

Universidad Autónoma de Baja California

Facultad de Ingeniería, Arquitectura y Diseño



Maestría y Doctorado en Ciencias e Ingeniería

Segmentación de estructuras anatómicas en imágenes de fundoscopia digital con herramientas de procesamiento digital y aprendizaje automático

Tesis

Presentada para cumplir parcialmente con los requisitos necesarios para obtener el grado de

MAESTRÍA EN INGENIERÍA

Por:

Laura Johana Gonzalez Zazueta

Ensenada, Baja California, Junio 2025.

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO

MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

Segmentación de estructuras anatómicas en imágenes de fundoscopia digital con herramientas de procesamiento digital y aprendizaje automático

TESIS

Que para obtener el grado de Maestría en Ingeniería presenta:

Laura Johana Gonzalez Zazueta

Aprobada por:



Dr. Juan Iván Nieto Hipólito
Director de tesis



Dr. Gener José Avilés Rodríguez
Codirector de tesis



Dr. Christian Xavier Navarro Cota
Miembro del comité



Dra. Mabel Vázquez Briseño
Miembro del comité



Dr. Jason Penniecook Sawyers
Miembro del comité

Ensenada Baja California, México, 16 de junio del 2025

Resumen de la tesis de Laura Johana González Zazueta, presentada para cumplir parcialmente con los requisitos necesarios para obtener el grado de MAESTRA EN INGENIERÍA, del programa de Maestría y Doctorado en Ciencias e Ingeniería (MYDCI) de la UABC. Ensenada, Baja California, México, junio de 2025.

Resumen aprobado por:



Dr. Juan Iván Nieto Hipólito
Director de Tesis

En este trabajo se aborda la segmentación de estructuras anatómicas en imágenes de fundoscopia digital mediante la integración de técnicas de procesamiento digital de imágenes y algoritmos de aprendizaje automático. El objetivo es apoyar la detección temprana y la toma de decisiones clínicas en contextos con acceso limitado a servicios oftalmológicos. La metodología enfatiza la selección y evaluación de algoritmos de segmentación que sean tanto clínicamente relevantes como computacionalmente eficientes. Este enfoque surge de una colaboración transdisciplinaria entre las ciencias computacionales, las tecnologías de salud digital y la oftalmología clínica. Los resultados contribuyen al campo del Diagnóstico Asistido por Computadora y se alinean con los principios de la investigación traslacional, ofreciendo una estrategia escalable y adaptable para el análisis de imágenes médicas en entornos reales de atención sanitaria.

Palabras clave: *Estructuras anatómicas, Imágenes de fundoscopia, Procesamiento digital de imágenes, Aprendizaje automático, Técnicas de segmentación, Oftalmología, imágenes médicas.*

Dedicatoria

A Betsaida Covarrubias, mi entrañable amiga y confidente, por acompañarme con lealtad, sabiduría y generosidad en cada etapa de este camino. Gracias por tu escucha atenta, tus palabras sinceras y por ser un pilar firme en los momentos de incertidumbre. Tu amistad ha sido un regalo invaluable, este logro también es tuyo.

A Antonio Esquer, mi prometido, por tu amor incondicional, tu paciencia infinita y por caminar a mi lado, aún en los días más complejos. Gracias por recordarme lo lejos que puedo llegar, por celebrar cada pequeño logro como si fuera propio y por sostenerme en cada tropiezo sin soltar mi mano.

A mi familia, por ser mi primer hogar, mi mayor ejemplo y mi fuente constante de fuerza. Su amor, sus enseñanzas y su confianza en mí han sido la base sobre la cual he podido construir este sueño. Nada de esto habría sido posible sin ustedes.

Agradecimientos

Expreso mi más profundo y sincero agradecimiento a mis directores de tesis, el Dr. Juan Iván Nieto Hipólito y el Dr. Gener José Avilés Rodríguez. Su guía académica, su compromiso constante y su generosidad intelectual han sido pilares fundamentales a lo largo de este proceso. Gracias por creer en mí, por impulsarme a superar mis propios límites y por brindarme siempre una mirada crítica, constructiva y humana. Más allá del conocimiento, me han enseñado con el ejemplo el valor de la ética, la disciplina y la pasión por la ciencia.

Agradezco sinceramente a los integrantes de mi comité de tesis, por sus valiosas observaciones, su tiempo y el compromiso con la mejora continua de este trabajo. Cada retroalimentación ha representado una oportunidad para crecer y aprender.

Al Programa de Maestría y Doctorado en Ciencias e Ingeniería, por brindarme un entorno de formación académica riguroso, interdisciplinario y estimulante. Este espacio ha sido clave en mi desarrollo como científica.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti), por el respaldo financiero y su firme apuesta por el fortalecimiento de las vocaciones científicas. Su apoyo hizo posible esta etapa de formación.

Finalmente, dedico un agradecimiento muy especial a todas las mujeres que, con su ejemplo, sus enseñanzas y su presencia, han contribuido a mi formación en la ciencia. A quienes abrieron camino antes que yo, a quienes caminan a mi lado y a quienes vienen detrás: gracias por demostrar que la ciencia también se escribe en femenino.

Tabla de contenidos

1. Introducción	9
1.2 Planteamiento del problema	10
1.3 Pregunta de investigación	10
1.4 Hipótesis	10
1.5 Objetivo general	11
1.6 Objetivos específicos	11
1.7 Estructura de la tesis	11
2. Revisión de la literatura	13
2.1 Fundamentos físicos y anatómicos	13
2.1.1 Principios ópticos aplicados en fundoscopia	13
2.1.2 Estructuras clave del ojo	14
2.2 Historia y evolución tecnológica	16
2.2.1 Orígenes de la fundoscopia	16
2.2.2 Evolución hacia la fotografía de fondo de ojo	17
2.2.3 Desarrollo de la cámara de fondo de ojo moderna	18
2.3 Aplicaciones clínicas y computacionales actuales	19
2.3.1 Avances tecnológicos en la evaluación de la calidad de imágenes de fondo de ojo	19
2.3.2 Importancia clínica de la fundoscopia	20
2.3.3 El salto de la oftalmología análoga a la oftalmología digital	20
2.3.4 Identificación de estructuras anatómicas en fundoscopia digital en el proceso clínico.	21
2.3.5 Algoritmos de segmentación en imágenes de fondo de ojo con técnicas de cómputo inteligente	22
3. Desarrollo de la estrategia de investigación	23
3.1 Metodología para la validación	23
3.2 Marco de Conceptos Generales (General Concepts Framework)	24
3.2.1 Diagnóstico asistido por computadora	24
3.2.2 Fundoscopia digital	25
3.2.3 Segmentación de imágenes	25
3.2.4 Inteligencia artificial	25
3.3 Marco de Modelado Integrado (Integrating Modeling Framework)	25
3.3.1 Análisis de imágenes médicas	26
3.3.2 Modelos de Aprendizaje automático	26
3.3.3 Modelos de Aprendizaje Profundo	27
3.3.4 Segmentación de estructuras anatómicas	27
3.4 Marco Computacional (Computational Framework)	27
3.4.1 Exploración de algoritmos	27
3.4.2 Recopilación y Preparación de Datos	28
3.4.3 Preprocesamiento	28
3.4.4 Modelos de Aprendizaje Profundo y aprendizaje automático	28
3.4.5 Segmentación de estructuras anatómicas	29

3.4.6 Selección de algoritmos	29
3.4.7 Optimización de Hiperparámetros	29
3.4.8 Evaluación de algoritmos	30
3.4.9 Comparación de Algoritmos	30
3.4.10 Comparación de resultados para su uso clínico	31
4. Resultados	32
4.1 Revisión Sistemática de la Literatura	32
4.1.1 Metodología de búsqueda y selección (PRISMA)	32
4.2 Aprendizaje supervisado	36
4.2.1 Redes neuronales convolucionales (CNN)	37
4.2.2 Redes completamente convolucionales (FCN)	43
4.2.3 U-Net	46
4.2.4 Otras arquitecturas	52
4.2.5 Modelos basados en aprendizaje automático	57
4.2.6 Procesamiento digital de imágenes	60
4.3 Modelos basados en aprendizaje no supervisado y semi-supervisado	62
4.4 Recolección y análisis de bases de datos	65
4.4.1 Descripción de las bases de datos utilizadas	66
4.4.2 Justificación de la base de datos seleccionada.	69
4.4.3 APTOS 2019 Blindless detection	70
4.5 Clasificación y comparación de algoritmos existentes	71
4.5.1 Tabla comparativa de algoritmos encontrados en la revisión del estado del arte.	71
4.5.2 Descripción general del dataset y características visuales iniciales	73
4.6 Análisis visual preliminar de las imágenes	73
4.7 Evaluación automatizada de la calidad de imagen mediante análisis computacional	76
4.8 Mejoras preliminares de contraste	80
4.8.1 Mejora de contraste mediante CLAHE	80
4.8.2 Análisis de canales RGB con mejora de contraste mediante CLAHE	82
4.9 Aplicación del algoritmo Canny para la detección de estructuras anatómicas	83
4.9.1 Combinación de CLAHE y Detección de bordes mediante el algoritmo de Canny	85
4.10 Preprocesamiento avanzado en el canal verde, extracción de bordes y detección de estructuras	87
4.10.1 Preprocesamiento secuencial sobre el canal verde para realce y detección de estructuras anatómicas	87
4.10.2 Preprocesamiento avanzado: umbralización adaptativa, apertura morfológica y detección de bordes en canal verde	89
4.11 Evaluación de algoritmos seleccionados	91
4.12 Implementación de los modelos seleccionados.	91
4.12.1 U-Net	91
4.12.2 Metodología del algoritmo U-Net	93
4.12.3 Resultados con U-Net	96
4.12.4 nnUnet	98
4.12.5 Metodología del algoritmo nnU-Net	101

4.12.6 Resultados con nnU-Net	104
4.12.7 Segment Anything Model (SAM)	106
4.12.8 Metodología del algoritmo SAM	108
4.12.9 Resultados con algoritmo SAM	112
4.13 Evaluación de desempeño	115
4.13.1 Resultados de U-Net	115
4.13.2 Resultados con UNet (Imagen original y predicción)	116
4.13.3 Resultados de nnU-Net	118
4.13.4 Resultados con nnU-Net (Imagen original y predicción)	119
4.13.5 Resultados de SAM	122
4.13.6 Resultados con SAM (Imagen original y predicción)	123
4.14 Propuesta de sistema de segmentación aplicable al contexto regional	124
5. Discusión y conclusiones	126
5.1 Discusión	126
5.2 Conclusiones	128
5.3 Trabajo futuro	129
5.4 Producción académica	130
6. Bibliografía	133

Lista de Figuras

Figura 2.1 Principio de Gullstrand.	15
Figura 2.2 Esquema detallado del ojo humano	16
Figura 2.3 Estructuras anatómicas relevantes	17
Figura 2.4 Oftalmoscopio original de Hermann von Helmholtz	18
Figura 2.5 Sistema de iluminación de la cámara de fondo de ojo Zeiss	19
Figura 3.1 Esquema general de la metodología DIMA	25
Figura 3.2. Diagrama general del flujo metodológico de la tesis	32
Figura 4.1 Diagrama de flujo PRISMA.	36
Figura 4.2 Artículos seleccionados en el estado del arte	37
Figura 4.3. Mapa conceptual de redes neuronales convolucionales (CNN) aplicadas a la segmentación de imágenes de fondo de ojo.	38
Figura 4.4. Mapa conceptual de redes totalmente convolucionales (FCN) aplicadas a la segmentación de imágenes de fondo de ojo.	45
Figura 4.5 Mapa conceptual de U-Nets aplicadas a la segmentación de imágenes de fondo de ojo.	48
Figura 4.6 Mapa conceptual de modelos con otras arquitecturas aplicadas a la segmentación de imágenes de fondo de ojo.	54
Figura 4.7 Mapa conceptual de modelos de aprendizaje automático aplicados a la segmentación de imágenes de fondo de ojo.	59
Figura 4.8 Mapa conceptual de procesamiento digital de imágenes aplicado a la segmentación de imágenes de fondo de ojo.	62
Figura 4.9 Mapa conceptual de modelos de aprendizaje no supervisado y semi-supervisado aplicadas a la segmentación de imágenes de fondo de ojo.	65
Figura 4.10 Imágenes representativas de las bases de datos más utilizadas en estudios de segmentación de imágenes de fondo de ojo	70
Figura 4.11 Imágenes representativas de las cinco clases de severidad de retinopatía diabética presentes en la base de datos	72
Figura 4.12 Ejemplos de imágenes de fondo de ojo extraídas de la base de datos APTOS.	73
Figura 4.13 Imagen de fondo de ojo original utilizada para mostrar el proceso del análisis preliminar.	76
Figura 4.14 Conversión de una imagen de fondo de ojo a escala de grises.	76
Figura 4.15 Histograma de niveles de gris correspondiente a la imagen original.	77
Figura 4.16 Máscara binaria aplicada a la imagen en escala de grises, resaltando zonas de intensidad.	77

Figura 4.17 Se muestra el análisis visual de una imagen de fondo de ojo	78
Figura 4.18 Visualización generada por la aplicación desarrollada en Streamlit.	79
Figura 4.19 Ejemplo de imagen clasificada como “usable”	80
Figura 4.20 Ejemplo de imagen clasificada como “oscura”	80
Figura 4.21 Ejemplo de imagen clasificada como “velada”	81
Figura 4.22 Estructura de regiones en una imagen aplicando CLAHE	82
Figura 4.23 Ejemplos de mejora de contraste mediante CLAHE en imágenes de fondo de ojo.	84
Figura 4.24 Comparación visual del efecto de CLAHE aplicado individualmente a los canales rojo, verde y azul en imágenes de fondo de ojo.	85
Figura 4.25 Ejemplo de la aplicación del algoritmo Canny	86
Figura 4.26 Aplicación del algoritmo Canny sobre imágenes de fondo de ojo	87
Figura 4.27 Flujo de trabajo de preprocesamiento combinado: mejora de contraste y detección de bordes.	88
Figura 4.28 Ejemplo representativo de dos imágenes del flujo de preprocesamiento avanzado aplicado sobre el canal verde	90
Figura 4.29 Ejemplo representativo de dos imágenes del flujo de preprocesamiento avanzado aplicado sobre el canal verde.	91
Figura 4.30 Comparación entre la imagen original de fondo de ojo y el resultado final después de un preprocesamiento compuesto	92
Figura 4.31 Arquitectura de U-Net	94
Figura 4.32 Imagen representativa del proceso de anotaciones	95
Figura 4.33 Imagen de fondo de ojo y su correspondiente máscara de segmentación de arcada vascular	96
Figura 4.34 Resultado de segmentación vascular obtenido mediante U-Net.	98
Figura 4.35 Resultado de segmentación vascular obtenido mediante U-Net.	99
Figura 4.36 Resultado de segmentación vascular obtenido mediante U-Net.	99
Figura 4.37 Resultado de segmentación vascular obtenido mediante U-Net.	99
Figura 4.38 Diagrama de flujo del modelo nnU-Net	102
Figura 4.39 Herramienta de Anotación para Visión por Computadora con el conjunto de datos	103
Figura 4.40 Muestra de máscaras manuales de las estructuras anatómicas	104
Figura 4.41 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net.	106
Figura 4.42 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net.	107
Figura 4.43 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo	107

nnU-Net.	
Figura 4.44 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net.	107
Figura 4.45 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net.	108
Figura 4.46 Representación del flujo de trabajo del modelo Segment Anything (SAM).	109
Figura 4.47 Imagen de fondo de ojo en una matriz tridimensional en formato RGB utilizada como entrada para el modelo SAM.	110
Figura 4.48 Imagen de fondo de ojo con máscara de la arcada vascular.	111
Figura 4.49 Visualización de imagen con caja delimitadora y máscara	112
Figura 4.50 Resultados de segmentación obtenidos con el modelo SAM del entrenamiento.	114
Figura 4.51 Resultados de segmentación obtenidos con el modelo SAM del entrenamiento.	115
Figura 4.52 Resultados de segmentación obtenidos con el modelo SAM del entrenamiento.	115
Figura 4.53 Resultados de segmentación obtenidos con el modelo SAM del entrenamiento.	115
Figura 4.54 Resultados de segmentación obtenidos con el modelo SAM del entrenamiento.	115
Figura 4.55 Resultados de segmentación obtenidos con el modelo SAM de la inferencia.	116
Figura 4.56 Resultados de segmentación obtenidos con el modelo SAM de la inferencia.	116
Figura 4.57 Resultados de segmentación obtenidos con el modelo SAM de la inferencia.	116
Figura 4.58 Resultados de segmentación obtenidos con el modelo SAM de la inferencia.	117
Figura 4.59 Resultados de segmentación obtenidos con el modelo SAM de la inferencia.	117
Figura 4.60 Resultado de segmentación obtenido con el modelo UNet.	119
Figura 4.61 Resultado de segmentación obtenido con el modelo UNet.	119
Figura 4.62 Resultado de segmentación obtenido con el modelo UNet.	120
Figura 4.63 Resultado de segmentación obtenido con el modelo nnU-Net.	121
Figura 4.64 Resultado de segmentación obtenido con el modelo nnU-Net.	122
Figura 4.65 Resultado de segmentación obtenido con el modelo nnU-Net.	122
Figura 4.66 Resultado de segmentación obtenido con el modelo SAM.	125
Figura 4.67 Resultado de segmentación obtenido con el modelo SAM.	126
Figura 4.68 Resultado de segmentación obtenido con el modelo SAM.	126
Figura 5.1 Evidencia de la capacitación en fundamentos generales de oftalmología	128
Figura 5.2 Participación en el Seminario de Inteligencia Artificial y Cómputo Cuántico	129
Figura 5.3 Constancia de ponencia en modalidad cartel en el 3er. Coloquio de Aplicaciones de la Inteligencia Artificial en Ciencia, CICESE.	129
Figura 5.4 Constancia de participación como ponente en modalidad cartel en el 2do. Coloquio de Informática Médica: Innovación y Actualización en Salud y Tecnología	130

Lista de Tablas

Tabla 4.1 La tabla muestra para cadenas de Búsqueda por Base de Datos	33
Tabla 4.2 Muestra los aspectos técnicos de los modelos del tipo de CNN	42
Tabla 4.3 Sintetiza la información recabada de los modelos CNN empleados	43
Tabla 4.4 Muestra los aspectos técnicos de los modelos del tipo de FCN	46
Tabla 4.5 Sintetiza la información recabada de los modelos FCN empleados	47
Tabla 4.6. Muestra los aspectos técnicos de los modelos del tipo de U-Net	51
Tabla 4.7 Sintetiza la información recabada de los modelos U-Net empleados	52
Tabla 4.8 Muestra los aspectos técnicos de los modelos de diversas arquitecturas	56
Tabla 4.9 Sintetiza la información recabada de los modelos de diversas arquitecturas empleados	57
Tabla 4.10 Muestra los aspectos técnicos de los modelos de aprendizaje automático	60
Tabla 4.11 Sintetiza la información recabada de los modelos de aprendizaje automático empleados	60
Tabla 4.12 Muestra los aspectos técnicos de los modelos de procesamiento digital de imágenes	63
Tabla 4.13 Sintetiza la información recabada de los modelos de procesamiento digital de imágenes empleados	63
Tabla 4.14 Muestra los aspectos técnicos de los modelos de aprendizaje no supervisado y aprendizaje semi-supervisado	66
Tabla 4.15 Sintetiza la información recabada de los modelos modelos de aprendizaje no supervisado y aprendizaje semi-supervisado empleados	66
Tabla 4.16 Relación de autores, modelos propuestos y bases de datos utilizadas para la segmentación de estructuras anatómicas en imágenes de fondo de ojo.	67
Tabla 4.17 Frecuencia de uso de las bases de datos en los estudios revisados.	69
Tabla 4.18 Comparación de algoritmos encontrados en la revisión del estado del arte para la segmentación de estructuras anatómicas en imágenes de fundoscopia digital.	72
Tabla 4.19 La tabla de las métricas en Streamlit	80
Tabla 4.20 Requerimientos para U-Net	96
Tabla 4.21 Tabla de métricas U-Net	117
Tabla 4.22 Tabla de métricas nnU-Net generales del algoritmo	120
Tabla 4.23 Tabla de métricas nnU-Net en arcada vascular	122
Tabla 4.24 Tabla de métricas nnU-Net en nervio óptico	123
Tabla 4.25 Tabla de métricas de SAM	123
Tabla 4.26 Resumen las características de los tres modelos evaluados	126

1. Introducción

La segmentación automática de estructuras anatómicas en las imágenes de fondo de ojo se ha consolidado como una herramienta que ha sido clave para agilizar y apoyar los procesos diagnósticos en oftalmología, sobre todo en entornos donde el acceso a especialistas es limitado, como lo son las Regiones marginadas o rurales, en este caso, como muchas del estado de Baja California, enfrentan barreras importantes para acceder a atención oftalmológica oportuna, lo que retrasa la detección de enfermedades visuales tratables [1].

Ante este panorama, la integración de técnicas de procesamiento digital de imágenes y aprendizaje automático permite identificar de forma precisa regiones clave como el disco óptico, la mácula y la arcada vascular, contribuyendo al triaje automatizado y a la priorización de pacientes en riesgo. Esta tesis se enfoca en la selección y evaluación de algoritmos capaces de segmentar dichas estructuras de forma robusta, eficiente y clínicamente relevante, con el objetivo de fortalecer el trabajo del personal médico generalista y ampliar el alcance de la salud visual en contextos donde la oftalmología no está presente.

1.1 Motivación

Este proyecto surge como una respuesta a la necesidad de aportar a la mejora continua en el tema del análisis automatizado de imágenes oftalmológicas de fondo de ojo, en particular para situaciones donde el acceso a servicios especializados es limitado o restringido. El uso de imágenes de fundoscopia digital tiene un gran potencial para facilitar el diagnóstico y seguimiento de diversas enfermedades oculares. No obstante, su aprovechamiento efectivo depende en gran medida tanto de la calidad de las imágenes como de la capacidad del equipo de expertos en oftalmología para segmentar con precisión las estructuras fundamentales del fondo de ojo.

En el estado de Baja California, como en muchas otras regiones del país, persiste una marcada desigualdad en el acceso a servicios de salud visual, especialmente en comunidades rurales y marginadas [2]. La escasez de especialistas en oftalmología, sumada a la concentración de recursos diagnósticos en zonas urbanas, limita la detección oportuna de enfermedades visuales prevenibles o tratables[3]. Ante esta realidad, la integración de tecnologías accesibles como la fundoscopia digital, apoyada por algoritmos de análisis automatizado, representa una alternativa prometedora. Este enfoque no solo podría ampliar la cobertura diagnóstica, sino también apoyar al personal médico de primer contacto mediante herramientas que prioricen casos con hallazgos de riesgo, optimizando así los recursos disponibles y mejorando la equidad en la atención oftalmológica. La motivación principal de esta investigación radica en identificar, analizar y seleccionar algoritmos apropiados para la segmentación y el procesamiento digital de imágenes que posean relevancia clínica y sean computacionalmente viables.

1.2 Planteamiento del problema

El acceso oportuno a diagnósticos oftalmológicos de calidad sigue siendo un desafío en las regiones de México debido a la escasez de especialistas [3], las limitaciones tecnológicas y la inequidad en la repartición de recursos de salud [4]. En este contexto, las tecnologías de imagen, como la fundoscopia digital, han adquirido un papel cada vez más relevante, siendo uno de los exámenes básicos en la práctica clínica, junto con el examen de agudeza visual. La fundoscopia digital permite la visualización no invasiva del fondo de ojo y apoya el trabajo clínico de los especialistas. La interpretación de imágenes de fundoscopia digital puede facilitar el diagnóstico de diversas enfermedades oculares; sin embargo, factores como la baja calidad de las imágenes y la complejidad anatómica del fondo de ojo dificultan la segmentación precisa de estructuras clave como la mácula, el disco óptico y los vasos sanguíneos.

Dado que hay una limitación significativa de expertos en salud visual [5] disponibles para atender a toda la población, se vuelve especialmente atractivo explorar el uso de algoritmos de cómputo inteligente. Estos algoritmos pueden tener como objetivo el facilitar procesos de gestión del riesgo clínico, permitiendo así manejar de manera adecuada y segura los flujos de pacientes que hay en un entorno clínico, especialmente cuando la demanda y las necesidades clínicas superan los recursos disponibles. A este enfoque también se le conoce como *triaje automatizado* [6]. Los algoritmos también pueden ayudar a identificar, a partir de estudios de fundoscopia digital, a los pacientes que presenten indicadores asociados a patologías o a un alto riesgo de desarrollarlas. Para abordar estos desafíos, los profesionales de la salud visual evalúan las imágenes de fondo de ojo comenzando por identificar las estructuras anatómicas clave, para luego avanzar en la detección de patrones patológicos. En este contexto, resulta imprescindible elegir los algoritmos adecuados de procesamiento digital de imágenes y aprendizaje automático que permitan alcanzar la tarea de segmentación de manera precisa y robusta, sin olvidar la optimización al mismo tiempo en el uso de los recursos computacionales. La implementación de estas tecnologías contribuiría a mejorar la detección y el seguimiento de enfermedades oculares, reduciendo la carga de trabajo clínico y ampliando el acceso a servicios de salud visual, especialmente en contextos con recursos limitados.

1.3 Pregunta de investigación

¿Cómo seleccionar y adaptar algoritmos de segmentación automática para identificar con precisión las estructuras anatómicas del ojo en imágenes de fundoscopia digital de baja calidad, optimizando el recurso computacional y su aplicabilidad en entornos clínicos similares al contexto regional?

1.4 Hipótesis

Es posible identificar con precisión estructuras anatómicas clave en imágenes de fundoscopia digital de calidad variable mediante la selección y adaptación de algoritmos de segmentación automática, lo que permite una implementación eficiente en contextos clínicos regionales.

1.5 Objetivo general

Investigar y validar técnicas de segmentación aplicadas a imágenes de fondo de ojo, mediante el uso de algoritmos de aprendizaje automático y procesamiento digital de imágenes, con el fin de identificar de manera precisa las principales estructuras anatómicas relevantes en el contexto clínico.

1.6 Objetivos específicos

1. Buscar y recabar las bases de datos de fundoscopia digital que sean representativas de imágenes de fondo de ojo en la práctica clínica, priorizando el contexto regional.
2. Catalogar los algoritmos utilizados en el estado del arte para atacar la tarea de segmentación de estructuras anatómicas en imágenes de fundoscopia digital.
3. Obtener capacitación en conceptos generales de oftalmología.
4. Seleccionar los algoritmos con mejor rendimiento en la tarea de segmentación de estructuras anatómicas en imágenes de fundoscopia digital de acuerdo a la revisión del estado del arte.
5. Validar experimentalmente los algoritmos de segmentación seleccionados utilizando imágenes representativas de la práctica clínica regional, evaluando su rendimiento mediante métricas encontradas en la revisión de la literatura
6. Proponer una aproximación algorítmica apropiada para realizar la tarea de segmentación de estructuras anatómicas en imágenes de fundoscopia digital similares a las utilizadas en la región.

1.7 Estructura de la tesis

Este trabajo está dividido en los siguientes capítulos:

Capítulo 1: En este capítulo se presenta una introducción breve al proyecto de tesis, se expone una justificación del enfoque adoptado, se plantea el problema de investigación y se describe la secuencia de contenidos de los capítulos siguientes.

Capítulo 2: Se presenta una revisión de los fundamentos ópticos aplicados a la fundoscopia, las estructuras anatómicas relevantes, y la evolución de la imagenología ocular. Posteriormente, se revisa literatura entre 2014 y 2024 sobre algoritmos de segmentación y análisis automático en imágenes de fondo de ojo, con el fin de seleccionar aquellos con mayor aplicabilidad clínica para este proyecto.

Capítulo 3: Se describe la estrategia de investigación seguida en este trabajo, incluyendo el marco teórico, los conceptos clave, los algoritmos considerados, el diseño metodológico, así como las decisiones relacionadas con la recopilación de datos, el preprocesamiento de imágenes, la selección y evaluación de modelos, y su vinculación con necesidades clínicas.

Capítulo 4: Se presentan los resultados obtenidos a lo largo del proyecto. Se incluye una revisión sistemática de literatura, el análisis exploratorio del conjunto de datos, la evaluación de calidad de imagen, la aplicación de técnicas de preprocesamiento, y los resultados obtenidos con los modelos

U-Net, nnU-Net y SAM. También se incluyen comparaciones entre los algoritmos y una propuesta final de sistema de segmentación aplicable a entornos clínicos.

Capítulo 5: Este capítulo discute los hallazgos obtenidos, analiza las fortalezas y debilidades de cada modelo, y examina la aplicabilidad clínica de los enfoques propuestos. Asimismo, se identifican las limitaciones del estudio y se sugieren posibles líneas de trabajo futuro, así como las conclusiones generales del trabajo.

Capítulo 6: Este capítulo contiene las referencias bibliográficas completas de todas las fuentes consultadas y citadas en el desarrollo de la tesis, organizadas conforme a las normas establecidas para garantizar la correcta identificación y consulta de la información científica y técnica utilizada.

2. Revisión de la literatura

2.1 Fundamentos físicos y anatómicos

2.1.1 Principios ópticos aplicados en fundoscopia

El proceso visual inicia en el ojo humano, que puede ser considerado como un instrumento óptico extremadamente robusto y relativamente simple, que está compuesto por solo dos lentes positivas llamadas córnea y cristalino, las cuales proyectan las imágenes de alta resolución en la retina, es un sistema muy simple que no puede ser comparado en robustez con un sistema óptico artificial [7]. Primeramente hay que comprender cómo interactúa la luz con las estructuras oculares para poder interpretar las imágenes capturadas en los dispositivos modernos y cómo desarrollar algoritmos de análisis automatizado. La luz visible, que comprende longitudes de onda entre los 400 y 700 nanómetros, se comporta como una onda electromagnética que puede ser reflejada, refractada, absorbida o dispersada al interactuar con distintos medios [8]. En el caso del ojo humano, los cambios en el índice de refracción, es decir, cuánto se reduce la velocidad de la luz al pasar por ese material en comparación con su velocidad en el vacío, suceden en la interfase córnea-humor acuoso, y en la del cristalino-vítreo. El cristalino al tener curvas cambiantes y la córnea con su curvatura tienen la combinación perfecta de factores para la refracción y enfoque correcto de la luz en la retina [9].

La propagación de la luz en los tejidos biológicos, como los del ojo humano, está influenciada por varios fenómenos ópticos que determinan la calidad y precisión de las imágenes obtenidas. La manera por la cual funciona es muy similar a una cámara fotográfica, donde el ojo va a enfocar la luz que proviene de algún tipo de objeto en el mundo exterior sobre la retina, aquí es donde se encuentran las células fotorreceptoras. Estos fotorreceptores sumergidos en la retina están dispuestos en una geometría esférica, a diferencia del detector plano que se encuentra en una cámara [9], [10].

La capacidad del ojo para enfocar objetos a diferentes distancias se debe al poder óptico variable del cristalino, mientras que el poder de enfoque general del ojo se logra gracias al poder óptico fijo de la córnea [10]. Alrededor de dos tercios de la potencia óptica del ojo están en la córnea y un tercio en el cristalino. Por lo tanto, conforme a la ley de Snell [8], esta diferencia genera una mayor refracción de la luz al ingresar al ojo, lo que convierte a la córnea en el principal responsable del enfoque inicial de los rayos luminosos hacia la retina.

Para comprender con mayor detalle cómo estas propiedades ópticas se combinan y afectan la formación de la imagen en la retina, en esta área cabe mencionar un modelo que destaca, este es el modelo de ojo de Gullstrand [11], se utiliza para modelar con mayor precisión el comportamiento óptico del ojo humano, especialmente en aplicaciones clínicas y computacionales.

El modelo de ojo de Gullstrand que se observa en la Figura 2.1 describe al ojo como un sistema compuesto por medios ópticos con superficies esféricas y diferentes índices de refracción, asignando espesores específicos a cada uno de estos componentes. En este modelo Gullstrand propone una representación más realista que incorpora la curvatura y el espesor de la córnea, el índice de refracción del humor acuoso, el cristalino cuyo índice va variando de forma gradual y el humor vítreo [12], [13]. Además, considera la distancia entre las distintas superficies ópticas, como la separación

entre la córnea y el cristalino. Debido a esto se puede estimar con mayor exactitud la posición de la imagen formada en la retina, así como analizar fenómenos ópticos relevantes como la aberración esférica y la capacidad del ojo para enfocar objetos a diferentes distancias [14]. Estos aspectos resultan fundamentales en la óptica fisiológica, como en este caso en los sistemas de adquisición de imágenes del fondo de ojo.

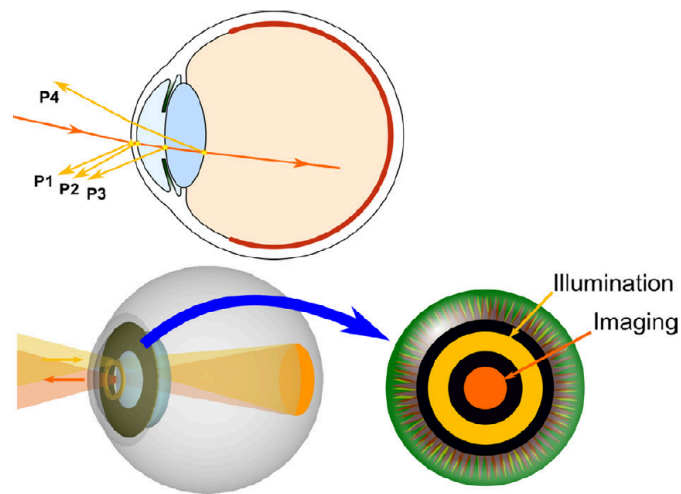


Figura 2.1 Principio de Gullstrand. Del lado izquierdo se encuentran las cuatro superficies principales del ojo según el modelo de Gullstrand (córnea anterior, córnea posterior, cristalino anterior y cristalino posterior) y en el lado derecho se observan las trayectorias de los rayos de luz que se usan en fundoscopias o en microscopios oftálmicos [11].

2.1.2 Estructuras clave del ojo

El ojo humano es un órgano de alta especialización y complejidad que ha sido catalogado como la ventana al alma humana [9]. El globo ocular de un adulto, frecuentemente conocido como esfera, en realidad posee una forma aproximadamente esférica, con un diámetro máximo de 24 mm en la dirección antero-posterior. La Figura 2.2 presenta un esquema detallado del ojo humano [15].

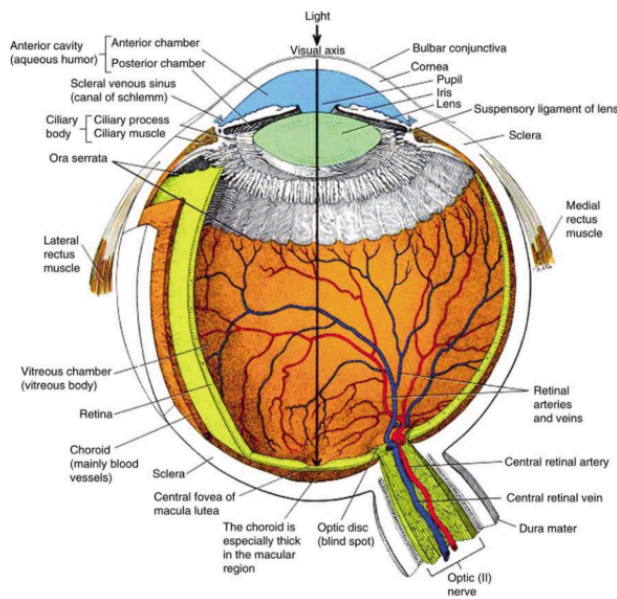


Figura 2.2 Esquema detallado del ojo humano [15].

En términos anatómicos generales, el ojo humano se divide en dos grandes áreas: el segmento anterior y el segmento posterior, cada área contiene estructuras así como condiciones patológicas particulares [15].

En el contexto de la práctica clínica, las tres estructuras de mayor relevancia diagnóstica en una imagen de fondo de ojo son la mácula, el nervio óptico y la arcada vascular y pueden ser observadas en la Figura 2.3. Estas áreas no solo constituyen elementos esenciales en la estructura y funcionamiento de la retina, sino que también proporcionan atributos visuales únicos que simplifican su segmentación y evaluación automatizada a través de herramientas computacionales. Además, representan los primeros indicadores que el oftalmólogo analiza durante el análisis del fondo de ojo, dado que su condición debe ser apropiada para poder seguir con la detección de posibles señales de patología. Así pues, su adecuada identificación y valoración son fundamentales tanto para el diagnóstico clínico [16].

Retina

La retina es la región principal que se captura en una fundoscopia digital, lo que la posiciona como un campo esencial para el diagnóstico y la investigación en oftalmología computacional. Anatómicamente es una estructura delgada de aproximadamente 0.5 mm de espesor y 30–40 mm de diámetro. Se trata de un tejido sensorial fotosensible localizado en la parte posterior del ojo. Está compuesta por millones de células agrupadas en tres tipos principales: células fotorreceptoras, células neuronales y células gliales. Su función es convertir la energía lumínica en que es transmitida al cerebro para su interpretación visual [15], [17].

Nervio óptico

La estructura que es conocida como el nervio óptico, es encargada de conectar el ojo con el cerebro. Transfiere los impulsos eléctricos generados en la retina hacia la corteza visual, donde son procesados

para ser interpretados como imágenes. Este alberga alrededor de 1.2 millones de fibras nerviosas. En la región donde el nervio óptico emerge se encuentra el disco óptico o punto ciego, una zona carente de células fotorreceptoras [15], [17].

Mácula

La mácula es una región de una forma oval y pigmentada que se encuentra en la retina, esta es esencial para la visión central de alta resolución. La fovea es central a la mácula y en su núcleo, presenta una depresión que alberga únicamente conos fotorreceptores, encargados de proveer la señal necesaria para la agudeza visual requerida para tareas como la lectura, el manejo y la apreciación fina de detalle [15], [17].

Arcada vascular

Los vasos sanguíneos de la coroides, la cual es una de las capas del ojo, ubicada entre la retina y la esclerótica, es decir, la parte blanca del ojo. son visibles incluso en el área foveal, mientras que los vasos retinianos rara vez se encuentran en esta región. Normalmente, los vasos sanguíneos centrales de la retina se dividen en dos ramas superiores e inferiores, dando lugar a cuatro ramas arteriales y cuatro venosas que emergen del disco óptico [15], [17].

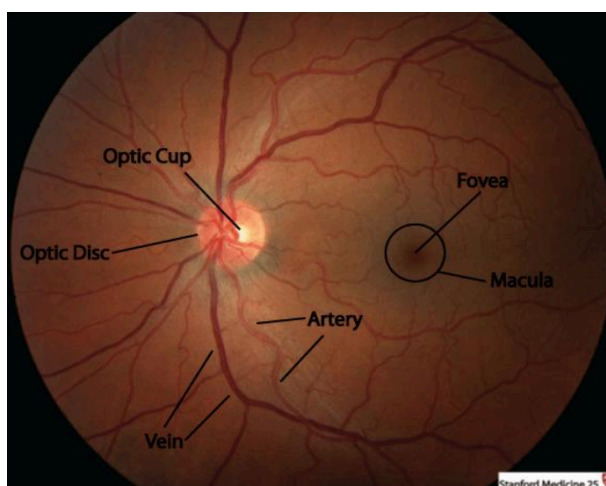


Figura 2.3 Estructuras anatómicas relevantes a lo largo de esta investigación. [18]

2.2 Historia y evolución tecnológica

2.2.1 Orígenes de la fundoscopia

Durante los siguientes tres siglos y medio, el entendimiento de la visión y la creación de imágenes en el ojo ha experimentado un avance significativo. Este avance ha evolucionado paralelamente al avance de tecnologías de imagen cada vez más avanzadas. A lo largo de la historia, la exploración del fondo de ojo ha sufrido una significativa transformación tecnológica. Desde los primeros dispositivos de óptica, como lo es el principio Helmholtz que facilitaban una representación elemental de la retina, se ha progresado hacia sistemas de imagen digital de alta resolución [11].

El denominado principio de Helmholtz, propuesto por Hermann von Helmholtz en el siglo XIX representa un hito fundamental en la comprensión de la percepción visual. Este principio defiende que los objetos son vistos como si estuvieran en el campo visual de la manera requerida para producir la misma impresión en el sistema nervioso, en circunstancias normales. En resumen, la percepción visual no se restringe a capturar de manera pasiva la imagen retinal, sino que involucra un proceso de interpretación cerebral, que determina en función de la experiencia y el contexto, qué objeto externo podría haber originado dicha imagen. Este método planteó la noción de que ver es, en cierta medida, un acto de inferencia, lo que se convirtió en esencial para el avance de instrumentos ópticos, este instrumento se puede observar en la Figura 2.4. Gracias a sus contribuciones, Helmholtz logró por primera vez la visualización directa del fondo de ojo humano, sentando así las bases de la fundoscopia moderna [19].

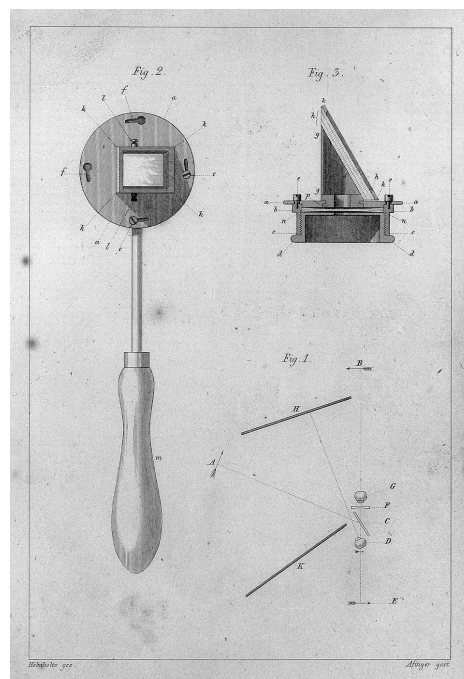


Figura 2.4 Oftalmoscopio original de Hermann von Helmholtz (1851), instrumento pionero que permitió por primera vez la visualización del fondo del ojo en pacientes vivos.[20]

2.2.2 Evolución hacia la fotografía de fondo de ojo

La historia de la fotografía del fondo de ojo se remonta a 1862, cuando el Dr. Henry Drury Noyes logró capturar por primera vez una imagen del fondo de ojo en un conejo. Ese mismo año, el Dr. Abner Mulholland Rosebrugh fotografió la retina de un gato, marcando los primeros pasos en la documentación visual del sistema ocular. No fue sino hasta 1886 que se registraron los primeros intentos en humanos: el Dr. Lucien Howe obtuvo una imagen de la retina humana, aunque con baja calidad, y en ese mismo año, Jackman y Webster realizaron también fotografías del fondo de ojo, consolidando este avance como un hito importante en la oftalmología moderna [21].

En 1905, la compañía Zeiss, en cooperación con el Dr. Oskar Dimmer, creó un innovador oftalmoscopio que incluía una cámara fotográfica. observado en la Figura 2.5, Este fue un progreso importante para solucionar varios problemas ópticos que impactaron a los aparatos previos, como la existencia de artefactos en las fotografías. Este diseño revolucionario estableció los cimientos para la creación de los primeros oftalmoscopios de uso comercial, señalando un hito en el progreso de las técnicas de documentación visual en el campo de la oftalmología [22].

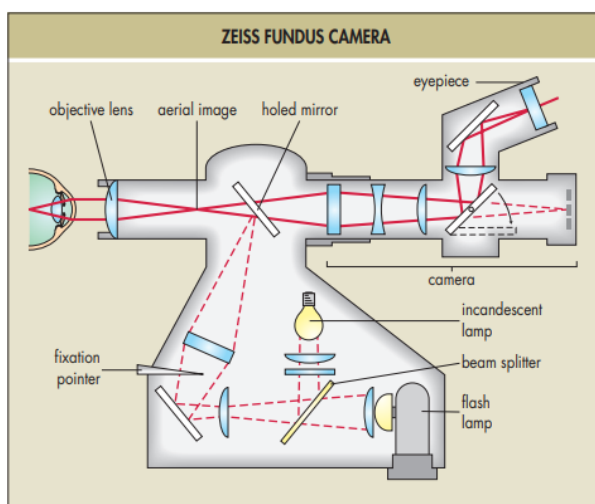


Figura 2.5 Sistema de iluminación de la cámara de fondo de ojo Zeiss. La luz incandescente y el flash se combinan mediante un divisor de haz y se dirigen al ojo a través de un espejo con orificio, mientras que la lente objetiva enfoca la luz en la pupila. La luz infrarroja se usa para enfocar la retina sin estimularla antes de tomar la fotografía [17].

2.2.3 Desarrollo de la cámara de fondo de ojo moderna

Durante el siglo XX se registraron avances significativos en la fotografía del fondo de ojo. En 1926, Zeiss introdujo la primera cámara comercial para el fondo de ojo. Metzger, un año después, en 1927, publicó las primeras imágenes estereoscópicas del fondo de ojo, expandiendo las oportunidades de diagnóstico a través de una percepción tridimensional. En 1953, Hansell y Beeson lograron integrar exitosamente un tubo electrónico de flash en la cámara oftálmica, lo que incrementó significativamente la calidad de la iluminación y la visibilidad de las estructuras retinianas. En 1960, la cámara Equator-Plus, creada por Pomerantzeff, posibilitó la toma de fotografías con un campo de visión de hasta 148°, lo que aumentó de manera notable la cobertura visual de la retina. En 1975, Steven Sasson, ingeniero de Eastman Kodak, creó la primera cámara digital, lo que marcó un hito hacia la época de la documentación digital en imágenes médicas [22].

Estos avances establecieron los fundamentos tecnológicos para el progreso de la fundoscopia digital, ya que el éxito de la obtención de las imágenes de fondo de ojo depende de la interacción entre la óptica de la cámara de fondo de ojo y la óptica del propio ojo. Las cámaras de fondo de ojo utilizan un diseño de lente objetivo esférica que, al combinarse con la óptica del ojo del sujeto, adapta el plano de enfoque a la curvatura del fondo de ojo, la trayectoria de la luz que sale de la cámara de fondo de ojo pasa a través de la óptica refractiva de la córnea y el cristalino, que suelen enfocarse a distancia tras la dilatación. El control de enfoque de la cámara se utiliza para compensar los errores de refracción del

ojo del sujeto, mientras que la luz necesaria se suministra de manera axial, este sistema óptico perteneciente a la cámara También proyecta la forma de un anillo de luz siguiendo los principios de Gullstrand anteriormente mencionados para separar la iluminación entrante y saliente. Debido a los avances de este salto tecnológico, la cámara de fondo de ojo moderna se transforma en un recurso imprescindible, no solo para la documentación visual, sino también como plataforma fundamental para el progreso de tecnologías fundamentadas en inteligencia artificial [23], [24].

2.3 Aplicaciones clínicas y computacionales actuales

2.3.1 Avances tecnológicos en la evaluación de la calidad de imágenes de fondo de ojo

La revolución digital ha sido un avance clave en la fundoscopia moderna, permitiendo el registro, almacenamiento y análisis digital de imágenes retinianas. Esta transformación ha facilitado no solo el seguimiento longitudinal de patologías oculares, sino también la implementación de técnicas de procesamiento digital para apoyar el diagnóstico clínico [11].

Desde finales del siglo XX, múltiples investigaciones han sugerido técnicas automatizadas para valorar la calidad de las imágenes del fondo de ojo, con la finalidad de garantizar su aptitud para el análisis clínico o computacional. Dentro de las primeras investigaciones sobresale *Automatic Retinal Image Quality Assessment and Enhancement* [25], que implementó el índice de calidad Q como medida cuantitativa y objetiva para evaluar la calidad de la imagen retiniana. En esta misma línea, *Automatic Visual Quality Assessment in Optical Fundus Images* [26], planteó un enfoque basado en la identificación de características comunes a imágenes de buena calidad.

Posteriormente, *Diabetic Retinal Image Quality Based on Clarity and Field Definition* [27], propuso un método específico para imágenes de un solo campo de 45°, centrado en el diagnóstico de retinopatía diabética. Desde aquel momento, se han creado instrumentos más sofisticados, como el enfoque sugerido por Kohler et al, que emplea la segmentación vascular para medir el desenfoque y el ruido sin requerir una imagen de referencia en 2013 denominado *Automatic no-reference quality assessment for retinal fundus images using vessel segmentation* [28].

En 2014, desarrollaron el sistema *Automatic Retinal Interest Evaluation System (ARIES)*[29], que alcanzó una precisión del 99.54% al identificar imágenes retinianas dentro de grandes conjuntos de datos. En años posteriores, surgieron métodos basados en extracción de características genéricas[30], redes neuronales convolucionales inspiradas en la neurobiología [31], y combinaciones de modelos supervisados y no supervisados para la clasificación de calidad [32].

Además, Shao et al [33], propuso una evaluación automática basada en la iluminación, naturalidad y estructura de las imágenes, y más recientemente, Avilés et al [34], integró el análisis de datos topológicos (TDA) con técnicas de aprendizaje automático para mejorar la evaluación de calidad en fundoscopia digital.

Estas investigaciones han evidenciado la eficacia de métodos como el aprendizaje automático, la obtención de características, el análisis topológico de datos y métricas objetivas concretas para incrementar la exactitud en el diagnóstico. Sin embargo, el análisis de la calidad de las imágenes de fondo de ojo continúa siendo un reto fundamental, dado que la calidad de las imágenes impacta directamente en la capacidad de los algoritmos para detectar estructuras anatómicas con exactitud. Es

crucial la calidad de la imagen para asegurar diagnósticos precisos y a tiempo, dado que imágenes de baja resolución pueden obstaculizar la correcta segmentación de zonas esenciales como la mácula, el disco óptico y los vasos sanguíneos.

2.3.2 Importancia clínica de la fundoscopia

La fundoscopia proporciona el único medio para la inspección directa de arterias, venas o el sistema nervioso central en el paciente vivo e intacto, mayormente se utiliza para registrar afecciones médicas, seguir la progresión de enfermedades y crear ilustraciones para publicaciones y fines educativos.[35] Eso se hace por medio de un examen, el cual como se menciona es tan rutinario como el examen básico como el examen de agudeza visual, todo médico puede practicarlo no solamente el oftalmólogo. Este consta de una inspección solamente donde uno observa por un oftalmoscopio, este es simplemente una luz con varias modificaciones ópticas, incluyendo lentes donde el instrumento ilumina la retina a través del defecto normal del iris que es la pupila. Los rayos de luz que forman la imagen de la retina se reemergen a través de la pupila. La abertura de visualización del oftalmoscopio contiene una lente que modifica los rayos de luz para ayudar al usuario. El resultado final son las estructuras que se encuentran en el aspecto más interno del globo ocular, es a lo que se le conoce como fondo de ojo, estas estructuras son retina, vasos sanguíneos y nervio óptico [23], [36].

Muchas de las enfermedades sistematizadas más comunes tienen cambios característicos en el fondo de ojo, esto quiere decir que el diagnóstico oportuno al tomar una simple imagen de fondo de ojo puede revelar signos tempranos de enfermedades neurológicas, cardiovasculares, infecciosas, hematológicas y metabólicas, muchas de las cuales requieren intervención urgente. Esta es la función principal de la fundoscopia digital ya que la información que aporta es clave para la prevención de complicaciones graves, incluyendo la ceguera, daño neurológico irreversible y muerte [37], [38].

2.3.3 El salto de la oftalmología análoga a la oftalmología digital

Las imágenes de fondo de ojo se han desarrollado rápidamente durante los últimos 160 años, convirtiéndose en un pilar fundamental para la atención clínica, especialmente al tratar a pacientes con algún tipo de anomalía o enfermedades asociadas a la retina, así como aquellas derivadas de enfermedades sistémicas [38]. El progreso de la oftalmología ha dependido en gran medida del desarrollo de tecnologías de imagen cada vez más precisas. A lo largo de gran parte del siglo XX, la oftalmología depende de herramientas analógicas para el diagnóstico visual, sin embargo, a pesar de existir el oftalmoscopio directo e indirecto, estos solamente son enfocados en la observación de la retina en tiempo real, sin la posibilidad de poder capturar las imágenes, por lo tanto este proceso depende fuertemente del criterio de especialistas, sin posibilidad de análisis retrospectivo o de comparación longitudinal precisa [39].

Actualmente, las cámaras utilizadas para la obtención de imágenes del fondo de ojo se clasifican principalmente en dos tipos: cámaras no midriáticas y cámaras midriáticas [40]. Las cámaras no midriáticas son aquellas que permiten el tomar una imagen de fondo de ojo sin tener que dilatar la pupila, utilizan una luz infrarroja para el enfoque y un flash visible para la captura, no son muy espaciosas, Algunas versiones modernas permiten obtener imágenes a través de pupilas de hasta 3.3 mm, facilitando un estudio detallado de la papila óptica. Por otro lado, las cámaras midriáticas requieren la dilatación de la pupila ya que utilizan una luz visible, Muchas permiten capturas simultáneas, lo que mejora la precisión diagnóstica al evitar errores por movimiento entre

exposiciones, son comúnmente usadas en estereofotografía retiniana. Ambas son ampliamente utilizadas en el diagnóstico de enfermedades oculares [21]. Debido a la complejidad en el diseño y a los altos estándares clínicos requeridos para su fabricación, sólo un número limitado de fabricantes produce estos instrumentos, entre ellos: Topcon, Zeiss, Canon, Nidek, Kowa, CSO y CenterVue [41].

Al momento de introducirse la oftalmología digital, existieron cambios considerables ya que ahora se debía de tomar en cuenta el poder realizar la captura, almacenamiento y procesamiento computacional de imágenes del fondo de ojo permitió no solo mejorar la documentación clínica, sino también integrar herramientas de diagnóstico asistido por computadora. Por otro lado, otro cambio que hubo para la evolución de la oftalmología con estas nuevas herramientas fue la identificación de características de referencia del fondo de una manera más amena, las más importantes sobre todo, como el disco óptico, la fovea y los vasos retinianos como coordenadas de referencia, las cuales son un requisito previo antes de que los sistemas puedan realizar tareas más complejas, como la identificación de entidades patológicas [42].

Sin embargo, al momento de realizar el análisis de las imágenes de fondo de ojo existen ciertos sujetos que se deben de tomar en cuenta, Este análisis abarcaba diversas áreas clave que se dividen en 3 ramas, primero, se debe de hacer una evaluación de la calidad de imagen, donde se verifica la calidad general de las imágenes, identificando artefactos y lesiones iatrogénicas, como cicatrices de láser, después existe una localización y segmentación de estructuras retinianas, donde se analizan los vasos retinianos, determinando su diámetro y clasificándolos en arterias y venas, así como detectando oclusiones vasculares. También se identifica la fovea y el disco óptico, evaluando características como la copa y el borde del disco, esta parte es esencial, ya que si no se tienen una buena localización y segmentación no se pueden realizar tareas más específicas como lo es la segmentación de anomalías como lo son hemorragias, microaneurismas, neovascularizaciones e infartos de fibras nerviosas también llamadas manchas algodonosas, esto es en el caso de los vasos sanguíneos, generalmente también se identifican alteraciones del epitelio pigmentario, incluyendo drusas y cambios en la pigmentación, y anomalías coroideas, como nevos, melanomas y lesiones asociadas a uveítis [38].

Más que nada, esta transición que se obtuvo en oftalmología, transformó la manera en la cual se adquieren y almacenan las imágenes de fondo de ojo. abriéndole la puerta a una nueva era de análisis automatizado y diagnóstico asistido por computadora. Esta evolución ha permitido mejorar la precisión diagnóstica, estandarizar evaluaciones clínicas y facilitar el monitoreo longitudinal de patologías oculares como se menciona anteriormente, por lo tanto no solo se observa una mejora tecnológica, más que nada es un cambio de paradigma que sienta las bases para el uso de técnicas de inteligencia artificial en la práctica clínica oftalmológica.

2.3.4 Identificación de estructuras anatómicas en fundoscopia digital en el proceso clínico.

La localización y segmentación de las estructuras dentro de la retina resultan tareas cruciales para los sistemas de análisis automático de imágenes de fondo de ojo. Como lo es el nervio óptico que facilita la extracción automática del resto de estructuras anatómicas y patológicas. Además, su posición y diámetro pueden utilizarse como referencia de posición de otras estructuras de la retina. Son entre ellas puntos de encuentro ya que la distancia aproximadamente constante entre el nervio óptico y la fovea se puede utilizar para estimar la ubicación de la mácula, así como el nervio óptico también es el punto de partida para otras estructuras, como es información útil para la segmentación capilar, ya que

el árbol vascular nace de él, junto con la arcada vascular que sirve para identificar el ojo si se está hablando del derecho o el izquierdo al que se le ha hecho la imagen de fondo de ojo. El cambio que existe en estas estructuras, forma o color, es un indicador de algún tipo de patología, por lo que sus dimensiones se utilizan para medir características anormales causadas por ciertas enfermedades [40], [43].

No obstante, existen otras consideraciones como ciertos riesgos, uno de ellos es la infección cruzada entre pacientes, especialmente en aquellos con infecciones oculares activas, para las personas con fotofobia la luz intensa del flash puede resultar molesta e intolerable. También, al momento de adquirir la imagen pueden existir diversos artefactos que comprometen la calidad del diagnóstico debido a imperfecciones en las fotografías, entre los más comunes puede ser el tener reflejos naranjas o brillantes, estos se presentan en los bordes de la imagen producto de la iluminación del iris en pupilas poco dilatadas, las pestañas del paciente en la imagen también llamados velos difusos, neblina o pérdida de detalle que es generada por no tomar las imágenes a una distancia adecuada y manchas recurrentes por no tener la higiene adecuada con la cámara que se esté utilizando. Mayormente estas complicaciones pueden evitarse si hay un uso adecuado de los equipos y las normas de higiene de la institución que realiza el examen [42].

2.3.5 Algoritmos de segmentación en imágenes de fondo de ojo con técnicas de cómputo inteligente

La segmentación de imágenes mediante técnicas de aprendizaje automático ha revolucionado el análisis de imágenes médicas, especialmente en áreas donde la precisión y la rapidez en el diagnóstico son críticas. En el área oftalmológica este tipo de herramientas para la segmentación automática en imágenes de fondo de ojo tiene relevancia notable ya que se enfoca en identificar y delimitar regiones de interés, como vasos sanguíneos, disco óptico y mácula, con alta precisión. Este avance no solo agiliza el proceso diagnóstico, reduciendo el tiempo que un especialista necesita para analizar cada imagen, sino que también incrementa la objetividad al minimizar la variabilidad entre observadores. Al integrar técnicas computacionales en estos procesos, se logra mejorar la detección temprana de enfermedades que amenazan la visión, como la retinopatía diabética y el glaucoma, contribuyendo así a una atención médica más eficaz y accesible [1], [44].

En los últimos años, el aprendizaje automático ha experimentado un resurgimiento notable, impulsado por el aumento de la potencia computacional y la disponibilidad de extensos conjuntos de datos. A medida que se escalan a volúmenes mayores de datos, los modelos de aprendizaje profundo han mostrado mejoras significativas en su rendimiento, superando a muchos métodos clásicos de aprendizaje automático en precisión y eficacia. Este avance permite desarrollar modelos de diagnóstico más precisos y eficientes, lo cual resulta particularmente relevante en el campo de la oftalmología, donde la aplicación de aprendizaje profundo en imágenes oftálmicas puede facilitar diagnósticos automatizados que optimizan los tiempos y reducen los costos [1].

3. Desarrollo de la estrategia de investigación

3.1 Metodología para la validación

Para abordar esta tesis se eligió una metodología que permite combinar de una manera intuitiva los marcos de modelado en el ámbito conceptual, de integración y computacionales. Por ello, la metodología DIMA (*Decentralized Integrated Modeling Approach*) la cual, como se describe en [45], es especialmente adecuada para este propósito, ya que facilita la construcción de modelos complejos mediante la integración de diversos enfoques y la descentralización de decisiones, manteniendo un enfoque para la segmentación automática de estructuras anatómicas, y puede ser aplicada al caso de imágenes de fundoscopia digital.

Una de las principales ventajas de DIMA es su capacidad para implementar el principio de modelado completo, abarcando las dimensiones conceptual, computacional y operativa. Esto permite una transición fluida entre las distintas etapas del desarrollo del proyecto, desde el diseño teórico hasta la validación práctica. Asimismo, DIMA contempla una visión holística del sistema al combinar distintos enfoques de modelado, coordinar los niveles de planificación y ejecución, y articular los distintos niveles de toma de decisiones de manera coherente

En el contexto de esta investigación, buscando cumplir el mantener un enfoque técnico alineado con las necesidades clínicas desde las primeras etapas, de forma que los resultados obtenidos puedan ser utilizados en escenarios reales de atención a pacientes así como la capacidad de adaptar y combinar diferentes modelos y algoritmos permite responder eficazmente a las variaciones en la calidad de las imágenes y a las restricciones computacionales, asegurando así la viabilidad y eficacia del sistema propuesto, DIMA ofrece la flexibilidad y modularidad que son fundamentales. La Figura 3.1 muestra un diagrama esquemático de la metodología DIMA aplicada paso a paso al desarrollo de esta tesis. A continuación, se presenta una descripción detallada de cada uno de los componentes mostrados en el diagrama.

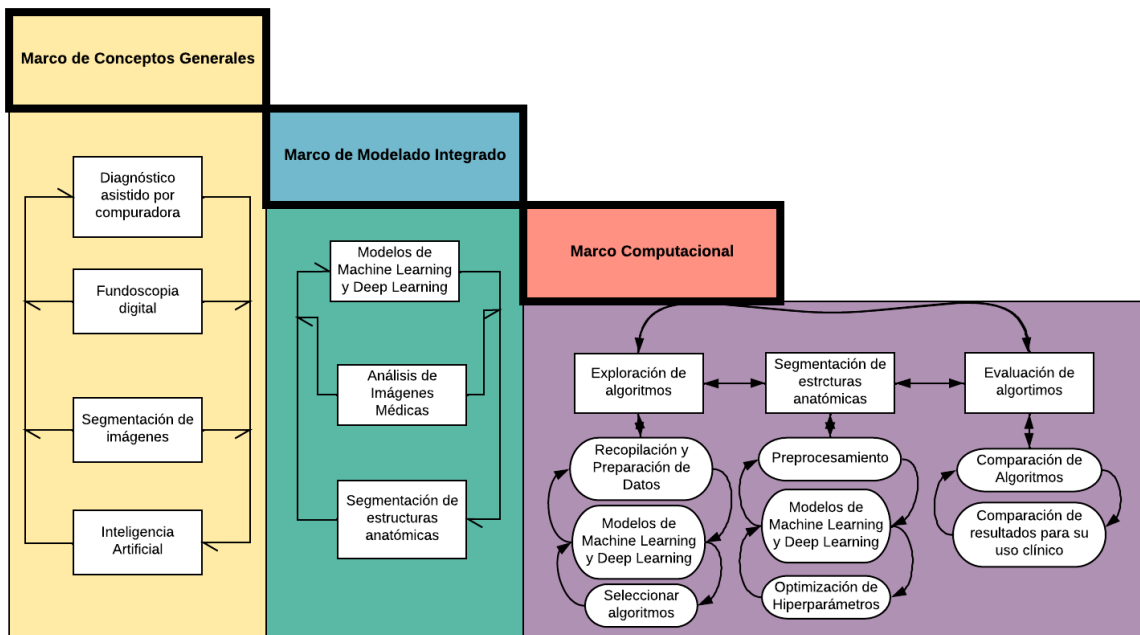


Figura 3.1 Esquema general de la metodología DIMA adaptada para el desarrollo de un sistema de segmentación automática de estructuras anatómicas en imágenes de fundoscopia digital.

3.2 Marco de Conceptos Generales (General Concepts Framework)

Esta sección introduce los conceptos clave que sustentan el enfoque metodológico adoptado en esta investigación. Cada uno de estos elementos es necesario para comprender la estructura, los objetivos y la implementación del sistema propuesto desde el punto de vista metodológico.

3.2.1 Diagnóstico asistido por computadora

El diagnóstico asistido por computadora (CAD, por sus siglas en inglés) se centra en el uso de herramientas computacionales que ayudan al personal de salud con la interpretación de la información médica, mayormente se utiliza en imágenes. Usualmente estos sistemas utilizan métodos de aprendizaje automático para analizar tanto datos de imagen como datos clínicos no visuales, provenientes de muestras de casos anteriores de una población de pacientes, todo esto con la finalidad de poder desarrollar modelos que puedan resolver tareas como hacer proyecciones o construir una relación entre los datos facilitados y los resultados en el contexto de atención a la salud [46].

Los sistemas CAD, que tradicionalmente son basados en aprendizaje automático, se construyen mediante diseño manual de descriptores de características, extraídos a través de técnicas de procesamiento de imágenes y conocimiento clínico especializado. Este tipo de características pueden representar en una imagen su forma, textura e intensidad dependiendo de qué parte de la imagen se este observando [47].

3.2.2 Fundoscopia digital

La fundoscopia digital es ahora la modalidad existente de imagen estándar tanto en clínicas oftalmológicas como en centros de detección. Esta es una técnica no invasiva utilizada para capturar imágenes de alta resolución del fondo de ojo, incluyendo estructuras como la retina, el disco óptico, la mácula y los vasos sanguíneos. Desde que se ha tomado en cuenta la toma de las imágenes por medio de una cámara digital junto al uso de cómputo y la capacidad de almacenamiento, se espera que el CAD aporte en el proceso de diagnósticos oculares, especialmente en regiones con baja cobertura oftalmológica. Es una parte rutinaria del examen del ojo de todo médico, no solo del oftalmólogo. Consiste en la inspección donde uno observa las estructuras anatómicas del ojo humano a través de una cámara automática de fondo de ojo [11], [35].

3.2.3 Segmentación de imágenes

La segmentación de imágenes es una tarea fundamental en el ámbito de diagnóstico asistido por computadora y procesamiento digital, su objetivo principal es dividir una imagen en regiones significativas, o más bien etiquetar cada píxel según su pertenencia a una categoría específica. Cada tarea en el proceso de segmentación se puede dividir en aproximaciones semánticas, segmentación de instancias y segmentación panóptica [48].

Cuando se habla de segmentación semántica hace referencia a asignarle una etiqueta a cada píxel de la imagen de acuerdo a categorías definidas, esto sin distinguir entre diferentes objetos de la misma clase. Por otro lado, la segmentación de instancias extiende su enfoque a identificar y delimitar cada objeto individual dentro de una imagen, incluso si pertenece a la misma clase. Finalmente, la segmentación panóptica que combina ambos enfoques, ofreciendo una comprensión integral tanto a nivel de clase como de instancia [49].

3.2.4 Inteligencia artificial

La inteligencia artificial en términos generales, se refiere a un conjunto de métodos y algoritmos que están diseñados para imitar los comportamientos inteligentes propios de los seres humanos, como lo es el razonamiento, el aprendizaje, la toma de decisiones y la percepción visual. A lo largo de la historia, la inteligencia artificial tiene principalmente dos enfoques teóricos, el computacionalismo y el conexionismo. Donde el computacionalismo es una perspectiva más lógica, ya que se centra en reglas explícitas y axiomas para poder llegar a conclusiones de manera formal. Funcionan a través de algoritmos estructurados, de una manera análoga al funcionamiento de las computadoras tradicionales. Por otro lado, el conexionismo se inspira en cómo funciona el cerebro humano y propone modelos basados en redes de unidades simples, que se interconectan masivamente, como lo son las redes neuronales [50].

3.3 Marco de Modelado Integrado (Integrating Modeling Framework)

En este apartado se describen los modelos computacionales empleados en la presente investigación, los cuales se integran dentro de la metodología DIMA como unidades funcionales descentralizadas. Cada modelo fue seleccionado y ajustado con base en su capacidad para procesar imágenes de

fundoscopia digital y extraer con precisión las estructuras anatómicas relevantes. A través de esta integración, se busca construir un sistema robusto, adaptable a variaciones clínicas y viable en términos computacionales para su posible aplicación en entornos reales.

3.3.1 Análisis de imágenes médicas

El análisis de imágenes médicas conlleva al uso de técnicas computacionales para poder interpretar los datos generados por dispositivos clínicos, con el fin de recabar información estructural o que sea relevante. En el contexto de esta tesis, dicho análisis se aplica específicamente a imágenes de fondo de ojo obtenidas por fundoscopia digital, el cual como se mencionó anteriormente, es un método no invasivo que permite observar la retina y otras estructuras internas del globo ocular. Se considera esta etapa un componente esencial dentro del sistema que se propone, ya que permite comprender las particularidades visuales y clínicas del conjunto de datos, y a partir de ello, diseñar el preprocesamiento más adecuado para mejorar la calidad de las imágenes trayendo consigo la planeación para poder diseñar mejoras en cuanto a la normalización del brillo y contraste, la eliminación de ruido, la conversión a escalas adecuadas o la segmentación preliminar por mencionar algunas, las cuales son decisivas para asegurar el rendimiento de los modelos de aprendizaje automático y profundo aplicados posteriormente.

También, el análisis detallado de las imágenes permite identificar patrones frecuentes, variabilidad en la calidad y condiciones de captura, lo que guía la selección y configuración de los algoritmos de segmentación más adecuados. En otros términos, este análisis actúa como un puente entre la adquisición de datos clínicos y el desarrollo computacional, garantizando que cada etapa posterior esté contextualizada y optimizada para lograr resultados clínicamente útiles.

3.3.2 Modelos de Aprendizaje automático

Dentro del campo del aprendizaje automático, los modelos tradicionales, es decir, que no utilizan arquitecturas profundas ni redes neuronales complejas siguen siendo relevantes por su simplicidad y efectividad en ciertos contextos, han sido ampliamente utilizados en el análisis de imágenes médicas, particularmente cuando se dispone de conjuntos de datos limitados o cuando se busca mantener bajos requerimientos computacionales. Hablando de la segmentación de estructuras anatómicas en imágenes de fundoscopia digital, estos modelos han mostrado ser útil en tareas como clasificación de calidad de imagen, localización preliminar del disco óptico, y separación de regiones mediante técnicas basadas en características manuales.

En esta tesis, se contempla la revisión y comparación de diversos algoritmos supervisados, tales como los que se pueden encontrar en la literatura, Máquina de soporte vectorial (por sus siglas en inglés, SVM), Bosques aleatorios y k-Vecinos cercanos (por sus siglas en inglés, k-NN), los cuales han sido reportados en estudios previos por su simplicidad, rapidez y facilidad de implementación. Estos modelos son relevantes para contextos clínicos donde los recursos computacionales pueden ser limitados y donde la interpretabilidad del modelo es deseable. La revisión de estos enfoques permitirán ser considerados para su evaluación, su aplicabilidad y desempeño en relación con los requerimientos específicos del problema, lo que contribuirá a establecer un marco comparativo con técnicas más avanzadas.

3.3.3 Modelos de Aprendizaje Profundo

El uso de modelos de aprendizaje profundo ha marcado un antes y un después en el campo de la segmentación de imágenes médicas durante la última década. Debido a su capacidad para aprender representaciones complejas directamente de los datos, estos modelos superan muchas de las limitaciones asociadas al aprendizaje automático clásico, especialmente en tareas que requieren segmentación precisa a nivel de píxel. En este estudio se considera el análisis de arquitecturas de redes neuronales convolucionales (CNNs), con énfasis en aquellas diseñadas específicamente para tareas de segmentación biomédica. Dado su potencial y la evidencia reportada en la literatura, estas redes han demostrado ser especialmente eficaces en imágenes de fondo de ojo, logrando detectar estructuras clave como vasos sanguíneos, mácula y disco óptico, incluso en condiciones de iluminación o calidad variables.

3.3.4 Segmentación de estructuras anatómicas

El enfoque principal de este estudio es la segmentación de estructuras anatómicas en imágenes de fundoscopia digital, pues facilita la identificación de áreas de interés las cuales son esenciales para el diagnóstico oportuno. A partir del análisis inicial de las imágenes, como se mencionó anteriormente, se determina el preprocesamiento adecuado para realzar los rasgos relevantes y minimizar artefactos que puedan interferir con la detección automática. Este paso previo garantiza que los modelos de segmentación reciban datos consistentes y adecuados a su arquitectura.

La segmentación se aborda principalmente desde la perspectiva de la segmentación semántica, en la que cada píxel de la imagen es clasificado según la estructura anatómica a la que pertenece. Para ello, se emplean tanto algoritmos clásicos de aprendizaje automático como modelos de aprendizaje profundo, particularmente redes convolucionales diseñadas para tareas biomédicas, como U-Net y sus variantes. Estos modelos permiten una segmentación precisa incluso en imágenes de calidad variable, lo que es crucial para escenarios clínicos con condiciones heterogéneas.

Los resultados de segmentación se evalúan mediante métricas estándar, las cuales se explican unas secciones adelante, asegurando que los algoritmos seleccionados no solo funcionen computacionalmente, sino que también sean clínicamente relevantes. La robustez y exactitud de estas segmentaciones son fundamentales para que el sistema pueda integrarse como herramienta de apoyo al triaje automatizado o a la priorización de casos sospechosos en contextos sin acceso inmediato a oftalmología especializada.

3.4 Marco Computacional (Computational Framework)

El desarrollo computacional de este proyecto se organiza siguiendo los principios del modelo DIMA, integrando módulos funcionales que corresponden a cada etapa del flujo de procesamiento. A continuación, se describe cada uno de los componentes que conforman el marco computacional propuesto.

3.4.1 Exploración de algoritmos

Como parte inicial del proyecto, se realizará una revisión exhaustiva de la literatura con la finalidad de poder catalogar los algoritmos de segmentación aplicados a imágenes de fondo de ojo mayormente

utilizados dentro de esta década. Se incluirán tanto métodos clásicos de aprendizaje automático como modelos de aprendizaje profundo. Esta exploración tiene como objetivo identificar qué algoritmos se han utilizado previamente en contextos similares y cuáles han demostrado mejor rendimiento en tareas de segmentación de estructuras anatómicas.

3.4.2 Recopilación y Preparación de Datos

Se realizará una búsqueda exhaustiva en la literatura para poder identificar las bases de datos públicas de imágenes de fundoscopia digital que representen casos clínicos reales y que cumplan los requisitos necesarios para poder ser aplicados dentro de un algoritmo de segmentación. Habrá enfoque en aquellas que sean similares a las tomas que se obtendrían si estas fueran tomadas en un contexto clínico regional, sin importar si las tomas son provenientes de cámaras midriáticas y no midriáticas para simular la diversidad presente en la práctica clínica. Los datos serán organizados y almacenados en formato compatible con los *frameworks* de entrenamiento y evaluación de modelos algunos ejemplos de ello pueden ser PyTorch.

3.4.3 Preprocesamiento

El preprocesamiento de las imágenes incluirá varias técnicas esenciales para mejorar la calidad y homogeneidad de los datos, tales como:

- Normalización de intensidad y contraste.
- Creación de máscaras
- Conversión a escala de grises (cuando sea conveniente).
- Alineación y recorte centrado en regiones de interés.
- Redimensionamiento uniforme a resoluciones aceptadas por las arquitecturas de red.
- Aumentación de datos

Las transformaciones se aplicarán de forma controlada para no perder información anatómica importante. Esta etapa se desarrolla y ejemplifica principalmente en la sección de resultados, donde se presentan los efectos visuales y cuantitativos de dichas técnicas aplicadas sobre las imágenes.

3.4.4 Modelos de Aprendizaje Profundo y aprendizaje automático

En cuanto a los modelos de aprendizaje automático, se contempla la evaluación de algoritmos como lo que se mencionaron anteriormente en la sección de modelos de aprendizaje automático, los cuales serán analizados como posibles candidatos para tareas de clasificación básica o segmentación supervisada de estructuras retinianas, en especial cuando se busque reducir el recurso computacional. Por otro lado, dentro de los modelos de aprendizaje profundo, se proyecta la revisión y posible implementación de arquitecturas ampliamente utilizadas en el dominio médico, tales como:

- U-Net [51], debido a su eficacia comprobada en segmentación biomédica.
- ResUNet [52], una variante mejorada que incorpora bloques residuales para mayor profundidad y robustez.
- DeepLabV3+ [53], conocida por su capacidad de segmentación semántica mediante convoluciones dilatadas y contextos multiescala.

La selección final de los modelos que serán implementados experimentalmente se definirá con base en los resultados de la revisión sistemática de literatura que se presenta en el capítulo de resultados. Esta revisión permitirá identificar qué modelos han demostrado mejor desempeño en tareas comparables, considerando tanto su precisión como su viabilidad clínica y computacional en contextos similares al abordado en esta tesis.

3.4.5 Segmentación de estructuras anatómicas

La segmentación automática se centrará en las estructuras clave del fondo de ojo. Para este fin, se utilizarán modelos entrenados con subconjuntos de datos que cuenten con anotaciones manuales de referencia, los formatos utilizados para estas anotaciones manuales serán utilizados de acuerdo al algoritmo que se esté presentando, permitiendo evaluar la capacidad del sistema para identificar con precisión estas regiones anatómicas. Estas estructuras son fundamentales para integrar sistemas de diagnóstico asistido por computadora (CAD) como herramienta de triaje automatizado, la automatización de su segmentación busca aliviar la carga de trabajo del personal de salud, acelerar el proceso de evaluación y facilitar su implementación en contextos con escasez de especialistas o equipamiento.

3.4.6 Selección de algoritmos

La selección de algoritmos se basará en su capacidad para segmentar estructuras con precisión, pero también en su viabilidad clínica. Se considerarán criterios como:

- Precisión obtenida en métricas cuantitativas.
- Robustez ante variaciones en la calidad de imagen.
- Requerimientos computacionales.
- Potencial de aplicación en entornos clínicos reales con recursos limitados

En la sección de evaluación de los algoritmos seleccionados, se explicarán de manera detallada las métricas cuantitativas utilizadas, con el fin de interpretar adecuadamente los resultados obtenidos y su relevancia en un contexto clínico.

3.4.7 Optimización de Hiperparámetros

La optimización de los hiperparámetros de los modelos se llevará a cabo utilizando la técnica de validación cruzada. Esta estrategia permite evaluar el desempeño de un modelo bajo distintas particiones del conjunto de entrenamiento, reduciendo el riesgo de sobreajuste y mejorando la capacidad de generalización.

Dependiendo del modelo y del tamaño del conjunto de datos, se aplicará k -fold (participaciones) de validación cruzada, dividiendo los datos en k -subconjuntos. En cada iteración, uno de los subconjuntos se utiliza como conjunto de validación y los $k-1$ restantes como conjunto de entrenamiento. El proceso se repite k veces y el rendimiento final del modelo se obtiene como el promedio de las métricas obtenidas. Esta metodología asegura una selección más robusta de los hiperparámetros y permite comparar de forma más justa distintos algoritmos bajo condiciones similares.

3.4.8 Evaluación de algoritmos

El rendimiento de los modelos será evaluado utilizando métricas específicas para segmentación, tales como: IoU (Intersection over Union), coeficiente de Dice, precisión, sensibilidad, especificidad, tiempo de inferencia y uso de memoria (para valorar viabilidad computacional). Estas métricas permitirán comparar modelos tanto en desempeño técnico como en relevancia clínica. Todas estas métricas han sido seleccionadas conforme a criterios ampliamente reconocidos en la literatura especializada para segmentación médica, tal como se describe en [54].

- **IoU (Intersection over Union):** mide la superposición entre la segmentación predicha y la verdadera.

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

- **Coeficiente de Dice:** evalúa la similitud entre dos conjuntos.

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

- **Precisión:** proporción de verdaderos positivos entre los elementos clasificados como positivos.

$$Precisión = \frac{TP}{TP + FP}$$

- **Sensibilidad:** mide la capacidad del modelo para detectar los positivos reales.

$$Sensibilidad = \frac{TP}{TP + FN}$$

- **Exactitud:** capacidad del modelo para detectar correctamente los negativos reales.

$$Exactitud = \frac{TN}{TN + FP}$$

- **Tiempo de inferencia y uso de memoria:** se cuantificarán empíricamente durante las pruebas para valorar su viabilidad computacional en entornos clínicos.

Donde:

- **TP:** Verdaderos positivos (True Positives)
- **FP:** Falsos positivos (False Positives)
- **FN:** Falsos negativos (False Negatives)
- **TN:** Verdaderos negativos (True Negatives)

Estas métricas permitirán comparar modelos tanto en desempeño técnico como en relevancia clínica.

3.4.9 Comparación de Algoritmos

Se realizará una comparación sistemática entre los algoritmos seleccionados, considerando:

- Precisión y robustez de la segmentación.
- Adaptabilidad a diferentes calidades de imagen.
- Potencial de uso en sistemas de diagnóstico asistido.

Esta comparación guiará la recomendación final del modelo más adecuado para su implementación en un entorno clínico regional.

3.4.10 Comparación de resultados para su uso clínico

Una vez obtenidos los resultados de segmentación, se analizarán patrones de comportamiento entre modelos y casos clínicos. Se buscará identificar:

- Tendencias comunes de error de acuerdo a la estructura anatómica.
- Casos donde el modelo podría ser más útil como apoyo al triaje.
- Posibles ajustes necesarios para mejorar su aplicabilidad clínica.
- Exactitud al momento de realizar las segmentaciones.
- Similitud a lo que el clínico oftalmólogo proporciona de acuerdo a sus conocimientos.

Se discutirá también la posibilidad de integrar el modelo como herramienta de apoyo al personal de salud no especialista, se adjunta un diagrama del proceso a seguir en la Figura 3.2.

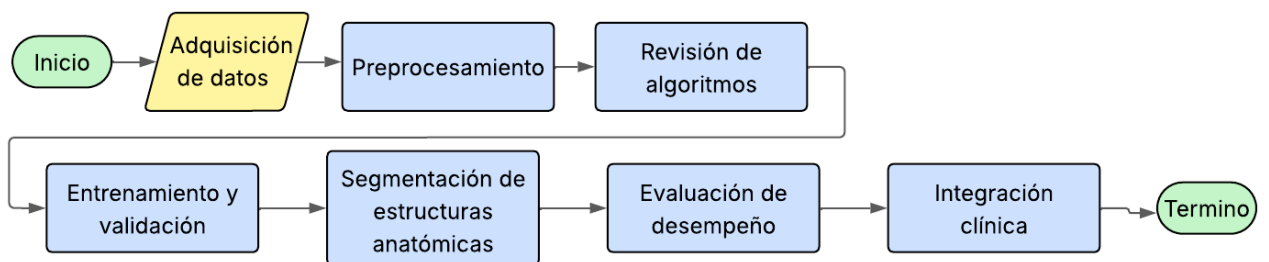


Figura 3.2 Diagrama general del flujo metodológico de la tesis, desde la adquisición y preprocesamiento de imágenes de fondo de ojo hasta la evaluación comparativa de los algoritmos de segmentación y su potencial uso clínico como herramienta de apoyo al diagnóstico.

4. Resultados

4.1 Revisión Sistemática de la Literatura

Se realizó una revisión de alcance para mapear sistemáticamente la investigación sobre la segmentación de estructuras anatómicas en imágenes de fundoscopia digital de calidad diversa, utilizando técnicas de procesamiento digital y aprendizaje automático. El objetivo es identificar las metodologías empleadas para la segmentación de imágenes de fondo de ojo, evaluar su eficacia y determinar las brechas existentes en el conocimiento de este campo para poder llegar a su aplicabilidad con una de las bases de datos sea seleccionada como la base de datos que sea similar de acuerdo al contexto regional.

4.1.1 Metodología de búsqueda y selección (PRISMA)

Años de publicación

Esta revisión sistemática examina la literatura publicada entre 2014 y 2024 en el ámbito de la segmentación de estructuras anatómicas en imágenes de fundoscopia digital utilizando técnicas de procesamiento digital y aprendizaje automático.

Para llevar a cabo esta revisión sistemática se siguió las pautas del PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [55], que permiten identificar, evaluar, seleccionar la elegibilidad y filtrar los estudios relevantes para este análisis. Se desarrolló una estrategia de búsqueda exhaustiva en varias bases de datos académicas como PubMed, Scopus e IEEE Xplore, estas bases de datos fueron consideradas debido a su robustez en bibliografía multidisciplinaria enfocada mayormente al área científica.

Cadenas de Búsqueda por Base de Datos

Se buscó en 3 bases de datos de disciplinas científicas, técnicas, médicas, biomédicas y tecnología cuyos nombres y cadenas de búsqueda, se realizó una búsqueda estructurada utilizando términos de búsqueda relevantes, sinónimos, palabras clave y filtros. Los detalles completos de estas búsquedas complementarias se pueden encontrar en la Tabla 4.1

Base de datos	Cadena de búsqueda	Resultados
SCOPUS	(TITLE-ABS-KEY (machine AND learning) AND TITLE-ABS-KEY (segmentation) AND TITLE-ABS-KEY (fundus AND image)) AND PUBYEAR > 2013 AND PUBYEAR < 2025 AND (LIMIT-TO (EXACTKEYWORD , "Image Segmentation") OR LIMIT-TO (EXACTKEYWORD , "Ophthalmology") OR LIMIT-TO (EXACTKEYWORD , "Eye Fundus") OR LIMIT-TO (EXACTKEYWORD , "Machine Learning") OR LIMIT-TO (EXACTKEYWORD , "Algorithm")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))	305

PUBMED	"machine learning and segmentation and fundus photography and image segmentation and Ophthalmology"	76
IEEE XPLORE	"machine learning and segmentation and fundus image and image segmentation and Ophthalmology"	67

Tabla 4.1 La tabla muestra para cada base de datos consultada su nombre, la cadena estructurada utilizada para la búsqueda, los filtros utilizados dentro las bases de datos y el número de artículos encontrados

A continuación, se describe el proceso de selección, los criterios aplicados y el resultado de la búsqueda en bases de datos científicas reconocidas.

Criterios de selección

Para poder realizar el proceso de selección de estudios se llevaron a cabo dos etapas principales para sólo incluir investigaciones relevantes y de alta calidad.

Primeramente, se realizó una revisión de los títulos y resúmenes de todos los artículos identificados a través de la búsqueda en bases de datos. Este paso tuvo como objetivo eliminar rápidamente aquellos estudios que claramente no cumplían con los criterios de inclusión predefinidos. En la segunda etapa, los textos completos de los estudios preseleccionados fueron revisados exhaustivamente para determinar su elegibilidad donde se observa la eficacia de las técnicas de segmentación utilizando aprendizaje automático en imágenes de fundoscopia. Se incluyeron sólo aquellos artículos que presentaban resultados completos y detallados, permitiendo una evaluación adecuada de las métricas y técnicas de segmentación. Además se registraron detalladamente los estudios excluidos junto con las razones de su exclusión, lo cual se presentó en un diagrama de flujo PRISMA. Finalmente se realizó una síntesis comparativa preliminar de los artículos revisados.

Criterios de elegibilidad

Se incluyeron estudios publicados en revistas revisadas por pares que utilicen técnicas de procesamiento de imágenes o algoritmos de aprendizaje automático para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Además, se consideraron aquellos estudios que reporten métricas de desempeño cuantitativas, las poblaciones de interés abarcan imágenes de fondo de ojo de individuos de cualquier grupo de edad, incluyendo tanto ojos sanos como aquellos con diversas patologías oculares. Las intervenciones incluyen el uso de técnicas de procesamiento de imágenes y la aplicación de algoritmos de aprendizaje automático. Se aceptaron comparaciones entre diferentes técnicas de segmentación, incluyendo métodos basados en aprendizaje profundo y métodos tradicionales, así como comparaciones entre imágenes de diferentes calidades (alta, media, baja). Los resultados incluyen métricas de segmentación, como precisión, sensibilidad, especificidad y área bajo la curva ROC (AUC-ROC), así como evaluaciones de la eficiencia computacional, como el tiempo de procesamiento o uso de recursos computacionales. Por otro lado, se excluyeron resúmenes, cartas al editor, opiniones, estudios de caso individuales y revisiones narrativas, así como estudios no revisados por pares, como artículos de conferencias no revisados e informes técnicos. También se descartaron aquellos estudios que no especifiquen la fuente o la calidad de las imágenes de fondo de ojo utilizadas, así como los que utilicen imágenes generadas sintéticamente y no imágenes reales. Se

excluyeron estudios que no utilicen técnicas de procesamiento de imágenes o algoritmos de aprendizaje automático, así como aquellos que solo describen técnicas sin proporcionar datos cuantitativos sobre su desempeño. Además, se descartaron estudios que no comparen diferentes técnicas de segmentación o que no reporten resultados de comparación. Los estudios que no reportaron métricas de desempeño cuantitativas claras y específicas, o que solo reportaron resultados cualitativos sin datos cuantitativos, también fueron excluidos. Finalmente, se excluyeron estudios publicados donde no se disponga de traducción adecuada y aquellos publicados hace más de 10 años, a menos que sean considerados pioneros y altamente citados en el campo.

Selección de estudios

Se identificaron un total de 448 registros a través de la búsqueda en bases de datos (PubMed, Scopus, IEEE Xplore), después de eliminar 33 duplicados, se revisaron 415 estudios en cribado para determinar su elegibilidad de los cuales se excluyeron 345 artículos debido a diversas razones, como la falta de información pertinente, el enfoque en la segmentación de patologías o su aplicación únicamente para la clasificación de enfermedades. Posteriormente se analizaron los estudio de texto completo de 70 artículos, donde $n = 1$ fue removido por el autor, $n = 2$ eran duplicados, $n = 11$ se enfocan solamente en regiones patológicas y $n = 34$ no aportan nada diferente a lo mencionado en otros artículos o son resultados muy similares a los encontrados en otros artículos, a continuación en la Figura 4.1 se puede observar el diagrama de flujo PRISMA .

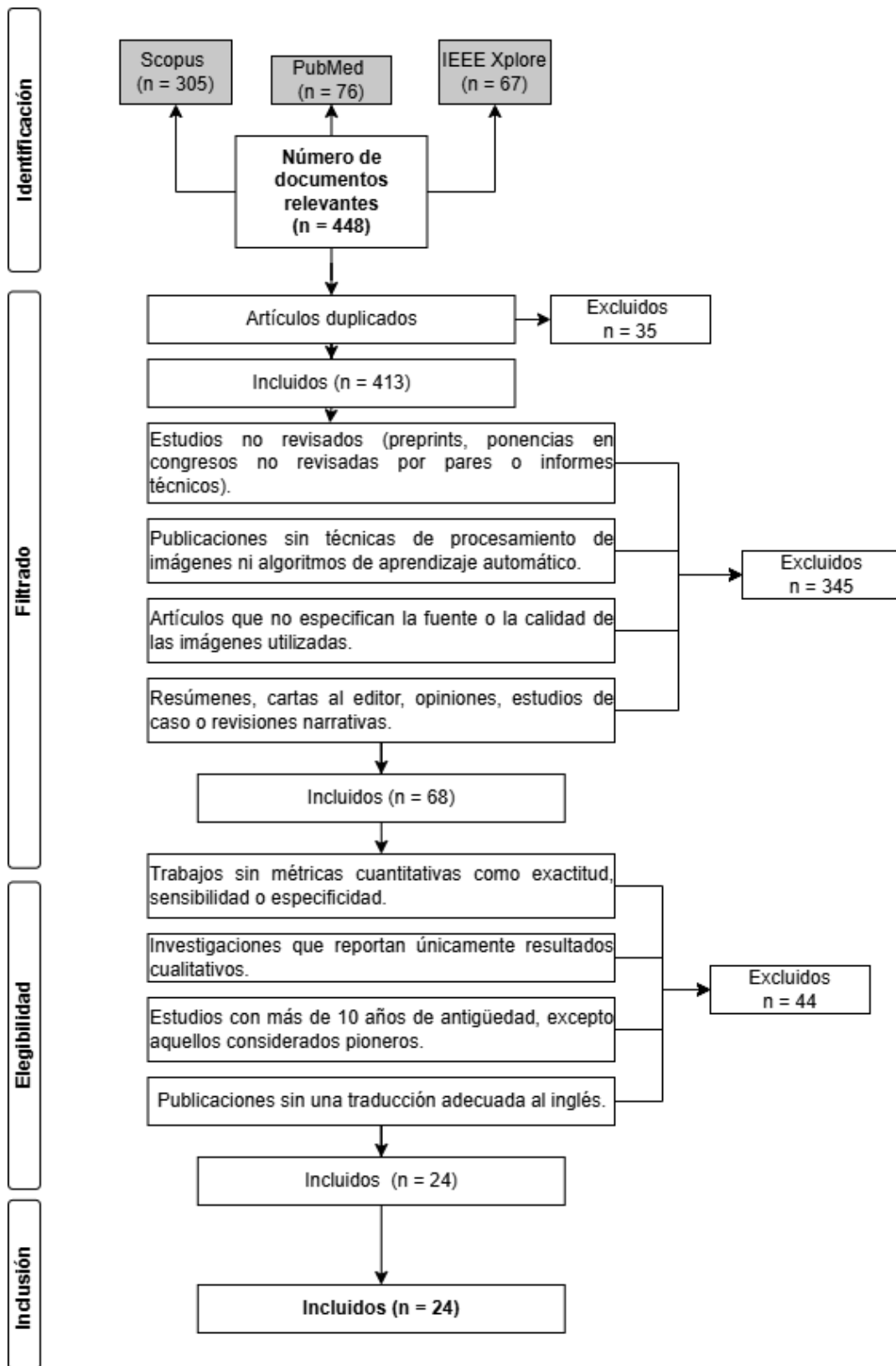


Figura 4.1 Diagrama de flujo PRISMA.

Después del proceso de selección de estudios, se obtuvieron 24 artículos. De estos, 22 emplean enfoques de aprendizaje supervisado, con 19 centrados en técnicas de deep learning. Dentro de estos 19 estudios, 7 utilizan redes neuronales convolucionales (CNN), 2 implementan redes completamente convolucionales (Fully Convolutional Networks, FCN), 5 emplean arquitecturas de U-Net, y los 4 restantes exploran diversas arquitecturas avanzadas de deep learning. Además, se identificaron dos artículos que aplican enfoques más tradicionales de aprendizaje automático. Uno de estos utiliza una Máquina de Aprendizaje Extremo (Extreme Learning Machine, ELM) y el otro se enfoca en un método basado en aprendizaje estructurado para el procesamiento de imágenes. En cuanto a técnicas de procesamiento digital de imágenes, se encontraron dos estudios donde se emplea un filtro de coincidencia modificado en uno y en otro la minería de datos.

Por otra parte, dos artículos se basan en el aprendizaje no supervisado, ambos utilizando deep learning. Uno de ellos emplea un enfoque semi-supervisado con redes profundas, mientras que el otro utiliza Redes Generativas Antagónicas (GAN). La Figura 4.2 Presenta un esquema que clasifica los 24 artículos revisados según los enfoques y técnicas utilizadas.

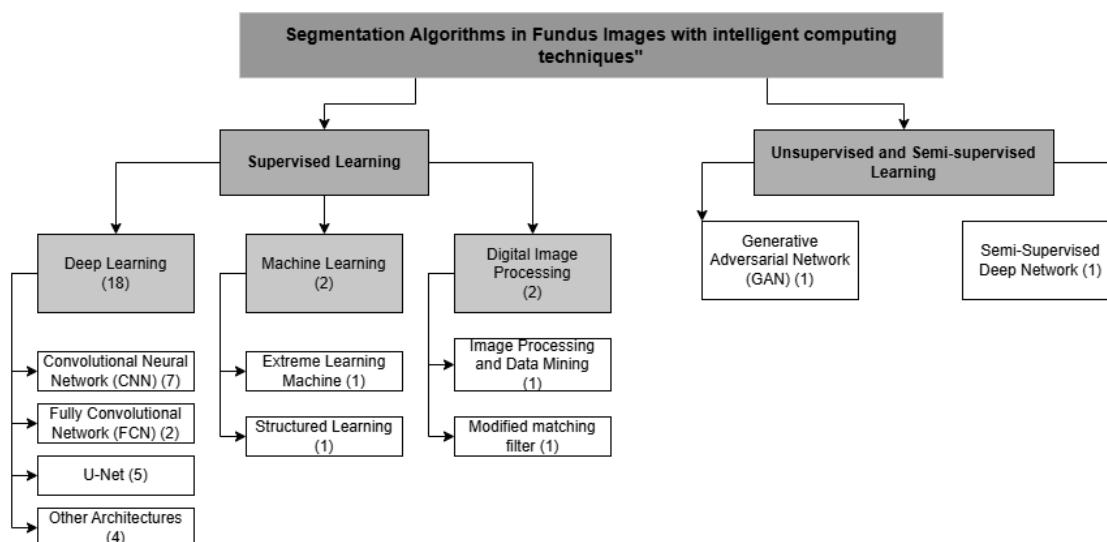


Figura 4.2 Artículos seleccionados en el estado del arte

Los artículos incluidos en esta revisión de la literatura tienen como objetivo identificar las metodologías más adecuadas para la segmentación de imágenes de fondo de ojo. Estas investigaciones buscan responder preguntas clave, cómo lograr una segmentación precisa de estructuras anatómicas, evaluar las técnicas de aprendizaje automático utilizadas y analizar el impacto del recurso computacional en el desarrollo de algoritmos de segmentación orientados a aplicaciones clínicas de bajo costo. Además, se exploran los resultados obtenidos en términos de precisión y eficiencia computacional.

4.2 Aprendizaje supervisado

Esta sección presenta el análisis de varias arquitecturas de aprendizaje profundo aplicadas a la segmentación de estructuras anatómicas en imágenes de fondo de ojo, incluyendo redes neuronales

convolucionales (CNN), redes completamente convolucionales (FCN), U-Nets y otras arquitecturas avanzadas.

4.2.1 Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales (CNN por sus siglas en inglés), un tipo de modelo de aprendizaje profundo para procesar datos que tienen un patrón de cuadrícula, como imágenes, que está inspirado en la organización de la corteza visual animal y diseñado para aprender de forma automática y adaptativa jerarquías espaciales de características, desde patrones de bajo a alto nivel. Las CNN, un tipo de redes neuronales artificiales que se han vuelto dominantes en diversas tareas de visión artificial, están despertando interés en diversos ámbitos, incluyendo la radiología. Las CNN están diseñadas para aprender de forma automática y adaptativa las jerarquías espaciales de las características mediante retropropagación utilizando múltiples componentes, como capas de convolución, capas de agrupación y capas completamente conectadas [56]. A continuación, se presenta un mapa conceptual en la Figura 4.3 que resume las arquitecturas de redes neuronales convolucionales aplicadas a la segmentación de estructuras anatómicas en imágenes de fondo de ojo, así como la estructura anatómica que segmenta.

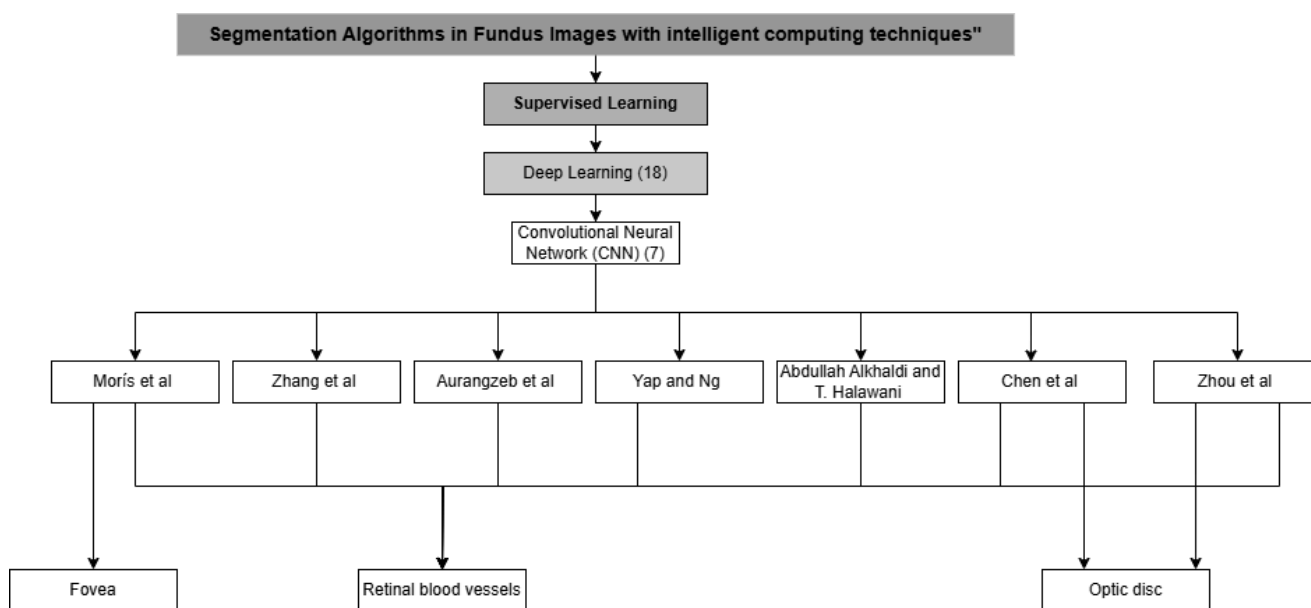


Figura 4.3 Mapa conceptual de redes neuronales convolucionales (CNN) aplicadas a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron diversas implementaciones basadas en CNN, cada una con enfoques específicos para la segmentación automática. A continuación, se discuten los trabajos encontrados destacándose sus hallazgos más representativos.

De acuerdo a Yap and Ng. [57], la propuesta principal es el algoritmo CAMContrast el cual consiste en 2 etapas, en la primera etapa, se entrena una red neuronal convolucional (CNN) para clasificar imágenes normales y anormales, generando mapas de activación de clase (CAM) que se utilizan para extraer información espacial sobre las regiones anormales y en la segunda etapa, se entrena un codificador de imágenes para maximizar la coincidencia entre pares de imágenes y mapas de calor derivados de la primera etapa. Se utiliza un conjunto de datos denominado como OIA-ODIR, es un conjunto de datos de fotografías de fondo ocular que contiene 10,000 imágenes anotadas con etiquetas de ocho categorías, este trabajo ha empleado hasta dos GPUs NVIDIA V100, cada una con 16 GB de memoria, CAMContrast supera a otros métodos en la mayoría de las tareas, mostrando su eficacia en el aprendizaje de representaciones a partir de etiquetas gruesas.

Por su parte, Chen et al. [58] la propuesta se centra en un marco novedoso que separa la regularización de forma y el mapa de segmentación. Se utiliza un extractor de regularización de forma probabilística basado en una transformación de blanqueo (WT-PSE) para suprimir interferencias de textura y estilo específicas del dominio así como obtener representaciones de forma robustas y de alta calidad. Además, se incorpora una red de distilación guiada por distancia de Wasserstein para transferir el conocimiento de la red docente (WT-PSE-T) a la red estudiante (WT-PSE-S) durante la inferencia, finalmente, se combina la regularización de forma con el mapa de segmentación para generar predicciones finales. Se utiliza la base de datos Fundos que consta de cuatro conjuntos de imágenes de fondo de ojo de diferentes centros de salud, con dos objetivos de segmentación: el disco óptico (OD) y la copa óptica (OC). El algoritmo se implementa utilizando PyTorch y se entrena en una GPU NVIDIA 3090, también, se utilizan técnicas de optimización como el optimizador Adam y se realizan experimentos durante 200 épocas, es decir, se refiere a que el modelo se entrena utilizando el conjunto de datos durante 200 ciclos completos a través de todo el conjunto de entrenamiento. Hubo una comparación entre 8 algoritmos similares de los cuales el método propuesto logró un promedio de Coeficiente de Similitud de Dice (DSC) de 83.11 y una Distancia de Superficie Promedio (ASD) de 13.04 para la Segmentación de la Copa Óptica (OC) y Segmentación del Disco Óptico (OD) el método alcanzó un promedio de DSC de 93.08 y un promedio de ASD de 10.47, estos resultados indican que el enfoque propuesto es efectivo para la segmentación de imágenes médicas, mostrando un rendimiento superior en comparación con otros métodos existentes.

Tal como se indica en Aurangzeb et al. [59], la propuesta se centra en el desarrollo de un modelo de aprendizaje profundo llamado ColonSegNet, se basa en una arquitectura de codificador-decodificador que incluye dos bloques de codificación y dos bloques de decodificación. Cada bloque de codificación consta de una convolución estriada de 3x3 entre los bloques residuales, mientras que cada bloque de decodificación contiene convoluciones transpuestas y conexiones de salto desde el bloque de codificación. Se busca optimizar el número de filtros en los bloques para reducir la complejidad computacional sin comprometer el rendimiento de segmentación, está enfocada en dispositivos de bajo rendimiento debido a su baja complejidad computacional y el número reducido de parámetros. Se utilizan 3 conjuntos de datos llamadas DRIVE, CHASE_DB1 y STARE, las cuales fueron seleccionadas para la segmentación de vasos. Ya que el propósito del estudio es reducir la complejidad computacional se enfatiza que el modelo ColonSegNet V2 ha logrado reducir el número de

parámetros de 5.00 millones a 982,177, lo que implica una disminución significativa en el recurso computacional y permite un uso más eficiente de los recursos computacionales. Se presentan comparaciones de rendimiento entre diferentes modelos de aprendizaje profundo, sin embargo, en el modelo propuesto la precisión en la segmentación de vasos retinianos se obtuvo un alto rendimiento, ya que los valores de sensibilidad superan e 0.83 en todas las bases de datos mencionadas, la especificidad también se mantuvo alta, alcanzando hasta 0.981 en STARE, lo que sugiere que el modelo es efectivo en la minimización de falsos positivos, cabe mencionar que la precisión se mantuvo por encima del 0.96, lo que refuerza la confiabilidad del modelo en la segmentación de vasos.

Según lo planteado en Moris et al. [60] la propuesta del algoritmo se divide en dos etapas, en el pre-entrenamiento del Context Encoder y el Ajuste fino para tareas específicas en la primera etapa entrena utilizando un proceso de imágenes transformadas en parches es decir, Patch-Wise Context Encoder (PW-CE) ,se Aplica un patrón de omisión en forma de tablero de ajedrez sobre la imagen completa ,Global Mask Context Encoder con patrón Checkerboard (GM-CE (CB)) y utiliza un patrón de omisión que proporciona más contexto al modelo, Global Mask Context Encoder con patrón Center-Surround (GM-CE (CS)), finalmente, Utiliza un patrón de omisión que proporciona más contexto al modelo y con ello resuelve tareas de segmentación en vasos retinianos y localización de fovea. Utiliza 3 conjuntos de bases de datos Isfahan MISP (50 imágenes para entrenamiento y 9 para validación), DRIVE (20 para entrenamiento y 20 para prueba) e IDRiD (413 para entrenamiento y 103 para prueba).El artículo no proporciona detalles específicos sobre el recurso computacional en términos de tiempo de procesamiento o recursos de hardware utilizados. Hablando de la reconstrucción de las imágenes Los enfoques GM-CE mostraron un mejor rendimiento en términos de similitud estructural (1-SSIM) en comparación con PW-CE, especialmente cuando se utilizó la función de pérdida SSIM, al momento de segmentar los vasos retinianos se utilizó el conjunto de datos DRIVE aquí se observó que el enfoque PW-CE con la función de pérdida SSIM obtuvo el mejor rendimiento, alcanzando un Área bajo la curva de precisión-recall (AUC-PR) superior al 91% con 20 imágenes de entrenamiento, por último, la base de datos utilizada para segmentar la fovea fue el conjunto de datos IDRiD donde el enfoque PW-CE con la función de pérdida de Error cuadrático medio (MSE) mostró el mejor rendimiento, con un error de distancia media que disminuyó a medida que aumentaba el número de imágenes de entrenamiento.

En otro enfoque, Abdullah Alkhalidi y T. Halawani [61] la propuesta del algoritmo es desarrollar un nuevo modelo llamado GOFED-RBVSC (Grasshopper Optimization with Fuzzy Edge Detection based Retinal Blood Vessel Segmentation and Classification) que combina la optimización por saltamontes, detección de bordes difusos, extracción de características ORB y un codificador-decodificador variacional condicional mejorado (ICAVE) para la segmentación y clasificación de vasos sanguíneos en imágenes de fondo retinal, previamente de preprocesamiento se utiliza la mejora de contraste de las imágenes. Para este estudio se utilizó un conjunto de datos de referencia de detección de retinopatía diabética disponible en Kaggle, que contiene imágenes clasificadas en cinco clases distintas. El artículo no proporciona detalles específicos sobre el recurso computacional en términos de tiempo de procesamiento o recursos de hardware utilizados. El modelo GOFED-RBVSC mostró un rendimiento superior en comparación con otros enfoques como ResNet-101, VGG-16 y VGG-19,el estudio de interés superó significativamente en precisión, sensibilidad y especificidad.En el conjunto de prueba (30% de los datos), alcanzó una precisión promedio del 97.92%, lo que indica que el modelo es muy efectivo para clasificar las imágenes de

fondo retinal, el area de interes son las imagenes de ojo sano, por lo tanto cabe denotar que la precisión tuvo un resultado de 97.92 %, sensibilidad del 76.15%, y especificidad del 97.31%, esto quiere decir el modelo es particularmente efectivo en la segmentación y clasificación de vasos sanguíneos retinianos .

De acuerdo a Zhang et al. [62] la propuesta de arquitectura para la segmentación automática de vasos sanguíneos retinianos se basa en un método de aprendizaje supervisado que combina un modelo de fusión de características multi-escala donde se extraen características mediante un operador de línea en una forma de multiresolución y la respuesta del filtro B-COSFIRE para aumentar el contraste de las estructuras vasculares y las áreas no vasculares, finalmente se utiliza un clasificador de máquina de soporte (SVM) para entrenar el modelo con las características extraídas de las imágenes retinianas para obtener la imagen de segmentación final. Se utilizan conjuntos de datos públicos para la evaluación, específicamente el conjunto de datos DRIVE (20 imágenes para entrenamiento y 20 imágenes para prueba) y STARE (14 imágenes para el entrenamiento y 6 imágenes para la prueba). El artículo no proporciona detalles específicos sobre el recurso computacional en términos de tiempo de procesamiento o recursos de hardware utilizados. En el conjunto DRIVE, se obtuvo una precisión de 0.8657, una sensibilidad de 0.7088, una especificidad de 0.9900 y una exactitud de 0.9660. Por otro lado, en el conjunto STARE, los resultados fueron ligeramente superiores en precisión con 0.8782, pero la sensibilidad fue más baja, alcanzando 0.6189, mientras que la especificidad fue de 0.9908 y la exactitud de 0.9494. Estos resultados indican un buen rendimiento general del método, aunque se observa una menor sensibilidad en el conjunto STARE, lo que sugiere áreas de mejora en la detección de vasos más delgados.

Finalmente, Zhou et al. [63] proponen el algoritmo AutoMorph utiliza la arquitectura EfficientNet-b4 en el módulo de clasificación de calidad de imagen. Además, emplea una red de segmentación adversarial para la segmentación de vasos binarios y una red de fusión de información para la segmentación de arterias/venas. Para la segmentación del disco óptico, se utiliza una red de aprendizaje profundo de coarse-to-fine. Se utilizaron varias bases de datos públicas para el desarrollo y validación externa de los modelos de aprendizaje profundo, para la clasificación de calidad de imagen EyePACS-Q, para la segmentación de vasos binarios y la segmentación de arterias/venas DRIVE, STARE, CHASEDB1, HRF, IOSTAR, LES-AV, para la validación externa y la segmentación del disco óptico REFUGE, GAMMA e IDRID. AutoMorph es compatible con Google Colaboratory, lo que permite su uso en entornos de computación en la nube con GPU gratuitas, sin embargo, el tiempo promedio de procesamiento para una imagen es de aproximadamente 20 segundos utilizando una tarjeta gráfica Tesla T4. Las comparaciones con otros métodos muestran que AutoMorph tiene un rendimiento competitivo, en la clasificación de calidad de imagen se reporta un F1-score de 0.86 en el conjunto de datos EyePACS-Q, así como la sensibilidad (0.85) y la especificidad (0.93) altos, lo que indica que el modelo tiene un buen equilibrio entre precisión y sensibilidad, o que significa que el modelo es efectivo para identificar correctamente las imágenes de calidad. También un AUC-ROC (0.97). Para la segmentación binaria de vasos, indica un rendimiento sólido al tener el F1-score es de 0.73 en AV-WIDE y 0.78 en DR HAGIS. en la segmentación de Arterias/Venas el modelo muestra un rendimiento moderado ya que el F1-score de 0.66 en el conjunto IOSTAR-AV, una sensibilidad de 0.64 y especificidad de 0.98 indica que funciona mayormente para evitar falsos positivos. Finalmente, para la segmentación del disco óptico con un F1-score de 0.94 en IDRID, el modelo demuestra una alta precisión en la segmentación del disco óptico.

En síntesis, los estudios revisados demuestran una amplia variedad de estrategias en el uso de redes neuronales convolucionales para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Las diferencias entre enfoques radica tanto en las arquitecturas empleadas como en los métodos de preprocesamiento, los descriptores utilizados y el contexto clínico de aplicación. Para facilitar la comparación entre estos trabajos, se presentan a continuación en la Tabla 4.2 y la Tabla 4.3, donde se encuentra una comparación sistemática entre las distintas propuestas y servirán como base para la selección de modelos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen (D)	Recurso computacional
Yap y Ng [57]	CAMContrast	Normalización de mapas de calor, supresión de fondos y activaciones negativas	Generación de mapas de calor mediante Class Activation Maps (CAM)	Características basadas en densidad, forma y topología	Dos GPUs NVIDIA V100, cada una con 16 GB de memoria
Chen et al. [58]	Shape Regularization Extractor (WT-PSE)	Transformación de Blanqueo (WT)	Destilación de conocimiento guiada por distancia de Wasserstein	Características basadas en forma	Entrenado con PyTorch en una GPU NVIDIA 3090
Aurangzeb et al. [59]	ColonSegNet V2	CLAHE (Ecuación de histograma adaptativa limitada por contraste)	SVM y Naïve Bayes	Características basadas en morfología	No reportado
Morís et al. [60]	Context Encoder	PW-CE (por parches), GM-CE (CB) (tablero de ajedrez) y GM-CE (CS) (centro-periferia)	U-Net, codificador-decodificador con conexiones residuales	ORB (FAST orientado y BRIEF rotado)	No reportado
Abdullah Alkhalidi y T. Halawani [61]	GOFED-RBVSC (Optimización de saltamontes con detección de bordes difusos para segmentación y clasificación de vasos retinianos)	CLAHE	ICAVE (Autoencoder condicional variacional mejorado)	No aplica / No reportado	No reportado
Zhang et al. [62]	Fusión de características multiescala	Filtro B-COSFIRE	SVM	Características multiescala e intensidad de líneas	No reportado

Zhou et al. [63]	AutoMorph	Umbralización	EfficientNet-B4	Tortuosidad, dimensión fractal, densidad vascular y calibre de la copa	No reportado
------------------	-----------	---------------	-----------------	--	--------------

Tabla 4.2 Muestra los aspectos técnicos de los modelos del tipo de CNN utilizada, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Yap y Ng [57]	OIA-ODIR, IDRiD-seg, REFUGE-seg, Vessel-seg	- AUC-PR promedio: $66,29 \pm 1,18$ - F1-score (REFUGE): $91,66 \pm 0,06$ - F1-score (vasos): $80,82 \pm 0,05$	No reportado
Chen et al. [58]	FUNDUS	Segmentación de la copa óptica (OC):- Dice Similarity Coefficient (DSC): $83,11$ - Distancia superficial promedio (ASD): $13,04$ Segmentación del disco óptico (OD):- DSC: $93,08$ - ASD: $10,47$	No reportado
Aurangzeb et al. [59]	DRIVE	Sensibilidad: $83,9\%$ Especificidad: $97,9\%$ Exactitud: $96,6\%$	No reportado
Morís et al [60]	CHASE_DB, STARE, DRIVE	CHASE_DB: Sensibilidad: $86,5\%$ Especificidad: $97,9\%$ Exactitud: $97,1\%$ STARE: Sensibilidad: $86,7\%$ Especificidad: $98,1\%$ Exactitud: $97,2\%$ DRIVE: AUC-ROC: $97,94\%$ AUC-PR: $91,17\%$	No reportado
Alkhaldi y Halawani [61]	Kaggle (no especificado)	Sensibilidad: $76,15\%$ Especificidad: $97,31\%$ Exactitud: $97,92\%$	No reportado
Zhang et al. [62]	DRIVE, STARE	DRIVE: Sensibilidad: $70,88\%$ Especificidad: 99%	No reportado

		Precisión: 86,56% Exactitud: 96,66% STARE: Sensibilidad: 61,89% Especificidad: 99,08% Precisión: 87,82% Exactitud: 94,94%	
Zhou et al.[63]	EyePACS-Q, IOSTAR-AV, IDRID	EyePACS-Q (calidad de imagen): Sensibilidad: 85% Especificidad: 93% Precisión: 87% Exactitud: 92% F1-Score: 86% AUC-ROC: 97% IOSTAR-AV (arteria/vena): F1-Score: 66% AUC-ROC: 95% IDRID (disco óptico): F1-Score: 94% AUC-ROC: 95%	No reportado

Tabla 4.3 Sintetiza la información recabada de los modelos CNN empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.2.2 Redes completamente convolucionales (FCN)

Las CNNs y las Redes Totalmente Convolucionales (FCN, por sus siglas en inglés) son muy similares. La diferencia principal radica en la última capa: en las FCN no se utiliza una capa completamente conectada que integre la información. En su lugar, se emplea una red neuronal convolucional para extraer características y se utilizan conexiones de salto para conservar la ubicación espacial de las características a nivel del mapa completo. En cambio, el número final de canales de salida se ajusta mediante capas convolucionales. La principal ventaja de este enfoque es que el modelo no está limitado por una capa totalmente conectada, lo que permite una mayor flexibilidad en el tamaño de entrada. En este modelo, el resultado se obtiene a nivel de píxel, por lo que la red es capaz de realizar tareas de segmentación [64]. A continuación, se muestra un mapa conceptual en la Figura 4.4, donde se resumen las arquitecturas Redes Totalmente Convolucionales (FCN) usadas para segmentar estructuras anatómicas en imágenes de fondo de ojo. También se indica qué parte anatómica segmenta cada una.

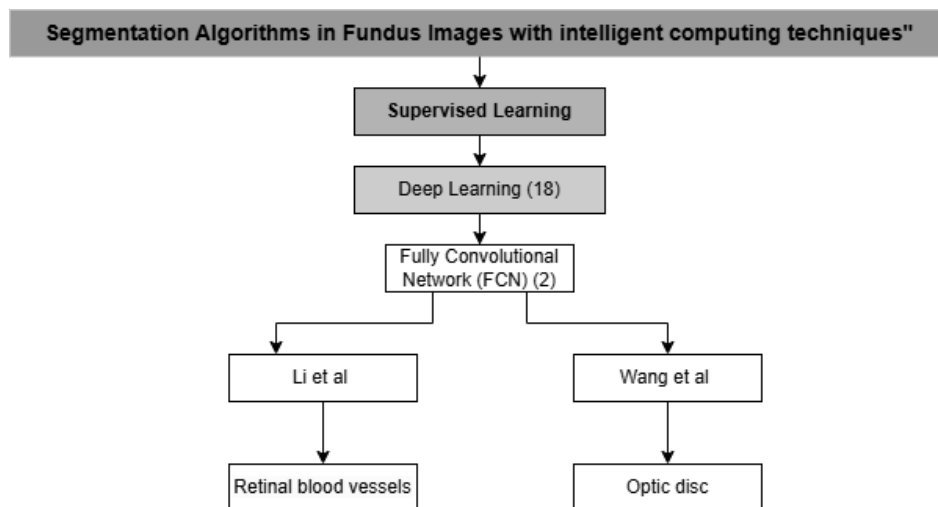


Figura 4.4. Mapa conceptual de redes totalmente convolucionales (FCN) aplicadas a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron varias implementaciones basadas en Redes Totalmente Convolucionales (FCN), cada una con enfoques particulares para la segmentación automática a nivel de píxel. A continuación, se presentan los trabajos revisados, destacando sus hallazgos más relevantes y la manera en que aplican esta arquitectura para segmentar estructuras anatómicas en imágenes de fondo de ojo.

De acuerdo a Li et al. [65] la propuesta es utilizar una red neuronal convolucional completamente convolucional (FCN) para la segmentación de imágenes de fondo de ojo, incluye una parte de convolución y una parte de deconvolución. La parte de convolución consta de tres redes neuronales convolucionales profundas (DCNN) paralelas, mientras que la parte de deconvolución realiza un up-sampling del mapa de características para restaurar el tamaño de la imagen original, esto combina una imagen en escala de grises preprocesada y la información de bordes procesada por operadores Sobel como entradas que permite mejorar la precisión de la segmentación al aprovechar características de diferentes fuentes con un clasificador softmax para la clasificación de píxeles, y se incorpora una capa de abandono (dropout) para prevenir el sobreajuste. La base de datos utilizada consiste en imágenes de fondo de ojo de 5,620 pacientes del Hospital Tongren, clasificadas por oftalmólogos, abarcan cuatro tipos de cataratas: normal, catarata leve, catarata moderada y catarata severa, representadas como grados 0, 1, 2 y 3, respectivamente. Sin embargo, solamente se toman en cuenta aquellas que se encuentran en grado 0 o normales. El entorno de hardware utilizado para el algoritmo de reconocimiento de vasos de imágenes de fondo de ojo incluye un CPU E5-2609, 8GB de RAM, una GPU Quadro K620 y el sistema operativo Ubuntu 16. El preprocesamiento de imágenes se implementó con la caja de herramientas de procesamiento de imágenes en Matlab, y se utilizó Caffe para construir el clasificador automático con la FCN. Hablando del rendimiento de extracción la precisión de 0.9491, la sensibilidad de 0.7621 y la especificidad de 0.9514 en imágenes normales (grado 0) refleja de la misma manera que la extracción de vasos entre más severa es la catarata más impreciso es el modelo. Finalmente, en precisión de segmentación de imágenes de fondo de ojo entre este estudio y otros es efectivo para la extracción de vasos en imágenes de fondo de ojo, especialmente en imágenes normales al tener unas métricas comparables.

A su vez, Wang et al. [66] la propuesta es un método de detección de disco óptico débilmente supervisado basado en una red neuronal convolucional (FCN), con G'generación de la matriz de características de la imagen, integrar la información semántica de alto nivel aprendida por la FCN con el conocimiento previo del fondo de la región del disco óptico para la detección del disco óptico y la decomposición de la matriz de características utilizando una matriz de ponderación construida a partir de la información previa. Se utilizan dos conjuntos de datos públicos para evaluar el método, DRISHTI-GS (50 imágenes de entrenamiento y 51 de prueba) e IDRiD (54 imágenes de entrenamiento y 27 de prueba). Este método se ejecuta en un sistema operativo Windows 10 con un entorno de hardware que incluye un CPU Intel Xeon y una GPU NVIDIA Quadro P400, con una velocidad de reloj de CPU de 4GHz y 32 GB de RAM. Los resultados en el conjunto de datos DRISHTI-GS en el método propuesto tiene el mayor valor de IoU (0.918), un recall de 0.968, un F-1 score de 0.955 y un MAE de 0.072 lo que indica que su detección del OD es más precisa en comparación con los otros métodos, tiene una mayor precisión en la detección a nivel de píxel, por lo tanto se infiere que es un método muy robusto. Por otro lado, los resultados en el conjunto de datos IDRiD en el método propuesto también supera a los demás en este conjunto de datos con un IoU de 0.872, un recall de 0.942, un F-1 score de 0.932 y un MAE de 0.106 indican que el método propuesto tiene un alto nivel de exactitud en sus predicciones, hay una detección de la mayoría de los discos ópticos presentes en las imágenes y tiene un rendimiento superior en la detección precisa del disco óptico.

En síntesis, los estudios revisados demuestran una amplia variedad de estrategias en el uso de Redes Totalmente Convolucionales (FCN) para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Los enfoques difieren en las arquitecturas empleadas, así como en los métodos de preprocesamiento, los descriptores utilizados y el contexto clínico de aplicación. Para facilitar la comparación entre estos trabajos, se presentan a continuación en la Tabla 4.4 y la Tabla 4.5 una comparación sistemática entre las distintas propuestas, la cual servirá de base para la selección de modelos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen	Recurso computacional
Li et al. [65]	Red Neuronal Totalmente Convolucional (FCN) con enfoque de fusión dual	Extracción de canales (RGB) y eualización de histograma	No reportado	Conjunto de características heurísticas no manuales	CPU E5-2609, 8 GB RAM, GPU Quadro K620, Ubuntu 16
Wang et al.[66]	FCN combinada con modelo WLRR (recuperación de matriz de bajo rango)	Extracción de canales (RGB), filtros de Gabor y SLIC (agrupamiento iterativo simple)	No reportado	Características de color, borde y textura	CPU Intel Xeon, GPU NVIDIA Quadro P400, 4 GHz, 32 GB RAM

Tabla 4.4 Muestra los aspectos técnicos de los modelos del tipo de FCN utilizada, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Li et al. [65]	Pacientes del Hospital Tongren	Sensibilidad: 72.15% Especificidad: 95.76% Exactitud: 92.10%	Precisión de extracción: Sensibilidad: 76.21% Especificidad: 95.14% Exactitud: 94.91%
Wang et al. [66]	DRISHTI-GS	IoU: 91.8% Precisión: 93.8% Recall: 96.8% F1 Score: 95.5% MAE: 7.2%	No reportado

Tabla 4.5 Sintetiza la información recabada de los modelos FCN empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.2.3 U-Net

La estructura básica de una arquitectura U-Net consta de dos rutas. En la primera ruta se realiza una contracción, También se le conoce como el codificador o la ruta de análisis, que es similar a una red de convolución regular y proporciona información de clasificación. La segunda se le conoce como ruta de expansión, también se le conoce como el decodificador o la ruta de síntesis, que consiste en convoluciones ascendentes y concatenaciones con características de la ruta de contracción. Esta expansión permite que la red aprenda información de clasificación localizada. Además, la ruta de expansión también aumenta la resolución de la salida, que luego puede pasar a una capa convolucional final para crear una imagen completamente segmentada. Lo interesante de este modelo de segmentación es que la red resultante es casi simétrica, lo que le da una forma similar a una U. La principal tarea canónica realizada por la mayoría de las redes convolucionales es clasificar toda la imagen en una sola etiqueta. U-net es particularmente útil al momento de la creación de mapas de segmentación altamente detallados utilizando muestras comerciales muy limitadas lo cual la hace fundamental en el ámbito de las imágenes médicas, ya que las imágenes correctamente etiquetadas suelen ser limitadas. Esto se logra mediante el uso de deformación elástica aleatoria en los datos de entrenamiento, lo que permite a la red aprender estas variaciones sin necesidad de nuevos datos etiquetados [67]. A continuación, se muestra un mapa conceptual en la Figura 4.5, donde se resumen las arquitecturas basadas en U-Net utilizadas para segmentar estructuras anatómicas en imágenes de fondo de ojo. También se indica qué parte anatómica segmenta cada una.

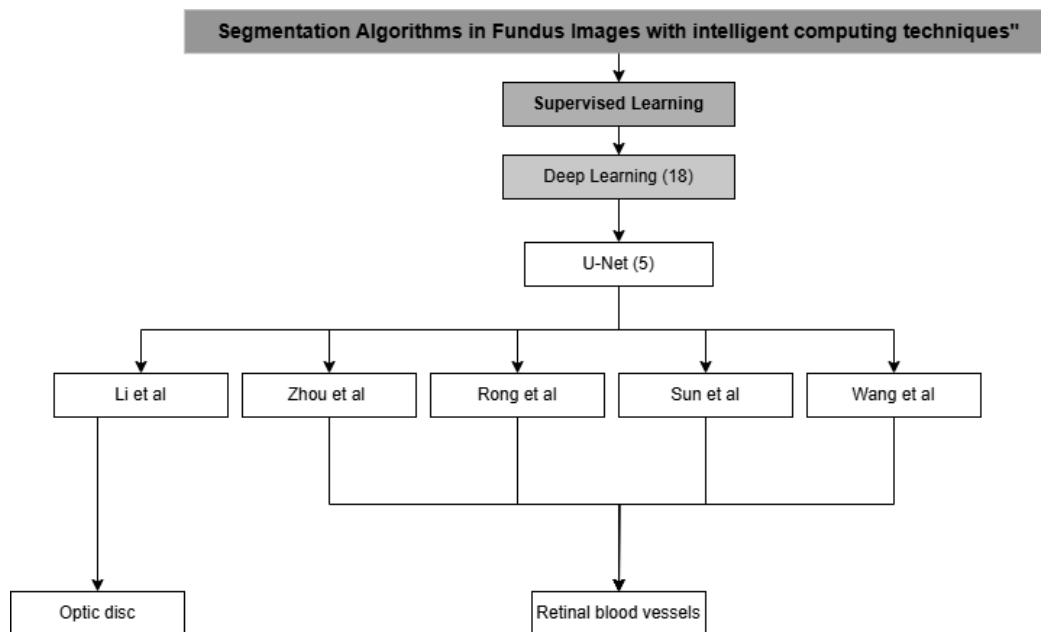


Figura 4.5 Mapa conceptual de U-Nets aplicadas a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron múltiples implementaciones basadas en la arquitectura U-Net, cada una con enfoques particulares para la segmentación automática a nivel de píxel. A continuación, se presentan los trabajos revisados, destacando sus hallazgos más relevantes y la manera en que aplican esta arquitectura para segmentar estructuras anatómicas en imágenes de fondo de ojo.

Tal como se indica en Zhou et al. [68] el algoritmo propuesto, CF-Loss, incorpora características vasculares relevantes clínicamente como funciones de pérdida en un enfoque de extremo a extremo. Se centra en dos características específicas: la densidad de vasos y la dimensión fractal. El algoritmo mide la densidad de vasos y utiliza un conteo de cajas suave para optimizar la segmentación de los vasos en las imágenes de fondo de retina. Se utilizan dos arquitecturas de red como ejemplos para la segmentación de vasos: U-Net y BF-Net. Estas arquitecturas se emplean para evaluar el rendimiento del CF-Loss en la segmentación de vasos multi-clase. Se utilizan tres conjuntos de datos públicos para verificar el rendimiento de la segmentación de vasos multi-clase y la medición de características, DRIVE-AV: 40 fotografías de fondo de ojo, con 20 imágenes para entrenamiento y 20 para prueba, LES-AV: 22 imágenes, con 11 para entrenamiento y 11 para prueba y HRF-AV: 45 imágenes, con 24 para entrenamiento y 21 para prueba. El código se implementa utilizando Pytorch 1.9 y se ejecuta en una GPU Tesla T4 (16 GB). Se realizan operaciones de aumento de datos, como rotación y mejora de color, durante el entrenamiento. En el análisis, BF-Net mostró un desempeño general superior, alcanzando una sensibilidad de hasta 72.91% en el conjunto de datos DRIVE-AV, en comparación con el 71.37% de U-Net. Aunque, en cuanto al F1-score, U-Net tuvo una ligera ventaja al alcanzar un 73.22%, en comparación con el 73.04% de BF-Net en el conjunto de datos DRIVE-AV nuevamente. Respecto al acuerdo de las características vasculares, el Coeficiente de Correlación Intraclass (ICC)

fue más alto en el conjunto LES-AV, con un 0.92 para la dimensión fractal y un 0.95 para la densidad de vasos.

Por su parte, Li et al. [69] se propone TUNet, una red híbrida basada en CNN y Transformer con una arquitectura en forma de U. Esta red aprovecha las características contextuales globales codificadas por Transformers y la información espacial detallada extraída por U-Net, es decir, extraer las regiones de interés de la imagen de fondo original y utilizar imágenes del dominio fuente y del dominio objetivo para el aprendizaje invariante al dominio. Además, se introduce un clasificador auxiliar para mejorar la discriminación en los detalles de segmentación. Se realizaron experimentos en tres conjuntos de datos públicos de imágenes de fondo retinal: REFUGE, Drishti-GS y RIM-ONE-r3. Los experimentos se llevaron a cabo en un servidor equipado con 2 GPUs Nvidia 2080Ti. Se utilizó el código base de segmentación para implementar el modelo, y se aplicaron técnicas de aumento de datos durante el proceso de entrenamiento. Los resultados de segmentación de disco óptico y copa óptica en los conjuntos de datos Drishti-GS y RIM-ONE-r3 muestran un rendimiento superior en ambas métricas, lo que indica una mejor precisión en la segmentación de las regiones ya que logra un Coeficiente de Dice (DC) de 0.9025 para la copa y 0.9727 para el disco en Drishti-GS, y un Coeficiente de Dice (DC) de 0.8618 para la copa y 0.9690 para el disco en RIM-ONE-r3, hablando de el rendimiento destacado en comparación con REFUGE, el modelo alcanza un DC de 0.9006 para la copa y 0.9613 para el disco, con un δ de 0.0337, se reporta que el modelo propuesto (TUNet) logra un rendimiento superior en ambas las métricas CDR (Cup-to-Disc Ratio) y RADR (Rim-to-Disc Area Ratio) en comparación con otros métodos.

De acuerdo a Rong et al. [70] la propuesta se basa en la modificación de la arquitectura U-Net, incorporando convoluciones auto-calibradas (SCC) permiten que la red aprenda representaciones desde diferentes campos receptivos al separar los filtros de convolución en múltiples porciones y calibrar la transformación de convolución de otro filtro utilizando información de características de baja dimensión para reemplazar las convoluciones tradicionales y un módulo de atención espacial mejorado (ISAM) se utiliza para conectar las partes de codificación y decodificación de la U-Net, preservando información más completa sobre los vasos sanguíneos, especialmente los delgados. Esto permite que la U-Net aprenda representaciones discriminativas desde diferentes campos receptivos. Se realizaron pruebas en dos bases de datos públicas para la segmentación de vasos retinianos, DRIVE (20 se utilizaron para entrenamiento y 20 para prueba) y CHASE DB1 (20 se utilizaron para entrenamiento y 8 para prueba). El entrenamiento de las redes se realizó en una computadora equipada con una GPU NVIDIA TITAN Xp con 12GB de memoria, utilizando los marcos Keras y TensorFlow. Se utilizó el descenso de gradiente estocástico para minimizar la función de pérdida. Se comparan el rendimiento del modelo propuesto (U-Net+SCC+ISAM) con otros métodos de segmentación de vasos retinianos en los conjuntos de datos DRIVE y CHASE DB1, la combinación propuesta en el método de SCC y ISAM proporciona el mejor rendimiento, con ACC de 0.9680, SE de 0.8036, SP de 0.9840, F1 de 0.8138 y AUC de 0.9840 en DRIVE y también ofrece el mejor rendimiento en este conjunto de datos, con ACC de 0.9756, SE de 0.8118, SP de 0.9867, F1 de 0.8068 y AUC de 0.9888 para CHASE DB1, Esto indica que las modificaciones propuestas son efectivas para la segmentación de vasos.

De acuerdo a Sun et al. [71] se propone un método de segmentación de vasos retinianos basado en un modelo mejorado de U-Net, denominado SDAU-Net, la cual es una red de codificador-decodificador que reemplaza el módulo de convolución en U-Net por SDC. Este modelo incorpora una estructura de

convolución deformable en serie (SDC) y mecanismos de atención (LAM y DAM) en cada capa de convolución del módulo decodificador para mejorar la extracción de características de los vasos sanguíneos con morfología compleja y mejorar la segmentación en imágenes con bajo contraste. Se utilizaron cuatro conjuntos de datos para la evaluación del modelo, DRIVE (20 imágenes para entrenamiento y 20 para prueba), STARE (10 con lesiones y 10 sin lesiones), CHASE DB1 (28 imágenes de fondo de ojo) e IOSTAR: (30 imágenes mediante la técnica de tomografía de coherencia óptica, por lo tanto esta base de datos no se tomará en cuenta). Los experimentos se realizaron utilizando las plataformas Tensorflow y Keras, en un CPU Intel Xeon E5-2678 V3 y una GPU NVIDIA GeForce RTX 2080Ti. Existe una comparación con otros métodos existentes, sin embargo, los resultados indican que SDAU-Net tiene una buena capacidad de generalización y mantiene un rendimiento competitivo, en la capacidad del modelo para identificar correctamente los vasos sanguíneos, STARE con 0.8973 es la base de datos que es más sobresaliente, también en la capacidad del modelo para identificar correctamente el fondo, en la proporción de predicciones correctas sobre el total de predicciones y en la capacidad del modelo para distinguir entre clases.

Finalmente, en Wang et al. [72] se propone el modelo Context Spatial U-Net (CSU-Net) para la segmentación de vasos sanguíneos en imágenes de fondo de ojo. Este modelo cuenta con un codificador de dos canales: un canal de contexto que utiliza convoluciones a múltiples escalas para capturar un campo receptivo más amplio y un canal espacial que utiliza un kernel grande para retener información espacial. Además, se incorpora un módulo de fusión de características (FFM) y un módulo de atención (ASM) para mejorar la representación de características y reducir la redundancia. El modelo CSU-Net se entrena y evalúa en tres conjuntos de datos, DRIVE (20 imágenes para entrenamiento y 20 para prueba), STARE (10 imágenes para entrenamiento y 10 para prueba), CHASE DB1 (20 imágenes para entrenamiento y 8 para prueba). El modelo se implementa utilizando el marco de trabajo Pytorch en una plataforma Ubuntu 16.04 (64 bits) equipada con una GPU NVIDIA GTX 2080Ti. Durante la etapa de entrenamiento, se utiliza el optimizador Adam con un tamaño de minibatch de 16 y se aplican estrategias de ajuste de tasa de aprendizaje. El modelo mantiene un rendimiento alto en imágenes normales, en el caso de DRIVE con un F1-score 0.8263, una sensibilidad de 0.8088, una especificidad de 0.9783, una exactitud de 0.9568 y un AUC de 0.9798, así como en la base de datos STARE un F1-score 0.8696, una sensibilidad de 0.8467, una especificidad de 0.9889, una exactitud de 0.9748 y un AUC de 0.9828, esto sugiere que es capaz de manejar variaciones en la calidad de las imágenes, también puede ser empleado para detección de enfermedades oculares.

En síntesis, los estudios revisados evidencian una amplia adopción de arquitecturas basadas en U-Net para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Las propuestas varían en términos de modificaciones arquitectónicas, técnicas de preprocesamiento, descriptores empleados y escenarios clínicos de aplicación. Para facilitar la comparación entre estos trabajos, se presentan a continuación en la Tabla 4.6 y la Tabla 4.7 una comparación sistemática entre las distintas propuestas, la cual servirá de base para la selección de modelos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen	Recurso computacional
Zhou et al. [68]	U-Net, BF-Net	No reportado	Análisis discriminante lineal (LDA), k vecinos más cercanos (kNN) y bosque aleatorio (Random Forest)	Características basadas en densidad, forma, topología y heurísticas	PyTorch 1.9, ejecutado en una GPU Tesla T4 (16GB)
Li et al. [69]	Red híbrida basada en CNN y Transformer	Extracción de regiones de interés (ROI)	Red ResNet-18	No reportado	2 GPUs Nvidia 2080Ti
Rong et al. [70]	Convoluciones auto-calibradas (SCC)	No reportado	Módulo de atención espacial mejorado (ISAM)	Características jerárquicas	GPU NVIDIA TITAN Xp con 12GB
Sun et al. [71]	SDAU-Net	Fusión de canales rojo-verde, ecualización de histograma y CLAHE (Ecuación de Histograma con Limitación de Contraste)	No reportado	Convoluciones deformables en serie (SDC), módulos de atención ligera (LAM) y módulos de doble atención (DAM)	TensorFlow y Keras, en CPU Intel Xeon e5-2678 V3 y GPU NVIDIA GeForce RTX 2080Ti
Wang et al. [72]	Context Spatial U-Net	CLAHE (Ecuación de Histograma con Limitación de Contraste)	Módulo de atención espacial (SAM) y módulo de fusión de características (FFM)	ORB (FAST orientado) y BRIEF rotado)	PyTorch sobre Ubuntu 16.04 (64-bit), equipado con GPU NVIDIA GTX 2080Ti

Tabla 4.6. Muestra los aspectos técnicos de los modelos del tipo de U-Net utilizada, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Zhou et al. [68]	DRIVE-AV	U-Net: Sensibilidad: $71.37 \pm 0.75\%$, F1-score: $73.22 \pm 0.98\%$, IOU: $58.25 \pm 1.21\%$, MSE: $2.85 \pm 0.01\%$, Error de Betti: $7.92 \pm 1.02\%$ BF-Net: Sensibilidad: $72.91 \pm 1.27\%$, F1-score: $73.04 \pm 0.58\%$, IOU: $57.99 \pm 0.70\%$, MSE: $2.93 \pm 0.06\%$, Error de Betti: $7.75 \pm 1.21\%$	ICC U-Net: Dim. fractal: 0.78 (0.45–0.92), Densidad vascular: 0.72 (0.36–0.93)BF-Net: Dim. fractal: 0.84 (0.64–0.92), Densidad vascular: 0.78 (0.56–0.92)
	LES-AV	U-Net: Sensibilidad: $62.21 \pm 2.14\%$, F1-score: $65.93 \pm 1.32\%$, IOU: $50.66 \pm 1.51\%$, MSE: $2.61 \pm 0.25\%$, Error de Betti: $4.76 \pm 1.15\%$ BF-Net: Sensibilidad: $67.06 \pm 1.76\%$, F1-score: $69.87 \pm 1.56\%$, IOU: $54.98 \pm 1.61\%$, MSE: $2.32 \pm 0.10\%$, Error de Betti: $3.04 \pm 0.66\%$	ICC U-Net: Dim. fractal: 0.72 (0.33–0.94), Densidad vascular: 0.71 (0.32–0.90)BF-Net: Dim. fractal: 0.92 (0.83–0.97), Densidad vascular: 0.95 (0.88–0.98)
	HRF-AV	U-Net: Sensibilidad: $69.41 \pm 1.75\%$, F1-score: $72.17 \pm 0.66\%$, IOU: $57.74 \pm 0.73\%$, MSE: $1.91 \pm 0.02\%$, Error de Betti: $6.61 \pm 0.52\%$ BF-Net: Sensibilidad: $67.61 \pm 2.48\%$, F1-score: $71.19 \pm 0.58\%$, IOU: $56.48 \pm 0.73\%$, MSE: $1.96 \pm 0.03\%$, Error de Betti: $6.82 \pm 1.16\%$	ICC U-Net: Dim. fractal: 0.72 (0.33–0.94), Densidad vascular: 0.71 (0.32–0.90)BF-Net: Dim. fractal: 0.86 (0.71–0.96), Densidad vascular: 0.91 (0.82–0.96)
Li et al. [69]	REFUGE	CDR: 0.9639, RADR: 0.9639,	Copa disco: 0.9006 Disco óptico: 0.9613 Error CDR: 0.0337
	Drishti-GS	CDR: 0.9286, RADR: 0.9286,	Copa disco: 0.9025, Disco óptico: 0.9727 Error CDR: 0.0428
	RIM-ONE-r3	CDR: 0.8125, RADR: 0.7500	Copa disco: 0.8618 Disco óptico: 0.9690 Error CDR: 0.0315
Rong et al. [70]	DRIVE-AV	Sensibilidad: 80.36%, Especificidad: 98.40%, Exactitud: 96.80%, F1-score: 81.38%, AUC-ROC: 98.40%	No reportado
	CHASE DB1	Sensibilidad: 81.18%, Especificidad: 98.67%, Exactitud: 97.56%, F1-score: 80.68%, AUC-ROC: 98.88%	No reportado
Sun et al. [71]	DRIVE-AV	Sensibilidad: 79.55%, Especificidad: 98.48%, Exactitud: 96.82%, AUC-ROC: 98.34%	No reportado
	CHASE DB1	Sensibilidad: 83.21%, Especificidad: 98.25%, Exactitud: 97.32%, AUC-ROC: 98.58%	No reportado

	STARE	Sensibilidad: 89.73%, Especificidad: 99.03%, Exactitud: 98.33%, AUC-ROC: 99.63%	No reportado
Wang et al. [72]	DRIVE-AV	Sensibilidad: 80.71%, Especificidad: 97.82%, Exactitud: 95.65%, F1-score: 82.51%, AUC-ROC: 98.01%	No reportado

Tabla 4.7 Sintetiza la información recabada de los modelos U-Net empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.2.4 Otras arquitecturas

En la última década se le ha dado enfoque a modelos que combinan múltiples paradigmas del aprendizaje automático y profundo, integrando elementos como redes convolucionales, mecanismos de atención, redes basadas en grafos y módulos de aprendizaje clásico por mencionar algunos. A estas se les denomina como arquitecturas híbridas que no encajan completamente en las categorías mencionadas en esta sección de la tesis, por lo tanto se les ha dado un apartado. Sin embargo de igual manera que los apartados anteriores su diseño busca aprovechar lo mejor de cada enfoque para abordar desafíos específicos en la segmentación de imágenes de fondo de ojo, como la preservación de estructuras finas, la integración de contexto global y la mejora de la robustez ante variaciones de contraste y calidad de imagen[73]. A continuación, se muestra un mapa conceptual en la Figura 4.6, donde se resumen las arquitecturas híbridas utilizadas para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Estas arquitecturas combinan distintos enfoques de aprendizaje profundo y tradicional, con el objetivo de mejorar la precisión y la generalización del modelo. Asimismo, se indica la estructura anatómica que cada una de ellas segmenta, lo que permite visualizar cómo estas soluciones multidisciplinarias abordan los desafíos específicos de cada tarea.

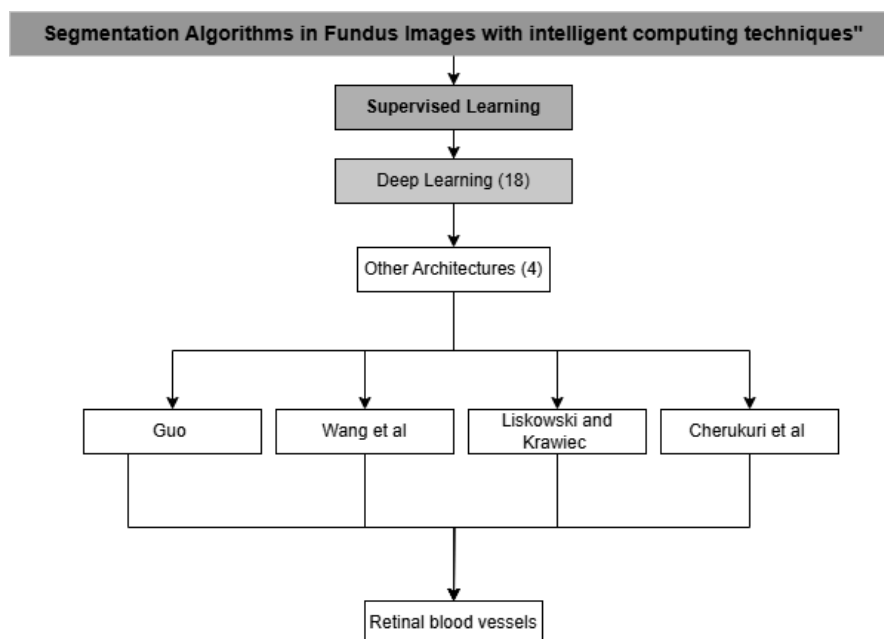


Figura 4.6 Mapa conceptual de modelos con otras arquitecturas aplicadas a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron diversas implementaciones basadas en arquitecturas híbridas, que combinan elementos de distintas técnicas de aprendizaje profundo y machine learning. Estas soluciones adoptan enfoques variados para la segmentación automática y no encajan estrictamente en categorías convencionales. A continuación, se presentan los trabajos revisados, destacando sus hallazgos más relevantes y cómo aplican estas arquitecturas híbridas para segmentar estructuras anatómicas en imágenes de fondo de ojo.

De acuerdo a Guo [74] se propone una red llamada CSGNet (Cascade Semantic Guided Net) que sigue una arquitectura de bajo a alto. Esta red consta de tres ramas que procesan características a diferentes escalas (1/4, 1/2 y 1). La arquitectura incluye un módulo de aprendizaje de características multi-escala y multi-direccional (M2FL) que utiliza convoluciones en tiras y convoluciones dilatadas para alinear mejor con la forma de los vasos y codificar información contextual, se incorpora un módulo guiado por semántica (SGM) para mejorar el aprendizaje de características de alta resolución. Se realizaron experimentos utilizando cuatro conjuntos de datos públicos, DRIVE, STARE, CHASE DB1 y HRF, enfocados en fondo de ojo. Los experimentos se llevaron a cabo utilizando el marco Caffe en una GPU NVIDIA RTX 3090. Se utilizó la inicialización de pesos de Xavier y el optimizador Adam. El tamaño del lote se estableció en 1, y se realizaron diferentes configuraciones de tasa de aprendizaje y número de iteraciones según el conjunto de datos. El estudio incluye un enfoque de aprendizaje de características de bajo a alto y el uso de módulos M2FL, contribuye a su rendimiento superior, especialmente en la identificación de vasos delgados y de bajo contraste, las métricas se calcularon tanto con máscaras de campo de visión (FOV) como sin ellas, para proporcionar una evaluación más completa del rendimiento del modelo ya que ayudan a mejorar la precisión de las métricas de evaluación al asegurarse de que los cálculos se realicen solo dentro de la región donde se espera encontrar los vasos. Generalmente, CSGNet ha demostrado un rendimiento comparable o superior en comparación con otros modelos de segmentación ya que mostró valores de Sen, Spe, Acc, AUC y F1-score que son consistentemente altos en todos los conjuntos de datos.

Según lo planteado en Wang et al. [75] se propone una nueva arquitectura de aprendizaje profundo llamada Hard Attention Net (HANet) para la segmentación automática de vasos retinianos. HANet está compuesta por una red de codificación y tres redes de decodificación. La primera red de decodificación genera un resultado de segmentación de vasos "grosero" los cuales han sido segmentados de manera inicial y que pueden requerir un refinamiento adicional, que se utiliza para identificar dinámicamente las regiones de la imagen que son "difíciles" o "fáciles" de segmentar. Las otras dos redes de decodificación están diseñadas específicamente para segmentar las regiones "difíciles" y "fáciles". Además, se introducen mecanismos de atención en HANet para reforzar las características de imagen en las regiones "difíciles". Se realizaron experimentos utilizando cuatro conjuntos de datos públicos, DRIVE, STARE, CHASE DB1 y HRF, enfocados en fondo de ojo. HANet supera o se compara favorablemente con otros modelos de segmentación de vasos retinianos en los múltiples conjuntos de datos, la implementación de decodificadores separados para regiones "difíciles" y "fáciles", junto con mecanismos de atención, permite al modelo enfocarse en características relevantes y mejorar la segmentación de vasos delgados y de difícil detección.

Después, Liskowski and Krawiec [76] donde el algoritmo propuesto se basa en redes neuronales profundas (DNN) que realizan la segmentación de vasos sanguíneos en imágenes de fondo ocular. Se

utiliza un enfoque de aprendizaje supervisado, donde la red se entrena con ejemplos de parches de imágenes etiquetados, de las cuales se utilizaron dos configuraciones principales: PLAIN y NO-POOL. La arquitectura PLAIN incluye capas de max-pooling, mientras que NO-POOL no las utiliza. Ambas arquitecturas están compuestas por múltiples capas de convolución seguidas de capas completamente conectadas. Se evaluaron también configuraciones adicionales como GCN, ZCA, AUGMENT y BALANCED, que se basan en la arquitectura PLAIN para mejorar la precisión de la segmentación al considerar las relaciones espaciales entre píxeles. Se realizaron experimentos utilizando tres conjuntos de datos públicos, DRIVE, STARE y CHASE DB1, sin embargo, a pesar de utilizar los conjuntos mencionados se reporta como el mejor rendimiento aquellos utilizados con la base de datos STARE con un total de 20 imágenes. El entrenamiento de las redes se realizó en configuraciones de hardware que incluyen procesadores Intel Core i7 y tarjetas gráficas NVIDIA GTX Titan y Tesla K20c. Se utilizó el marco de trabajo Caffè para realizar los cálculos en GPU en aritmética de precisión simple. Esta arquitectura CNN con configuración NO-POOL-SP mostró resistencia a artefactos como el reflejo central del vaso. En general, el tamaño de ventana $S=5$ demostró ser el mejor para la segmentación en las bases de datos DRIVE y STARE. Con esta configuración, el modelo logró un balance óptimo entre la detección precisa de los vasos sanguíneos y la reducción de errores de falsos positivos. Esto se debe a que una ventana de salida de $S=5$ permite al modelo capturar de manera efectiva las relaciones espaciales entre píxeles vecinos sin perder precisión en la detección de vasos finos y detalles específicos. Además, el índice de kappa obtenido en las pruebas indica una alta concordancia entre la segmentación automática y el estándar de referencia, validando la efectividad y la precisión del modelo en la detección de estructuras vasculares.

Por su parte, en Cherukuri et al. [77] se propone una arquitectura de red profunda denominada Multi-Scale Regularized Deep Network for Vessel Segmentation (MSR-DNVS). Esta arquitectura consta de dos componentes principales, primeramente una red de representación la cual aprende características geométricas (específicamente curvilíneas) adaptadas a imágenes retinianas y una red de tarea que utiliza las características obtenidas de la capa de representación para realizar la segmentación a nivel de píxel. Se evalúa el MSR-DNVS en dos conjuntos de datos públicos ampliamente utilizados para la segmentación de vasos retinianos, DRIVE (20 imágenes de entrenamiento y 20 de prueba) y STARE donde se repite el ciclo de entrenamiento y prueba 20 veces, reservando una imagen para prueba en cada iteración. En este modelo existe solamente una comparación entre 4 métodos similares y el propuesto, donde en ambos conjuntos de datos la arquitectura propuesta es superior a las demás.

En síntesis, los estudios revisados evidencian una amplia variedad de arquitecturas híbridas aplicadas a la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Estas propuestas se distinguen por combinar diferentes enfoques de aprendizaje profundo y técnicas clásicas, así como por sus estrategias de preprocesamiento, descriptores de imagen y contextos clínicos específicos. Para facilitar la comparación entre estos trabajos, se presentan a continuación en la Tabla 4.8 y la Tabla 4.9 una comparación sistemática entre las distintas propuestas, la cual servirá de base para la selección de modelos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen	Recurso computacional
Guo [74]	CSGNet	Normalización y aumento de datos	Módulo de Aprendizaje Multiescala y Multidireccional (M2FL)	Convoluciones en tiras y convoluciones dilatadas	GPU NVIDIA RTX 3090
Wang et al.[75]	HAnet	Ecualización adaptativa del histograma limitada por contraste y ajuste gamma (gamma = 1.2)	Red codificadora para extracción de características y red decodificadora para mapeo de características	Intensidad de color, respuesta a filtros wavelet, bordes y líneas	PyTorch, una GPU Nvidia GeForce Titan X
Liskowski y Krawiec [76]	Redes Neuronales Profundas (DNN)	Blanqueamiento de Componentes de Fase Cero (ZCA Whitening) y aumento de datos	CNNs para clasificar píxeles y predecir múltiples píxeles simultáneamente	Características aprendidas y extraídas automáticamente mediante capas convolucionales	Procesadores Intel Core i7, GPU NVIDIA GTX Titan y Tesla K20c
Cherukuri et al. [77]	Red Profunda Regularizada Multiescala para Segmentación de Vasos (MSR-DNVS)	Blanqueamiento de Componentes de Fase Cero (ZCA Whitening) y aumento de datos	CNN en capa de representación, capa de tareas y filtros convolucionales a diferentes escalas	Características geométricas curvilíneas y filtros para vasos sanguíneos retinianos	GPU NVIDIA Titan X (12GB), usando TensorFlow

Tabla 4.8 Muestra los aspectos técnicos de los modelos de diversas arquitecturas, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Guo [74]	DRIVE-AV	Sensibilidad: 0.7984 ± 0.0029 Especificidad: 0.9875 ± 0.0003 Exactitud: 0.9709 ± 0.0000 F1-Score: 0.8312 ± 0.0004 AUC-ROC: 0.9881 ± 0.0002	No reportado
	CHASE DB1	Sensibilidad: 0.7945 ± 0.0026 Especificidad: 0.9902 ± 0.0001 Exactitud: 0.9779 ± 0.0001 F1-Score: 0.9923 ± 0.0001 AUC-ROC: 0.8246 ± 0.0009	No reportado
	STARE	Sensibilidad: 0.8298 ± 0.0031 Especificidad: 0.9855 ± 0.0007 Exactitud: 0.9692 ± 0.0005 F1-Score: 0.8493 ± 0.0002 AUC-ROC: 0.9895 ± 0.0006	No reportado
	HRF	Sensibilidad: 0.7828 ± 0.0028 Especificidad: 0.9839 ± 0.0004 Exactitud: 0.9659 ± 0.0002 F1-Score: 0.9905 ± 0.0003 AUC-ROC: 0.8332 ± 0.0021	No reportado
Wang et al. [75]	DRIVE-AV	Sensibilidad: 79.91% Especificidad: 98.13% Exactitud: 95.81% F1-Score: 82.93% AUC-ROC: 98.23%	No reportado
	CHASE DB1	Sensibilidad: 82.39% Especificidad: 98.13% Exactitud: 96.70% F1-Score: 81.91% AUC-ROC: 98.71%	No reportado
	STARE	Sensibilidad: 81.86% Especificidad: 98.44% Exactitud: 96.73% F1-Score: 83.79% AUC-ROC: 98.81%	No reportado
	HRF	Sensibilidad: 78.03% Especificidad: 98.43% Exactitud: 96.54% F1-Score: 80.74% AUC-ROC: 98.54%	No reportado
Liskowski y Krawiec [76]	DRIVE-AV	Sensibilidad: 78.11% Especificidad: 98.07% Exactitud: 95.35% AUC-ROC: 97.90% Kappa: 79.10%	No reportado
	STARE	Sensibilidad: 85.54% Especificidad: 98.62% Exactitud: 97.29% AUC-ROC: 99.28% Kappa: 85.07%	No reportado

Cherukuri et al. [77]	DRIVE-AV	F1-Score: 80.87% Exactitud: 96.95%AUC-ROC: 98.13%	No reportado
-----------------------	----------	--	--------------

Tabla 4.9 Sintetiza la información recabada de los modelos de diversas arquitecturas empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.2.5 Modelos basados en aprendizaje automático

El aprendizaje automático es un área en las ciencias dentro del campo de la inteligencia artificial que se centra en el diseño y desarrollo de algoritmos que permiten a las computadoras aprender patrones y tomar decisiones a partir de datos, sin ser programadas explícitamente para cada tarea. Estos modelos surgen en la intersección de las estadísticas y la informática, con su enfoque en el diseño de algoritmos eficientes y escalables buscando inferir relaciones y estructuras subyacentes en los datos. Ha demostrado ser especialmente poderoso en tareas donde la programación tradicional resulta limitada, como el reconocimiento de patrones, la predicción de comportamientos, la clasificación automática y la toma de decisiones basada en datos, todo esto motivado por los crecientes desafíos que implica construir modelos capaces de aprender a partir de volúmenes masivos de información, que pueden incluir miles de millones o incluso billones de puntos de datos [78]. A continuación, se presenta un mapa conceptual en la Figura 4.7, que resume los enfoques de aprendizaje automático aplicados a la segmentación de estructuras anatómicas en imágenes de fondo de ojo, así como la estructura anatómica específica que se segmenta en cada caso. Este mapa incluye los métodos tradicionales de clasificación y agrupamiento más comúnmente utilizados, junto con sus combinaciones de técnicas de preprocesamiento, extracción de características y tipos de clasificadores empleados.

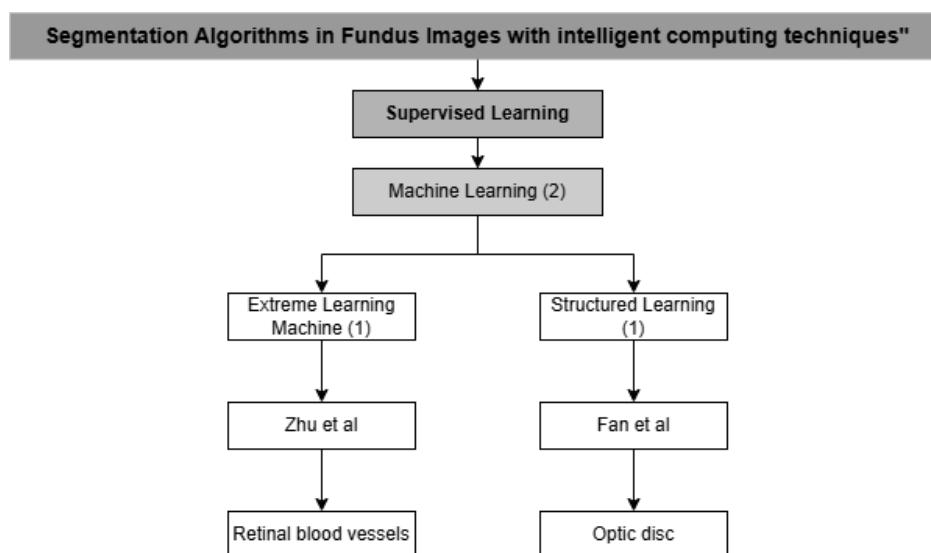


Figura 4.7 Mapa conceptual de modelos de aprendizaje automático aplicados a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron diversas implementaciones basadas en técnicas de aprendizaje automático, cada una con estrategias particulares para abordar la segmentación automática de estructuras anatómicas en imágenes de fondo de ojo. Estas propuestas emplean clasificadores tradicionales, algoritmos de agrupamiento y métodos supervisados, combinados con distintas técnicas de preprocesamiento y extracción de características. A continuación, se presentan los trabajos revisados, resaltando sus hallazgos más relevantes y la manera en que aplican estos enfoques para lograr una segmentación precisa.

Según lo planteado en Zhu et al. [79] se propone un método para la segmentación de vasos retinianos en imágenes de fondo de ojo, el cual incluye componentes de entrenamiento y segmentación, donde se extraen vectores de características para cada píxel de la imagen de entrenamiento donde hay una extracción de un vector de características de 39 dimensiones para cada píxel, que incluye características locales, morfológicas, congruencia de fase, Hessian y divergencia de campos vectoriales, la construcción de una matriz de entrenamiento basada en los vectores de características y las etiquetas manuales, finalmente se utiliza un clasificador ELM para segmentar los vasos. Se evalúa el modelo en dos conjuntos de datos públicos ampliamente utilizados para la segmentación de vasos retinianos, DRIVE (20 imágenes de entrenamiento y 20 de prueba) y RIS (15 imágenes con vasos retinianos etiquetados manualmente). Los experimentos se realizaron en un entorno con un CPU Intel i7-4790K a 4.0 GHz y 32 GB de memoria. El tiempo de entrenamiento del clasificador ELM es de aproximadamente 283 segundos para la base de datos DRIVE y 1800 segundos para la base de datos RIS. El método propuesto en la base de datos DRIVE tiene una sensibilidad de 0.7140, especificidad de 0.9868 y precisión de 0.9607, que son competitivos en comparación con otros métodos, algunos de los cuales tienen un rendimiento similar pero requieren mucho más tiempo para la segmentación.

En otro enfoque, Fan et al. [80] donde el algoritmo propuesto para la detección del disco óptico (OD) se basa en el aprendizaje estructurado, utilizando un modelo de clasificador entrenado para detectar bordes en imágenes de fondo, se aplica un método de umbralización (Otsu) para crear una imagen binaria del OD a partir del mapa de bordes. La transformada de Hough circular se utiliza para aproximar el contorno del OD a una forma circular y finalmente se emplea un bosque aleatorio como detector de bordes, que tiene la capacidad de seleccionar automáticamente un conjunto óptimo de características de la imagen original, incluyendo los canales verde, rojo y azul. Se evalúa el modelo en tres conjuntos de datos ampliamente utilizados para la segmentación de disco óptico siendo imágenes de alta resolución, MESSIDOR, DRIONS y ONHSD, estas últimas con una segmentación manual por especialistas. Los experimentos se realizaron en una PC equipada con un procesador Intel (R) Core (TM) i-5 4210 M a 2.60 GHz y 4 GB de RAM. El tiempo promedio de computación para la segmentación del OD fue de aproximadamente 1.7494 segundos. Las métricas empleadas para esta evaluación incluyen Área de Superposición (AOL), coeficiente de Dice, exactitud, Tasa de Verdaderos Positivos (TPF) y Tasa de Falsos Positivos (FPF). De acuerdo con los resultados obtenidos, el mejor desempeño se logró en la base de datos ONHSD.

En síntesis, los estudios revisados evidencian una amplia aplicación de técnicas de aprendizaje automático en la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Estas propuestas se caracterizan por el uso de clasificadores tradicionales, técnicas de extracción de características, y diversas estrategias de preprocesamiento adaptadas al dominio clínico. La variedad de métodos empleados refleja la versatilidad del aprendizaje automático en este campo. Para facilitar la comparación entre estos trabajos, se presentan a continuación en la Tabla 4.10 y la Tabla 4.11 una comparación sistemática entre las distintas propuestas, la cual servirá de base para la selección de modelos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen	Recurso computacional
Zhu et al. [79]	Extreme Learning Machine (ELM)	Extracción de canales (RGB), transformación bottom-hat y filtros gaussianos	ELM	Características de Hessian, transformaciones morfológicas, congruencia de fase y divergencia de campo vectorial	CPU Intel i7-4790K a 4.0 GHz y 32 GB de memoria
Fan et al. [80]	Aprendizaje estructurado	Detección de bordes, umbralización y Transformada de Hough circular	Random Forest	Color, magnitud, pares y vectores característicos	PC con procesador Intel Core i5-4210M a 2.60 GHz y 4 GB RAM

Tabla 4.10 Muestra los aspectos técnicos de los modelos de aprendizaje automático utilizados, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Zhu et al. [79]	DRIVE-AV	Exactitud: 96,07% Sensibilidad: 71,40% Especificidad: 98,68%	No reportado
	RIS	Exactitud: 96,28% Sensibilidad: 72,05% Especificidad: 97,66%	No reportado
Fan et al. [80]	MESSIDOR	No reportado	AOL: 0.8636 (± 0.1268) S: 0.9196 (± 0.1019) Ac: 0.9770 (± 0.0284) TPF: 0.9212 (± 0.1213) FPF: 0.0106 (± 0.0129)

Tabla 4.11 Sintetiza la información recabada de los modelos de aprendizaje automático empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.2.6 Procesamiento digital de imágenes

El procesamiento digital de imágenes es una disciplina que se encarga del análisis y la manipulación de imágenes digitales mediante algoritmos computacionales, esto con el objetivo de obtener una mejor calidad de imagen, poder extraerle la información relevante, así como el poder facilitar su interpretación. Este campo es amplio ya que abarca desde la adquisición y representación de imágenes digitales, hasta técnicas de mejora, restauración, segmentación y compresión, aplicándose en múltiples áreas como la medicina, la visión por computadora y el reconocimiento de patrones. ya que permite imágenes en datos que pueden ser procesados automáticamente para realizar tareas, lo cual lo hace fundamental en aplicaciones clínicas y de diagnóstico asistido por computadora[81]. A continuación, se presenta un mapa conceptual en la Figura 4.8, que resume los principales enfoques de procesamiento digital de imágenes aplicados a la segmentación de estructuras anatómicas en imágenes de fondo de ojo, así como la estructura anatómica específica que se segmenta en cada caso. Este mapa incluye técnicas de mejora, filtrado, segmentación y extracción de características más comúnmente utilizadas, junto con sus combinaciones de métodos de preprocesamiento y análisis que permiten optimizar la calidad y precisión en la segmentación.

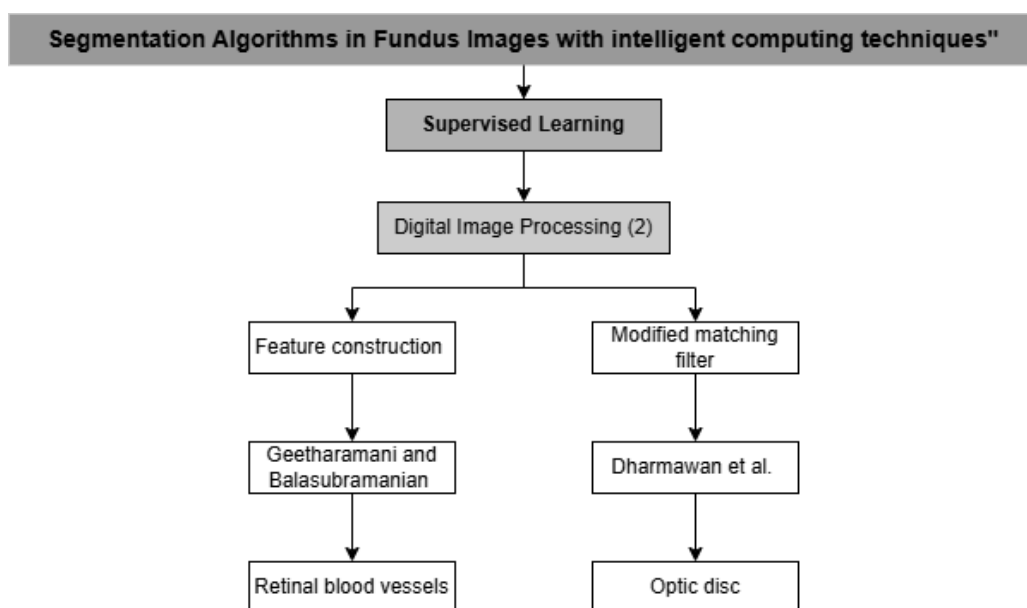


Figura 4.8 Mapa conceptual de procesamiento digital de imágenes aplicado a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron diversas implementaciones basadas en técnicas de procesamiento digital de imágenes, cada una con estrategias particulares para abordar la segmentación automática de estructuras anatómicas en imágenes de fondo de ojo. Estas propuestas emplean métodos de mejora de imagen, filtrado, técnicas de realce, segmentación basada en umbrales y detección de bordes, en combinación con diversas estrategias de preprocesamiento y extracción de características. A continuación, se presentan los trabajos revisados, resaltando sus hallazgos más relevantes y la manera en que aplican estos enfoques para lograr una segmentación precisa y robusta en el dominio clínico.

De acuerdo a Geetharamani y Balasubramanian [82], en la metodología propuesta en el artículo se basa en una arquitectura que combina técnicas de procesamiento de imágenes y minería de datos. Esta arquitectura incluye módulos para la recolección de datos de imágenes retinianas, procesamiento de imágenes (como conversión de espacio de color así como la extracción de canales, mejora de contraste mediante CLAHE y filtrado de Gabor), y técnicas de minería de datos que abarcan análisis de componentes principales, agrupamiento (K-means) y clasificación (Naïve-Bayes y C4.5 con técnicas de bagging). La base de datos utilizada para la evaluación de la metodología es el STARE (Structured Analysis of the Retina), que contiene 20 imágenes de fondo de ojo, de las cuales 10 son imágenes sanas y 10 están afectadas por diversas enfermedades retinianas donde se dispone de segmentaciones manuales con observadores humanos. El procesamiento de imágenes se realizó en Matlab, mientras que las técnicas de minería de datos se implementaron en Tanagra. La precisión en la detección de vasos y no vasos en imágenes sanas fue del 95,20% y la precisión promedio de la segmentación de vasos en imágenes patológicas fue del 94,89%. A pesar de estos resultados, se observaron algunos vasos desconectados, lo que llevó a realizar operaciones de conexión adicionales, logrando una sensibilidad del 68,59% y una especificidad del 97,17%. En general, la sensibilidad alcanzó un máximo del 81,46%, y la especificidad alta refleja una notable reducción en los falsos positivos.

Por su parte, Dharmawan et al. [83] donde el algoritmo propuesto para la segmentación automática del disco óptico (OD) en imágenes de fondo se basa en tres fases principales las cuales son, la estimación del tamaño del OD donde se estima el radio del OD utilizando fórmulas que consideran el área de visualización retinal, la localización del OD con métodos independientes basados en plantillas, basado en el mapa de densidad de vasos y basado en la entropía máxima, y la detección del contorno del OD. Se utiliza un filtro coincidente modificado de Dolph-Chebyshev tipo I para detectar candidatos a los bordes del OD, seguido de la Transformada de Hough Circular y la aproximación B-spline para obtener el contorno final del OD. El algoritmo se implementó en un ordenador con un procesador Intel Core i5 a 2.30 GHz y 4GB de RAM. Se menciona que el tiempo de computación del algoritmo propuesto es competitivo en comparación con otros métodos en la literatura. Los métodos basados en plantillas, densidad de vasos y entropía máxima tuvieron tasas de éxito más bajas en comparación con el método propuesto, lo que indica que la combinación de estos enfoques en el método propuesto mejora significativamente la localización del OD, este estudio reporta que el método propuesto localizó el 100% de los OD en el conjunto de datos DRIVE (20 de 20) y el 99.33% en MESSIDOR (1192 de 1200).

En síntesis, los estudios revisados evidencian una amplia aplicación de técnicas de procesamiento digital de imágenes para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Estas estrategias combinan métodos clásicos y avanzados de mejora, filtrado, segmentación y extracción de características, adaptándose a las particularidades del dominio clínico. Las soluciones analizadas destacan por su capacidad para optimizar la calidad de las imágenes y facilitar la segmentación precisa, incluso frente a desafíos como el ruido, la variabilidad anatómica y las condiciones de adquisición. Para facilitar la comparación entre estas propuestas, se presentan a continuación en la Tabla 4.12 y la Tabla 4.13 una síntesis estructurada de sus principales características, la cual servirá como base para la selección de métodos en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptores de imagen	Recurso computacional
Geetharamani y Balasubramanian [82]	Construcción de características mediante Análisis de Componentes Principales (PCA)	Conversión de espacio de color, extracción de canales y mejora de contraste usando CLAHE	K-Means, Naïve Bayes y C4.5	Análisis de Componentes Principales (PCA), filtrado Gabor y técnica de mejora de contraste	Matlab r2008a y Tanagra
Dharmawan et al [83]	Filtro adaptado modificado Dolph-Chebyshev	Extracción de Regiones de Interés (ROI)	Método basado en plantillas, método basado en densidad de vasos y método basado en máxima entropía	No reportado	Computadora con procesador Intel Core i5 de 2.30 GHz y 4 GB de RAM

Tabla 4.12 Muestra los aspectos técnicos de los modelos de procesamiento digital de imágenes utilizado, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Geetharamani y Balasubramanian [82]	STARE	Precisión (Accuracy): 95.20% Sensibilidad (Sensitivity): 71.34% Especificidad (Specificity): 81.46%	No reportado
Dharmawan et al [83]	DRIVE	No reportado	AOL: 0.873 Índice de Dice (DC): 0.932 Precisión (AC): 0.997 Verdaderos Positivos (TPF): 0.914 Falsos Positivos (TFF): 0.001 Sensibilidad: 100%

Tabla 4.13 Sintetiza la información recabada de los modelos de procesamiento digital de imágenes empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.3 Modelos basados en aprendizaje no supervisado y semi-supervisado

A diferencia del aprendizaje supervisado, el aprendizaje no supervisado tiene como objetivo detectar patrones en un conjunto de datos y categorizar instancias individuales en el conjunto de datos en dichas categorías, los algoritmos de aprendizaje no supervisado son denominados de esta manera porque los patrones que pueden existir o no en un conjunto de datos no están informados por un objetivo y quedan a criterio del algoritmo. Así como la predicción de comportamientos, la

clasificación automática y la toma de decisiones basada en datos. Por otro lado, el aprendizaje semi-supervisado puede considerarse que es un punto medio entre el aprendizaje supervisado y no supervisado, lo cual lo hace especialmente útil para conjuntos de datos que tienen todas las características están presentes, pero no todas tienen objetivos asociados, es decir, contienen datos etiquetados y no etiquetados. Ya que estas situaciones suelen surgir cuando etiquetar imágenes requiere mucho tiempo o no hay una autorización para realizarlo [84]. A continuación, se presenta un mapa conceptual en la Figura 4.9, que resume los enfoques de aprendizaje no supervisado y semi supervisado aplicados a la segmentación de estructuras anatómicas en imágenes de fondo de ojo, así como la estructura anatómica específica que se segmenta en cada caso. Este mapa incluye los métodos de agrupamiento, reducción de dimensionalidad, y modelos de autoaprendizaje más comúnmente utilizados, junto con sus combinaciones de técnicas de preprocesamiento y extracción de características.

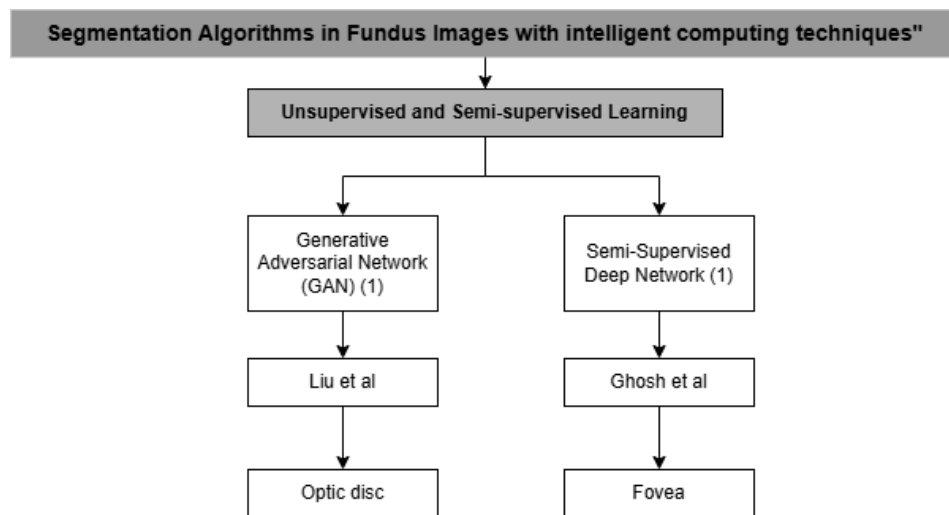


Figura 4.9 Mapa conceptual de modelos de aprendizaje no supervisado y semi-supervisado aplicadas a la segmentación de imágenes de fondo de ojo. Se incluyen arquitecturas clave y estudios representativos revisados en esta tesis.

A partir de los estudios analizados, se identificaron diversas implementaciones basadas en técnicas de aprendizaje no supervisado y semi supervisado, cada una con estrategias particulares para abordar la segmentación automática de estructuras anatómicas en imágenes de fondo de ojo. Estas propuestas emplean algoritmos de agrupamiento, técnicas de reducción de dimensionalidad, autoencoders, y métodos de propagación de etiquetas, en combinación con diversas estrategias de preprocesamiento y extracción de características. A continuación, se presentan los trabajos revisados, resaltando sus hallazgos más relevantes y la manera en que aplican estos enfoques para lograr una segmentación precisa con supervisión limitada o sin supervisión.

De acuerdo a Liu et al. [85] la arquitectura propuesta en el artículo consiste en un marco de trabajo basado en redes generativas adversariales condicionales (cGANs) para la segmentación conjunta del disco óptico y la copa. Este marco incluye tres componentes principales, los cuales son la red de segmentación (S) que produce mapas de predicción multi-clase para las regiones del disco y la copa a

partir de imágenes de fondo, el generador (G) el cual convierte los mapas de segmentación en imágenes de fondo pseudo generadas y el discriminador (D) que se enfoca en distinguir entre pares de imágenes y etiquetas reales y falsos. El objetivo es minimizar la función de valor que describe el juego minimax entre S, G y D. Se utilizan tanto datos etiquetados como no etiquetados en un marco de aprendizaje semi-supervisado para mejorar la precisión de la segmentación. Se utilizan dos conjuntos de datos para evaluar el rendimiento del método propuesto, ORIGA la cual tiene 650 imágenes de fondo (550 imágenes para entrenamiento y 100 para prueba) , siendo de estas 482 imágenes de ojos sanos y REFUGE la cual tiene 400 imágenes de fondo (300 imágenes para entrenamiento y 100 para prueba) , siendo de estas 360 imágenes de ojos sanos. El entrenamiento del método se realiza en una sola GPU NVIDIA GTX 1080Ti, y el proceso de entrenamiento toma aproximadamente 4 horas para 200 iteraciones. Durante la fase de prueba en línea, el tiempo de procesamiento para producir el mapa de segmentación para una imagen de fondo es de aproximadamente 0.0016 s para ORIGA y 0.0019 s para REFUGE. El método logra resultados superiores en comparación con otros enfoques en términos de segmentación del disco óptico y la copa, también, la utilización de datos no etiquetados mejora significativamente el rendimiento del modelo, lo que es crucial dado que los datos etiquetados son escasos en este dominio.

Por otro lado, Ghosh et al. [86] donde el algoritmo se basa en un enfoque de aprendizaje semi-supervisado que combina datos etiquetados y no etiquetados. La arquitectura propuesta para la segmentación de la fovea es DeepLabv3+, que utiliza EfficientNet-B3 como extractor de características. DeepLabv3+ emplea un módulo de agrupamiento espacial atrous y una red de codificación-decodificación, lo que permite ajustar el campo de visión de los filtros sin afectar la resolución de las respuestas de características. Se utiliza una combinación ponderada de la pérdida de entropía cruzada binaria (BCE) y la pérdida de Tversky (FTL) para abordar el problema del desequilibrio de clases. Se utilizaron un total de 484 imágenes etiquetadas de cinco conjuntos de datos diferentes, DRIVE, IDRiD , MESSIDOR, DIARETDB0 ,DIARETDB1, Además, se incorporaron 1200 imágenes no etiquetadas del conjunto de datos MESSIDOR para el aprendizaje semi-supervisado. La arquitectura DeepLabv3+ con EfficientNet-B3 permite un rendimiento eficiente con menos parámetros, lo que reduce los recursos computacionales necesarios, sin embargo, se realizaron ajustes de hiperparámetros a través de un método de búsqueda en cuadrícula para optimizar la combinación de pérdidas. La implementación de un enfoque de aprendizaje semi-supervisado ha demostrado ser efectiva para mejorar la segmentación de la fovea, el modelo propuesto muestra un rendimiento superior en la mayoría de las métricas evaluadas en comparación con otros métodos existentes, especialmente en términos de sensibilidad y especificidad.

En síntesis, los estudios revisados evidencian una creciente incorporación de técnicas de aprendizaje no supervisado y semi-supervisado en la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Estas estrategias permiten aprovechar tanto datos etiquetados como no etiquetados, ampliando así el alcance y la eficacia de los modelos propuestos. Las soluciones analizadas destacan por el uso de arquitecturas avanzadas, combinaciones personalizadas de funciones de pérdida y mecanismos de aprendizaje que maximizan el rendimiento incluso en escenarios con recursos limitados. Para facilitar la comparación entre estas propuestas, se presentan a continuación en la Tabla 4.14 y la Tabla 4.15 una síntesis estructurada de sus principales características, la cual servirá como base para la selección de enfoques en el desarrollo experimental de esta tesis.

Artículo	Modelo	Preprocesamiento	Clasificadores	Descriptor de imagen	Recurso computacional
Ghosh et al [86]	DeepLabv3+	Filtrado Gaussiano, Umbralización binaria, Operaciones morfológicas y Aumento de datos	EfficientNet-B3	Entropía Cruzada Binaria (BCE), Índice de Tversky (TI) y Pérdida Focal de Tversky (FTL)	No reportado
Liu et al [85]	Redes Generativas Antagónicas Condicionales (cGANs)	Segmentación por interés y redimensionamiento	U-Net, M-Net, SegNet y cGANs	-	Una GPU NVIDIA GTX 1080Ti

Tabla 4.14 Muestra los aspectos técnicos de los modelos de aprendizaje no supervisado y aprendizaje semi-supervisado utilizados, los procesos de preprocesamiento, los clasificadores complementarios empleados, los descriptores utilizados y el recurso computacional reportado.

Artículo	Base de datos	Métricas del algoritmo	Métricas clínicas
Ghosh et al [86]	DRIVE, MESSIDOR, IDRiD, DIARETDB0, DIARETDB1	DICE: 82,43% MIoU: 70,52% Sensibilidad (Sn): 91,74% Especificidad (Sp): 99,75% Exactitud (Ac): 99,57%	No reportado
Liu et al [85]	ORIGA	IoU disco óptico (od): $0,9420 \pm 0,0011$ IoU copa óptica (oc): $0,7812$ MIoU: $0,8460 \pm 0,0025$	No reportado

Tabla 4.15 Sintetiza la información recabada de los modelos de aprendizaje no supervisado y aprendizaje semi-supervisado empleados, las bases de datos empleadas, las métricas utilizadas tanto a nivel de algoritmo como en el contexto clínico, permitiendo valorar la aplicabilidad real de cada estudio en entornos de salud.

4.4 Recolección y análisis de bases de datos

No se puede seleccionar el algoritmo adecuado sin antes pensar en la base de datos empleada, el uso de la base de datos es fundamental para el entrenamiento, validación y evaluación del modelo más óptimo en la segmentación, sobre todo en el ámbito de oftalmología. Por ello, se hizo un análisis de

las bases de datos que contienen imágenes de retina obtenidas mediante técnicas de fundoscopia digital y permiten el desarrollo de algoritmos con aplicaciones en la práctica clínica. Para cumplir con el objetivo específico de esta tesis se realizó una recopilación de las bases de datos en los artículos seleccionados para poder analizar las más relevantes y frecuentemente utilizadas en estudios relacionados con segmentación de estructuras anatómicas en imágenes de fondo de ojo.

4.4.1 Descripción de las bases de datos utilizadas

En esta sección se describen las bases de datos recopiladas, se enfoca en el modelo utilizado, el autor o autores responsables, y el nombre de la base de datos. Esta información se resume en la Tabla 4.16, la cual permite observar de forma estructurada la variedad de fuentes empleadas en los diferentes enfoques de segmentación. Mayormente las bases de datos son de acceso público, esto quiere decir que son ampliamente utilizadas en la comunidad científica, sin embargo, también hay otras que son de instituciones médicas específicas o iniciativas regionales. Esta recopilación y clasificación sienta las bases para una selección informada de las bases de datos más adecuadas en el desarrollo experimental posterior, teniendo en cuenta la representatividad clínica, el tipo de estructuras anatómicas segmentadas y la disponibilidad de anotaciones expertas que sirvan como referencia para evaluación.

Artículo	Modelo	Base(s) de datos utilizadas
Zhu et al. [79]	Extreme Learning Machine (ELM)	DRIVE RIS
Fan et al.[80]	Aprendizaje estructurado	MESSIDOR
Li et al. [65]	Fully Convolutional Network (FCN)	Tongren Hospital
Wang et al.. [66]	FCN combinada con modelo WLRR (recuperación de matriz de bajo rango)	DRISHTI-GS
Zhou et al. [68]	U-Net, BF-Net	DRIVE-AV LES-AV HRF-AV
Li et al. [69]	CNN + Transformer	REFUGE Drishti-GS RIM-ONE-r3
Rong et al.[70]	Convoluciones auto-calibradas (SCC)	DRIVE-AV CHASE_DB1
Sun et al. [71]	SDAU-Net	DRIVE-AV CHASE_DB1 STARE
Wang et al.[72]	Context Spatial U-Net	DRIVE-AV
Guo [74]	CSGNet	DRIVE-AV CHASE_DB1 STARE HRF
Wang et al. [75]	HANet	DRIVE-AV

		CHASE_DB1 STARE HRF
Liskowski y Krawiec [76]	Deep Neural Networks (DNN)	DRIVE-AV STARE
Cherukuri et al. [77]	MSR-DNVS (Multi-Scale Regularized Deep Network)	DRIVE-AV
Ghosh et al. [86]	DeepLabv3+ + EfficientNet-B3 semi-supervisado	DRIVE MESSIDOR IDRiD DIARETDB0, DIARETDB1
Liu et al. [85]	Conditional GANs	ORIGA REFUGE
Geetharamani y Balasubramanian [82]	PCA + KMeans, Naïve Bayes, C4.5	STARE
Dharmawan et al. [83]	Modified Dolph-Chebyshev matched filter	DRIVE
Yap y Ng [57]	CAMContrast	OIA-ODIR, IDRiD-seg, REFUGE-seg, Vessel-seg
Zhou et al.[63]	AutoMorph	EyePACS-Q IOSTAR IDRiD
Chen et al. [58]	Shape Regularization Extractor (WT-PSE)	FUNDUS
Aurangzeb et al. [59]	ColonSegNet V2	DRIVE
Morís et al. [60]	Context Encoder	CHASE_DB, STARE, DRIVE
Alkhaldi y Halawani [61]	GOFED-RBVSC (Optimización de saltamontes con detección de bordes difusos para segmentación y clasificación de vasos retinianos)	Kaggle (no especificado)
Zhang et al. [62]	Fusión de características multiescala	DRIVE, STARE

Tabla 4.16 Relación de autores, modelos propuestos y bases de datos utilizadas para la segmentación de estructuras anatómicas en imágenes de fondo de ojo. Esta Tabla resume los enfoques revisados en esta tesis, destacando la diversidad de fuentes de datos y metodologías empleadas.

Las bases de datos más utilizadas entre los estudios revisados son DRIVE, STARE, CHASE_DB1, MESSIDOR y REFUGE, la primera razón por la que se consideran de las más relevantes todas ellas son de acceso público, lo cual facilita la reproducibilidad de los experimentos, la comparación entre

distintos enfoques y la colaboración entre grupos de investigación. Además, estas bases se destacan por ofrecer anotaciones expertas de alta calidad, realizadas generalmente por oftalmólogos y retinólogos. Estas segmentaciones manuales son cruciales ya que permiten establecer un referencia de mayor precisión y confiabilidad con el cual evaluar objetivamente la precisión y eficacia de los modelos, ya sea en tareas de segmentación. La frecuencia con la que estas bases de datos han sido utilizadas en los estudios revisados se resume en la Tabla 4.17, donde se muestra el número de veces que cada base aparece reportada en la literatura analizada. Esta información permite identificar cuáles son las bases más representativas y utilizadas en el desarrollo y validación de modelos de segmentación en imágenes de fondo de ojo.

Base de datos	Número de veces que aparece	Observaciones
DRIVE	13	Es la más usada. Contiene imágenes de retina con anotaciones de vasos sanguíneos, útil para segmentación vascular.
STARE	10	Frecuente en segmentación de vasos sanguíneos; incluye anotaciones expertas.
CHASE_DB1	6	Similar a DRIVE, también utilizada en tareas de segmentación vascular.
MESSIDOR	6	Contiene imágenes asociadas a retinopatía diabética; incluye grados clínicos.
REFUGE	5	Base usada principalmente para segmentación de disco óptico y copa.

Tabla 4.17 Frecuencia de uso de las bases de datos en los estudios revisados.

Sin embargo, estas bases de datos presentan una diversidad clínica significativa, ya que DRIVE, STARE Y CHASE_DB1 están orientadas principalmente a la segmentación vascular, y contienen imágenes de pacientes sanos y con patologías, MESSIDOR es una base de referencia para el diagnóstico automatizado de retinopatía diabética, incluyendo anotaciones de severidad y REFUGE está enfocada en la segmentación de disco y copa óptica. Por estas razones han sido fundamentales en contextos académicos y para prototipos clínicos y herramientas de diagnóstico asistido por computadora. En la Figura 4.10 Se presentan imágenes representativas de las bases de datos más utilizadas en los estudios revisados, específicamente DRIVE, STARE, CHASE_DB1, REFUGE y MESSIDOR.



Figura 4.10 Imágenes representativas de las bases de datos más utilizadas en estudios de segmentación de imágenes de fondo de ojo: **DRIVE**, **STARE**, **CHASE_DB1**, **MESSIDOR** y

REFUGE. Se observan diferencias en calidad, resolución y tipo de estructuras anatómicas destacadas, las cuales reflejan el propósito clínico específico de cada base de datos.

4.4.2 Justificación de la base de datos seleccionada.

Si bien como se mencionaba anteriormente las bases de datos más representativas en esta revisión de la literatura han sido ampliamente utilizadas y son consideradas en diversas investigaciones internacionales, presentan ciertas limitaciones cuando se trata de extrapolar sus resultados a contextos clínicos regionales. Estas bases, en su mayoría, contienen imágenes capturadas en condiciones ideales, con equipos de alta gama, personal especializado y una calidad de imagen generalmente elevada. Por ello, estas condiciones no representan fielmente los escenarios reales observados en entornos clínicos de la región, donde se enfrentan desafíos particulares como la variabilidad en la calidad de las imágenes, la complejidad anatómica en poblaciones diversas, y la necesidad de identificar cambios patológicos en etapas tempranas. En muchos casos, las imágenes locales son tomadas con dispositivos de menor resolución o bajo condiciones de iluminación subóptima, lo que impacta directamente en la aplicabilidad de los modelos entrenados con bases de datos ideales. De acuerdo a la finalidad de este proyecto el cual propone una solución adapta las condiciones y recursos similares al entorno regional, donde se busca desarrollar o ajustar modelos algorítmicos que no dependan exclusivamente de imágenes de calidad óptima, sino que sean capaces de funcionar eficazmente en contextos clínicos reales, considerando las limitaciones técnicas y los desafíos diagnósticos específicos de la región se propone una base de datos similar a las encontradas en la revisión de la literatura pero que cumple con los requerimientos para el entorno propuesto.

En este contexto, se seleccionó la base de datos APTOS 2019 Blindness Detection (por sus siglas en inglés, Asia Pacific Tele-Ophthalmology Society) por kaggle [87], como punto de partida para el desarrollo y prueba del enfoque propuesto, debido a que presenta características que se asemejan en mayor medida a las condiciones encontradas en entornos clínicos reales y regionales ya que a diferencia de las otras bases de datos mencionadas anteriormente APTOS 2019 Blindness Detection incluye imágenes con variaciones en calidad, contraste, iluminación y enfoque, lo cual representa un desafío técnico más cercano a lo que ocurre en la práctica clínica diaria en regiones con limitaciones de infraestructura tecnológica. Además, APTOS 2019 Blindness Detection contiene una amplia variabilidad en la calidad de las imágenes ya que no fueron tomadas por el mismo instrumento médico ni en el mismo lugar ni el mismo personal de la salud, contiene un amplio y diverso panorama de imágenes de retina desde ojos completamente sanos hasta fondos de ojo con una variabilidad de patología severa, cabe decir esta base de datos ha sido utilizada en competencias internacionales, como las organizadas por Kaggle, lo que garantiza cierto nivel de validación y reconocimiento por parte de la comunidad científica. En la Figura 4.11, se muestran ejemplos visuales correspondientes a cada una de las clases de severidad presentes en la base de datos APTOS.

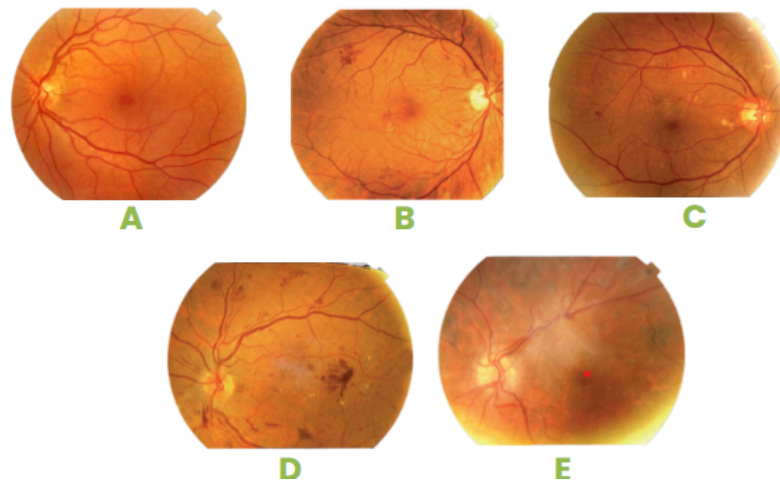


Figura 4.11 Imágenes representativas de las cinco clases de severidad de retinopatía diabética presentes en la base de datos **APTOS**: **A** (sin patología), **B** (patología leve), **C**(patología moderada), **D** (patología severa) y **E**(retinopatía proliferativa).

El uso de esta base de datos permite desarrollar un modelo más robusto ante las condiciones no ideales proporcionadas por el entorno clínico regional, permitiendo evaluar el desempeño de algoritmos en un escenario más representativo de la realidad. En consecuencia, su incorporación está alineada con el propósito central de este proyecto: proponer soluciones algorítmicas adaptadas al contexto regional, sin depender exclusivamente de imágenes estandarizadas de alta calidad.

4.4.3 APTOS 2019 Blindless detection

La base de datos APTOS contiene un amplio conjunto de imágenes de retina tomadas mediante fotografía de fondo de ojo en una variedad de condiciones de imagen. Contiene imágenes clasificadas en A - sin patología, B- Patología Leve, 2- Patología moderada, 3 - Patología Grave y 4- Patología proliferativa. Como se mencionó anteriormente, las imágenes tienen variabilidad en la calidad de las imágenes, algunas cuentan con desenfoque, contienen artefactos, por mencionar alguna de las situaciones. También se debe de tomar en cuenta su toma con diversos artefactos y en divergentes tiempos, lo que agrega una mayor variación. Las imágenes son de acceso público, pertenecen a un concurso lanzado en 2019 por KAGGLE, Las imágenes se dividen en: Etiquetas de entrenamiento y de prueba e Imágenes de entrenamiento y prueba, Las consta de 3662 fondos de ojo y 1928 fondos de ojo para entrenamiento y prueba respectivamente [87]. Como se evidencia en la Figura 4.12, las imágenes provenientes de la base de datos APTOS presentan una notable variabilidad en calidad, enfoque, contraste y condiciones de captura. En dicha Figura se muestran cuatro ejemplos representativos que ilustran esta diversidad, la cual responde a diferencias en los dispositivos utilizados, iluminación, presencia de artefactos, y momentos de adquisición.

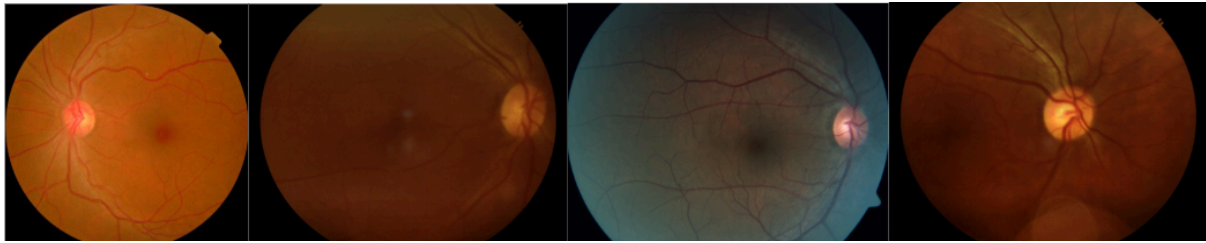


Figura 4.12 Ejemplos de imágenes de fondo de ojo extraídas de la base de datos APTOS. Se observa la variabilidad en calidad de imagen, enfoque, iluminación y presencia de artefactos, lo cual refleja condiciones de captura clínicas diversas y no estandarizadas.

4.5 Clasificación y comparación de algoritmos existentes

Una vez identificada la base de datos más representativa de la región para realizar la segmentación estructuras anatómicas en imágenes de fondo de ojo, es necesario analizar los algoritmos aplicados en la revisión de la literatura para seleccionar el más acorde a uno de los objetivos específicos para esta investigación el cual es catalogar los algoritmos utilizados en el estado del arte para atacar la tarea de segmentación mencionada anteriormente. Por o tanto en esta sección se presentan de forma sistemática los modelos, tareas abordadas en este caso la estructura segmentada, y sus requerimientos computacionales. Esta categorización permite comparar el rendimiento de los distintos enfoques así como constituye la base para seleccionar los modelos más adecuados para el desarrollo experimental de esta tesis, considerando el contexto clínico regional previamente descrito.

4.5.1 Tabla comparativa de algoritmos encontrados en la revisión del estado del arte.

La Tabla 4.18 presenta un resumen comparativo de los algoritmos identificados en la revisión del estado del arte, organizados según el modelo utilizado, el tipo de algoritmo al que pertenecen, la tarea principal que abordan en este caso segmentación de qué estructura anatómica principal, como dato extra si existe una clasificación y si requiere el uso de GPU para su implementación. También en esta Tabla se puede observar de manera clara la diversidad de enfoques aplicados en el procesamiento de imágenes de fondo de ojo y ofrece un punto de partida para identificar los modelos más adecuados en función de los recursos disponibles y las necesidades clínicas del contexto regional.

Artículo	Modelo	Tipo de algoritmo	Tarea principal	¿Requiere GPU?
Li et al [65]	FCN con fusión dual	Deep Learning (FCN)	Segmentación disco y copa	Sí
Wang et al [66]	FCN + WLRR	Deep Learning (FCN)	Segmentación vascular	Sí
Zhou et al [68]	U-Net, BF-Net	Deep Learning (U-Net)	Segmentación vascular	Sí
Li et al [7]	CNN + Transformer	Deep Learning (U-Net)	Segmentación disco y copa	Sí
Rong et al [70]	SCC + ISAM	Deep Learning (U-Net)	Segmentación vascular	Sí

Sun et al [71]	SDAU-Net	Deep Learning (U-Net)	Segmentación vascular	Sí
Wang et al [72]	Context Spatial U-Net	Deep Learning (U-Net)	Segmentación vascular	Sí
Guo [74]	CSGNet	Deep Learning (Otra)	Segmentación vascular	Sí
Wang et al [75]	HANet	Deep Learning (Otra)	Segmentación vascular	Sí
Liskowski y Krawiec [76]	DNN	Deep Learning (Otra)	Segmentación vascular	Sí
Cherukuri et al [77]	MSR-DNVS	Deep Learning (Otra)	Segmentación vascular	Sí
Zhu et al [79]	ELM	Machine Learning	Segmentación vascular	No
Fan et al [79]	Structured Learning	Machine Learning	Segmentación disco óptico	No
Ghosh et al [86]	DeepLabv3+	No supervisado /Semi-supervisado	Segmentación de la fovea	No reportado
Liu et al [85]	cGANs	No supervisado /Semi-supervisado	Segmentación disco y copa	Sí
Geetharamani y Balasubramanian [82]	PCA + K-Means	Procesamiento digital de imágenes	Segmentación vascular	No
Dharmawan et al [83]	Filtro adaptado	Procesamiento digital de imágenes	Segmentación vascular	No
Aurangzeb et al [59]	ColonSegNet V2	Deep Learning (CNN)	Segmentación vascular	No reportado
Morís et al [60]	Context Encoder	Deep Learning (CNN)	Segmentación vascular y fovea	No reportado
Alkhalidi y Halawani [61]	GOFED-RBVSC	Deep Learning (CNN)	Segmentación y clasificación	No reportado
Zhang et al [62]	SVM + B-COSFIRE	Deep Learning (CNN)	Segmentación vascular	No reportado
Zhou et al [63]	AutoMorph	Deep Learning (CNN)	Segmentación disco	No reportado
Yap y Ng [57]	CAMContrast	Deep Learning (CNN)	Clasificación y localización vascular	Sí
Chen et al [58]	WT-PSE	Deep Learning (CNN)	Segmentación disco y copa	Sí

Tabla 4.18 Comparación de algoritmos encontrados en la revisión del estado del arte para la segmentación de estructuras anatómicas en imágenes de fundoscopia digital. Se incluye el modelo

propuesto, el tipo de algoritmo, la tarea principal abordada y si requiere GPU para su implementación,.

Como puede observarse en la Tabla, la mayoría de los modelos recientes se basan en arquitecturas de aprendizaje profundo, particularmente redes convolucionales como U-Net, FCN o variantes híbridas (otras). El enfoque de estos modelos es la segmentación fina, dominan la extracción de vasos sanguíneos o el contorno del disco óptico, y han demostrado un desempeño sobresaliente en bases de datos públicas y escenarios controlados. Sin embargo, tiene una consideración la cual implica que su implementación generalmente requiere mayor capacidad computacional, incluyendo el uso de GPU. Por otro lado, los modelos basados en aprendizaje automático y procesamiento digital de imágenes tienden a ser más ligeros computacionalmente, lo que los hace atractivos en entornos con recursos limitados, Pero su rendimiento puede verse afectado frente a imágenes de baja calidad o estructuras anatómicas complejas ya que al ser más ligeros, son menos precisos lo cual puede afectar la segmentación de estructuras tan detalladas como lo es la arcada vascular.

Debido a esta comparación y considerando amplia adopción en aplicaciones médicas, así como su versatilidad para adaptarse a distintas tareas de segmentación, en este proyecto se opta por trabajar con arquitecturas basadas en U-Net como punto de partida para las pruebas experimentales. Estas redes ofrecen un compromiso favorable entre precisión, estructura modular y posibilidad de personalización, lo que las convierte en una opción sólida para escenarios clínicos reales.

4.5.2 Descripción general del dataset y características visuales iniciales

Previamente a analizar la base de datos con el algoritmo de segmentación seleccionados, se realizó un análisis visual preliminar con el objetivo de identificar patrones comunes en las imágenes de fondo de ojo, Este análisis permite comprender mejor qué tipo de características estructurales y visuales predominantes se están enfrentando, lo cual es esencial para ajustar adecuadamente las técnicas de mejora de contraste y extracción de bordes. Además, se incluyó una evaluación automatizada de la calidad de imagen para respaldar cuantitativamente las observaciones visuales y facilitar una selección informada de las imágenes que serán utilizadas en la validación experimental.

4.6 Análisis visual preliminar de las imágenes

Como se mencionaba anteriormente, en esta sección se realiza un análisis como etapa preliminar, aquí se se realizó una evaluación visual y cuantitativa de las imágenes del conjunto de datos, con el objetivo de determinar la variabilidad de su calidad y así poder establecer los criterios para su inclusión o descarte en las etapas posteriores del procesamiento. Se considera este paso como fundamental debido a la información que nos arroja KAGGLE, ya que una de las problemáticas es como afecta directamente el rendimiento de los algoritmos de segmentación automática debido a la calidad de la imagen a continuación vemos en la Figura 4.13 la imagen original a evaluar.

Imagen de Fundoscopia Digital
Base de Datos APTOS

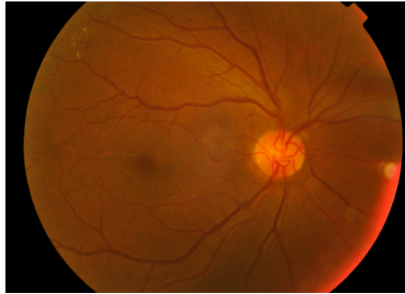


Figura 4.13 Imagen de fondo de ojo original utilizada para mostrar el proceso del análisis preliminar.

Cada imagen fue procesada en tres etapas:

1. Conversión a escala de grises

Cada imagen fue convertida a escala de grises (Figura 4.14) para, de esta manera, hacer el análisis más simple utilizando valores de luminiscencia. Ya que la representación en un solo canal permite observar la distribución de brillo en la imagen de manera directa, facilitando el cálculo de métricas y la visualización de artefactos o zonas oscuras.

Imagen de Fundoscopia en Escala de Grises
Base de Datos APTOS

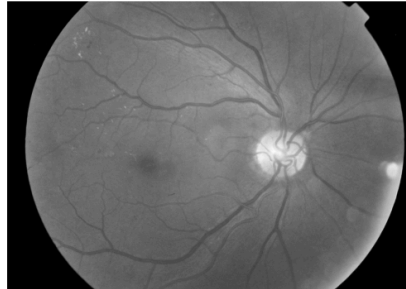


Figura 4.14 Conversión de una imagen de fondo de ojo a escala de grises.

2. Cálculo del histograma

Después, para cada imagen en escala de grises se se generó su histograma (Figura 4.15), representando la frecuencia de aparición de cada valor de nivel de gris, este rango permaneció dentro del rango 0 a 255, permitió observar la distribución de luminancia de la imagen y detectar si predominan valores muy oscuros o saturaciones excesivas. Aquí se observó que un grupo importante de imágenes presentaban un pico de intensidad dominante dentro del rango de 35 a 100, lo cual indica una predominancia de tonos oscuros.

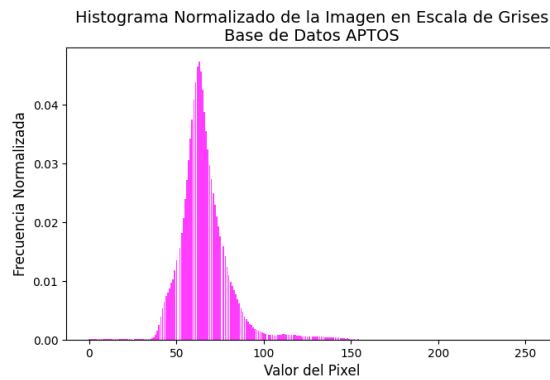


Figura 4.15 Histograma de niveles de gris correspondiente a la imagen anterior.

3. Aplicación de máscara

Para apoyar esta evaluación, se generó una máscara binaria (Figura 4.16) que resalta únicamente los píxeles con valores de intensidad entre 50 y 100. Este rango fue elegido como un criterio inicial para identificar zonas con iluminación moderada, útiles para evaluar si la imagen podría ser considerada de calidad diagnóstica. La máscara resultante permite visualizar de forma rápida las regiones potencialmente útiles y descartar aquellas con baja visibilidad si son demasiado oscuras o brillantes.



Figura 4.16 Máscara binaria aplicada a la imagen en escala de grises, resaltando zonas de intensidad.

Este análisis se aplicó a todas las imágenes del conjunto, generando una visualización compuesta por cuatro elementos por imagen: la versión original en color, la imagen en escala de grises, el histograma y la máscara de luminancia. La Figura 4.17 muestra un ejemplo representativo del proceso aplicado.

Análisis Visual de Imágenes de Fundoscopia

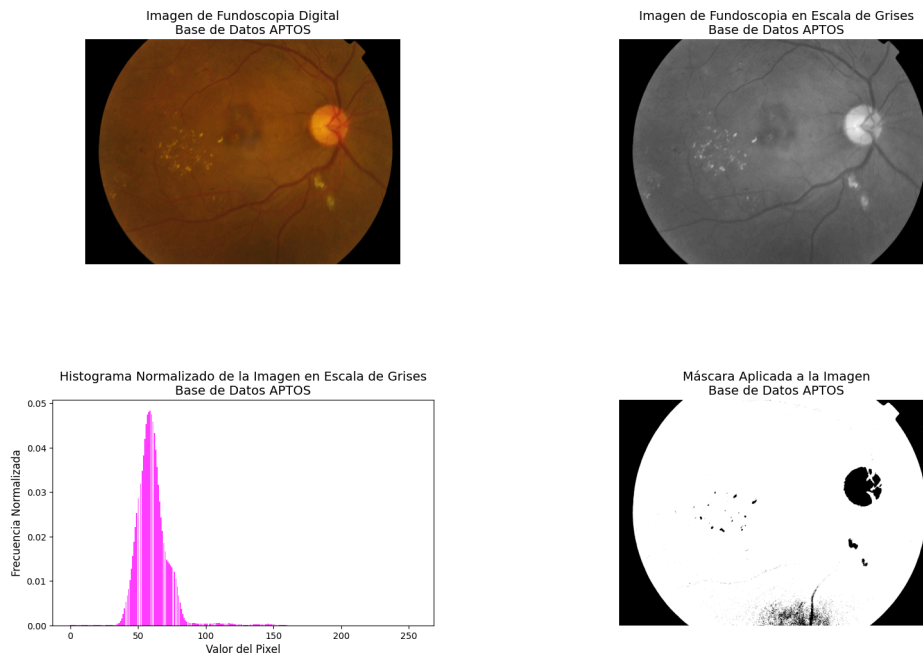


Figura 4.17 Se muestra otro ejemplo del proceso aplicado a una imagen representativa, que incluye: La imagen original en color, su versión en escala de grises, el histograma de luminancia y la máscara binaria aplicada al rango [50, 100].

Generalmente, se puede observar que las imágenes se mantuvieron en un rango de 35 a 100 en el valor pixelar sobre el rango establecido de 0 a 255, esto indica que la mayoría de los píxeles tienen baja intensidad de luminancia, lo cual implica que se tienen mayormente imágenes oscuras o sobreexpuestas. Esto en imágenes de fundoscopia puede ocultar las estructuras anatómicas de interés y complicar la segmentación automática y visual. Por ello, se considera que muchas imágenes requieren una corrección de iluminación, mediante técnicas de mejora de contraste localizadas, que ajustan dinámicamente la distribución de intensidades en regiones específicas de la imagen, procesamiento frecuentemente reportado en la literatura como paso previo a la segmentación. Este tipo de procesamiento de igual manera es uno de los que se reportaron en la mayoría de los casos en la revisión de la literatura, sin embargo su utilidad se confirma para la base de datos utilizada en esta investigación, sin embargo, en algunos casos ciertas imágenes podrían ser descartadas si no cumplen con un umbral mínimo de calidad para análisis automatizado.

4.7 Evaluación automatizada de la calidad de imagen mediante análisis computacional

Después de obtener un análisis preliminar de las imágenes de manera visual y con el objetivo de facilitar una evaluación sistemática y reproducible de la calidad de las imágenes de fondo de ojo, se desarrolló una herramienta interactiva utilizando *Streamlit* una herramienta de Python que permite crear aplicaciones web interactivas de forma rápida y sencilla, la plataforma se observa en la Figura 4.18.

Imágenes de Fondo de Ojo

Representación y Clasificación de Imágenes

Esta herramienta proporciona una representación de:

- La **imagen original**.
- Su versión en **escala de grises**.
- La imagen mejorada con **CLAHE**.
- El **histograma** correspondiente.

Métricas de Clasificación:

1. **Brillo promedio** (ignora píxeles con valor 0): Mide la luminosidad general de la imagen.
2. **Contraste** (ignora píxeles con valor 0): Evalúa la diferencia entre las intensidades más altas y más bajas.
3. **Nitidez** (ignora píxeles con valor 0): Analiza la claridad de los bordes y detalles.

Categorías de Clasificación:

- **Oscura:** Brillo promedio ≤ 40 .
- **Usable:** $40 < \text{Brillo} \leq 150$, $\text{Contraste} \geq 30$ y $\text{Nitidez} \geq 20$.
- **Velada:** No cumple con las condiciones anteriores.

Imágenes para su análisis:

Selecciona las imágenes

Drag and drop files here
Limit 200MB per file • PNG

Figura 4.18 Visualización generada por la aplicación desarrollada en *Streamlit*.

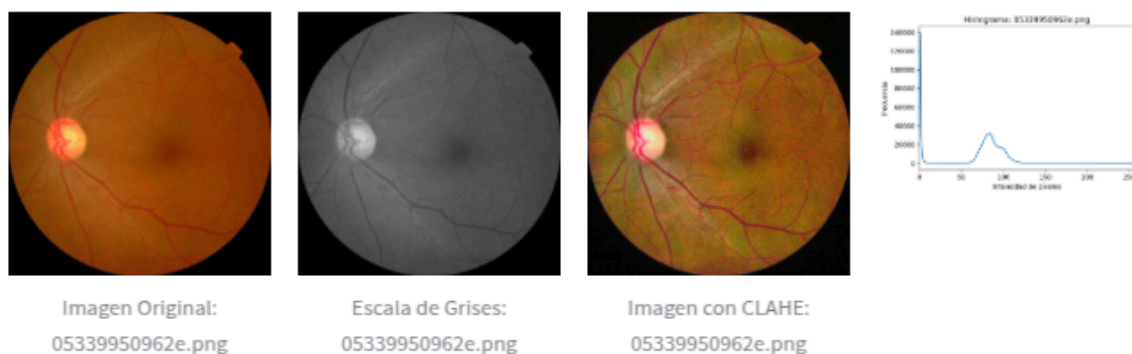
Esta aplicación se desarrolló con objetivos muy puntuales, los cuales fueron la carga de imágenes de la base de datos para obtener tanto un análisis visual como cuantitativo en tiempo real. Este sistema arroja una visualización donde se encuentra la imagen original al color que se quiere analizar, su conversión a escala de grises, la imagen mejorada mediante CLAHE (Por sus siglas en inglés, Contrast Limited Adaptive Histogram Equalization) el cual se detallará en la sección siguiente, y el histograma de distribución de niveles de intensidad. Además, este sistema proporciona tres métricas claves para la visualización de calidad las cuales son brillo promedio el cual fue estimado como la media de los valores de luminancia, excluyendo los píxeles con valor cero, contraste que es calculado mediante la desviación estándar de los niveles de la escala de grises y nitidez medida mediante la varianza del operador Laplaciano[88], que evalúa la claridad de los bordes presentes en la imagen.

De acuerdo a estas métricas, cada imagen se clasificaba automáticamente en cada una de las siguientes categorías:

- **Oscura:** brillo ≤ 40 .
- **Usable:** brillo entre 40 y 150, contraste ≥ 30 y nitidez ≥ 40 .
- **Velada:** imágenes que no cumplen los criterios anteriores y presentan calidad visual comprometida.

Esta clasificación permite una filtración preliminar automatizada, que resulta útil para excluir imágenes de baja calidad antes de aplicar algoritmos más costosos de segmentación anatómica.A

continuación, se muestran en las Figuras 4.19 a 4.21 ejemplos de salida de la herramienta para una imagen clasificada como “Usable”, “Oscura” y “Velada” con sus visualizaciones y métricas correspondientes.



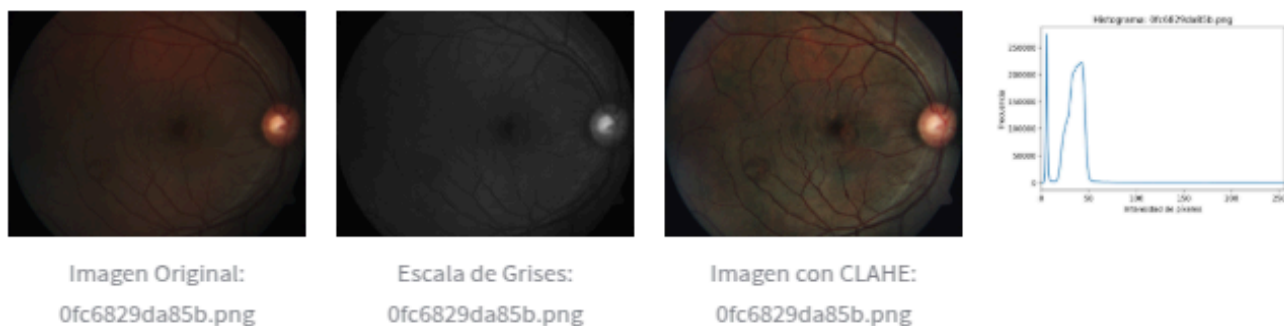
Brillo Promedio: 72.06

Contraste: 36.73

Nitidez: 20.06

Clasificación: Usable

Figura 4.18 Ejemplo de imagen clasificada como “usable”, Visualización de la imagen original, su escala de grises y el histograma correspondiente. La imagen presenta un brillo promedio adecuado, buen contraste y nitidez suficiente



Brillo Promedio: 33.67

Contraste: 13.52

Nitidez: 3.86

Clasificación: Oscura

Figura 4.19 Ejemplo de imagen clasificada como “oscura”. La imagen muestra una baja intensidad de luminancia y escasa visibilidad de estructuras anatómicas. Las métricas indican que no cumple con el umbral mínimo para ser considerada útil en el análisis automatizado.

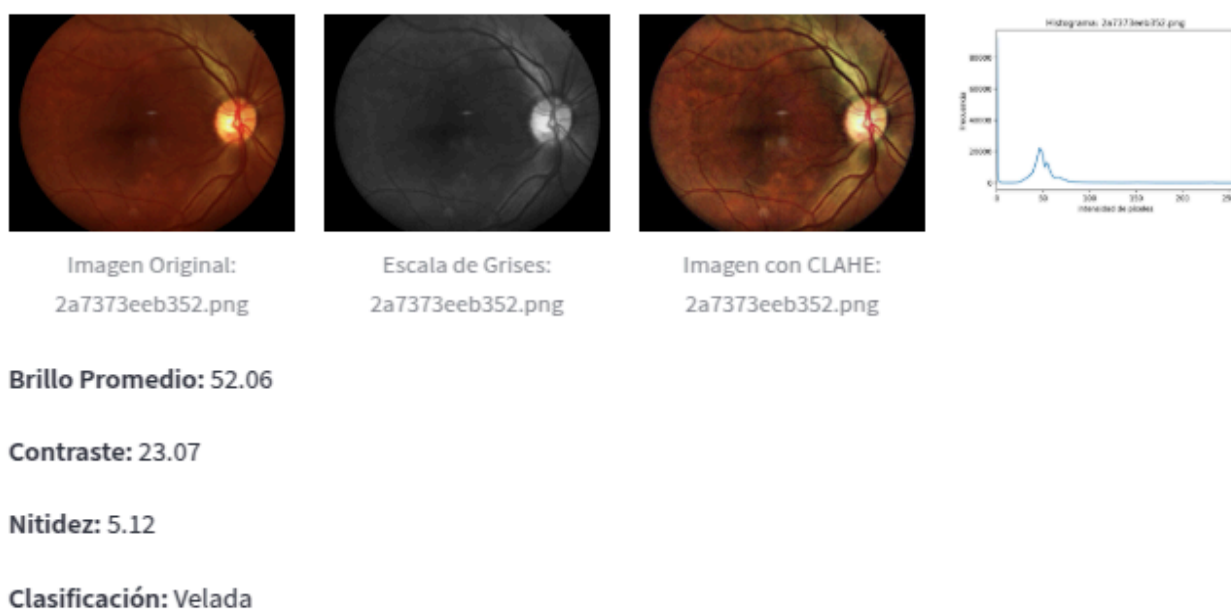


Figura 4.20 Ejemplo de imagen clasificada como “velada”. Esta imagen presenta deficiencias en contraste y nitidez a pesar de tener un brillo intermedio.

Esta clasificación permite una filtración preliminar automatizada, que resulta útil para excluir imágenes de baja calidad antes de aplicar algoritmos más costosos de segmentación anatómica. A continuación, se presentan las métricas claves mencionadas anteriormente en una Tabla comparativa 4.19 para de esta manera, observar la diferencia entre los resultados arrojados en cada categoría, es muy importante recalcar que la imagen 05339950962e.png perteneciente a la clase “Usable” presenta valores superiores en todas las métricas, indicando una mayor calidad en términos de luminancia, contraste y definición, lo que la hace más adecuada para análisis detallados, por otra parte las imágenes 2a7373eeb352.png y 0fc6829da85b.png muestran limitaciones en estos aspectos, que podrían comprometer su utilidad en aplicaciones clínicas o diagnósticas. Aquí se muestra un acercamiento a los requerimientos indispensables para poder validar experimentalmente los algoritmos de segmentación seleccionados utilizando imágenes representativas de la práctica clínica regional.

Imagen	Brillo Promedio	Contraste	Nitidez	Clasificación
2a7373eeb352.png	52.06	23.07	5.12	Velada
0fc6829da85b.png	33.67	13.52	3.86	Oscura
05339950962e.png	72.06	36.73	20.06	Usable

Tabla 4.19 La tabla muestra las métricas de brillo, contraste y nitidez, junto con la clasificación cualitativa asignada a cada imagen.

4.8 Mejoras preliminares de contraste

Tras realizar la clasificación inicial de calidad de imagen y evaluar sus características clave de brillo, contraste y nitidez, se identificaron aquellas imágenes que cumplen con los requisitos mínimos para su uso en procesos de segmentación anatómica se procede con algunas mejoras preliminares de la imagen, también denominada como preprocesamiento con el objetivo de optimizar la visualización de las estructuras de interés, si bien las imágenes se determinaron como "Usables" pueden agregarse ajustes adicionales.

Para esta etapa, se emplearon métodos de realce de contraste localizados, como CLAHE (Contrast Limited Adaptive Histogram Equalization), técnica que será descrita en la siguiente sección.

4.8.1 Mejora de contraste mediante CLAHE

La ecualización de histograma adaptativa con límite de contraste (CLAHE, por sus siglas en inglés) ha demostrado producir buenos resultados en imágenes médicas, ya que está enfocado especialmente cuando la imagen se representa con un rango estrecho de valores de intensidad. Este método se enfoca en el ajuste de intensidades distribuido mediante un histograma utilizando todo el rango de intensidades de forma uniforme mejorando el contraste de la imagen [89]. Se basa en dividir la imagen en regiones no supuestas en tamaños proporcionales, en la Figura 4.22 se puede ver un ejemplo de una imagen de 512 x 512 píxeles, con un total de 64 regiones dividido equitativamente en 8 partes en cada dirección.

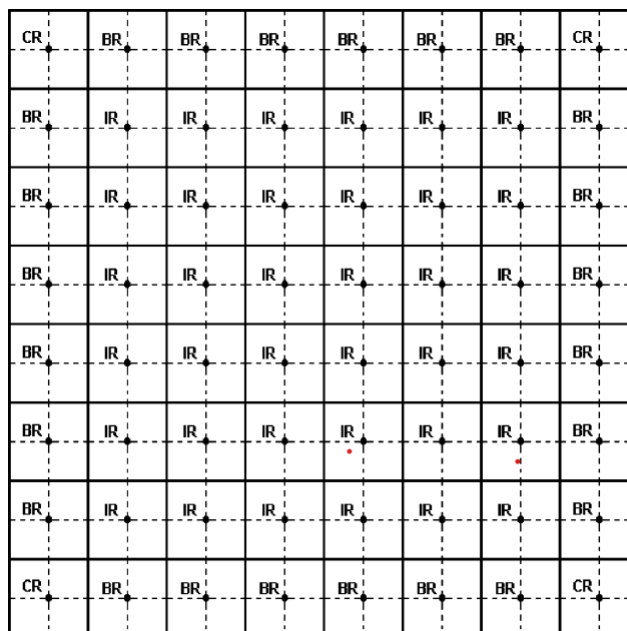


Figura 4.22 Estructura de regiones en una imagen cuadrada de 512 x 512 píxeles que ha sido dividida en 64 regiones cuadradas de igual tamaño [90].

De acuerdo a la Figura 4.21 da lugar existencia de tres grupos de regiones diferentes que se definen de acuerdo al número de regiones calculadas mediante los histogramas, Las regiones de esquina (CR) son un grupo de sólo 4 regiones, Regiones de borde (BR) son las que están compuestas por 24 regiones donde todas las regiones ubicadas en el borde de la imagen (IR), excluyendo las esquinas, pertenecen a esta clase y las regiones internas que está compuesto de las 36 regiones restantes [90].

Dado que las imágenes del fondo de ojo en el conjunto de datos APTOS presentan una alta variabilidad en cuanto a condiciones de iluminación, contraste y presencia de artefactos, es fundamental aplicar técnicas de mejora de imagen que permitan resaltar estructuras anatómicas relevantes. La revisión de la literatura demuestra en la mayoría de los estudios que CLAHE mejora el desempeño de algoritmos de segmentación y clasificación al proporcionar entradas más claras y con mejor definición de bordes. En esta tesis su implementación se considera una etapa fundamental de preprocesamiento para asegurar que los algoritmos de segmentación posteriores operen sobre imágenes con características más uniformes y representativas de la práctica clínica.

Una vez tomada esta decisión, se seleccionó un conjunto de imágenes de la base de datos APTOS con los parámetros correctos para aplicar este preprocesamiento. El proceso consistió en la separación de los canales RGB de cada imagen, seguida de la aplicación independiente de CLAHE a cada canal utilizando los parámetros $\text{clipLimit} = 2.0$ y $\text{tileGridSize} = (8, 8)$, lo que permitió controlar la amplificación del contraste de forma localizada. Posteriormente, los canales procesados fueron recombinados para reconstruir la imagen mejorada. Como parte de la validación visual, se generó una imagen compuesta que muestra la versión original y la mejorada dispuestas horizontalmente, debidamente etiquetadas para facilitar la comparación. Todos los resultados fueron almacenados automáticamente en una carpeta de salida predefinida.

Este procesamiento fue automatizado mediante un script en Python, utilizando bibliotecas como **OpenCV** y **NumPy**. Las imágenes procesadas, en formato **.png** y **.jpeg**, fueron previamente organizadas en carpetas específicas para facilitar su tratamiento.

A continuación, en la Figura 4.23, se presentan tres ejemplos representativos de imágenes originales provenientes de la base de datos APTOS, se observa la imagen original y, a su lado, la imagen resultante tras la mejora de contraste CLAHE.

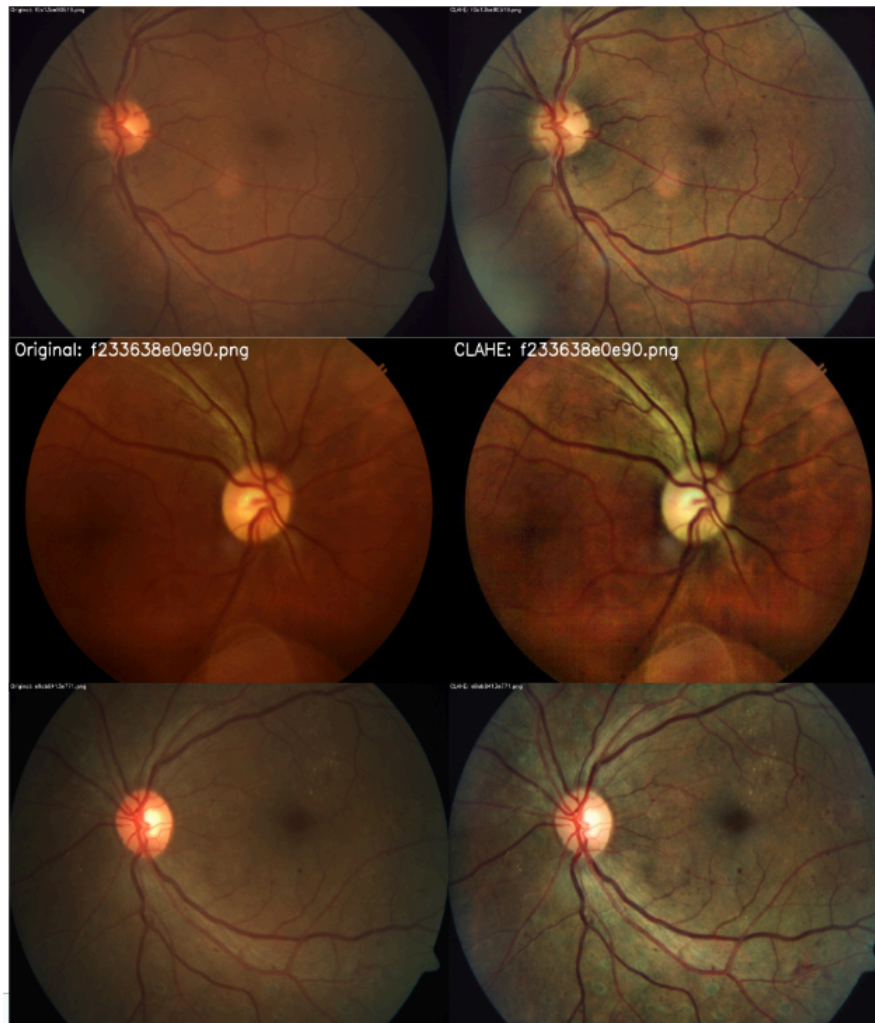


Figura 4.23 Ejemplos de mejora de contraste mediante CLAHE en imágenes de fondo de ojo. Cada fila muestra una imagen original de la base de datos APTOS (izquierda) y su correspondiente versión procesada con CLAHE (derecha).

4.8.2 Análisis de canales RGB con mejora de contraste mediante CLAHE

Al momento de realizar la técnica de preprocesamiento anteriormente mencionada, se decide seguir realizando un análisis exploratorio de la base de datos con el fin de obtener la mayor información posible antes de su evaluación, esto se hace por medio de una extracción de características mediante los canales RGB, el proceso mediante el cual se separan los tres componentes principales de color (Rojo, Verde y Azul) de una imagen digital en color. Cada canal representa la intensidad de un color específico en cada píxel, y al analizarlos por separado se pueden realizar mejoras, análisis o transformaciones específicas que luego pueden recombinarse para reconstruir la imagen original o una versión modificada [81]. Por lo tanto, se realizó un análisis individualizado de los canales rojo, verde y azul (RGB) antes y después de aplicar la técnica de mejora de contraste CLAHE.

A partir de las imágenes previamente seleccionadas se realizó un procesamiento individual de los tres canales de color que componen cada imagen (rojo, verde y azul). Este procedimiento comenzó con la

separación de canales RGB utilizando la biblioteca OpenCV, respetando el orden BGR que utiliza por defecto esta herramienta. Posteriormente, se aplicó la técnica de CLAHE de forma independiente a cada canal, utilizando los parámetros $\text{clipLimit} = 2.0$ y $\text{tileGridSize} = (8, 8)$, con el objetivo de mejorar localmente el contraste en cada componente de color.

Parte del análisis fueron las imágenes donde se generaron las que incluyen los canales rojo, verde y azul en su estado original, sus versiones mejoradas tras aplicar CLAHE, la imagen original en color y la imagen reconstruida a partir de los canales procesados, esto podemos observar en la Figura 4.24.

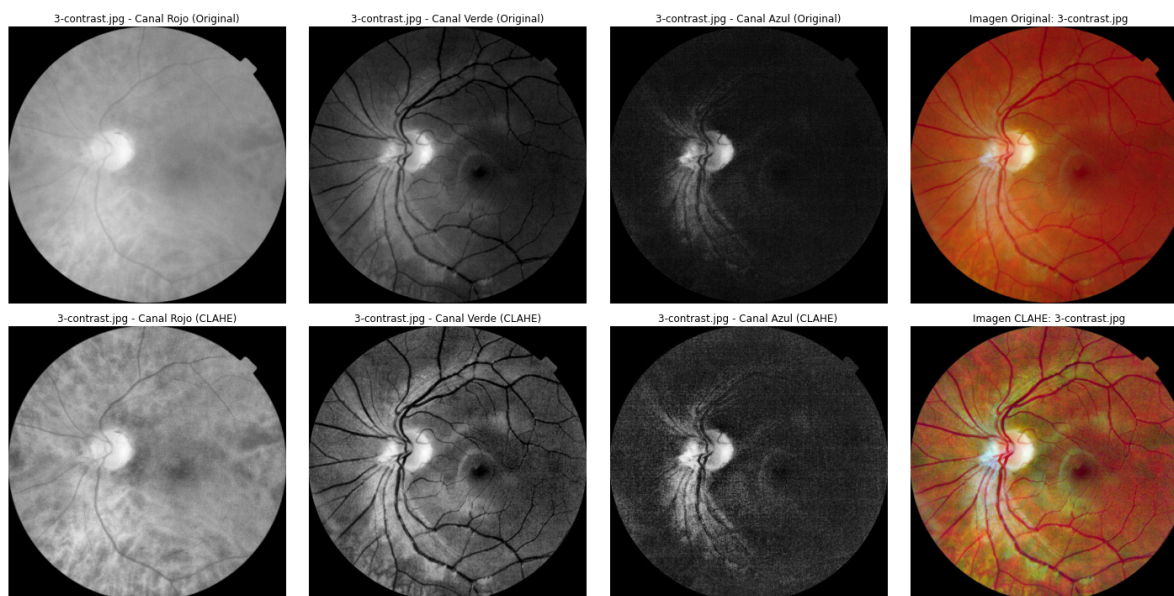


Figura 4.24 Comparación visual del efecto de CLAHE aplicado individualmente a los canales rojo, verde y azul en imágenes de fondo de ojo. Se muestran (de derecha a izquierda): los canales RGB originales, los canales mejorados con CLAHE, la imagen original en color y la imagen reconstruida a partir de los canales procesados.

De acuerdo al análisis realizado se pueden hacer varias inferencias. El canal verde demostró ser el más eficaz para resaltar la arcada vascular y los márgenes del nervio óptico, lo cual es consistente con hallazgos reportados en la revisión de la literatura, donde reportan que los vasos sanguíneos retinianos tienen mayor visibilidad en longitudes de onda intermedias. Por otro lado, el canal rojo, aunque suele estar sobreexpuesto, puede contener información útil en regiones más profundas, en este caso resalta de una manera más luminiscente el nervio óptico, finalmente el canal azul tiende a presentar mayor ruido y menor utilidad diagnóstica. Esto quiere decir que el canal verde arroja información pertinente para la segmentación de las estructuras por lo tanto esta combinación se puede seguir explorando en el análisis de los datos.

4.9 Aplicación del algoritmo Canny para la detección de estructuras anatómicas

Con el propósito de destacar los contornos relevantes en las imágenes de fondo de ojo y evaluar su utilidad para la segmentación de estructuras anatómicas, se aplicó el algoritmo de detección de bordes

de Canny Edge Detector, es uno de los métodos más populares para la detección de bordes en imágenes. No solo detecta bordes, sino que también los define de manera precisa, al mismo tiempo que elimina el ruido que podría generar bordes falsos. Es un proceso de varias etapas, Primeramente se elimina el ruido de la imagen por medio de un filtro Gaussiano para suavizar la imagen, después se le detectan los cambios en la intensidad de los píxeles, lo que indica la presencia de un borde. Se aplica el filtro Sobel en dos direcciones: horizontal (X) y vertical (Y). De aquí se obtiene la magnitud del gradiente que es el que indica la fuerza que tiene el borde y la dirección del gradiente que indica hacia donde va a estar orientado el cambio de intensidad del mismo. A continuación se eliminan los píxeles irrelevantes para que los bordes resultantes sean más delgados y precisos. Se escanea cada píxel para verificar si es un máximo local en la dirección del gradiente. Si no lo es, el píxel se suprime, es decir, se convierte en cero (Figura 4.25) [91], [92].

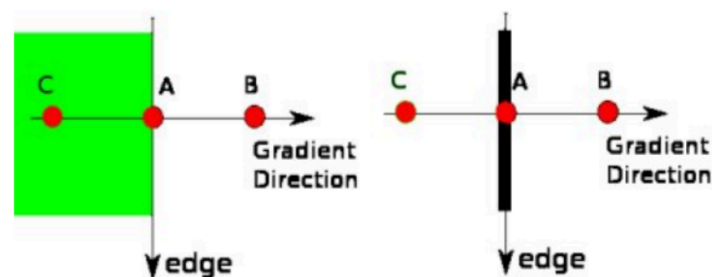


Figura 4.25 Un ejemplo de ello es esta imagen donde el punto A está en el borde (en dirección vertical). La dirección del gradiente es normal al borde. Los puntos B y C están en direcciones de gradiente. Por lo tanto se verifica el punto A con los puntos B y C para ver si forma un máximo local. Si es así, se considera para la siguiente etapa; de lo contrario, se suprime (se pone a cero) [92].

Finalmente se se definen los bordes reales y se eliminan los falsos positivos, es decir, se establecen los umbrales, el umbral alto donde los píxeles con magnitudes mayores a este valor son bordes fuertes y el umbral bajo donde los píxeles con magnitudes entre el umbral bajo y el alto son bordes débiles, y se mantienen solo si están conectados a un borde fuerte. Resultando en una imagen binaria donde los píxeles blancos representan los bordes y los píxeles negros corresponden al fondo [91].

En este caso, con el propósito de destacar los contornos relevantes en las imágenes de fondo de ojo y evaluar su utilidad para la segmentación de estructuras anatómicas, se aplicó el algoritmo de detección de bordes de Canny Edge Detector al conjunto de imágenes seleccionadas anteriormente y tras el preprocesamiento con CLAHE y el análisis de los canales individuales, la detección de bordes permitía identificar contornos significativos es otra técnica que puede ser aprovechada por los algoritmos posteriores. Para ello se tomó el conjunto de imágenes, se convirtieron en escala de grises y cada imagen fue suavizada utilizando un filtro Gaussian Blur con kernel de tamaño (5, 5) y desviación estándar $\sigma = 1.5$, con el fin de reducir el ruido de alta frecuencia. Posteriormente, se aplicó el algoritmo Canny Edge Detection con umbrales ajustados a 30 (inferior) y 70 (superior), valores determinados empíricamente tras un proceso exploratorio. Los resultados se muestran en la Figura 4.26 donde se encuentran las imágenes originales y sus versiones con bordes destacados, esto fue mediante el uso de OpenCV.

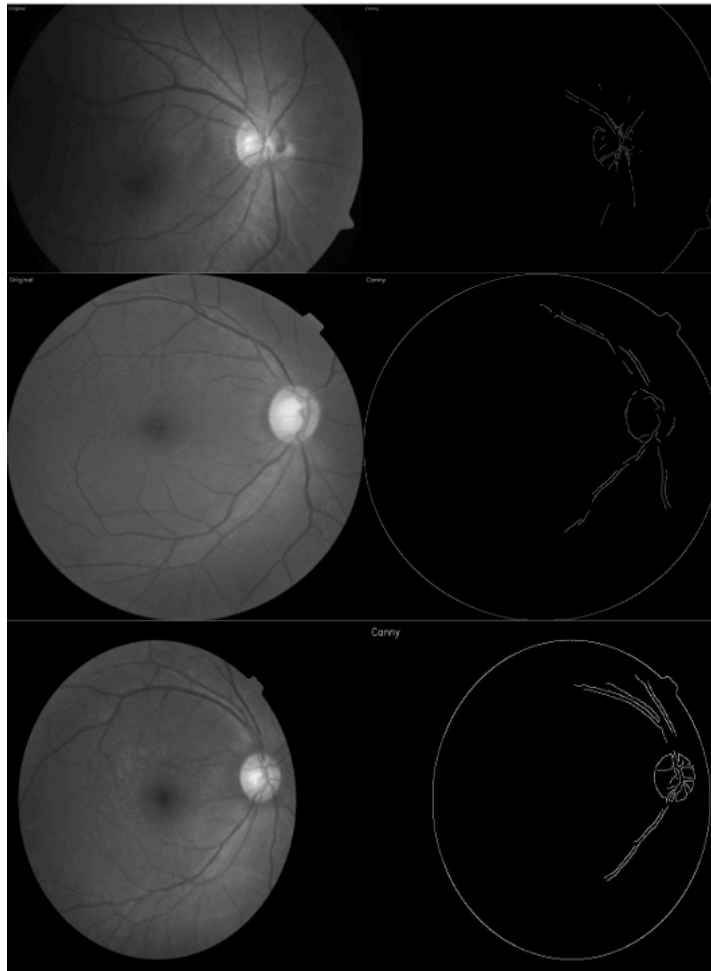


Figura 4.26 Aplicación del algoritmo Canny sobre imágenes de fondo de ojo. Se muestran, de izquierda a derecha, las imágenes originales en escala de grises y sus correspondientes versiones con los bordes detectados.

En general, la aplicación del algoritmo de Canny demostró ser una técnica de preprocesamiento útil ya que homogeniza de manera binaria todas las imágenes, es un apoyo visual en el análisis clínico preliminar, sin embargo, debido a la calidad de la imagen no existen bordes continuos que delimiten las estructuras de la imagen de una manera exitosa, quizá canny edge detector es óptima si es que se le aplica alguna otra técnica que mejore la calidad de la imagen.

4.9.1 Combinación de CLAHE y Detección de bordes mediante el algoritmo de Canny

Aunque la aplicación de Canny sobre imágenes en escala de grises permitió detectar bordes generales muy superficiales, se observó que muchas imágenes carecían de contraste o la calidad no era la adecuada o suficiente para una detección clara de estructuras. Por ello, se propuso aplicar primero CLAHE sobre cada canal RGB de la imagen original, para realzar los detalles locales, y posteriormente aplicar Canny sobre la imagen combinada con alto contraste.

Tras observar que el canal verde ofrece una mayor visibilidad de las estructuras anatómicas clave, como los son arcada vascular y el disco óptico, se decidió centrar el preprocesamiento avanzado

exclusivamente en este canal. La siguiente sección describe un flujo de procesamiento secuencial diseñado para resaltar dichas estructuras mediante técnicas de realce de contraste, segmentación y detección de bordes.

El flujo de trabajo para esta sección es el siguiente, se utiliza el conjunto de imágenes propuesto anteriormente y se realiza la separación de canales RGB y aplicación de CLAHE de forma independiente a cada uno, con parámetros clipLimit=2.0 y tileGridSize=(8,8) como se realizó anteriormente en las secciones previas. Hubo una combinación de los canales procesados en una nueva imagen RGB con contraste mejorado. Después se aplica el algoritmo Canny con los umbrales (lower_threshold=30, upper_threshold=70) sobre la imagen CLAHE y se convierte de la imagen de bordes a formato RGB para su visualización junto con las otras imágenes, es decir, imagen original, imagen con CLAHE aplicado, y la imagen de bordes detectados esto se puede visualizar en la Figura 4.27.

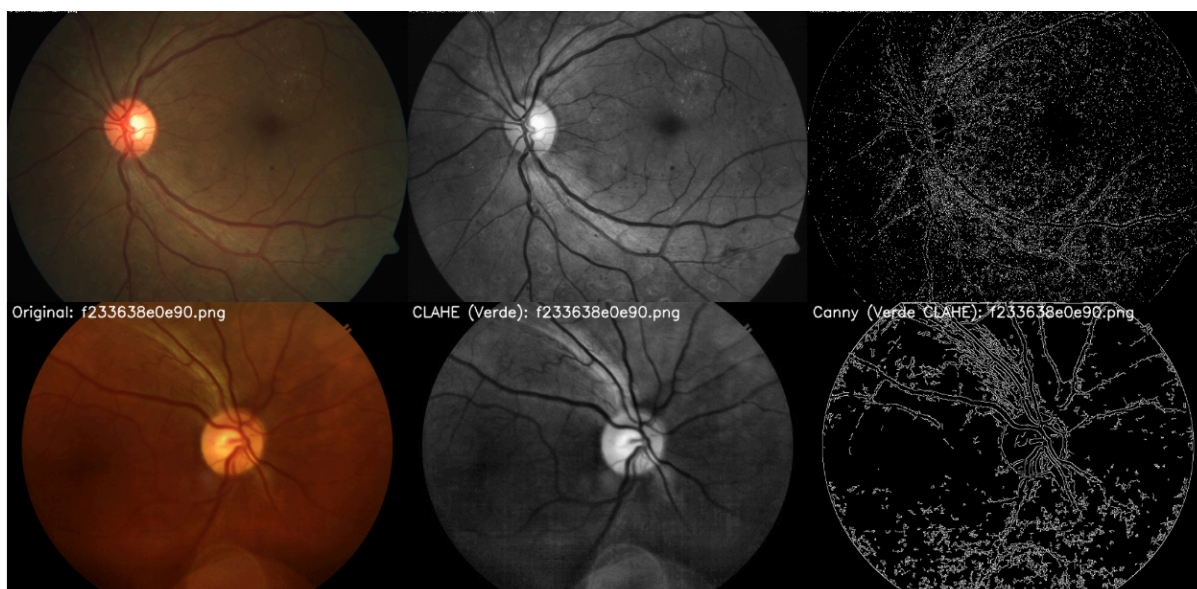


Figura 4.27 Flujo de trabajo de preprocesamiento combinado: mejora de contraste y detección de bordes. De izquierda a derecha se muestran: la imagen original de fondo de ojo, la imagen con contraste mejorado mediante CLAHE aplicado de forma independiente a los canales RGB, y la imagen resultante tras la detección de bordes mediante el algoritmo Canny.

Si bien la estrategia de aplicar CLAHE sobre los canales RGB seguida de la detección de bordes mediante el algoritmo de Canny permitió definir los contornos anatómicos como el disco óptico y segmentos de la arcada vascular, también generó un nivel considerable de ruido visual el cual hace que se pierdan estos contornos. Este ruido, presente principalmente en regiones con texturas irregulares o variaciones de iluminación, dificulta la identificación precisa de estructuras en algunas imágenes.

4.10 Preprocesamiento avanzado en el canal verde, extracción de bordes y detección de estructuras

Dado que el canal verde es algo que en todas las técnicas utilizadas ha destacado en la visibilidad de las estructuras anatómicas, se exploró un flujo de preprocesamiento avanzado, incorporando técnicas de segmentación binaria y filtrado morfológico con el objetivo de mejorar la delimitación de las estructuras de interés y reducir el ruido presente en imágenes de la base de datos APTOS.

4.10.1 Preprocesamiento secuencial sobre el canal verde para realce y detección de estructuras anatómicas

Para mejorar la visibilidad de las estructuras en las imágenes de fondo de ojo y facilitar la posterior segmentación, se implementó un procedimiento de preprocesamiento que incluye las siguientes etapas, extracción y selección del canal verde, aplicación de CLAHE, detección de bordes por el algoritmo Canny y esta vez se utiliza la umbralización binaria. Esta técnica (convierte una imagen en escala de grises en una imagen binaria es decir a blanco y negro) usando un valor de umbral. Los píxeles con valores superiores al umbral se asignan a un valor máximo el cual será blanco, mientras que los píxeles con valores inferiores se asignan a 0 es decir, negro [93].

Este flujo de trabajo de la misma manera que se viene manejando desde las técnicas pasadas fue aplicado a un conjunto de imágenes de la base de datos seleccionada, donde se toma el canal verde de cada imagen RGB, sobre esa imagen se aplica CLAHE la cual mejora el contraste local sin simplificar excesivamente el ruido el cual era uno de los problemas anteriores. Para este proceso se utilizó un clip limit de 2.0 y un tamaño de mosaico de 8x8 píxeles. Después se introduce esta técnica llamada umbralización binaria donde la imagen resultante de CLAHE fue sometida a una umbralización binaria simple, utilizando un umbral inferior de 50, esto quiere decir que todos los números menores del umbral se vuelven negro y superior de 255, donde todos los píxeles superiores a este umbral se vuelvan blanco, Esta operación permite separar regiones de alta intensidad, que son las regiones de interés, con esta técnica se espera resaltar los contornos para que con el algoritmo Canny y los umbrales 30 como mínimo y 70 como máximo puedan obtener los contornos de la arcada vascular y el nervio óptico. A continuación en la Figura 4.28 se puede observar la imagen original, CLAHE desde el canal Verde, la umbralización binaria creada en la imagen anterior y finalmente el algoritmo Canny para los bordes.

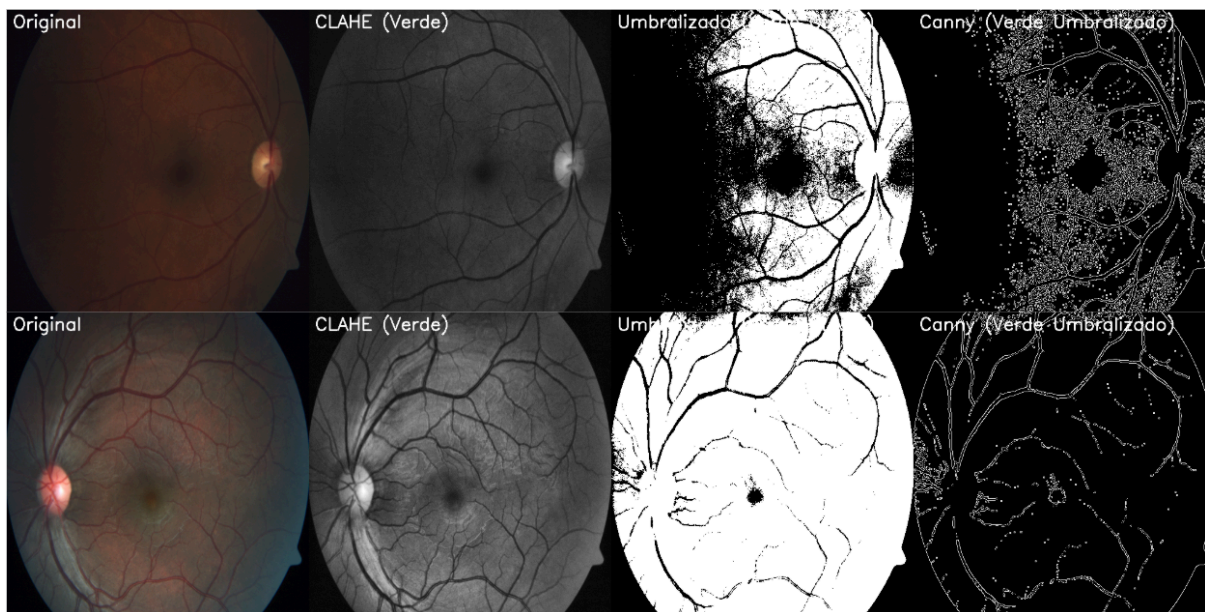


Figura 4.28 Ejemplo representativo de dos imágenes del flujo de preprocesamiento avanzado aplicado sobre el canal verde. De izquierda a derecha: imagen original en color, canal verde mejorado mediante CLAHE, imagen umbralizada binaria del canal verde, y resultado de la detección de bordes con Canny.

El uso combinado de CLAHE, umbralización binaria y el algoritmo Canny resultó en bordes más definidos y una mejor separación entre fondo y estructuras relevantes. No obstante, en algunas imágenes se observó la aparición de ruido constante o bordes discontinuos, o también solo definió la mitad de la imagen más cercana al nervio óptico, especialmente cuando el contraste inicial era bajo. A pesar de estas limitaciones, este flujo de trabajo constituye una base más estable para etapas posteriores de segmentación automatizada.

También, existe otra técnica llamada umbralización adaptativa la cual es una técnica de binarización en procesamiento de imágenes que se utiliza cuando la iluminación no es uniforme en toda la imagen, a diferencia de la umbralización binaria que aplica el mismo valor de corte a todos los píxeles, este calcula un umbral local diferente para cada píxel, basado en la intensidad de los píxeles vecinos, es decir Se toma una pequeña vecindad alrededor del píxel, se calcula una estadística local y se aplica el valor constante [94].

Este flujo de trabajo mantiene la misma estructura secuencial aplicada en etapas previas, basada en la extracción del canal verde, el realce de contraste mediante CLAHE y la posterior detección de bordes con el algoritmo de Canny. Sin embargo, en lugar de aplicar la umbralización binaria se incorpora la umbralización adaptativa, la cual permite una binarización más sensible a las variaciones locales de intensidad, especialmente útil en imágenes con iluminación no uniforme como se explicaba en el párrafo anterior. Cada una de las etapas fue visualizada y etiquetada para su análisis comparativo, generando una composición que incluye: la imagen original, la imagen del canal verde con CLAHE, el resultado de la umbralización adaptativa invertida y la imagen final con bordes detectados. Esta visualización se presenta en la Figura 4.29, la cual permite observar con claridad la evolución del procesamiento y la mejora en la detección de estructuras anatómicas.

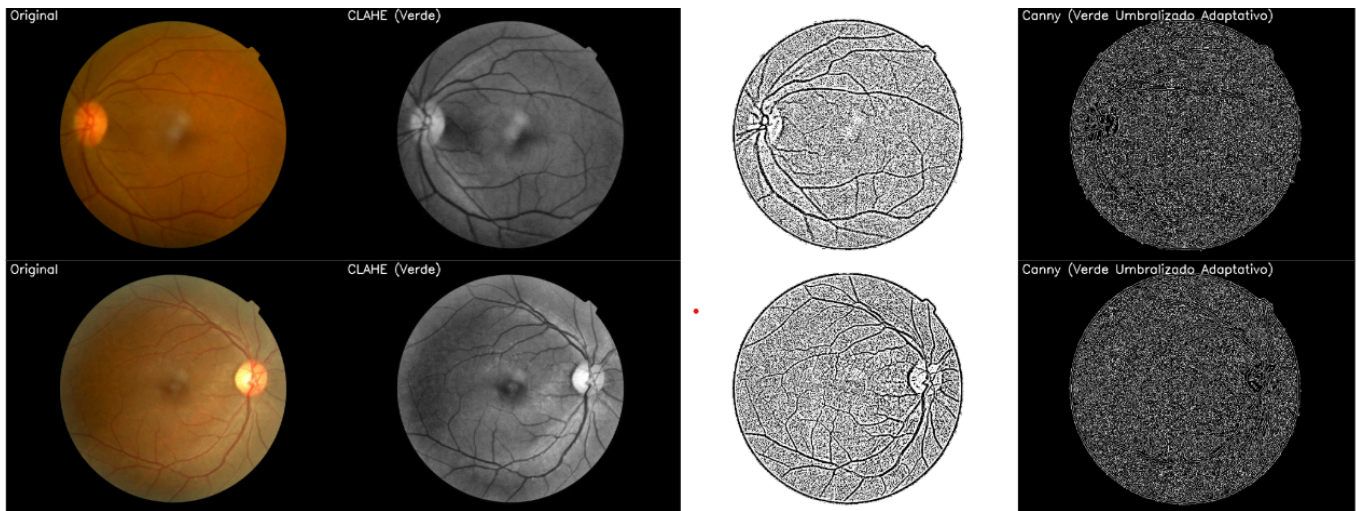


Figura 4.29 Ejemplo representativo de dos imágenes del flujo de preprocesamiento avanzado aplicado sobre el canal verde. De izquierda a derecha: imagen original, canal verde con CLAHE, umbralización adaptativa invertida, y bordes detectados mediante Canny.

A pesar de los avances logrados mediante el uso de CLAHE en el canal verde para mejorar el contraste y de la aplicación de umbralización binaria y consecuentemente adaptativa, seguida de detección de bordes con el algoritmo de Canny, los resultados obtenidos no fueron óptimos para una segmentación de las características confiable de las estructuras anatómicas en las imágenes de fondo de ojo. En el caso de la umbralización binaria, el método mostró una alta sensibilidad a variaciones de iluminación y contraste entre imágenes, lo que resultó en contornos incompletos o fragmentados, así como el constante ruido esporádico. Por otro lado, aunque la umbralización adaptativa introdujo una mayor flexibilidad al ajustar el umbral localmente, esto también generó una sobre segmentación de regiones irrelevantes, amplificando el ruido estructural y reduciendo la especificidad del proceso.

4.10.2 Preprocesamiento avanzado: umbralización adaptativa, apertura morfológica y detección de bordes en canal verde

Si bien las etapas anteriores con CLAHE y el algoritmo Canny permitieron una visualización general de los bordes anatómicos aunque fueron discontinuos o difusos, aún persistía la presencia de ruido estructural y artefactos que podían entorpecer el análisis automático. Por ello, aquí se integraron técnicas adicionales como umbralización adaptativa la cual se menciono anteriormente, inversión de binarios donde se realiza una umbralización adaptativa solamente que ahora utilizando un umbral inferior de 50, esto quiere decir que todos los numeros menores del umbral se vuelven blanco y superior de 255 se vuelven negro, donde todos los pixeles superiores a este umbral se vuelvan blanco, y operaciones morfológicas que permiten limpiar la imagen antes de aplicar el detector de bordes.

Las operaciones morfológicas, en este caso se centran en una denominada como Apertura, es una operación fundamental en el procesamiento morfológico de imágenes binarias la cual se centra

principalmente para eliminar ruido pequeño como lo son los puntos aislados como vimos en la umbralización adaptativa, esto con la finalidad de no afectar las estructuras de interés y suavizar los bordes de los objetos sin afectar demasiado su forma general[95].

Este flujo de trabajo de la misma manera que se viene manejando desde las técnicas pasadas fue aplicado a un conjunto de imágenes de la base de datos seleccionada, se extrae canal verde debido a su contraste para las regiones de interés, se aplica CLAHE, con parámetros clipLimit=2.0 y tileGridSize=(8,8), para mejorar el contraste local, Se realiza una umbralización adaptativa la cual ajusta Se utilizó cv2.adaptiveThreshold con método Gaussiano y valores de bloque 11 y constante 2 para binarizar la imagen, en este caso se realiza una inversión binaria donde de la imagen resultante, de modo que las estructuras de interés aparezcan en blanco sobre fondo negro. Después se realiza la apertura morfológica esto se hace por medio de un kernel 3 x 3 píxeles para eliminar objetos pequeños aislados y ruido espurio, conservando las estructuras anatómicas relevantes y se termina con la detección de bordes mediante el algoritmo de Canny, sus umbrales se mantienen de igual manera que las secciones anteriores (Figura 4.30).

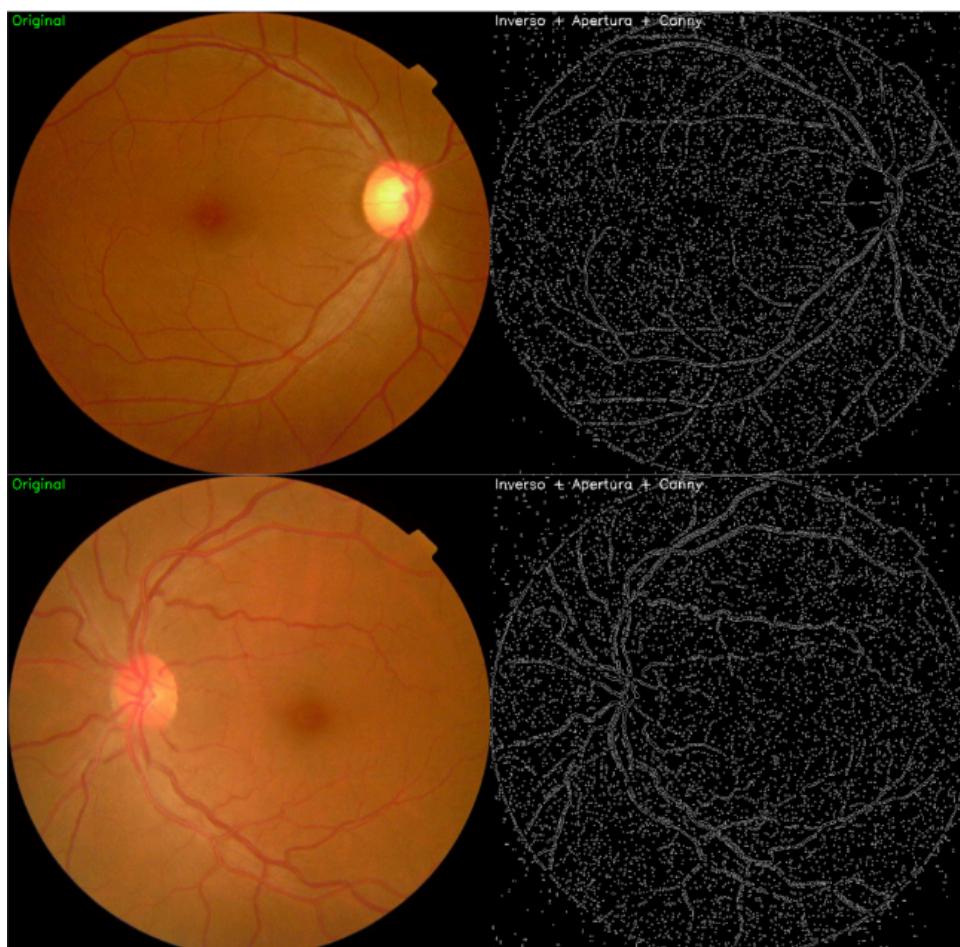


Figura 4.30 Comparación entre la imagen original de fondo de ojo (izquierda) y el resultado final tras aplicar un flujo de preprocesamiento compuesto por extracción del canal verde, realce de contraste

con CLAHE, umbralización adaptativa, inversión binaria, apertura morfológica y detección de bordes con el algoritmo de Canny (derecha).

Después de implementar y evaluar distintas de las combinaciones mencionadas para el análisis exploratorio de los datos y preprocesamiento, se observó que, si bien estas técnicas permiten una visualización general de las estructuras, no son lo suficientemente robustas para aproximarse a una segmentación confiable de las estructuras anatómicas de interés. Muchos de los resultados fueron inconsistentes, ya que las imágenes creadas eran mayormente discontinuas, afectadas por ruido y con baja capacidad de distinguir estructuras finas como vasos sanguíneos del fondo. Incluso tras aplicar aperturas morfológicas, muchas regiones presentaban bordes fragmentados o amplificación de artefactos.

Sin embargo, el uso de CLAHE sobre el canal verde permitió que muchas imágenes que originalmente no tenían una calidad muy buena, o que eran oscuras pasaran a ser clasificadas como de buena o aceptable calidad. En conjunto, los hallazgos sugieren que estas técnicas tradicionales, aunque útiles para un preprocesamiento preliminar y exploratorio, no son suficientes por sí solas para alcanzar una segmentación anatómica precisa. De esta manera, se hace aún más evidente el uso de métodos más avanzados como lo son las técnicas basadas en aprendizaje automático o aprendizaje profundo que puedan manejar la variabilidad y complejidad de las imágenes de fondo de ojo.

4.11 Evaluación de algoritmos seleccionados

Una vez terminada la fase donde se realizó el análisis exploratorio de los datos, se se procedió a la implementación y evaluación de distintos algoritmos de segmentación con el objetivo de identificar estructuras anatómicas relevantes en imágenes de fondo de ojo, tales como la arcada vascular y el nervio óptico. Se reitera que la selección de estos algoritmos fue en base a la relevancia que tuvieron en la revisión de la literatura. El objetivo de esta fase fue validar experimentalmente los algoritmos de segmentación seleccionados utilizando imágenes representativas de la práctica clínica regional, evaluando su rendimiento mediante métricas comúnmente utilizadas en estudios previos, como la exactitud, sensibilidad, especificidad y el coeficiente de Dice, todos estos mencionados en la revisión de la literatura . Este enfoque permitió contextualizar los resultados obtenidos en función de las necesidades locales y los estándares de calidad.

4.12 Implementación de los modelos seleccionados.

4.12.1 U-Net

De acuerdo a la comparación que se realizó en la sección de la revisión de la literatura, se identificó que las arquitecturas basadas en U-Net propuesta en 2015 [51], son las más ampliamente adoptadas en el ámbito médico, ya que demuestran ser altamente eficaces en la detección y delimitación de estructuras el cuerpo humano. Todo esto debido a su estructura modular y flexible, lo que permite adaptar su arquitectura a diferentes resoluciones, canales de entrada y requerimientos computacionales.

Una U-Net es una arquitectura de red neuronal convolucional diseñada específicamente para tareas de segmentación semántica, como identificar y delinear estructuras anatómicas en imágenes médicas. Se caracteriza por su forma de 'U', que combina un camino de compresión o encoder, para extraer características clave, y un camino de expansión o decoder, que reconstruye una segmentación detallada píxel a píxel [96].

De una manera detallada en la Figura 4.31 se puede observar cómo se desarrolla su arquitectura, donde consta de una ruta de contracción (lado izquierdo) y una ruta de expansión (lado derecho). La **ruta de contracción** sigue la arquitectura típica de una red convolucional. Consiste en la aplicación repetida de dos convoluciones de 3x3 (convoluciones sin relleno), cada una seguida por una unidad lineal rectificadora (ReLU) y una operación de max pooling de 2x2 con paso (stride) 2 para la reducción de la resolución. En cada paso de reducción, se duplica el número de canales de características. Cada paso en la **ruta de expansión** consiste en un muestreo ascendente (upsampling) del mapa de características, seguido por una convolución de 2x2 (“up-convolution”) que reduce a la mitad el número de canales de características, una concatenación con el mapa de características correspondiente recortado desde la ruta de contracción, y dos convoluciones de 3x3, cada una seguida por una ReLU. El recorte es necesario debido a la pérdida de píxeles en los bordes en cada convolución. En la capa final se utiliza una convolución de 1x1 para mapear cada vector de características de 64 componentes al número deseado de clases. En total, la red tiene 23 capas convolucionales [51].

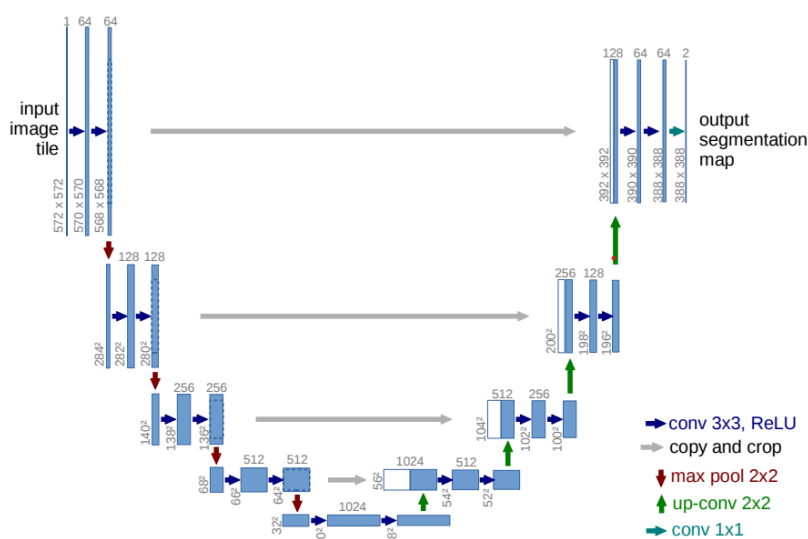


Figura 4.31 Arquitectura de U-Net , cada caja azul que contiene es un mapa de características multicanal. En la parte superior de cada caja se encuentra el número de canales y el tamaño en X y Y se tiene en la esquina de abajo de cada caja. Por otro lado, las cajas blancas representan mapas de características copiados, y las flechas indican las operaciones [51].

En resumen, el desempeño de la U-Net está directamente relacionado con la cantidad y calidad de las imágenes etiquetadas usadas para entrenarla. Esto pudo ser un reto en el caso de esta tesis, considerando la disponibilidad limitada de datos etiquetados en la región. Para abordar la segmentación de la arcada vascular en imágenes de fondo de ojo, se implementó esta arquitectura, El

flujo de trabajo incluyó tres etapas principales: preprocesamiento de datos, entrenamiento del modelo y evaluación de resultados.

4.12.2 Metodología del algoritmo U-Net

Formato de las imágenes

El proceso de entrenamiento propuesto en la arquitectura U-Net original se basa en el uso de imágenes junto con sus respectivas máscaras de segmentación. La base de datos elegida APTOS no cuenta con máscaras, por lo tanto el primer paso para utilizar este algoritmo fue la creación de las mismas. Para llevar a cabo el proceso de etiquetado, se utilizó la plataforma Roboflow, ampliamente utilizada en proyectos de visión artificial, se realizó una anotación manual en cada una de las imágenes seleccionadas, en este caso fueron 100 imágenes tomadas aleatoriamente en la base de datos APTOS. Se realizó una anotación manual en cada imagen, utilizando las herramientas principales de esta plataforma, en el caso de esta tesis solamente se utilizó polígonos debido a que esta herramienta facilitaba segmentar con mayor precisión las regiones anatómicas relevantes, se integró una única clase de interés, la arcada vascular, principalmente por razones prácticas relacionadas con la simplicidad del entrenamiento. A continuación en la Figura 4.32 se puede ver el proceso de anotaciones.

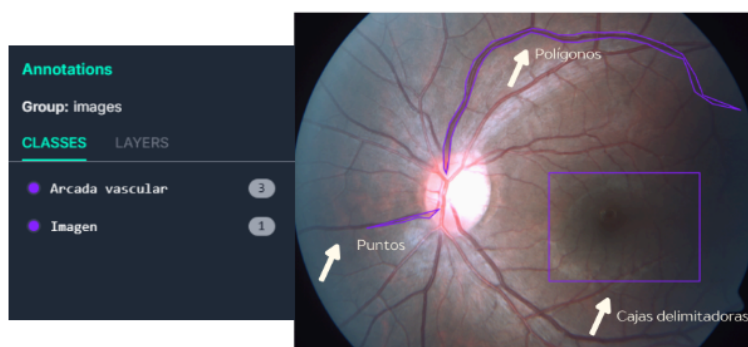


Figura 4.32 Imagen representativa del proceso de etiquetar las 100 imágenes requeridas para el entrenamiento de U-Net. Podemos ver las distintas herramientas proporcionadas por la plataforma y la división de clases del algoritmo.

Una vez que se validaron las etiquetas para garantizar su precisión y completitud, el conjunto de datos anotado se exportó en el formato YOLOv8 (You Only Look Once versión 8), compatible con los frameworks de aprendizaje profundo utilizados en este trabajo. Este formato hace que cada imagen etiquetada se asocie a un archivo de texto .txt con el mismo nombre, donde se describen los objetos presentes mediante líneas que contienen los siguientes elementos:

- **Clase:** Número de clase que identifica el tipo de objeto.
- **Coordenadas normalizadas** del centro del objeto: x_center y y_center , expresadas en relación al ancho y alto de la imagen, con valores entre 0 y 1.
- **Dimensiones normalizadas del objeto:** width y height, también en valores entre 0 y 1.

Una vez finalizado el proceso de etiquetado y exportadas las anotaciones se procedió a la generación de las máscaras binarias correspondientes a cada imagen del conjunto de datos. A partir de las coordenadas y clases especificadas en los archivos .txt, se crearon máscaras de segmentación en formato imagen. El proceso de conversión se automatizó mediante scripts en Python utilizando librerías como OpenCV y NumPy, que permiten interpretar las coordenadas normalizadas, proyectarlas sobre el espacio de píxeles de la imagen original y construir la máscara correspondiente. En la Figura 4.33 se puede observar algunos de los resultados de las máscaras creadas con segmentaciones manuales.

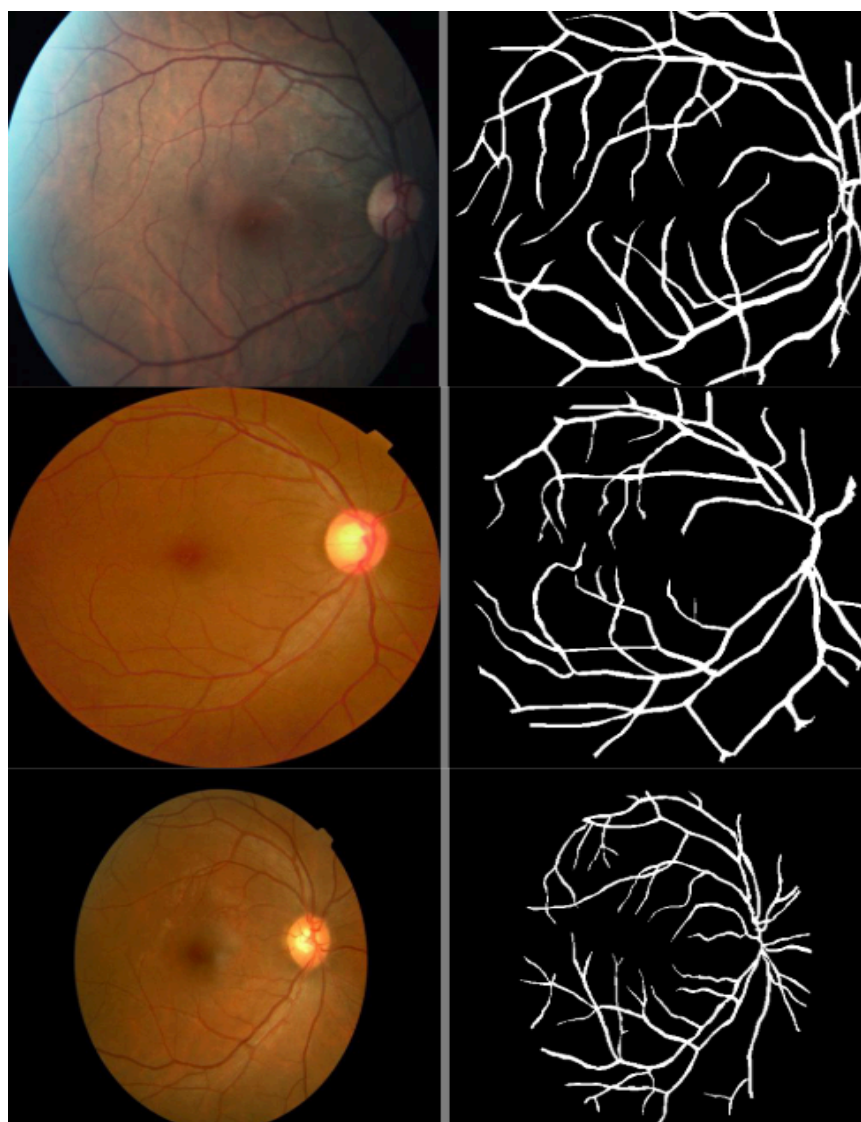


Figura 4.33 Imagen de fondo de ojo (izquierda) y su correspondiente máscara de segmentación de la arcada vascular (derecha) generada a partir de las anotaciones en formato YOLOv8.

Instalación de requerimientos

Para esta sección se muestra la siguiente Tabla 4.20 con los requerimientos para el funcionamiento de la U-Net.

Paquete	Versión recomendada
torch	≥ 1.8
torchvision	≥ 0.9
opencv-python	≥ 4.5
numpy	≥ 1.19
matplotlib	≥ 3.3
albumentations	≥ 0.5

Tabla 4.20 Requerimientos para U-Net implementada en este trabajo.

Preprocesamiento

Antes de entrenar el modelo, fue necesario aplicar un conjunto de transformaciones a las imágenes con el fin de mejorar la calidad visual, reducir el ruido y facilitar la segmentación de las estructuras vasculares. Este preprocesamiento se aplicó tanto a las imágenes originales como a las máscaras asociadas.

Primeramente las imágenes fueron convertidas a formato RGB y las máscaras a escala de grises para la compatibilidad de las funciones de PyTorch, después tanto las imágenes como sus máscaras fueron redimensionadas a una resolución estándar de 512×512 píxeles debido a las especificaciones de la arquitectura U-Net, se normalizaron los valores pixelares de las imágenes en un rango de $[0,1]$ dividiendo por 255. Las imágenes y máscaras fueron reorganizadas en el formato tensorial requerido por PyTorch: las imágenes RGB como tensores de tamaño $(3, H, W)$ y las máscaras binarias como $(1, H, W)$.

Entrenamiento

El entrenamiento del modelo U-Net se llevó a cabo utilizando el *framework* PyTorch, Se implementó un enfoque supervisado, en el cual el modelo aprendió a segmentar estructuras vasculares mediante la exposición a pares de imágenes de fondo de ojo y sus correspondientes máscaras binarias que fueron construidas en la sección anterior, Se utilizaron un total de 100 imágenes de las cuales se dividieron de la siguiente manera: 70% entrenamiento, 20% prueba y 10% validación.

Para este proceso, se establecieron los siguientes hiperparámetros: las imágenes de entrada fueron redimensionadas a 512×512 píxeles, se utilizó un tamaño de lote (batch size) de 2 imágenes, y el entrenamiento se realizó durante un total de 50 épocas. Se empleó el optimizador Adam con una tasa de aprendizaje inicial de 1×10^{-4} . La función de pérdida seleccionada fue una combinación entre Dice Loss y Binary Cross-Entropy Loss (DiceBCELoss), con el objetivo de mejorar la segmentación tanto a nivel local como global.

El flujo de entrenamiento siguió una estructura clara. Primero, las imágenes y sus máscaras fueron cargadas y organizadas en lotes utilizando los DataLoaders de PyTorch, lo que permitió una alimentación eficiente del modelo durante el proceso de entrenamiento. Posteriormente, cada lote de imágenes fue procesado a través del modelo U-Net para generar las predicciones correspondientes (forward pass). A continuación, se calculó la pérdida entre las predicciones del modelo y las máscaras reales empleando la función DiceBCELoss. Los gradientes de error resultantes fueron transmitidos hacia capas anteriores (backpropagation) a través de la red, permitiendo la actualización de los pesos mediante el optimizador Adam. Una vez completada cada época, se evaluó el rendimiento del modelo utilizando el conjunto de validación, con el fin de monitorear su comportamiento y mitigar el riesgo de sobreajuste. Si se observaba una mejora en el rendimiento durante la validación, los pesos del modelo eran almacenados en disco, permitiendo su recuperación posterior para la fase de prueba.

4.12.3 Resultados con U-Net

A continuación, se presentan ejemplos visuales de los resultados obtenidos por el modelo U-Net tras el proceso de entrenamiento (Figuras 4.34 a la 4.37). En cada una de estas imágenes se observa, de izquierda a derecha, la imagen de fondo de ojo original, la máscara de referencia construida manualmente y la máscara segmentada predicha por el modelo. En general, se aprecia que la red U-Net logra reproducir con precisión la morfología de los vasos sanguíneos principales. Sin embargo, se identifican algunas limitaciones recurrentes, particularmente en la detección de ramificaciones vasculares más delgadas o de bajo contraste, lo que sugiere una menor sensibilidad en estructuras de pequeño calibre. Estos resultados permiten valorar visualmente la calidad de segmentación alcanzada, así como identificar áreas específicas donde se podrían incorporar mejoras, como técnicas de realce, ajustes en la función de pérdida o esquemas de entrenamiento más sofisticados.



Figura 4.34. Resultado de segmentación vascular obtenido mediante U-Net. Se muestra la imagen original (izquierda), la máscara de referencia (centro) y la máscara predicha por el modelo (derecha). La predicción reproduce adecuadamente la morfología de los vasos principales, con algunas pérdidas en ramificaciones delgadas.



Figura 4.35 Resultado de segmentación vascular obtenido mediante U-Net. Se muestra la imagen original (izquierda), la máscara de referencia (centro) y la máscara predicha por el modelo (derecha). La predicción reproduce adecuadamente la morfología de los vasos principales, con algunas pérdidas en ramificaciones delgadas.

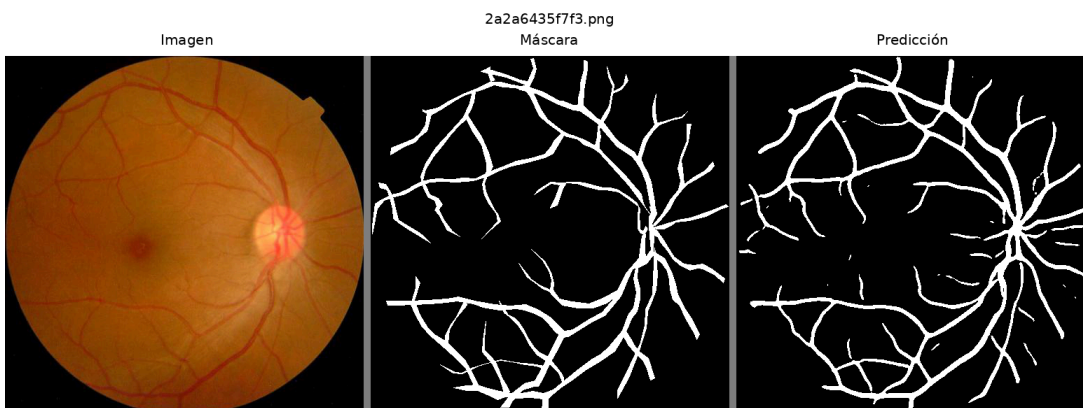


Figura 4.36 Resultado de segmentación vascular obtenido mediante U-Net. Se muestra la imagen original (izquierda), la máscara de referencia (centro) y la máscara predicha por el modelo (derecha). La predicción reproduce adecuadamente la morfología de los vasos principales, con algunas pérdidas en ramificaciones delgadas.



Figura 4.37 Resultado de segmentación vascular obtenido mediante U-Net. Se muestra la imagen original (izquierda), la máscara de referencia (centro) y la máscara predicha por el modelo (derecha). La predicción reproduce adecuadamente la morfología de los vasos principales, con algunas pérdidas en ramificaciones delgadas.

4.12.4 nnUnet

Aunque inicialmente se implementó la arquitectura U-Net para la segmentación de vasos sanguíneos, se decidió migrar a la arquitectura nnU-Net ya que debido a sus capacidades de auto-configuración, permiten optimizar automáticamente la arquitectura, los hiperparámetros y el preprocesamiento según las características específicas del conjunto de datos. En esta arquitectura se realizó la segmentación de 2 estructuras anatómicas de interés: Arcada vascular y nervio óptico.

En este contexto, cabe destacar el trabajo publicado en 2021 titulado “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”[97], presentó un marco automatizado que adapta una arquitectura basada en U-Net que originalmente fue diseñada para imágenes médicas con intervención manual mínima. Esta arquitectura, reconocida por su forma simétrica en “U”, combina un codificador y un decodificador conectados mediante conexiones de salto, lo que permite preservar detalles espaciales durante el proceso de segmentación.

Tradicionalmente la implementación de un modelo para una tarea específica requiere un proceso previo de diseño y configuración manual. Este proceso demanda un alto nivel de conocimiento especializado y experiencia, ya que pequeños errores pueden provocar caídas significativas en el rendimiento. Además, los ajustes del flujo de trabajo óptimos para un conjunto de datos pueden no generalizarse a otros, lo que hace necesario optimizarlos nuevamente para cada conjunto de datos de manera individual. nnU-Net aborda este desafío mediante configuraciones automatizadas rápidas y eficientes, ya que comprende una ejecución de reglas predefinidas que minimizan la necesidad de decisiones manuales. Se trata de un método holístico, ya que cubre todo el proceso de segmentación, incluyendo el preprocesamiento, arquitectura de red, entrenamiento y posprocesamiento. Su funcionamiento se basa en un proceso de configuración neutral a la tarea, lo que reduce considerablemente el número de opciones de diseño que deben evaluarse empíricamente al enfrentar un nuevo caso. Esto lo hace adaptable a distintos escenarios sin necesidad de modificar su estructura base, facilitando su aplicación a nuevos conjuntos de datos. Para ello, el diseño de nnU-Net se fundamenta en tres principios principales para la gestión de decisiones:

1. **Parámetros fijos:** aquellos que permanecen constantes entre diferentes conjuntos de datos, como la arquitectura base del modelo.
2. **Parámetros basados en reglas:** configurados automáticamente según las características específicas de cada conjunto de datos.
3. **Parámetros empíricos:** ajustados mediante validación durante el proceso de entrenamiento para optimizar el desempeño.

La arquitectura nnUnet no realiza una optimización específica para cada tarea, lo que puede restringir su desempeño en casos particulares. También, está diseñado para maximizar la métrica Dice, lo que podría no ser ideal para tareas que requieren métricas especializadas. Su configuración automática, aunque potente, no siempre se adapta de forma óptima a segmentaciones muy complejas o específicas,

lo que puede demandar ajustes manuales. Por último, debido a la gran diversidad de datos biomédicos, nnU-Net no reemplaza la necesidad de realizar evaluaciones amplias y específicas en diferentes tipos de datos. Sin embargo, su enfoque automatizado y adaptable ha representado un avance importante en la facilitación de la segmentación en una amplia variedad de imágenes médicas sin necesidad de diseño manual extenso [97]. La Figura 4.38 ilustra de manera detallada el flujo de trabajo del modelo nnU-Net.

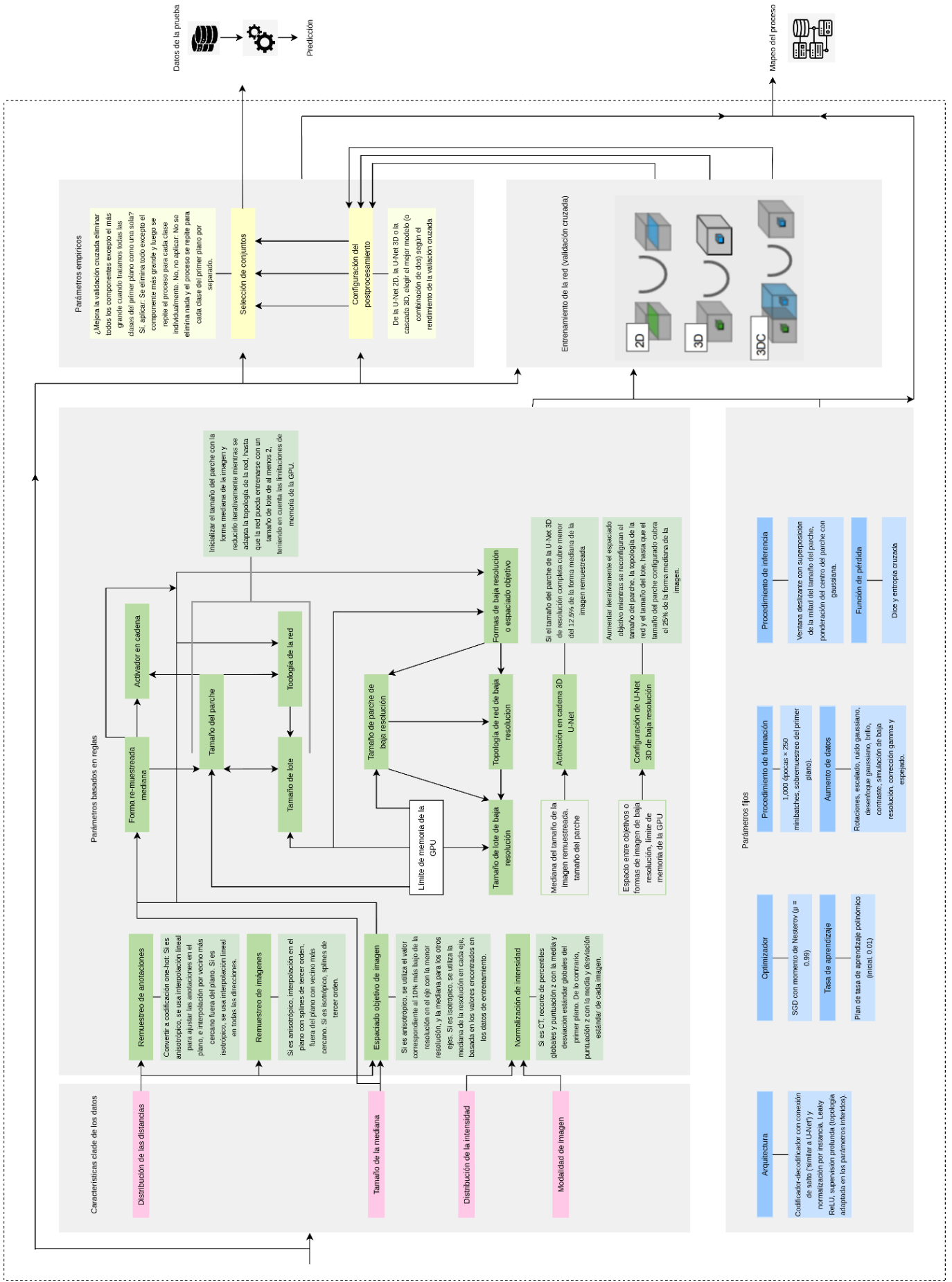


Figura 4.38 Diagrama de flujo del modelo nnU-Net, adaptado del original presentado en [97] .

4.12.5 Metodología del algoritmo nnU-Net

Elección de modelo

Se seleccionó nnU-Net debido a su enfoque automatizado y adaptativo, que facilita la segmentación en imágenes médicas sin requerir una configuración manual extensa. Se realiza la estrategia de entrenamiento según sus características y objetivos. nnU-Net se entrena desde cero porque fue diseñado para esa tarea, adaptándose a nuevos datos fácilmente.

Formato de imagen

Aunque la base de datos original contiene 3,662 muestras, para el presente estudio se seleccionó aleatoriamente un conjunto de 500 imágenes de la base de datos APTOS. El conjunto de imágenes para las máscaras fue etiquetado manualmente utilizando la Herramienta de Anotación para Visión por Computadora (CVAT, por sus siglas en inglés), A continuación en la Figura 4.39 se observa el funcionamiento de la plataforma y el conjunto de imágenes etiquetadas con sus clases.

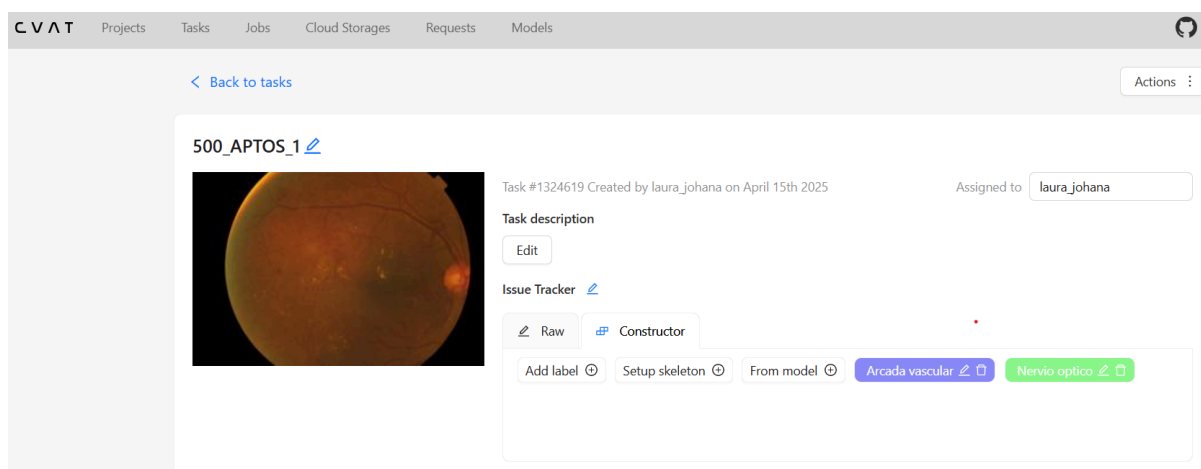


Figura 4.39 Herramienta de Anotación para Visión por Computadora con el conjunto de datos APTOS de 500 imágenes.

A partir de este proceso de anotación, se generaron máscaras con dos clases de etiquetas: máscara vascular y nervio óptico. Estas anotaciones se descargan en formato JSON ,es decir, una estructura de texto que representa datos organizados mediante pares clave-valor, listas y objetos anidados en este caso conteniendo únicamente las coordenadas asociadas a cada imagen, las cuales fueron procesadas posteriormente y convertidas al formato PNG para su uso en el modelo. La Figura 4.40 es una muestra de las máscaras obtenidas.

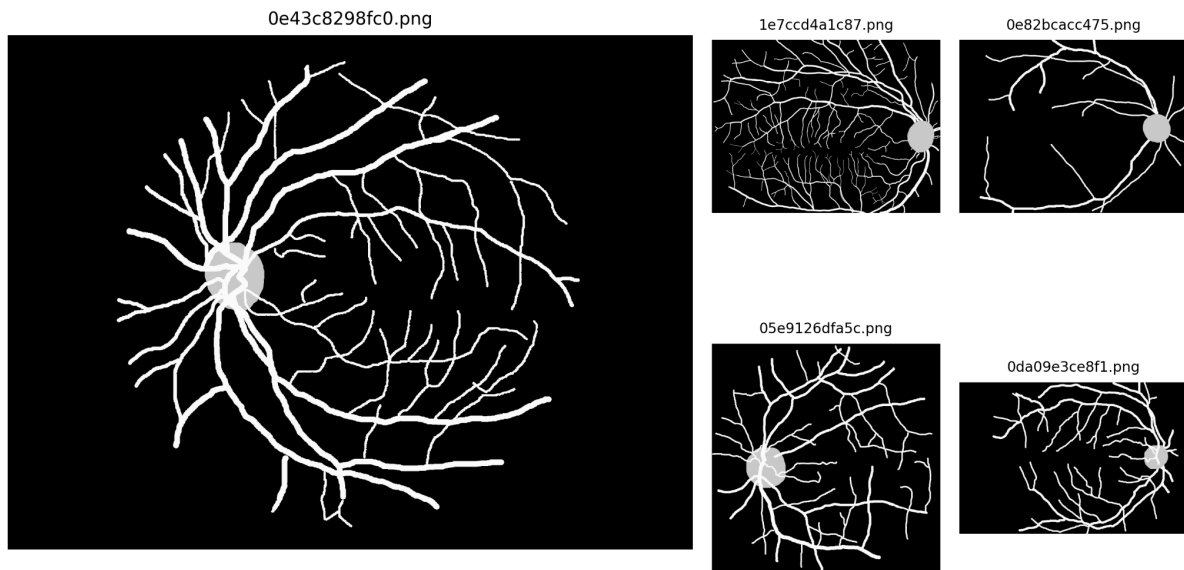


Figura 4.40 Muestra de máscaras manuales de las estructuras anatómicas: arcada vascular y nervio óptico.

Las imágenes originales descargadas de APTOS en una matriz tridimensional en formato RGB fueron convertidas a escala de grises. Esta decisión se tomó tanto por razones técnicas como prácticas: durante el preprocesamiento, se presentaron errores al cargar imágenes en una matriz tridimensional en formato RGB debido a inconsistencias en el número de canales esperados por nnU-Net. Al convertir las imágenes a escala de grises, cada una pasa a tener un solo canal, ajustándose al formato estándar de nnU-Net. Esto evitó errores relacionados con el número de canales, y eliminó la necesidad de especificar múltiples entradas. Dado que la segmentación de la arcada vascular puede realizarse de manera efectiva en escala de grises, como es común en la literatura, esta alternativa resulta más simple para evitar complicaciones durante el flujo de trabajo. Además, la arcada vascular se caracteriza principalmente por sus contrastes y formas, más que por su información cromática.

Para cumplir con los requisitos de entrada de nnUN-Net, se integraron las dos clases de interés (arcada vascular y nervio óptico) en una única máscara por imagen, donde cada píxel fue etiquetado con un valor entero que representa una clase específica según la siguiente codificación:

1. **Arcada vascular:** 1
2. **Nervio óptico:** 2
3. **Fondo(sin clases):** 0

Instalación de requerimientos

Aunque nnU-Net automatiza gran parte del flujo de trabajo de segmentación, es necesario realizar una fase inicial de preparación de datos. Esta fase consiste en adecuar el conjunto de datos a una estructura específica que nnU-Net reconoce, lo cual permite que el sistema realice automáticamente el preprocesamiento y entrenamiento, así como el almacenamiento correcto de datos y resultados. Esta fase incluyó los siguientes pasos:

1. Creación de las carpetas que reconoce nnU-Net

nnUNet_raw se utiliza para almacenar los datos originales y la información de entrada al modelo; nnUNet_preprocessed contiene los datos que ya han sido procesados; nnUNet_results almacena los modelos entrenados y los resultados de validación, mientras que nnUNet_predictions se emplea para guardar las inferencias realizadas sobre nuevos datos.

2. Instalación de requisitos

Se instaló python 3.11, ya que nnU-Net requiere una versión mínima de Python 3.9 para funcionar correctamente. Y para asegurar la compatibilidad, se instalaron manualmente algunas bibliotecas esenciales: nibabel (versión 5.2.1), SimpleITK (2.3.1) y numpy (1.26.4).

3. División del conjunto de datos para entrenar/validar/probar

70 % de los datos se asignaron al conjunto de entrenamiento, utilizado para ajustar los pesos del modelo. El 20 % se destinaron a validación, permitiendo evaluar el rendimiento durante el entrenamiento y ajustar hiperparámetros. Por último, 10 % se reservaron para el conjunto de prueba.

4. Renombramiento archivos para compatibilidad con nnU-Net

Dentro de la ruta de nnUNet_raw se crea una carpeta principal nombrada con un ID único de 200 (debe estar entre 0 y 999) y un nombre descriptivo. Esta carpeta contiene tres subcarpetas: imagesTr que almacena las imágenes para entrenamiento, labelsTr la cual contiene las máscaras para entrenamiento, e imagesTs que guarda las imágenes para prueba.

5. Creación del archivo Json en nnUNet_raw con la siguiente información:

Se define el nombre y el número de canales de las imágenes en channel_names, especificando un único canal en escala de grises bajo la etiqueta "gray". Las clases se detallan en labels, donde "background" tiene la etiqueta 0, "arcada_vascular" la etiqueta 1 y "nervio_optico" la etiqueta 2. Además, se indica un total de 350 casos para entrenamiento en numTraining, y se establece la extensión de los archivos de imagen como ".png" en file_ending.

6. Instalación de nnU-Net

Se clonó el repositorio oficial de nnU-Net [98] y se instaló en modo editable mediante el comando pip install -e. Esta modalidad permite realizar modificaciones directas en el código y facilita su prueba sin necesidad de reinstalar el paquete. También, se instaló la librería hiddenlayer desde su repositorio GitHub para habilitar el uso de capas ocultas durante el entrenamiento y visualización[99].

7. Exportación de las rutas de las carpetas como variables de entorno

Con el objetivo de facilitar la organización, reutilización y portabilidad del proyecto, se definieron las rutas principales del conjunto de datos, máscaras, resultados y modelos entrenados como variables de entorno.

Preprocesamiento

Para terminar esta parte, se inicia el preprocesamiento donde se ejecuta `nnUNetv2_plan_and_preprocess` para organizar y verificar la integridad del dataset usando la opción `--verify_dataset_integrity`, que comprueba que los metadatos entre los archivos coincidan. Se define el ID del dataset con `-d`, la modalidad de entrenamiento como 2d con `-c 2d` y el número de procesos simultáneos con `-np 1`.

Entrenamiento del modelo

Para iniciar el entrenamiento, se emplea el comando `nnUNetv2_train`, seguido del ID del dataset, la configuración 2d, el índice de división del conjunto de datos (0 en este caso, lo que indica que la primera quinta parte del conjunto será usada para validación y no para entrenamiento), y el nombre del entrenador con la cantidad de 250 épocas, como `nnUNetTrainer_250epochs`.

4.12.6 Resultados con nnU-Net

En las Figuras 4.41 a la 4.45 se muestran ejemplos representativos de los resultados de segmentación vascular generados por el modelo nnU-Net. Cada conjunto visual incluye la imagen original de fondo de ojo, la máscara de referencia y la predicción del modelo. Al igual que U-Net, nnU-Net logra una segmentación precisa de las estructuras vasculares principales. No obstante, se observa una mejora en la continuidad de los vasos y una mayor sensibilidad para detectar ramificaciones delgadas, en comparación con U-Net. Esta capacidad puede atribuirse al proceso de auto-configuración que caracteriza a nnU-Net, el cual ajusta automáticamente tanto los hiperparámetros como los esquemas de preprocesamiento y postprocesamiento según las características del conjunto de datos.

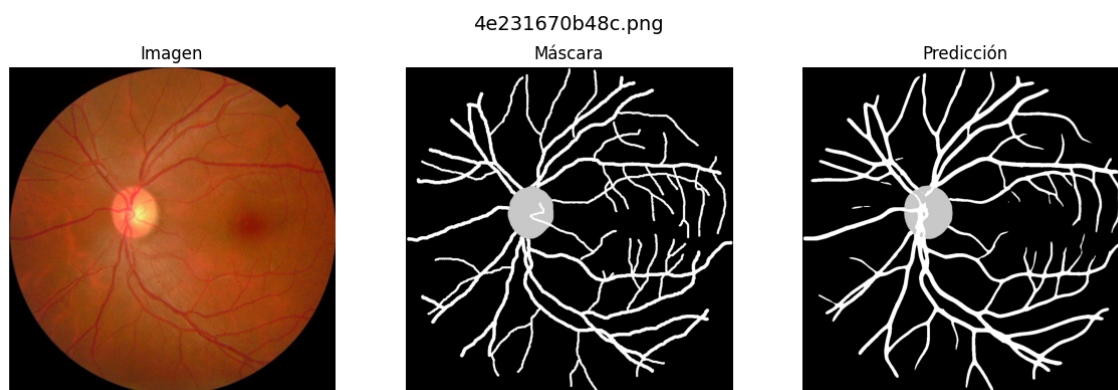


Figura 4.41 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net. De izquierda a derecha se presenta la imagen original, la máscara real y la predicción. El modelo muestra una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.

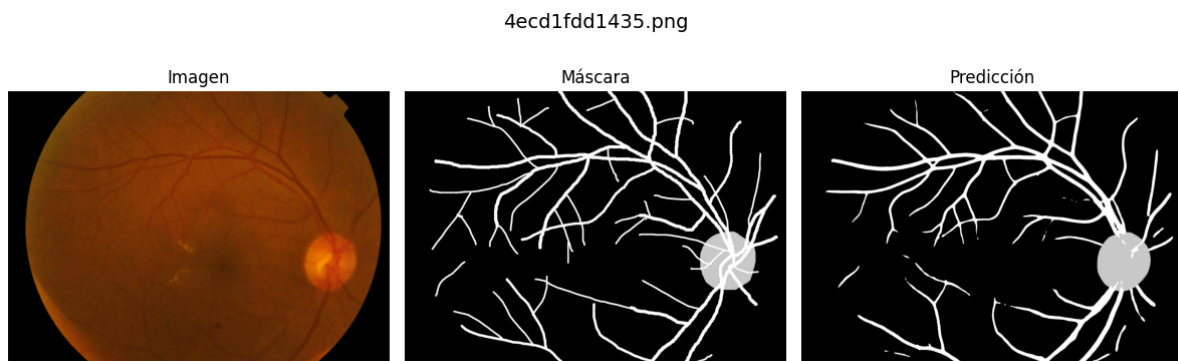


Figura 4.42 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net. De izquierda a derecha se presenta la imagen original, la máscara real y la predicción. El modelo muestra una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.



Figura 4.43 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net. De izquierda a derecha se presenta la imagen original, la máscara real y la predicción. El modelo muestra una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.

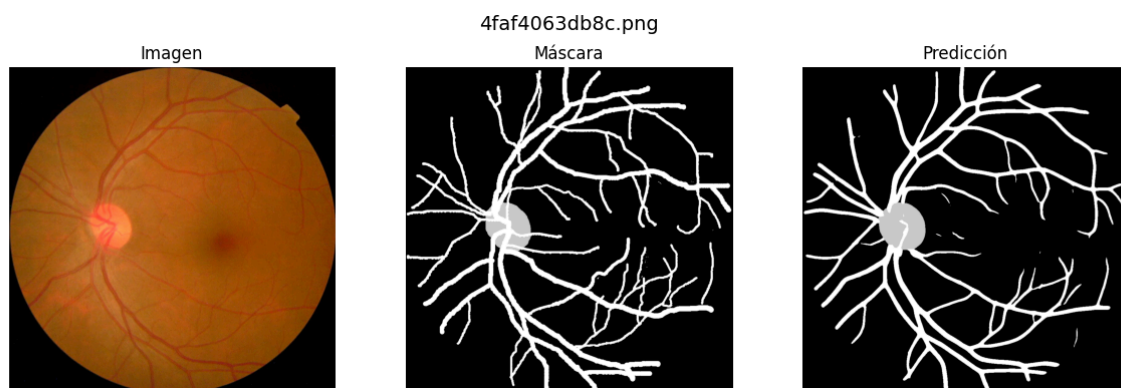


Figura 4.44 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net. De izquierda a derecha se presenta la imagen original, la máscara real y la predicción. El modelo muestra una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.

una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.

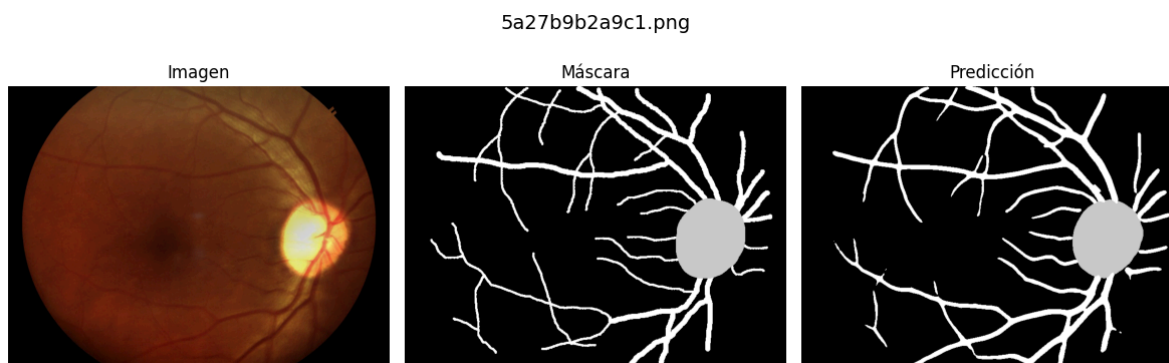


Figura 4.45 Segmentación de la arcada vascular y nervio óptico obtenida con el modelo nnU-Net. De izquierda a derecha se presenta la imagen original, la máscara real y la predicción. El modelo muestra una mayor cobertura y continuidad en la segmentación de vasos finos, con una fidelidad topológica superior respecto a U-Net, resultado de su capacidad de autoajuste al conjunto de datos.

4.12.7 Segment Anything Model (SAM)

Tras la implementación de nnU-Net, que permitió adaptar automáticamente la arquitectura y los parámetros del modelo a las particularidades de las imágenes de fondo de ojo, se exploró una nueva línea de trabajo utilizando el Segment Anything Model (SAM). En 2023, Meta IA presentó SAM un modelo de segmentación de imágenes. A diferencia de enfoques tradicionales como U-Net o nnU-Net, que se basan en arquitecturas de redes neuronales convolucionales con forma de U, el Segment Anything Model (SAM) utiliza una arquitectura basada en transformadores y mecanismos de atención. Esta diferencia arquitectónica permite a SAM realizar segmentación de manera generalizada, sin necesidad de entrenamiento adicional sobre conjuntos de datos específicos, y segmentar cualquier objeto a partir de indicaciones mínimas como puntos, cajas o máscaras.

La tarea se abordó con una idea de segmentación mediante indicaciones, que consiste en utilizar indicaciones para guiar el proceso de segmentación. Estas indicaciones pueden incluir puntos, recuadros o descripciones en lenguaje natural. Por otro lado, el modelo está diseñado para cumplir tres requisitos: soporte para indicaciones más flexibles, capacidad de generar máscaras en una respuesta rápida e interactiva y la capacidad para generar múltiples segmentaciones posibles ante ambigüedades. Para ello, la arquitectura de SAM está compuesta por un codificador de imágenes, basado en un Transformador de Visión (ViT, por sus siglas en inglés) pre entrenado con Masked Autoencoders (MAE), que genera una representación de los datos de entrada en un espacio latente, un codificador con indicaciones que interpreta la indicación dada y un decodificador de máscaras eficiente, que combina eficientemente ambas entradas para producir una máscara segmentada a través de un bloque de transformadores con atención cruzada y un clasificador dinámico por píxel. Por último, en cuestión de los datos, fue necesario contar con un conjunto de datos lo suficientemente amplio y diverso para lograr una fuerte capacidad de generalización. Dado que las máscaras de

segmentación no están ampliamente disponibles, se desarrolló un “motor de datos”, para la generalización de anotaciones en tres etapas:

1. **Asistencia manual:** SAM se utiliza como asistente para apoyar a los anotadores en la creación manual de las segmentaciones.
2. **Fase semiautomática:** SAM genera predicciones iniciales de máscaras, y los anotadores completan o corrigen los objetos restantes.
3. **Fase automática:** SAM realiza la anotación de forma prácticamente autónoma, requiriendo mínima intervención humana.

Este proceso dió origen a SA-1B, un conjunto de datos con más de 1.000 millones de máscaras sobre 11 millones de imágenes. Este conjunto de datos, creado completamente mediante un proceso automatizado de generación de datos es 400 veces más grande que cualquier conjunto de datos previo en el ámbito de la segmentación de imágenes hasta el momento[100].

Los modelos tradicionales de segmentación están diseñados para tareas específicas, como segmentación semántica (clasificar cada píxel en una categoría), por instancia (distinguir objetos individuales) y panóptica (combinación de ambas), como se mencionó en las secciones anteriores. Y solo tienen un rendimiento significativo en las tareas para las que fueron entrenados. En cambio, SAM en lugar de entrenarse para tareas fijas, aprende a responder a indicaciones, lo que le permite adaptarse a nuevas tareas sin entrenamiento adicional, pudiendo integrarse fácilmente en sistemas más complejos.

SAM puede tener limitaciones en segmentar detalles finos y no siempre es tan preciso como otros métodos especializados (entrenados para tareas específicas). Además, su desempeño no es totalmente robusto en tareas como segmentación semántica o panóptica, y su procesamiento no siempre es completamente en tiempo real. Sin embargo, SAM introduce avances importantes: un enfoque de segmentación mediante indicaciones, un modelo versátil y un extenso conjunto de datos[100]. La Figura 4.46 presenta una visión general del modelo SAM.

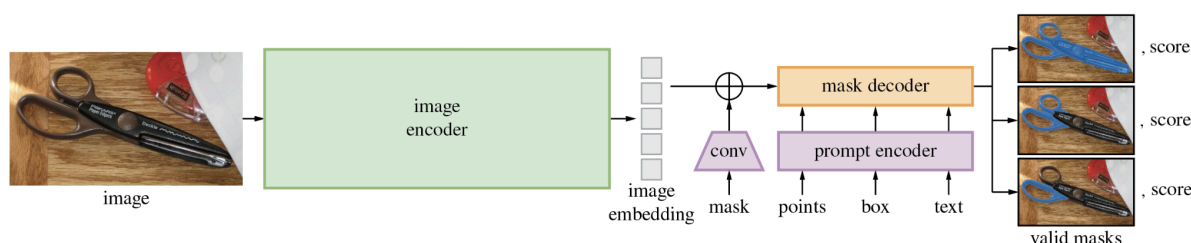


Figura 4.46 Representación del flujo de trabajo del modelo Segment Anything (SAM). El modelo recibe una imagen junto con una indicación y genera una segmentación correspondiente de la región de interés[101].

La inclusión de SAM se debe a su enfoque innovador basado en segmentación mediante indicaciones. Aunque no fue diseñado específicamente para el dominio de imágenes médicas, se decidió explorar su aplicación en este ámbito dada su versatilidad y el uso del paradigma de atención sobre el que está diseñado. SAM cuenta con tres versiones principales, diferenciadas por el tamaño de arquitectura base basada en ViT[100]:

- vit_h: Modelo SAM con ViT-Base de tamaño H (el más grande y potente).
- vit_l: Modelo SAM con ViT de tamaño L (un tamaño intermedio).
- vit_b: Modelo SAM con ViT de tamaño B (el más pequeño y rápido).

En este trabajo, se utilizó la versión vit_b, debido a su tamaño compacto y menores requerimientos computacionales, lo que facilita su implementación. Originalmente SAM funciona como un modelo generalista, diseñado para segmentar una amplia variedad de objetos sin distinción, por lo que es más fácil adaptarlo a datos específicos mediante sintonización fina de parámetros (fine tuning), de esta forma se aprovecha el conocimiento adquirido previamente por el modelo, ya que fue entrenado con más de 1.1 mil millones de máscaras en 11 millones de imágenes, lo que le otorga una base sólida para tareas de segmentación. Por esta razón, en este trabajo se optó por realizar sintonización con SAM_B, ya que permite especializar el modelo en tareas de segmentación de manera más práctica y eficiente.

4.12.8 Metodología del algoritmo SAM

Formato de las imágenes

La imagen se representa como una matriz tridimensional en formato RGB, donde cada canal codifica la intensidad de los colores rojo, verde y azul, ya que fue entrenado principalmente con imágenes RGB y su arquitectura basada en ViT's espera entradas de tres canales, por lo que mantener el formato original resulta compatible y adecuado, A continuación se puede observar la imagen de fondo de ojo con los canales RGB en la Figura 4.47

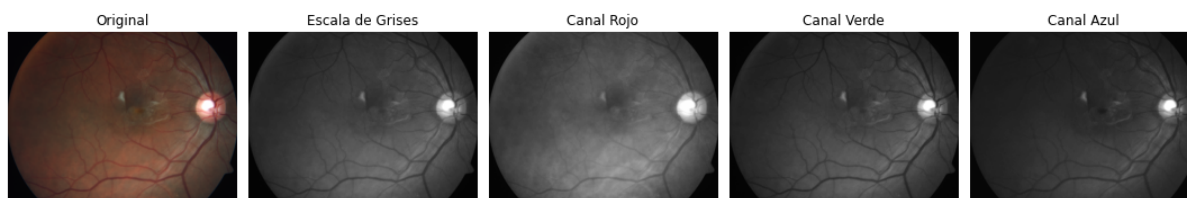


Figura 4.47 Imagen de fondo de ojo en una matriz tridimensional en formato RGB utilizada como entrada para el modelo SAM. Se conserva el formato en tres canales (rojo, verde y azul)

Selección de clase

Se integró una única clase de interés: la arcada vascular, principalmente por razones prácticas relacionadas con la simplicidad del entrenamiento y la reducción de complejidad en la segmentación. SAM fue diseñado para detectar objetos individuales usando indicaciones, por lo que para aplicar fine-tuning multiclase se requiere adaptar el flujo de trabajo. En particular:

1. Requiere máscaras codificadas por clase (con valores de píxel diferentes por categoría).
2. Adaptar el criterio de pérdida (o loss function).
3. Adaptar el pipeline del modelo, ya que SAM está diseñado originalmente para segmentación binaria.

En este caso, se utilizó el valor 255 (blanco) para etiquetar los píxeles correspondientes a la clase de interés, diferenciándolos del fondo, que se representa con el valor 0. Esto se debe a que SAM identifica como clase todos los píxeles cuyo valor sea mayor que 0. La Figura 4.48 muestra una representación de la imagen junto con su máscara correspondiente, en la que se visualiza exclusivamente la arcada vascular, excluyendo el nervio óptico.

0e43c8298fc0.png

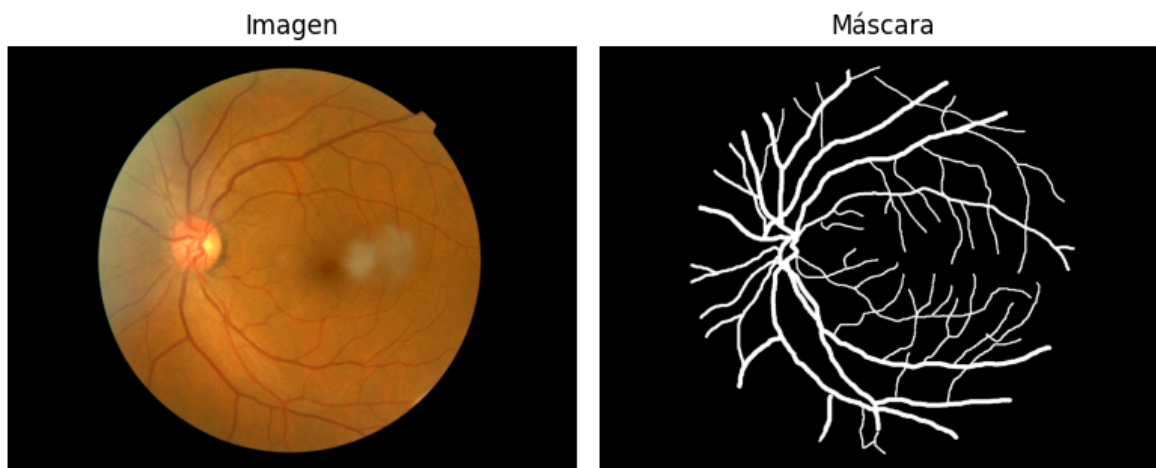


Figura 4.48 Imagen de fondo de ojo con máscara de la arcada vascular.

Instalación de requerimientos

Se instalan las librerías esenciales para trabajar con SAM y deep learning: la implementación oficial de Segment Anything[101], Transformers de Hugging Face[102] para modelos preentrenados, datasets para manejo eficiente de datos [102], y MONAI[103], especializada en imágenes médicas y funciones de pérdida. Además, se instala scikit-learn, una biblioteca para aprendizaje automático. Después se realizaron los siguientes pasos:

1. Cargar imágenes y máscaras en formato compatible

Para almacenar las imágenes y máscaras utilizamos la biblioteca Numpy, el cual permite manejar todas las imágenes y máscaras juntas como matrices homogéneas para facilitar su procesamiento, lo que permite manipularlas y procesarlas fácilmente como conjuntos de datos numéricos. En NumPy no se pueden almacenar imágenes con tamaños variables, este requiere que todos los elementos de un arreglo multidimensional tengan la misma forma para crear un array (arreglo numérico) homogéneo. Por lo tanto, en este proyecto, debido a la variabilidad en los tamaños de las imágenes y máscaras, se redimensionan a 256×256 píxeles. Esto se hace considerando que, aunque el modelo SAM redimensiona automáticamente las imágenes de entrada a 1024×1024 píxeles para su procesamiento interno, las predicciones que genera (etiquetas) son de tamaño 256×256 por defecto, y las máscaras y las predicciones deben tener las mismas dimensiones. Para facilitar el uso del conjunto de datos en el

modelo, las imágenes y máscaras fueron cargadas y convertidas a un formato compatible con la biblioteca datasets de Hugging Face. Esto permitió estructurar los datos como pares (imagen, máscara) correctamente alineados, manteniendo coherencia en tamaño y formato. Evitando posibles errores de entrada y asegurar compatibilidad y habilitar el uso de funciones avanzadas como transformaciones, partición de datos y carga directa en modelos pre entrenados (como es necesario en este caso).

2. Una integración fluida con herramientas como Transformers.

Al convertir imágenes y máscaras al formato compatible con datasets, se asegura una integración fluida en el pipeline, reduciendo errores de entrada o incompatibilidades de formato.

3. Generación de indicaciones

Las indicaciones utilizadas en este ejercicio fueron cajas delimitadoras (bounding boxes) generadas automáticamente a partir de las máscaras de segmentación. Para ello se extraen las coordenadas mínimas y máximas que encierran los píxeles etiquetados en la máscara binaria y se añade una pequeña perturbación aleatoria que expande ligeramente los bordes de la caja. Esto permite capturar mejor el contexto alrededor de la región de interés y proporciona variabilidad útil durante el fine-tuning, como se muestra en la Figura 4.49.

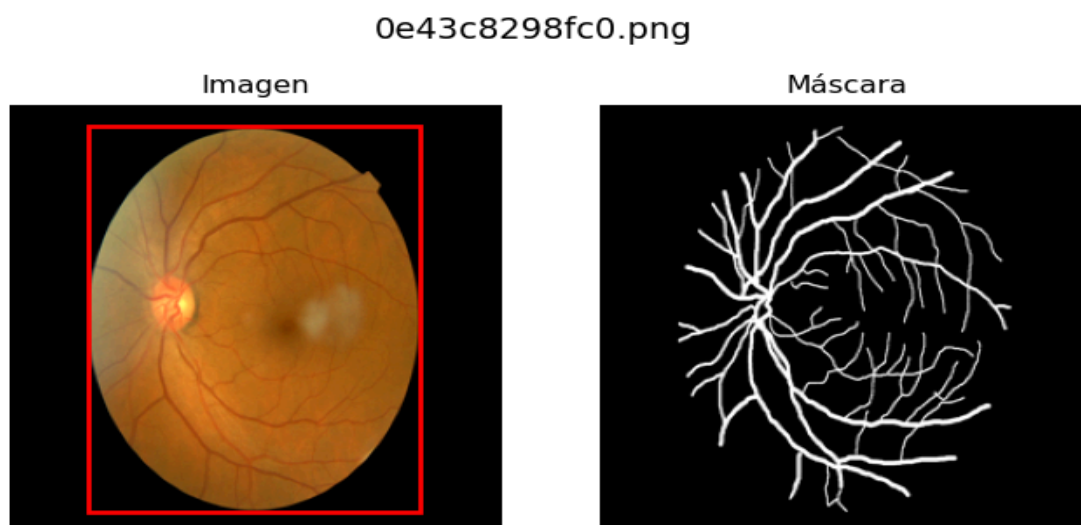


Figura 4.49 Visualización de imagen con caja delimitadora y máscara redimensionada.

Configuración del entrenamiento usando PyTorch y MONAI

PyTorch es un marco de trabajo utilizado en aprendizaje profundo. En esta tesis, se utiliza para construir, entrenar y optimizar el modelo con datos en forma de tensores, que son estructuras numéricas multidimensionales que el modelo puede procesar de forma eficiente. PyTorch puede integrarse con otros marcos como MONAI, que ofrece funcionalidades especializadas para el aprendizaje profundo en imágenes médicas.

La implementación del entrenamiento con el modelo SAM se estructuró como una serie de pasos secuenciales, diseñados para adaptar eficientemente la arquitectura preentrenada a la tarea específica de segmentación médica. A continuación, se describen los principales componentes del pipeline:

1. Inicialización del procesador

Se carga un procesador preentrenado para SAM. Este SamProcessor se encarga de preparar las imágenes y las indicaciones para la segmentación en el formato que el modelo SAM espera.

2. Preparación del dataset

Se crea un dataset personalizado (SAMDataset) que toma los datos originales y los transforma con el procesador para que sean compatibles con SAM.

3. DataLoader para entrenamiento

Se crea un DataLoader, que es una clase de PyTorch, para manejar el flujo de datos durante el entrenamiento, alimentando el modelo con lotes (`batch_size=2`) del dataset, barajándolos aleatoriamente (`shuffle=True`).

4. Se carga el modelo SAM preentrenado.

El modelo SAM preentrenado se carga desde el repositorio de modelos de Hugging Face[102], descargándose automáticamente desde esta plataforma que aloja modelos ya entrenados listos para usar.

5. Congelación de parámetros

Durante el entrenamiento, se congelan los parámetros del encoder de visión y del encoder de indicaciones, lo que significa que no se calculan sus gradientes y, por tanto, no se actualizan. Esta configuración se realiza con PyTorch y permite enfocar el entrenamiento únicamente en el decodificador de máscaras (`mask_decoder`), lo que reduce el recurso computacional y ayuda a prevenir el sobreajuste.

6. Optimizador y función de pérdida

Se utiliza el optimizador Adam de PyTorch para ajustar exclusivamente los parámetros del módulo `mask_decoder`, con una tasa de aprendizaje baja de $1e-5$ y sin regularización por peso. La función de pérdida empleada es DiceCELoss, proporcionada por MONAI, la cual combina Dice Loss y Cross-Entropy Loss, siendo especialmente adecuada para tareas de segmentación médica. Esta función se configura con `sigmoid=True` para producir una salida binaria por píxel, `squared_pred=True` para suavizar el cálculo de la Dice Loss, y `reduction='mean'` para promediar la pérdida sobre todo el lote.

Entrenamiento

Se entrenó el modelo SAM con el conjunto de 500 imágenes durante 250 épocas. En cada época, itera sobre los lotes del conjunto de entrenamiento usando un DataLoader. Para cada lote, realiza una propagación hacia adelante (forward pass) donde el modelo genera las máscaras segmentadas a partir de las imágenes y las cajas de entrada. Luego, calcula la función de pérdida comparando las máscaras predichas con las máscaras reales (ground truth). A continuación, realiza la retropropagación (backward pass) para calcular los gradientes de los parámetros respecto a la pérdida. Posteriormente, actualiza los parámetros del modelo usando el optimizador Adam. Se guarda la pérdida de cada lote para calcular la pérdida promedio de la época.

4.12.9 Resultados con algoritmo SAM

Las Figuras 4.50 –4.59 ilustran los resultados de segmentación obtenidos mediante el modelo SAM. En cada Figura se presentan, en orden, la imagen original de fondo de ojo, la máscara de referencia manual (si está disponible), la máscara predicha por el modelo y el mapa de probabilidad generado. Este último se incluye con el fin de mostrar el nivel de confianza del modelo en cada región segmentada, permitiendo una interpretación más detallada del resultado. A diferencia de los modelos entrenados específicamente para esta tarea, SAM aplica segmentación a partir de indicaciones mínimas, por lo que el mapa de probabilidad resulta útil para visualizar la ambigüedad o certeza con la que el modelo identificó las estructuras vasculares. Las primeras cinco imágenes corresponden a la visualización del proceso de inferencia utilizando imágenes del conjunto de entrenamiento, mientras que las siguientes cinco muestran la inferencia sobre imágenes sin anotación manual, lo que permite evaluar la capacidad generalizada del modelo para aplicar segmentación en escenarios sin supervisión directa..

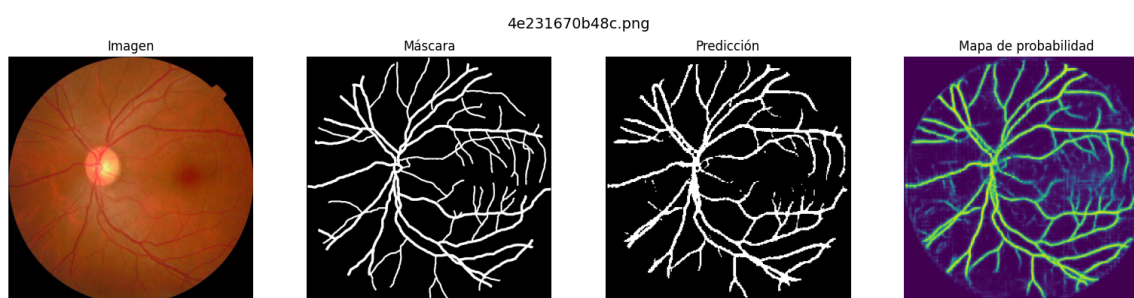


Figura 4.50 Resultados de segmentación obtenidos con el modelo SAM. Se muestra la imagen original, la máscara real y el mapa de probabilidad producido por el modelo.

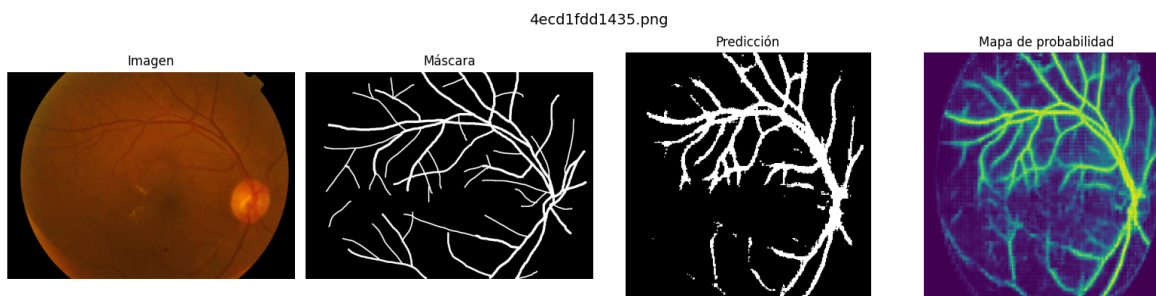


Figura 5.51 Resultados de segmentación obtenidos con el modelo SAM. Se muestra la imagen original, la máscara real y el mapa de probabilidad producido por el modelo.

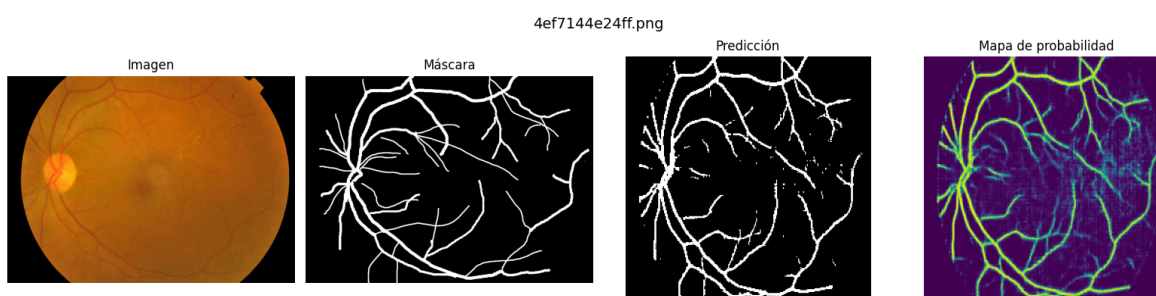


Figura 4.52 Resultados de segmentación obtenidos con el modelo SAM. Se muestra la imagen original, la máscara real y el mapa de probabilidad producido por el modelo.

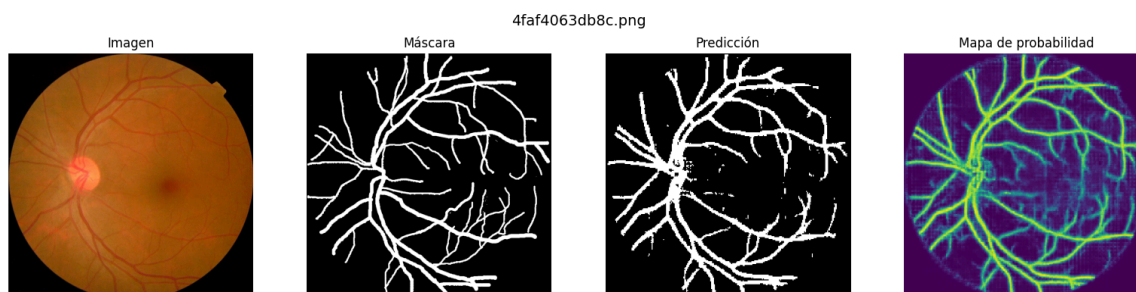


Figura 4.53 Resultados de segmentación obtenidos con el modelo SAM. Se muestra la imagen original, la máscara real y el mapa de probabilidad producido por el modelo.

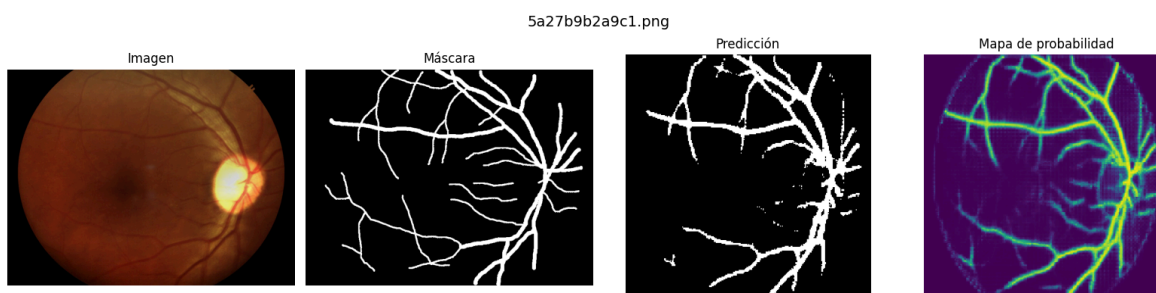


Figura 4.54 Resultados de segmentación obtenidos con el modelo SAM. Se muestra la imagen original, la máscara real y el mapa de probabilidad producido por el modelo.

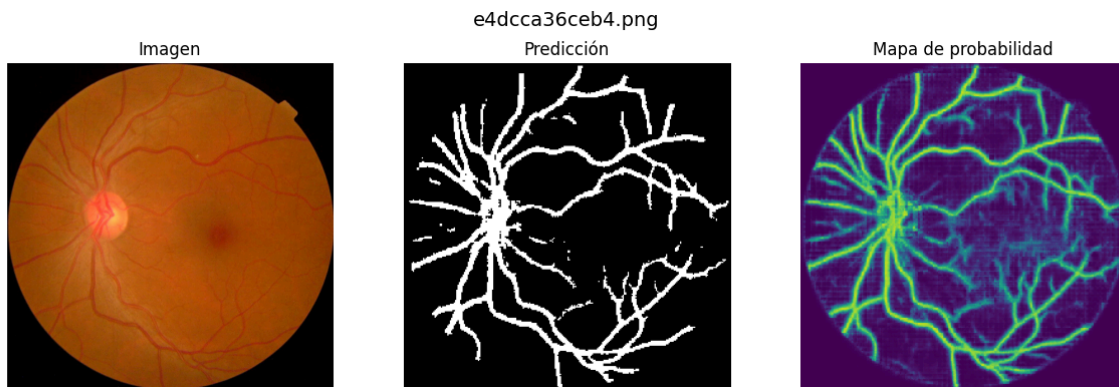


Figura 4.55 Resultados de segmentación obtenidos con el modelo SAM. De izquierda a derecha se muestra la imagen original de fondo de ojo, la máscara de segmentación generada automáticamente y el mapa de probabilidad asociado.

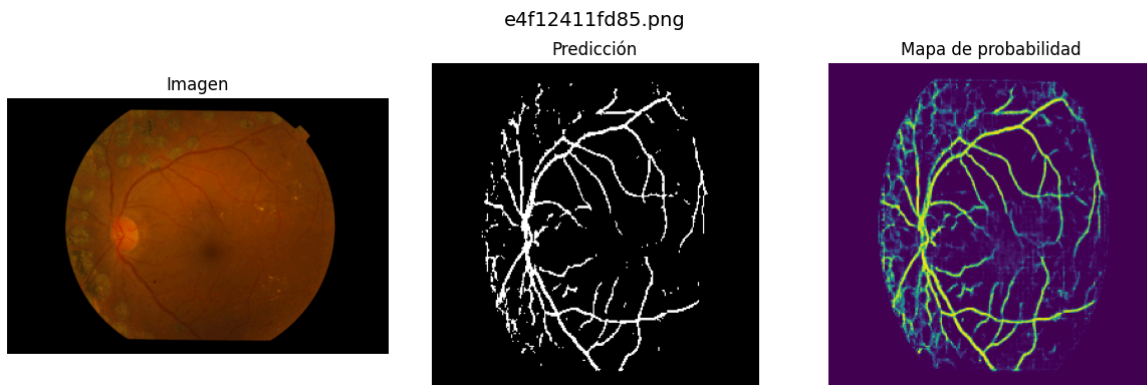


Figura 4.56 Resultados de segmentación obtenidos con el modelo SAM. De izquierda a derecha se muestra la imagen original de fondo de ojo, la máscara de segmentación generada automáticamente y el mapa de probabilidad asociado.

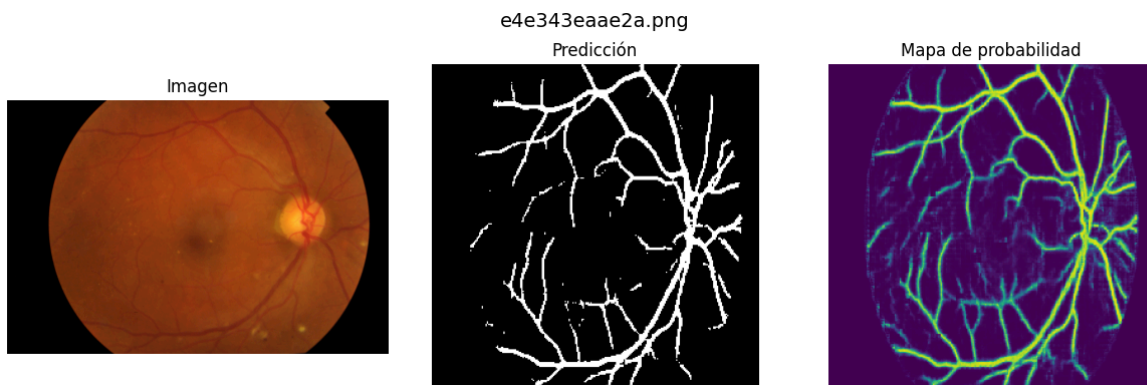


Figura 4.57 Resultados de segmentación obtenidos con el modelo SAM. De izquierda a derecha se muestra la imagen original de fondo de ojo, la máscara de segmentación generada automáticamente y el mapa de probabilidad asociado.

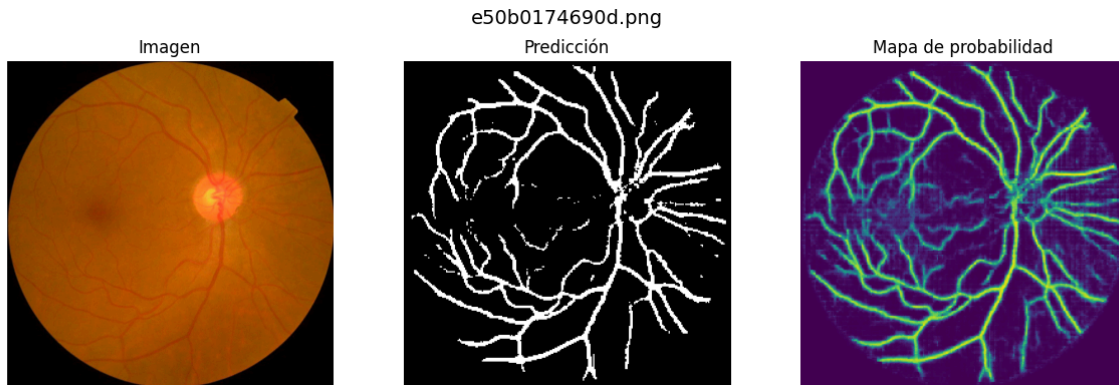


Figura 4.58 Resultados de segmentación obtenidos con el modelo SAM. De izquierda a derecha se muestra la imagen original de fondo de ojo, la máscara de segmentación generada automáticamente y el mapa de probabilidad asociado.

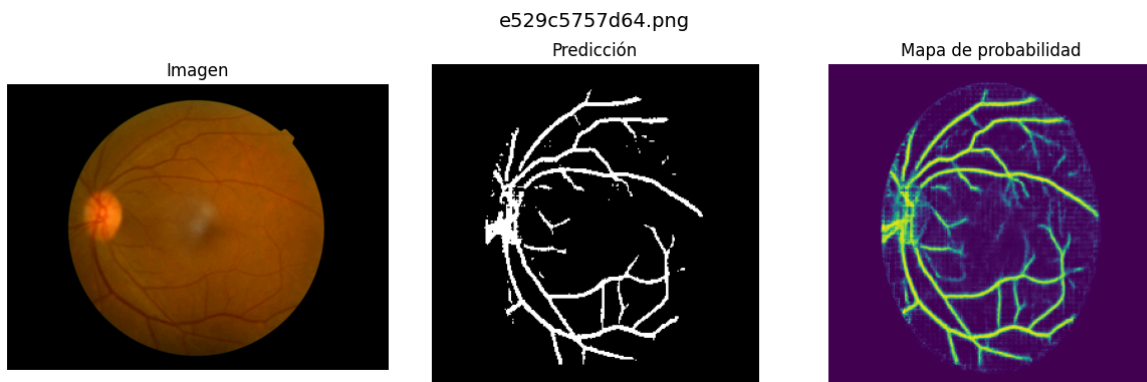


Figura 4.59 Resultados de segmentación obtenidos con el modelo SAM. De izquierda a derecha se muestra la imagen original de fondo de ojo, la máscara de segmentación generada automáticamente y el mapa de probabilidad asociado.

4.13 Evaluación de desempeño

4.13.1 Resultados de U-Net

Inicialmente la U-Net mostró un desempeño muy competitivo en la tarea de segmentación, ya que las métricas se muestran equilibradas de una manera aceptable. En especial el valor de F1 Score fue de 69.76% lo que indica una buena armonía entre la capacidad del modelo para identificar correctamente las regiones de interés (sensibilidad) y su capacidad para evitar falsos positivos (precisión). Esto concuerda con el índice de Dice que fue de 69.92%, esta métrica es comúnmente utilizada en segmentación médica debido a su enfoque en la superposición entre las regiones predichas y las

reales. Hablando de sensibilidad alcanzó un 68.32% lo que implica que el modelo logró detectar aproximadamente dos tercios de las regiones relevantes, aunque dejó escapar una parte importante de los verdaderos positivos. Por otro lado, la precisión fue del 73.45 %, lo que sugiere que la mayoría de las regiones detectadas como positivas por el modelo fueron efectivamente verdaderas. Algo que queda destacar es la relación entre sensibilidad y precisión ya que indica que el modelo tiende a ser ligeramente conservador, es decir, prefiere evitar falsos positivos incluso si eso significa no detectar todas las regiones positivas.

Hablando generalmente del modelo, este obtuvo una exactitud del 94.94% es elevado, pero dado que esta métrica considera también las regiones negativas en ese caso el fondo de la imagen, esto puede influir en el desbalance de clases, lo cual es muy normal en la segmentación del area medica debido al tamaño de las regiones de interés. Finalmente, el índice de Jaccard (IoU) fue de 53.74 %. Aunque este valor es más bajo que el índice de Dice, es consistente con su comportamiento, ya que el IoU es una métrica más estricta que penaliza con mayor severidad los errores en la superposición entre las predicciones y la verdad de terreno. La segmentación es aceptable, no obstante, existe margen de mejora en términos de sensibilidad. A continuación se encuentra la Tabla 4.21 que presenta un resumen de las métricas de resultados para U-Net.

Métrica	Resultado
F1 Score	69.76%
Dice	69.92%
Sensibilidad	68.32%
Precisión	73.45%
Exactitud	94.94%
IoU	53.74%

Tabla 4.21 Tabla de métricas U-Net

4.13.2 Resultados con UNet (Imagen original y predicción)

Además, en las Figuras 4.60 a 4.62 se presentan ejemplos visuales que permiten apreciar la calidad de la segmentación realizada por el modelo U-Net. En cada una de estas imágenes se incluye la imagen original junto con la correspondiente predicción generada por el modelo, facilitando la comparación visual entre la realidad y la segmentación automática obtenida. Estas representaciones gráficas complementan las métricas cuantitativas y evidencian tanto los aciertos como las áreas susceptibles de mejora.

1b8ad0afe9fb.png

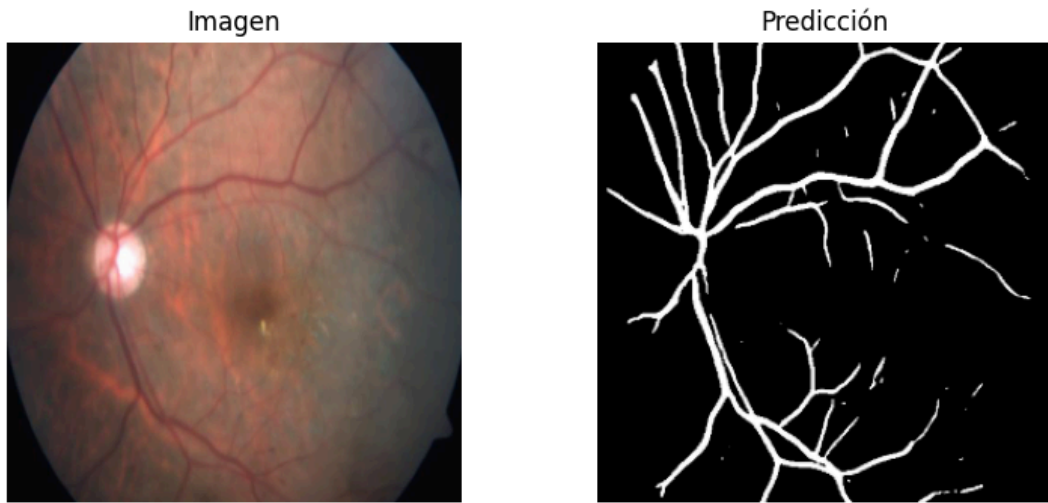


Figura 4.60 Resultado de segmentación obtenido con el modelo UNet. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

1b329a127307.png

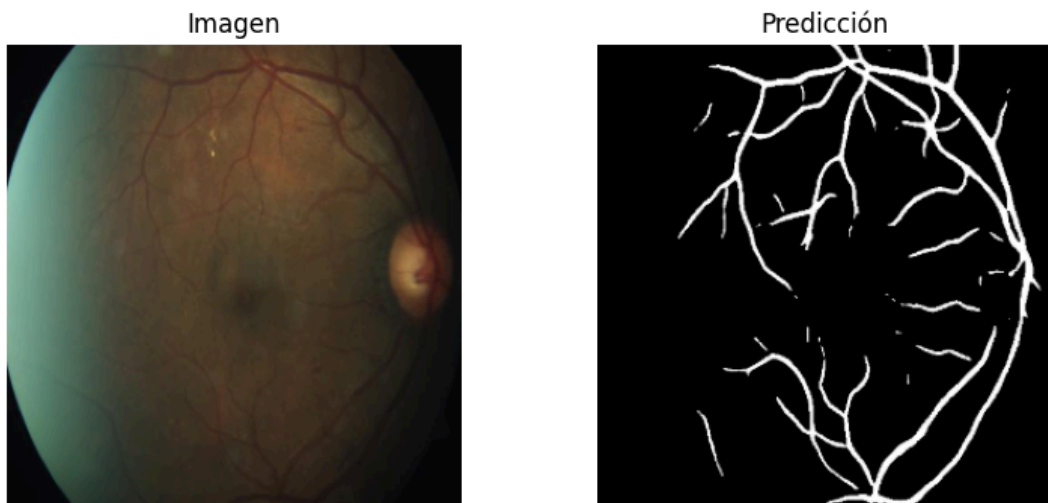


Figura 4.61 Resultado de segmentación obtenido con el modelo UNet. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

1b862fb6f65d.png

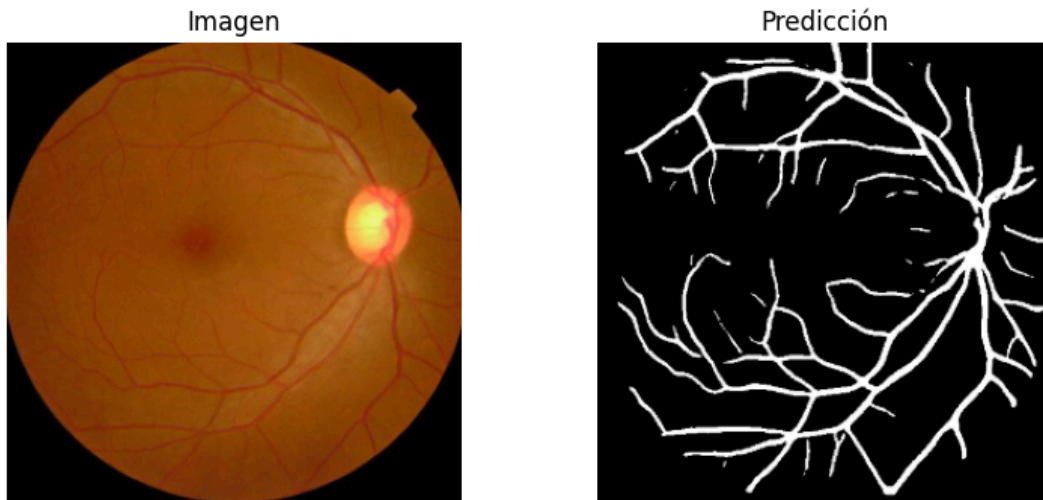


Figura 4.62 Resultado de segmentación obtenido con el modelo UNet. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

4.13.3 Resultados de nnU-Net

General

El modelo nnU-Net alcanzó un desempeño moderado en la tarea de segmentación, con métricas que reflejan un comportamiento equilibrado pero aún con espacio para mejoras, especialmente en cuanto a la capacidad del modelo para detectar correctamente las regiones positivas. Iniciando con el valor del índice de Dice fue de 69.98 %, esto quiere decir que el nivel que está manejando es aceptable entre las segmentaciones predichas por el modelo y las referencias reales, es un valor que concuerda con las imágenes observadas anteriormente. Por otro lado, el índice de Jaccard (IoU), una métrica más estricta que el Dice, alcanzó 55.70 %, lo cual confirma correspondencia razonable entre la segmentación automatizada y la referencia manual. Siguiendo con sensibilidad de 58.00 %, lo que indica que identificó poco más de la mitad de las regiones positivas reales, continuando con la precisión que fue de 58.54 %, lo que revela que, entre todas las regiones marcadas como positivas por el modelo, una proporción ligeramente superior a la mitad eran correctas, en conjunto con el F1 Score de 58.26 %, reflejan que el modelo tiene un rendimiento moderado, con una tendencia a equivocarse tanto al omitir regiones relevantes como al predecir regiones incorrectas. Finalmente la exactitud general fue de 95.02 %, lo cual podría parecer elevada; sin embargo, en problemas de segmentación, esta métrica puede estar sesgada por la presencia mayoritaria de la clase negativa es decir, fondo, de la misma manera que la U-Net. A continuación se encuentra la Tabla 4.22 que presenta un resumen de las métricas de resultados para nnU-Net.

Métrica	Resultado
Dice	69.98%
IoU	55.70%
Sensibilidad	58.00%
Precisión	58.54%
Exactitud	95.02%
F1 Score	58.26

Tabla 4.22 Tabla de métricas nnU-Net

4.13.4 Resultados con nnU-Net (Imagen original y predicción)

De manera similar, en las Figuras 4.63 a 4.65 se muestran imágenes originales junto con las predicciones generadas por el modelo nnU-Net. Estas Figuras ilustran la capacidad del modelo para segmentar las estructuras de interés, evidenciando mejoras en la precisión y detalle respecto a U-Net en ciertos casos. La comparación visual entre las predicciones y la verdad de terreno permite identificar áreas donde el modelo performa de forma satisfactoria, así como aquellas que requieren ajustes para optimizar la sensibilidad.

e50b0174690d.png

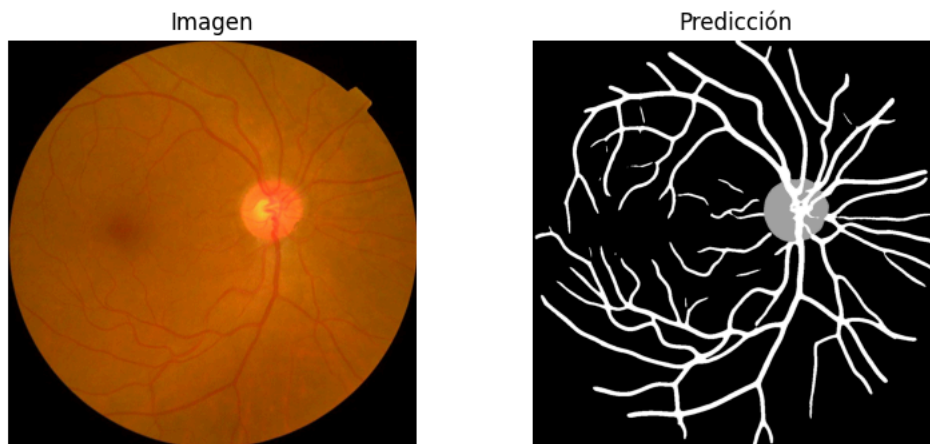


Figura 4.63 Resultado de segmentación obtenido con el modelo nnU-Net. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

e4dcca36ceb4.png

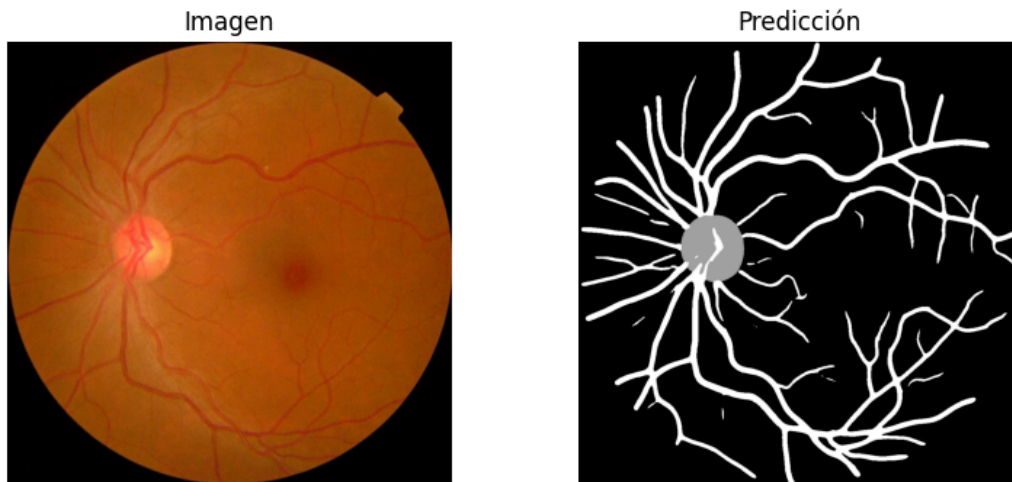


Figura 4.64 Resultado de segmentación obtenido con el modelo nnU-Net. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

e4f12411fd85.png

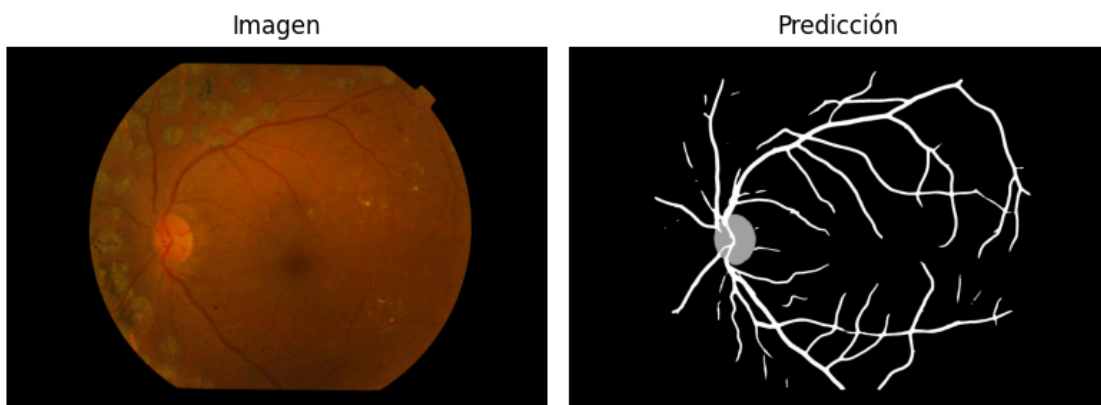


Figura 4.65 Resultado de segmentación obtenido con el modelo nnU-Net. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

Clase 1 (Arcada vascular)

En el caso específico de la clase uno, correspondiente a la arcada vascular, el modelo nnU-Net mostró un desempeño moderado. El índice de Dice fue de 58.17%, lo que indica un nivel de solapamiento aceptable entre la segmentación automatizada y la referencia manual, aunque menor en comparación con los valores obtenidos en la segmentación general. El índice de Jaccard (IoU) alcanzó un 41.45%, evidenciando que el modelo tuvo una menor precisión en la superposición estricta, esto quiere decir que tiene una dificultad inherente de delinear estructuras vasculares delgadas y complejas. Debido a que la sensibilidad fue de 53.33% cerca de la mitad de las estructuras vasculares verdaderas fueron

identificadas por el modelo, mientras que la precisión alcanzó 56.06 %, indicando que una proporción similar de las predicciones positivas fueron correctas. Hablando del equilibrio entre la sensibilidad y la precisión el F1 Score fue de 54.64% lo cual refleja una capacidad limitada para segmentar la arcada vascular de manera exacta. Finalmente, hablando de la exactitud general fue de 90.74 %, una cifra elevada pero que debe interpretarse de manera prudente debido a la posible desproporción entre las regiones positivas en este caso los vasos sanguíneos y negativas como se ha estado diciendo en las secciones anteriores el fondo. A continuación se encuentra la Tabla 4.23 que presenta un resumen de las métricas de resultados para nnU-Net enfocado en la clase 1, también llamada arcada vascular.

Métrica	Resultado
Dice	58.17%
IoU	41.45%
Sensibilidad	53.33%
Precisión	56.06%
Exactitud	90.74%
F1 Score	54.64%

Tabla 4.23 Tabla de métricas nnU-Net específicamente en Arcada vascular

Clase 2 (Nervio óptico)

En la segmentación de la clase dos, correspondiente al nervio óptico, el modelo nnU-Net presentó un desempeño destacado. El índice de Dice alcanzó un 81.79 %, lo que indica un alto grado de solapamiento entre las segmentaciones predichas y las referencias manuales, esto quiere decir que el modelo es muy congruente a la morfología del nervio óptico, sobretodo en comparación los resultados de la arcada vascular. El índice de Jaccard (IoU) fue de 69.95 %, lo cual confirma una coincidencia acertada entre las regiones segmentadas automáticamente y las referencias, al mismo tiempo se obtuvo una sensibilidad de 90.34 %, demostrando una elevada capacidad para identificar correctamente la mayoría de los píxeles positivos que le corresponden al nervio óptico. Sin embargo, la precisión fue de 71.43% lo que arroja algunas predicciones incorrectas, es decir falsos positivos, pero mayormente las segmentaciones fueron acertadas. Con un F1 Score de 79.68% se reafirma como ha sido este modelo robusto para esta clase. Asimismo, la exactitud global (Accuracy) fue de 99.31 %, reflejando un desempeño sobresaliente a nivel general, aun así puede tratarse de una desproporción de clases. A continuación se encuentra la Tabla 4.24 que presenta un resumen de las métricas de resultados para nnU-Net enfocada en la clase dos, también llamada nervio óptico.

Métrica	Resultado
Dice	81.79%
IoU	69.95%
Sensibilidad	90.34%
Precisión	71.43%
Exactitud	99.31%
F1 Score	79.68%

Tabla 4.24 Tabla de métricas nnU-Net específicamente en nervio óptico

4.13.5 Resultados de SAM

El modelo SAM obtuvo un desempeño moderado en la segmentación de las estructuras de interés en imágenes de fondo de ojo. De acuerdo al índice de Dice que fue de 61.96 %, lo cual indica un nivel aceptable de superposición entre la segmentación automatizada y la referencia manual, aunque inferior al rendimiento observado en nnU-Net. Hablando el índice de Jaccard (IoU) fue de 45.35 % lo cual refleja un grado de coincidencia más conservador pero evidencia mejor margen de mejora en la precisión de las delimitaciones. De acuerdo a la capacidad de detección SAM alcanzó una sensibilidad de 58.82 %, lo que sugiere que poco más de la mitad de las regiones de interés fueron correctamente identificadas. Sin embargo, la precisión fue de 66.68 %, lo que indica que una proporción importante de las predicciones positivas fueron correctas, aun así se puede observar que una gran cantidad fue de falsos positivos. El F1 Score fue de 62.48% teniendo un equilibrio razonable entre precisión y sensibilidad. Finalmente la exactitud global (accuracy) alcanzó un 93.48 %, lo cual es elevado, pero como en otros casos, este valor puede estar influenciado por el desequilibrio entre clases en las imágenes de igual manera que U-Net y nnU-Net. Estos resultados sugieren que SAM, aunque es un modelo generalista con buena capacidad de segmentación sin necesidad de entrenamiento específico, sin embargo, puede ser que se vea muy influenciado por la variabilidad de la calidad de las imágenes. A continuación se encuentra la Tabla 4.25 que presenta un resumen de las métricas de resultados para SAM.

Métrica	Resultado
Dice	61.96%
IoU	45.35%
Sensibilidad	58.82%
Precisión	66.68%
Exactitud	93.48%
F1 Score	62.48%

Tabla 4.25 Tabla de métricas de SAM

4.13.6 Resultados con SAM (Imagen original y predicción)

Finalmente, las Figuras 4.66 a 4.68 presentan la segmentación realizada por el modelo SAM. En estas imágenes se comparan las imágenes originales con las segmentaciones obtenidas, permitiendo evaluar visualmente la efectividad del modelo en la identificación de las regiones de interés, en este caso arcada vascular. Aunque SAM presenta fortalezas particulares sobre todo visualmente, las imágenes evidencian algunas inconsistencias en la superposición que podrían ser abordadas para mejorar la exactitud clínica del sistema.

e50b0174690d.png

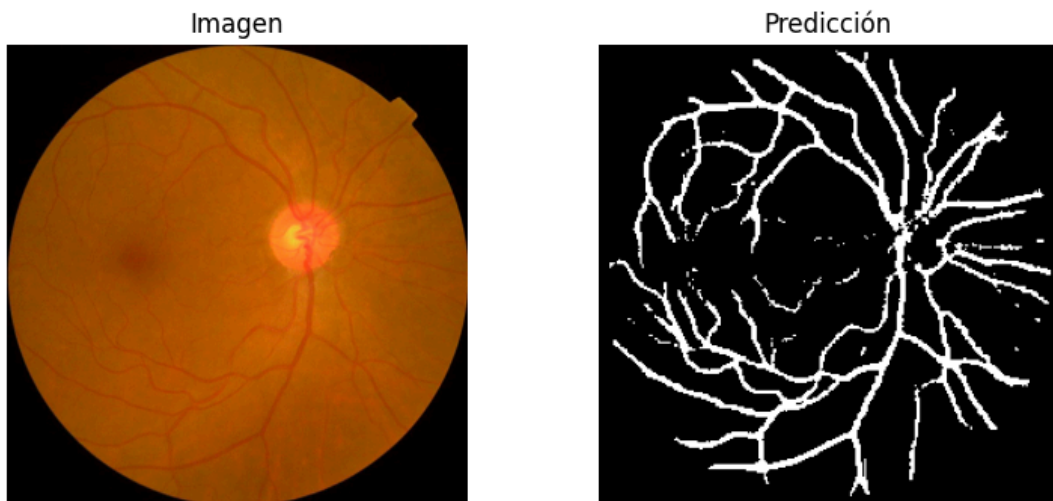


Figura 4.66 Resultado de segmentación obtenido con el modelo SAM. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

e4f12411fd85.png

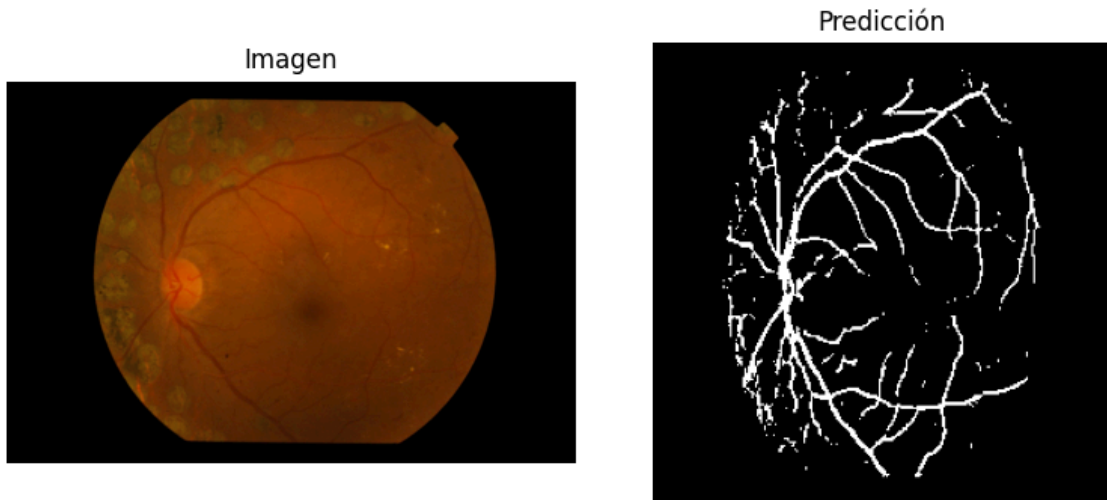


Figura 4.67 Resultado de segmentación obtenido con el modelo SAM. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

e529c5757d64.png

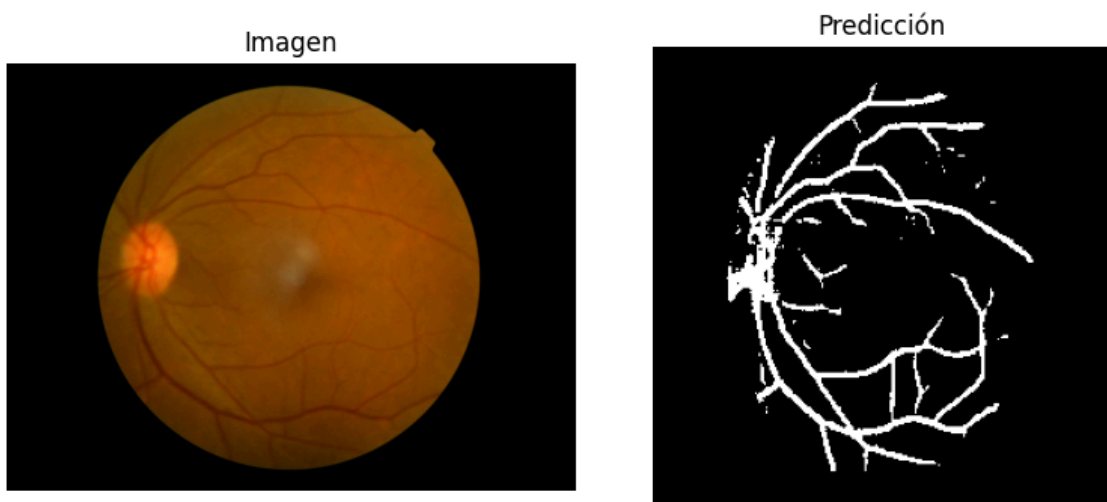


Figura 4.68 Resultado de segmentación obtenido con el modelo SAM. A la izquierda se muestra la imagen original de fondo de ojo, y a la derecha la máscara de segmentación predicha por el modelo.

4.14 Propuesta de sistema de segmentación aplicable al contexto regional

En esta sección se presenta una propuesta de sistema de segmentación de estructuras anatómicas en imágenes de fundoscopia digital, considerando la diversidad de condiciones en el ámbito tecnológico,

clínico y operativo presentes en el contexto regional. Aquí se propone una aproximación algorítmica apropiada para la segmentación de estructuras clave en este caso de las segmentadas por los algoritmos en secciones anteriores, Todo esto teniendo en cuenta los recursos disponibles, la precisión requerida y los tiempos de procesamiento tolerables. A continuación, se resumen las características a considerar en la siguiente Tabla 4.26:

Métrica	UNet (%)	nnU-Net (%)	SAM (%)
Dice	69.92	69.98	61.96
IoU	53.74	55.70	45.35
Sensibilidad	68.32	58.00	58.82
Precisión	73.45	58.54	66.68
F1 Score	69.76	58.26	62.48
Exactitud	94.94	95.02	93.48
Recurso computacional utilizado	Bajo (PyTorch en Intel UHD Graphics 630)	Alto (Alienware Aurora R15 AMD Ryzen 9 7900X)	Alto (Alienware Aurora R15 AMD Ryzen 9 7900X)
Tiempo de entrenamiento	10 horas	6 horas	15 horas
Tiempo de inferencia	1:20 min (20 imágenes)	< 1 min (50 imágenes)	15 min (500 imágenes)

Tabla 4.26 Resumen las características de los tres modelos evaluados: **UNet**, **nnU-Net** y **SAM**

A partir del análisis comparativo que se realizó en la Tabla de características a considerar donde se encuentran métricas de desempeño, tiempos de entrenamiento e inferencia, así como del recurso computacional requerido por cada arquitectura, se propone que al momento de elegir el algoritmo de segmentación más adecuado se deben de tomar en cuenta ciertos factores como lo son las condiciones técnicas, clínicas y operativas de la región. Esta propuesta se enfoca más que nada en priorizar las necesidades del contexto en el cual se necesiten este tipo de modelos ya que el que el rendimiento algorítmico no es el único criterio, sino que se pondera también la viabilidad de implementación en entornos reales del sistema de salud. Hablando en términos computacionales, el recurso varía significativamente entre los modelos, iniciando con U-Net este modelo fue entrenado con especificaciones modestas: Pytorch 2.5.1, compilado en una estación con tarjeta gráfica integrada Intel UHD Graphics 630, arquitectura CometLake-S. Esta configuración es un entorno común en la mayoría de las instituciones por lo tanto es algo muy alcanzable en la mayoría de los sistemas de salud. En este caso, el entrenamiento tomó aproximadamente 10 horas, y la inferencia de 20 imágenes se logró en 1 minuto y 20 segundos. En este escenario, U-Net puede ser viable para contextos regionales donde no se puede obtener acceso a una GPU. Por otro lado, nnU-Net y SAM fueron entrenadas en equipos de alto rendimiento, en este caso Alienware Aurora R15 AMD, con procesador AMD Ryzen 9 7900X de 12 núcleos, arquitectura de 64 bits y múltiples puentes PCIe de alto ancho de banda. En este contexto, nnU-Net logró completar su entrenamiento en 6 horas, con inferencias inferiores a 1 minuto por 50 imágenes, demostrando su eficiencia operativa cuando se dispone de

recursos computacionales adecuados. Finalmente, SAM por su parte, aunque se entrenó con el mismo equipo que nnU-Net presentó un tiempo de entrenamiento de 15 horas, más prolongado y un tiempo inferencia mayor 15 minutos por 500 imágenes lo que indica una carga computacional intensiva, posiblemente atribuida a su arquitectura generalista de segmentación y su necesidad de ajustes adicionales para imágenes médicas especializadas.

Desde una justificación clínica, se observa que debe de considerarse los objetivos de cada sistema de salud para poder realizar la selección del algoritmo más adecuado, de acuerdo a las necesidades y alcances computacionales, por ejemplo si es que busca hacer monitoreos con frecuencia de las estructuras anatómicas, en el caso de esta investigación arcada vascular y nervio óptico, en zonas rurales o unidades móviles con equipos limitados, U-Net ofrece un balance aceptable entre precisión y factibilidad operativa. Por otro lado, si la prioridad de la institución clínica es alta precisión diagnóstica, como en centros especializados donde se toman decisiones terapéuticas críticas a partir de imágenes, nnU-Net sería la elección más robusta, gracias a su rendimiento superior en métricas de segmentación que da más peso a las regiones donde hay coincidencia real entre el área segmentada y el área real y al momento de dibujar los bordes correctos de la estructura anatómica. Finalmente SAM, aunque tiene un rendimiento general menor en métricas globales, podría tener aplicaciones específicas donde se busque una segmentación inicial de grandes volúmenes de datos o cuando se requiera una arquitectura flexible capaz de adaptarse a diversas estructuras, siempre que se disponga de infraestructura avanzada otra observación de este modelo es se ue genera segmentaciones visualmente más detalladas, con contornos anatómicos finos y ajustados, especialmente en estructuras vasculares pequeñas o con ramificaciones complejas. Esta característica puede ser clínicamente útil en escenarios donde se requiera mayor resolución estructural o donde la revisión visual sea parte del proceso diagnóstico, sin embargo, requiere más atención en la sintonización de parámetros para desarrollar el desempeño de las métricas. En conclusión, no existe un único algoritmo óptimo para todas las condiciones posibles. La elección debe estar guiada por una evaluación integral de los recursos computacionales disponibles, los requerimientos clínicos del entorno y la naturaleza de las imágenes a procesar. Esta propuesta considera que la implementación efectiva de sistemas de segmentación en el contexto regional debe ser adaptable, permitiendo seleccionar el algoritmo según lo que se tiene en la infraestructura y lo que se necesita como cuánta precisión clínica se requiere.

5. Discusión y conclusiones

5.1 Discusión

El utilizar aprendizaje automático ha cobrado gran relevancia en el ámbito de la oftalmología ya que es una herramienta que hace hincapié a la oportunidad de tener diagnósticos más precisos y eficaces debido a la segmentación y clasificación de imágenes de fundoscopia digital. En este contexto, diversos estudios recientes han propuesto algoritmos que abordan los desafíos específicos de esta tarea mediante técnicas innovadoras de aprendizaje profundo y procesamiento de imágenes, la revisión de estos algoritmos permite identificar enfoques prometedores, Para el caso de esta tesis se abordó dicha tarea considerando enfoques clásicos, avanzados y generalistas como U-Net, nnU-Net y SAM, evaluando su rendimiento sobre estructuras anatómicas clave como la arcada vascular y el nervio óptico.

En términos generales, los métodos propuestos en la literatura pueden agruparse en distintas ramas del aprendizaje automático, cada una con características particulares que responden a diferentes necesidades clínicas y técnica, por lo tanto no queda más que destacar sus fortalezas y debilidades para la implementación de esta tesis antes de discutir los modelos seleccionados. Entre las técnicas más utilizadas se encuentran las Redes Neuronales Convolucionales (CNNs) las cuales han sido fundamentales en la evolución del procesamiento de imágenes médicas. Estos modelos destacan por su habilidad para aprender representaciones espaciales complejas, lo que las hace eficaces para identificar las estructuras anatómicas como es este caso, Sin embargo, su desempeño óptimo suele estar asociado a una alta demanda computacional, lo cual puede limitar su implementación en entornos clínicos de bajos recursos. Por otro lado, surgió FCN como una evolución de CNN ya que aquí se eliminan las capas más densas permitiendo que haya una segmentación a nivel píxel. Sin embargo, al momento de realizar segmentaciones detalladas presentan limitaciones. Por su parte, los modelos basados en UNet y sus variantes han ganado gran popularidad por su efectividad en segmentar estructuras pequeñas y detalladas. Su diseño simétrico de codificación y decodificación permite recuperar la información espacial de forma eficiente, sin embargo, sus requerimientos computacionales están para considerarse. También, las otras arquitecturas que no pertenecen a ninguna categoría del aprendizaje profundo debido a sus arquitecturas híbridas mantienen un enfoque en resolver retos específicos, como la segmentación de vasos delgados o de bajo contraste, pero en general requieren un preprocesamiento más complejo. En cuanto al aprendizaje automático, las mencionadas técnicas tradicionales han sido útiles por su bajo consumo de memoria y su velocidad de procesamiento, lo que los hace viables en aplicaciones clínicas en tiempo real. Sin embargo, su precisión puede ser limitada en comparación con las redes profundas. Finalmente otros de los modelos mencionados son los que surgen de enfoques semi-supervisados y no supervisados, útiles cuando se dispone de pocos datos etiquetados han demostrado ser prometedoras en contextos donde se busca un balance entre rendimiento y eficiencia, aunque su entrenamiento puede ser más complejo e inestable. En resumen, los diferentes enfoques de segmentación revisados presentan diversas fortalezas y limitaciones, y su aplicabilidad depende del equilibrio entre la precisión que ofrecen y los recursos computacionales que requieren. Independientemente del tipo de arquitectura utilizada, ya sea profunda, tradicional o híbrida, es claro que el rendimiento de cualquier modelo está directamente influenciado por la calidad de los datos de entrada. Es por ello que, previo al entrenamiento o inferencia, resulta indispensable aplicar técnicas de preprocesamiento.

Un aspecto transversal a todos estos enfoques es la importancia del preprocesamiento, especialmente en imágenes con bajo contraste. El uso de técnicas como la conversión a escala de grises, la normalización y el aumento de datos están presentes en la mayoría de los estudios mencionados y estos han demostrado ser esencial para reducir la complejidad de la imagen y mejorar la robustez del modelo, sin embargo, CLAHE destaca como la técnica de preprocesamiento más empleada en este conjunto, ya que aborda uno de los principales desafíos en la segmentación de imágenes médicas: el bajo contraste en estructuras, especialmente en vasos de baja visibilidad. Por otro lado, la interacción entre los clasificadores y los descriptores es otro aspecto crítico, ya que la eficacia de un clasificador puede variar dependiendo de los descriptores utilizados, en el caso de estos estudios hubo mucha variabilidad entre cada uno de ellos a pesar de que muchos trabajan con arquitecturas similares, esto es debido a que deben de ser elegidos en función de factores como la precisión necesaria, la naturaleza de los datos y los recursos computacionales disponibles, la variabilidad de opciones permite personalizar los modelos para maximizar el rendimiento en cada aplicación específica.

Abordando generalmente del recurso computacional, existen arquitecturas con alta precisión en la segmentación, especialmente para estructuras complejas como el disco óptico y la arcada vascular, donde la sensibilidad es clave, sin embargo requieren altos recursos computacionales durante el entrenamiento debido a sus complejas arquitecturas y el ajuste de múltiples parámetros y funciones de pérdida, esto quiere decir, que necesitan servidores de alto rendimiento o potentes GPU durante el entrenamiento, lo cual puede ser costoso en tiempo y recursos. Sin embargo, resultan ideales para aplicaciones clínicas y de diagnóstico en tiempo real tras el entrenamiento inicial. Finalmente, elegir el modelo adecuado dependerá del contexto de uso y la disponibilidad de recursos, balanceando entre el recurso computacional y la precisión requerida para cada tarea específica en este caso de la estructura anatómica que se está segmentando en las imágenes retinianas.

A partir de los resultados obtenidos en esta tesis, se evidencia que no existe un modelo único que cumpla con todos los requisitos clínicos, computacionales y contextuales de manera absoluta de igual manera es importante señalar que el recurso computacional no debe entenderse únicamente en función del tipo de hardware utilizado, sino también como una combinación de factores que incluyen el tiempo de entrenamiento, la frecuencia de actualización del modelo, el consumo energético y los recursos humanos necesarios para su implementación y mantenimiento. En cambio, los hallazgos respaldan la necesidad de una estrategia flexible, basada en la evaluación crítica de los modelos disponibles y su alineación con las condiciones reales del entorno donde se pretende implementar. Modelos como lo son U-Net y nnU-Net destacan por la precisión que tienen al momento de segmentar las regiones de interés, puntualmente en la arcada vascular, sin embargo, esta precisión viene acompañada de un recurso computacional considerable, especialmente durante la etapa de entrenamiento, lo cual puede representar una barrera en instituciones sin acceso a infraestructura de cómputo especializada, su rendimiento superior y la disponibilidad de herramientas preconfiguradas como en el caso de nnU-Net los convierten en una opción recomendable en fases iniciales de desarrollo, siempre que el entrenamiento pueda realizarse de forma centralizada o en colaboración con instituciones con mayor capacidad técnica. Por su parte SAM, que forma parte de arquitecturas generalistas abren una nueva perspectiva en cuanto a segmentación automática sin entrenamiento supervisado intensivo. No obstante, su uso aún presenta desafíos relacionados con la infraestructura necesaria para realizar inferencias en tiempo real, así como la falta de validación clínica específica en el dominio oftalmológico.

Finalmente, se reitera que la propuesta de esta tesis no busca establecer un modelo superior de forma absoluta, sino resaltar que la elección del algoritmo debe ser estratégica, contextual y adaptativa, en función de los recursos disponibles, el tipo de estructura a segmentar y el propósito clínico específico. Esta perspectiva flexible permite una implementación realista y escalable de herramientas de inteligencia artificial en escenarios diversos, favoreciendo una atención médica más equitativa y tecnológicamente sostenible.

5.2 Conclusiones

La presente tesis tuvo como objetivo principal identificar, evaluar y comparar diversos algoritmos de segmentación para estructuras anatómicas clave en imágenes de fundoscopia digital, considerando no solo el rendimiento algorítmico, sino también la viabilidad computacional en contextos clínicos con recursos limitados, como los que se presentan en regiones marginadas del estado de Baja California. A través de esta investigación, se demostró el como la integración de diversas técnicas como lo son los

modelos de aprendizaje automático y aprendizaje profundo permiten una segmentación eficiente para las estructuras anatómicas de interés, en ese caso, arcada vascular y nervio óptico, los cuales son fundamentales en el diagnóstico oftalmológico. Los resultados experimentales revelaron que, si bien modelos como nnU-Net ofrecen un alto desempeño en términos de precisión y robustez, también implican un mayor recurso computacional, lo que puede dificultar su implementación en entornos clínicos con capacidades tecnológicas limitadas. Por otro lado, modelos más livianos como UNet, aunque presentan un rendimiento ligeramente inferior, resultan más adecuados para escenarios donde se prioriza la accesibilidad y la operatividad sobre la sofisticación tecnológica. Asimismo, SAM se posiciona como una arquitectura generalista con potencial para aplicaciones exploratorias o escenarios donde se requiere adaptabilidad a distintos tipos de estructuras.

Por lo tanto, en este trabajo como argumento final se confirma que no existe un único modelo para todas las condiciones de los sistemas de salud posibles, la elección del modelo utilizado debe realizarse a partir de una evaluación contextual que contemple los objetivos clínicos, la infraestructura disponible y la naturaleza de las imágenes a procesar. Sin embargo, algo que se sostiene es como la segmentación automatizada de estructuras de fondo de ojo representa una herramienta prometedora para apoyar la equidad en salud visual, siempre y cuando se adapten sus implementaciones a las realidades locales. En resumen, esta tesis aporta al área de la oftalmología computacional al sugerir criterios técnicos y clínicos para la elección de algoritmos de segmentación, enfatizando la relevancia de encontrar soluciones asequibles, escalables y adaptadas al contexto de uso.

5.3 Trabajo futuro

A partir de los resultados obtenidos y del análisis crítico realizado, se identifican diversas líneas de trabajo que pueden fortalecer la investigación y ampliar su aplicabilidad clínica, primero que nada se debe de optimizar el proceso de generación de máscaras, hay que buscar una manera de mejorar la elaboración de las máscaras de segmentación utilizadas para entrenar los modelos. Contar con segmentaciones manuales de alta calidad y mejor definidas para mejorar la ambigüedad de las regiones anatómicas. También, se sugiere continuar con el crecimiento del conjunto de imágenes para entrenamiento, ya sea mediante la adquisición local de nuevas imágenes o a través de la incorporación de técnicas de aumento de datos, También buscar bases de datos para evaluación e inferencia, con el fin de validar la robustez de los modelos en poblaciones y condiciones clínicas diversas. En el caso de esta investigación, se abordó la segmentación de la arcada vascular y, en el caso de nnU-Net, también del nervio óptico. Se propone como línea futura extender el uso de U-Net y SAM para segmentar tanto el nervio óptico como la fovea, ya que como se mencionó a lo largo de esta tesis estas estructuras son clave en el diagnóstico oftalmológico. En el caso de que los modelos empleados no fuera posible abordar la segmentación de la fovea debido a que representa un paradigma distinto en cuanto a localización y representación. Será necesario evaluar si los modelos empleados son suficientemente robustos para esta tarea o si debe considerarse la adopción de otros modelos referenciados en la literatura, particularmente aquellos diseñados para identificar regiones centrales y estructuras circulares pequeñas. Finalmente, se propone realizar un análisis más detallado del impacto que tienen los recursos computacionales sobre el desempeño de los modelos, esta exploración permitiría entender si existen mejoras significativas en la eficiencia algorítmica que justifiquen la inversión en infraestructura de mayor capacidad.

5.4 Producción académica

Con el fin de cumplir el objetivo específico de obtener capacitación en conceptos generales de oftalmología, se realizó una formación complementaria teórico-práctica orientada a comprender las estructuras anatómicas del ojo, su representación en imágenes de fondo de ojo, su relevancia clínica en el diagnóstico y la manera adecuada de utilizar el equipo para la toma de imágenes de fondo de ojo.

Como parte de esta capacitación, se llevó a cabo una estancia formativa en Oftalmomédica en Ensenada, Baja California, donde se tuvo contacto directo con especialistas en salud visual y se observó el proceso clínico de captura e interpretación de imágenes retinianas. Esta experiencia permitió fortalecer la comprensión del contexto clínico real y las limitaciones que enfrentan los profesionales en entornos locales así como las dificultades que tienen para obtener las imágenes sin imperfecciones. Además, se completaron dos certificaciones en línea a través de la plataforma Cybersight, orientadas a la interpretación de imágenes de fondo de ojo y a la orientación para identificar de una manera eficiente las estructuras anatómicas más importantes en el fondo de ojo. Estas certificaciones ofrecieron una base sólida para identificar estructuras clave como el disco óptico, la copa, la mácula, la fovea y los vasos sanguíneos retinianos.

Esta formación fue esencial para guiar la selección de regiones anatómicas de interés en el diseño del sistema propuesto, así como para interpretar los resultados de segmentación desde una perspectiva clínica y no únicamente computacional. En la Figura 5.1, se presentan evidencias de dicha capacitación: dos imágenes de la clínica Oftalmomédica Ensenada, donde se llevó a cabo la estancia formativa, y las constancias de las dos certificaciones obtenidas en la plataforma Cybersight, enfocadas en el análisis clínico de imágenes de fondo de ojo.



Figura 5.1 Evidencia de la capacitación en fundamentos generales de oftalmología: fotografía de la clínica Oftalmomédica Ensenada y constancias de certificación en línea a través de la plataforma Cybersight.

Como parte de las actividades de formación académica vinculadas a esta tesis, se participó activamente en el *Seminario de Inteligencia Artificial y Cómputo Cuántico*, organizado por el CITED-IPN. Esta participación se llevó a cabo en colaboración con el equipo de *Medicina Computacional* de la Universidad Autónoma de Baja California, fortaleciendo las capacidades interdisciplinarias del proyecto y permitiendo el intercambio de conocimientos con investigadores y estudiantes de otras instituciones del país. Durante el evento se abordaron temas como el uso de algoritmos cuánticos para análisis de datos biomédicos, redes neuronales aplicadas a imágenes médicas, y perspectivas futuras de la inteligencia artificial en el ámbito de la salud, en la Figura 5.2 se encuentra la imagen del programa.



Figura 5.2 Participación en el *Seminario de Inteligencia Artificial y Cómputo Cuántico* en el CITED-IPN junto al equipo de Medicina Computacional de UABC.


Se participó como ponente en modalidad cartel con la presentación titulada “Segmentación de estructuras anatómicas en imágenes de fundoscopia digital con herramientas de procesamiento digital y aprendizaje automático”, en el marco del 3er. Coloquio de Aplicaciones de la Inteligencia Artificial en Ciencia, organizado por el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) el día 26 de junio de 2024. Este espacio académico permitió difundir los avances de la investigación de tesis, así como recibir retroalimentación de expertos en inteligencia artificial aplicada a las ciencias (Figura 5.3).

Otorga el presente

Reconocimiento A: Laura Johana Gonzalez Zazueta

Por su participación como ponente en modalidad cartel "Segmentación de estructuras anatómicas en imágenes de fundoscopia digital con herramientas de procesamiento digital y aprendizaje automático"

en el marco del 3er. Coloquio de aplicaciones de la Inteligencia Artificial en ciencia, llevado a cabo el 26 de Junio 2024.


Dra. Ana Isabel Martínez García
Jefa del Departamento
de Ciencias de la Computación



Dr. Eugenio Méndez Méndez
Director de Física Aplicada

Figura 5.3 Constancia de ponencia en modalidad cartel en el 3er. Coloquio de Aplicaciones de la Inteligencia Artificial en Ciencia, CICESE.

Finalmente, se participó como ponente en modalidad cartel con la presentación titulada "Segmentación de estructuras anatómicas en imágenes de fundoscopia digital con herramientas de procesamiento digital y aprendizaje automático", en el marco del 2do. Coloquio de Informática Médica: Innovación y Actualización en Salud y Tecnología, organizado por el Hospital General Ensenada, el día 8 de noviembre de 2024. Este evento académico facilitó la exhibición de avances en investigaciones relacionadas con la salud y la tecnología, además de permitir la retroalimentación de expertos en informática médica y novedad en el uso de herramientas digitales y aprendizaje automático en la salud (Figura 5.4).



Figura 5.4 Constancia de participación como ponente en modalidad cartel en el 2do. Coloquio de Informática Médica: Innovación y Actualización en Salud y Tecnología, organizado por el Hospital General Ensenada.

6. Bibliografía

- [1] X. Huang *et al.*, “Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy,” *Front. Endocrinol.*, vol. 13, p. 946915, Sep. 2022, doi: 10.3389/fendo.2022.946915.
- [2] “Primer informe sobre desigualdades en salud en Mexico.” Accessed: May 21, 2025. [Online]. Available: http://sidss.salud.gob.mx/site2/docs/1_informe_desigualdad_mexico.pdf
- [3] Organización Mundial de la Salud, *Informe mundial sobre la visión*. Ginebra: Organización Mundial de la Salud, 2020. Accessed: May 06, 2025. [Online]. Available: <https://iris.who.int/handle/10665/331423>
- [4] R. Yao, W. Zhang, R. Evans, G. Cao, T. Rui, and L. Shen, “Inequities in Health Care Services Caused by the Adoption of Digital Health Technologies: Scoping Review,” *J. Med. Internet Res.*, vol. 24, no. 3, p. e34144, Mar. 2022, doi: 10.2196/34144.
- [5] “La precariedad laboral: el elefante no visibilizado en el Sistema de Salud Mexicano - Portal INSP.” Accessed: May 13, 2025. [Online]. Available: <https://www.insp.mx/informacion-relevante/la-precariedad-laboral-el-elefante-no-visibilizado-en-el-sistema-de-salud-mexicano>
- [6] W. Soler, M. Gómez Muñoz, E. Bragulat, and A. Álvarez, “El triaje: herramienta fundamental en urgencias y emergencias,” *An. Sist. Sanit. Navar.*, vol. 33, pp. 55–68, 2010.
- [7] P. Artal, “Optics of the eye and its impact in vision: a tutorial,” *Adv. Opt. Photonics*, vol. 6, no. 3, pp. 340–367, Sep. 2014, doi: 10.1364/AOP.6.000340.
- [8] E. Hecht, *Optics*, 5 ed. Boston: Pearson Education, Inc, 2017.
- [9] “Oftalmología en la práctica de la medicina general, 5e | AccessMedicina | McGraw Hill Medical.” Accessed: May 14, 2025. [Online]. Available: <https://accessmedicina.mhmedical.com/book.aspx?bookID=2801>
- [10] D. Atchison, *Optics of the Human Eye*, 2nd ed. Boca Raton: CRC Press, 2023. doi: 10.1201/9781003128601.
- [11] E. Trucco, T. MacGillivray, and Y. Xu, Eds., *Computational retina image analysis: tools, applications and perspectives*. in Elsevier and MICCAI Society book series. London: Academic Press, 2019.
- [12] A. A. Awad, H. A. Alkorbi, and H. Abu Serhan, “Allvar Gullstrand: The Only Ophthalmologist Who Won the Nobel Prize,” *Cureus*, vol. 16, no. 6, p. e62513, doi: 10.7759/cureus.62513.
- [13] I. C. Mann, “Notes on the Anatomy of the living Eye, as revealed by the Gullstrand Slitlamp,” *J. Anat.*, vol. 59, no. Pt 2, pp. 155–165, Jan. 1925.
- [14] F. Ribeiro, P. Ceia, and L. Jud, “Pseudophakic Eye Models,” in *Intraocular Lens Calculations*, J. Aramberri, K. J. Hoffer, T. Olsen, G. Savini, and H. J. Shamma, Eds., Cham: Springer International Publishing, 2024, pp. 121–134. doi: 10.1007/978-3-031-50666-6_5.
- [15] K. Irsch and D. L. Guyton, “Anatomy of Eyes,” in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds., Boston, MA: Springer US, 2009, pp. 11–16. doi: 10.1007/978-0-387-73003-5_253.
- [16] D. Lamb, “Anatomy of the Eye,” *Am. Assoc. Ind. Nurses J.*, vol. 12, no. 9, pp. 6–8, Sep. 1964, doi: 10.1177/216507996401200901.
- [17] “Ophthalmology / editores Myron Yanoff y Jay S. Duker ... [et al.].” Accessed: May 13, 2025. [Online]. Available: <http://biblioteca.ufm.edu/opac/record/177799>
- [18] E. Dhiravidachelvi and V. Rajamani, “A SURVEY ON ANATOMICAL STRUCTURES: IN FUNDUS RETINAL IMAGES,” *Int. J. Eng. Sci. Adv. Technol.*, vol. 3, no. 3, pp. 183–187.
- [19] R. L. Gregory, “Helmholtz’s Principle,” *Perception*, vol. 36, no. 6, pp. 795–796, Jun. 2007,

- doi: 10.1068/p3606ed.
- [20] *Helmholtz's ophthalmoscope*. Accessed: May 14, 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File%3AHelmholtz%27s_ophthalmoscope_Wellcome_M0010474.jpg?utm_source=chatgpt.com
- [21] E. C. Pérez Molina and L. León Veitía, “La fotografía de fondo de ojo como método de diagnóstico en el glaucoma,” *Medicentro Electrónica*, vol. 21, no. 1, pp. 3–10, Mar. 2017.
- [22] N. Panwar *et al.*, “Fundus Photography in the 21st Century—A Review of Recent Technological Advances and Their Implications for Worldwide Healthcare,” *Telemed. E-Health*, vol. 22, no. 3, pp. 198–208, Mar. 2016, doi: 10.1089/tmj.2015.0068.
- [23] T. J. Bennett and C. J. Barry, “Ophthalmic imaging today: an ophthalmic photographer’s viewpoint – a review,” *Clin. Experiment. Ophthalmol.*, vol. 37, no. 1, pp. 2–13, 2009, doi: 10.1111/j.1442-9071.2008.01812.x.
- [24] R. H. Corr, “Fundoscopy in the smartphone age: current ophthalmoscopy methods in neurology,” *Arq. Neuropsiquiatr.*, vol. 81, no. 5, pp. 502–509, May 2023, doi: 10.1055/s-0043-1763489.
- [25] S. C. Lee and Y. Wang, “Automatic retinal image quality assessment and enhancement,” presented at the Medical Imaging ’99, K. M. Hanson, Ed., San Diego, CA, May 1999, pp. 1581–1590. doi: 10.1117/12.348562.
- [26] “(PDF) Automatic image quality assessment in optical fundus images,” in *ResearchGate*, Accessed: May 04, 2025. [Online]. Available: https://www.researchgate.net/publication/251621420_Automatic_image_quality_assessment_in_optical_fundus_images
- [27] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, “Automated assessment of diabetic retinal image quality based on clarity and field definition,” *Invest. Ophthalmol. Vis. Sci.*, vol. 47, no. 3, pp. 1120–1125, Mar. 2006, doi: 10.1167/iovs.05-1155.
- [28] “Automatic no-reference quality assessment for retinal fundus images using vessel segmentation | IEEE Conference Publication | IEEE Xplore.” Accessed: May 04, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/6627771>
- [29] F. Yin *et al.*, “Automatic retinal interest evaluation system (ARIES),” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2014, pp. 162–165, 2014, doi: 10.1109/EMBC.2014.6943554.
- [30] Z. Yao, Z. Zhang, L.-Q. Xu, Q. Fan, and L. Xu, “Generic features for fundus image quality evaluation,” in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Sep. 2016, pp. 1–6. doi: 10.1109/HealthCom.2016.7749522.
- [31] D. Mahapatra, P. K. Roy, S. Sedai, and R. Garnavi, “A CNN based neurobiology inspired approach for retinal image quality assessment,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2016, pp. 1304–1307, Aug. 2016, doi: 10.1109/EMBC.2016.7590946.
- [32] F. Yu, J. Sun, A. Li, J. Cheng, C. Wan, and J. Liu, “Image quality classification for DR screening using deep learning,” in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2017, pp. 664–667. doi: 10.1109/EMBC.2017.8036912.
- [33] F. Shao, Y. Yang, Q. Jiang, G. Jiang, and Y.-S. Ho, “Automated Quality Assessment of Fundus Images via Analysis of Illumination, Naturalness and Structure,” *IEEE Access*, vol. 6, pp. 806–817, 2018, doi: 10.1109/ACCESS.2017.2776126.
- [34] G. J. Avilés-Rodríguez *et al.*, “Topological Data Analysis for Eye Fundus Image Quality Assessment,” *Diagnostics*, vol. 11, no. 8, p. 1322, Jul. 2021, doi: 10.3390/diagnostics11081322.
- [35] H. Schneiderman, “The Fundoscopic Examination,” in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., Boston: Butterworths, 1990. Accessed: May 15, 2025. [Online]. Available:

- <http://www.ncbi.nlm.nih.gov/books/NBK221/>
- [36] P. Roux, “Ophthalmoscopy for the general practitioner,” *South Afr. Fam. Pract.*, vol. 46, no. 5, pp. 10–14, Jun. 2004, doi: 10.1080/20786204.2004.10873079.
- [37] I. P. Chatziralli, E. D. Kanonidou, P. Keryttopoulos, P. Dimitriadis, and L. E. Papazisis, “The Value of Fundoscopy in General Practice,” *Open Ophthalmol. J.*, vol. 6, pp. 4–5, Mar. 2012, doi: 10.2174/1874364101206010004.
- [38] M. D. Abràmoff, M. K. Garvin, and M. Sonka, “Retinal Imaging and Image Analysis,” *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, Jan. 2010, doi: 10.1109/RBME.2010.2084567.
- [39] R. Anguita, L. Ferro Desideri, A. Loewenstein, and M. Zinkernagel, “The digital age in retinal practice,” *Int. J. Retina Vitre.*, vol. 10, no. 1, p. 67, Sep. 2024, doi: 10.1186/s40942-024-00580-2.
- [40] “Fundus Photography - an overview | ScienceDirect Topics.” Accessed: May 17, 2025. [Online]. Available: https://www-science-direct-com.translate.google.com/topics/medicine-and-dentistry/fundus-photography?x_tr_sl=en&x_tr_tl=es&x_tr_hl=es&x_tr_pto=sge
- [41] “Top 11 Fundus Cameras Companies in Global 2025 | Global Growth Insights.” Accessed: May 26, 2025. [Online]. Available: <https://www.globalgrowthinsights.com/blog/fundus-cameras-companies-597>
- [42] C. Mishra and K. Tripathy, “Fundus Camera,” in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: May 17, 2025. [Online]. Available: <http://www.ncbi.nlm.nih.gov/books/NBK585111/>
- [43] Y. García García, R. Rodríguez Guillén, and A. Taboada Crispi, “Mapeo de imágenes digitales de fondo de ojo atendiendo a rasgos de textura,” *Rev. Cuba. Cienc. Informáticas*, vol. 11, no. 1, pp. 106–121, Mar. 2017.
- [44] J.-P. O. Li *et al.*, “Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective,” *Prog. Retin. Eye Res.*, vol. 82, p. 100900, May 2021, doi: 10.1016/j.preteyeres.2020.100900.
- [45] “Adaptive Supply Chain Management | SpringerLink.” Accessed: May 18, 2025. [Online]. Available: <https://link.springer.com/book/10.1007/978-1-84882-952-7>
- [46] H.-P. Chan, L. M. Hadjiiski, and R. K. Samala, “Computer-Aided Diagnosis in the Era of Deep Learning,” *Med. Phys.*, vol. 47, no. 5, pp. e218–e227, Jun. 2020, doi: 10.1002/mp.13764.
- [47] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, “Deep Learning in Medical Image Analysis,” *Adv. Exp. Med. Biol.*, vol. 1213, pp. 3–21, 2020, doi: 10.1007/978-3-030-33128-3_1.
- [48] “Image Segmentation Using Deep Learning: A Survey.” Accessed: May 20, 2025. [Online]. Available: <https://libcon.rec.uabc.mx:2279/document/9356353>
- [49] A. Bali and S. N. Singh, “A Review on the Strategies and Techniques of Image Segmentation,” in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, Feb. 2015, pp. 113–120. doi: 10.1109/ACCT.2015.63.
- [50] A. Barragán-Montero *et al.*, “Artificial intelligence and machine learning for medical imaging: a technology review,” *Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB*, vol. 83, pp. 242–256, Mar. 2021, doi: 10.1016/j.ejmp.2021.04.016.
- [51] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
- [52] Z. Zhang, Q. Liu, and Y. Wang, “Road Extraction by Deep Residual U-Net,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018, doi: 10.1109/LGRS.2018.2802944.
- [53] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” Aug. 22, 2018, *arXiv:arXiv:1802.02611*. doi: 10.48550/arXiv.1802.02611.

- [54] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool,” *BMC Med. Imaging*, vol. 15, no. 1, p. 29, Aug. 2015, doi: 10.1186/s12880-015-0068-x.
- [55] M. J. Page *et al.*, “PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews,” *BMJ*, p. n160, Mar. 2021, doi: 10.1136/bmj.n160.
- [56] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, “Convolutional neural networks: an overview and application in radiology,” *Insights Imaging*, vol. 9, no. 4, Art. no. 4, Aug. 2018, doi: 10.1007/s13244-018-0639-9.
- [57] B. P. Yap and B. K. Ng, “Coarse-to-fine visual representation learning for medical images via class activation maps,” *Comput. Biol. Med.*, vol. 171, p. 108203, Mar. 2024, doi: 10.1016/j.compbiomed.2024.108203.
- [58] K. Chen, T. Qin, V. H.-F. Lee, H. Yan, and H. Li, “Learning Robust Shape Regularization for Generalizable Medical Image Segmentation,” *IEEE Trans. Med. Imaging*, vol. 43, no. 7, pp. 2693–2706, Jul. 2024, doi: 10.1109/tmi.2024.3371987.
- [59] K. Aurangzeb, R. S. Alharthi, S. I. Haider, and M. Alhussein, “Systematic Development of AI-Enabled Diagnostic Systems for Glaucoma and Diabetic Retinopathy,” *IEEE Access*, vol. 11, pp. 105069–105081, 2023, doi: 10.1109/access.2023.3317348.
- [60] D. I. Morís, Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, “Context encoder transfer learning approaches for retinal image analysis,” *Comput. Biol. Med.*, vol. 152, p. 106451, Jan. 2023, doi: 10.1016/j.compbiomed.2022.106451.
- [61] N. Abdullah Alkhalidi and H. T. Halawani, “Intelligent Machine Learning Enabled Retinal Blood Vessel Segmentation and Classification,” *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 399–414, 2023, doi: 10.32604/cmc.2023.030872.
- [62] T. Zhang, L. Wei, N. Chen, and J. Li, “Learning based multi-scale feature fusion for retinal blood vessels segmentation,” *J. Algorithms Comput. Technol.*, vol. 16, p. 174830262110653, Jan. 2022, doi: 10.1177/17483026211065369.
- [63] Y. Zhou *et al.*, “AutoMorph: Automated Retinal Vascular Morphology Quantification Via a Deep Learning Pipeline,” *Transl. Vis. Sci. Technol.*, vol. 11, no. 7, p. 12, Jul. 2022, doi: 10.1167/tvst.11.7.12.
- [64] S.-Y. Huang, W.-L. Hsu, R.-J. Hsu, and D.-W. Liu, “Fully Convolutional Network for the Semantic Segmentation of Medical Images: A Survey,” *Diagnostics*, vol. 12, no. 11, p. 2765, Nov. 2022, doi: 10.3390/diagnostics12112765.
- [65] J. Li, Q. Hu, A. Imran, L. Zhang, J. Yang, and Q. Wang, “Vessel Recognition of Retinal Fundus Images Based on Fully Convolutional Network,” in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, Japan: IEEE, Jul. 2018, pp. 413–418. doi: 10.1109/compsac.2018.10268.
- [66] S. Wang *et al.*, “Optic disc detection based on fully convolutional network and weighted matrix recovery model,” *Med. Biol. Eng. Comput.*, vol. 61, no. 12, pp. 3319–3333, Dec. 2023, doi: 10.1007/s11517-023-02891-2.
- [67] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021, doi: 10.1109/ACCESS.2021.3086020.
- [68] Y. Zhou *et al.*, “CF-Loss: Clinically-relevant feature optimised loss function for retinal multi-class vessel segmentation and vascular feature measurement,” *Med. Image Anal.*, vol. 93, p. 103098, Apr. 2024, doi: 10.1016/j.media.2024.103098.
- [69] Z. Li, C. Zhao, Z. Han, and C. Hong, “TUNet and domain adaptation based learning for joint optic disc and cup segmentation,” *Comput. Biol. Med.*, vol. 163, p. 107209, Sep. 2023, doi: 10.1016/j.compbiomed.2023.107209.
- [70] Y. Rong *et al.*, “Segmentation of retinal vessels in fundus images based on U-Net with self-calibrated convolutions and spatial attention modules,” *Med. Biol. Eng. Comput.*, vol. 61, no. 7, pp. 1745–1755, Jul. 2023, doi: 10.1007/s11517-023-02806-1.
- [71] K. Sun, Y. Chen, Y. Chao, J. Geng, and Y. Chen, “A retinal vessel segmentation method based

- improved U-Net model,” *Biomed. Signal Process. Control*, vol. 82, p. 104574, Apr. 2023, doi: 10.1016/j.bspc.2023.104574.
- [72] B. Wang, S. Wang, S. Qiu, W. Wei, H. Wang, and H. He, “CSU-Net: A Context Spatial U-Net for Accurate Blood Vessel Segmentation in Fundus Images,” *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1128–1138, Apr. 2021, doi: 10.1109/jbhi.2020.3011178.
- [73] J. Chen *et al.*, “TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation,” Feb. 08, 2021, *arXiv*: arXiv:2102.04306. doi: 10.48550/arXiv.2102.04306.
- [74] S. Guo, “CSGNet: Cascade semantic guided net for retinal vessel segmentation,” *Biomed. Signal Process. Control*, vol. 78, p. 103930, Sep. 2022, doi: 10.1016/j.bspc.2022.103930.
- [75] D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, “Hard Attention Net for Automatic Retinal Vessel Segmentation,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 12, pp. 3384–3396, Dec. 2020, doi: 10.1109/jbhi.2020.3002985.
- [76] P. Liskowski and K. Krawiec, “Segmenting Retinal Blood Vessels With Deep Neural Networks,” *IEEE Trans. Med. Imaging*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016, doi: 10.1109/tmi.2016.2546227.
- [77] V. Cherukuri, V. K. B.G., R. Bala, and V. Monga, “Multi-Scale Regularized Deep Network for Retinal Vessel Segmentation,” in *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan: IEEE, Sep. 2019, pp. 824–828. doi: 10.1109/icip.2019.8803762.
- [78] R. C. Deo, “Machine Learning in Medicine,” *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.
- [79] C. Zhu *et al.*, “Retinal vessel segmentation in colour fundus images using Extreme Learning Machine,” *Comput. Med. Imaging Graph.*, vol. 55, pp. 68–77, Jan. 2017, doi: 10.1016/j.compmedimag.2016.05.004.
- [80] Z. Fan *et al.*, “Optic Disk Detection in Fundus Image Based on Structured Learning,” *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 224–234, Jan. 2018, doi: 10.1109/jbhi.2017.2723678.
- [81] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Fourth edition, Global edition. New York: Pearson, 2017.
- [82] R. Geetharamani and L. Balasubramanian, “Automatic segmentation of blood vessels from retinal fundus images through image processing and data mining techniques,” *Sadhana*, vol. 40, no. 6, pp. 1715–1736, Sep. 2015, doi: 10.1007/s12046-015-0411-5.
- [83] D. A. Dharmawan, B. P. Ng, and S. Rahardja, “A new optic disc segmentation method using a modified Dolph-Chebyshev matched filter,” *Biomed. Signal Process. Control*, vol. 59, p. 101932, May 2020, doi: 10.1016/j.bspc.2020.101932.
- [84] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, “Introduction to Machine Learning, Neural Networks, and Deep Learning,” *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 14, doi: 10.1167/tvst.9.2.14.
- [85] S. Liu *et al.*, “Joint optic disc and cup segmentation using semi-supervised conditional GANs,” *Comput. Biol. Med.*, vol. 115, p. 103485, Dec. 2019, doi: 10.1016/j.compbiomed.2019.103485.
- [86] A. Ghosh, S. Khose, Y. S. Kamath, N. I. R. Kuzhupilly, and J. R. Harish Kumar, “Fovea Segmentation Using Semi-Supervised Learning,” in *2023 IEEE 20th India Council International Conference (INDICON)*, Hyderabad, India: IEEE, Dec. 2023, pp. 590–595. doi: 10.1109/indicon59947.2023.10440760.
- [87] “APTOS 2019 Blindness Detection.” Accessed: May 24, 2025. [Online]. Available: <https://kaggle.com/aptos2019-blindness-detection>
- [88] “Laplacian Operator - an overview | ScienceDirect Topics.” Accessed: May 25, 2025. [Online]. Available: <https://www.sciencedirect.com/topics/computer-science/laplacian-operator>
- [89] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, “Contrast-limited adaptive histogram equalization: speed and effectiveness,” in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, May 1990, pp. 337–345. doi: 10.1109/VBC.1990.109340.
- [90] A. M. Reza, “Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE)

- for Real-Time Image Enhancement,” *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 38, no. 1, pp. 35–44, Aug. 2004, doi: 10.1023/B:VLSI.0000028532.53893.82.
- [91] Y. Lu, L. Duanmu, Z. (John) Zhai, and Z. Wang, “Application and improvement of Canny edge-detection algorithm for exterior wall hollowing detection using infrared thermal images,” *Energy Build.*, vol. 274, p. 112421, Nov. 2022, doi: 10.1016/j.enbuild.2022.112421.
- [92] “OpenCV: Canny Edge Detection.” Accessed: May 25, 2025. [Online]. Available: https://docs.opencv.org/4.x/da/d22/tutorial_py_canny.html
- [93] “OpenCV: Image Thresholding.” Accessed: May 25, 2025. [Online]. Available: https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html
- [94] “Adaptive thresholding: A comparative study.” Accessed: May 25, 2025. [Online]. Available: <https://libcon.rec.uabc.mx:2279/document/6993140>
- [95] “OpenCV: Transformaciones morfológicas.” Accessed: May 25, 2025. [Online]. Available: https://docs.opencv.org/4.x/d9/d61/tutorial_py_morphological_ops.html
- [96] A. AL Qurri and M. Almekkawy, “Improved UNet with Attention for Medical Image Segmentation,” *Sensors*, vol. 23, no. 20, p. 8589, Oct. 2023, doi: 10.3390/s23208589.
- [97] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nat. Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.
- [98] M. E. A. Mokhtari, *amine0110/nnUNet*. (May 25, 2024). Accessed: May 27, 2025. [Online]. Available: <https://github.com/amine0110/nnUNet>
- [99] F. Isensee, *FabianIsensee/hiddenlayer*. (Mar. 20, 2025). Python. Accessed: May 27, 2025. [Online]. Available: <https://github.com/FabianIsensee/hiddenlayer>
- [100] A. Kirillov *et al.*, “Segment Anything,” Apr. 05, 2023, *arXiv*: arXiv:2304.02643. doi: 10.48550/arXiv.2304.02643.
- [101] *facebookresearch/segment-anything*. (May 28, 2025). Jupyter Notebook. Meta Research. Accessed: May 27, 2025. [Online]. Available: <https://github.com/facebookresearch/segment-anything>
- [102] “facebook/sam-vit-base · Hugging Face.” Accessed: May 27, 2025. [Online]. Available: <https://huggingface.co/facebook/sam-vit-base>
- [103] M. J. Cardoso *et al.*, *MONAI: An open-source framework for deep learning in healthcare*. (Nov. 2022). Python. doi: <https://doi.org/10.48550/arXiv.2211.02701>.