

MMDetection también incluye funcionalidades para diversas técnicas de preprocesamiento de datos y métricas de evaluación, lo que lo convierte en un completo conjunto de herramientas para tareas de segmentación y detección de objetos. Admite conjuntos de datos populares como COCO, Pascal VOC y Cityscapes, y proporciona protocolos de evaluación estandarizados, lo que facilita la evaluación comparativa y la comparación de diferentes modelos.

4.4 PyTorch

PyTorch es una plataforma de aprendizaje profundo de código abierto que proporciona un camino fluido desde la creación de prototipos de investigación hasta la implementación de producción. Desarrollado por el grupo de Investigación de Inteligencia Artificial (FAIR) de Facebook, PyTorch ofrece un amplio conjunto de funciones que lo convierten en una opción popular entre los investigadores y desarrolladores en el campo de la inteligencia artificial y el aprendizaje automático.

Una de las características clave de PyTorch es su gráfico computacional dinámico (también conocido como estrategia de definición por ejecución), que contrasta con el gráfico de cálculo estático (estrategia de definición y ejecución) utilizado por algunas otras bibliotecas de aprendizaje profundo. En un gráfico computacional dinámico, la estructura del gráfico se define sobre la marcha durante la ejecución. Esto proporciona más flexibilidad y facilita el flujo de control dinámico complejo. Esto es particularmente ventajoso para los modelos que involucran un flujo de control dinámico no uniforme, como las redes neuronales recurrentes (RNN) y las redes neuronales recursivas[60][5.

Se trabajó con PyTorch en su versión 1.13.1.

4.5 Keras

Keras es una biblioteca de aprendizaje profundo de alto nivel y código abierto diseñada para agilizar el proceso de creación, capacitación y evaluación de redes neuronales. Presentado en 2015 por François Chollet, investigador de inteligencia artificial e ingeniero de software de Google, Keras se desarrolló con el objetivo de facilitar la experimentación rápida y ofrecer una interfaz fácil de usar. Desde entonces, ha obtenido una adopción generalizada en las comunidades de investigación de aprendizaje automático e inteligencia artificial, así como en aplicaciones industriales[61].

La biblioteca Keras simplifica el proceso de desarrollo al abstraer conceptos subyacentes complejos y proporcionar bloques de construcción modulares y componibles para crear redes neuronales. Ofrece una amplia gama de capas predefinidas, funciones de activación, optimizadores y otras utilidades, lo que permite a los investigadores y profesionales crear rápidamente prototipos y experimentar con diversas arquitecturas de redes neuronales. Además, Keras admite varios modelos preentrenados para tareas como clasificación de imágenes, segmentación y procesamiento de lenguaje natural, que se pueden ajustar para casos de uso específicos.

Keras version 2.4.3

4.6 Roboflow

Roboflow es una plataforma de software innovadora que permite la gestión de conjuntos de datos hasta la implementación de modelos. Reconociendo la naturaleza a menudo compleja y lenta del desarrollo y la gestión de aplicaciones de visión por computadora, Roboflow se estableció con el objetivo de simplificar y acelerar este proceso [62].

Roboflow proporciona un conjunto integrado de herramientas que ayudan a los usuarios a manejar varias tareas asociadas con el desarrollo de modelos de visión por computadora. Estas tareas incluyen, entre otras, anotación de imágenes, control de versiones de conjuntos de datos, preprocesamiento y aumento, entrenamiento de modelos y, en última instancia, implementación. Su interfaz fácil de usar y su amplia gama de funcionalidades lo convierten en una herramienta

ideal tanto para principiantes como para investigadores experimentados en el campo de la visión artificial.

Una característica clave de Roboflow es su capacidad para administrar y versionar conjuntos de datos. Esto incluye capacidades para importar datos de una amplia variedad de fuentes, anotar imágenes con cuadros delimitadores o máscaras de segmentación y crear versiones de conjuntos de datos para realizar un seguimiento de los cambios. Roboflow también es compatible con una variedad de técnicas de preprocesamiento y aumento, que pueden mejorar la solidez de los modelos y permitirles generalizar mejor los datos ocultos.

4.7 Base de Datos

4.7.1 ORIGA

Este estudio utiliza un repositorio en línea conocido como ORIGA-light, que fue diseñado para difundir al público imágenes clínicas reales de la retina. Este repositorio promueve el acceso abierto para los investigadores, ofreciendo una plataforma para validar y comparar sus algoritmos de segmentación asistida por computadora. ORIGA-light se construyó con la ayuda de una herramienta interna de segmentación y clasificación de imágenes, lo que facilitó enormemente su construcción.

ORIGA-light comprende 650 imágenes de la retina de resolución 3072 x 2048 en formato JPG, que han sido clasificadas por profesionales capacitados del Instituto de Investigación del Ojo de Singapur. Donde 482 imágenes se encuentran etiquetadas como Ojos Sanos y 168 como Ojos con Glaucoma [34].

El uso de una base de datos de este tipo es fundamental para aprovechar el conocimiento clínico integrado en las imágenes del fondo de retina. Además, el proceso sistemático de evaluación comparativa descrito garantiza el desarrollo y la validación continuos de algoritmos de segmentación asistidos por computadora de alto rendimiento, con el objetivo final de mejorar la precisión diagnóstica en enfermedades oculares. Este trabajo subraya el potencial de integrar técnicas de aprendizaje profundo y bases de datos de imágenes a gran escala para avanzar en el campo de la investigación de imágenes oftálmicas.

4.7.2 Retina

El conjunto de datos Retina, obtenido de un repositorio de GitHub disponible públicamente, constituye un componente importante del material utilizado en esta investigación. Este conjunto de datos se compone de 601 imágenes retinianas de alta resolución (2052 x 1728 píxeles) en

formato JPG, que han sido clasificadas por expertos en cuatro categorías: sin lesión (NL), catarata, glaucoma y lesión retiniana.

La composición del conjunto de datos está equilibrada para garantizar una investigación exhaustiva a través de diferentes condiciones oculares. Hay 300 imágenes que representan la categoría Sin lesión, sirviendo, así como muestras de control contra las cuales se comparan las clases patológicas. Las 301 imágenes restantes se distribuyen uniformemente en las tres condiciones patológicas: catarata, glaucoma y lesión retiniana, con 100, 101 y 100 imágenes respectivamente [63].

Esta composición del conjunto de datos permite la investigación de detección de biomarcadores de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo para cada una de estas condiciones. Además, el uso de imágenes de alta resolución en este estudio facilita la extracción de características detalladas necesarias para la detección precisa de biomarcadores, lo que contribuye significativamente a la solidez de los modelos desarrollados.

4.7.3 DRISHTI-GS1

El conjunto de datos Drishti-GS1 cuenta con 101 imágenes de resolución 2048 x 1750 en formato PNG. Las imágenes incorporadas en el conjunto de datos Drishti-GS fueron recopiladas y anotadas meticulosamente por el renombrado Aravind Eye Hospital, ubicado en Madurai, India[64].

En particular, el conjunto de datos presenta una característica única en el sentido de que se origina en una sola población: todos los sujetos que contribuyen a las imágenes oculares en este conjunto de datos son personas indias diagnosticadas con glaucoma. Esta homogeneidad en el conjunto de datos, aunque restrictiva en términos de diversidad, brinda un enfoque concentrado en la detección y caracterización del glaucoma dentro de este grupo demográfico específico.

4.8 Hardware

En esta investigación, se empleó una configuración computacional robusta para el entrenamiento de algoritmos de aprendizaje profundo, que incluyen un amplio procesamiento de imágenes y detección de biomarcadores en imágenes de fondo de ojo. La configuración de hardware utilizada en este estudio está diseñada para manejar altas demandas computacionales de manera efectiva.

- **Colab:** El servidor en la configuración más avanzada permite el uso de una GPU NVIDIA A100 con 49 GB de VRAM. Esta GPU está diseñada específicamente para centros de datos e investigación de IA, y ofrece una velocidad y una potencia computacional

notables. La gran VRAM permite el procesamiento de grandes lotes de datos simultáneamente, lo que la convierte en un excelente recurso para entrenar y validar modelos complejos de aprendizaje profundo.

- **PC-1:** La estación está equipada con una CPU AMD Ryzen 7 5800HS y una GPU RTX 3050 4 GB de VRAM y 24 GB de RAM, sirve como recurso computacional primario.
- **PC-2:** La estación está equipada con una CPU AMD Ryzen 7 5800X y una GPU RTX 3080 10 GB de VRAM y 32 GB de RAM, sirve como recurso computacional secundario.

4.9 Metodología

La Figura 16 proporciona una representación visual de la metodología robusta y sistemática empleada en este estudio de investigación. Esta metodología, que sirve como columna vertebral de la investigación, consta de nueve fases. Cada una de estas etapas contribuye al objetivo general de esta investigación, facilitando una comprensión integral del problema de investigación y permitiendo el desarrollo de soluciones efectivas.

En términos generales, estas fases abarcan colectivamente la gama de actividades realizadas a lo largo de este estudio, desde la investigación inicial de las prácticas actuales hasta el análisis final y las conclusiones. Su ejecución secuencial asegura una progresión lógica de la investigación, que finalmente conduce al cumplimiento de los objetivos de la tesis.

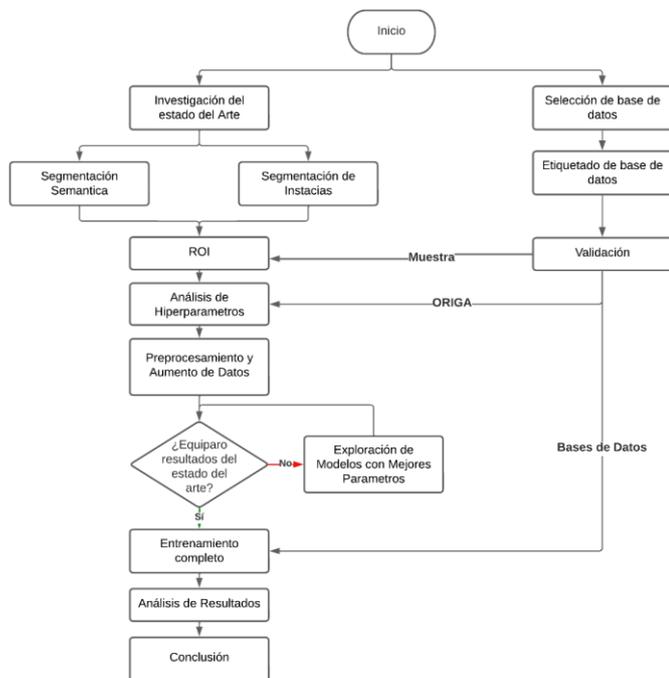


Figura 16. Diagrama de flujo para desarrollo de modelo de segmentación y clasificación de APP.

Investigación del estado del arte: La investigación comienza con una revisión exhaustiva de las metodologías utilizadas en segmentación de imágenes. Esta etapa fundamental requiere una investigación meticulosa de los modelos de segmentación tanto semánticos como de instancias, con especial énfasis en sus fortalezas, debilidades y posibles aplicaciones en el campo del análisis de imágenes de fondo de ojo. Esta investigación servirá como base del desarrollo del trabajo de investigación, proporcionando una comprensión integral de las prácticas y desarrollos actuales en el campo.

Búsqueda en bases de datos públicas: Al mismo tiempo, se lleva a cabo una búsqueda sistemática de bases de datos de imágenes de fondo de ojo disponibles públicamente. Esto implica una búsqueda cuidadosa de varias fuentes para identificar bases de datos que ofrezcan imágenes representativas, diversas y de alta calidad. El objetivo es garantizar un conjunto de datos completo y fiable para las etapas posteriores de la investigación que cumplan con los objetivos establecidos.

Etiquetado de bases de datos: Una vez que se identifican las bases de datos, se lleva a cabo un riguroso proceso de etiquetado. Este es un paso crucial para proporcionar etiquetas reales para las tareas de segmentación. Implica la delimitación meticulosa de las áreas de interés, específicamente la atrofia peripapilar Alfa y Beta, en las imágenes del fondo de ojo. Este conjunto de datos etiquetados será empleado para las etapas de entrenamiento, validación y prueba de los modelos de segmentación.

Selección de modelos: La investigación luego procede a la etapa de selección del modelo. Aquí, se realiza un análisis comparativo entre los modelos de segmentación semántica y de instancia utilizando un subconjunto representativo de la base de datos etiquetada y previamente validada. Esta comparación tiene como objetivo determinar el enfoque de segmentación más adecuado para la investigación, considerando factores como el rendimiento del modelo, la eficiencia computacional y la idoneidad para la tarea en cuestión.

Evaluación de técnicas de preprocesamiento: Después de la selección del enfoque de segmentación, se procede a la exploración de diversas técnicas de preprocesamiento. Estas técnicas tienen como objetivo mejorar el rendimiento del modelo de segmentación generando un impacto positivo en la calidad y la interpretabilidad de las imágenes de entrada. Esto podría implicar técnicas de redimensionamiento, mejora de contraste, extracción de las regiones de interés, entre otras.

Evaluación de Hiperparámetros: Como cualquier modelo de inteligencia artificial, la exploración de los hiperparámetros es un punto crucial para incrementar el desempeño de los modelos para la realización de las tareas específicas de investigación. La propuesta de hiperparámetros van en relación con el análisis de puntuaciones obtenidas en las métricas, así como un análisis de errores y observaciones cualitativas de las máscaras de predicción de los modelos.

Validación de la Base de Datos: Simultáneamente con la evaluación del preprocesamiento, la investigación continúa con la validación de las bases de datos seleccionadas. Este proceso asegura la calidad y confiabilidad de los datos utilizados en la investigación, lo cual es fundamental para la credibilidad de los resultados de la investigación.

Comentado [GAF22]: Tilde

Aumento de datos y exploración de modelos: Con la selección de la técnica de preprocesamiento óptima, la investigación continúa con la implementación de estrategias de aumento de datos. Estas estrategias aumentan la diversidad y el volumen de los datos de entrenamiento, mejorando así la capacidad de generalización del modelo. Además, se exploran varios modelos que se alinean con el enfoque de segmentación seleccionado, junto a una exploración de hiperparámetros con el objetivo de identificar candidatos potenciales para una mayor experimentación.

Entrenamiento Top: Los modelos de mayor rendimiento, identificados en la etapa anterior, se someten luego a un entrenamiento riguroso utilizando múltiples bases de datos. Este paso es fundamental para evaluar la solidez y adaptabilidad de los modelos, ya que los expone a una amplia variedad de datos y condiciones.

Análisis de resultados y conclusión: La investigación culmina con un análisis exhaustivo de los resultados obtenidos de los modelos entrenados. Este análisis tiene en cuenta varias métricas de rendimiento para comparar y contrastar la eficacia de cada modelo. La investigación concluye con la formulación de hallazgos clave, que brindan información sobre las capacidades y limitaciones del estado actual del estado del arte en la segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo. Además, ofrece recomendaciones valiosas para el trabajo futuro en este dominio, lo que podría catalizar nuevos avances en el campo.

4.10 Implementación

En la subsiguiente sección, se delinea meticulosamente cada fase del proceso, conforme a la metodología previamente expuesta, para llevar a cabo la implementación de un modelo avanzado de inteligencia artificial destinado a la clasificación y segmentación de la atrofia peripapilar.

4.10.1 Propuesta de Investigación

En la exploración rigurosa de técnicas de vanguardia para segmentar la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo, se identificaron una tendencia creciente hacia el uso de modelos de aprendizaje profundo para abordar este complejo desafío de segmentación. Entre las redes de segmentación semántica que exhibieron resultados favorables se encuentran FCN, SegNet, Unet, mientras que el enfoque de segmentación de instancias estuvo representado principalmente por Mask R-CNN.

Comentado [GAF23]: Debería tener un breve encabezado de dos o tres renglones.

4.10.2 Búsqueda de Base de Datos

Para la determinación de conjuntos de datos adecuados para la investigación, se priorizo la búsqueda en bases de datos de acceso público con un enfoque principal en el diagnóstico de glaucoma, asegurando la presencia de atrofia peripapilar Alfa y Beta. Además, buscamos casos de atrofia en otras patologías, lo que resultó en la incorporación de imágenes que presentaban cataratas y lesiones retinianas. En consecuencia, se eligieron las bases de datos ORIGA, Retina y Drishti-GS1 para este estudio. Para delimitar las máscaras de segmentación de las atrofiás peripapilares se empleó la herramienta de anotación de Roboflow [62] como se muestra en la Figura 17, lo que permite exportar anotaciones en el formato JSON y Mask PNG requerido para el entrenamiento de modelos de segmentación semántica y de instancias.

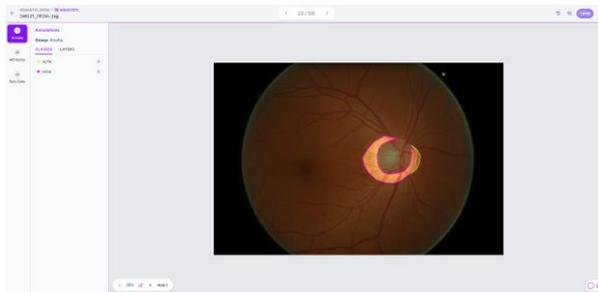


Figura 17. Framework Roboflow.

Luego de la anotación exitosa de aproximadamente el 40% de la base de datos ORIGA con regiones de atrofia peripapilar Alfa y Beta, se procedió a iniciar la validación de las máscaras de segmentación través de colaboraciones con expertos del Instituto Mexicano de Oftalmología, obteniendo un balance de clases de 58.80% para la clase Beta y 41.20% para la clase Alfa en la muestra de la base de datos.

4.10.3 Selección de línea de segmentación

Con la muestra de la base de datos validada por especialistas, se procedió a realizar experimentos preliminares con nuestros modelos propuestos, comparando la segmentación semántica y las técnicas de segmentación de instancias. Donde para los modelos de segmentación semántica como FCN, Segnet y Unet se establecieron variaciones en sus redes troncales para tener un contexto más amplio de los alcances de estos modelos, en la Tabla 2 se muestra en detalle cada una de las configuraciones exploradas. Para una observación más detallada de estos experimentos se pueden visualizar los resultados obtenidos en la sección de Resultados. Como resolución de los conocimientos obtenidos en esta primera experimentación, los resultados orientaron nuestra

Comentado [GAF24]: Tilde

Comentado [GAF25]: Tilde

Comentado [GAF26]: Singular

investigación hacia la exploración de modelos de segmentación de instancias, específicamente Mask R-CNN.

Tabla 2. Configuración de modelos de segmentación semántica

Modelo	Modelo Base	Modelo de Segmentación
Unet	Vanilla CCN	U-Net
Unet VGG	VGG 16	U-Net
Unet Resnet50	Resnet50	U-Net
Unet Mobilenet	MobileNet	U-Net
Unet mini	Vanilla Mini CCN	U-Net
Segnet	Vanilla CCN	Segnet
Segnet VGG	VGG 16	Segnet
Segnet Resnet50	Resnet50	Segnet
Segnet Mobilenet	MobileNet	Segnet
FCN 8	VGG 16	FCN 8
FCN 8 VGG	VGG 16	FCN 8
FCN 8 Mobilenet	MobileNet	FCN 8
FCN 32	VGG 16	FCN 32
FCN 32 VGG	VGG 16	FCN 32
FCN 32 Mobilenet	MobileNet	FCN 32

Comentado [GAF27]: Revisa todas las semánticas!!

4.10.3 Técnicas de Preprocesamiento

Con la dirección de investigación establecida y la muestra de la base de datos validada por expertos, se procede a profundizar en analizar diversas técnicas de preprocesamiento.

La tarea de preprocesamiento en cualquier análisis computacional de imágenes es indispensable. El propósito de este proceso es mejorar la calidad de los datos que se alimentarán en los modelos, reduciendo así posibles discrepancias y errores durante el proceso de aprendizaje. En el contexto de la segmentación de la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo, las técnicas de preprocesamiento utilizadas son de vital importancia.

4.10.3.1 Extracción de región de interés (ROI)

La extracción de ROI implica identificar y aislar la región en la imagen que contiene la información más relevante para la tarea en cuestión. En el contexto de las imágenes de fondo de ojo, esta sería típicamente la región alrededor del disco óptico donde está presente la atrofia peripapilar (zonas Alfa y Beta). El aislamiento de estas áreas permite agilizar el proceso de segmentación mediante la eliminación de datos innecesarios y la reducción de la complejidad computacional. Esta técnica permite que los modelos se concentren en regiones críticas, mejorando así su precisión y eficiencia.

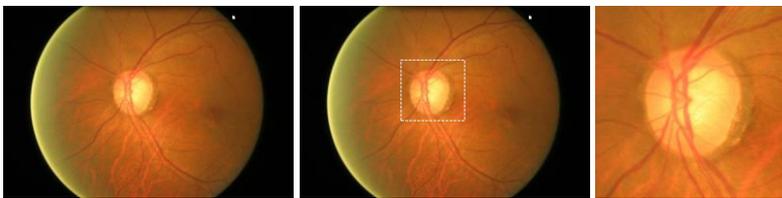


Figura 18. Extracción de la región de interés [54].

4.10.3.2 Aumento de datos volteo horizontal y vertical

Este método tiene como objetivo aumentar la variabilidad del conjunto de entrenamiento para ayudar al modelo a aprender características más sólidas y generalizadas. Las técnicas pueden incluir aumentos de imagen como rotación, escalado, volteo, desenfoco y ajuste de brillo. Estas transformaciones pueden ayudar a que el modelo sea menos sensible a estas variaciones en los datos del mundo real, lo que en última instancia puede mejorar el rendimiento del modelo, especialmente cuando se trata de un conjunto de datos limitado.

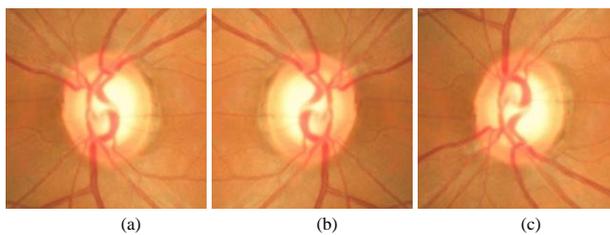


Figura 19. Aumento de datos: original(a), Flip Horizontal(b) y Flip Vertical(c).

4.10.3.3 Ajuste de contraste

Las imágenes de fondo de ojo pueden exhibir una amplia gama de calidad y diferentes niveles de contraste. Las imágenes de bajo contraste pueden plantear un desafío para que un modelo distinga las diferencias entre la APP Alfa y Beta. En este contexto, ajustar el contraste de la imagen se convierte en un paso de preprocesamiento con gran potencial para mejorar la visibilidad de variaciones sutiles y, en consecuencia, mejorar el rendimiento de la segmentación.

En este trabajo, se evaluó el impacto de emplear la técnica de estiramiento de contraste, o normalización, esta técnica brinda una mejora de imagen simple y eficaz que incrementa la calidad visual de una imagen ampliando el rango de valores de intensidad que contiene para abarcar un rango deseado de valores, generalmente el rango completo de intensidades de píxeles que el tipo de imagen es capaz de mostrar.

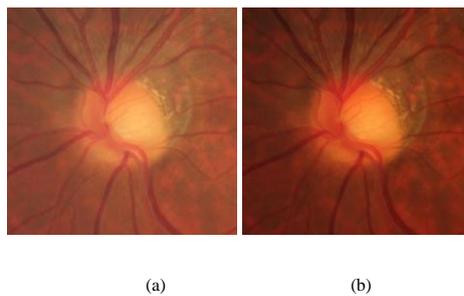


Figura 20. Comparación de imagen original (a) vs ajuste de Estiramiento de Contraste (b).

4.10.3.4 Blur

La adquisición de imágenes de fondo de ojo a menudo encuentra problemas relacionados con el enfoque, lo que hace que algunas imágenes sean menos nítidas que otras. Estas imágenes desenfocadas o borrosas presentan desafíos únicos para los modelos de segmentación basados en IA, que se entrenan principalmente en imágenes de alta resolución bien enfocadas. En estas circunstancias, el rendimiento del modelo puede deteriorarse, al no poder identificar y segmentar con precisión los casos de atrofia peripapilar (APP).

Para mejorar la solidez de nuestro modelo y su capacidad para manejar la variabilidad del mundo real, introdujimos deliberadamente 'desenfoco' como un paso de preprocesamiento. Esta técnica consiste en desenfocar artificialmente nuestras imágenes de entrenamiento para simular las variaciones de enfoque que pueden ocurrir durante el proceso de captura de imágenes. De esta forma, aumentamos nuestro conjunto de datos con una gama más amplia de condiciones de imagen, incorporando las variaciones que el modelo puede encontrar en un entorno práctico.

Comentado [GAF28]: Parece la misma imagen volteada. Especifica cuál es cuál antes y después del ajuste del contraste.

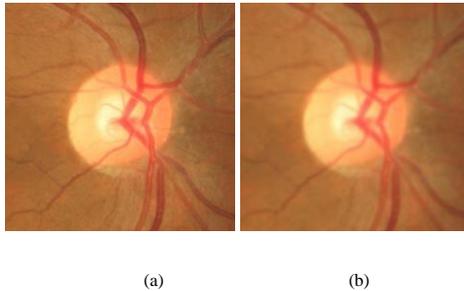


Figura 21. Comparación de imagen original (a) vs. Ajuste de desenfoque (b).

Comentado [GAF29]: Especificar antes y después

4.10.3.5 Ajuste de exposición de imagen

Los distintos niveles de brillo e iluminación en las imágenes del fondo de ojo pueden afectar la visibilidad de la atrofia peripapilar. Ajustar la exposición de las imágenes dentro de un rango de +/- 10% puede ayudar para tener en cuenta las variaciones en el brillo de la imagen debido a las diferencias en las condiciones de imagen o los dispositivos. Por lo que resulta apropiado explorar el comportamiento del modelo frente a estas variaciones, como estrategia para mejorar potencialmente su robustez y generalización.

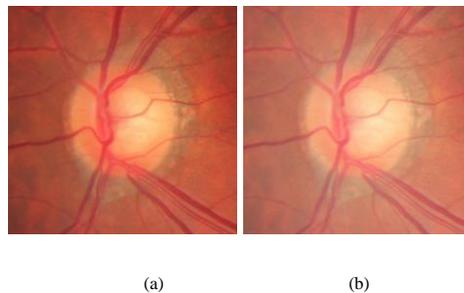


Figura 22. Comparación de imagen original (a) vs. ajuste de exposición de imagen + 10%.

Comentado [GAF30]: Idem a las descripciones de imágenes anteriores y ya revisa de aquí en adelante, donde quiera que establezcas una comparación.

Las imágenes de fondo de ojo, capturadas con cámaras de fondo de ojo especializadas a menudo muestran un alto grado de variabilidad en términos de calidad, brillo y contraste debido a una variedad de factores que incluyen las características del paciente, las variaciones del equipo de imágenes y las condiciones de iluminación. Esta variabilidad representa un desafío para la capacidad del modelo de IA para segmentar y clasificar con precisión las estructuras patológicas, en particular cuando son sutiles, como en el caso de las instancias de APP.

Los resultados obtenidos a través de la comparación de diferentes técnicas de preprocesamiento nos orientaron hacia la utilización de la extracción de regiones de interés (ROI) y el aumento de datos, específicamente donde se incorporaron todas las técnicas en conjunto. Esta estrategia no solo amplió nuestro conjunto de datos, sino que también promovió la variancia de las características de la imagen, lo cual es de suma importancia en la detección del glaucoma, dado que la enfermedad puede manifestarse a través de patrones espaciales variables.

Vale la pena señalar que la selección de estas técnicas de preprocesamiento se basó en su potencial para mejorar las características y estructuras clave relevantes para la segmentación de la atrofia peripapilar Alfa y Beta.

4.10.4 Exploración de Hiperparámetros

Como se ha mencionado la primera parte de la metodología implicó la exploración de hiperparámetros en modelos de segmentación semántica. Estos experimentos iniciales demostraron que las variaciones en la estructura de la columna vertebral pueden mejorar considerablemente el rendimiento de los modelos de segmentación semántica. En consecuencia, se orienta la investigación hacia la optimización de hiperparámetros en modelos de segmentación de instancias. Analizamos el peso en la pérdida de máscaras, funciones de pérdidas para clasificación y regresión, entrenamientos en diferentes períodos de épocas, optimizadores, estructuras troncales y tamaño de cuadros delimitadores. Esto con base a un análisis de errores, métricas cuantitativas y cualitativas. Con el objetivo de mejorar el rendimiento de los modelos.

4.10.4.1 Análisis Mask Loss Weight:

El peso de pérdida de máscara es un hiperparámetro que controla la contribución de la pérdida de máscara a la pérdida total de la red. Un valor demasiado alto podría obligar al modelo a enfatizar demasiado la precisión de la máscara a costa de otros aspectos como la precisión de Bbox [65]. Evaluamos varios valores de pérdida de peso de máscara para encontrar un equilibrio que ofreciera el rendimiento óptimo. Esto se llevó a cabo a través de prueba y error, respaldado por el enfoque de búsqueda aleatoria guiada por el comportamiento de las diferentes funciones de pérdida en el entramiento de la red.

4.10.4.2 Análisis de función de pérdida: Cross Entrophy y Focal Loss

Las funciones de pérdida desempeñan un papel fundamental en el entrenamiento de modelos de aprendizaje profundo, ya que guían el proceso de optimización al proporcionar una medida del rendimiento del modelo. En esta experimentación, se examinaron dos funciones de pérdida: Cross Entrophy y Focal Loss.

Comentado [GAF31]: Tilde

Comentado [GAF32]: tilde

Comentado [GAF33]: tilde

Comentado [GAF34]: tilde

Comentado [GAF35]: Tilde, revisar todas!!

- **Cross Entrophy**

La pérdida de Cross Entrophy es una función de pérdida ampliamente utilizada para problemas de clasificación. Calcula la disimilitud entre la distribución de probabilidad predicha (resultado del modelo) y la distribución real [66]. Para un problema de clasificación de N clases, la pérdida de entropía cruzada se define como:

$$CE = -\sum_{i=1}^{i=N} y_i \cdot \log(\hat{y}_i) \quad \text{Ec. 12}$$

Dónde:

CE : Cross Entrophy

y_i : Valor real

\hat{y}_i : Predicción

Si bien es efectivo en muchos casos, Cross Entrophy puede tener problemas en situaciones donde hay un desequilibrio de clases o cuando las clases son difíciles de distinguir, un problema común en imágenes médicas.

- **Focal Loss**

Para superar estos desafíos, se considera la función Focal Loss. Introducido por Lin et al. en su artículo de 2017, Focal Loss está diseñado para abordar el desequilibrio de clase en las tareas de detección de objetos. Modula la pérdida de Cross Entrophy con un factor que reduce el peso de los ejemplos fáciles y se enfoca en los ejemplos difíciles [67]. Se define como:

$$FL = -\sum_{i=1}^{i=N} \alpha_i (1 - p_i)^{\gamma} \hat{y}_i \cdot \log(\hat{y}_i) \quad \text{Ec. 13}$$

Dónde:

FL : Focal Loss

α_i : Factor de ponderación de desbalance de clases

γ : Parámetro de enfoque

$(1 - p_i)$: Parámetro de reducción de peso

\hat{y}_i : Predicción

La adición de este parámetro de enfoque permite que el modelo se concentre en instancias más difíciles de clasificar, lo que lo convierte en una opción adecuada para esta investigación.

Comentado [GAF36]: Espacio de más

Comentado [GAF37]: Numerar las ecuaciones.

4.10.4.3 Análisis de Bbox Loss: L1 Loss, SmoothL1, DIoU, CIoU

Como se ha mencionado las funciones de pérdida juegan un papel crucial en el entrenamiento de modelos de aprendizaje profundo, incluidos modelos de detección de objetos como Mask RCNN. Estas cuantifican qué tan bien se alinean las predicciones del modelo con los valores reales y proporcionan una medida que el modelo busca minimizar durante el entrenamiento. En el contexto de la detección de objetos, un componente crítico de la tarea del modelo es la regresión del cuadro delimitador, donde el modelo predice las coordenadas de un cuadro que abarca un objeto en una imagen[68].

En la exploración de hiperparámetros para reducir error de localización para modelos de detección de objetos, se han propuesto diferentes funciones de pérdida utilizadas en la regresión de cuadro delimitador (Bbox): Loss L1, SmoothL1, DIoU y CIoU.

- **Loss L1**

También conocida como pérdida de error absoluta, es la diferencia absoluta entre una predicción y el valor real, calculada para cada ejemplo en un conjunto de datos. Esta pérdida es un cálculo de error para cada ejemplo, lo que proporciona una métrica de qué tan bien predijimos para esa observación[69].

$$L1 = \sum_{i=1}^{i=N} |y_i - \hat{y}_i| \quad \text{Ec. 14}$$

Dónde:

L1: Loss L1

y_i : Valor real

\hat{y}_i : Predicción

L1 Loss no es sensible a los valores atípicos, ya que es simplemente la diferencia absoluta.

- **Smooth L1**

Smooth L1 Loss, también conocida como Huber Loss, es una modificación de las pérdidas estándar L1 y L2. La función de pérdida L1 se basa en la diferencia absoluta, mientras que la L2 utiliza la diferencia cuadrada. Smooth L1 Loss combina esencialmente las fortalezas de ambas. Funciona de manera diferente dependiendo de la magnitud de la diferencia absoluta entre el valor predicho y el valor real [70]. La representación matemática de Smooth L1 Loss es la siguiente:

$$Smooth\ L1 = \begin{cases} \frac{1}{2}x^2, & x < 1 \\ |x| - \frac{1}{2}, & x \geq 1 \end{cases} \quad Ec. 15$$

Dónde:

Smooth L1: Smooth Loss L1

x : $|y_i - \hat{y}_i|$

y_i : Valor real

\hat{y}_i : Predicción

Smooth L1 Loss se comporta como L1 Loss para errores más grandes y como L2 Loss para errores más pequeños. En consecuencia, es menos sensible a los valores atípicos en comparación con L2 Loss y no experimenta el problema de los gradientes explosivos, lo que garantiza un proceso de aprendizaje estable.

Comentado [GAF38]: La mencionas pero no la explicas, sirve para entender mejor la explicación previa.

- **GIoU (Intersección Generalizada sobre Union)**

GIoU Loss es una extensión de IoU loss, propuesta para abordar algunas de sus deficiencias. A diferencia de la pérdida de IoU, que solo considera el área de superposición entre los cuadros delimitadores predichos y reales, la pérdida de GIoU la ecuación tiene en cuenta los casos que no se superponen [71].

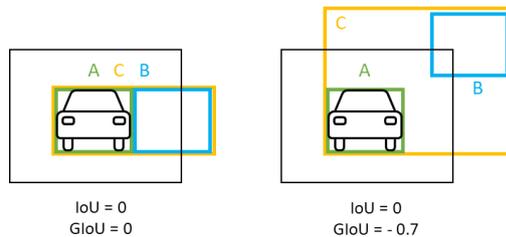


Figura 23. Diagrama esquemático de las pérdidas IoU Y GIoU.

$$L_{GIoU} = 1 - IoU + \frac{|C/BUA|}{|C|} \quad Ec. 16$$

Dónde:

IoU : Intersección sobre la unión de A y B

C: Cuadro más pequeño que encierre A y B.

A: Valor real.
 B: Predicción.

Una deficiencia de la pérdida de IoU es que se convierte en cero cuando no hay superposición entre los cuadros delimitadores predichos y reales, lo que no proporciona un gradiente para la mejora del modelo. La pérdida de GIoU, al considerar el área fuera de los dos cuadros, pero dentro del cuadro más pequeño, puede continuar brindando gradientes y mejorando el rendimiento del modelo incluso cuando no hay superposición.

- **DIoU (Intersección Generalizada sobre Union)**

DIoU Loss introduce un término de distancia del punto central a la pérdida de IoU estándar, con el objetivo de tener en cuenta la superposición y la distancia del punto central simultáneamente. La idea principal es acercar el cuadro delimitador predicho a la ground truth, no solo en términos de IoU sino también con respecto a sus puntos centrales [72].

Comentado [GAF39]: Creo que utilizar la terminología en inglés estará bien, solo ponerla en cursiva.

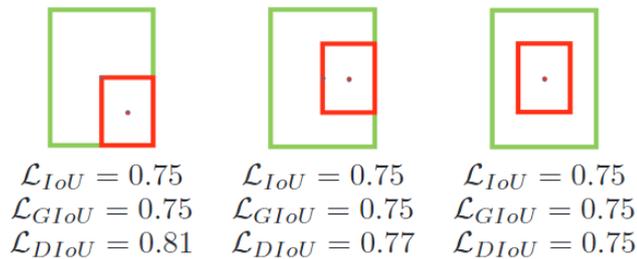


Figura 24. Diagrama esquemático de las pérdidas IoU, GIoU y DIoU [72].

Comentado [GAF40]: Pon aquí también la referencia al documento de donde sacaste las imágenes.

$$L_{DIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} \quad \text{Ec. 17}$$

Dónde:

IoU: Intersección sobre la unión de A y B

ρ : Distancia euclidiana entre los puntos centrales del cuadro delimitador de valor real y el cuadro delimitador predicho.

c: Cuadro más pequeño que encierre A y B.

b: Predicción.

b_{gt} : Valor real.

DIoU Loss alienta al modelo a no solo aumentar la superposición entre los cuadros delimitadores predichos y reales, sino también a minimizar la distancia entre sus centros. Esta característica única lleva a mejorar la precisión del cuadro delimitador y la tasa de convergencia.

- **CIoU (Intersección Completa sobre Union)**

CIoU Loss representa un avance en la evolución de las funciones de pérdida basadas en IoU. Amplía la función de pérdida de intersección de distancia sobre unión (DIoU) al incorporar un término adicional que explica la relación de aspecto de los cuadros delimitadores, lo que da como resultado una regresión de cuadro delimitador más precisa. La pérdida de CIoU es una combinación de tres componentes: la IoU, la distancia entre los puntos centrales de los cuadros delimitadores predichos y reales, y la consistencia de la relación de aspecto entre estos cuadros [73].

Matemáticamente, la pérdida de CIoU se puede representar de la siguiente manera:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b_{gt})}{c^2} + \alpha v \quad \text{Ec. 18}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad \text{Ec. 19}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad \text{Ec. 20}$$

Dónde:

IoU : Intersección sobre la unión de A y B

ρ : Distancia euclidiana entre los puntos centrales del cuadro delimitador de valor real y el cuadro delimitador predicho.

c : Distancia diagonal del cuadro más pequeño que encierre A y B.

b : Punto central de bbox de predicción.

b_{gt} : Punto central de bbox de valor real.

v : Parámetro de consistencia de la relación de aspecto.

α : Parámetro de compensación positivo.

En la configuración experimental, cada función de regresión de cuadro delimitador se evaluó meticulosamente y su desempeño se midió con una serie de métricas. Curiosamente, nuestros

hallazgos revelaron que, a pesar de su simplicidad, la función de pérdida de L1 manifestó un rendimiento superior en comparación con otras funciones de regresión consideradas.

El rendimiento superior de la función de pérdida L1 podría estar relacionado con su papel en la configuración original de la red, lo que sugiere que la red puede optimizarse inherentemente para esta función de pérdida en particular. Sin embargo, es crucial interpretar estos resultados con precaución, ya que pueden depender del conjunto de datos específico, la arquitectura del modelo y la metodología de entrenamiento empleada.

Además, vale la pena señalar que, si bien la función de pérdida de L1 mostró el mejor rendimiento en nuestros experimentos, esto no excluye la eficacia potencial de otras funciones de regresión de cuadro delimitador en diferentes contextos o entornos. Por lo tanto, la elección de la función de regresión de cuadro delimitador debe guiarse no solo por la evidencia experimental sino también por una comprensión integral del dominio del problema, los requisitos específicos de la tarea y las características de los datos disponibles.

4.10.4.5 Análisis de Optimizador: SGD y Adam

La elección del algoritmo de optimización juega un papel fundamental en el entrenamiento del modelo de aprendizaje automático. No solo afecta la velocidad de convergencia del modelo, sino también su rendimiento final. En este trabajo, investigamos dos algoritmos de optimización populares: Stochastic Gradient Descent (SGD) y Adaptive Moment Estimation (Adam), en el contexto del modelo Mask R-CNN para segmentar instancias de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo.

- **SGD (Descenso de gradiente estocástico)**

El descenso de gradiente estocástico es una variante del algoritmo de descenso de gradiente tradicional, un método de optimización de primer orden que utiliza el gradiente de la función objetivo para encontrar el mínimo. La diferencia radica en la cantidad de ejemplos utilizados para calcular el gradiente en cada paso. Mientras que Gradient Descent usa todos los ejemplos (es decir, un "lote"), SGD usa solo un ejemplo elegido al azar por paso, de ahí el nombre "estocástico"[74].

La fórmula matemática de SGD se expresa de la siguiente forma. Para una función de costo dada $J(\theta)$, se realiza la siguiente actualización para cada instancia de entrenamiento i hasta la convergencia:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)}) \quad \text{Ec. 21}$$

Dónde:

θ : Parámetros de entrenamiento.

η : Tasa de aprendizaje.

$\nabla_{\theta} J(\theta; x^{(i)}, y^{(i)})$: Gradiente de la función de costo

$x^{(i)}$: Variable de entrada.

$y^{(i)}$: Etiqueta de la variable de entrada.

- **Adam (Adaptive Moment Estimation)**

Adam es un algoritmo de optimización de la tasa de aprendizaje adaptativo. Aprovecha el poder de dos extensiones de SGD: Adaptive Gradient Algorithm (AdaGrad) [75] y Root Mean Square Propagation (RMSProp)[76]. Adam calcula las tasas de aprendizaje adaptativo para diferentes parámetros al considerar una estimación del primer momento (la media) y el segundo momento bruto (la varianza no centrada) de los gradientes [77].

Comentado [GAF41]: Definir siglas y referenciar estos optimizadores.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t \quad \text{Ec. 22}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad \text{Ec. 23}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad \text{Ec. 24}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \text{Ec. 25}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad \text{Ec. 26}$$

Dónde:

θ : Parámetros de entrenamiento.

g_t : Gradiente en el paso de tiempo t .

η, ϵ : Tasa de aprendizaje y constante para mantener estabilidad numérica.

m_t, v_t : Estimaciones del primer momento (media) y segundo momento (varianza no centrada) de los gradientes respectivamente.

\hat{m}_t, \hat{v}_t : Estimaciones corregidas por sesgo.

β_2, β_1 : Tasas de decaimiento exponencial para las estimaciones de momento

En el caso concreto del modelo Mask R-CNN aplicado a la segmentación de atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo, entran en juego varias consideraciones. Se llevó a cabo una gran cantidad de experimentos, empleando ambos optimizadores en múltiples escenarios para garantizar una comparación completa y confiable. La determinación entre SGD y Adam no fue claramente evidente en las métricas obtenidas, lo que implica que el rendimiento de ambos optimizadores estuvo casi a la par. Estos hallazgos subrayan el comportamiento matizado de los algoritmos de optimización, que está fuertemente influenciado por la tarea específica, la distribución de datos y la arquitectura del modelo en cuestión.

A pesar de la actuación muy disputada entre SGD y Adam, la decisión finalmente se inclinó hacia el uso de SGD para entrenamientos completos. Esta elección estuvo guiada no solo por los resultados experimentales, sino también por la consideración de varios factores, incluida la estabilidad de la convergencia y resultados cualitativos.

4.10.4.6 Análisis de Backbone: ResNet50 y ResNet101

El Backbone es un componente fundamental de cualquier modelo de red neuronal convolucional, principalmente responsable de la extracción de características de una imagen de entrada. En esta investigación, consideramos ResNet50 y ResNet101, dos variantes de Residual Networks (ResNet), como la columna vertebral de nuestros modelos de segmentación.

Los modelos ResNet son reconocidos por su capacidad para gestionar el problema de la desaparición de gradientes en redes profundas. Están diseñados con conexiones de acceso directo (o conexiones de salto) que facilitan el desvío de algunas capas en la red neuronal, lo que permite que el modelo aprenda una función de identidad que mitiga el problema de la pérdida de información [47].

- **ResNet50**

ResNet50, como sugiere su nombre, consta de 50 capas, incluida 1 capa de entrada, 48 capas convolucionales y 1 capa totalmente conectada. Dada su profundidad relativamente moderada, ResNet50 es computacionalmente eficiente y puede entrenarse con hardware menos potente sin comprometer la calidad de los resultados. Este modelo sirve como un excelente punto de partida para nuestros experimentos, ya que ofrece un equilibrio entre la complejidad computacional y el rendimiento.

- **ResNet101**

Comentado [GAF42]: Pudieras revisar esto?, tenía entendido que Adam proporcionaba mayor estabilidad en la convergencia.

Comentado [JA43R42]: La selección de SGD fue tomada por su comportamiento en nuestra experimentación, donde denotó un mejor desempeño que Adam.

Comentado [GAF44]: Yo creo que puedes poner backbone, es terminología aceptada en el campo.

Por otro lado, ResNet101 consta de 101 capas. Esta arquitectura más profunda permite que el modelo aprenda características más complejas de los datos, lo que podría mejorar la precisión de las tareas de segmentación. Sin embargo, entrenar ResNet101 requiere más recursos computacionales en comparación con ResNet50. Este aumento de la complejidad se consideró justificable en nuestra investigación, dadas las posibles mejoras en la segmentación de la atrofia peripapilar.

En los experimentos, ambas redes troncales se evaluaron en función de su capacidad para mejorar el rendimiento de los modelos de segmentación de instancias. Esto permitió discernir si profundizar en las estructuras del modelo puede conducir a mejoras en la tarea de segmentar la atrofia peripapilar.

4.10.4.7 Ajuste de cuadros de anclaje

En modelos de detección de objetos se utiliza el concepto de cuadros de anclaje, que son cuadros delimitadores predefinidos de una determinada altura y anchura. Ajustar el tamaño y la relación de aspecto de estos cuadros ancla para que se ajusten mejor a su conjunto de datos específico puede reducir el error de localización [78].

El análisis de los gráficos de anchos, alturas y relaciones de aspecto del cuadro delimitador puede proporcionar información sobre el tamaño y la forma típicos de los objetos en su conjunto de datos. Así es como puede interpretar los resultados:

- **Anchuras y Alturas**

Analizando la distribución de los cuadros delimitadores de nuestra base de datos, es posible identificar la frecuencia de los tamaños más comunes. Si el pico de anchos y altos está en valores más pequeños, entonces los objetos típicos en el conjunto de datos son más pequeños y es posible que necesite usar cuadros de ancla más pequeños en su configuración. Si el pico está en valores más grandes, entonces los objetos son más grandes y se debería considerar usar cuadros de anclaje más grandes.

- **Relaciones de aspecto**

La relación de aspecto es la relación entre el ancho y la altura. Si el pico de la distribución está cerca de 1, eso significa que sus objetos suelen ser cuadrados. Si el pico es inferior a 1, los objetos suelen ser más altos que anchos y es posible que se desee utilizar cuadros

Comentado [GAF45]: El?

de anclaje con una relación de aspecto similar. Si el pico es mayor que 1, los objetos suelen ser más anchos que altos.

Después de analizar estas distribuciones, fue posible establecer propuestas iniciales para la generación de cuadros delimitadores que después de un análisis se generaron ajustes acordes a resultados expresados en las métricas de evaluaciones.

4.10.5 Análisis de modelos: Cascade Mask RCNN y Mask Scoring

En este estudio, examinamos dos modelos de segmentación de instancias particulares: Cascade Mask R-CNN y Mask Scoring R-CNN.

- **Cascade Mask R-CNN**

Cascade Mask R-CNN es una evolución del modelo Mask R-CNN. Este modelo incorpora un enfoque de detección de objetos de varias etapas que refina las predicciones en cada etapa. En otras palabras, Cascade Mask R-CNN entrena varios modelos de Mask R-CNN, cada uno de los cuales se centra en predicciones de alta calidad de la etapa anterior. Este mecanismo en cascada mejora la precisión del modelo, especialmente para casos desafiantes donde los objetos a segmentar son pequeños o tienen formas irregulares, como es el caso de la atrofia peripapilar.

- **Mask Scoring R-CNN**

Mask Scoring R-CNN también se basa en Mask R-CNN al agregar una rama para predecir la calidad de las máscaras de segmentación. Esta rama adicional, que calcula una puntuación IoU de máscara, permite que el modelo tenga en cuenta tanto la puntuación de clasificación como la puntuación de calidad de la máscara al realizar predicciones. En consecuencia, Mask Scoring R-CNN puede generar máscaras más precisas en comparación con Mask R-CNN estándar, lo que es beneficioso para segmentar estructuras complicadas como la atrofia peripapilar Alfa y Beta.

4.10.6 Entrenamiento General

Esta etapa del proceso de investigación está dedicada a la formación integral de los modelos de alto rendimiento identificados en la fase exploratoria anterior. El proceso de entrenamiento de un modelo es una fase central y vital en el proceso general, ya que tiene la clave de la solidez, la

adaptabilidad y el éxito general de los modelos en la segmentación eficaz de las regiones de interés en las imágenes del fondo de ojo.

Los modelos seleccionados están sujetos a una amplia formación, utilizando varias bases de datos. El uso de múltiples bases de datos no es simplemente una cuestión de amplificar el volumen de datos de entrenamiento, sino que es una decisión estratégica que expone los modelos a una diversa gama de datos. Estas bases de datos ofrecen variaciones en términos de las características de las imágenes, incluidas, entre otras, las diferencias en el equipo utilizado para la captura de datos, la población de la que se recopilan los datos, la calidad de la imagen y la presencia de biomarcadores específicos. Esta diversidad es crítica ya que imita la variabilidad del mundo real, equipando al modelo para generalizar mejor y funcionar de manera confiable en diferentes circunstancias.

V. RESULTADOS Y DISCUSIÓN

Este capítulo revela los hallazgos de nuestros experimentos realizados con el objetivo de crear un modelo de inteligencia artificial capaz de segmentar de manera efectiva la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo. Nuestra exploración comenzó con el etiquetado y validación de una muestra de la base de datos ORIGA, que permitiera establecer los primeros experimentos para definir la dirección de la investigación enfocado a modelos de segmentación semántica o de instancias. Con ello desarrollar una experimentación de los diferentes métodos de preprocesamiento y combinación de hiperparámetros que mejor se ajusten a la tarea de investigación.

5.1 Adquisición de la base de datos

Como se ha mencionado anteriormente la selección de una base de datos adecuada para el etiquetado y la validación es un paso crucial en nuestra metodología. En este caso, inicialmente se eligió la base de datos ORIGA de la que se extrajo un subconjunto de 268 imágenes para nuestro estudio. Esta extracción dio como resultado una base de datos compuesta por 207 máscaras específicas para la atrofia peripapilar Beta y 145 máscaras para la atrofia peripapilar Alfa. Las imágenes utilizadas de la base de datos ORIGA eran de alta resolución, concretamente de 3072 x 2048 píxeles, y guardadas en formato JPG.

Cabe destacar que esta muestra fue utilizada para determinar la línea de segmentación a emplear durante la investigación y análisis de la extracción de la región de interés. Posterior a ello, para tener un contexto real de las características de la base de datos frente a los diferentes hiperparámetros explorados y demás técnicas de preprocesamiento se trabajó con la base de datos

Comentado [GAF46]: Creo que lleva tilde.

Comentado [GAF47]: Tilde.

completamente validada y finalmente en la experimentación final se introdujeron las bases de datos de RETINA y DRISHTI-GS1.

5.2 Resultados de Segmentación Semántica

Comentado [GAF48]: Tilde.

Las investigaciones iniciales se centraron en la exploración de una gama de modelos de segmentación semántica. La segmentación semántica se refiere a la tarea de clasificar cada píxel de una imagen en una clase específica, lo que da como resultado una comprensión integral de la imagen a nivel de píxel.

En esta fase de nuestra investigación se emplearon los modelos de redes totalmente convolucionales (FCN), SegNet y Unet. Cada uno de los cuales aporta características y ventajas únicas a la tarea de segmentación. Los resultados se midieron en función de métricas como precisión, recall, F1 score e Intersección sobre Unión (IoU).

Se hizo uso de modelos disponibles en el repositorio de GitHub [79] y [80], demostrando resultados de segmentación satisfactorios y ejemplificando la implementación simple pero impactante de modelos reconocidos en la literatura. Es crucial notar que, durante el proceso de comparación, extrajimos los modelos para evaluación del repositorio sin alterar la estructura algorítmica de ninguno. Además, seleccionamos una base de datos que comprende imágenes enmascaradas que delinean tres clases distintas por imagen (fondo, Alfa y Beta) a través de la distinción de píxeles.

Los experimentos se realizaron en la estación de trabajo PC-2. Los parámetros de entrenamiento incluyeron un tamaño de lote de dos y diez épocas de entrenamiento, incorporando la técnica de parada temprana como protección contra el sobreajuste. Esta técnica monitorea constantemente el valor de pérdida por época, sirviendo como criterio decisivo para la terminación. El entrenamiento también utilizó optimizador Adam y la función de pérdida de Cross Entropy.

Para reforzar la imparcialidad de la comparación de modelos, se implementó una estrategia de validación cruzada de cinco, que ha demostrado reducir el sesgo al evaluar la efectividad del modelo [81]. Esto solamente para los modelos FCN, SegNet y Unet.

Tras la implementación de los modelos de segmentación semántica mencionados anteriormente. Se estableció una tabla comparativa de desempeño para facilitar su análisis. La Tabla 3 sirve como un resumen completo de las métricas de rendimiento de cada modelo y presenta información pertinente que refleja su perfil de rendimiento. El esquema organizativo de la tabla presenta el nombre de la arquitectura del modelo en la primera columna, seguido de la sensibilidad, la

precisión y la puntuación F1 en las columnas siguientes. La columna final muestra la puntuación de Intersección sobre Unión (IoU), una métrica crítica que indica la superposición entre la segmentación predicha y la realidad del terreno.

Cabe destacar que cada valor representado en la Tabla 3 representa la puntuación métrica media de las clases de interés (Alfa, Beta, Fondo).

Comentado [GAF49]: Entiendo lo que dices pero lo leo raro, si pudieras redactar mejor.

Tabla 3. Métricas de evaluación promedio de Segmentación semántica.

Modelo	Sensitividad	Precisión	F1	IoU
Unet	0.422872538	0.578870096	0.445871648	0.371745326
Unet VGG	0.424668700	0.656726427	0.456783617	0.404989830
Unet Resnet50	0.450021967	0.726202610	0.494157316	0.431584518
Unet Mobilenet	0.509541283	0.694882211	0.566212375	0.480521759
Unet mini	0.331530693	0.332429490	0.331530801	0.333333224
Segnet	0.402613668	0.469181336	0.408177434	0.375648983
Segnet VGG	0.495574293	0.425872150	0.434548325	0.390540518
Segnet Resnet50	0.461445894	0.496516901	0.433895397	0.393720801
Segnet Mobilenet	0.482338111	0.589697126	0.495186544	0.434488554
FCN 8	0.423276589	0.593136416	0.441259860	0.395021014
FCN 8 VGG	0.361831946	0.515973570	0.372145080	0.353339558
FCN 8 Mobilenet	0.450834337	0.692384935	0.484729554	0.426341171
FCN 32	0.372461821	0.462829075	0.390366992	0.362941986
FCN 32 VGG	0.340752777	0.351304121	0.343261754	0.337405914
FCN 32 Mobilenet	0.407678427	0.509537711	0.424556399	0.385652328

Tras la comparación detallada de los diferentes modelos, resulta evidente que Unet Mobilenet se destaca como el modelo más prometedor, obteniendo la mayoría de los puntajes superiores en nuestras métricas de evaluación. Específicamente, Unet Mobilenet logró resultados de 0.5095, 0.5662 y 0.4805 correspondientes a la sensibilidad, el puntaje F1 y la métrica IoU respectivamente.

Finalmente, vale la pena destacar también el rendimiento del modelo Unet Resnet50. A pesar de no encabezar la lista en todas las categorías, este modelo demostró su fortaleza con la métrica de precisión más alta, logrando un impresionante puntaje de 0.7262.

Para una evaluación más granular de los resultados, la Tabla 4 se presenta a continuación, mostrando en detalle las métricas obtenidas para cada clase individual. Esta desagregación permite apreciar con mayor precisión el rendimiento de cada modelo en la segmentación de las diferentes clases (fondo, Alfa y Veta) siguiendo el orden presentado en la Tabla 3.

Tabla 4. Métricas de evaluación por clase de Segmentación semántica.

Modelo	Sensitividad	Precisión	F1	IoU
	[Fondo, Alfa, Beta]	[Fondo, Alfa, Beta]	[Fondo, Alfa, Beta]	[Fondo, Alfa, Beta]
Unet	[0.998, 0.049, 0.220]	[0.995, 0.306, 0.434]	[0.997, 0.081, 0.258]	[0.994, 0.027, 0.192]
Unet-VGG	[0.999, 0.030, 0.244]	[0.996, 0.410, 0.564]	[0.997, 0.053, 0.320]	[0.995, 0.028, 0.192]
Unet-Resnet50	[0.999, 0.078, 0.272]	[0.996, 0.468, 0.715]	[0.998, 0.111, 0.373]	[0.995, 0.061, 0.238]
Unet-Mobilenet	[0.999, 0.152, 0.377]	[0.997, 0.350, 0.738]	[0.998, 0.205, 0.496]	[0.996, 0.114, 0.332]
Unet-mini	[1.000, 0.000, 0.000]	[0.995, 0.000, 0.000]	[0.997, 0.000, 0.000]	[0.995, 0.000, 0.000]
Segnet	[0.998, 0.000, 0.209]	[0.996, 0.000, 0.412]	[0.997, 0.000, 0.228]	[0.994, 0.000, 0.133]
Segnet -VGG	[0.990, 0.005, 0.492]	[0.997, 0.009, 0.271]	[0.994, 0.006, 0.304]	[0.988, 0.003, 0.181]
Segnet-Resnet50	[0.987, 0.000, 0.397]	[0.997, 0.097, 0.396]	[0.992, 0.001, 0.309]	[0.984, 0.000, 0.197]
Segnet-Mobilenet	[0.999, 0.097, 0.351]	[0.997, 0.186, 0.587]	[0.998, 0.070, 0.418]	[0.995, 0.038, 0.270]
FCN 8	[0.999, 0.048, 0.223]	[0.996, 0.356, 0.428]	[0.997, 0.073, 0.253]	[0.994, 0.039, 0.152]
FCN 8 VGG	[1.000, 0.027, 0.059]	[0.995, 0.166, 0.386]	[0.997, 0.028, 0.092]	[0.995, 0.014, 0.051]
FCN 8 Mobilenet	[0.999, 0.038, 0.316]	[0.996, 0.436, 0.645]	[0.998, 0.050, 0.400]	[0.995, 0.030, 0.254]
FCN 32	[0.999, 0.009, 0.109]	[0.995, 0.047, 0.347]	[0.997, 0.012, 0.162]	[0.994, 0.006, 0.089]
FCN 32 VGG	[1.000, 0.000, 0.022]	[0.995, 0.000, 0.059]	[0.997, 0.000, 0.033]	[0.862, 0.000, 0.018]
FCN 32 Mobilenet	[0.999, 0.009, 0.215]	[0.996, 0.048, 0.485]	[0.997, 0.013, 0.263]	[0.995, 0.007, 0.156]

Al analizar las Tablas 3 y 4, y teniendo en cuenta las métricas basadas en clases, se hace evidente que ciertas redes lidian con la clasificación eficiente de la clase Alfa. Esta ineficiencia podría atribuirse a la ausencia de máscaras donde se expresen de forma detallada los márgenes fronterizos entre atrofas. Estableciendo una generalización excesiva, lo que llevó a una clasificación errónea común de la mayoría de las APP Alfas como Betas.

Los modelos Unet, aunque arrojaron resultados alentadores en general, el modelo Unet Mini, en particular, expresó mínimas puntuaciones en todas las métricas principales (precisión, sensibilidad, puntaje F1 e IoU) con valores respectivos de 0.3315, 0.3324, 0.3315 y 0.3333. Estas cifras señalaron una sorprendente ausencia de segmentación de APP en los resultados, lo que

destaca la importancia de evaluar diferentes versiones de un modelo para identificar la más adecuada para abordar el problema específico en cuestión.

Para ilustrar aún más el rendimiento y la eficacia de estos modelos, se seleccionaron aleatoriamente muestras del conjunto de prueba para demostrar visualmente los resultados de la segmentación. Estas muestras, representadas en la Figura 25 subrayan las fortalezas y debilidades de los diversos modelos, proporcionando representaciones tangibles de sus capacidades en la tarea de segmentar la atrofia peripapilar Alfa y Beta.

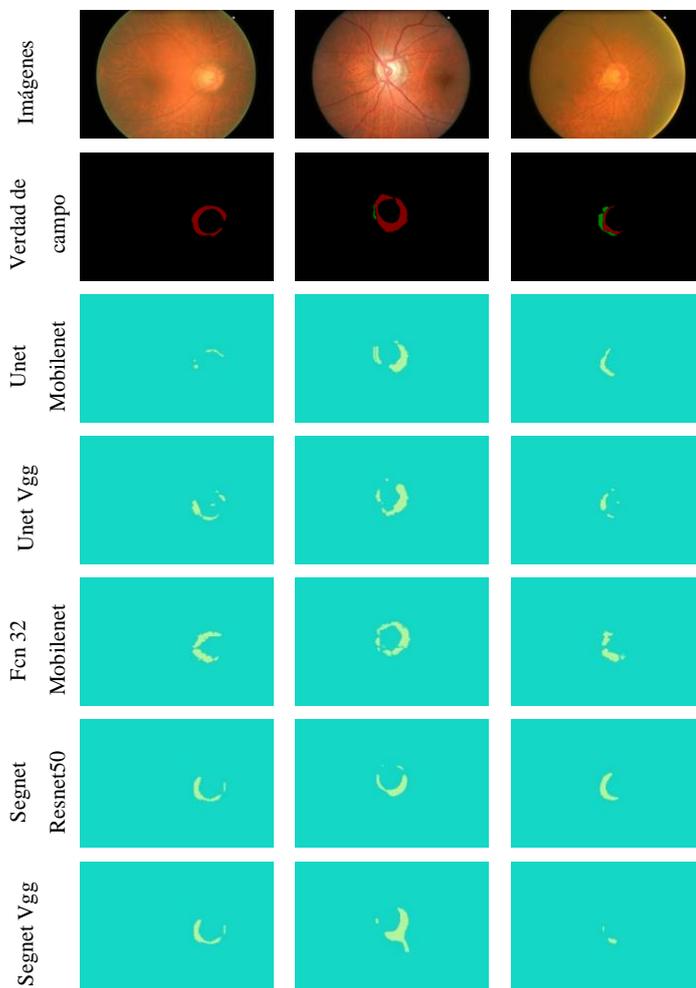
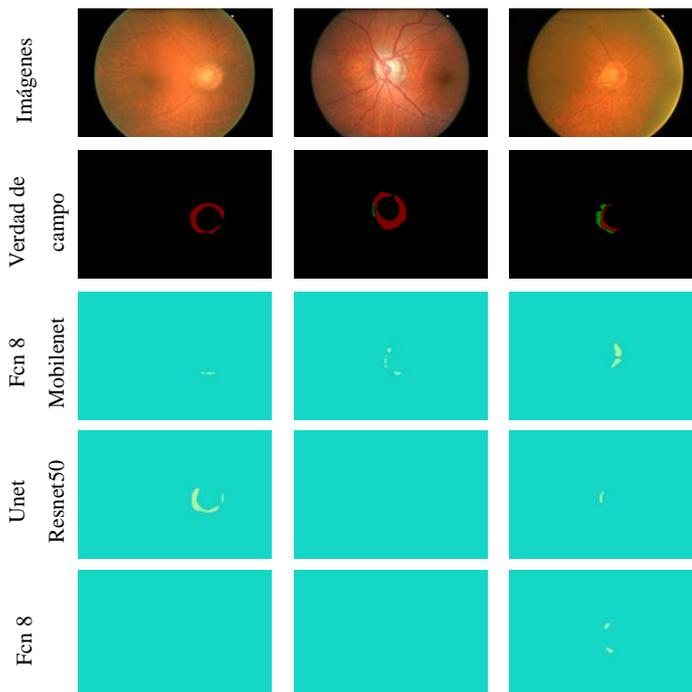


Figura 25. Resultados de segmentaciones de los modelos con altas métricas vs la verdad de campo.

Las imágenes retinianas originales están en la primera fila. Las áreas de APP de las máscaras de verdad se muestran en la segunda y las predicciones de los modelos se encuentran en las siguientes filas.

La Figura 25 muestra una comparación entre los resultados de segmentación y la verdad de campo de los modelos con mayor puntuación en las métricas. Las predicciones obtenidas demostraron una variabilidad en la segmentación de las imágenes. No obstante, Unet Mobilenet expresó resultados que se acercan a la estructura morfológica y localización de los biomarcadores Alfa y Beta, las máscaras predichas son casi idénticas a su correspondiente verdad de campo. Sin embargo, se obtuvieron resultados experimentales con baja similitud, como se aprecian en la Figura 26.



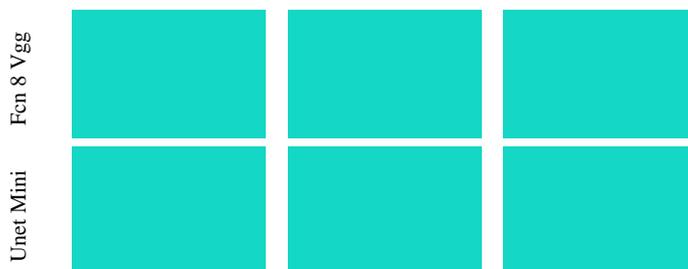


Figura 26. Resultados de segmentaciones de los modelos con bajas métricas vs la verdad de campo.

Estos resultados refuerzan la importancia de realizar una búsqueda más amplia de modelos de segmentación.

5.3 Resultados de Segmentación de Instancias

Procediendo con la experimentación, se analiza el desempeño del modelo Mask RCNN, realizado a través de MMDetection en el entorno de desarrollo de Google Colab. El modelo Mask R-CNN es uno de los modelos de segmentación de instancias destacados que se utilizan en tareas de visión por computadora, y su eficacia en el manejo de tareas complejas de segmentación de imágenes formó la base de su selección para este estudio.

El conjunto de datos seleccionado para esta fase experimental se dividió cuidadosamente en subconjuntos de Entrenamiento, Validación y Prueba siguiendo la relación de distribución estándar de 70 %, 20 % y 10 %, respectivamente. Esta división estratificada que se representa en la Figura 27 aseguró que el modelo tuviera una cantidad significativa de datos para aprender los detalles intrincados de la tarea (conjunto de entrenamiento), ajustar sus parámetros (conjunto de validación) y finalmente evaluar su rendimiento general de manera imparcial (conjunto de prueba).

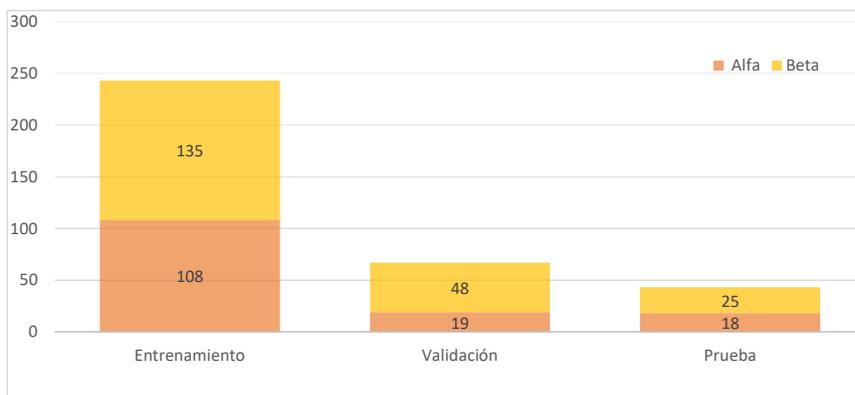


Figura 27. Balance de clases por subconjuntos de datos.

El proceso de entrenamiento del modelo Mask R-CNN requería establecer hiperparámetros específicos. Los valores elegidos para estos fueron consistentes con la configuración predeterminada del repositorio MMDetection en GitHub, que ha sido ampliamente aceptado por la comunidad de investigación. Estos incluyeron un tamaño de lote de dos y una tasa de aprendizaje de 0,0025 siguiendo una política de tasa de aprendizaje "Step" en un curso de 12 épocas. La ventaja de adherirse a esta configuración predeterminada es la reducción de la necesidad de un ajuste extenso de hiperparámetros, lo que simplifica significativamente el proceso de capacitación y garantiza la replicabilidad de los resultados.

Es importante destacar que la fase de entrenamiento se inició con pesos preentrenados del modelo Mask R-CNN, previamente entrenado en el conjunto de datos COCO a gran escala. Aprovechar estos pesos preexistentes aprovecha el poder del aprendizaje de transferencia. Como el conjunto de datos COCO proporciona una gran variedad de datos de entrenamiento, permite que el modelo Mask R-CNN aprenda características comunes y discriminatorias en diferentes categorías de objetos, mejorando así su capacidad de generalización [19].

Tabla 5. Métricas de evaluación por clase de Segmentación de Instancias.

Modelo	Prec	Rec	IoU	F1	AP	Clase
Mask RCNN	44.23	67.64	36.5	53.48	10.70	Beta
	31.57	50.0	24.0	38.70	6.20	Alfa

La Tabla 5 presenta las métricas de rendimiento del modelo Mask R-CNN cuando se aplica a la tarea de delinear zonas de atrofia Alfa y Beta en imágenes de fondo de ojo. Cada fila refleja las métricas de rendimiento del modelo para la clase de atrofia correspondiente.

Al analizar los resultados, vemos que el modelo se desempeñó considerablemente mejor en la detección de la atrofia Beta en comparación con la atrofia Alfa. Esta diferencia en el rendimiento se puede atribuir a varios factores, como la complejidad y la variabilidad en la apariencia y estructura de los tipos de atrofia y el posible desequilibrio en los datos de entrenamiento.

En conclusión, mientras que el modelo Mask R-CNN demuestra un cierto nivel de eficacia en la delimitación de las zonas de atrofia Beta, presenta desafíos considerables para detectar la atrofia Alfa. Esto podría deberse a una combinación de características del conjunto de datos, la arquitectura del modelo y la elección de hiperparámetros, y justifica una mayor investigación y perfeccionamiento del proceso de entrenamiento del modelo y quizás de la arquitectura del modelo en sí.

5.4 Selección de línea de segmentación

Realizando un análisis exhaustivo y una comparación de los dos modelos principales aplicados en esta investigación: el modelo de segmentación semántica, Unet con una Backbone MobileNet, y el modelo de segmentación de instancias, Mask R-CNN. Se establece la Tabla 6 como comparativa de las métricas obtenidas en cada modelo.

Tabla 6. Métricas de evaluación por clase de Segmentación de Instancias vs Semántica.

Modelo	Prec	Rec	IoU	F1	AP	Clase
Mask RCNN	44.23	67.64	36.5	53.48	10.70	Beta
	31.57	50.0	24.0	38.70	6.20	Alfa
Unet-MobileNet	73.8	37.7	33.2	49.6	-	Beta
	35.00	15.2	11.4	20.50	-	Alfa

La precisión, calculada como la relación entre los verdaderos positivos y la suma de los verdaderos positivos y los falsos positivos, destaca la capacidad del modelo para evitar detecciones de falsos positivos. Para la clase de atrofia Beta, Unet-MobileNet supera notablemente a Mask R-CNN con una precisión del 73,8 % frente al 44,23 %. Esta mayor precisión indica que

Comentado [GAF50]: Tilde.

el modelo Unet-MobileNet identifica con mayor precisión los casos de atrofia Beta, minimizando los falsos positivos.

Por el contrario, en lo que respecta a la atrofia Alfa, Mask R-CNN supera a Unet-MobileNet, aunque marginalmente, con una precisión del 31,57 % frente al 35 % de Unet-MobileNet. Esto sugiere que, para la atrofia Alfa, Mask R-CNN proporciona un poco menos de falsos positivos.

El recall se define como la relación entre los verdaderos positivos y la suma de los verdaderos positivos y los falsos negativos. Significa la capacidad del modelo para detectar todos los posibles positivos. Para ambas clases, atrofia Beta y Alfa, el modelo Mask R-CNN supera a Unet-MobileNet en recall. Para la atrofia Beta, Mask R-CNN y Unet-MobileNet ofrecen un recall del 67,64 % y el 37,7 % respectivamente, y para la atrofia Alfa, Mask R-CNN produce un recall del 50 % en comparación con el 15,2 % de Unet-MobileNet. Esto indica que Mask R-CNN es más hábil para capturar todos los casos de atrofia, independientemente del tipo.

La métrica de IoU ofrece una medida objetiva de la superposición entre las áreas de verdad predichas y de campo. La puntuación F1, por otro lado, proporciona una perspectiva equilibrada del rendimiento del modelo, combinando precisión y recuperación.

En términos de IoU, Mask R-CNN logra mejores puntajes para atrofia Alfa y Beta (36,5% y 24,0%) que Unet-MobileNet (33,2% y 11,4%). Esto demuestra que Mask R-CNN funciona de manera superior cuando se trata de superponer con precisión las áreas predichas con los valores reales.

Si bien las puntuaciones F1 para la atrofia Beta son comparables entre los dos modelos, Mask R-CNN supera a Unet-MobileNet para la atrofia Alfa, lo que sugiere que Mask R-CNN proporciona un rendimiento más equilibrado en términos de precisión y recall de la atrofia Alfa.

La precisión promedio (AP) es una métrica valiosa en las tareas de segmentación de instancias que agrega la precisión lograda en diferentes umbrales de recuperación. Como Unet-MobileNet es un modelo de segmentación semántica, no produce instancias separadas y no se pudo calcular AP.

Para Mask R-CNN, los valores de AP se calcularon como 10,70 % para atrofia Beta y 6,20 % para atrofia Alfa. Estos valores destacan el mejor rendimiento de Mask R-CNN en la detección de instancias de atrofia Beta sobre atrofia Alfa, aunque mejora en ambos.

En resumen, la capacidad única de Mask R-CNN para la segmentación de instancias lo convierte en una potente herramienta para la segmentación de la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo. La capacidad de delinear y clasificar con precisión instancias individuales de atrofia ofrece un nivel de detalle y precisión que es invaluable para los análisis médicos. El trabajo futuro será dirigido para mejorar el rendimiento de este modelo, lo que sugiere enfocarse en estrategias para mejorar la precisión, como el ajuste del modelo, el aumento de datos, exploración de modelos con línea de segmentación de instancias.

Comentado [GAF51]: Tilde.

5.5 Análisis de Errores

Dada la complejidad de las tareas de detección de objetos, es importante realizar un análisis exhaustivo de los errores para comprender mejor los tipos de errores que comete un modelo. Para los modelos de detección de objetos utilizados en este estudio, empleamos la herramienta TIDE, un completo marco de análisis de errores diseñado para categorizar y cuantificar los distintos tipos de errores encontrados durante la detección de objetos.

El principio clave detrás de la categorización de errores de TIDE es el uso de la intersección sobre la unión (IoU). Utilizando dos umbrales IoU, a saber, el umbral de primer plano (T_f) y el umbral de fondo (T_b), TIDE asigna una única categoría de error a cada predicción de salida, o la reconoce como correcta.

Las categorías de error y sus definiciones propuestas por TIDE, ilustradas en la Figura 28, son las siguientes:

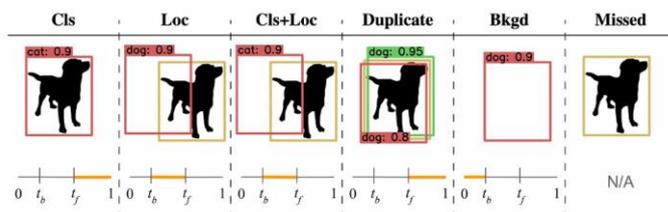


Figura 28. Tipos de Errores en Detección de Objetos[83]

- **Error de clasificación (CLS):** $\text{IoU} \geq T_f$ para el objetivo de la clase incorrecta (es decir, localizado correctamente, pero clasificado incorrectamente).
- **Error de localización (LOC):** $T_b \leq \text{IoU} < T_f$ para el objetivo de la clase correcta (es decir, clasificado correctamente, pero localizado incorrectamente).

Comentado [GAF52]: Referencia de dónde sacaste la imagen

- **Error de Cls y Loc (CLS y LOC):** $Tb \leq IoU < Tf$ para el objetivo de la clase incorrecta (es decir, clasificado y localizado incorrectamente).
- **Error de detección de duplicados (DUP):** $IoU \geq Tf$ para el objetivo de la clase correcta pero otra detección de puntuación más alta ya coincidió con el objetivo (es decir, sería correcto si no fuera por una detección de puntuación más alta).
- **Error de fondo (BKG):** $IoU < Tb$ para todos los objetivos (es decir, fondo detectado como primer plano).
- **Error de objetivo perdido (MISS):** todos los objetivos no detectados (falsos negativos) que aún no están cubiertos por un error de clasificación o localización.

Al aplicar la herramienta TIDE al modelo Mask RCNN utilizado en este estudio, observamos una incidencia destacada de errores de localización como se muestra en la Figura 29.

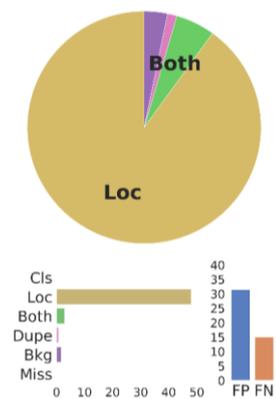


Figura 29. Tipos de Errores en Mask RCNN

Esto sugiere que, si bien el modelo fue capaz de identificar correctamente la clase del objeto (APP Alfa o Beta), a menudo falló a la hora de señalar con precisión la ubicación del objeto dentro de la imagen, lo que dio lugar a un elevado número de falsos positivos.

5.5.1 ROI

Como se ha mencionado en la presente investigación se realizó un análisis comparativo entre modelos de segmentación de instancias entrenados sobre imágenes originales y aquellos entrenados sobre imágenes procesadas mediante ROI. Como una estrategia para reducir el tipo de

error de localización, basándonos en un análisis de mapas de calor de la presencia de las atrofas peripapilares en las imágenes de fondo de ojo como se muestran en la Figura 30.

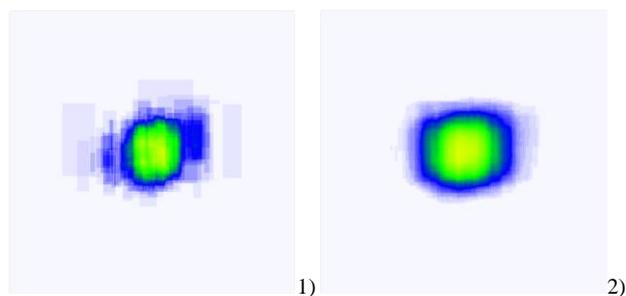


Figura 30. Mapas de calor de localización de atrofia Alfa (1) y Beta (2).

Como es posible observar que las máscaras de atrofia Alfa presentan una menor densidad y mayor variabilidad de localización en comparación a la atrofia Beta. Para evitar la pérdida de máscaras de segmentación se extrajo una región de interés de dimensionamiento de 840x840 con ayuda de un modelo de detección de objetos previamente entrenado para la detección de disco óptico [84].

Tabla 7. Métricas de evaluación p.

Base de Datos	IoU	F1	Prec	Recall	AP	AP:50	Clase
ORIGINAL	36.5	53.48	44.23	67.64	10.70	25.00	Beta
	24.0	38.70	31.57	50.0	6.20		Alfa
ROI	38.33	55.42	45.10	71.88	17.6	32.2	Beta
	23.91	38.60	32.35	47.83	3.20		Alfa

Al examinar los resultados presentados en la Tabla 7, observamos una mejora notable en las métricas de rendimiento del modelo cuando se aplica el enfoque basado en el ROI. La mejora es particularmente pronunciada en la precisión promedio (AP) para la clase de atrofia Beta, donde se observa un aumento significativo de 10.70 en la imagen original a 17.60 en la imagen modificada con ROI. Al mismo tiempo, se observa una mejora en la puntuación de IoU, indicativa de una segmentación más precisa, para la clase de atrofia Beta, que aumenta de 36.50 a 38.33.

La puntuación F1, que armoniza la precisión y la recuperación, ve un aumento de 53.48 a 55.42 en la clase Beta cuando se realiza la transición a imágenes de ROI. Esto sugiere que la compensación entre la capacidad del modelo para identificar correctamente los verdaderos

positivos (precisión) y su capacidad para detectar una mayor cantidad de verdaderos positivos del grupo disponible (recall) se equilibra mejor con la aplicación del ROI.

Sin embargo, la puntuación AP e IoU para la clase de atrofia Alfa no experimentó una mejora similar con el uso de ROI, lo que indica una posible dificultad en la capacidad del modelo para distinguir instancias de atrofia Alfa, lo que indica que esta clase puede ser más desafiante para el modelo de segmentación.

No obstante, el enfoque basado en el ROI contribuye a mejorar el rendimiento de los modelos de segmentación de instancias, como lo corrobora la mejora en las métricas de rendimiento clave en la clase de atrofia Beta. El enfoque específico proporcionado por el aislamiento de la región de interés dirige el modelo a las partes esenciales de la imagen, mejorando su capacidad de segmentación precisa. Sin embargo, el desafío presentado por la clase de atrofia Alfa sugiere la necesidad de aumento de datos o métodos de conjunto para mejorar el rendimiento en ambas clases.

5.6 Análisis de Hiperparámetros

5.6.1 Análisis de peso de pérdida de máscara

El estudio del comportamiento de pérdida a lo largo del entrenamiento del modelo es un componente esencial para evaluar y mejorar la eficacia del modelo. Como parte de nuestro análisis profundo del modelo Mask RCNN, examinamos la pérdida de máscara, un factor crítico que rige el rendimiento del modelo en las tareas de segmentación de instancias.

Como se muestra en la Figura 31, la pérdida de máscara supera considerablemente otras formas de pérdidas durante el proceso de entrenamiento del modelo. Esta observación realza un interés en explorar los efectos potenciales de manipular el peso de pérdida de máscara, el factor de escala que modula la importancia relativa de la pérdida de máscara durante el proceso de optimización.

En esta exploración, se investigaron dos alternativas al valor de peso original de 1, específicamente 0,5 y 2. La hipótesis subyacente estableció que estos cambios en el peso podrían ayudar a mejorar el rendimiento de aprendizaje del modelo, como se refleja en la pérdida de máscara durante las iteraciones.

Comentado [GAF53]: Tildes w hiperparámetro arriba también

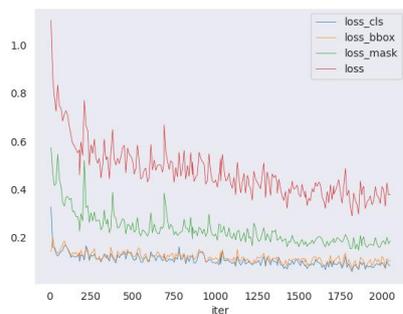


Figura 31. Pérdida vs Iteraciones Mask Loss Weight 1.0

La Figura 32 proporciona una comparativa de la evolución de la pérdida de máscara entre las dos modificaciones del peso de máscara.

Comentado [GAF54]: Lleva tilde. Revisar todos.

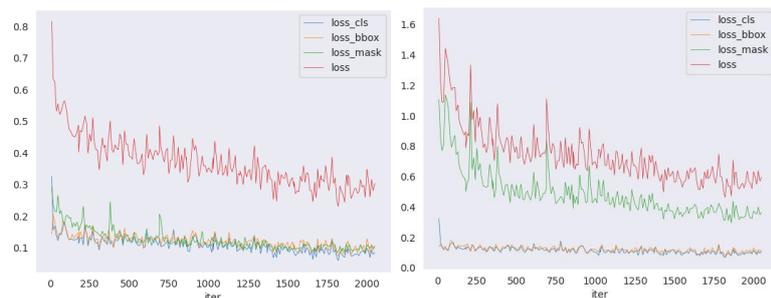


Figura 32. Pérdida vs Iteraciones Mask Loss Weight 0.5 y 2.0

Se observa que cuando el peso se redujo a 0,5, el modelo mostró una convergencia más rápida y valores de pérdida general más bajos. Esta observación insinuó una mayor eficiencia en el aprendizaje. Por el contrario, aumentar el peso a 2,0 condujo a una pérdida inicial superior a 1 y una eventual convergencia a valores mínimos más altos alrededor de 0,4, lo que indica posibles dificultades en el proceso de aprendizaje del modelo en comparación con las otras configuraciones.

Sin embargo, un examen más detallado de las métricas de evaluación en la Tabla 8 sugiere una narrativa contrastante.

Tabla 8. Métricas de evaluación de Mask Loss Weight

Mask Loss Weight	IoU	F1	Prec	Recall	AP	AP:50	Clase
1	35.53	52.43	45.00	62.79	20.40	31.40	Beta
	18.00	30.51	26.09	36.73	3.10		Alfa
0.5	31.48	47.89	39.84	60.00	16.30	25.5	Beta
	12.90	22.86	21.43	24.49	0.70		Alfa
2.0	32.52	49.07	40.15	63.10	20.10	38.3	Beta
	14.29	25.00	24.53	25.49	2.20		Alfa

A pesar de la mejora observada en el comportamiento de pérdida de máscara con una ponderación de 0.5, esto no se tradujo en un rendimiento superior de acuerdo con las métricas de evaluación. Ya que se observa una reducción de la ponderación de cada una de las métricas, no obstante, para la pérdida de máscara con una ponderación de 2.0, se observa un incremento notable para AP:50 pasando de un valor de 31.40 a 38.3, pero, sorprendentemente, la máscara de pérdida de peso original de 1 demostró el rendimiento más equilibrado en las clases de atrofia Alfa y Beta.

Este examen ilustra la complejidad de la optimización del modelo de aprendizaje profundo y enfatiza que las mejoras en un aspecto (en este caso, la pérdida de máscara) no conducen necesariamente a un mejor rendimiento general del modelo.

5.6.2 Análisis de funciones de pérdida en regresión de Bbox

Tras nuestro análisis exhaustivo de los comportamientos de pérdida, centramos nuestra atención en las funciones de pérdida específicas de la regresión de cuadro delimitador, un componente central del modelo Mask RCNN. Específicamente, evaluamos el rendimiento del modelo bajo diferentes funciones de pérdida de regresión de cuadro delimitador, a saber, L1 Loss, Smooth L1 Loss, Distance-IoU (DIoU) y Complete IoU (CIoU).

Cada función de pérdida ofrece una perspectiva única sobre cómo penalizar la discrepancia entre los cuadros delimitadores predichos y reales. Como tal, la elección de la función de pérdida de regresión del cuadro delimitador puede influir significativamente en la capacidad del modelo para ubicar los objetos de interés con precisión.

En nuestro análisis, mantuvimos constantes todos los hiperparámetros en todos los experimentos, excepto la función de pérdida de regresión del cuadro delimitador. Este enfoque sistemático nos

permitió aislar el efecto de la función de pérdida en el desempeño del modelo. En la Figura 33 se observa el análisis de errores para cada una de las funciones evaluadas.

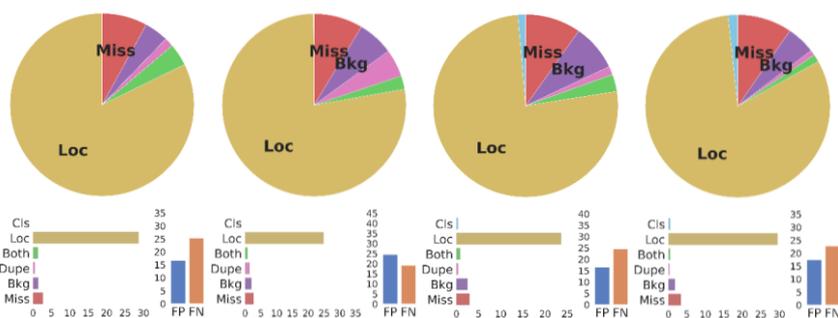


Figura 33. Análisis de Errores para Loss L1, DIoU, CIoU y SmoothL1

Para la configuración base, empleamos la pérdida L1 estándar como la función de pérdida de regresión del cuadro delimitador en el modelo Mask RCNN. Esta elección arrojó una puntuación de error de localización de aproximadamente 30, que fue notablemente más alta que la de DIoU y CIoU, aunque comparable a Smooth L1.

Para una comprensión más a detalle de las funciones se compara en la Tabla 9 las métricas de evaluación.

Tabla 9. Métricas de evaluación de Bbox Loss

Bbox Loss	IoU	F1	Prec	Recall	AP	AP:50	Clase
LossL1	35.53	52.43	45.00	62.79	20.40	31.40	Beta
	18.00	30.51	26.09	36.73	3.10		Alfa
SmoothL1	33.54	50.24	39.85	67.95	18.70	32.10	Beta
	16.84	28.83	27.12	30.77	2.70		Alfa
DIoU	24.09	38.83	28.65	60.23	13.50	25.00	Beta
	7.23	13.48	13.95	13.04	1.60		Alfa
CIoU	33.95	50.69	40.74	67.07	19.80	33.4	Beta
	17.95	30.43	33.33	28.00	3.30		Alfa

Curiosamente, a pesar de que CIOU no mostró los valores más altos en las métricas consideradas, ofreció un rendimiento equilibrado que compite con los mejores. La pérdida de CIOU utiliza una perspectiva más completa sobre la regresión del cuadro delimitador, ya que incorpora el aspecto de la superposición, la distancia del centroide y la relación de aspecto en su cálculo. Esta perspectiva matizada, en teoría, lo hace mejor equipado para manejar diferentes tipos de desafíos de localización. De hecho, CIOU demostró una reducción notable en el valor del error de localización en comparación con L1 Loss sin afectar significativamente los falsos positivos (FP) y los falsos negativos (FN).

DIOU, por otro lado, presentó un comportamiento intrigante, reduciendo tanto el error de localización como los falsos negativos, a expensas de un mayor número de falsos positivos. Esta tendencia se puede atribuir al enfoque de DIOU en la distancia entre los centros de los cuadros delimitadores, lo que puede conducir a una mejor recall, pero a costa de la precisión.

Smooth L1 Loss, a pesar de estar diseñado para ser resistente a valores atípicos en comparación con L1 Loss, no produjo una mejora sustancial en las métricas consideradas.

Este análisis comparativo de las funciones de pérdida de regresión de cuadro delimitador subraya las complejas compensaciones involucradas en la selección de la función de pérdida. Aunque no produce los valores métricos más altos, el rendimiento equilibrado de CIOU, subrayado por sus criterios de evaluación integrales, lo marca como la mejor opción para la regresión de cuadro delimitador en el contexto de la segmentación de la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo.

5.6.3 Análisis de Épocas

En el ámbito del aprendizaje profundo, la cantidad de épocas de entrenamiento, que se refiere a la cantidad de veces que el modelo aprende de todo el conjunto de datos, es un hiperparámetro crucial. A la luz de esto, nuestro próximo análisis examina el impacto de variar el número de épocas de entrenamiento en el rendimiento del modelo Mask RCNN.

Para este análisis, investigamos tres duraciones de entrenamiento diferentes: un entrenamiento corto de 12 épocas, un entrenamiento intermedio de 36 épocas y un entrenamiento extendido de 50 épocas. Estas épocas se eligieron para proporcionar información sobre la evolución del rendimiento del modelo a medida que el entrenamiento avanza con el tiempo.

Los resultados obtenidos en cuanto a las métricas se expresan en la Tabla 10.

Tabla 10. Métricas de evaluación de Épocas

Épocas	IoU	F1	Prec	Recall	AP	AP:50	Clase
12	33.95	50.69	40.74	67.07	19.80	33.4	Beta
	17.95	30.43	33.33	28.00	3.30		Alfa
36	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
50	34.04	50.79	48.98	52.75	17.90	31.10	Beta
	18.10	30.65	26.76	35.85	2.50		Alfa

Como se presenta en la Tabla 10, extender la duración del entrenamiento a 36 épocas arrojó resultados prometedores, con métricas como IoU y puntaje F1 que demostraron una mejora notable con respecto al modelo inicial de 12 épocas.

Sin embargo, es importante tener en cuenta que aumentar el número de épocas de entrenamiento no siempre conduce a un mejor rendimiento. Por ejemplo, una mayor extensión del período de entrenamiento a 50 épocas no mostró ninguna mejora significativa. En cambio, resultó en métricas de rendimiento comparables y, en algunos casos, más bajas en comparación con el modelo de 36 épocas.

Para contrarrestar la posibilidad de sobreajuste, una situación en la que el modelo se vuelve excesivamente complejo y comienza a aprender el ruido de los datos en lugar de los patrones subyacentes, utilizamos una estrategia de selección de modelos basada en la métrica de segmentación de precisión promedio (AP). En concreto, seleccionamos el modelo que alcanzó la puntuación AP más alta durante el entrenamiento. Para el modelo de 12 épocas, esto ocurrió en la época 12; para el modelo de 36 épocas, en la época 30; y para el modelo de 50 épocas, en la época 24.

Curiosamente, los modelos con mejor rendimiento no eran de la época final de cada duración del entrenamiento. Esta observación no es infrecuente en el aprendizaje profundo, donde el rendimiento del modelo puede estabilizarse o incluso degradarse si el entrenamiento se extiende en exceso.

En conclusión, nuestro análisis indica que una duración de entrenamiento de 36 épocas ofrece una compensación beneficiosa entre el rendimiento y la eficiencia computacional para la tarea de investigación.

5.6.4 Análisis de Función de pérdida

Como se ha mencionado la función de pérdida sirve como guía para el proceso de optimización, lo que lleva al modelo al estado que mejor se adapta a los datos. En este estudio, exploramos el impacto de emplear diferentes funciones de pérdida (Entropía cruzada y Pérdida focal)

La función de pérdida Cross Entropy es un estándar ampliamente adoptado en tareas de clasificación debido a su efectividad. Por ende, es apropiado su selección como parámetro base de la configuración del modelo Mask RCNN.

Por otro lado, la función Focal Loss está especialmente diseñada para abordar el problema del desequilibrio de clases y centrarse más en ejemplos difíciles y mal clasificados. Al reducir la contribución de los ejemplos fáciles. En nuestros experimentos, el parámetro α en Focal Loss, que se introdujo para manejar el desequilibrio de clases, se exploró con valores de 0,5 y 0,25.

Al analizar las distribuciones de error que se muestran en la Figura 34, notamos que el empleo de Focal Loss resultó en una disminución en las instancias de "Miss", las instancias que el modelo no pudo identificar. Además, Focal Loss demostró una reducción en los Falsos Negativos (FN), aunque con un ligero aumento en los Falsos Positivos (FP), en comparación con la función de pérdida Cross Entropy. Este comportamiento se alinea con el diseño y las características de la función Focal Loss, que maneja inherentemente ejemplos difíciles y desequilibrio de clases.

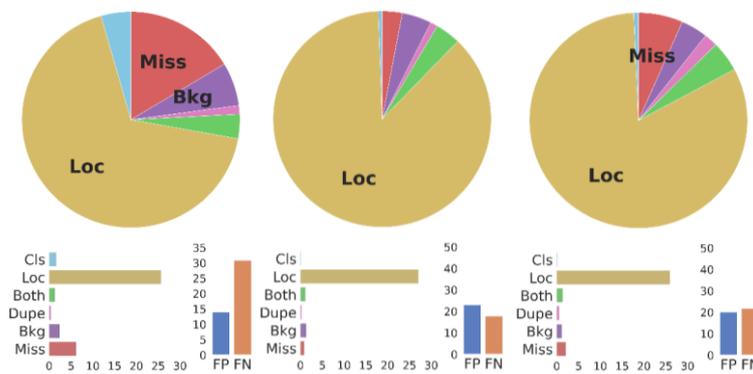


Figura 34. Análisis de Errores para Cross Entropy y Focal Loss con $\alpha:0.5$ y 0.25

Sin embargo, contrastar la eficacia teórica de Focal Loss con los resultados empíricos revela una discrepancia. Como se muestra en la Tabla 11, a pesar de sus atractivas propiedades teóricas, Focal Loss no superó a Cross Entropy en la mayoría de las métricas de evaluación de nuestros experimentos.

Tabla 11. Métricas de evaluación de Épocas

Loss Function	IoU	F1	Prec	Recall	AP	AP:50	Clase
Cross Entropy	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
Focal Loss $\alpha: 0.5$	30.94	47.26	37.09	65.12	15.00	28.90	Beta
	8.08	14.95	14.81	15.09	1.80		Alfa
Focal Loss $\alpha: 0.25$	29.21	45.22	35.86	61.18	18.50	31.40	Beta
	9.38	17.14	17.65	16.67	2.40		Alfa

La función de pérdida de Cross Entropy logró generar un rendimiento superior en términos de Intersección sobre unión (IoU), puntaje F1, precisión, recuperación y precisión promedio (AP).

En conclusión, aunque Focal Loss se ha propuesto como una solución robusta para abordar problemas relacionados con el desequilibrio de clases y ejemplos difíciles, nuestros experimentos con el modelo Mask RCNN sugieren que la función de pérdida Cross Entropy todavía tiene una ventaja en términos de rendimiento para la tarea.

5.6.5 Análisis de Optimizador

Los algoritmos de optimización en el aprendizaje automático juegan un papel importante en el perfeccionamiento de los parámetros del modelo para minimizar la función de pérdida. La selección de un optimizador es un paso importante en el proceso de entrenamiento del modelo y tiene un impacto directo en el rendimiento del modelo. En este estudio, buscamos investigar el impacto de dos optimizadores diferentes, Stochastic Gradient Descent (SGD) y AdamW, en el rendimiento del modelo Mask RCNN.

El optimizador SGD, la base estándar del modelo Mask RCNN, se empleó con la siguiente configuración: una tasa de aprendizaje de 0,0025, un impulso de 0,9, una caída de peso de 0,0001 y un programador de tasa de aprendizaje "Paso", que reduce la tasa de aprendizaje en las épocas 28 y 34. SGD también utilizó iteraciones de calentamiento de 500 y un factor de calentamiento de 0,001. SGD es uno de los algoritmos de optimización más básicos pero efectivos en el aprendizaje automático. A pesar de su simplicidad, SGD ha demostrado tener mucho éxito en el entrenamiento de redes profundas debido a su capacidad para atravesar gradientes ruidosos y escapar de mínimos locales poco profundos.

Por otro lado, se utilizó el optimizador AdamW, un método de tasa de aprendizaje adaptativo propuesto como una variante de Adam, con la siguiente configuración: valores β (0.9, 0.999), ϵ de $1e-08$, una tasa de aprendizaje de 0.00025, un "CosineAnnealing" Programador de tasa de aprendizaje, 1000 iteraciones de calentamiento y un factor de calentamiento de 0.1. Se eligió la variante AdamW debido a su notable desempeño en tareas de segmentación del disco óptico.

La Figura 35 ofrece una vista clara del impacto de los optimizadores en el rendimiento del modelo, donde se encontró que el uso de AdamW aumenta los errores de ubicación en comparación con SGD, en contra de las expectativas.

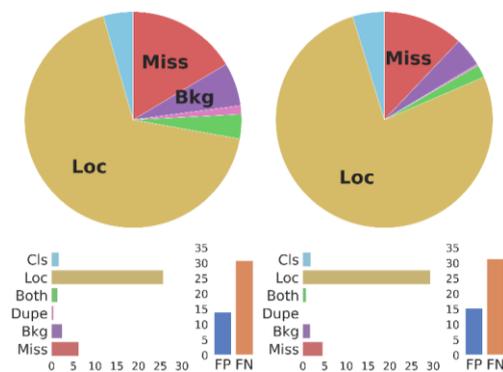


Figura 35. Análisis de Errores para SGD y AdamW.

En la Tabla 12 se presenta una comparación adicional entre SGD y AdamW, que muestra las métricas de evaluación para ambos optimizadores. Las métricas revelan que SGD supera a AdamW en todas de las métricas.

Tabla 12. Métricas de evaluación de Optimizador

Optimizador	IoU	F1	Prec	Recall	AP	AP:50	Clase
SGD	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
AdamW	31.65	48.09	45.83	50.57	15.70	30.30	Beta
	12.90	22.86	21.05	25.00	3.10		Alfa

5.6.6 Análisis de Backbone

Una Backbone bien elegida puede extraer características útiles de los datos, lo que posteriormente mejora el poder predictivo del modelo. En este estudio, se compararon dos redes troncales populares: ResNet50 y ResNet101.

ResNet50 y ResNet101 fueron seleccionados para este análisis debido a su accesibilidad dentro del marco de MMDetection y sus pesos previamente entrenados en el conjunto de datos COCO. Un principio clave en el aprendizaje profundo sugiere que las redes más profundas pueden aprender características más matizadas, mejorando así el rendimiento de los modelos. Sin embargo, la compensación radica en el mayor costo computacional requerido para entrenar estas redes más profundas.

El análisis de errores que se muestra en la Figura 36 revela una reducción en los errores de localización cuando se usa ResNet101 en comparación con ResNet50. Esta mejora podría deberse a la capacidad de ResNet101 para extraer características más detalladas debido a su profundidad. Sin embargo, esto resultó en un aumento en los errores de tipo *Both*, donde los errores de localización y clasificación ocurren simultáneamente. Además, el uso de ResNet101 conduce a un aumento de los falsos positivos (FP), pero a una disminución notable de los falsos negativos (FN).

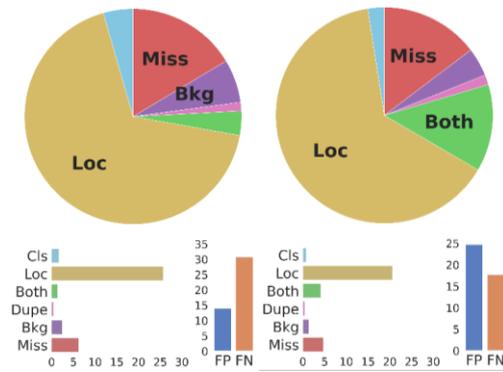


Figura 36. Análisis de Errores para ResNet50 y ResNet101

La Tabla 13 proporciona un resumen de las métricas de rendimiento para ambas redes troncales. A pesar de la anticipación de que una red más profunda como ResNet101 brindaría métricas superiores en todos los aspectos, esto no se observó de manera consistente.

Tabla 13. Métricas de evaluación de Backbone

Backbone	IoU	F1	Prec	Recall	AP	AP:50	Clase
ResNet50	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
ResNet101	38.28	55.37	55.68	55.06	21.70	36.10	Beta
	20.62	34.19	31.25	37.74	2.10		Alfa

Si bien el modelo ResNet101 superó a ResNet50 en la mayoría de las métricas de atrofia Alfa, no produjo una mejora significativa para la atrofia Beta. Sin embargo, considerando su impacto positivo en la reducción de errores de localización, ResNet101 sigue siendo un fuerte competidor para futuras investigaciones y entrenamiento de modelos.

Sin embargo, es fundamental tener en cuenta que el uso de redes más profundas como ResNet101 tiene un costo computacional considerable, con tiempos de entrenamiento que pueden duplicarse en comparación con redes menos profundas como ResNet50. Por lo tanto, se debe lograr un equilibrio entre mejorar el rendimiento del modelo y administrar los recursos computacionales.

5.6.7 Análisis de Bbox Size

El modelo Mask RCNN utiliza un enfoque de cuadro delimitador para localizar objetos de interés dentro de las imágenes. La generación de cuadros delimitadores juega un papel crucial en la segmentación exitosa de la atrofia peripapilar Alfa y Beta en imágenes de fondo de ojo. En esta investigación, se llevaron a cabo una serie de configuraciones experimentales para optimizar la generación de cuadros delimitadores basados en el análisis de características de los cuadros delimitadores, seguido de propuestas por conocimientos empíricos.

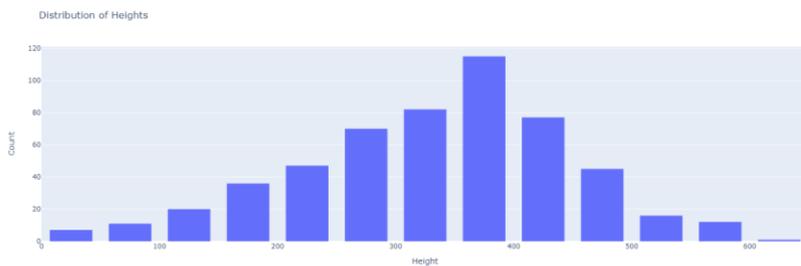
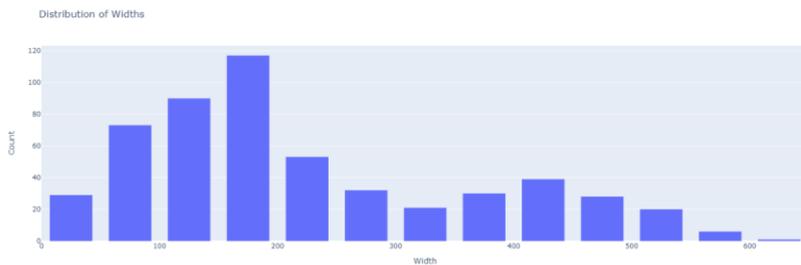
Los parámetros para modificar son: Scales, Ratio y Strides. Los cuales se definen como:

- **Scales:** estos valores denotan las escalas de anclaje y están dictados principalmente por el tamaño de los objetos en el conjunto de datos. El valor de la escala corresponde al tamaño base del ancla, que se calcula como la zancada multiplicada por la escala. Por ejemplo, una escala de 8 significa un tamaño base de 8 píxeles para cada cuadro de anclaje.

- **Ratio:** Las relaciones de aspecto de los anclajes están representadas por estos valores. Las proporciones como [0.5, 1.0, 2.0] implican que el modelo generará cuadros ancla con proporciones de aspecto de 1:2, 1:1 y 2:1 (alto:ancho).
- **Stride:** estos valores dictan los pasos del mapa de características sobre las imágenes de entrada, determinando así la densidad de los puntos de anclaje.

Al proponer nuevas configuraciones de cuadros delimitadores, este estudio analizó las características del ancho, la altura y las proporciones de los cuadros delimitadores en los conjuntos de datos de entrenamiento, validación y prueba. Las observaciones clave de estos conjuntos de datos incluyen:

Conjunto de datos de entrenamiento: los valores de ancho oscilaron entre 25 (mín.) y 650 (máx.), con una moda de 200. Los valores de altura exhibieron un rango similar, con un mínimo de 25, un máximo de 650 y una moda de 370. las proporciones variaron de 0,1 (mín) a 2,8 (máx), con una moda de 0,5.



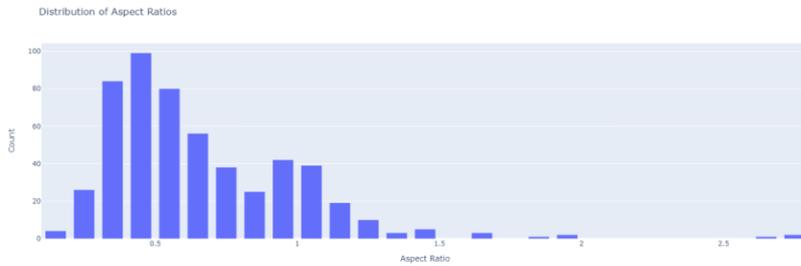
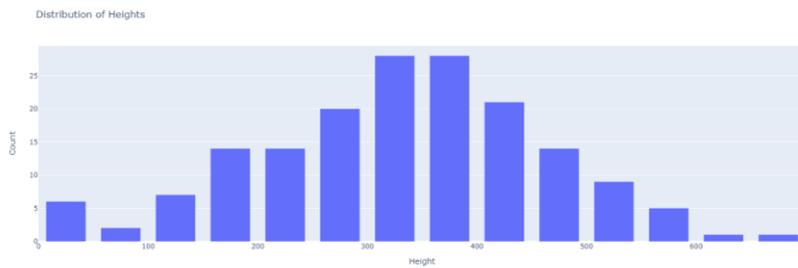
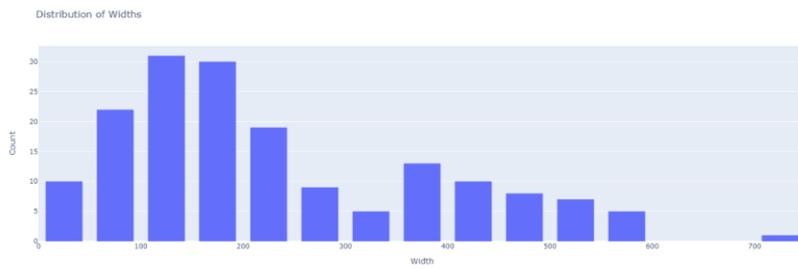


Figura 37. Distribución de Ancho, Alto y Ratio de conjunto de entrenamiento.

Conjunto de datos de validación: los valores de ancho variaron de 25 (mín.) a 750 (máx.), con una moda de 150. Los valores de altura variaron de 25 (mín.) a 700 (máx.), con una moda de 350. Las proporciones variaron de 0,1 (min) a 4 (max), con una moda de 0,5.



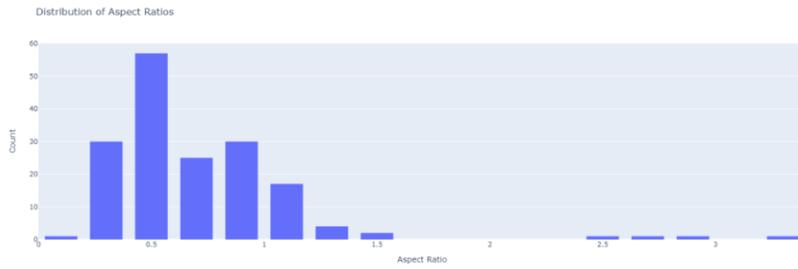
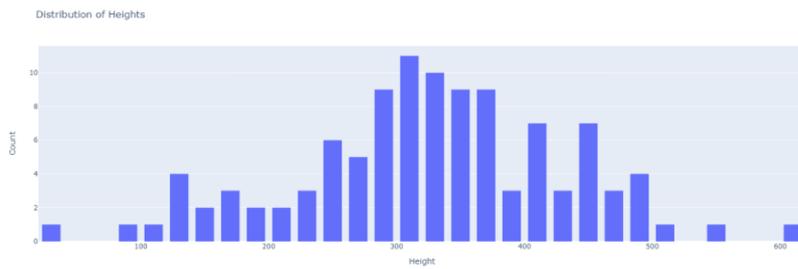
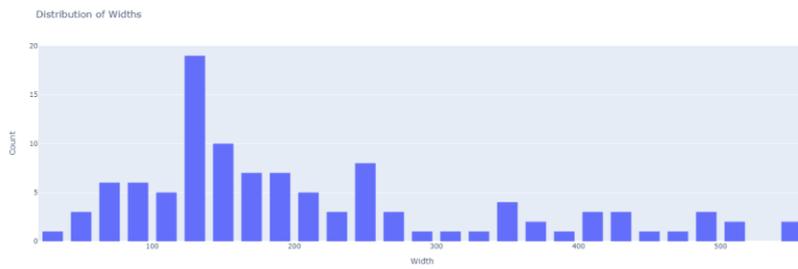


Figura 38. Distribución de Ancho, Alto y Ratio de conjunto de validación.

Conjunto de datos de prueba: los valores de ancho variaron de 20 (mín.) a 560 (máx.), con una moda de 130. Los valores de altura variaron de 20 (mín.) a 620 (máx.), con una moda de 320. Las proporciones variaron de 0,1 (min) a 4 (max), con una moda de 0,5.



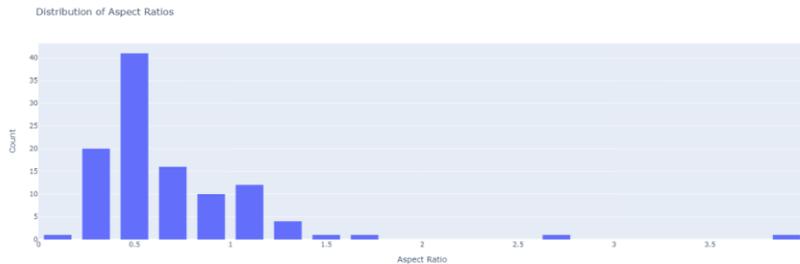


Figura 39. Distribución de Ancho, Alto y Ratio de conjunto de prueba

Visualmente, es evidente que las distribuciones de estas funciones en los conjuntos de datos de entrenamiento, validación y prueba son muy similares, lo que respalda la adopción de un enfoque de configuración de cuadro delimitador unificado para estos conjuntos de datos.

La Tabla 14 presenta los resultados de ocho configuraciones experimentales, con cada configuración variando los parámetros utilizados para generar los cuadros delimitadores, a saber, escalas, proporciones y pasos. La selección de estos parámetros estuvo guiada por las características del mínimo, máximo y modo en la distribución de los conjuntos de datos de entrenamiento, validación y prueba.

Tabla 14. Métricas de evaluación de Backbone

Propuestas	Clase	Mask	Mask	Mask	Mask	Mask	anchor_generator		
		AP	IoU	F1	Precision	Recall	Scales	Ratios	Strides
1	β :	0.1790	0.4237	0.5952	0.5882	0.6024	[8]	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
	α :	0.0360	0.1605	0.2766	0.2889	0.2653			
2	β :	0.125	0.2568	0.4086	0.8261	0.2714	[20, 130, 560, 320, 620]	[0.1, 0.3, 0.5, 1.0, 2.0, 3.0, 4.0]	[8, 16, 32, 64, 128]
	α :	0.016	0.0000	0.0000	0.0000	0.0000			
3	β :	0.155	0.3246	0.4901	0.5286	0.4568	[8, 20, 60, 130, 320, 620]	[0.1, 0.5, 1.0, 2.0, 3.0]	[4, 8, 16, 32, 64]
	α :	0.024	0.1212	0.2162	0.2963	0.1702			
4	β :	0.16	0.3782	0.5488	0.5233	0.5769	[8, 20, 130]	[0.3, 0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
	α :	0.026	0.1194	0.2133	0.3077	0.1633			
5	β :	0.152	0.3143	0.4783	0.4272	0.5432	[8, 20]	[0.5, 1.0, 2.0, 3.0]	[4, 8, 16, 32, 64]
	α :	0.026	0.1444	0.2524	0.2600	0.2453			
6	β :	0.144	0.3905	0.5616	0.6212	0.5125	[8, 15, 20, 40]	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
	α :	0.021	0.1148	0.2059	0.3043	0.1556			
7	β :	0.161	0.3125	0.4762	0.5263	0.4348	[8, 15, 20, 30]	[0.5, 1.0, 2.0, 3.0]	[4, 8, 16, 32, 64]
	α :	0.015	0.1290	0.2286	0.2143	0.2449			
8	β :	0.142	0.3158	0.4800	0.5294	0.4390	[4, 8, 15]	[0.5, 1.0, 2.0]	[4, 8, 16, 32, 64]
	α :	0.028	0.1154	0.2069	0.2308	0.1875			

El experimento 1 se basó en la configuración inicial con una escala de [8], proporciones de [0.5, 1.0, 2.0] y pasos de [4, 8, 16, 32, 64]. Esta configuración sirvió como base para comparaciones posteriores.

En el Experimento 2 se amplió el número de opciones para cada variable, la cual se basó en considerar valores mínimos, máximos y moda. Las escalas se establecieron en [20, 130, 560, 320,

620], proporciones en [0,1, 0,3, 0,5, 1,0, 2,0, 3,0, 4,0] y pasos en [8, 16, 32, 64, 128]. Sin embargo, esta mayor variabilidad condujo a una disminución en el rendimiento para las clases de atrofia Alfa y Beta.

El Experimento 3 tomó como referencia los valores iniciales propuestos, volviendo así a escalas de [8, 20, 60, 130, 320, 620], proporciones de [0.1, 0.5, 1.0, 2.0, 3.0] y los pasos mantenidos como [4, 8, 16, 32, 64].

Los experimentos 4 a 8 redujeron iterativamente la variabilidad de los valores y proporciones escalables mientras mantenían constantes los avances. Estas configuraciones tenían como objetivo centrarse en los cuadros delimitadores de escalas más bajas y enfatizar los tamaños más pequeños al mismo tiempo que se consideraban los tamaños más grandes.

En las ocho configuraciones experimentales, las métricas de rendimiento más altas se lograron en el Experimento 1, lo que destaca a la configuración base como mejor opción para la generación de cuadros delimitadores.

5.7 Análisis de técnicas de Preprocesamiento

5.7.1 Contraste

Dados los contornos inherentemente indistintos de las atrofias peripapilares Alfa y Beta en algunas imágenes de fondo de ojo, este estudio exploró el papel de las técnicas de preprocesamiento de imágenes, como el ajuste de contraste, para mejorar el rendimiento del modelo. Específicamente, se aplicó un ajuste de estiramiento de contraste al conjunto de datos y el rendimiento del modelo Mask RCNN se comparó con el conjunto de datos original sin ajustar.

La evaluación del rendimiento del ajuste de contraste se presenta en la Tabla 15. Los resultados destacan un efecto positivo y negativo del ajuste de contraste en la segmentación de las atrofias peripapilares Alfa y Beta.

Tabla 15. Métricas de evaluación de Contrast stretching

Base de Datos	IoU	F1	Prec	Recall	AP	AP:50	Clase
ORIGINAL	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
CONTRASTE	30.08	46.24	44.44	48.19	17.20	33.50	Beta

17.07	29.17	29.79	28.57	3.50	Alfa
-------	-------	-------	-------	------	------

El impacto del ajuste de contraste sugiere que puede no ser una técnica de preprocesamiento universalmente beneficiosa para todos los tipos de atrofas. Si bien mejora la segmentación de la atrofia Alfa, afecta negativamente a la segmentación de la atrofia Beta, lo que compromete el rendimiento general del modelo.

Por lo tanto, a pesar de sus beneficios potenciales, el ajuste de contraste como técnica de preprocesamiento no se recomienda universalmente para mejorar la segmentación de la atrofia peripapilar Alfa y Beta en las imágenes de fondo de ojo.

5.7.2 Aumento de Datos

Para crear un modelo de segmentación sólido capaz de gestionar la heterogeneidad inherente a los escenarios de captura de imágenes de fondo de ojo del mundo real, se examinaron varias técnicas de aumento de datos: volteo horizontal y vertical, ajuste de exposición de imagen, desenfoque y una combinación de todos estos métodos.

La eficacia de estas técnicas de aumento se compara en la Tabla 16, lo que proporciona una evaluación integral a través de varias métricas.

Tabla 16. Métricas de evaluación de Aumento de Datos

Aumento	IoU	F1	Prec	Recall	AP	AP:50	Clase
ORIGINAL	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
FLIP	36.29	53.25	52.33	54.22	19.90	31.10	Beta
	13.16	23.26	25.00	21.74	2.70		Alfa
EXPOSICIÓN	38.60	55.70	53.66	57.89	17.40	31.00	Beta
	12.12	21.62	32.00	16.33	2.80		Alfa
BLUR	35.48	52.38	52.38	52.38	18.10	32.7	Beta
	17.17	29.31	26.98	32.08	3.10		Alfa

	32.21	48.73	42.48	57.14	17.90		Beta
FLIP-EXP- BLUR	21.21	35.00	29.58	42.86	5.00	31.8	Alfa

Volteo horizontal y vertical (Flip): a pesar de que la lógica de esta técnica mejora la adaptabilidad del modelo a varias orientaciones, los resultados indican una disminución en todas las métricas para las atrofias Alfa y Beta a excepción de AP en comparación con el conjunto de datos original. Esto indica que voltear como técnica de aumento puede no mejorar universalmente el rendimiento del modelo.

Ajuste de exposición de imagen (Exposición): Los resultados también demuestran una disminución en todas las métricas para ambas clases de atrofias en comparación con el conjunto de datos original. Esto sugiere que, si bien el ajuste de la exposición a veces puede mejorar el detalle y la claridad de la imagen, su aplicación como técnica de aumento independiente no condujo a un mejor rendimiento de la segmentación en este contexto.

Desenfoco (Blur): la aplicación de desenfoque dio como resultado una mejora en el rendimiento de la atrofia Alfa en comparación con el conjunto de datos original, como lo demuestra el aumento en IoU, F1 y recuerdo. Sin embargo, para la atrofia Beta, el rendimiento disminuyó ligeramente. El desenfoque parece ser más efectivo para la segmentación de la atrofia Alfa, quizás debido a los límites intrínsecamente borrosos de esta clase.

Técnica compuesta (Flip-Exp-Blur): la técnica de aumento combinado demuestra resultados variables. Para la atrofia Alfa, el rendimiento mejoró en todas las métricas en comparación con el conjunto de datos original, con una mejora notable en la métrica AP. Por el contrario, el rendimiento de la atrofia Beta generalmente disminuyó. Esto indica la efectividad potencial de la técnica compuesta para la atrofia Alfa, pero no para la atrofia Beta.

En resumen, mientras que las técnicas de aumento de datos a veces pueden ayudar a mejorar la solidez del modelo, los resultados de este estudio sugieren que su eficacia depende del tipo de atrofia. Por lo que para una búsqueda de balance de desempeño entre atrofias peripapilares resulta alentador el uso de una técnica compuesta de aumento de datos.

Comentado [GAF56]: Ortografía

5.8 Análisis de Modelos

Con el objetivo de aumentar el rendimiento de nuestro modelo de segmentación, exploramos la implementación de Cascade Mask R-CNN y Mask Scoring R-CNN. Estos modelos sofisticados

brindan mejoras y soluciones para abordar los desafíos inherentes presentados por Mask R-CNN, lo que justifica su evaluación en esta investigación.

En la Tabla 17 se muestran de forma detallada las métricas evaluadas para cada entrenamiento donde también se comparó el uso de los modelos más profundos como ResNet 101 junto a ResNet50. Cabe destacar que fueron entradas con los hiperparámetros que ofrecieron el mejor desempeño.

Comentado [GAF57]: Tilde.

Tabla 17. Métricas de evaluación de Modelos de Segmentación de Instancias

Modelo	IoU	F1	Prec	Recall	AP	AP:50	Clase
Mask RCNN ResNet50	42.37	59.52	58.82	60.24	17.90	34.60	Beta
	16.05	27.66	28.89	26.53	3.60		Alfa
Mask RCNN ResNet101	38.28	55.37	55.68	55.06	21.70	36.10	Beta
	20.62	34.19	31.25	37.74	2.10		Alfa
Cascade Mask RCNN ResNet50	39.81	56.94	66.13	50.00	17.40	35.10	Beta
	16.42	28.21	34.38	23.91	4.10		Alfa
Cascade Mask RCNN ResNet101	37.59	54.64	56.18	53.19	21.70	33.40	Beta
	16.36	28.13	22.78	36.73	2.60		Alfa
Mask Scoring RCNN ResNet50	33.09	49.73	48.94	50.55	18.50	30.50	Beta
	19.35	32.43	30.00	35.29	2.20		Alfa
Mask Scoring RCNN ResNet101	40.57	57.72	65.15	51.81	21.40	34.90	Beta
	15.28	26.51	27.50	25.58	3.30		Alfa

Al examinar la Tabla 17, que muestra las métricas de rendimiento de estos modelos, podemos obtener varias conclusiones. El modelo Mask R-CNN con un Backbone ResNet50 demuestra un rendimiento significativo para la segmentación de la atrofia Beta, mientras que su homólogo con ResNet101 muestra un rendimiento mejorado en términos de puntuación F1, precisión y recall de la atrofia Alfa. En particular, el modelo Cascade Mask R-CNN, independientemente de la Backbone, demostró consistentemente un equilibrio entre la atrofia Beta y Alfa, mientras que el Mask Scoring R-CNN funcionó admirablemente para la atrofia Alfa con una columna vertebral ResNet50 y la atrofia Beta con una columna vertebral ResNet101.

Estas observaciones se pueden entender mejor cuando consideramos las propiedades y los mecanismos inherentes de cada modelo.

El modelo Mask R-CNN emplea un procedimiento sencillo de dos etapas que primero identifica objetos y luego refina estas detecciones con segmentaciones a nivel de píxeles. Las diferencias de rendimiento entre las redes troncales ResNet50 y ResNet101 sugieren que la arquitectura ResNet50 más simple podría ser más adecuada para la atrofia Beta menos compleja, mientras que ResNet101, más profundo, proporciona mejores capacidades de extracción de características para la segmentación de atrofia Alfa más compleja.

Cascade Mask R-CNN funciona al reducir la desalineación entre el puntaje de clasificación y la Intersección sobre unión (IoU) observada en Mask R-CNN estándar. Lo hace empleando una arquitectura en cascada con múltiples etapas de regresión y refinamiento de segmentación. Este mecanismo podría resultar particularmente beneficioso para los límites complejos de la atrofia Alfa.

Por último, Mask Scoring R-CNN mejora al modelo Mask R-CNN de máscaras mediante la introducción de un cabezal de máscara-IoU que predice la calidad de las máscaras de instancia predichas, que luego se utilizan para volver a puntuar las casillas de detección. Esta funcionalidad adicional cierra la brecha entre la calidad de la máscara y los puntajes de la máscara, lo que lleva a una segmentación más precisa. El rendimiento relativamente mejor de Mask Scoring R-CNN con ResNet50 para la atrofia Alfa y ResNet101 para la atrofia Beta podría atribuirse a este mecanismo, lo que permite una mejor delimitación de la atrofia Alfa compleja y límites más precisos en la segmentación de la atrofia Beta.

5.9 Entrenamiento de los mejores modelos

A lo largo de la investigación, la experimentación realizada fue sobre la base de datos ORIGA, debido a que fue la única base de datos en ser validada por expertos glaucomatólogos, lo que garantiza la precisión y la confiabilidad de los datos utilizados para el entrenamiento y las pruebas de los modelos.

Pero en un intento para establecer un modelo robusto frente a las diferentes características de las atrofias peripapilares, se diseñó y ejecutó un experimento integral para desafiar la solidez de estos modelos de segmentación de instancias frente a un conjunto de datos más grande y diverso. Este conjunto de datos, que consta de 1349 imágenes de fondo de ojo, se reunió unificando las tres bases de datos: ORIGA, RETINA y DRISHTI-GS1 extrayendo la región de interés, generando una dimensión de 840 x 840 en todas las imágenes. A pesar de que solo se validó ORIGA, las

Comentado [GAF58]: Validada?

Comentado [GAF59]: Tilde.

Comentado [GAF60]: Ortografía.

Comentado [JA61R60]: 1. adj. Que comprende todos los elementos o aspectos de algo. *Panorámica integral. Educación integral.*

Comentado [GAF62]: Tilde.

bases de datos RETINA y DRISHTI-GS1 se etiquetaron bajo un conocimiento previo en identificación de APP Alfa y Beta.

En este experimento, se aprovecharon los tres modelos de segmentación de instancias: Mask R-CNN, Cascade Mask R-CNN y Mask Scoring R-CNN. Cada modelo estaba respaldado por una Backbone ResNet50 para la extracción de funciones y una configuración de hiperparámetros previamente seleccionados para un mejor desempeño, los cuales se muestran en la Tabla 18. Lo que ofrece un equilibrio adecuado entre eficiencia y rendimiento computacional.

Tabla 18. Métricas de evaluación de Modelos de Segmentación de Instancias

Hiperparámetros	Valores
Peso de Pérdida de Máscara	1
Función de Pérdida	Entropía Cruzada
Función de Regresión	CIoU
Épocas	36
Optimizador	SGD
Backbone	ResNet50
Bbox Generador	Parámetros base

Comentado [GAF63]: Tildes.

Comentado [GAF64]: Tilde.

Pese a que el entrenamiento fue integrado con todos los conjuntos de datos, cada base de datos fue dividida en 70%, 20% y 10%. Para tomar en cuenta para trabajos futuros donde se amplie la validación a todas las bases de datos. Con lo que para destacar el resultado confiable los modelos solo fueron probados para el conjunto de datos prueba de la base de datos ORIGA.

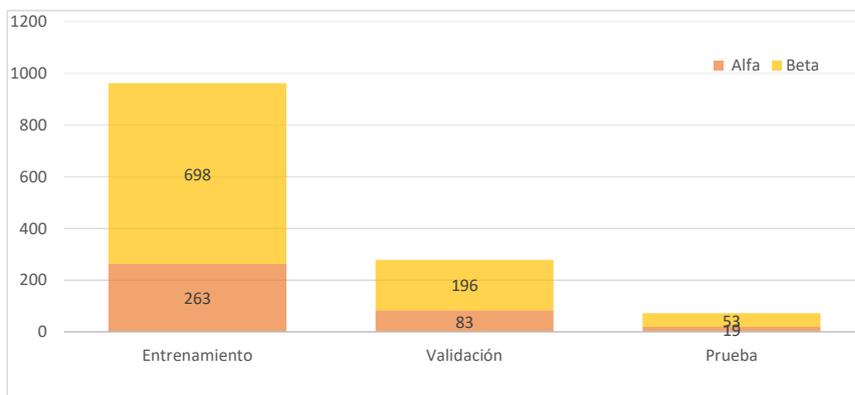


Figura 40. Balance de clases por distribución de subconjuntos en base de datos integral.

En la Tabla 19 se muestra una comparativa de los tres modelos con un Backbone ResNet50 y con un registro de hiperparámetros de mejor desempeño.

Comentado [GAF65]: tilde

Tabla 19. Métricas de evaluación de Modelos de Segmentación de Instancias

Modelo	IoU	F1	Prec	Recall	AP	AP:50	Clase
Mask RCNN ResNet50	36.94	53.95	54.67	53.25	20.60	35.70	Beta
	13.64	24.00	33.33	18.75	3.50		Alfa
Cascade Mask RCNN ResNet50	34.69	51.52	47.66	56.04	20.40	32.00	Beta
	17.48	29.75	25.35	36.00	2.50		Alfa
Mask Scoring RCNN ResNet50	36.28	53.25	54.67	51.90	20.60	32.7	Beta
	15.49	26.83	31.43	23.40	2.70		Alfa

Los resultados del experimento, tal como se resumen en la Tabla 19, revelaron ideas interesantes. En la segmentación de clase Beta, Mask R-CNN y Mask Scoring R-CNN exhibieron un rendimiento casi idéntico, logrando un IoU de 36,94 % y 36,28 % respectivamente. Por otro lado, Cascade Mask R-CNN mostró una capacidad mejorada para identificar la clase Alfa más compleja, registrando el IoU más alto del 17,48 %.

Estos hallazgos sugieren que mientras Mask R-CNN y Mask Scoring R-CNN podrían proporcionar un rendimiento ligeramente superior en clases más comunes como Beta, Cascade Mask R-CNN demuestra una capacidad para manejar mejores clases más complejas, lo que podría atribuirse a su mecanismo único en cascada.

No obstante, al comparar las tablas 17 y 19, podemos observar algunos cambios intrigantes en las métricas de rendimiento de nuestros modelos al incorporar bases de datos no validadas, RETINA y DRISHTI-GS1, junto con la base de datos ORIGA validada.

El resultado más sorprendente es la disminución del rendimiento en todos los modelos cuando se evalúan en el conjunto de datos aumentado, particularmente en términos de Intersección sobre Unión (IoU) y puntaje F1. Por ejemplo, Mask R-CNN con una Backbone ResNet50 experimentó una reducción en la puntuación de IoU para la clase Beta del 42,37 % al 36,94 %. También se puede ver una tendencia similar con los modelos Cascade Mask R-CNN y Mask Scoring R-CNN.

Curiosamente, la clase Alfa en el modelo Mask Scoring R-CNN presentó una caída de rendimiento menor en comparación con los otros modelos, ya que esta se benefició en algunas métricas. Este resultado podría indicar una solidez superior de Mask Scoring R-CNN en el manejo de clases complejas o menos representadas cuando se exponen a datos más diversos.

Por el contrario, la precisión promedio (AP) para la clase Beta ha mejorado ligeramente en los modelos Mask R-CNN y Mask Scoring R-CNN, lo que sugiere que los modelos pueden tener una capacidad ligeramente mejorada para discriminar los verdaderos positivos de los falsos positivos en la clase Beta cuando se entrena en un conjunto de datos más diverso.

Sin embargo, las métricas de rendimiento más bajas, especialmente en IoU y F1, indican que la introducción de datos no validados podría haber introducido ruido en el conjunto de entrenamiento, lo que dificulta el proceso de aprendizaje del modelo y, en última instancia, su rendimiento.

Esto destaca la importancia de utilizar bases de datos de alta calidad validadas por expertos para entrenar modelos de IA en el análisis de imágenes médicas. Aunque aumentar los datos con bases de datos no validadas puede proporcionar un conjunto de datos más grande y posiblemente más diverso, es esencial garantizar la confiabilidad y la coherencia de todos los datos utilizados.

CONCLUSIONES

En esta investigación se ha llevado a cabo un profundo análisis sobre la aplicación de modelos de inteligencia artificial para la clasificación y segmentación de las atrofas peripapilares Alfa y Beta en imágenes de fondo de ojo. Esta tarea es de vital importancia dada la relevancia clínica de estas clases en patologías como el glaucoma y la miopía. Hemos liderado el desarrollo y la validación de la primera base de datos dedicada a la segmentación de la atrofia peripapilar Alfa y Beta, la cual servirá como un invaluable recurso para futuras investigaciones en este campo.

Comentado [GAF66]: Ortografía.

Comentado [GAF67]: Tilde.

La investigación ha revelado desafíos clave en este ámbito, tales como la estructura irregular y difusa de la atrofia peripapilar, junto con el problema del desequilibrio de clases y el tamaño menor de la atrofia Alfa. Estos desafíos resaltan la complejidad asociada a la segmentación de estas clases y subrayan la necesidad de modelos de segmentación robustos y versátiles.

Para afrontar estos retos, hemos empleado avanzados modelos de aprendizaje profundo, incluyendo Mask RCNN, Cascade Mask RCNN y Mask Scoring RCNN, con distintas arquitecturas de Backbone (ResNet50 y ResNet101). Hemos sido pioneros en la incorporación de técnicas de detección de objetos para resolver la tarea de clasificación y segmentación de las atrofas peripapilares. Estos modelos han sido rigurosamente evaluados, considerando múltiples métricas de rendimiento como IoU, F1 Score, precisión, sensibilidad y precisión promedio.

A pesar de que estos modelos son eficaces en ciertos aspectos, no están exentos de limitaciones, especialmente cuando se enfrentan a los desafíos mencionados. Por esta razón, realizamos una extensa exploración de varias técnicas de preprocesamiento, incluyendo el desenfoque, ajuste de contraste, modificación de la exposición y volteo. De todas estas técnicas, la extracción de una región de interés demostró ser la más efectiva para mejorar el rendimiento del modelo. Además, realizamos una optimización exhaustiva de los hiperparámetros del modelo para lograr el mejor ajuste que maximizara el rendimiento, donde destacó la función de CIoU para la generación de las cajas delimitadoras, el uso del optimizador SGD, y la función de pérdida Cross Entropy.

Mientras que numerosos modelos líderes en el campo han logrado resultados notables en términos de rendimiento general, es esencial señalar que ninguno de estos trabajos no establece diferencia entre las atrofas peripapilares Alfa y Beta.

Nuestro estudio, por otro lado, ha logrado avances sustanciales en esta segmentación específica, aportando una contribución única y valiosa al ámbito de la oftalmología y la inteligencia artificial. Esta particularidad no solo resalta la originalidad y relevancia de nuestro enfoque, sino que también sienta las bases para futuras investigaciones especializadas en esta dirección.

Aunque hemos avanzado, la investigación demuestra que el manejo de estructuras irregulares y el desequilibrio de clases en la segmentación de imágenes sigue siendo un desafío. Por lo tanto, los trabajos futuros podrían centrarse en el desarrollo de modelos y técnicas más sólidos para superar los problemas de desequilibrio de clases y el uso de más bases de datos validadas por especialistas. Estos hallazgos subrayan la importancia del ajuste de hiperparámetros y las técnicas de preprocesamiento para mejorar el rendimiento de los modelos de segmentación, así como la posible integración de los modelos Transformers.

En resumen, nuestra investigación representa un avance significativo en la aplicación de la inteligencia artificial para la subclasificación y segmentación de la atrofia peripapilar en imágenes de fondo de ojo. A medida que continuamos refinando estos modelos y profundizando en nuestra comprensión, se espera explorar más en profundidad el camino para el desarrollo de herramientas de asistencia diagnóstica más robustas que puedan ser de gran utilidad en el manejo y diagnóstico del glaucoma.

REFERENCIAS

- [1] P. Riordan-Evan and E. T. Cunningham Jr., “VAUGHAN Y ASBURY Oftalmología general,” 2012.
- [2] C. A. Romo Arpio *et al.*, “Prevalencia de glaucoma primario de ángulo abierto en pacientes mayores de 40 años de edad en un simulacro de campaña diagnóstica,” *Revista Mexicana de Oftalmología*, vol. 91, no. 6, pp. 279–285, Nov. 2017, doi: 10.1016/j.mexoft.2016.08.003.
- [3] Organización Mundial de la Salud, “Informe mundial sobre la visión,” 2020.
- [4] M. Fallon, “Protocolo Meta-analisis Dispositivos de Imagen y Retinografía en Glaucoma”, doi: 10.13140/RG.2.1.1850.8564.
- [5] M. Fingeret, F. A. Medeiros, R. Susanna, and R. N. Weinreb, “Five rules to evaluate the optic disc and retinal nerve fiber layer for glaucoma,” 2005.
- [6] A. Sharma, M. Agrawal, S. Dutta Roy, V. Gupta, P. Vashisht, and T. Sidhu, “Deep learning to diagnose Peripapillary Atrophy in retinal images along with statistical features,” *Biomed Signal Process Control*, vol. 64, Feb. 2021, doi: 10.1016/j.bspc.2020.102254.

- [7] M. K. Song, K. R. Sung, J. W. Shin, J. Kwon, J. Y. Lee, and J. M. Park, “Progressive change in peripapillary atrophy in myopic glaucomatous eyes,” *British Journal of Ophthalmology*, vol. 102, no. 11, pp. 1527–1532, Nov. 2018, doi: 10.1136/bjophthalmol-2017-311152.
- [8] Y. X. Wang, S. Panda-Jonas, and J. B. Jonas, “Optic nerve head anatomy in myopia and glaucoma, including parapapillary zones alpha, beta, gamma and delta: Histology and clinical features,” *Progress in Retinal and Eye Research*, vol. 83. Elsevier Ltd, Jul. 01, 2021. doi: 10.1016/j.preteyeres.2020.100933.
- [9] Hector Darío Forero Angel, Julio Cesar Bernal Serna, and Angela María Garcés Valencia, “Optic Nerve and Peripapillary Retina Characteristics in Primary Open Angle Glaucoma,” 2015.
- [10] Y. Chai, H. Liu, and J. Xu, “A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images,” *Applied Soft Computing Journal*, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105890.
- [11] “Tomografía de coherencia óptica OCT Medellín Envigado Sabaneta.”.
- [12] J. Mccarthy, “WHAT IS ARTIFICIAL INTELLIGENCE?,” 2007. [Online]. Available: <http://www-formal.stanford.edu/jmc/>
- [13] M. A. Al-Antari, “Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology!,” *Diagnostics*, vol. 13, no. 4. MDPI, Feb. 01, 2023. doi: 10.3390/diagnostics13040688.
- [14] A. Grzybowski, *Artificial Intelligence in Ophthalmology*, 1st ed., vol. 1. Springer Cham, 2021.
- [15] R. Tolšius, O. Kurasova, and J. Bernataviciene, “Semantic segmentation of eye fundus images using convolutional neural networks,” *Informacijos Mokslai*, vol. 85, 2019, doi: 10.15388/IM.2019.85.20.
- [16] U. Schmidt-Erfurth, A. Sadeghipour, B. S. Gerendas, S. M. Waldstein, and H. Bogunović, “Artificial intelligence in retina,” *Progress in Retinal and Eye Research*, vol. 67. Elsevier Ltd, pp. 1–29, Nov. 01, 2018. doi: 10.1016/j.preteyeres.2018.07.004.
- [17] J. M. Molina Casado, “Detección y Localización Automática de Estructuras Anatómicas en Imágenes de Retina utilizando Técnicas de Visión Artificial.”
- [18] D. Y. Tratamiento and E. Recomendaciones, *GUÍA DE PRÁCTICA CLÍNICA GPC DE GLAUCOMA PRIMARIO DE ANGULO ABIERTO*. [Online]. Available: <http://www.cenetec.salud.gob.mx/contenidos/gpc/catalogoMaestroGPC.html>
- [19] A. Sharma, M. Agrawal, S. Dutta Roy, V. Gupta, P. Vashisht, and T. Sidhu, “Deep learning to diagnose Peripapillary Atrophy in retinal images along with statistical

- features,” *Biomed Signal Process Control*, vol. 64, Feb. 2021, doi: 10.1016/j.bspc.2020.102254.
- [20] L. Cheng-Kai, T. Tong Boon, and F. M. Alan, “Automatic Parapapillary Atrophy Shape Detection and Quantification in Colour Fundus Images,” 2010.
- [21] Muramatsu *et al.*, *Computerized Detection of Peripapillary Chorioretinal Atrophy by Texture Analysis*. 2011. doi: 10.0/Linux-x86_64.
- [22] J. Cheng *et al.*, “Peripapillary atrophy detection by sparse biologically inspired feature manifold,” *IEEE Trans Med Imaging*, vol. 31, no. 12, pp. 2355–2365, 2012, doi: 10.1109/TMI.2012.2218118.
- [23] A. Septiarini, R. Pulungan, A. Harjoko, and R. Ekantini, “Peripapillary Atrophy Detection in Fundus Images Based on Sectors with Scan Lines Approach.”
- [24] A. Septiarini, D. M. Khairina, A. H. Kridalaksana, and H. Hamdani, “Automatic glaucoma detection method applying a statistical approach to fundus images,” *Healthc Inform Res*, vol. 24, no. 1, pp. 53–60, Jan. 2018, doi: 10.4258/hir.2018.24.1.53.
- [25] L. Hanxiang, K. Jieliang, F. Yunlong, X. Jie, and L. Huiqi, “Automatic segmentation of PPA in retinal images,” 2018.
- [26] STMIK AKAKOM Yogyakarta, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, *2nd ISRITI 2019 proceeding : the 2nd International Seminar on Research of Information Technology and Intelligent Systems 2019 : “The future & challenges of extended intelligence” : Yogyakarta, Indonesia, 05-06 December 2019*.
- [27] Y. Chai, H. Liu, and J. Xu, “A new convolutional neural network model for peripapillary atrophy area segmentation from retinal fundus images,” *Applied Soft Computing Journal*, vol. 86, Jan. 2020, doi: 10.1016/j.asoc.2019.105890.
- [28] F. Z. Zulfira and S. Suyanto, “Detection of Multi-Class Glaucoma Using Active Contour Snakes and Support Vector Machine,” in *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 650–654. doi: 10.1109/ISRITI51436.2020.9315372.
- [29] F. Z. Zulfira, S. Suyanto, and A. Septiarini, “Segmentation technique and dynamic ensemble selection to enhance glaucoma severity detection,” *Comput Biol Med*, vol. 139, Dec. 2021, doi: 10.1016/j.compbiomed.2021.104951.
- [30] M. Li, H. Zhao, J. Xu, and H. Li, “Peripapillary Atrophy Segmentation Based on ASM Loss,” in *Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society*, 2022. doi: 10.1109/ISBI52829.2022.9761687.

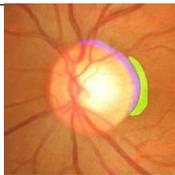
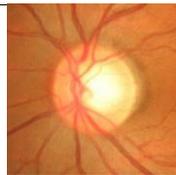
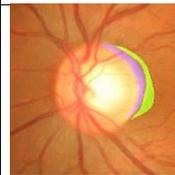
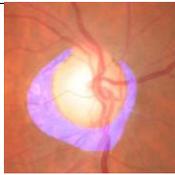
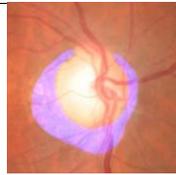
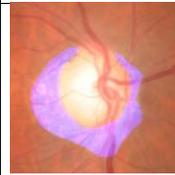
- [31] A. Almansour *et al.*, “Peripapillary atrophy classification using CNN deep learning for glaucoma screening,” *PLoS One*, vol. 17, no. 10 October, Oct. 2022, doi: 10.1371/journal.pone.0275446.
- [32] A. Ahmed Gasm Elseid and A. Osman Mohammed Hamza, “Computer-Aided Glaucoma Diagnosis System.”
- [33] R. Chrastek, H. Niemann, L. Kubecka, J. Jan, V. Derhartunian, and G. Michelson, “Optic nerve head segmentation in multimodal retinal images,” in *Medical Imaging 2005: Image Processing*, SPIE, Apr. 2005, p. 1604. doi: 10.1117/12.594492.
- [34] Z. Zhang *et al.*, “ORIGA-light: An online retinal fundus image database for glaucoma analysis and research,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC’10*, IEEE Computer Society, 2010, pp. 3065–3068. doi: 10.1109/IEMBS.2010.5626137.
- [35] J. B. Jonas, “Clinical implications of peripapillary atrophy in glaucoma.”
- [36] B. Mehlig, “Machine learning with neural networks,” Jan. 2019, doi: 10.1017/9781108860604.
- [37] S. S. Haykin and S. S. Haykin, *Neural networks and learning machines*. Prentice Hall/Pearson, 2009.
- [38] M. Nielsen, “Neural Networks and Deep Learning.” [Online]. Available: <http://neuralnetworksanddeeplearning.com>
- [39] I. El Naqa and M. J. Murphy, “What Is Machine Learning?,” in *Machine Learning in Radiation Oncology*, Springer International Publishing, 2015, pp. 3–11. doi: 10.1007/978-3-319-18305-3_1.
- [40] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [41] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” Nov. 2015, [Online]. Available: <http://arxiv.org/abs/1511.08458>
- [42] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” *Artif Intell Rev*, vol. 52, no. 2, pp. 1089–1106, Aug. 2019, doi: 10.1007/s10462-018-9641-3.
- [43] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” Nov. 2014, [Online]. Available: <http://arxiv.org/abs/1411.4038>
- [44] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” Nov. 2015, [Online]. Available: <http://arxiv.org/abs/1511.00561>
- [45] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>

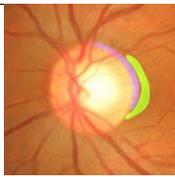
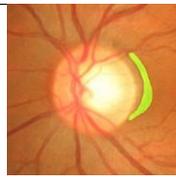
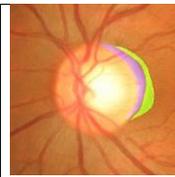
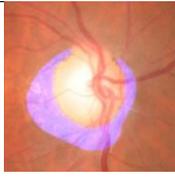
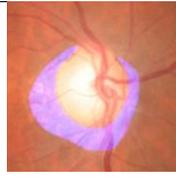
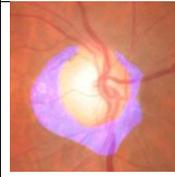
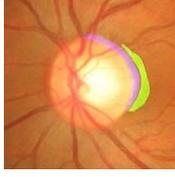
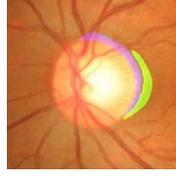
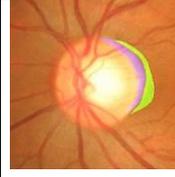
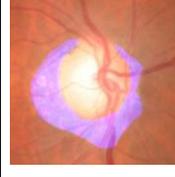
- [46] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” May 2015, [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [47] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Dec. 2015, [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [48] A. G. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” Apr. 2017, [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [49] H. Zhang, H. Sun, W. Ao, and G. Dimirovski, “A survey on instance segmentation: Recent advances and challenges,” *International Journal of Innovative Computing, Information and Control*, vol. 17, no. 3, pp. 1041–1053, 2021, doi: 10.24507/ijicic.17.03.1041.
- [50] X. Zhang, Y. H. Yang, Z. Han, H. Wang, and C. Gao, “Object class detection: A survey,” *ACM Computing Surveys*, vol. 46, no. 1. Oct. 2013. doi: 10.1145/2522968.2522978.
- [51] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” Nov. 2013, [Online]. Available: <http://arxiv.org/abs/1311.2524>
- [52] R. Girshick, “Fast R-CNN,” Apr. 2015, [Online]. Available: <http://arxiv.org/abs/1504.08083>
- [53] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Jun. 2015, [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [54] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” 2017.
- [55] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, “Mask Scoring R-CNN,” Mar. 2019, [Online]. Available: <http://arxiv.org/abs/1903.00241>
- [56] V. Yeghiazaryan and I. Voiculescu, “Family of boundary overlap metrics for the evaluation of medical image segmentation,” *Journal of Medical Imaging*, vol. 5, no. 01, p. 1, Feb. 2018, doi: 10.1117/1.jmi.5.1.015006.
- [57] Google Research, “Google Colaboratory,” 2017.
- [58] Guido van Rossum, “Python Software Foundation,” 2001.
- [59] K. Chen *et al.*, “MMDetection: Open MMLab Detection Toolbox and Benchmark,” Jun. 2019, [Online]. Available: <http://arxiv.org/abs/1906.07155>
- [60] Facebook AI Research, “PyTorch.” 2016.
- [61] François Chollet, “Keras_ Deep Learning for humans.” 2015.
- [62] B. Dwyer, J. Nelson, and J. Solawetz, “Roboflow,” 2022.
- [63] “retina_dataset_dataset at master · yiweichen04_retina_dataset · GitHub”.

- [64] by Jayanthi Sivaswamy *et al.*, “Drishit-GS: Retinal Image Dataset for Optic Nerve Head(ONH) Segmentation DRISHTI-GS: RETINAL IMAGE DATASET FOR OPTIC NERVE HEAD(ONH) SEGMENTATION,” 2014.
- [65] X. Zhao, J. Xiao, B. Zhang, Q. Zhang, and A.-N. Waleed, “Journal Logo Weight-Guided Loss for Long-Tailed Object Detection and Instance Segmentation,” 2023, [Online]. Available: <https://ssrn.com/abstract=4054228>
- [66] Z. Zhang and M. R. Sabuncu, “Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels,” May 2018, [Online]. Available: <http://arxiv.org/abs/1805.07836>
- [67] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” Aug. 2017, [Online]. Available: <http://arxiv.org/abs/1708.02002>
- [68] J. He, S. Erfani, X. Ma, J. Bailey, Y. Chi, and X.-S. Hua, “Alpha-IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression.”
- [69] Amit Shekhar, “What Are L1 and L2 Loss Functions?,” Aug. 01, 2019. <https://amitshekhar.me/blog/l1-and-l2-loss-functions> (accessed Jun. 15, 2023).
- [70] Z. H. Feng, J. Kittler, M. Awais, P. Huber, and X. J. Wu, “Wing Loss for Robust Facial Landmark Localisation with Convolutional Neural Networks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2018, pp. 2235–2245. doi: 10.1109/CVPR.2018.00238.
- [71] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression.”
- [72] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” 2016. [Online]. Available: <https://github.com/Zzh-tju/DIoU>.
- [73] Z. Zheng *et al.*, “Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation,” May 2020, [Online]. Available: <http://arxiv.org/abs/2005.03572>
- [74] S. Ruder, “An overview of gradient descent optimization algorithms,” Sep. 2016, [Online]. Available: <http://arxiv.org/abs/1609.04747>
- [75] J. Duchi JDUCHI and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization * Elad Hazan,” 2011.
- [76] Y. N. Dauphin, J. Chung, and Y. Bengio BENGIOY, “RMSProp and equilibrated adaptive learning rates for non-convex optimization SpeechBrain View project Oracle Performance for Visual Captioning View project Harm de Vries Université de Montréal RMSProp and

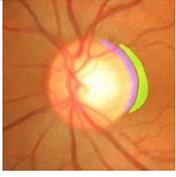
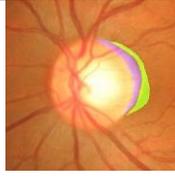
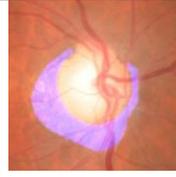
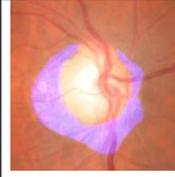
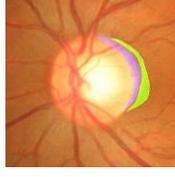
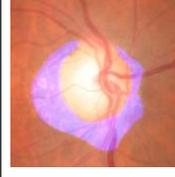
- equilibrated adaptive learning rates for non-convex optimization Harm de Vries,” 2015. [Online]. Available: <https://www.researchgate.net/publication/272423025>
- [77] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Dec. 2014, [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [78] “Selecting the Right Bounding Box Using Non-Max Suppression (with implementation).” [Online]. Available: <https://pjreddie.com/darknet/yolov1/>
- [79] G. Divam, “Image-segmentation-keras_ Implementation of Segnet, FCN, UNet , PSPNet and other models in Keras.” <https://github.com/divamgupta/image-segmentation-keras> (accessed Sep. 17, 2022).
- [80] “RoboFlow.”
- [81] D. Berrar, “Cross-validation,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, Elsevier, 2018, pp. 542–545. doi: 10.1016/B978-0-12-809633-8.20349-X.
- [82] W. Zhang, C. Witharana, A. K. Liljedahl, and M. Kanevskiy, “Deep convolutional neural networks for automated characterization of arctic ice-wedge polygons in very high spatial resolution aerial imagery,” *Remote Sens (Basel)*, vol. 10, no. 9, Sep. 2018, doi: 10.3390/rs10091487.
- [83] D. Bolya, S. Foley, J. Hays, and J. Hoffman, “TIDE: A General Toolbox for Identifying Object Detection Errors,” Aug. 2020, [Online]. Available: <http://arxiv.org/abs/2008.08115>
- [84] G. Alfonso-Francia *et al.*, “Performance Evaluation of Different Object Detection Models for the Segmentation of Optical Cups and Discs,” *Diagnostics*, vol. 12, no. 12, Dec. 2022, doi: 10.3390/diagnostics12123031.

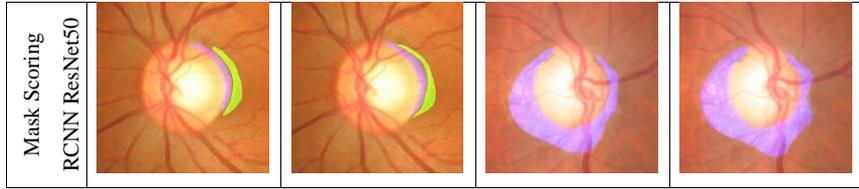
ANEXOS

	ResNet50	ResNet101	Ground Truth
Mask RCNN			
			

Mask Cascade RCNN			
			
Mask Scoring RCNN			
			

Anexo 5. Tabla comparativa de predicciones vs Ground truth de Análisis de Modelos.

	Predicción	Ground Truth	Predicción	Ground Truth
Mask RCNN ResNet50				
Cascade RCNN ResNet50				



Anexo 6. Tabla comparativa de predicciones vs Ground Truth de los Mejores Modelos.