



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

Extracción y clasificación automática de características de glaucoma en imágenes de retina

Tesis

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

Presenta:

Hiram José Sandoval Cuellar

Dirigido por:

Dr. Saúl Tovar Arriaga

Co. Dirigido por:

Dr. Miguel Ángel Vázquez Membrillo



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

Extracción y clasificación automática de características de glaucoma en imágenes de retina

Tesis

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

Presenta:

Hiram José Sandoval Cuellar

Dirigido por:

Dr. Saúl Tovar Arriaga

Co. Dirigido por:

Dr. Miguel Ángel Vázquez Membrillo

Dr. Saúl Tovar Arriaga
Presidente

Dr. Miguel Ángel Vázquez Membrillo
Secretario

Dr. Jesús Carlos Pedraza Ortega
Vocal

Dr. Juan Manuel Ramos Arreguín
Suplente

M.C. Gendry Alfonso Francia
Suplente

Centro Universitario, Querétaro, QRO

México.

Mes y año de aprobación por el consejo

Dirección General de Bibliotecas UAQ

Esta tesis está dedicada a mi familia

Agradecimientos

Le agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por proveer los fondos necesarios para la realización de esta investigación. Así mismo, se le agradece a la Universidad Autónoma de Querétaro (UAQ) y a la Dirección de Investigación y Posgrado de la Facultad de Ingeniería por su apoyo y contribución para mi formación como estudiante de la maestría. Quiero extender mi agradecimiento a todos mis profesores, ya que fueron una pieza medular para que yo pudiera llevar a buen fin esta investigación, sin su apoyo y conocimiento no hubiera sido posible, especialmente a mi asesor de tesis, el Dr. Saúl Tovar Arriaga. Agradezco también al Dr. Miguel Ángel Vázquez Membrillo y al Dr. Edgar Daniel Fuentes Pérez por haberme brindado su apoyo en la parte médica, así como a todo el equipo de especialistas que me ayudaron y al Instituto Mexicano de Oftalmología por permitirme hacer una estancia y mejorar la investigación. Finalmente, todo esto no hubiera sido posible sin el apoyo de madre y mi familia, que siempre estuvo apoyándome durante esta etapa.

Abstract

Glaucoma is a disease that affects the vision of those who suffer from it, causing an irreversible loss of this sense. The disc cup relationship is one of the most important characteristics used to diagnose this condition.

Fundus images are widely used by ophthalmologists to evaluate the retina and detect glaucoma, which is, according to studies by the World Health Organization (WHO), the second leading cause of blindness in the world.

In this thesis, algorithms of Convolutional Neural Networks are studied for the evaluation of glaucoma through fundus images. First, a method based on an image-level classification is proposed, that is, it classifies whether a patient has glaucoma or not. This method consists of proposing a new architecture to carry out this task. This network consists of 15 layers to achieve a better extraction of the characteristics sought. In addition to that, a map of characteristics of each image is obtained, to be able to visualize in which part the network is focusing to perform a classification, helping to verify how good the architecture is.

An important step to achieve good performance in any type of network is the preprocessing of the images, within this step is the location of the area of interest. A new technique is proposed to locate the optic disc and thus cut the image to reduce processing time and focus only on the area of interest.

The database used was ORIGA, which contains 650 images annotated by specialists. From this database, manual segmentation was carried out in conjunction with the Mexican Institute of Ophthalmology, 100 disk images, and 50 cup images to train the second proposed network.

In the second architecture, a variation in the power supply of the network is proposed that is responsible for performing a classification at the pixel level. This change consists of converting the images to polar coordinates, reducing the class imbalance, and increasing the performance of the network. The architecture used was the well-known U-Net and residual blocks were added to it. This algorithm is responsible for segmenting the disc and the optic cup.

Resumen

El glaucoma es una enfermedad que afecta la visión de quienes lo padecen, provocando una pérdida irreversible de este sentido. La relación copa disco es una de las características más importantes y utilizadas para diagnosticar este padecimiento.

Las imágenes del fondo de ojo son ampliamente utilizadas por los oftalmólogos para evaluar la retina y detectar el glaucoma, que es, según estudios de la Organización Mundial de la Salud (OMS), la segunda causa de ceguera en el mundo.

En esta tesis se estudian algoritmos de Redes Neuronales Convolucionales para la evaluación del glaucoma mediante imágenes de fondo de ojo. Primero se propone un método basado en una clasificación a nivel imagen, es decir, clasifica si un paciente tiene glaucoma o no. Este método consiste en proponer una arquitectura nueva para realizar dicha tarea. Esta red consta de 15 capas para lograr una mejor extracción de las características buscadas. Aunado a eso, se obtiene un mapa de características de cada imagen, esto con el fin de poder visualizar en que parte se está centrando la red para realizar una clasificación, ayudando a comprobar que tan buena es la arquitectura.

Un paso importante para lograr un buen desempeño en cualquier tipo de red es el preprocesamiento de las imágenes, dentro de este paso esta la ubicación del área de interés. Se propone una nueva técnica para localizar el disco óptico, y de esa manera, recortar la imagen para lograr disminuir el tiempo de procesamiento, además de enfocarnos únicamente en el área de interés.

La base de datos utilizada fue ORIGA, que contiene 650 imágenes anotadas por especialistas. De esta base de datos se realizó una segmentación manual en conjunto con el Instituto Mexicano de Oftalmología, de 100 imágenes de disco y 50 de copa, para poder entrenar la segunda red propuesta.

En la segunda arquitectura se propone una variación en la alimentación de la red que se encarga de realizar una clasificación a nivel píxel. Este cambio consiste en convertir las imágenes a coordenadas polares, logrando disminuir el desbalance de clases, y aumentando el desempeño de la red. La arquitectura utilizada fue la ya conocida U-Net y se le agregaron bloques residuales. Este algoritmo se encarga de segmentar el disco y la copa óptica.

Índice General

Agradecimientos	II
Abstract	III
Resumen	V
Índice	VII
Índice de figuras	X
Índice de tablas	XII
Índice de algoritmos	XIII
Abreviaturas y siglas	XIV
1. Introducción	1
1.1. Glaucoma	1
1.2. Descripción del problema	4
1.3. Justificación	5
1.4. Hipótesis	6
1.5. Objetivos	7
1.5.1. Objetivo general	7
1.5.2. Objetivos particulares	7
2. Antecedentes	8
2.1. Estado del arte	8
2.1.1. Clasificación a nivel imagen	8
2.1.2. Clasificación a nivel píxel	10
2.2. Redes Neuronales Convolucionales	13

2.3.	Capas que conforman una CNN	14
2.3.1.	Convolución	14
2.3.2.	Capa de agrupación	17
2.3.3.	Funciones de activación	18
2.3.3.1.	Funciones lineales	19
2.3.3.2.	Funciones no lineales	20
2.3.3.2.1.	Función sigmoideal	20
2.3.3.2.2.	Función Tanh	21
2.3.3.2.3.	Función ReLu	22
2.3.3.2.4.	Función <i>Leaky</i> ReLu	23
2.3.4.	Capas totalmente conectadas (FC)	24
2.4.	Funciones de pérdida	25
2.4.1.	Entropía cruzada	25
2.4.2.	Función de pérdida Euclidiana	25
2.4.3.	Función de pérdida <i>Hinge</i>	25
2.5.	Técnicas para evitar el sobreajuste	26
2.5.1.	<i>Dropout</i>	27
2.5.2.	<i>Drop-Weights</i>	27
2.5.3.	Normalización por lotes (Batch Normalization)	27
2.5.4.	Aumento de datos	28
2.5.5.	Parada anticipada	28
2.6.	Selección del optimizador	28
2.6.1.	Descenso de gradiente	29
2.6.2.	Descenso de gradiente estocástico	30
2.6.3.	<i>Momentum</i>	30
2.6.4.	AdaGrad	30
2.6.5.	RMSProp	31
2.6.6.	Estimación de Momento Adaptivo (Adam)	31
2.7.	Arquitectura U-Net para segmentación de imágenes	31
2.8.	Bloques residuales	32

2.9.	Visualización para Redes Neuronales Convolucionales	33
2.10.	Síntesis de imágenes médicas usando CNN	35
3.	Metodología	37
3.1.	Adquisición de imágenes	37
3.1.1.	Base de datos ORIGA	38
3.2.	Recursos computacionales	39
3.3.	Preprocesamiento de las imágenes	39
3.4.	Clasificación a nivel imagen	43
3.5.	Clasificación a nivel píxel	46
3.5.1.	Segmentación de las imágenes utilizadas	46
3.5.2.	Transformación de imágenes a coordenadas polares	50
3.5.3.	Arquitectura propuesta	52
4.	Resultados y discusión	56
4.1.	Resultados para la región de interés	56
4.2.	Resultados para la clasificación a nivel imagen	59
4.3.	Resultados para la clasificación a nivel píxel	63
5.	Conclusiones y trabajo futuro	68
5.1.	Trabajo futuro	69
	Referencias	71
	Apéndice	79
A.	Anexos	79
B.	Artículo publicado	84
C.	Estancia académica	85
D.	Certificado de inglés	86

Índice de figuras

1.1. Imagen centrada en el nervio óptico	2
1.2. Regla ISNT	4
2.1. Ejemplo de un <i>kernel</i> de 2 x 2	15
2.2. Representación de la operación de convolución	16
2.3. Representación de la operación de convolución con un paso de 3	17
2.4. Representación de la operación de agrupamiento máximo	18
2.5. Función lineal	20
2.6. Función sigmoïdal	21
2.7. Función tanh	22
2.8. Función ReLu	23
2.9. Función <i>Leaky</i> ReLu	24
2.10. Capas FC	24
2.11. Modelo insuficientemente ajustado	26
2.12. Modelo sobreajustado	26
2.13. Modelo ajustado	26
2.14. Arquitectura sin <i>Dropout</i>	27
2.15. Arquitectura con <i>Dropout</i>	27
2.16. Tasa de aprendizaje pequeño	29
2.17. Tasa de aprendizaje grande	29
2.18. Buena elección de tasa de aprendizaje	29
2.19. Arquitectura U-Net	32
2.20. Bloque residual	33
3.1. Metodología propuesta para la clasificación a nivel imagen	37
3.2. Metodología propuesta para la clasificación a nivel píxel	38
3.3. Ejemplo de imágenes de pacientes sanos de la base de datos ORIGA	39
3.4. Ejemplo de imágenes de pacientes con glaucoma de la base de datos ORIGA ...	39

3.5. Centro de la imagen que será recortada	41
3.6. Resultado después de recortar la imagen y obtener el ROI	42
3.7. Arquitectura propuesta	44
3.8. Contorno del disco óptico después de colocar los puntos para segmentar	47
3.9. Disco óptico segmentado	48
3.10. Copa óptica marcada	48
3.11. Copa óptica segmentada	48
3.12. Representación de cambio de coordenadas	50
3.13. Imagen original y segmentada en coordenadas cartesianas	52
3.14. Imagen original y segmentada en coordenadas polares	52
3.15. La arquitectura propuesta está compuesta por una U-Net con bloques residuales ...	53
3.16. Bloque residual en este trabajo	54
4.1. Transformación a escala de grises UIT-R BT. 601-7	56
4.2. Transformación a escala de grises propuesta	57
4.3. Desempeño de la red con las diferentes tasas de aprendizaje utilizadas	61
4.4. Área bajo la curva	61

Índice de tablas

1.1. Lesiones por glaucoma en el nervio óptico	3
2.1. Trabajos más importantes en la clasificación a nivel imagen	10
2.2. Trabajos más importantes en la clasificación a nivel píxel	12
3.1. Comparación entre la imagen original y la imagen segmentada	49
4.1. Resultados para la obtención de la región de interés en la base de datos ORIGA ..	57
4.2. Obtención del ROI	58
4.3. Desempeño sin la obtención del ROI	60
4.4. Desempeño con la obtención del ROI	60
4.5. Mapa de activación en distintas imágenes	62
4.6. Comparación de nuestro método contra los del estado del arte	63
4.7. Resultados de especificidad y sensibilidad	64
4.8. Resultados de exactitud y error de superposición	65
4.9. Comparación del método propuesto con los vistos en la literatura	65
4.10. Comparación entre métodos utilizando la misma base de datos	66
4.11. Algunos resultados obtenidos comparando la imagen segmentada con la imagen resultante	66

Índice de algoritmos

3.1. Región de interés	42
------------------------------	----

Dirección General de Bibliotecas UAQ

Abreviaturas y siglas

Adam - Adaptive moment estimation - Estimación del momento adaptivo

ANN - Artificial Neural Network - Red Neuronal Artificial

AUC - Area under the curve - Área bajo la curva

CDR - Cup-to-disc ratio - Relación copa-disco

CNN - Convolutional Neural Network - Redes Neuronales Convolucionales

DL - Deep Learning - Aprendizaje profundo

FC - Fully connected - Totalmente conectado

FN - False negatives - Falsos negativos

FP - False positives - Falsos positivos

GPU - Graphics Processing Unit - Unidad de procesamiento gráfico

HRF - High Resolution Fundus - Fondo de ojo de alta resolución

IA - Inteligencia artificial

MLP - Multilayer perceptron - Perceptrón multicapa

OD - Optic disc - Disco óptico

OMS - Organización Mundial de la Salud

ORIGA - Online retinal fundus image database for glaucoma analysis and research - Base de datos en línea de imágenes del fondo de ojo de la retina para el análisis y la investigación del glaucoma

ReLU - Rectified Linear Units - Unidad Lineal Rectificada

RGB - Red Green Blue - Rojo Verde Azul

ROI - Region of interest - Región de interés

SGD - Stochastic Gradient Descent - Descenso de gradiente estocástico

SVM - Support Vector Machine - Máquina de Soporte de Vectores

TN - True negatives - Verdaderos negativos

TP - True positives - Verdaderos positivos

XGB - Extreme Gradient Boost - Aumento de gradiente extremo

Dirección General de Bibliotecas UAQ

Introducción

1.1 Glaucoma

El glaucoma es la segunda causa de enfermedades de ceguera en el mundo (Pascolini, Mariotti, Pokharel, Pararajasegaram, Etya'ale, Négrel y Resnikoff, 2004), siendo las cataratas el número uno. Debido a la que la enfermedad no puede ser curada y es irreversible, la condición es considerada más severa que las cataratas (Kazi, Ajmera, Sukhija y Devadkar, 2018).

El glaucoma se refiere a un grupo de enfermedades clínicas que comparten una característica común, que consiste en una profundización o excavación de la cabeza del nervio óptico (Pinto, 2019). Esta excavación es causada por la pérdida de axones en las células ganglionares de la retina que forman las fibras del nervio óptico. Cuando la pérdida de tejido del nervio óptico es importante, el campo visual se reducirá gradualmente, lo que resultará en una ceguera total si la pérdida de fibras continúa (Piñero, Lora y Andrés, 2005). La base biológica de la enfermedad aún no se comprende completamente y los factores que contribuyen a su progresión aún no están completamente caracterizados. Sin un tratamiento adecuado, el glaucoma puede progresar a una discapacidad visual, y eventualmente, a ceguera (Weinreb y Khaw, 2004).

Las imágenes del fondo de ojo de la retina son utilizadas por los oftalmólogos en el diagnóstico de enfermedades relacionadas con los ojos. La etapa de análisis de la imagen de la retina que sigue a la captura de la imagen se considera una piedra angular en el proceso de diagnóstico general (Gopalakrishnan, Almazroa, Raahemifar y Lakshminarayanan, 2015).

El análisis de imágenes de la retina incluye localizar y extraer muchas estructuras anatómicas de la retina en una vista separada, lo que facilita el diagnóstico, brinda más información y, por lo tanto, mejora la precisión del diagnóstico (Gopalakrishnan et al., 2015).

En la figura 1.1 se muestra una imagen de retina centrada en el nervio óptico donde se ilustran las partes más importantes.

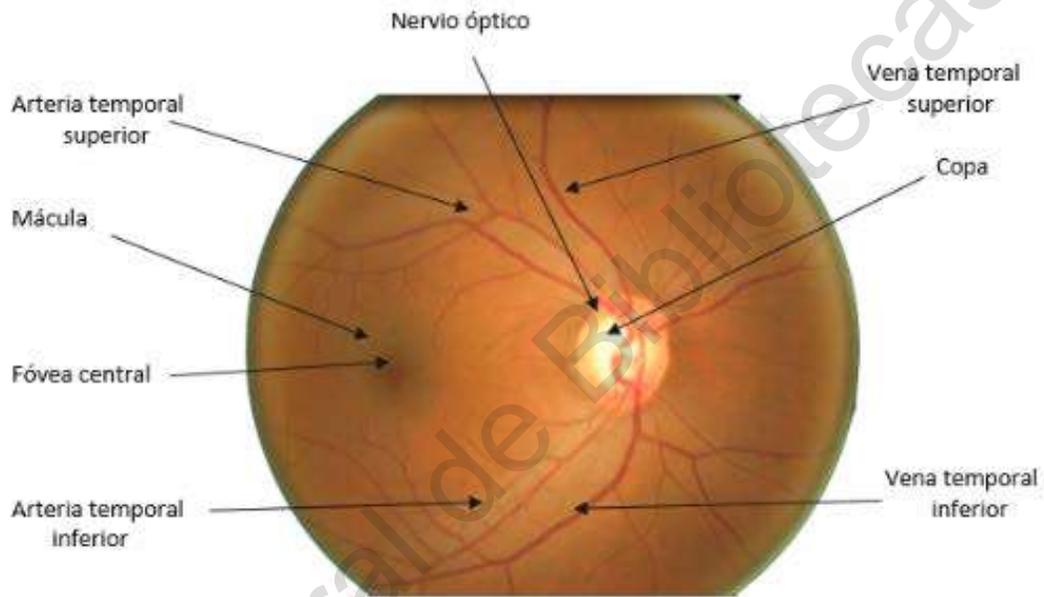


Figura 1.1: Imagen centrada en el nervio óptico. ORIGA (Zhang, Yin, Liu, Wong, 2010)

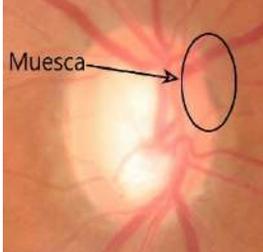
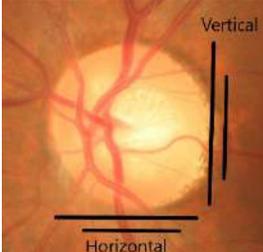
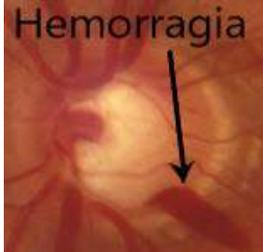
Las principales formas de glaucoma son el glaucoma de ángulo abierto, de ángulo cerrado, tensión normal y congénito (Pinto, 2019). El glaucoma de ángulo abierto también se denomina glaucoma primario, o crónico, y es el tipo más común de glaucoma, al menos el 90% de todos los casos (Weinreb et al., 2004).

Debido a que los síntomas iniciales del glaucoma no son obvios, los pacientes a menudo no experimentan dolor ni pérdida de la visión. El padecimiento puede notarse solo cuando la enfermedad progresa a un grado significativo, presentando pérdida de la visión periférica o ceguera completa (Huang y Hsiang, 2020). La detección temprana y el tratamiento oportuno son claves para prevenir la pérdida de visión de los pacientes, por lo

que deben desarrollarse métodos de diagnóstico efectivos. En términos de diagnóstico, la relación copa - disco (CDR) es normalmente usada para la detección de glaucoma (Yin, Liu, Wong, Tan, Cheung, Baskaran y Wong, 2012).

Además de la relación copa - disco existen otras características que se buscan en una imagen, dichas características se muestran en la Tabla 1.1.

Tabla 1.1: Lesiones por glaucoma en el nervio óptico

Descripción	Imagen
<p>Relación copa - disco. Esta relación debe ser menor a 0.5 para considerar que no se presenta glaucoma.</p>	
<p>Muecas en el disco óptico. Si alrededor del disco óptico se aprecian pequeñas muescas se podría sospechar que la persona podría presentar glaucoma.</p>	
<p>Si la relación copa - disco horizontal y vertical es mayor a 0.2 se podría sospechar que el paciente presenta glaucoma.</p>	
<p>Hemorragias en el disco óptico es otra señal de alerta de posible glaucoma.</p>	

Existe una prueba muy útil que se realiza para diferenciar un nervio óptico normal de uno que presenta glaucoma llamada ISNT. Normalmente la distancia Inferior entre el nervio óptico y la copa es mayor, seguido por la Superior, luego la Nasal y la más pequeña es la Temporal.

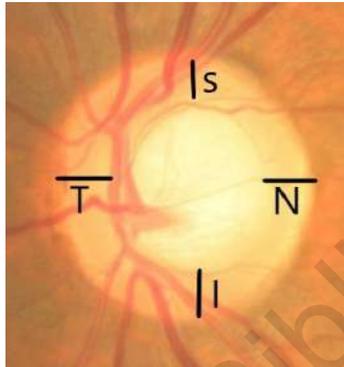


Figura 1.2: Regla ISNT. ORIGA (Zhang et al., 2010)

1.2 Descripción del problema

En abril del 2019 la *Food and Drug Administration* (FDA) publicó un documento donde propone una regulación para softwares que se basan en Inteligencia artificial como dispositivo médico. La idea detrás de esto es ofrecer una funcionalidad segura y eficaz que mejore la calidad de atención que recibirán los pacientes (Xu, Xue y Zhang, 2019).

Solamente existen dos sistemas en uso que apliquen técnicas de aprendizaje profundo en la identificación de imágenes, uno de ellos es IDx-DR por la FDA (Xu *et al.*, 2019), y el otro es EyeArt System, para Canada y la Unión Europea (Bhaskaranand *et al.*, 2019). Ambos sistemas detectan retinopatía diabética, he ahí la importancia de ampliar el rango de enfermedades que puedan detectar sistemas como estos. Estos equipos son sistemas bloqueados, es decir, no los puede usar cualquier persona y eso dificulta poder hacer una comparación.

Uno de los retos que se enfrentan a la hora de querer implementar un sistema que detecte el glaucoma es la falta de imágenes de entrenamiento, así como la diferencia que existe entre ellas. Es decir, no todas las imágenes de las bases de datos tienen las mismas características como lo son tamaño, niveles de luz, etc. Esto dificulta desarrollar un algoritmo que sea capaz de homogenizar todas estas características diferentes.

Todo el proceso que se lleva a cabo para poder dar un diagnóstico correcto es demasiado tardado y representa una tarea que requiere de mucho tiempo, además de ser muy tediosa para los oftalmólogos. Lo cual representa un costo muy alto ya que esta interpretación de imágenes las tiene que hacer médicos especializados (Sarhan, Rokne y Alhajj, 2019).

1.3 Justificación

Muchas enfermedades oculares, cerebrales, del sistema cardiovascular, y algunas otras enfermedades sistémicas se manifiestan en la retina, una de ellas es el glaucoma (Sertkaya, Ergen y Togacar, 2019).

Los estudios han indicado que el número de pacientes con glaucoma en todo el mundo aumentará a 111.8 millones para 2040 (Tham, Li, Wong, Quigley, Aung y Cheng, 2014) y, por lo tanto, es importante orientar la detección, el tratamiento, y el diseño de estrategias de salud pública relacionadas.

En México, existen un total de 2.237 millones de personas con discapacidad visual, algunas de las cuales quedan irreversiblemente ciegas como causa principal de glaucoma. Se estima que aproximadamente 1.5 millones de personas en México están afectadas por esta enfermedad, y hasta 50,000 casos de ceguera debido a su tardío descubrimiento. El glaucoma es una enfermedad que afecta al 2% de las personas mayores de 40 años y al 14% de las personas mayores de 60 años. El problema es que se estima que solo la mitad de los pacientes con glaucoma saben que tienen esta enfermedad porque no produce síntomas perceptibles en un inicio, y es difícil que los pacientes detecten la pérdida de visión periférica en una etapa temprana de la afección (Instituto de Salud para el Bienestar, 2020).

El tamaño y la forma del disco óptico son algunas de las estructuras importantes de la imagen del fondo de ojo, y son el parámetro auxiliar para juzgar diversas enfermedades oftálmicas (Maninis, Pont-Tuset, Arbeláez, y Van Gool, 2016). El disco también es una parte clave para detectar otras estructuras retinianas (Mittapalli, y Kande, 2016). Por lo tanto, el análisis del disco óptico retiniano en imágenes de fondo de ojo es un tema de investigación muy importante.

Aunque el glaucoma no se puede curar, si se diagnostica en una etapa temprana, su progresión puede ralentizarse o detenerse, llegando a salvarle la visión a una persona. Por lo tanto, la detección masiva de glaucoma es crucial para prevenir cualquier daño adicional (Gopalakrishnan et al., 2015). Las técnicas automatizadas para determinar la relación copa-disco pueden ser una ayuda útil para los médicos durante las evaluaciones, ya que apoyarán acelerando el diagnóstico, y aumentando la probabilidad de salvarle la visión a un individuo.

Si una persona presenta la enfermedad, es posible que integrantes de su familia la puedan desarrollar, por este motivo tendrán que realizarse exámenes regulares, aunque no presenten síntomas, para descartar cualquier riesgo (Neeru, 2015). Esto quiere decir que no solo personas con la patología deben de realizarse este estudio, sino también personas sanas que lo puedan desarrollar, esto hace mucho más complicada la labor de los especialistas en el diagnóstico de la enfermedad.

La segmentación del disco óptico en imágenes del fondo de ojo de la retina juega un papel fundamental en el diagnóstico de una variedad de patologías y anomalías oculares (Mohan, Kumar y Seelamantula 2019). La mayoría de las anomalías relacionadas con el disco óptico provocan cambios estructurales en las zonas interna y externa (Joshi, Masilamani, Alike, Suresh y Kumaresh 2019). La segmentación del disco óptico aumenta la sensibilidad de detección de diferentes enfermedades (Almotiri, Elleithy y Elleithy, 2018).

En los últimos años, el aprendizaje profundo se ha utilizado ampliamente en campo de la visión por computadora, que se beneficia del rápido desarrollo de red neuronal convolucional (Pan, Lu, Chen, y Xue, 2020). Varios procesamientos exitosos de imágenes médicas se han aplicado utilizando aprendizaje profundo.

1.4 Hipótesis

Es posible detectar algunas de las peculiaridades que tienen imágenes tomadas de fondo de ojo, a través de un preprocesamiento y técnicas de Inteligencia Artificial, logrando una mayor exactitud y sensibilidad en la detección de glaucoma en comparación con trabajos actuales.

1.5 Objetivos

1.5.1 Objetivo general

Desarrollar un algoritmo, que sea de apoyo para los oftalmólogos, y permita detectar las peculiaridades que presenta un ojo con glaucoma, usando imágenes de fondo de retina, a través de redes neuronales.

1.5.2 Objetivos particulares

- Seleccionar las bases de datos que se usarán durante el desarrollo del proyecto.
- Realizar un preprocesamiento de las imágenes para obtener la región de interés en la imagen.
- Realizar el diseño del algoritmo que sea capaz de detectar los rasgos distintivos que tienen imágenes con glaucoma.
- Realizar la comparación con trabajos previos para comparar los resultados obtenidos.

Antecedentes

2.1 Estado del arte

En el presente trabajo se abordan dos enfoques diferentes para la detección de glaucoma utilizando imágenes de retina. El primero es una clasificación a nivel de imagen, en este método el modelo de aprendizaje se entrena con imágenes previamente clasificadas por un experto (Kamber y Pei, 2006), en este caso, un especialista en retina. El segundo enfoque es una clasificación a nivel píxel, este método realiza una segmentación tanto del disco óptico, como de la copa. Las imágenes segmentadas utilizadas en el proceso de entrenamiento también deben de hacerse por especialistas del área. Debido a que son métodos diferentes, el estado del arte se dividió en trabajos relacionados a esos dos tipos de clasificaciones.

2.1.1 Clasificación a nivel imagen

Existen varios trabajos relacionados con este tipo de clasificación. Varios de estos métodos han logrado obtener métricas bastante buenas.

(Chen, Wong, Wong y Liu, 2015) proponen una arquitectura CNN de seis capas, donde cuatro son convolucionales y las dos últimas están completamente conectadas. En este estudio se utilizan ORIGA y SCES como base de datos. De la base de datos ORIGA, se seleccionaron aleatoriamente 99 imágenes para entrenamiento, y las 551 imágenes restantes se dejaron para la fase de prueba con resultados de un área bajo la curva (AUC) de 0.831. En un segundo experimento, se utilizan 650 imágenes de la base de datos ORIGA para el entrenamiento y 1676 imágenes de la base de datos SCES para las pruebas, el área bajo la curva obtenida fue 0.887.

(Acharya, Ng, Eugene, Noronha, Min, Nayak y Bhandary, 2015) proponen utilizar una máquina de soporte de vectores (SVM) para la clasificación, y la transformada de Gabor que notará los cambios sutiles en el fondo de la imagen. La base de datos utilizada fue una privada de Kasturba Medical College, Manipal, India, con 510 imágenes. El 90% de las imágenes se utilizaron para entrenamiento, mientras que el 10% restante para pruebas. Los resultados obtenidos fueron una exactitud del 93,10%, una sensibilidad del 89,75% y una especificidad del 96,20%.

(Raghavendra, Fujita, Bhandary, Gudigar, Tan y Acharya, 2018) proponen realizar el reconocimiento automático del glaucoma utilizando una red neuronal convolucional de 18 capas. Este trabajo consta de una CNN estándar, con capas de convolución y agrupación máxima, y una capa completamente conectada donde se realiza la clasificación. Inicialmente, el 70% de las muestras seleccionadas al azar se utilizan para entrenamiento y el 30% para pruebas. Se utilizaron 589 imágenes de pacientes sanos y 837 con glaucoma. El proceso se repitió cincuenta veces con particiones de prueba y entrenamiento aleatorias.

(Gómez-Valverde, Antón, Fatti, Liefers, Herranz, Santos y Ledesma-Carbayo, 2019) utilizan varias arquitecturas como VGG19, ResNet, GoogleNet y Denet Disc. Los mejores resultados obtenidos se obtuvieron utilizando VGG19 con aprendizaje de transferencia, obteniendo 0.9420 de AUC, 0.8701 de sensibilidad y 89.01 de especificidad.

(Diaz-Pinto, Morales, Naranjo, Köhler, Mossi y Navea, 2019) utilizan 5 bases de datos agregando un total de 1707 imágenes. Realizan la experimentación con cada una de las bases de datos por separado, pero los mejores resultados se obtuvieron juntando todas las imágenes disponibles. Alcanzan un AUC de 0.9605, una especificidad de 0.858 y una sensibilidad de 0.9346 utilizando la arquitectura Xception.

(Gour y Khanna, 2020) proponen un sistema automático de detección de glaucoma utilizando SVM para la clasificación. Combina GIST y PHOG para extraer características en las imágenes. Esta técnica elimina la necesidad de segmentar la imagen. En cambio, funciona con un sistema de diagnóstico que hace uso de características como la textura y la forma para

detectar la enfermedad. Este método arrojó una exactitud del 83.4% utilizando las bases de datos Drishti-GS1 y la de Fondo de Ojo de Alta Resolución (HRF).

(Gheisari, Shariflou, Phu, Kennedy, Agar, Kalloniatis y Golzan, 2021) proponen implementar 2 arquitecturas, VGG16 y ResNet, concatenando bloques LSTM. Para determinar cuál es la mejor, realizan varios experimentos variando el número de épocas y la tasa de aprendizaje. Se eligen 20 épocas y una tasa de aprendizaje de 0.001 como valores para obtener los mejores resultados. Los mejores resultados se logran con la red VGG16, logrando una sensibilidad del 95% y una especificidad del 96%.

Tabla 2.1: Trabajos más importantes en la clasificación a nivel imagen

Autor	Año	Método
Chen <i>et al.</i>	2015	CNN de 6 capas
Acharya <i>et al.</i>	2015	Extracción de características con transformada de Gabor (SVM)
Raghavendra <i>et al.</i>	2018	CNN de 18 capas
Gómez <i>et al.</i>	2019	VGG19 + Aprendizaje de transferencia
Díaz <i>et al.</i>	2019	Xception
Gour <i>et al.</i>	2020	Extracción de características utilizando GIST y PHOG (SVM)
Gheisari <i>et al.</i>	2021	VGG16 + LSTM

2.1.2 Clasificación a nivel píxel

La relación copa - disco (CDR) es estimada manualmente por un oftalmólogo profesional. Esto requiere mucho trabajo y tiempo. Con el fin de automatizar la cuantificación precisa de CDR y ayudar al diagnóstico de glaucoma, la segmentación de OD y OC está atrayendo mucha atención. En general, una CDR más grande sugiere un mayor riesgo de glaucoma (Joshi, Sivaswamy y Krishnadas, 2011).

(Zilly, Buhmann y Mahapatra, 2015) proponen procesar la representación de características y entrenar la CNN para generar un conjunto de filtros. El autor también utiliza muestreo de entropía y filtros de mejora para mejorar el rendimiento.

(Sevastopolsky, 2017) diseñó una U-Net mejorada para generar mapas de segmentación de la copa y el disco.

(Cheng, Liu, Xu, Yin, Wong, Tan y Wong, 2013) propusieron una red utilizando 2326 imágenes de fondo de ojo, incluidas 650 del Estudio de ojos de Malasia de Singapur (SiMES), y 1676 del Estudio de ojos de China de Singapur (SCES). En las imágenes de los conjuntos de datos, que tienen un tamaño de 3072 x 2048, la segmentación del disco óptico y la copa fueron realizadas manualmente por profesionales capacitados. Los métodos propuestos de segmentación fueron evaluados utilizando los límites manuales como verdad fundamental.

(Ronneberger, Fischer y Brox, 2015) proponen una arquitectura que consta de un camino de contracción y un camino de expansión. Incluye la aplicación repetida de dos convoluciones de 3 x 3, cada convolución seguida de una unidad lineal rectificadora (ReLU) y una operación de agrupación de 2 x 2. En cada paso de muestreo descendente, se duplicará el número de canales funcionales. Cada paso en la ruta de expansión incluye un muestreo del mapa de características, en seguida se realiza una convolución de 2 x 2, concatenando los correspondientes mapas de características. En la última capa, se usa convolución 1x1 para mapear cada vector al número requerido de clases.

(Fu, Cheng, Xu y Liu, 2019) propusieron un flujo de imágenes global y una red de guía segmentada, el aumento de datos se completó en el conjunto de entrenamiento mediante rotación aleatoria (0/90/180/270 grados). Para las transformaciones polares, se ajustan los parámetros de transformación a través del ángulo polar ($\varphi = 0/90/180/270$ grados) para controlar el aumento de datos. Durante el entrenamiento, utilizaron el descenso de gradiente estocástico para optimizar el modelo de profundidad. La tasa de aprendizaje que usaron disminuye gradualmente.

(Joshua, Nelwamondo y Mabuza-Hocquet, 2019) utilizan una arquitectura mejorada de red neuronal convolucional (CNN) U-net para segmentar el disco óptico y la copa óptica

de la imagen del fondo de ojo. Los mejores resultados obtenidos fueron 0.21 de error de superposición en la copa y 0.12 en el disco.

(Yin, Yuan, Cheng y Wu, 2020) proponen una arquitectura que consta de una red en forma de U, módulos de filtros guiados multiescala y capas de salida laterales. La red tiene imágenes RGB como entrada. La arquitectura es similar a U-Net, es decir, tiene una ruta de codificación y una ruta de decodificación. ReLu se utiliza como activación. La ruta de decodificación genera un mapa de características usando convoluciones.

(Afolabi, Mabuza-Hocquet, Nelwamondo y Paul, 2021) proponen un trabajo basado en 2 fases. En la primera fase se lleva a cabo el proceso de segmentación, y en la segunda, la detección. Para la segmentación, se utiliza un modelo U-Net *Lite* mientras que la detección se realiza con un aumento de gradiente extremo (XGB). La arquitectura de la red consta de la capa de codificación de muestreo descendente y la capa de decodificación de muestreo ascendente. La capa de codificación está formada por dos lotes de capas convolucionales de 3×3 conectadas a una capa de activación ReLU, y se conecta a una capa de agrupación máxima de 2×2 . La capa de decodificación concatena los mapas de características muestreados con la salida de la capa de codificación. El muestreo ascendente se realizó utilizando capas convolucionales de 2×2 .

Tabla 2.2: Trabajos más importantes en la clasificación a nivel píxel

Autor	Año	Método
Cheng <i>et al.</i>	2013	Superpixel
Ronneberger <i>et al.</i>	2015	UNet
Fu <i>et al.</i>	2019	MNet con transformación polar
Joshua <i>et al.</i>	2019	UNet modificada
Yin <i>et al.</i>	2020	<i>Deep Guidance Network</i>
Afolabi <i>et al.</i>	2021	UNet <i>Lite</i> con gradiente extremo

2.2 Redes Neuronales Convolucionales

El término aprendizaje profundo, o red neuronal profunda, se refiere a redes neuronales artificiales (ANN) con múltiples capas. Durante las últimas décadas, se ha considerado una de las herramientas más poderosas y se ha vuelto muy popular en la literatura, ya que es capaz de manejar una gran cantidad de datos. El interés por tener capas ocultas más profundas ha comenzado recientemente a superar el rendimiento de los métodos clásicos en diferentes campos, especialmente en el reconocimiento de patrones (Albawi, Mohammed y Al-Zawi, 2017). Una de las redes neuronales profundas más populares es la Red Neuronal Convolutiva (CNN). Toma este nombre de una operación matemática lineal entre matrices llamada convolución.

Aunque la CNN se hizo popular después del rendimiento récord de AlexNet (Krizhevsky, Sutskever y Hinton, 2012), en realidad se inició en 1980. Después de 2012, la CNN tomó el ritmo para hacerse cargo de diferentes campos de la visión por computadora, el procesamiento del lenguaje natural, y muchos más.

La fundación de la red neuronal convolutiva comenzó con el descubrimiento de Hubel y Wiesel. Según ellos, las células de la corteza visual animal reconocen la luz en el pequeño campo receptivo (Hubel y Wiesel, 1968). En 1980, inspirado por este trabajo, Kunihiko Fukushima propuso el Neocognitron. Esta red se considera el primer modelo teórico de CNN (Fukushima, 1980). En los 90's, se desarrolló el marco moderno de CNN llamado LeNet-5 (LeCun, Bottou, Bengio y Haffner, 1998) para reconocer dígitos escritos a mano. El algoritmo de entrenamiento por retro propagación (Rumelhart, Hinton y Williams, 1985) ayudó a LeNet-5 a reconocer patrones visuales en imágenes sin procesar directamente y sin utilizar ninguna extracción de características separadas.

Una red neuronal convolutiva tradicional está formada por bloques únicos, o múltiples, de capas de convolución y agrupación, seguidas de una o varias capas completamente conectadas (FC), y una capa de salida. La capa convolutiva es el bloque de construcción central de una CNN. Esta capa tiene como objetivo aprender las representaciones de entidades de la entrada. La capa convolutiva se compone de varios

filtros de convolución que aprenden y se utilizan para calcular diferentes mapas de características (Liu, Zou, Zhao y Liang, 2020). Cada unidad del mapa de características está conectado a un campo receptivo en la capa anterior. El nuevo mapa de características se produce aplicando una convolución a la entrada con los filtros, y aplicando la función de activación no lineal por elementos en el resultado. La propiedad de compartir parámetros de la capa convolucional reduce la complejidad del modelo (Ghosh, Sufian, Sultana, Chakrabarti y De, 2020). La capa de agrupación, o submuestreo, toma una pequeña región de la salida convolucional como entrada y la reduce para producir una única salida.

En el caso de la clasificación de imágenes, una CNN puede verse como una combinación de dos partes: la parte de extracción de características, y la parte de clasificación. Tanto las capas de convolución, como las de agrupación, realizan la extracción de características (Ghosh *et al.*, 2020). Las capas totalmente conectadas se agregan como clasificador y se asigna una probabilidad para clasificar la imagen.

2.3 Capas que conforma una CNN

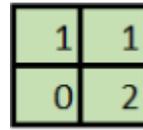
Las CNN se componen de múltiples capas, todas tienen una función y ayudan al correcto funcionamiento de la red.

2.3.1 Convolución

La capa convolucional es el componente más importante de cualquier arquitectura CNN. Contiene un conjunto de filtros convolucionales, que se combinan con la imagen de entrada para generar un mapa de características de salida (Zeiler y Fergus, 2014).

Con el objetivo de entender la convolución, primero se tiene que tener claro que es un filtro de convolución, o *kernel*. Un *kernel* se puede describir como una cuadrícula de valores o números discretos, donde cada valor se conoce como el peso de este filtro. Durante el inicio del proceso de entrenamiento de un modelo de CNN, todos los pesos de un *kernel* se asignan con números aleatorios. Luego, con cada época de entrenamiento, los pesos se ajustan y el filtro aprendió a extraer características significativas. Las CNN utilizan un

conjunto de múltiples filtros en cada capa de convolución para que cada filtro pueda extraer los diferentes tipos de funciones. La figura 2.1 muestra un ejemplo de un *kernel*.



1	1
0	2

Figura 2.1: Ejemplo de un *kernel* de 2 x 2.

Cada capa en una red convolucional es una red estructurada de 3 dimensiones que tiene altura (L_q), ancho (B_q) y profundidad(d_q). La profundidad de la capa no debe de ser confundida con la profundidad de la red en sí (Aggarwal, 2018).

La operación de convolución superpone el filtro en cada posible posición en la imagen y realiza un producto punto entre los parámetros del filtro y la red de entradas emparejadas. Las dimensiones de la imagen resultante, después de aplicar la operación de convolución, se pueden calcular con las ecuaciones 2.1 y 2.2 (Aggarwal, 2018).

$$L_{q+1} = L_d - F_d + 1 \quad (2.1)$$

$$B_{q+1} = B_d - F_d + 1 \quad (2.2)$$

El número de alineaciones entre el filtro y la imagen define la altura y el ancho, así como el número de filtros aplicados nos dará la profundidad de la siguiente capa oculta. Los filtros en las primeras capas se inclinan a detectar formas más primitivas, mientras que las últimas capas suelen localizar componentes más complejos.

En la figura 2.2 se muestra una representación de lo que sería la operación de convolución, aplicando un *kernel* de 2 x 2, a una imagen teórica de 5 x 5.

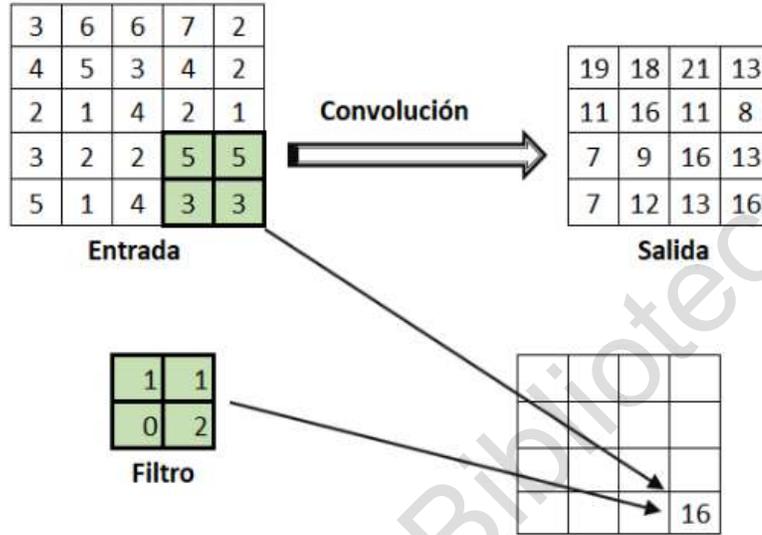


Figura 2.2: Representación de la operación de convolución.

Existen maneras de reducir el tamaño de la imagen tanto como necesitemos entre cada convolución, de esa manera se hace el proceso más rápido y se utilizan menos recursos. Uno de los métodos que se utilizan consiste en modificar el paso (S_d), esto quiere decir que ya no se moverá el filtro de píxel en píxel, sino que ahora se moverá tantos píxeles indique el paso.

Las dimensiones de la imagen resultante después de aplicar la operación de convolución con un paso mayor a 1 se puede calcular con las ecuaciones 2.3 y 2.4 (Aggarwal, 2018).

$$L_{q+1} = \frac{L_d - F_d}{S_d} + 1 \quad (2.3)$$

$$B_{q+1} = \frac{B_d - F_d}{S_d} + 1 \quad (2.4)$$

En las configuraciones que usualmente se usan, es común usar un paso de 1, aunque se podría llegar a usar de 2 o 3. También es recomendable usar imágenes que tengan tanto la

misma altura como el ancho, en otras palabras, que sean imágenes cuadradas. El número de filtros en cada capa es normalmente una potencia de 2 (Aggarwal, 2018).

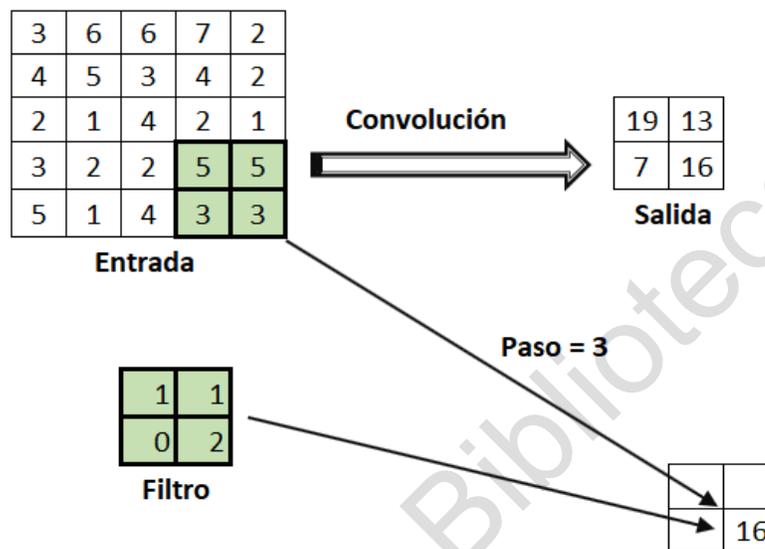


Figura 2.3: Representación de la operación de convolución con un paso de 3.

2.3.2 Capa de agrupación

Las capas de agrupación se utilizan para sub muestrear los mapas de características que son producidos después de la capa de convolución, es decir, toma los mapas de características de mayor tamaño y los reduce a mapas de características de menor tamaño. Al reducir los mapas de características, siempre conserva las características más dominantes en cada paso del grupo. La operación de agrupación se realiza especificando el tamaño de la región agrupada y el paso de la operación, similar a la operación de convolución. Hay diferentes tipos de técnicas de agrupación que se utilizan en diferentes capas de agrupación, como agrupación máxima, agrupación mínima, agrupación promedio, agrupación cerrada, agrupación de árboles, etc (LeCun, 1998). La agrupación máxima es la técnica de agrupación más popular y más utilizada. La capa de agrupación ayuda a una CNN a encontrar si una

característica específica está presente en la imagen de entrada dada o no, sin preocuparse por la posición correcta de esa característica.

Debido a la agrupación podemos concluir que nuestros resultados son invariantes a la translación, ya que un pequeño cambio a la imagen no afecta significativamente los parámetros de salida que se obtienen al aplicar dicha operación. De manera similar, puede identificar características cuando se encuentran desplazadas en algún lugar de la imagen.

Para representar el funcionamiento de la operación de agrupación, la figura 2.4 se muestra cómo sería un agrupamiento máximo (*max-pooling*).

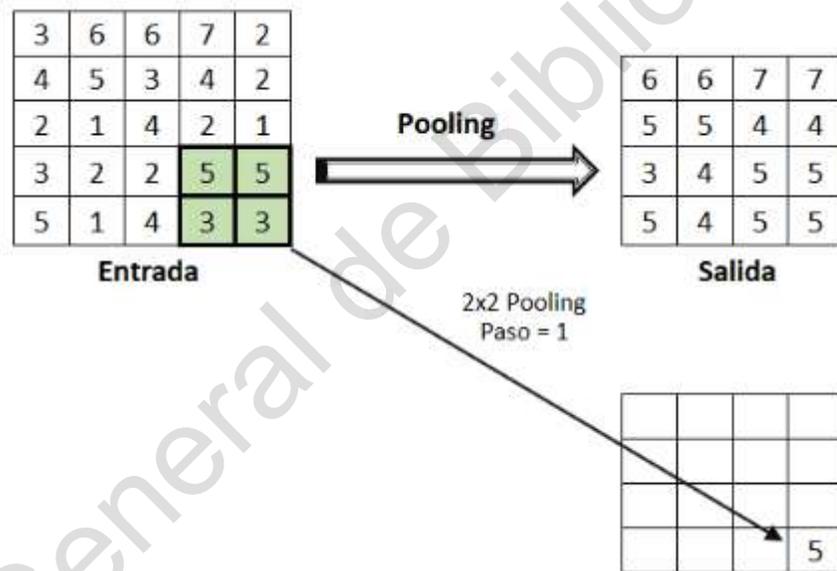


Figura 2.4: Representación de la operación de agrupamiento máximo.

2.3.3 Funciones de activación

Para una Red Neuronal Artificial, siempre hay muchas neuronas que funcionan en correspondencia de pesos, y a través de estas neuronas se pueden obtener los resultados de las entradas (Jain, Mao y Mohiuddin, 1996). Sin funciones de activación, las salidas pueden

ser cualquier cosa en el rango de números infinitos, por lo que las neuronas realmente no conocen los límites del valor.

La tarea principal de cualquier función de activación en un modelo basado en redes neuronales es mapear la entrada a la salida, donde el valor de salida se obtiene calculando la suma ponderada de la entrada. En otras palabras, la función de activación decide si una neurona funcionará o no para una entrada determinada al producir la salida correspondiente.

La función de activación también se conoce como función de transferencia. Las funciones de activación hacen posible que se actualicen los pesos, ya que los gradientes se proporcionan junto con el error para actualizarlos (LeCun, Boser, Denker, Henderson, Howard, Hubbard y Jackel, 1990).

De todas las funciones de activación no saturadas, la Unidad Lineal Rectificada (ReLU) es la más famosa (Nair y Hinton, 2010). G.E. Dahl mejoró su trabajo un 4,2% utilizando ReLU en un sistema de Redes Neuronales Profundas (Dahl, Sainath y Hinton, 2013). Kazuyuki Hara analizó dicha función y explicó la razón por la que ReLU mejora el proceso de aprendizaje en la teoría (Hara, Saito y Shouno, 2015).

Las funciones de activación se pueden dividir básicamente en 2 tipos, funciones de activación lineal y funciones de activación no lineal (Feng y Lu, 2019). Donde las funciones de activación lineal mantienen una constante, y las funciones de activación no lineales crean una variación que utiliza la construcción de la red neuronal.

2.3.3.1 Funciones lineales

La función de activación lineal tiene una ecuación similar a la de una línea recta, la activación es proporcional a la entrada, está dada por la ecuación 2.3.

$$f(x_i) = kx_i \quad (2.3)$$

Donde x_i es la entrada de la función de activación, y k es la constante fija. La representación gráfica de dicha función se puede observar en la figura 2.5.

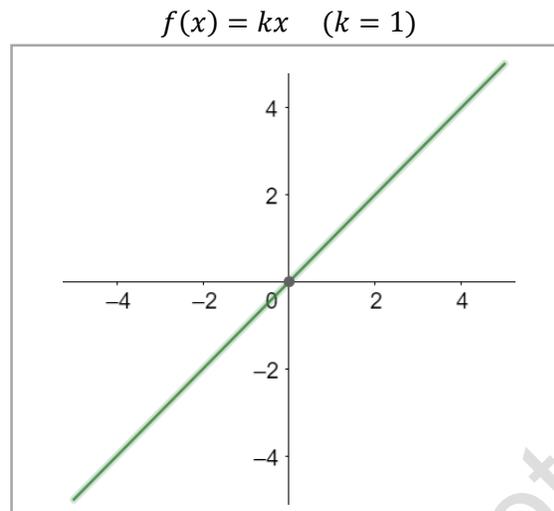


Figura 2.5: Función lineal.

2.3.3.2 Funciones no lineales

En comparación con las funciones de activación lineal, las funciones de activación no lineales son utilizadas más ampliamente. Facilitan que el modelo se generalice o se adapte con una variedad de datos y diferenciar entre los resultados (Feng *et al.*, 2019). Hay muchas funciones de activación no lineales típicas, y siempre se clasifican por su rango o curvas.

2.3.3.2.1 Función sigmoideal

El rango de esta función es [0 1] y su curva parece formar una S. Aunque es monótona, y su salida es muy pronunciada, pequeños cambios de entrada traerían grandes cambios de salida. En cualquiera de sus extremos, la salida tiende a 0 o 1. Esta señal sigue siendo muy popular en problemas de clasificación, especialmente en capas de salida donde la clasificación es binaria. La ecuación 2.4 muestra la fórmula que describe esta función.

$$f(x_i) = \frac{1}{1 + e^{-x_i}} \quad (2.4)$$

Donde x_i es la entrada de la función de activación. La representación gráfica de dicha función se puede observar en la figura 2.6.

$$f(x) = \frac{1}{1 + e^{-x}}$$

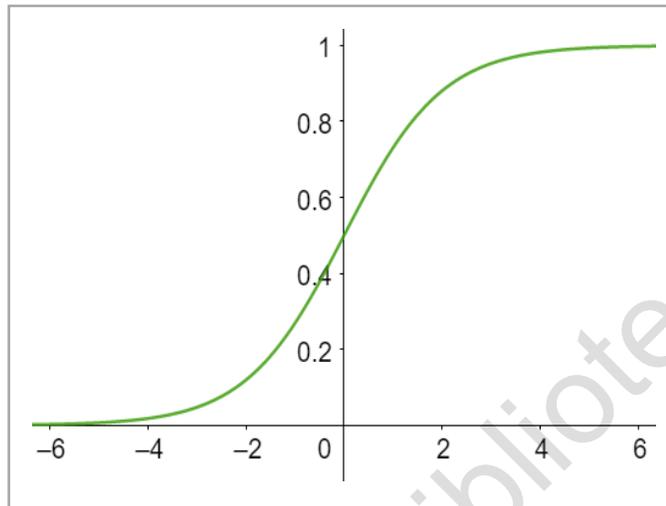


Figura 2.6: Función sigmoideal.

2.3.3.2.2 Función Tanh

Comparada con la función sigmoideal, la función Tanh tiene la diferencia de que su salida es centrada en cero y su gradiente es más fuerte. Su rango de $[-1, 1]$ hace que la media de las salidas sea cero, o muy cercana a cero, además, hace que los datos estén más concentrados y facilita mucho el aprendizaje, por lo que se suele utilizar en capas ocultas.

La ecuación 2.5 muestra el comportamiento de dicha función.

$$\tanh(x_i) = \frac{2}{1 + e^{-2x_i}} - 1 \quad (2.5)$$

Donde x_i es la entrada de la función de activación. La representación gráfica de dicha función se puede observar en la figura 2.7.

$$\tanh(x_i) = \frac{2}{1 + e^{-2x_i}} - 1$$

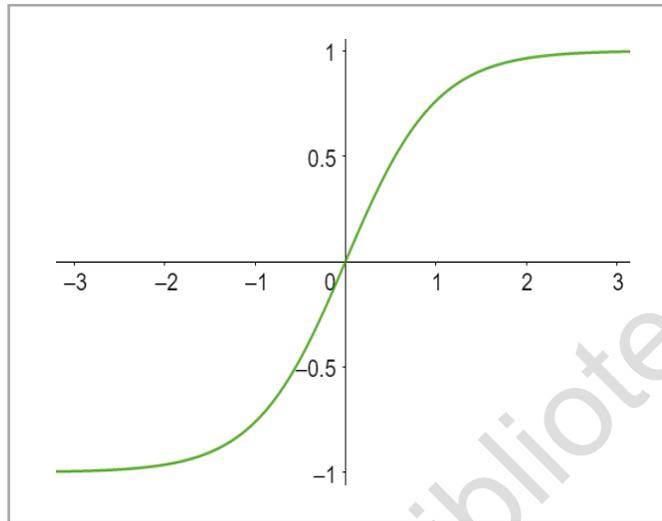


Figura 2.7: Función tanh.

2.3.3.2.3 Función ReLu

La función ReLu se ha vuelto muy popular en los últimos años, generalmente se implementa en capas ocultas de Redes Neuronales Artificiales (ANN), especialmente en casi todas las Redes Neuronales Convolucionales (CNN).

Esta función parece algo lineal, pero de hecho no lo es, porque la salida es 0 cuando $x_i < 0$. Además, evita y rectifica el problema de la desaparición del gradiente presentes en las funciones Sigmoidal y Tanh. Aunado a eso, tiene la característica de ser menos costosa computacionalmente que estas dos últimas.

La función ReLu no tiene límite superior, y debido a que el gradiente es 0 para las entradas negativas, los pesos no se ajustarán durante el descenso. Además, las neuronas que entran en ese estado dejarán de responder a las variaciones de las entradas, se atascan en un estado perpetuamente inactivo y “mueren”. esto se llama problema de “*Dying ReLu*” (Hu, 2018).

El comportamiento de esta función se puede obtener aplicando las ecuaciones 2.6.

$$f(x_i) = \max(0, x_i) = \begin{cases} x_i, & x_i \geq 0 \\ 0, & x_i < 0 \end{cases} \quad (2.6)$$

Donde x_i es la entrada de la función de activación. La representación gráfica de dicha función se puede observar en la figura 2.8.

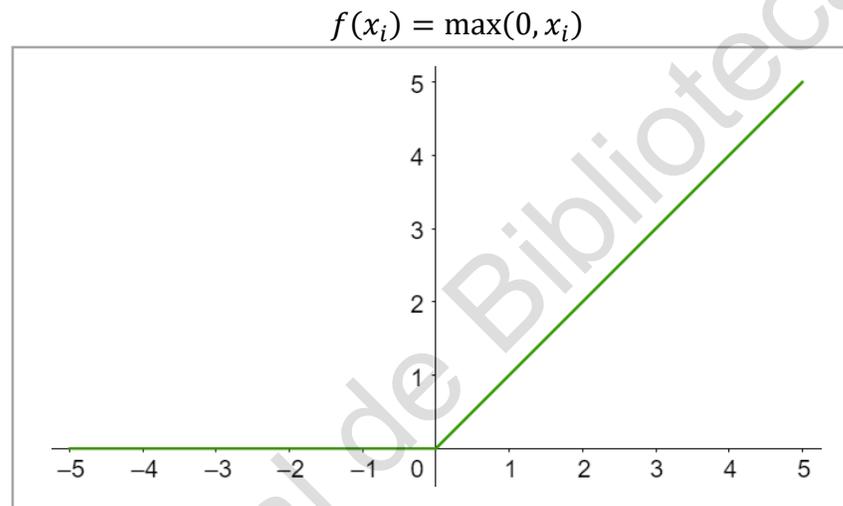


Figura 2.8: Función ReLu.

2.3.3.2.4 Función *Leaky ReLu*

El comportamiento de la función *Leaky ReLu* esta descrito por la ecuación 2.7.

$$f(x_i) = \begin{cases} x_i, & x_i \geq 0 \\ \alpha * x_i, & x_i < 0 \end{cases} \quad (2.7)$$

Donde α es el coeficiente por el cual se multiplican las entradas negativas, y es una pequeña constante, por lo general 0.01. Su salida no es 0 para entradas negativas, por lo que soluciona el problema de “*Dying ReLu*” que tiene la función ReLu normal (Hu, 2018).

La representación gráfica de dicha función se puede observar en la figura 2.9.

$$f(x_i) = \begin{cases} x_i, & x_i \geq 0 \\ \alpha * x_i, & x_i < 0 \end{cases} \quad (\alpha = 0.01)$$

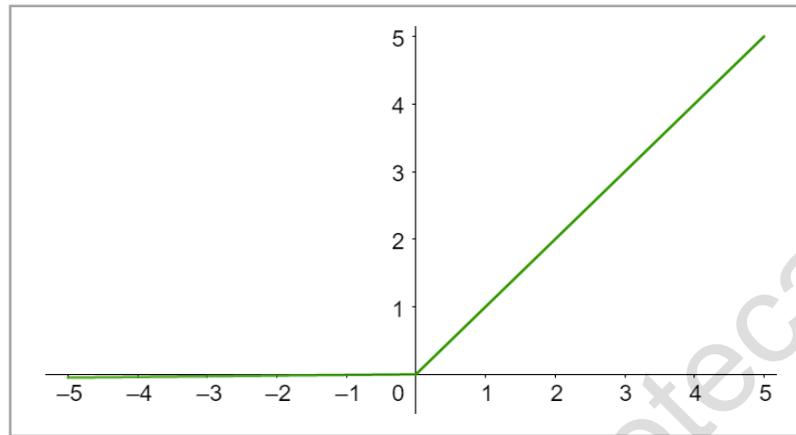


Figura 2.9: Función *Leaky ReLu*.

2.3.4 Capas totalmente conectadas (FC)

La última parte de cada arquitectura CNN consta de capas completamente conectadas, donde cada neurona dentro de una capa está conectada con cada neurona de su capa anterior. La última capa FC se utiliza como salida y es el clasificador de la arquitectura CNN. Las capas totalmente conectadas son un tipo de Red Neuronal Artificial de retroalimentación y siguen el principio de la red neuronal de percepción multicapa tradicional (MLP). Las capas FC toman información de la capa convolucional y generan un mapa de características que servirá para la toma de decisiones.

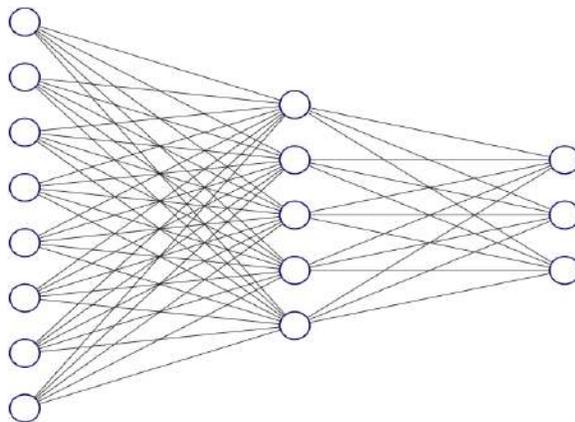


Figura 2.10: Capas FC

2.4 Funciones de pérdida

Se conoce que en la última capa de salida es donde se lleva a cabo la clasificación. En esta capa, calculamos el error de predicción generado por el modelo CNN sobre las muestras de entrenamiento usando alguna función de pérdida. Este error de predicción le dice a la red qué tan lejos está su clasificación de la salida real, y luego este error se optimizará durante el proceso de aprendizaje del modelo. La función de pérdida usa dos parámetros para calcular el error, el primer parámetro es la salida estimada del modelo o predicción, y el segundo es la salida real o etiqueta (Yessou, Sumbul y Demir, 2020). Existen diferentes tipos de funciones de pérdida que se utilizan en diferentes tipos de problemas.

2.4.1 Entropía cruzada

La función de pérdida de entropía cruzada, también llamada función de pérdida logarítmica, se usa ampliamente para medir el rendimiento del modelo CNN, cuya salida es la probabilidad $p \in \{0, 1\}$. Se utiliza ampliamente como una alternativa de la función de pérdida del error cuadrático en los problemas de clasificación donde se tienen múltiples clases (Nie, Hu y Li, 2018). Utiliza activaciones Soft-Max en la última capa para generar una salida dentro de una distribución de probabilidad.

2.4.2 Función de pérdida Euclidiana

La pérdida Euclidiana, también llamada error cuadrático medio, se usa ampliamente en problemas de regresión. El error cuadrático medio, entre la salida predicha $p \in R^N$ y la salida real $y \in R^N$ de cada neurona de la capa de salida de la CNN, son las variables usadas para realizar la predicción (Wang y Bovik, 2009). Se llama error cuadrático medio, ya que calcula el promedio de un conjunto de errores. Cuanto menor sea el error, mejor será el pronóstico.

2.4.3 Función de pérdida *Hinge*

La función de pérdida *Hinge* se utiliza ampliamente en problemas de clasificación binaria. Se utiliza en problemas de clasificación basados en márgenes máximos, sobre todo

para máquinas de soporte de vectores (SVM). Aquí, el optimizador intenta maximizar el margen entre dos clases objetivo.

2.5 Técnicas para evitar el sobreajuste

El desafío principal de los algoritmos de aprendizaje profundo es adaptarse adecuadamente a entradas nuevas o nunca antes vistas, extraídas de la misma distribución que los datos de entrenamiento, la capacidad para hacerlo se llama generalización. El principal problema para que un modelo de CNN logre una buena generalización es el sobreajuste (Salman y Liu, 2019). Cuando un modelo tiene un rendimiento excepcional en los datos de entrenamiento, pero falla en los datos de prueba, este tipo de modelo se denomina sobreajustado. Lo contrario es un modelo insuficientemente ajustado, que ocurre cuando el modelo no ha aprendido lo suficiente de los datos de entrenamiento, y cuando el modelo funciona bien tanto en los datos de entrenamiento como en los de prueba, este tipo de modelos se denominan modelo ajustado. En las figuras 2.11, 2.12 y 2.13 se pueden observar modelos con las características antes mencionadas.

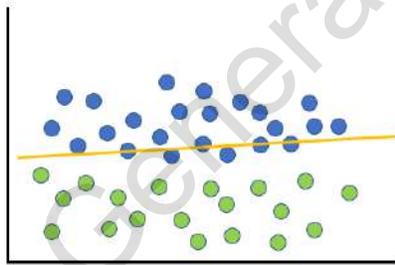


Figura 2.11: Modelo insuficientemente ajustado

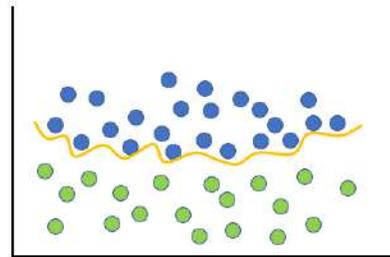


Figura 2.12: Modelo sobreajustado.

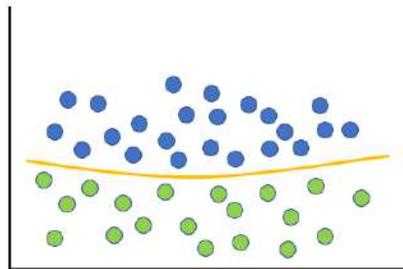


Figura 2.13: Modelo ajustado

2.5.1 Dropout

El *dropout* (Srivastava, Hinton, Krizhevsky, Sutskever y Salakhutdinov, 2014) es uno de los enfoques más utilizados para evitar el sobreajuste. En este método se apagan neuronas de la red al azar durante cada época de entrenamiento. Al apagar las neuronas, se intenta distribuir el poder de selección de características a todas las neuronas por igual, y se fuerza al modelo a aprender varias características independientes. Apagar una neurona significa que la unidad apagada no participará tanto en la propagación hacia adelante como en la propagación hacia atrás durante el proceso de entrenamiento.

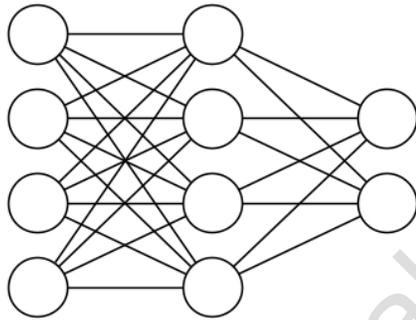


Figura 2.14: Arquitectura sin *Dropout*.

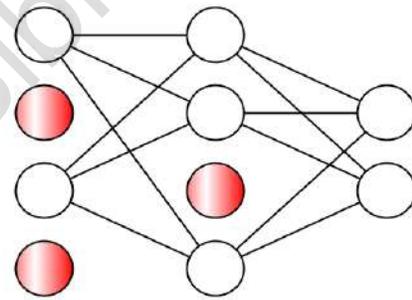


Figura 2.15: Arquitectura con *Dropout*.

2.5.2 Drop-Weights

Es muy similar al método de *Dropout*, pero en lugar de apagar neuronas, en este proceso se apagan pesos aleatorios en cada época del entrenamiento (Yildirim, Xu y Basaran, 2005).

2.5.3 Normalización por lotes (*Batch Normalization*)

La normalización por lotes (Ioffe y Szegedy 2015), asegura que las activaciones de salida de una red, seguirán una distribución gaussiana unitaria al normalizar la salida en cada capa, restando la media y dividiendo por la desviación estándar. La normalización por lotes se utiliza para reducir el cambio de covarianza interno de las capas de activación. El cambio

de covarianza interno se puede explicar como el cambio en la distribución de activaciones en cada capa. La operación de normalización por lotes se implementa como una capa en la arquitectura CNN. Reduce las posibilidades de sobreajuste porque tiene un ligero efecto de regularización.

2.5.4 Aumento de datos

La forma más fácil de evitar el sobreajuste es entrenar el modelo con una gran cantidad de datos. Existen varias técnicas para expandir artificialmente el tamaño del conjunto de datos de entrenamiento (Shorten y Khoshgoftaar, 2019). La más utilizada es la rotación de las imágenes.

2.5.5 Parada anticipada

En la parada anticipada, mantenemos una pequeña parte del conjunto de datos, que luego se utilizará para fines de validación cruzada, donde evaluamos el rendimiento del modelo entrenado sobre este conjunto en cada época del entrenamiento (Li, Soltanolkotabi y Oymak, 2020). Aquí utilizamos una estrategia que detiene el proceso de entrenamiento cuando el rendimiento en el conjunto de validación empeora en la época subsiguiente. A medida que aumenta el error de validación, la capacidad de generalización del modelo aprendido también disminuye.

2.6 Selección del optimizador

El objetivo principal de cualquier algoritmo de aprendizaje supervisado es minimizar el error, es decir, las diferencias entre la salida predicha y la salida real, basados en varios parámetros de aprendizaje como pesos, *bias*, etc. En los modelos de CNN, los métodos de aprendizaje basados en gradientes son una elección natural (Ghosh et al., 2020). Para reducir el error, los parámetros del modelo se actualizan continuamente durante cada época de entrenamiento y el modelo busca iterativamente la solución localmente óptima en cada época.

El tamaño de los pasos de actualización se denomina tasa de aprendizaje, y una iteración completa de actualización de parámetros, que incluye todo el conjunto de datos de

entrenamiento, al realizarse una vez, se denomina época. Aunque la tasa de aprendizaje es un hiperparámetro, debemos elegirlo con cuidado, de manera que no afecte de forma negativa el proceso de aprendizaje. En las figuras 2.16, 2.17 y 2.18 podemos ver los distintos casos que se pueden presentar al momento de elegir una tasa de aprendizaje.

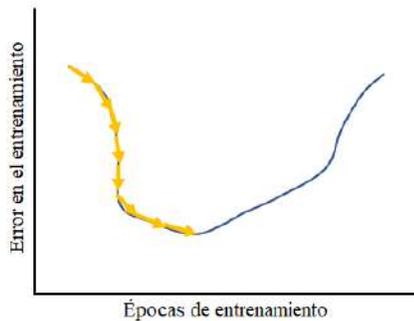


Figura 2.16: Tasa de aprendizaje pequeño

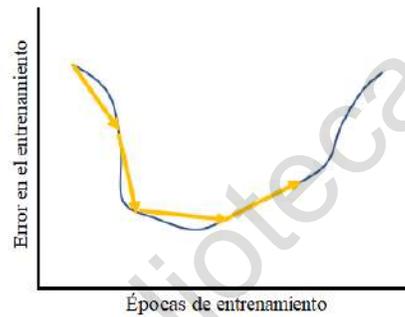


Figura 2.17: Tasa de aprendizaje grande



Figura 2.18: Buena elección de tasa de aprendizaje

2.6.1 Descenso de gradiente

El algoritmo de descenso de gradiente, actualiza los parámetros del modelo continuamente durante cada época para reducir el error de entrenamiento. Para actualizar esos parámetros de manera correcta, primero calcula la pendiente de la función objetivo usando la derivada de primer orden con respecto a los parámetros del modelo, después, para minimizar el error, actualiza el parámetro en la dirección opuesta a la pendiente (Hochreiter, Younger y Conwell, 2001). Este proceso de actualización de parámetros se realiza durante la

retropropagación del modelo, donde el gradiente en cada neurona se propaga nuevamente a todas las neuronas que pertenecen a su capa anterior.

2.6.2 Descenso de gradiente estocástico

En este método, los parámetros se actualizan para cada muestra de entrenamiento por separado (Bottou, 2010). Aquí se recomienda mezclar aleatoriamente las muestras de entrenamiento en cada época antes de realizarlo. El beneficio de usarlo, en lugar del descenso de gradientes, es que converge mucho más rápido en el caso de un gran conjunto de datos de entrenamiento y también es eficiente en memoria. Pero el problema es que, debido a las actualizaciones frecuentes, se toman pasos muy ruidosos hacia la solución que hacen que el comportamiento de la convergencia sea muy inestable.

2.6.3 Momentum

Momentum es una técnica utilizada en la función objetivo de las Redes Neuronales, mejora tanto la velocidad como la precisión del entrenamiento al agregar el gradiente calculado en el paso anterior ponderado por un parámetro llamado factor de *momentum* (Kanigoro y Salman, 2016). El principal problema del algoritmo de aprendizaje basado en gradientes es que se atasca fácilmente en un mínimo local, en lugar de un mínimo global, esto ocurre principalmente cuando el problema tiene un espacio de solución no convexo.

2.6.4 AdaGrad

AdaGrad, también conocido como método de tasa de aprendizaje adaptativo, actualiza cada parámetro de red de manera diferente en función de su importancia para el problema, es decir, realiza actualizaciones más grandes para parámetros poco frecuentes, y actualizaciones más pequeñas para parámetros más frecuentes. Se hace dividiendo la tasa de aprendizaje de cada parámetro con la suma del cuadrado de todos los gradientes pasados para cada parámetro en cada época de entrenamiento (Duchi, Hazan y Singer, 2011). En la práctica, AdaGrad es muy útil, especialmente en el caso de gradientes escasos, o cuando tenemos pocos datos de entrenamiento para Redes Neuronales a gran escala.

2.6.5 RMSProp

La propagación de raíz cuadrada media (RMSProp) también está diseñada para resolver el problema de tasas de aprendizaje radicalmente decrecientes. Fue desarrollado por el grupo de Geoffrey Hinton, intenta resolver este problema usando una media móvil sobre el gradiente cuadrado pasado (Ruder, 2016).

2.6.6 Estimación del Momento Adaptativo (Adam)

La Estimación del Momento Adaptativo (Da, 2014) es otra estrategia de aprendizaje, que calcula el factor de aprendizaje adaptativo para cada parámetro en la red, y combina las ventajas de *Momentum* y RMSprop manteniendo la media móvil exponencial de los gradientes, así como la media móvil exponencial de los gradientes cuadrados. La razón por la que Adam es tan usado es que es más eficiente en memoria que otros y también necesita menos poder computacional.

2.7 Arquitectura U-Net para segmentación de imágenes

U-Net es una arquitectura de Red Convolutiva para segmentación de imágenes rápida y precisa. Fue introducida por (Ronneberger *et al.*, 2015). Su arquitectura se modificó y amplió para trabajar con menos imágenes de entrenamiento y producir una segmentación más precisa.

La red U-Net consta de una ruta de contracción y una ruta expansiva, lo que le da la arquitectura en forma de U. La ruta de contracción es una red convolutiva típica que consiste en la aplicación repetida de convoluciones, cada una seguida de una unidad lineal rectificadora (ReLU) y una operación de agrupación máxima. Durante la contracción, la información espacial se reduce, mientras que la información de características aumenta. La vía expansiva, combina la característica y la información espacial a través de una secuencia de convoluciones ascendentes y concatenaciones con características de alta resolución.

Esta red se introdujo originalmente como una red neuronal totalmente convolucional, en la que los mapas de características se concatenan en profundidad a capas que se muestrean desde las capas inferiores. En términos simples, esta red recibe como entrada una imagen RGB y genera un mapa de probabilidad. La arquitectura de esta red se puede observar en la figura 2.19.

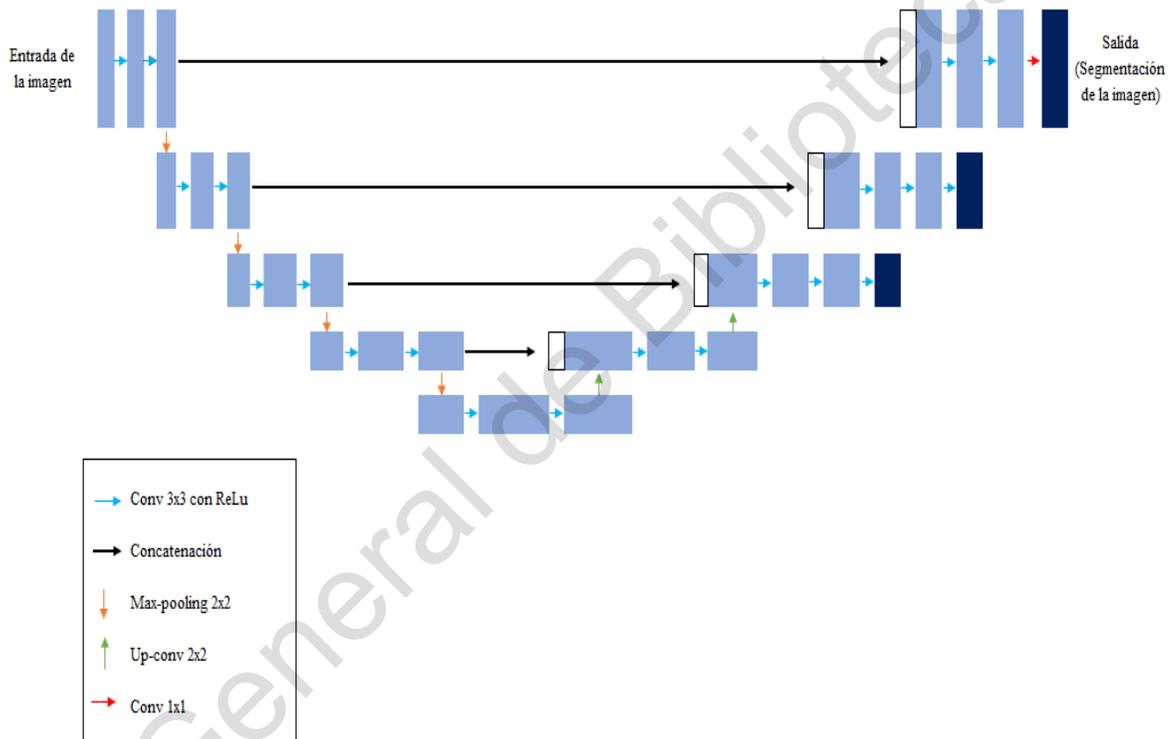


Figura 2.19: Arquitectura U-Net.

2.8 Bloques residuales

El bloque residual se utiliza para construir la propia red. Es una colección de convoluciones y capas agrupadas que conduce a una macro arquitectura. La descripción de este bloque y de toda la arquitectura fue publicada en el 2016 (He, Zhang, Ren y Sun, 2016). En este trabajo, demostraron que las redes extremadamente profundas se pueden entrenar

utilizando el optimizador estándar de descenso de gradiente estocástico (SGD) mediante el uso de bloques residuales.

Al utilizar la arquitectura clásica, asumimos que una capa solo necesita conocer la información generada por la capa anterior (García, 2018). Las conexiones residuales permiten acceder a toda la información generada por capas anteriores, incluidas las entradas. Un ejemplo de cómo funciona dichos bloques se muestra en la figura 2.20.

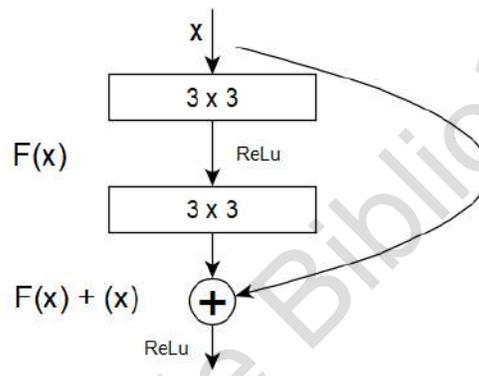


Figura 2.20 Bloque residual.

2.9 Visualización para Redes Neuronales Convolucionales

Un aspecto importante de las CNN es obtener características abstractas cuando la entrada se propaga hacia las capas más profundas (Albawi *et al.*, 2017). Por ejemplo, en la clasificación de imágenes, el borde podría detectarse en las primeras capas, después las formas más simples en las segundas, y luego las características de nivel superior en las capas posteriores. Sin embargo, sin un método complementario, las características que se identifican en cada capa son difíciles de localizar, esto podría causar dudas a la hora de interpretar algún resultado obtenido de la implementación de alguna arquitectura.

Para complementar los resultados obtenidos, se pueden utilizar métodos de visualización de neuronas, esto con el objetivo de darle valor agregado a los resultados y poder conocer cómo se está comportando la red internamente. Esto se podrá lograr

representando visualmente cuáles son las características en la imagen que se están aprendiendo.

Existen diferentes métodos para poder visualizar las neuronas en una CNN. El objetivo de todos ellos es representar cuáles son las neuronas que están teniendo mayor peso a la hora de tomar una decisión, de esta manera será posible identificar las características que provocan mayor activación en la imagen.

Ser capaz de visualizar las neuronas con mayor peso en la red nos da soporte y fundamentación de la predicción realizada por la arquitectura propuesta. También nos asegura que la red está aprendiendo de manera correcta y no está enfocándose en características erróneas. De igual manera, esta visualización nos podría ayudar a ajustar parámetros de la arquitectura con el fin de enfocarnos en las peculiaridades buscadas.

En imágenes médicas, los expertos tendrían mayor confianza de apoyarse en estos algoritmos, si cuentan con una representación gráfica que de más peso a la decisión realizada por la red.

Seleccionar bien que método se utilizará para la visualización también es muy importante, ya que no todos se adaptan a una arquitectura en concreto. Existen modelos que solo trabajan bien con redes muy específicas.

Existen diferentes métodos para la visualización de neuronas. Uno de los primeros métodos utiliza lo que se conoce como Maximización de Activación (Erhan, Bengio, Courville y Vincent, 2009), lo que propone es producir una representación que sea capaz de maximizar la activación de una de las neuronas que estamos tratando de observar. Esto nos indicará cuál es la particularidad que ha sido capaz de aprender. Aunque este método fue propuesto desde el 2009, la primera vez en ser utilizado en Redes Neuronales Convolucionales fue hasta 2013 en el trabajo que utiliza mapas de prominencia (Simonyan, Vedaldi y Zisserman, 2013).

Otro método muy conocido es el que utiliza Redes Neuronales Desconvolucionales (Zeiler, Taylor y Fergus, 2011), que básicamente lo que hace es detectar que patrones son los responsables de activar una neurona en concreto dentro de la CNN.

Uno de los métodos más populares y que se adapta bien a diferentes tipos de arquitecturas es el conocido como GradCAM (Selvaraju, Cogswell, Das, Vedantam, Parikh y Batra, 2017), que basa su funcionamiento en el procedimiento CAM (Zhou, Khosla, Lapedriza, Oliva y Torralba, 2016), pero hace una adaptación debido a que el método CAM tiene la desventaja que no se puede adaptar a cualquier tipo de red. GradCAM lo que propone es usar gradientes para evitar la dependencia de la arquitectura, y de este modo dar lugar a poder combinar los mapas de activación. Debido a lo amigable que puede llegar a ser este método nos da la posibilidad de visualizar cualquier neurona dentro de la red utilizada, no solo las neuronas en la última capa.

2.10 Síntesis de imágenes médicas usando CNN

Las CNN surgieron como una herramienta poderosa para la clasificación de imágenes y la segmentación semántica desde la competencia ImageNet en 2012, en la que el objetivo principal era estimar el contenido de imágenes naturales (Russakovsky, Deng, Su, Krause, Satheesh y Fei-Fei, 2015). Su éxito se debe al uso de GPU, rectificadores como ReLU, técnicas de aumento de datos, y nuevas técnicas de regularización como *Dropout* (Srivastava *et al.*, 2014). El poder principal de las arquitecturas de CNN se basa en su capacidad para extraer automáticamente características altamente discriminatorias en múltiples niveles de abstracción.

Las CNN están diseñadas para procesar datos que vienen en formato sin procesar, como las imágenes. Sin embargo, entrenar a una CNN desde cero no es una tarea fácil (Jiang, Duan, Cheng, Gu, Fu y Liu, 2019). Requieren una gran cantidad de datos etiquetados, un requisito difícil de cumplir en varias enfermedades oculares, además de necesitar de altos recursos computacionales.

Existen dos alternativas para entrenar una CNN desde cero que se han aplicado previamente a varias tareas de clasificación de imágenes médicas. La primera alternativa consiste en ajustar una CNN que ha sido entrenada utilizando un gran conjunto de datos etiquetados de una aplicación diferente. Un ejemplo de esta alternativa es el trabajo donde

demostraron que los modelos de CNN que fueron entrenados previamente en imágenes naturales (Carneiro, Nascimento y Bradley, 2015), como ImageNet, son útiles en aplicaciones de imágenes médicas, a pesar de las diferencias significativas en la apariencia de la imagen. Otro estudio realizado demostró que el uso de una CNN pre-entrenada y afinada para localizar planos estándar en imágenes de ultrasonido supera el estado del arte para el plano estándar abdominal fetal (Chen, Ni, Qin, Li, Yang, Wang y Heng, 2015). Otro ejemplo es el estudio en el que llevaron a cabo una serie de experimentos para cuatro aplicaciones de imágenes médicas que muestran el uso de una CNN pre-entrenada así como un CNN entrenada desde cero (Tajbakhsh, Shin, Gurudu, Hurst, Kendall, Gotway y Liang, 2016).

La segunda alternativa consiste en utilizar una CNN entrenada por ImageNet como extractor de características, donde se alimenta a la CNN con una imagen de entrada, y luego las características se extraen de una determinada capa oculta de la red. De esa manera, las características extraídas se utilizan para entrenar un nuevo clasificador, como máquinas de soporte de vectores (SVM), árboles de decisión, o vecino más cercano. (Bar, Diamant, Wolf y Greenspan, 2015) utilizan una CNN previamente entrenada que usaron como extractor de características para la identificación de patologías torácicas.

Metodología

En esta sección se mostrará la metodología seguida para realizar la clasificación a nivel imagen y la clasificación a nivel píxel. Básicamente ambas metodologías cuentan con la adquisición de imágenes de fondo de ojo, un preprocesamiento realizado a las imágenes, definir cuál será la arquitectura que tendrá la red neuronal, validación de resultados, e interpretación de ellos.

3.1 Adquisición de imágenes

Para poder desarrollar este trabajo se utilizará la base de datos pública ORIGA tanto para la clasificación a nivel píxel, como para la clasificación a nivel imagen.

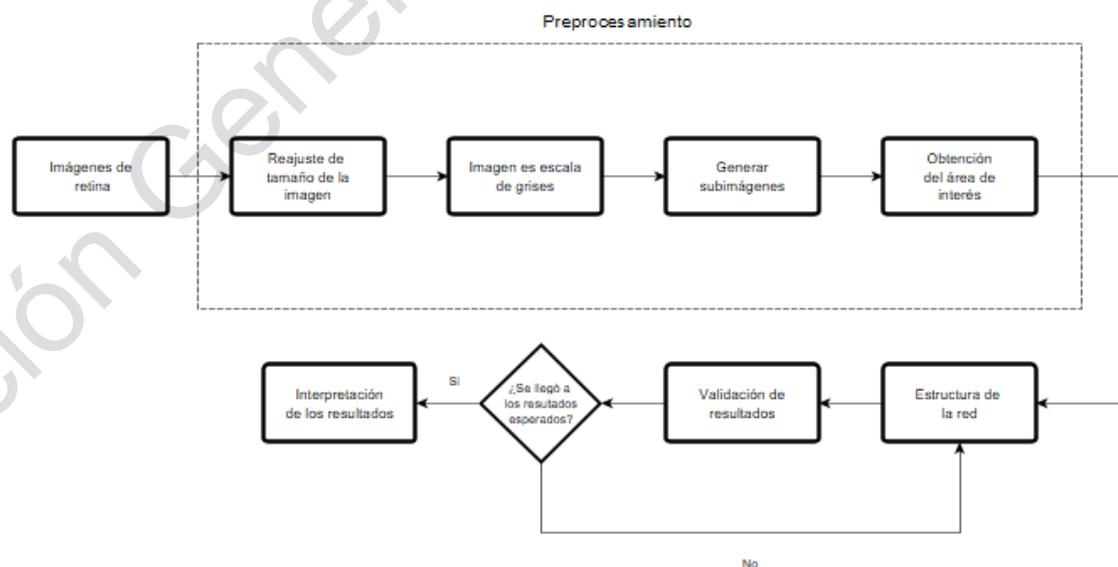


Figura 3.1: Metodología propuesta para la clasificación a nivel imagen.

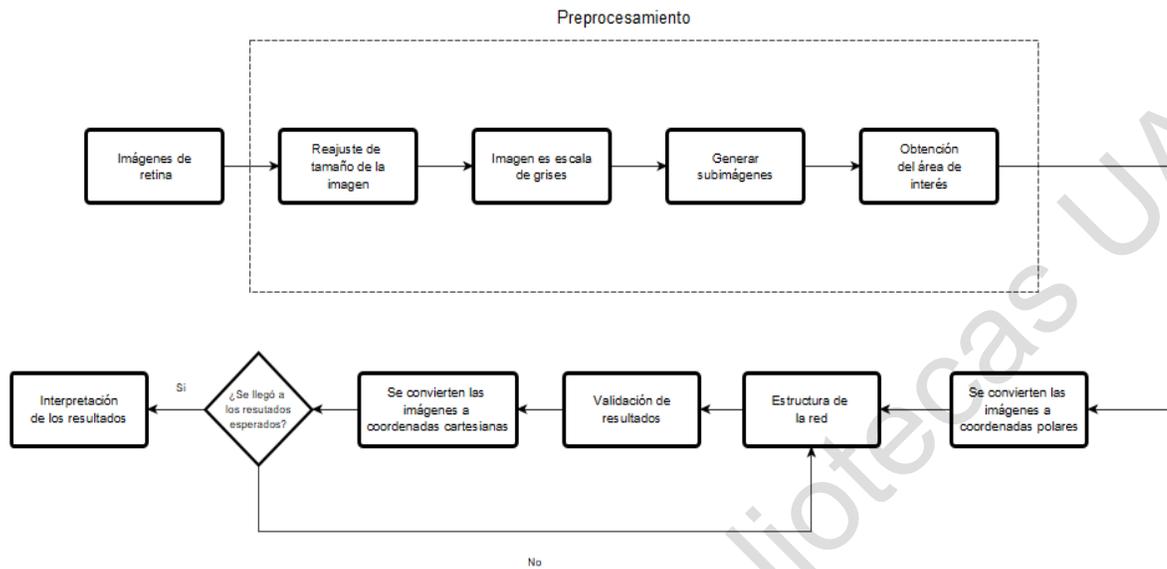


Figura 3.2: Metodología propuesta para la clasificación a nivel píxel.

3.1.1 Base de datos ORIGA

Hay relativamente pocos conjuntos de datos públicos para la evaluación del glaucoma en comparación con los conjuntos de datos disponibles para la retinopatía diabética (Decencièrè, Zhang, Cazuguel, Lay, Cochener, Trone & Klein, 2014) y la segmentación vascular (Brandon & Hoover, 2003). La base de datos ORIGA (Zhang et al., 2010) contiene 650 imágenes de retina anotadas por profesionales capacitados del Instituto de Investigación Ocular de Singapur. En dicha base de datos se observa una amplia colección de signos críticos para el diagnóstico de glaucoma. 482 imágenes corresponden a pacientes sanos y 168 imágenes a pacientes con glaucoma con una resolución de 3072 x 2048 píxeles. Esta base de datos se utiliza ampliamente como punto de referencia para métodos de clasificación de glaucoma.

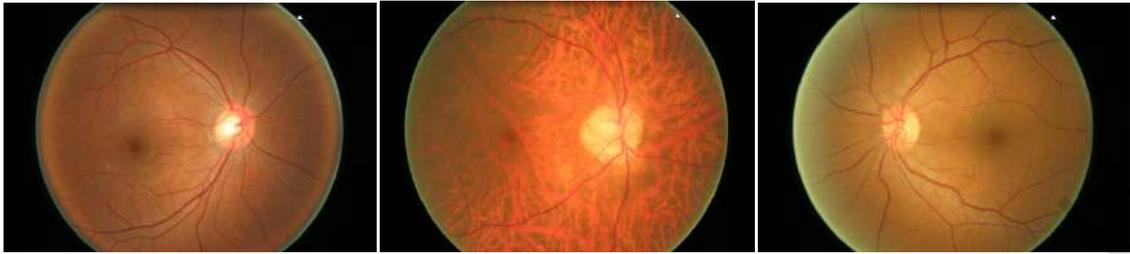


Figura 3.3: Ejemplo de imágenes de pacientes sanos de la base de datos ORIGA.

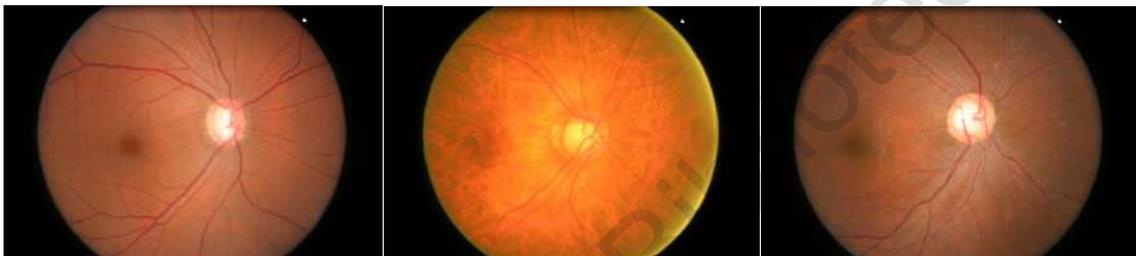


Figura 3.4: Ejemplo de imágenes de pacientes con glaucoma de la base de datos ORIGA.

3.2 Recursos computacionales

Todo el trabajo aquí expuesto se desarrolló en Python 3.0 y se ejecutó en una computadora con una CPU Intel Core i7-3160QM a 2,3 GHz y 8 GB de RAM.

3.3 Preprocesamiento de las imágenes

Todos los métodos diseñados para el análisis de imágenes retinianas, u otros objetos, constan de tres etapas principales: preprocesamiento, procesamiento y posprocesamiento. La etapa de preprocesamiento se considera preliminar, ya que influye en la calidad de las etapas posteriores y afecta el rendimiento general.

Esta sección se presenta un método para localizar el disco óptico de la retina, ya que en esta parte de la imagen se encuentran las características necesarias para diagnosticar

el glaucoma. Esta región de interés (ROI) será la entrada, tanto para la arquitectura propuesta para la clasificación a nivel imagen, como para la propuesta a nivel píxel.

Dado que el tamaño de las imágenes es de 3072 x 2048 píxeles, se reduce a cuatro veces su tamaño original, dejando las dimensiones de las imágenes en 768 x 512.

Las imágenes de entrada se convierten a escala de grises. De esta forma, es posible obtener un mayor contraste del disco óptico en comparación con la imagen original. Para ello se utilizan los canales rojo y verde, que son los que más impacto tienen en el disco óptico. Para hacer esto, se aplica la ecuación 3.1.

$$Img_{gray} = R * 0.9 + G * 0.5 \quad (3.1)$$

Donde Img_{gray} es la imagen en escala de grises, R y G corresponden a los canales rojo y verde de la imagen.

El siguiente paso es desplazar un filtro a través de la imagen para dividirlo en diferentes subimágenes. Lo que se busca es identificar el disco óptico, que sería la parte más brillante de las imágenes de retina, pero también se quiere evitar confundirlo con el exceso de brillo que se encuentra en otras partes.

Debido a que el disco óptico representa aproximadamente el 10% de la imagen original, se eligió el tamaño de filtro para garantizar que en alguna de las subimágenes estará localizado. El tamaño seleccionado fue de 192 x 192 píxeles, cuatro veces más pequeño que el ancho de la imagen después de la reducción de tamaño.

Para reducir el tiempo de procesamiento, el desplazamiento de filtro es 165 píxeles en la dirección horizontal y 150 en la dirección vertical. Con este paso, garantizamos que el disco óptico está en una de las subimágenes al menos una vez.

Se aplica la ecuación 3.2 para calcular las subimágenes, según el tamaño del cuadro de desplazamiento y el paso utilizado.

$$SubImg(x, y) = Img(x + x_0, y + y_0) \quad \forall \in R \quad (3.2)$$

Donde *SubImg* es la imagen dentro del filtro, *Img* es la imagen después de cambiar su tamaño, *x* es el punto de inicio horizontal del filtro, *y* el punto de inicio vertical, *x*₀ el punto final horizontal, es decir, *x* + 192, y *y*₀ el punto final vertical, es decir, *y* + 192.

Cada vez que se selecciona una nueva subimagen, se calcula el promedio de todos los píxeles aplicando la ecuación 3.3. Este promedio se almacena, y una vez analizadas todas las subimágenes, se ubica donde se encuentra el promedio más alto. De esta forma se elige la subimagen donde se predice está el disco óptico.

$$promPix = \frac{\sum_{i=1}^n p_i}{n} \quad (3.3)$$

Donde *promPix* es el promedio del valor de los píxeles, *p_i* es el valor actual del píxel, y *n* el número total de píxeles.

Con esta nueva subimagen seleccionada, se toma el valor promedio de los píxeles, pero ahora de cada columna y fila. Este promedio nos dará unas coordenadas (*x*, *y*). Éstas representan el nuevo centro de la imagen, que será recortada para obtener el ROI. Un ejemplo de esto se puede encontrar en la figura 3.5.

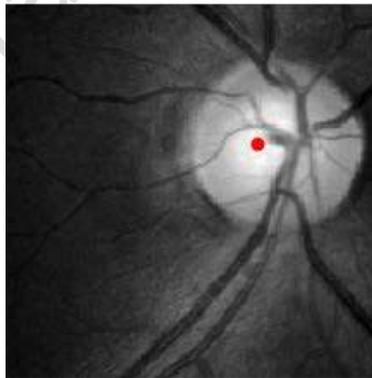


Figura 3.5: Centro de la imagen que será recortada.

Las coordenadas obtenidas con el centro localizado se escalan a la imagen original y se cortan a un tamaño de 800 x 800 píxeles, suficiente para cubrir el disco óptico en la nueva imagen. El resultado después de recortar una imagen se puede encontrar en la figura 3.6.

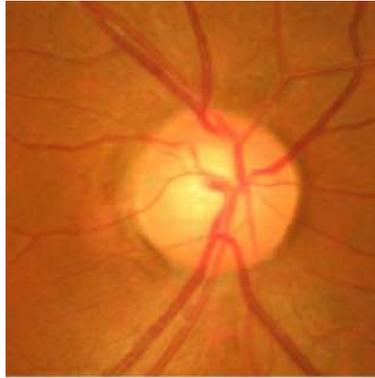


Figura 3.6: Resultado después de recortar la imagen y obtener el ROI.

Algoritmo 3.1: Región de interés

Entrada: Número de imágenes *Imágenes*

Output: Región de interés *ROI*

```
1: while Imágenes is not empty
2:   Imágenes = Imágenes/4
3:   Imágenes = (Canal rojo * 0.9) + (Canal verde * 0.5)
4:   filtro = 192x192
5:   if filtro ∈ Imágenes then
6:     valor promedio = valor promedio del filtro
7:     if valor promedio > promedio anterior
8:       promedio anterior = valor promedio
9:       Imágenes = filtro
10:    end
11:  end
12:  valMax = where (valMaxPix = Imágenes)
13:  x = mean(posValMaxX)
14:  y = mean(posValMaxY)
15:  ROI = centro(x, y)
16: end
```

El preprocesamiento descrito anteriormente se aplica a las 650 imágenes de la base de datos ORIGA para obtener el ROI.

Ya con la región de interés localizada, se procede a realizar una normalización a las imágenes, en este punto se encuentran entre valores de 0 y 255. Lo que se busca es que los valores se encuentren entre un rango más pequeño, ya que las Redes Neuronales Convolucionales no tienen un buen desempeño cuando los atributos numéricos a la entrada tienen un rango muy grande. La normalización que mejor desempeño tuvo fue la min-max (Mining, 2006).

$$valNorm = \frac{val - min_{img}}{max_{img} - min_{img}} \quad (3.4)$$

Donde $valNorm$ es el valor normalizado, val es el valor actual, min_{img} y max_{img} son los valores de imagen mínimo y máximo. Los valores después de la normalización, estarán en un intervalo de 0 a 1.

3.4 Clasificación a nivel imagen

Se propone una arquitectura que se alimenta con imágenes escaladas de un tamaño de 128 x 128 píxeles, las imágenes originales tienen un tamaño de 800 x 800, pero se redujeron para disminuir el tiempo de procesamiento. La red propuesta está formada por capas de convolución, funciones de activación ReLu, capas de agrupación máxima, y capas completamente conectadas. La salida de cada capa es la entrada de otra, lo que permite la extracción de características. Dado que se busca diferenciar entre características de imagen pequeñas y locales, que pueden diferenciar las características distintivas de una persona con glaucoma, se utilizan filtros pequeños de 3x3 y 5x5 en la red. Las dos últimas capas unirán estas características para hacer una clasificación.

La red neuronal propuesta consta de quince capas. Debemos tener en cuenta que, con una red menos profunda, la extracción de características puede ser insuficiente. Sin embargo, con una arquitectura más profunda, existe el riesgo de sobreajuste. Las capas convolucionales

representan características de alto nivel en los datos, y agregar capas completamente conectadas es una forma de aprender combinaciones no lineales de esas características. En la red se utilizan dos capas completamente conectadas, buscando un aprendizaje más significativo de estas combinaciones.

La arquitectura de la red propuesta se puede observar en la figura 3.7.

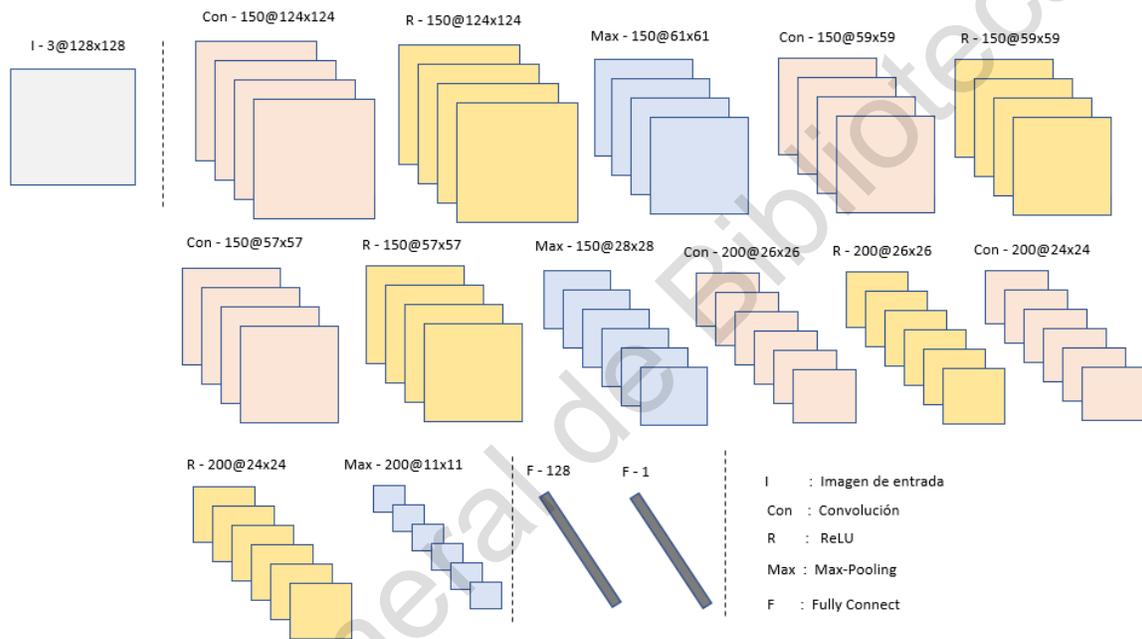


Figura 3.7: Arquitectura propuesta.

La validación cruzada es uno de los métodos más utilizados para la evaluación del desempeño de un modelo predictivo. Básicamente, según la división de datos, una parte de los datos se usa para entrenamiento y el resto de los datos se usa para pruebas (Berrar, 2019). Una versión de validación cruzada es el conocido como *k-fold*. En esta versión, los datos se dividen aleatoriamente en *k* subconjuntos de igual tamaño. *k* - 1 subconjuntos se utilizan para el entrenamiento y el subconjunto restante se utiliza para la prueba. El error de predicción promedio de cada división se obtiene para estimar el rendimiento del modelo (Zhang y Yang, 2015).

Basándose en lo antes mencionado, y para asegurar que los resultados obtenidos fueran correctos, se implementaron las pruebas con validación cruzada. Para esta validación el algoritmo se ejecutó diez veces para un *k-fold* con un valor de $k = 10$. Cada una de las ejecuciones tuvo 50 épocas. La ventaja de utilizar este método es que le da a nuestro modelo la oportunidad de realizar múltiples particiones de prueba y de entrenamiento, y con esto, mejorar el funcionamiento con datos no antes vistos.

De la base de datos utilizada, que cuenta con 650 imágenes, fueron utilizadas 59 para la etapa de validación, es decir, el 10% de las imágenes totales. Para la etapa de entrenamiento se utilizaron las 591 imágenes restantes. Las imágenes fueron elegidas de manera aleatoria para evitar cualquier sesgo en los resultados.

Como se sabe, no hay una manera analítica de calcular el mejor valor para la tasa de aprendizaje en una determinada arquitectura. Un enfoque alternativo es realizar un análisis de sensibilidad de las tasas de aprendizaje para el modelo elegido. Esto puede ayudar tanto a resaltar un orden de magnitud donde pueden residir buenas tasas de aprendizaje, como a describir la relación entre la tasa de aprendizaje y el rendimiento. También es buena idea revisar en la literatura cuáles son las mayormente utilizadas en este tipo de trabajos. Tomando en cuenta lo antes citado, en este trabajo se usaron 3 tasas de aprendizaje diferentes: 0.01, 0.001 y 0.0001.

Como función de activación a la salida de la red, se utilizó la función Sigmoidal. Se eligió esta función debido a que lo que nos interesa es que haga una clasificación binaria, por lo que esta función nos dará una probabilidad de a qué clase puede pertenecer. No hay que confundir esta función de activación con la usada entre cada capa convolucional, una es para agregar no linealidad a la red como lo es la ReLu, y otra para regularizar la salida como la Sigmoidal.

Como función de pérdida utilizamos Entropía Cruzada Binaria. Esto debió a que, como ya se mencionó, se busca realizar una clasificación binaria, esta función nos dará el error de predicción, el cual le indicará a la red que tan acertada esta su clasificación, lo que ayudará a la red a mejorar conforme pasen el número de épocas.

Para evitar el sobreajuste, se utilizó la técnica de *Dropout*, la cual va apagando aleatoriamente neuronas en cada época del entrenamiento, esto con el fin de distribuir el poder de selección de características a todas las neuronas por igual, de esta forma, se obliga al modelo a aprender características independientes.

Como optimizador se utilizó Estimación de Momento Adaptivo (Adam), este optimizador combina las ventajas de algunos otros, como lo es el de *Momentum*, manteniendo la media móvil exponencial de los gradientes, y el RMSprop, que mantiene la media móvil exponencial, pero de los gradientes al cuadrado. Además de necesitar menor poder computacional que algunos otros y se adecua bien a problemas con un gran número de parámetros.

3.5 Clasificación a nivel píxel

Muchos estudios han aplicado el aprendizaje profundo a la segmentación de la imagen del fondo de ojo. Sin embargo, algunos utilizan conjuntos de datos con muy pocas imágenes, lo que dificulta la posibilidad de generalizar a diferentes conjuntos de datos. Otro factor importante a tomar en cuenta es que se necesitan bases de datos que ya contengan las imágenes segmentadas, esto es difícil, ya que es complicado conseguir imágenes con estas características. Muchos trabajos encontrados para la segmentación del disco y la copa utilizan bases de datos privadas.

3.5.1 Segmentación de las imágenes utilizadas

Para solucionar con el problema de no contar con una base de datos previamente segmentada se hizo una estancia, de mes y medio aproximadamente, en el Instituto Mexicano de Oftalmología. Ahí, junto con médicos especializados en el área de glaucoma, se creó una base de datos propia, segmentando algunas de las imágenes contenidas en la base de datos pública ORIGA. Aún después de la estancia presencial, se siguió trabajando con ellos para segmentar las imágenes necesarias para el entrenamiento de la red.

Durante este tiempo se lograron segmentar 100 imágenes del disco óptico, y 50 de la copa. Las mismas que se utilizaron para entrenar la red propuesta en esta sección.

La segmentación se realizó utilizando el software ImageJ. Este software nos permite segmentar imágenes de cualquier tipo. El proceso, aunque muy tardado y tedioso, no es tan complicado. Se van colocando puntos en el contorno del disco o copa óptica para después segmentar el área seleccionada. El tiempo aproximado de segmentar un área en una imagen era aproximadamente de 30 minutos por imagen. Este proceso se puede observar en la figura 3.8.

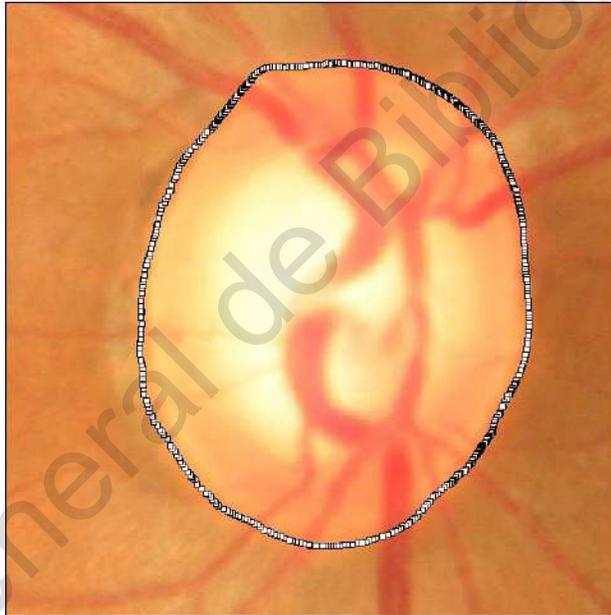


Figura 3.8: Contorno del disco óptico después de colocar los puntos para segmentar.

Se realizaba un zoom del 2400% para ir colocando cada punto e ir indicando el contorno del área deseada. Después de la segmentación la imagen quedaba como se observa en la figura 3.9.

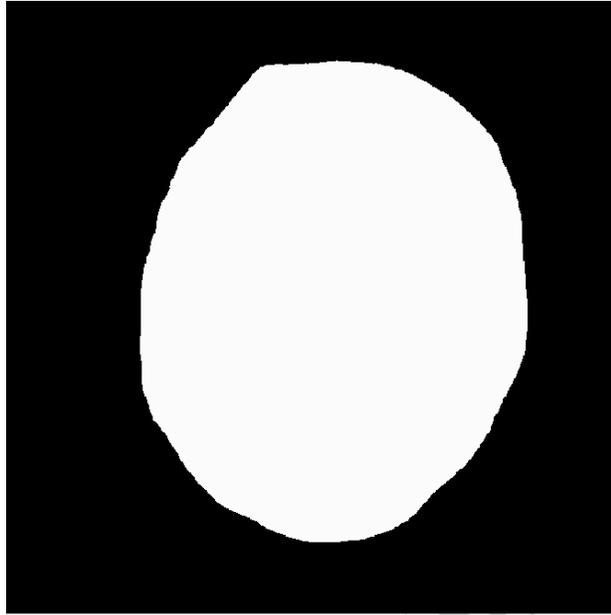


Figura 3.9: Disco óptico segmentado.

Este proceso se realizaba también para la copa óptica, en las figuras 3.10 y 3.11 podemos ver tanto la imagen antes de ser segmentada como el resultado.

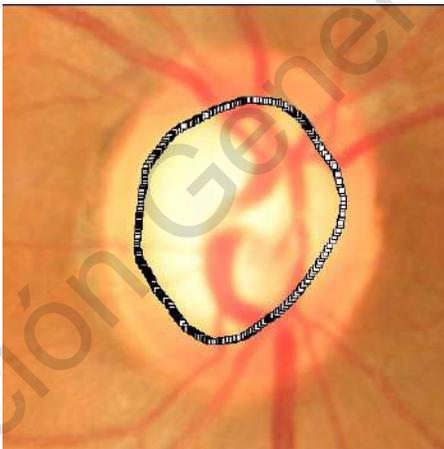


Figura 3.10: Copa óptica marcada.

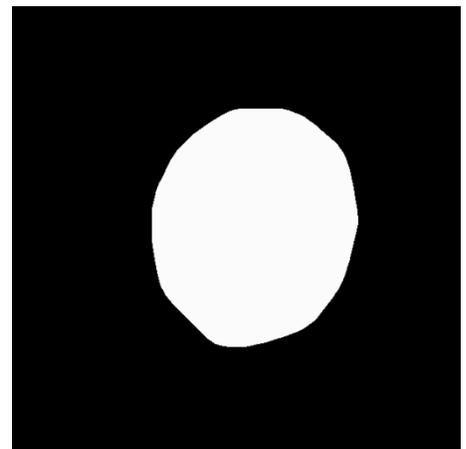
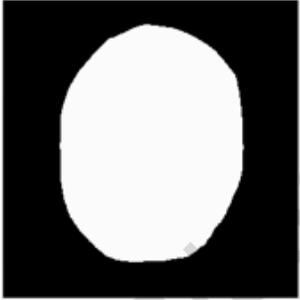
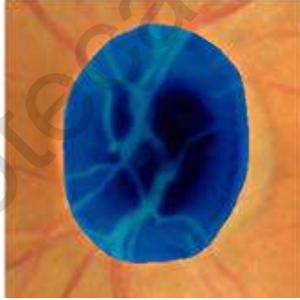
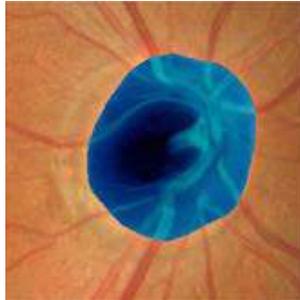
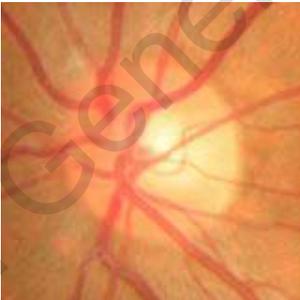
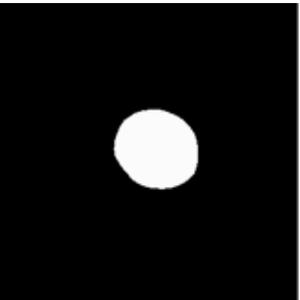
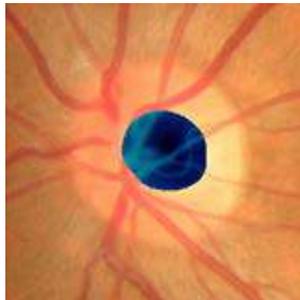
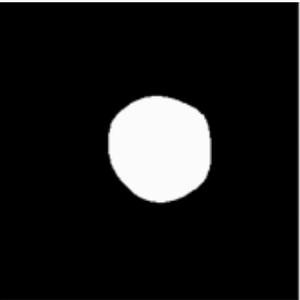
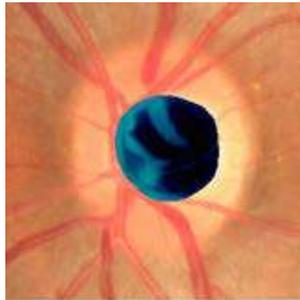


Figura 3.11: Copa óptica segmentada.

Algunas imágenes originales comparadas con su contraparte segmentada se pueden observar en la tabla 3.1.

Tabla 3.1: Comparación entre la imagen original y la imagen segmentada

	Imagen original	Imagen segmentada	Comparación de imágenes
Disco óptico			
			
Copa óptica			
			

3.5.2 Transformación de imágenes a coordenadas polares.

Para considerar el movimiento linealmente en el espacio de coordenadas polares, se necesita un proceso de transformación. Puede expresarse mediante la interpolación bilineal de los píxeles de la imagen en el nuevo dominio de los píxeles de la imagen de entrada. Los píxeles de la imagen operada deben estar en la trayectoria del movimiento de rotación (Morimoto, Isoe, Ohizumi, Al Maki y Sugimoto, 2011). Si se mantiene el radio y se transforma la coordenada polar de la imagen de entrada mediante el uso de interpolación bilineal, es posible introducir la operación invariante de desplazamiento en la imagen de entrada a lo largo del eje θ de la coordenada polar. Una representación gráfica de dicha conversión se puede apreciar en la figura 3.12.

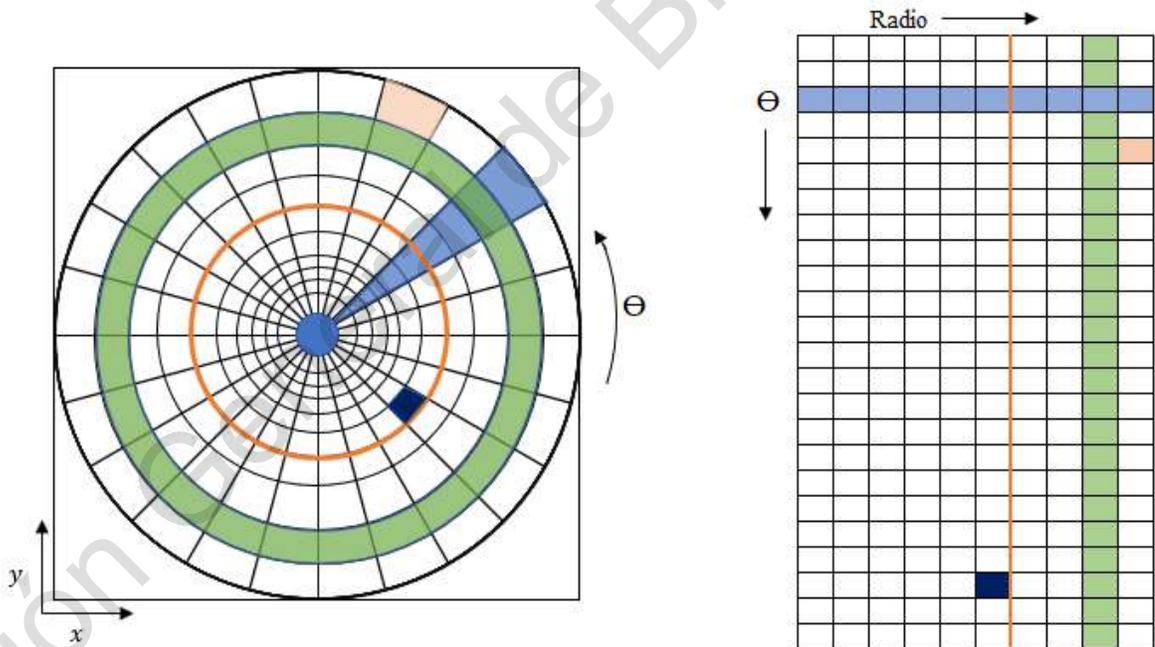


Figura 3.12: Representación de cambio de coordenadas.

Para alimentar nuestra arquitectura, la imagen se convierte a coordenadas polares, esto con el fin de aumentar la proporción que ocupan la copa y el disco, ya que, antes de la transformación, solo ocupaban el 5% y el 13% respectivamente. Después de la transformación, ocuparon, en promedio, el 21% y el 66% respectivamente. Equilibrar esta proporción evita un ajuste excesivo en el entrenamiento (Ronneberger *et al.*, 2015). Las ecuaciones utilizadas para convertir las imágenes a coordenadas polares se muestran en las ecuaciones 3.5 y 3.6.

$$u = u_0 + r * \cos(\theta + \varphi) \quad (3.5)$$

$$v = v_0 + r * \sin(\theta + \varphi) \quad (3.6)$$

Una vez que la red termina todo el procesamiento, entregará una imagen en coordenadas polares, debido a esto es necesario convertirla ahora de coordenadas polares a cartesianas. Para hacer esto se utilizan las ecuaciones 3.7 y 3.8.

$$r = \sqrt{(u - u_0)^2 + (v - v_0)^2} \quad (3.7)$$

$$\theta = \tan^{-1} \left(\frac{v - v_0}{u - u_0} \right) - \varphi \quad (3.8)$$

Donde $p(u, v)$ es el punto en el plano cartesiano, r y θ son el radio y el ángulo direccional, u_0 y v_0 son los puntos centrales del disco y φ es el ángulo polar.

La conversión de las imágenes de coordenadas cartesianas a polares se puede observar en la figura 3.13 y 3.14.

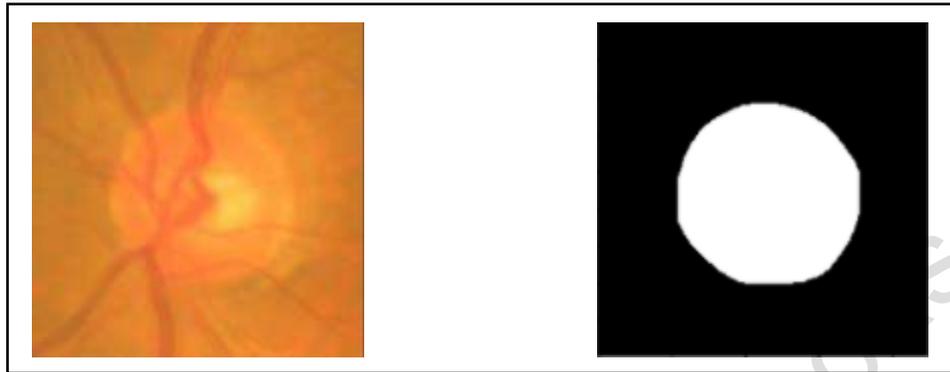


Figura 3.13: Imagen original y segmentada en coordenadas cartesianas.



Figura 3.14: Imagen original y segmentada en coordenadas polares.

3.5.3 Arquitectura propuesta.

La arquitectura utiliza una configuración de codificador-decodificador. Comienza con un módulo residual, luego se realiza una convolución de 3×3 , seguida de una unidad rectificadora (ReLU) y una operación de agrupación de 2×2 . Cada uno de los pasos del proceso de expansión muestra el mapa de características anterior. En la última capa se utiliza una convolución de 1×1 para que sea posible obtener una salida de 2 canales (He *et al.*, 2016), ya que, debido a la naturaleza de nuestro problema, la clasificación será binaria. La arquitectura se utiliza tanto para segmentar el disco óptico como para segmentar la copa.

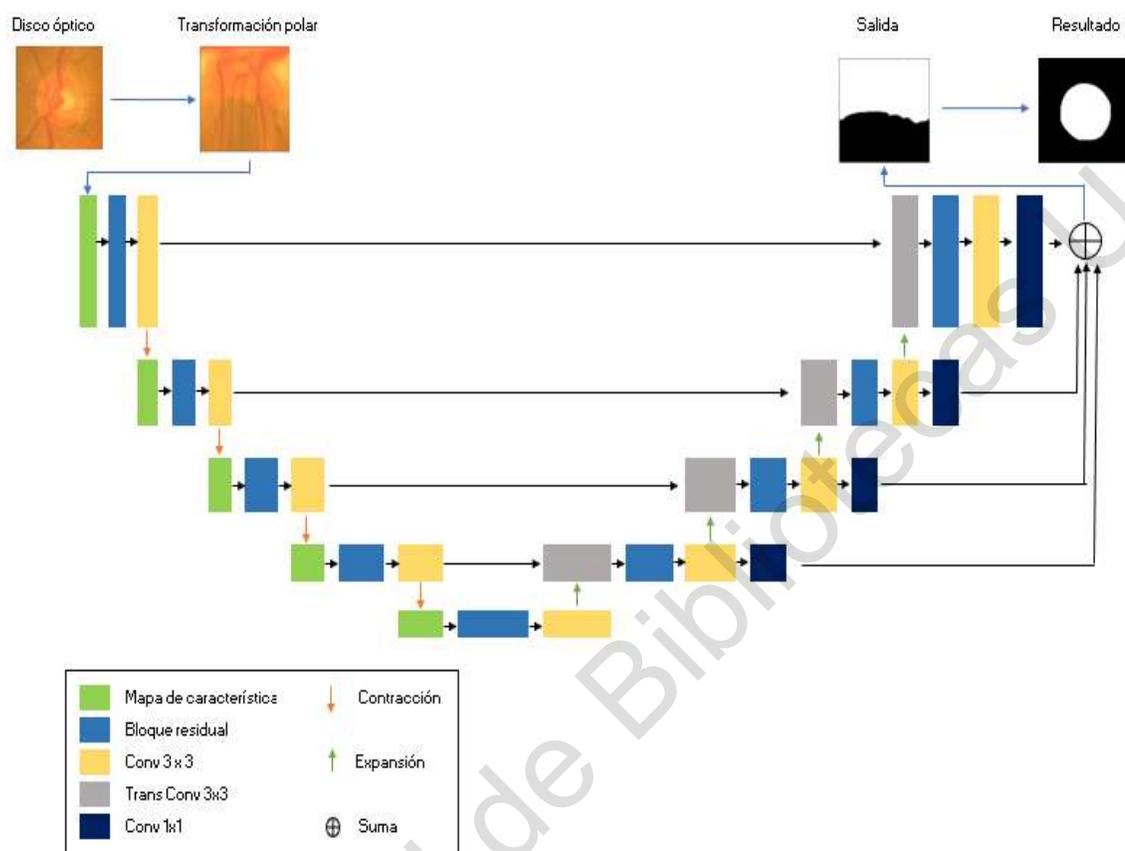


Figura 3.15: La arquitectura propuesta está compuesta por un U-Net con bloques residuales.

Se adoptó normalización por lote (*batch normalization*), justo después de cada convolución y antes de la activación, de igual manera como se realizó en el trabajo realizado por (Ioffe y Szegedy, 2015). La representación de estos bloques se puede apreciar en la figura 3.16.

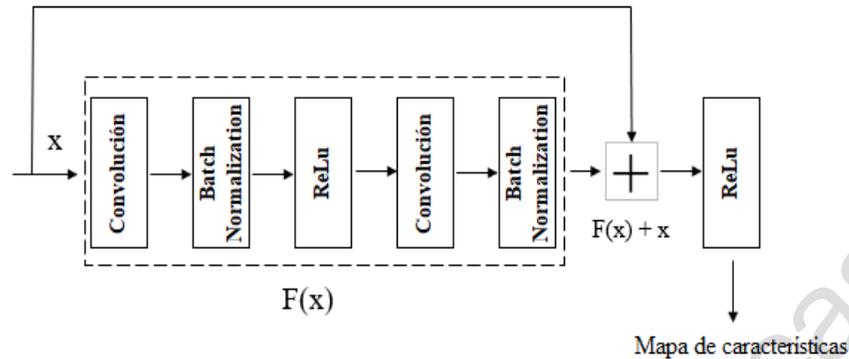


Figura 3.16: Bloque residual utilizado en este trabajo.

Del bloque residual mostrado $F(x) + x$ es el mapa de características que se espera tener como resultado, las dos operaciones de convolución con su correspondiente normalización son representadas por $F(x)$, y a dicha transformación, se le agrega la entrada original x .

Para la fase de entrenamiento se utilizaron 100 imágenes del disco óptico y 50 imágenes de la copa. Las imágenes originales se obtuvieron de la base de datos pública ORIGA. Estas imágenes fueron las mencionadas anteriormente, las cuales fueron segmentadas con ayuda de especialistas del Instituto Mexicano de Oftalmología.

Las pruebas se implementaron utilizando validación cruzada. Para esta validación el algoritmo se ejecutó un total de 5 veces y se utilizó k -fold con un valor de $k = 5$. Cada una de las ejecuciones tuvo 300 épocas con una parada anticipada de 25 iteraciones, con el objetivo de evitar un sobreajuste en el modelo.

De la base de datos utilizada, de la cual se cuenta con 100 imágenes segmentadas del disco óptico y 50 de la copa, se utilizaron 80% para la fase de entrenamiento y se dejó el 20% para las pruebas. Las imágenes fueron elegidas de manera aleatoria para evitar cualquier sesgo en los resultados.

La tasa de aprendizaje utilizada fue de 0.001, este fue con el que se tuvo mejores resultados. Como función de activación a la salida de la red se utilizó la función Sigmoidal, debido a que lo que se busca en el modelo es una clasificación binaria.

Como función de pérdida utilizamos Entropía Cruzada Binaria. Ya que lo que hará es una clasificación entre dos clases, esta función nos será útil para obtener el error de predicción y ajustar el modelo a que obtenga mejores resultados

Se utilizó la técnica de *Dropout*, la cual va apagando aleatoriamente neuronas en cada época del entrenamiento, ayudando al modelo aprender características independientes y para evitar el sobreajuste.

Como optimizador se utilizó el de Estimación de Momento Adaptivo (Adam), ya que combina las ventajas de otros optimizadores. Además de ser rápido en su ejecución, se adapta bien a las necesidades de la red.

Resultados y discusión.

En esta sección se mostrarán los resultados conseguidos en la obtención del área de interés, la clasificación a nivel imagen y la clasificación a nivel píxel. Se mostrarán las métricas utilizadas para evaluar los modelos propuestos y se realizará una comparación con los trabajos en el estado del arte.

4.1 Resultados para la región de interés.

Se observó que las imágenes de la base de datos disponible, afectan el desempeño de la localización del ROI cuando se realiza una transformación de la imagen a escala de grises, utilizando únicamente una proporción de los canales rojo y verde. Los coeficientes utilizados para calcular los valores de escala de grises de manera convencional son idénticos a los propuestos en la Recomendación UIT-R BT. 601-7 (Fernandez, Chavez y Kemper 2021). Las diferencias entre transformaciones se pueden observar en las figuras 4.1 y 4.2.



Figura 4.1: Transformación a escala de grises UIT-R BT. 601-7.

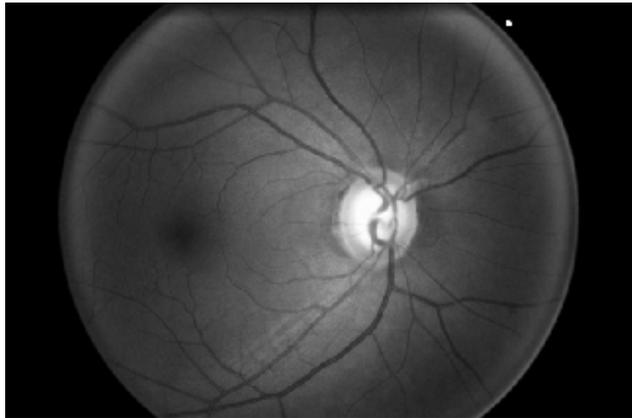


Figura 4.2: Transformación a escala de grises propuesta.

La comparación de los resultados utilizando estas 2 transformaciones de escala de grises se puede observar en la tabla 4.1.

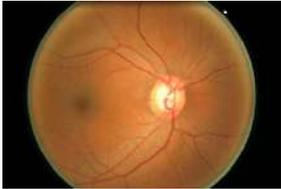
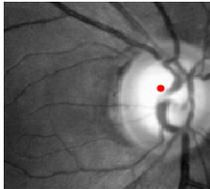
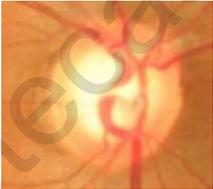
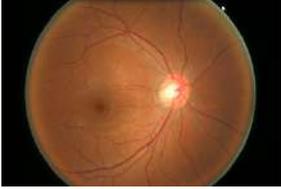
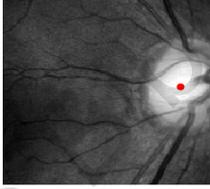
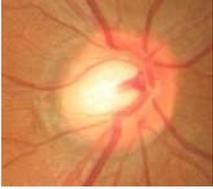
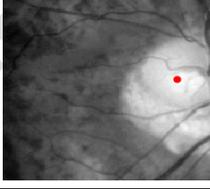
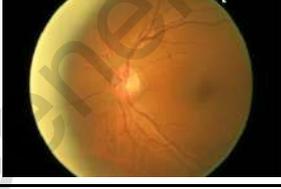
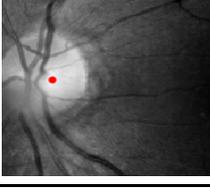
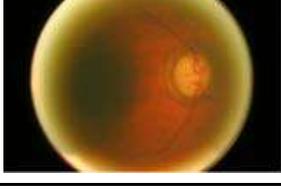
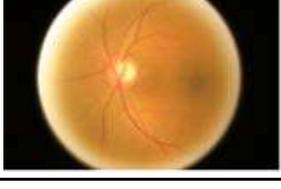
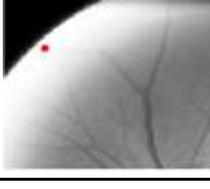
Tabla 4.1: Resultados para la obtención de la región de interés en la base de datos ORIGA.

Transformación a escala de grises	Recortadas correctamente	Recortadas incorrectamente	Resultados de la detección del ROI
UIT-R BT. 601-7	578	72	88.92%
Propuesto	648	2	99.69%

Analizando los resultados se puede observar que utilizando la transformación propuesta mejoran notablemente los resultados. Generalmente el área con más brillo en una imagen de retina suele ser el disco óptico, pero por distintos factores, como que el paciente se mueva a la hora de obtener la imagen, puede haber zonas con más brillo. Esto dificulta que se localice correctamente el área donde se encuentra en disco óptico. Aun así, se consiguieron muy buenos resultados, logrando localizar el ROI aún en imágenes complicadas. Utilizar una base de datos pública les da más peso a los resultados, ya que la

base de datos utilizada contiene tanto imágenes bien tomadas como imágenes no tan claras. En la tabla 4.2 podemos observar lo antes dicho.

Tabla 4.2: Obtención del ROI.

	Imagen original	Detección del centro	ROI
Imágenes claras			
			
Imágenes complicadas			
			
Imágenes no detectadas			No recortó la imagen
			No recortó la imagen

4.2 Resultados para la clasificación a nivel imagen.

Para poder evaluar los resultados obtenidos en este tipo de clasificación, se utilizaron las métricas de exactitud, sensibilidad y área bajo la curva (AUC). Para poder obtener estas métricas primero tenemos que tener claro algunos conceptos que se explicarán a continuación.

- Verdaderos positivos (TP): el valor predicho por el modelo fue positivo y concuerda con el valor verdadero.
- Verdadero negativo (TN): el valor predicho por el modelo fue negativo y concuerda con el valor verdadero.
- Falso positivo (FP): el modelo se clasifica como una clase positiva y el valor verdadero es negativo.
- Falso negativo (FN): el modelo se clasifica como una clase negativa y el valor real es positivo.

Con estos conceptos ya comprendidos, es posible obtener las métricas utilizadas en este trabajo.

La exactitud nos dice el porcentaje de predicciones que se hicieron correctamente durante la clasificación.

$$Exactitud = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.1)$$

La sensibilidad nos dice cuántos de los casos positivos el modelo pudo predecir correctamente.

$$Sensibilidad = \frac{TP}{TP + FN} \quad (4.2)$$

El área bajo la curva AUC es la métrica que nos muestra una curva de probabilidad, que se representa trazando el rango de verdaderos positivos contra el rango de falsos positivos en varios valores de umbral.

Para comprobar el impacto que tiene sobre la red la obtención del ROI, se ejecutó la arquitectura teniendo como entrada la imagen completa sin recortar.

Los resultados obtenidos, sin obtener el ROI, con las diferentes tasas de aprendizaje utilizadas, se muestra en la tabla 4.3.

Tabla 4.3: Desempeño sin la obtención del ROI.

Tasa de aprendizaje	Exactitud	Sensibilidad	AUC
0.01	89.93	90.13	90.76
0.001	90.42	90.97	91.48
0.0001	91.34	93.26	90.21

Los resultados obtenidos con diferentes tasas de aprendizaje y obteniendo el ROI se muestran en la tabla 4.4.

Tabla 4.4: Desempeño con la obtención del ROI.

Tasa de aprendizaje	Exactitud	Sensibilidad	AUC
0.01	92.1	93.23	93.11
0.001	92.84	93.96	93.58
0.0001	93.22	94.14	93.98

Además de obtener resultados inferiores sin obtener el ROI, el tiempo de entrenamiento fue mucho mayor, esto debido a que, con el recorte, la imagen se hace mucho más pequeña. Esto ayuda a reducir el tiempo que vaya a tener el algoritmo durante su entrenamiento.

Los resultados de la exactitud obtenida en cada uno de los k -folds, recordando que se utilizó un valor de k igual a 10, y con las diferentes tasas de aprendizaje utilizadas, se pueden observar en la figura 4.3.

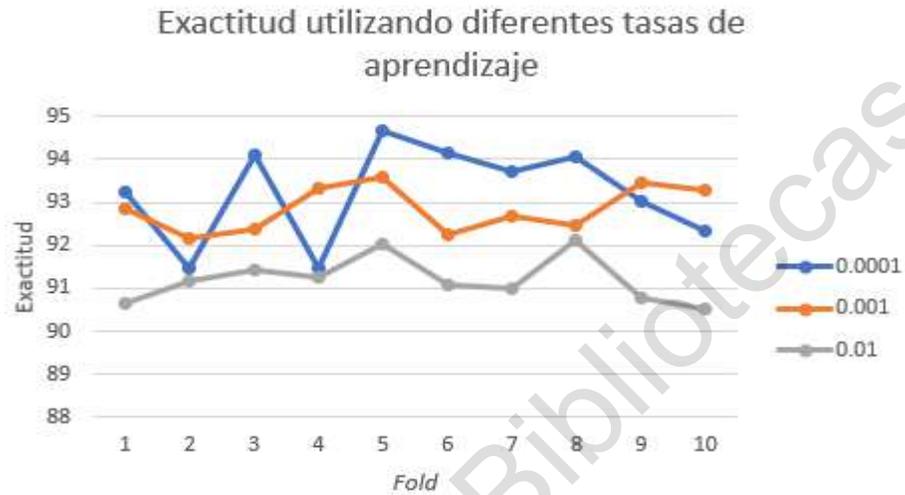


Figura 4.3: Desempeño de la red con las diferentes tasas de aprendizaje utilizadas.

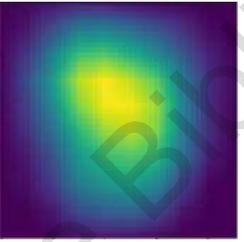
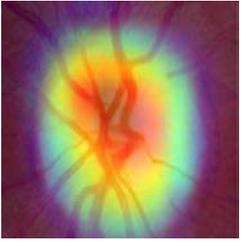
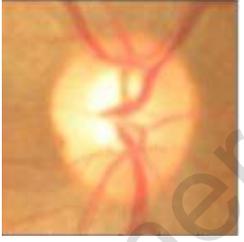
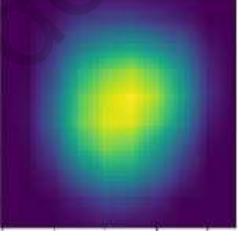
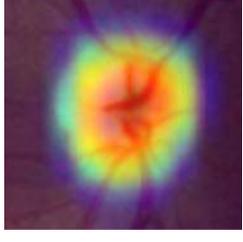
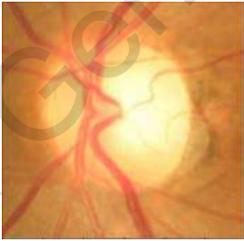
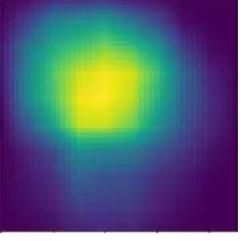
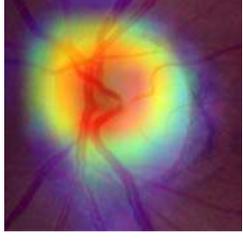
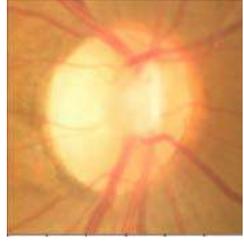
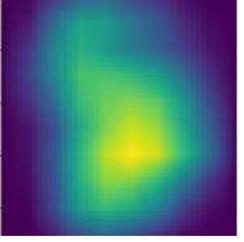
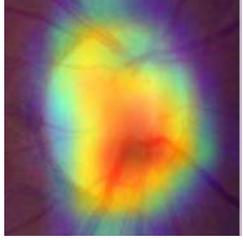
Para medir el desempeño de la red se obtuvo la representación del área bajo la curva. Esto se puede observar en la figura 4.4.



Figura 4.4: Área bajo la curva.

Aun cuando se obtuvieron buenos resultados al momento de realizar la clasificación, siempre queda la incertidumbre de en qué parte de la imagen se está enfocando la red para obtener dicha decisión. Esto podría ser de mucha ayuda cuando un especialista llegara a tener los resultados, para comprobar el enfoque de la red. Los resultados para obtener el mapa de características se muestran en la tabla 4.5. La técnica utilizada fue GradCAM (Selvaraju *et al.*, 2017) debido a su ventaja de adaptarse a muchos tipos de arquitecturas distintas.

Tabla 4.5: Mapa de activación en distintas imágenes.

	Imagen original	Mapa	Grad CAM
Sanas			
			
Glaucoma			
			

Existen varios trabajos relacionados con este tipo de clasificación, para comparar los resultados obtenidos, y medir el desempeño de la red propuesta contra otros trabajos encontrados en la literatura, se muestra una tabla comparativa donde se pueden apreciar que tan buenos fueron los resultados.

Tabla 4.6: Comparación de nuestro método contra los del estado del arte.

Autor	Año	Sensibilidad	Exactitud	AUC
Chen <i>et al.</i>	2015	-	-	0.887
Acharya <i>et al.</i>	2015	89.75	93.1	-
Gómez <i>et al.</i>	2018	87	-	0.942
Díaz <i>et al.</i>	2019	0.93	-	0.96
Raghavendra <i>et al.</i>	2019	98	98.13	-
Gour <i>et al.</i>	2020	-	83.4	0.88
Gheisari <i>et al.</i>	2021	95	-	-
Propuesto		94.14	93.22	0.93

4.3 Resultados para la clasificación a nivel píxel.

Para evaluar el algoritmo, se utilizaron nuevamente métricas como exactitud y sensibilidad, como en la clasificación a nivel imagen. Además de esas dos, se obtuvieron nuevas métricas como la especificidad y el error superpuesto. La ecuación 4.3 muestra la fórmula para obtener el error superpuesto.

$$Error = 1 - \frac{Area(S \cap G)}{Area(S \cup G)} \quad (4.3)$$

Donde G representa la imagen que se segmentó manualmente y S la imagen que el modelo arrojó como salida.

La especificidad, o tasa de verdaderos negativos, nos dice que proporción de la clase negativa fue clasificada correctamente. La ecuación 4.4 muestra la fórmula para obtener esta métrica, utilizando verdaderos negativos y falsos positivos como en la sección pasada.

$$\text{Especificidad} = \frac{TN}{TN + FP} \quad (4.4)$$

En el presente trabajo, el algoritmo se ejecutó segmentando individualmente el disco óptico y la copa. De esta forma, se obtuvieron mejores resultados que los obtenidos en trabajos en el estado del arte, que lo hacen conjuntamente.

Para valorar el efecto que tiene convertir las imágenes a coordenadas polares, los resultados obtenidos se reportan alimentando la red tanto con las imágenes originales, como con las imágenes aplicándoles dicha transformación.

La tabla 4.7 muestra los resultados de sensibilidad y especificidad para el disco y la copa óptica.

Tabla 4.7: Resultados de especificidad y sensibilidad.

Método	Sensibilidad	Especificidad	Sensibilidad	Especificidad
	Disco	Disco	Copa	Copa
Unet + bloque residual	0.82	0.97	0.74	0.96
Unet + bloque residual con transformación polar	0.85	0.99	0.83	0.97

La tabla 4.8 muestra los resultados de la exactitud y del error superpuesto aplicando la transformación polar y los obtenidos sin aplicar dicha transformación.

Tabla 4.8: Resultados de exactitud y error de superposición.

Método	Exactitud	Error de	Exactitud	Error de
	Disco	superposición	Copa	superposición
		Disco		Copa
Unet + bloque residual	93.21	0.171	90.4	0.0758
Unet + bloque residual con transformación polar	96.94	0.088	93.2	0.0706

Para comprobar los resultados obtenidos, se compararon contra algunos trabajos que se obtuvieron en el estado del arte. Dichos trabajos utilizan diferentes métodos para realizar la segmentación del disco y la copa óptica. Los resultados de la comparación se pueden observar en la tabla 4.9.

Tabla 4.9: Comparación del método propuesto con los vistos en la literatura.

Autor	Año	Exactitud	Error de	Exactitud	Error de
		Disco	superposición	Copa	superposición
			Disco		Copa
Ronneberger <i>et al.</i>	2013	95.9	0.115	90.1	0.287
Cheng <i>et al.</i>	2015	96.4	0.102	91.8	0.264
Fu <i>et al.</i>	2019	98.3	0.071	93.0	0.071
Joshua <i>et al.</i>	2019	-	0.12	-	0.21
Afolabi <i>et al.</i>	2020	-	0.1	-	0.21
Yin <i>et al.</i>	2021	-	0.071	-	0.220
Propuesto		96.94	0.088	93.2	0.0706

Los mejores resultados para la segmentación del disco óptico son los obtenidos por Fu *et al.* El método aquí propuesto logró los mejores resultados segmentando la copa. Ya que ambos métodos utilizan bases de datos diferentes, se replicó la red propuesta por Fu *et al.*, esto con el fin de poder realizar una mejor comparación entre las técnicas.

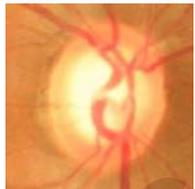
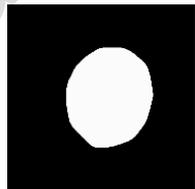
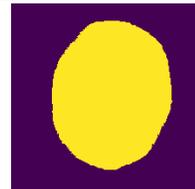
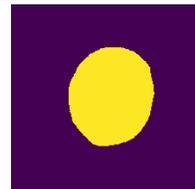
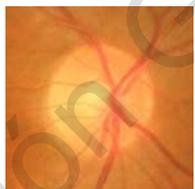
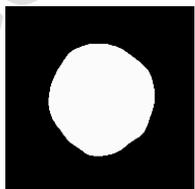
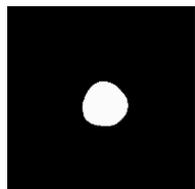
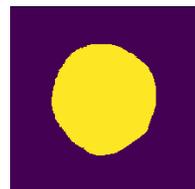
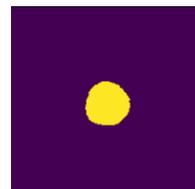
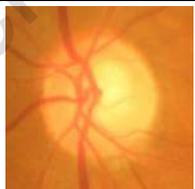
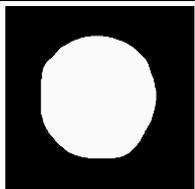
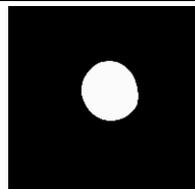
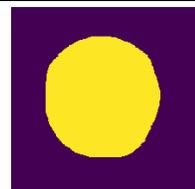
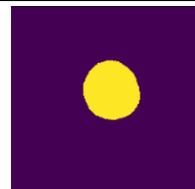
Debido a que únicamente se cuentan con 50 imágenes segmentadas de la copa óptica, únicamente se utilizaron esta cantidad de imágenes para entrenar y validar la red. La comparación de los resultados se muestra en la tabla 4.10

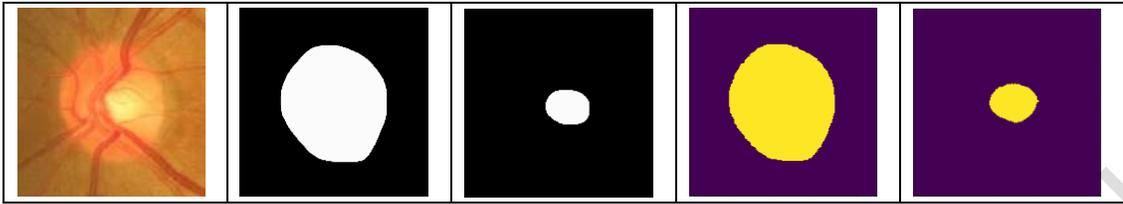
Tabla 4.10: Comparación entre métodos utilizando la misma base de datos.

Autor	Exactitud	Error de superposición	Exactitud	Error de superposición
	Disco	Disco	Copa	Copa
Fu <i>et al.</i>	97.1	0.082	91.9	0.194
Propuesto	96.94	0.088	93.2	0.0706

Los resultados de algunas de las imágenes que se lograron segmentar se muestran en la tabla 4.11.

Tabla 4.11: Algunos resultados obtenidos comparando la imagen segmentada con la imagen resultante.

Imagen	Segmentación manual del disco.	Segmentación manual de la copa.	Disco segmentado entregado por la red	Copa segmentada entregada por la red
				
				
				



Dirección General de Bibliotecas UAQ

Conclusión y trabajo futuro

Debido a la progresión que ha tenido el glaucoma a nivel nacional y mundial, y sabiendo que es una de las principales causas de ceguera en el mundo, es de vital importancia desarrollar nuevas técnicas y herramientas que ayuden a los oftalmólogos a realizar un diagnóstico más rápido. En este trabajo se propusieron dos enfoques diferentes para hacer esta tarea. La clasificación a nivel imagen alcanzo métricas bastante buenas con una red neuronal sin tantas capas, de esta manera podemos entrenar la red con mayor rapidez. Las métricas obtenidas fueron de una exactitud de 93.22, una sensibilidad de 94.14 y un área bajo la curva de 0.9398. Para poder comprobar el desempeño de la arquitectura propuesta, se hizo el análisis en la base de datos pública ORIGA. De esta forma, nos permite compararnos contra más trabajos porque es una de las bases de datos más utilizadas para el análisis de glaucoma.

Una de las técnicas propuestas fue la obtención del área de interés. Se obtuvieron resultados bastante buenos, únicamente no detectando el ROI en dos imágenes. Esto tuvo efecto en el algoritmo ayudándolo a tener mejor desempeño en la clasificación como se pudo observar en los resultados mostrados. Es importante haberse realizado sobre una base de datos pública, ya que de esta forma es uno de los métodos que son candidatos a usarse en futuros trabajos. Además de que la base de datos ORIGA contiene imágenes muy variadas en cuestión de calidad a la hora de obtener la imagen, dando más peso al desempeño del algoritmo.

Obteniendo el mapa de características de la imagen, pudimos observar en que parte de la imagen se estaba enfocando la red para realizar una clasificación. De esta manera, se le da un valor agregado a la red propuesta, y podría ser de mayor ayuda para un especialista al indicarle en que región se está enfocando para realizar una clasificación. Con el mapa de

características obtenido se puede concluir que la clasificación se centra en la región del disco y la copa óptica, esto es lo que se espera, ya que son las regiones donde se encuentran las peculiaridades para determinar si una persona padece de glaucoma o no.

Con respecto a la clasificación a nivel píxel, lo que se busca es lograr una correcta segmentación del disco y la copa óptica, ya que si bien, no es la única característica que presentan los pacientes con glaucoma, si es una de las más importantes y usadas para diagnosticar dicho padecimiento. En el presente trabajo, se puede observar cómo se obtuvieron mejores resultados, para la obtención de la copa, separando la segmentación del disco y la copa, que haciéndolo en conjunto. Hay que tener en cuenta que lograr una buena segmentación es una tarea que requiere un alto poder computacional. Al separar la segmentación en dos partes, puede hacer que la tarea sea menos exigente y el procesamiento más rápido.

Los resultados que se lograron con la base de datos utilizada muestra métricas muy competitivas. Algo que se debe resaltar, ya que gracias a eso se pudieron mejorar los resultados, fue el convertir las imágenes a coordenadas polares, puesto que esto nos ayudó a reducir el gran desequilibrio que se tenía entre clases. Una vez que se obtiene una imagen como salida, se regresa la imagen a su forma cartesiana original para visualizar de mejor manera los resultados.

Algo crucial para la investigación hecha fue la estancia realizada en el Instituto Mexicano de Oftalmología, durante ese tiempo se aclararon bastantes dudas sobre la enfermedad. Además de proporcionar una perspectiva diferente trabajar en conjunto con especialistas del área, ya que su amplio conocimiento del tema te ayuda a aprender de ellos y te da más ideas de que puedes implementar en la investigación.

5.1 Trabajo futuro.

- Seguir investigando nuevas formas de mejorar el rendimiento de la clasificación a nivel imagen y a nivel píxel, ya sea aplicando un procesamiento previo a las imágenes, o modificando las estructuras de las redes.

- Implementar un algoritmo que sea capaz de medir la relación copa disco que existen entre las imágenes segmentadas.
- Seguir segmentando las imágenes restantes de la base de datos ORIGA para tener un mayor volumen de datos que se puedan utilizar en futuros entrenamientos.
- Implementar el algoritmo de segmentación con base de datos diferentes.
- Analizar diferentes biomarcadores, además del disco y la copa óptica, que pueden ser útiles para realizar una mejor clasificación.

Referencias

- Acharya, U. R., Ng, E. Y. K., Eugene, L. W. J., Noronha, K. P., Min, L. C., Nayak, K. P., & Bhandary, S. V. (2015). Decision support system for the glaucoma using Gabor transformation. *Biomedical Signal Processing and Control*, 15, 18-26.
- Afolabi, O. J., Mabuza-Hocquet, G. P., Nelwamondo, F. V. y Paul, B. S. (2021). The Use of U-Net Lite and Extreme Gradient Boost (XGB) for Glaucoma Detection. *IEEE Access*, 9, 47411-47424.
- Aggarwal, C. C. (2018). *Neural networks and deep learning*. Springer, 10, 978-3.
- Albawi, S., Mohammed, T. A., y Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6).
- Almotiri, J., Elleithy, K., y Elleithy, A. (2018). An automated region-of-interest segmentation for optic disc extraction. In *2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (pp. 1-6). IEEE.
- Bar, Y., Diamant, I., Wolf, L. y Greenspan, H. (2015). Deep learning with non-medical training used for chest pathology identification. In *Medical Imaging 2015: Computer-Aided Diagnosis* (Vol. 9414, p. 94140V). International Society for Optics and Photonics.
- Berrar, D. (2019). Cross-validation. *Encyclopedia of bioinformatics and computational biology*, 1, 542-545.
- Bhaskaranand, M., Ramachandra, C., Bhat, S., Cuadros, J., Nittala, M. G., Sadda, S. R. y Solanki, K. (2019). The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes technology & therapeutics*, 21(11), 635-643.
- Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010* (pp. 177-186). Physica-Verlag HD.
- Brandon, L., y Hoover, A. (2003). Drusen detection in a retinal image using multi-level analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 618-625). Springer, Berlin, Heidelberg.
- Carneiro, G., Nascimento, J. y Bradley, A. P. (2015). Unregistered multiview mammogram analysis with pre-trained deep learning models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 652-660). Springer, Cham.

- Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T. y Heng, P. A. (2015). Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE journal of biomedical and health informatics*, 19(5), 1627-1636.
- Chen, X., Xu, Y., Wong, D. W. K., Wong, T. Y. y Liu, J. (2015). Glaucoma detection based on deep convolutional neural network. In 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 715-718). IEEE.
- Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D. W. K., Tan, N. M., ... y Wong, T. Y. (2013). Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE transactions on medical imaging*, 32(6), 1019-1032.
- Da, K. (2014). A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Dahl, G. E., Sainath, T. N. y Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8609-8613). IEEE.
- Daneshzand, M., Zoroofi, R. A., y Faezipour, M. (2014, April). MR image assisted drug delivery in respiratory tract and trachea tissues based on an enhanced level set method. In Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education (pp. 1-7). IEEE.
- Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., ... y Klein, J. C. (2014). Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3), 231-234.
- Diaz-Pinto, A., Morales, S., Naranjo, V., Köhler, T., Mossi, J. M., y Navea, A. (2019). CNNs for automatic glaucoma assessment using fundus images: an extensive validation. *Biomedical engineering online*, 18(1), 1-19.
- Duchi, J., Hazan, E. y Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
- Erhan, D., Bengio, Y., Courville, A., y Vincent, P. (2009). Visualizing higher-layer features of a deep network. *University of Montreal*, 1341(3), 1.
- Feng, J. y Lu, S. (2019). Performance analysis of various activation functions in artificial neural networks. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022030). IOP Publishing.
- Fernandez, V., Chavez, J. y Kemper, G. (2021). Device to evaluate cleanliness of fiber optic connectors using image processing and neural networks. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(4), 3093-3105.

- Fu, H., Cheng, J., Xu, Y. y Liu, J. (2019). Glaucoma detection based on deep learning network in fundus image. In *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics* (pp. 119-137). Springer, Cham.
- Fukushima, K. (1980). A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.*, 36, 193-202.
- García González, E. (2018). Segmentación de aorta en imágenes médicas 2D y 3D utilizando redes neuronales (Bachelor's thesis).
- Gheisari, S., Shariflou, S., Phu, J., Kennedy, P. J., Agar, A., Kalloniatis, M., y Golzan, S. M. (2021). A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Scientific reports*, 11(1), 1-11.
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A. y De, D. (2020). Fundamental concepts of convolutional neural network. In *Recent Trends and Advances in Artificial Intelligence and Internet of Things* (pp. 519-567). Springer, Cham.
- Gómez-Valverde, J. J., Antón, A., Fatti, G., Liefers, B., Herranz, A., Santos, A., ... y Ledesma-Carbayo, M. J. (2019). Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. *Biomedical optics express*, 10(2), 892-913.
- Gopalakrishnan, A., Almazroa, A., Raahemifar, K. y Lakshminarayanan, V. (2015). Optic disc segmentation using circular Hough transform and curve fitting. In *2015 2nd International Conference on Opto-Electronics and Applied Optics (IEM OPTRONIX)* (pp. 1-4). IEEE.
- Gour, N., y Khanna, P. (2020). Automated glaucoma detection using GIST and pyramid histogram of oriented gradients (PHOG) descriptors. *Pattern Recognition Letters*, 137, 3-11.
- Hara, K., Saito, D. y Shouno, H. (2015). Analysis of function of rectified linear unit used in deep learning. In *2015 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.
- He, K., Zhang, X., Ren, S. y Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hochreiter, S., Younger, A. S. y Conwell, P. R. (2001). Learning to learn using gradient descent. In *International Conference on Artificial Neural Networks* (pp. 87-94). Springer, Berlin, Heidelberg.

- Hu, H. (2018). vReLU Activation Functions for Artificial Neural Networks. In 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) (pp. 856-860). IEEE.
- Huang, G. R., y Hsiang, T. R. (2020). A Simplified Deep Network Architecture on Optic Cup and Disc Segmentation. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-7). IEEE.
- Hubel, D. H. y Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1), 215-243.
- Instituto de Salud para el Bienestar (2020). Día mundial del glaucoma. <https://www.gob.mx/insabi/articulos/12-de-marzo-dia-mundial-del-glaucoma?idiom=es>
- Ioffe, S. y Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). PMLR.
- Jain, A. K., Mao, J. y Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31-44.
- Jiang, Y., Duan, L., Cheng, J., Gu, Z., Xia, H., Fu, H. y Liu, J. (2019). JointRCNN: a region-based convolutional neural network for optic disc and cup segmentation. *IEEE Transactions on Biomedical Engineering*, 67(2), 335-343.
- Jonas, J. B., Bergua, A., Schmitz-Valckenberg, P., Papastathopoulos, K. I., y Budde, W. M. (2000). Ranking of optic disc variables for detection of glaucomatous optic nerve damage. *Investigative Ophthalmology & Visual Science*, 41(7), 1764-1773.
- Joshi, G. D., Sivaswamy, J. y Krishnadas, S. R. (2011). Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE transactions on medical imaging*, 30(6), 1192-1205.
- Joshi, P., KS, R. R., Masilamani, V., Alike, J., Suresh, K. y Kumares, K. (2019). Optic Disc Localization using Interference Map and Localized Segmentation. In 2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP) (pp. 1-4). IEEE.
- Joshua, A. O., Nelwamondo, F. V. y Mabuza-Hocquet, G. (2019). Segmentation of optic cup and disc for diagnosis of glaucoma on retinal fundus images. In 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA) (pp. 183-187). IEEE.
- Kamber, M. y Pei, J. (2006). *Data Mining*. Morgan Kaufmann.

- Kanigoro, B. y Salman, A. G. (2016). Recurrent gradient descent adaptive learning rate and momentum neural network for rainfall forecasting. In 2016 international seminar on application for technology of information and communication (ISemantic) (pp. 23-26). IEEE.
- Kazi, A., Ajmera, M., Sukhija, P., y Devadkar, K. (2018). Processing Retinal Images to Discover Diseases. International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-5). IEEE.
- Krizhevsky, A., Sutskever, I. y Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- LeCun, Y., Boser, B. E., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. E. y Jackel, L. D. (1990). Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems* (pp. 396-404).
- LeCun, Y., Bottou, L., Bengio, Y. y Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Li, M., Soltanolkotabi, M. y Oymak, S. (2020). Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks. In *International Conference on Artificial Intelligence and Statistics* (pp. 4313-4324). PMLR.
- Liu, Q., Zou, B., Zhao, Y. y Liang, Y. (2020). A Deep Gradient Boosting Network for Optic Disc and Cup Segmentation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 971-975). IEEE.
- Maninis, K. K., Pont-Tuset, J., Arbeláez, P., y Van Gool, L. (2016). Deep retinal image understanding. In *International conference on medical image computing and computer-assisted intervention* (pp. 140-148). Springer, Cham.
- Mining, W. I. D. (2006). *Data mining: Concepts and techniques*. Morgan Kaufmann.
- Mittapalli, P. S., y Kande, G. B. (2016). Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. *Biomedical Signal Processing and Control*, 24, 34-46.
- Mohan, D., Kumar, J. H. y Seelamantula, C. S. (2019). Optic disc segmentation using cascaded multiresolution convolutional neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 834-838). IEEE.
- Morimoto, T., Isoe, N., Ohizumi, Y., Al Maki, W. F. y Sugimoto, S. (2011). Restoration of Rotational Motion Blurred Images using the Inverse Filters. In *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its*

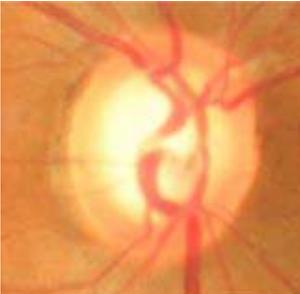
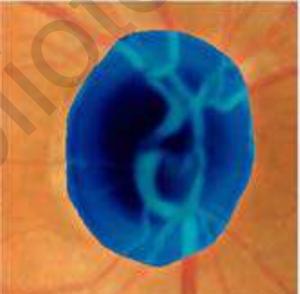
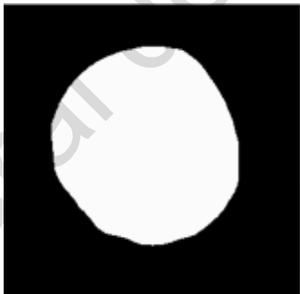
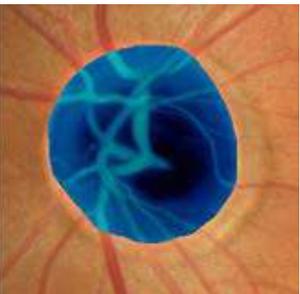
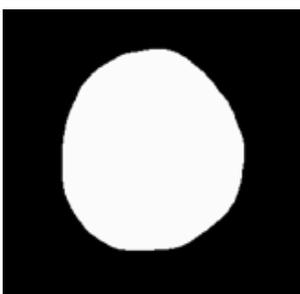
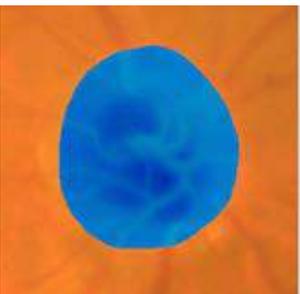
- Applications (Vol. 2011, pp. 264-270). The ISCIE Symposium on Stochastic Systems Theory and Its Applications.
- Nair, V. y Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml*.
- Neeru, G. (2015). Guías del Consejo Internacional de Oftalmología (ICO) para el Glaucoma. *Int Counc Ophthalmol*, 2-20.
- Nie, F., Hu, Z. y Li, X. (2018). An investigation for loss functions widely used in machine learning. *Communications in Information and Systems*, 18(1), 37-52.
- Pan, F., Lu, Z., Chen, D., y Xue, D. (2020). An optic disk semantic segmentation method based on weakly supervised learning. In *2020 Chinese Control And Decision Conference (CCDC)* (pp. 4791-4794). IEEE.
- Pascolini, D., Mariotti, S. P., Pokharel, G. P., Pararajasegaram, R., Etya'ale, D., Négrel, A. D., y Resnikoff, S. (2004). 2002 global update of available data on visual impairment: a compilation of population-based prevalence studies. *Ophthalmic epidemiology*, 11(2), 67-115.
- Pinto, A. Y. D. (2019). Machine learning for glaucoma assessment using fundus images (Doctoral dissertation, Universitat Politècnica de València).
- Piñero, R. T., Lora, M., y Andrés, M. I. (2005). Glaucoma: Patogenia, diagnóstico y tratamiento. *Offarm: farmacia y sociedad*, 24(2), 88-96.
- Raghavendra, U., Fujita, H., Bhandary, S. V., Gudigar, A., Tan, J. H. y Acharya, U. R. (2018). Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences*, 441, 41-49.
- Ronneberger, O., Fischer, P. y Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... y Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252.
- Salman, S. y Liu, X. (2019). Overfitting mechanism and avoidance in deep neural networks. *arXiv preprint arXiv:1901.06566*.

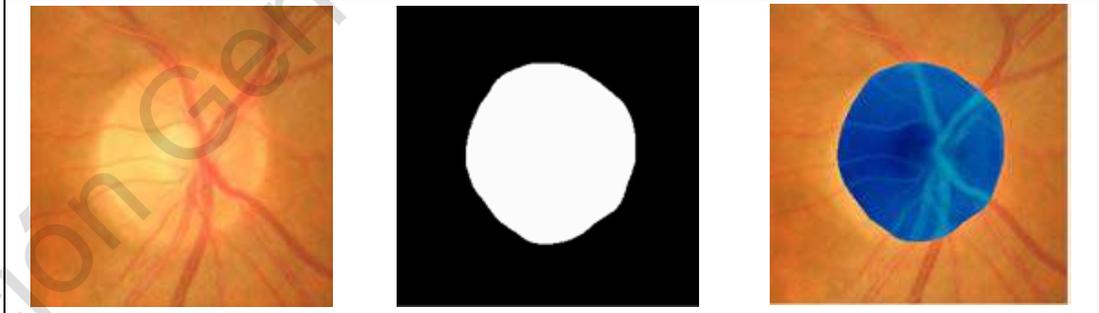
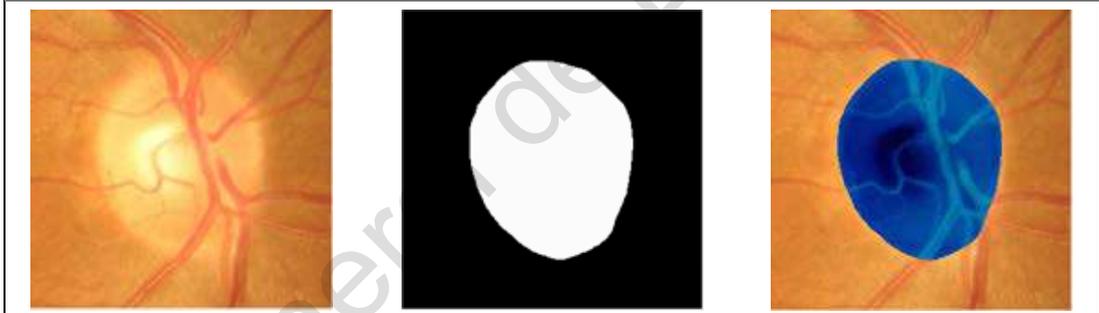
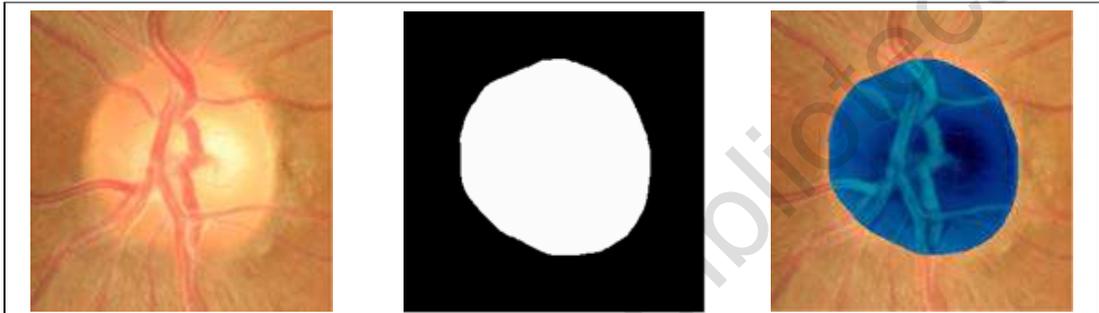
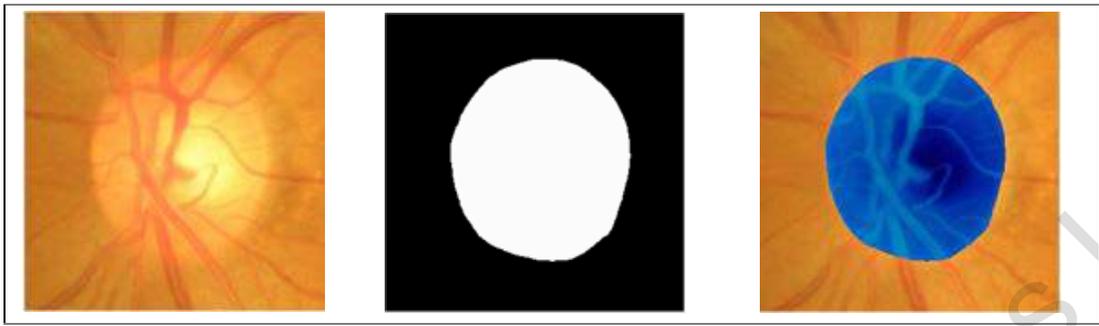
- Sarhan, A., Rokne, J., y Alhaji, R. (2019). Glaucoma detection using image processing techniques: A literature review. *Computerized Medical Imaging and Graphics*, 78, 101657.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., y Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).
- Sertkaya, M. E., Ergen, B., y Togacar, M. (2019, June). Diagnosis of eye retinal diseases based on convolutional neural networks using optical coherence images. In *2019 23rd International conference electronics* (pp. 1-5). IEEE.
- Sevastopolsky, A. (2017). Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Pattern Recognition and Image Analysis*, 27(3), 618-624.
- Shorten, C. y Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1-48.
- Simonyan, K., Vedaldi, A., y Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. y Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. y Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE transactions on medical imaging*, 35(5), 1299-1312.
- Tham, Y. C., Li, X., Wong, T. Y., Quigley, H. A., Aung, T., y Cheng, C. Y. (2014). Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11), 2081-2090.
- Wang, Z. y Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1), 98-117.
- Weinreb, R. N., y Khaw, P. T. (2004). Primary open-angle glaucoma. *The Lancet*, 363(9422), 1711-1720.
- Xiao, X., Lian, S., Luo, Z., y Li, S. (2018). Weighted res-unet for high-quality retina vessel segmentation. In *2018 9th international conference on information technology in medicine and education (ITME)* (pp. 327-331). IEEE.

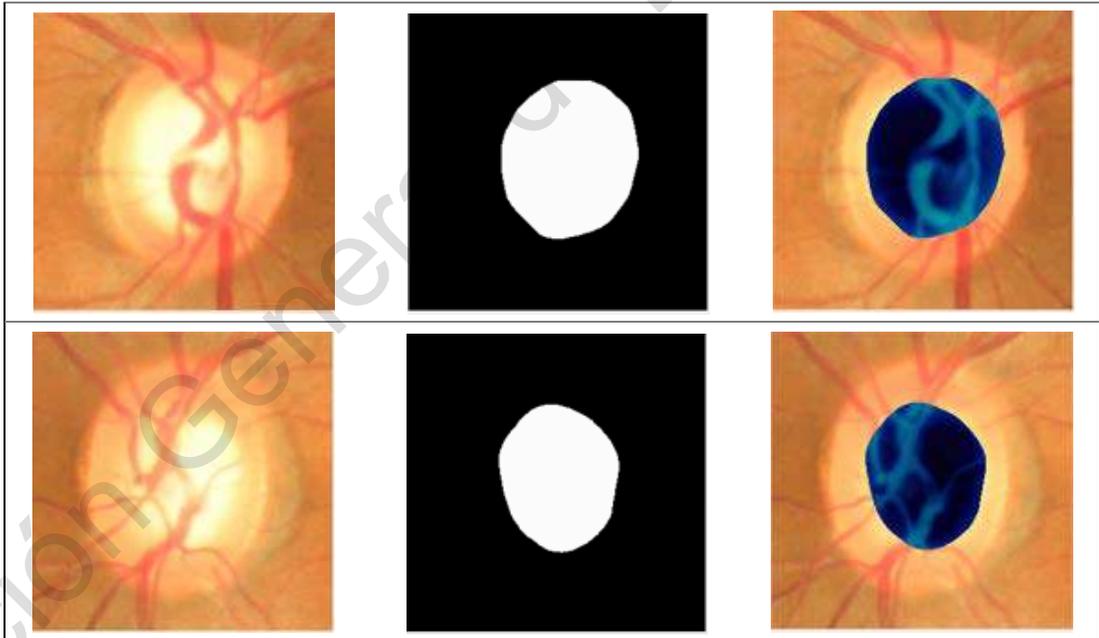
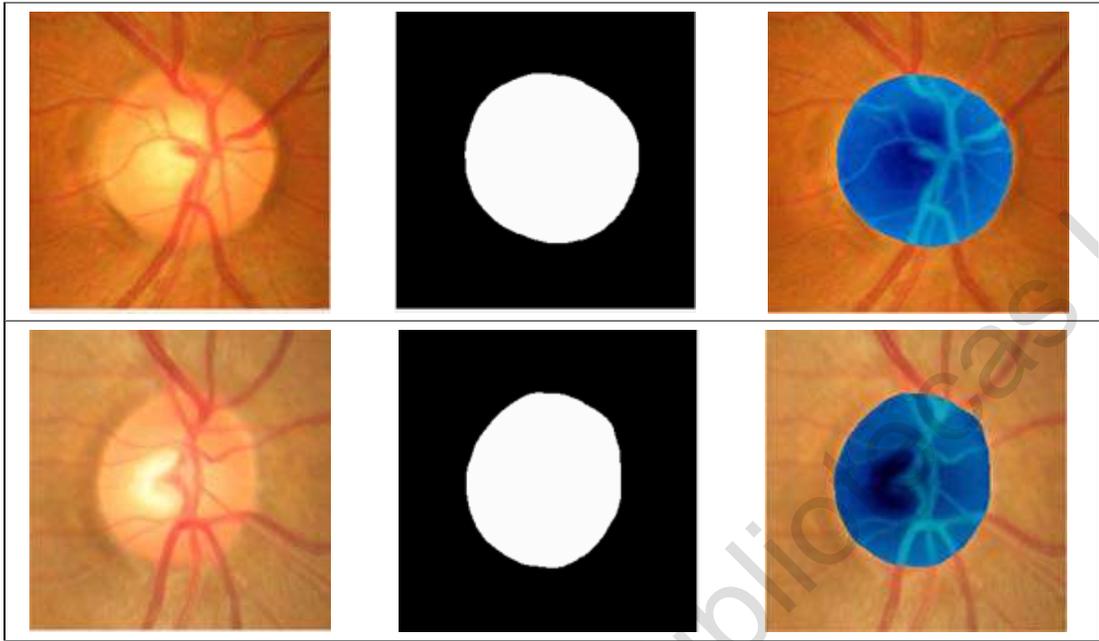
- Xu, J., Xue, K. y Zhang, K. (2019). Current status and future trends of clinical diagnoses via image-based deep learning. *Theranostics*, 9(25), 7556.
- Yessou, H., Sumbul, G. y Demir, B. (2020). A Comparative Study of Deep Learning Loss Functions for Multi-Label Remote Sensing Image Classification. arXiv preprint arXiv:2009.13935.
- Yildirim, O. E., Xu, Q. Basaran, O. A. (2005). Analysis of the drop weight method. *Physics of Fluids*, 17(6), 062107.
- Yin, F., Liu, J., Wong, D. W. K., Tan, N. M., Cheung, C., Baskaran, M., y Wong, T. Y. (2012). Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis. In 2012 25th IEEE international symposium on computer-based medical systems (CBMS) (pp. 1-6). IEEE
- Yin, P., Yuan, R., Cheng, Y. y Wu, Q. (2020). Deep Guidance Network for Biomedical Image Segmentation. *IEEE Access*, 8, 116106-116116.
- Zeiler, M. D. y Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.
- Zeiler, M. D., Taylor, G. W., y Fergus, R. (2011). Adaptive deconvolutional networks for mid and high level feature learning. In 2011 International Conference on Computer Vision (pp. 2018-2025). IEEE.
- Zhang, Y. y Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1), 95-112.
- Zhang, Z., Yin, F. S., Liu, J., Wong, W. K., Tan, N. M., Lee, B. H. y Wong, T. Y. (2010). Origa-light: An online retinal fundus image database for glaucoma analysis and research. Annual International Conference of the IEEE Engineering in Medicine and Biology (pp. 3065-3068). IEEE.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. y Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2921-2929).
- Zilly, J. G., Buhmann, J. M. y Mahapatra, D. (2015). Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images. In International workshop on machine learning in medical imaging (pp. 136-143). Springer, Cham.

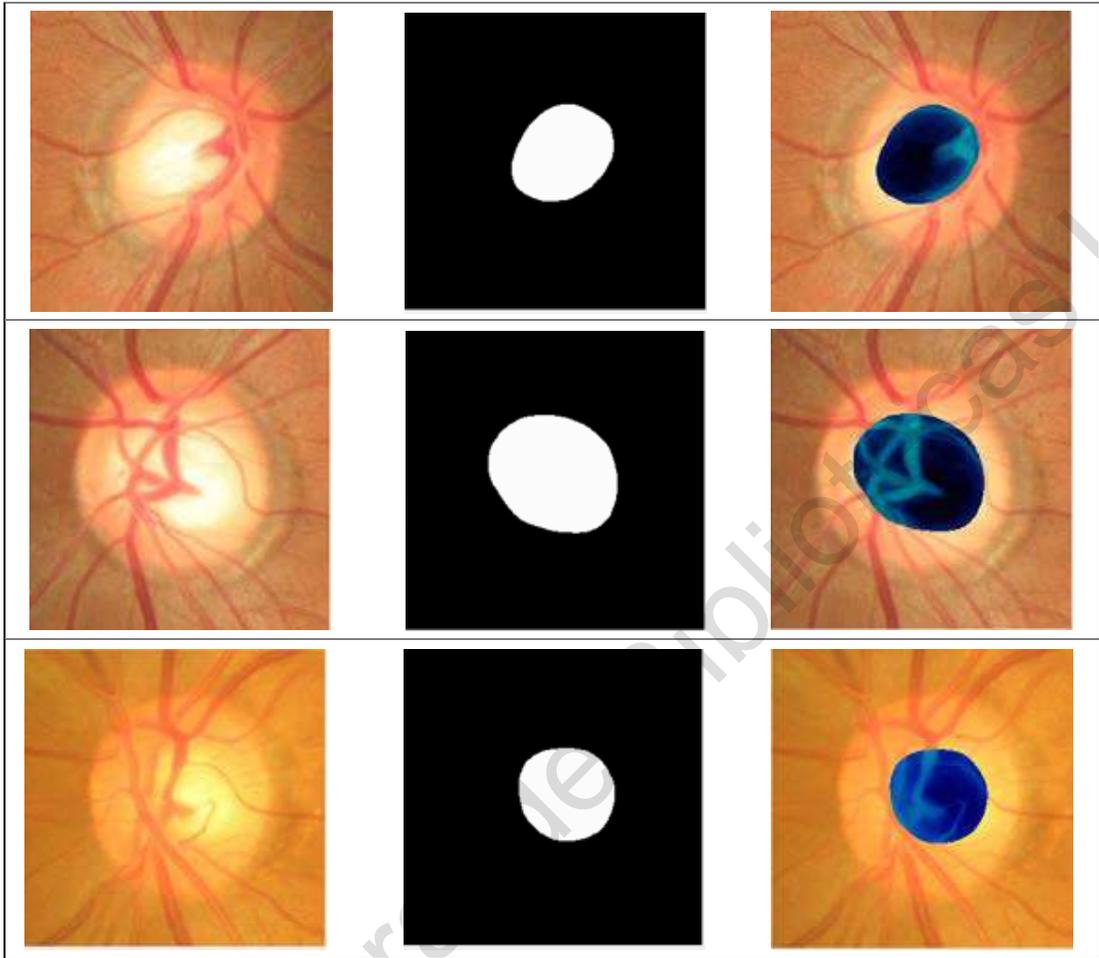
A. Anexos

- Segmentación manual de algunas imágenes

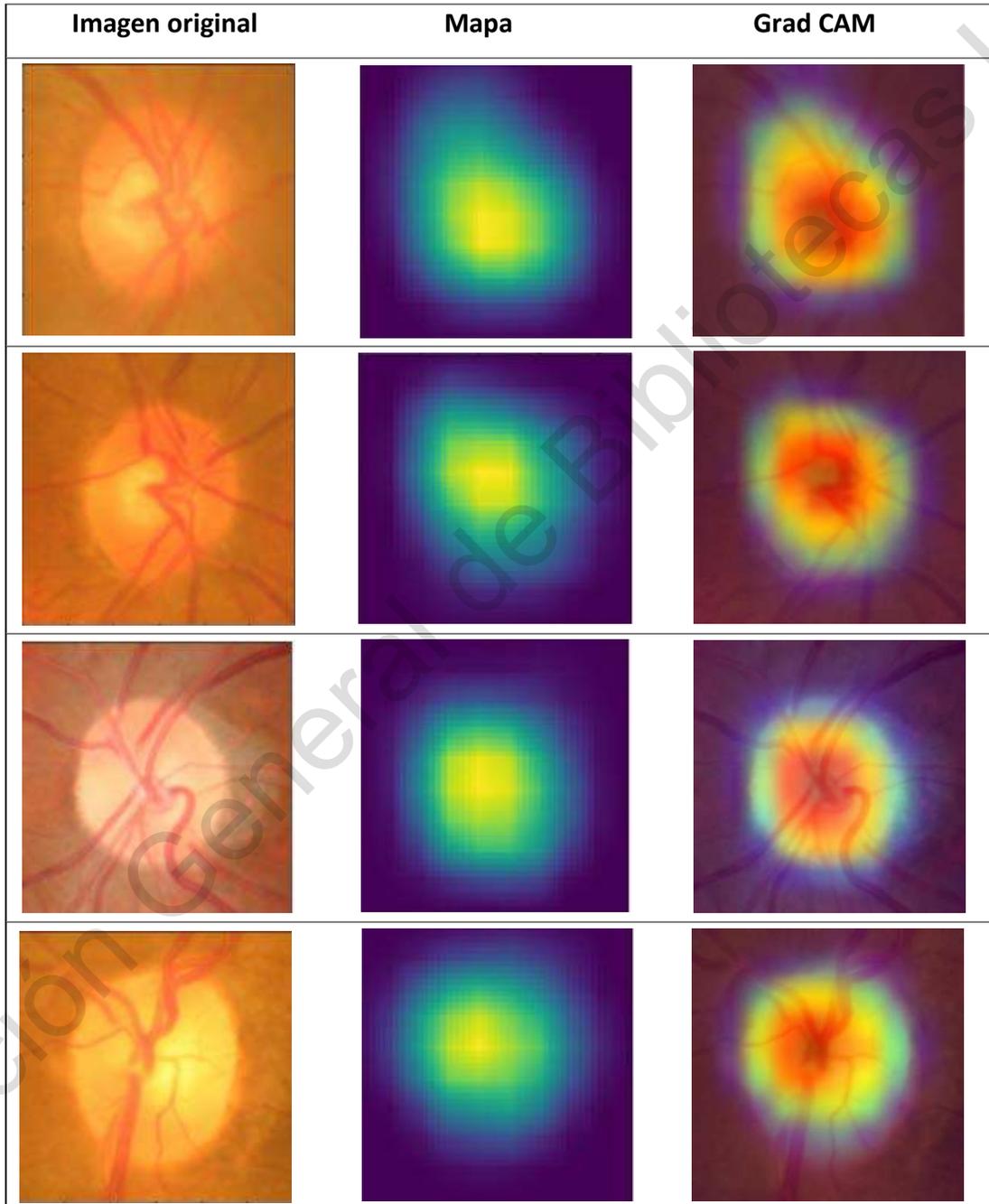
Imagen original	Imagen segmentada	Comparación de imágenes
		
		
		

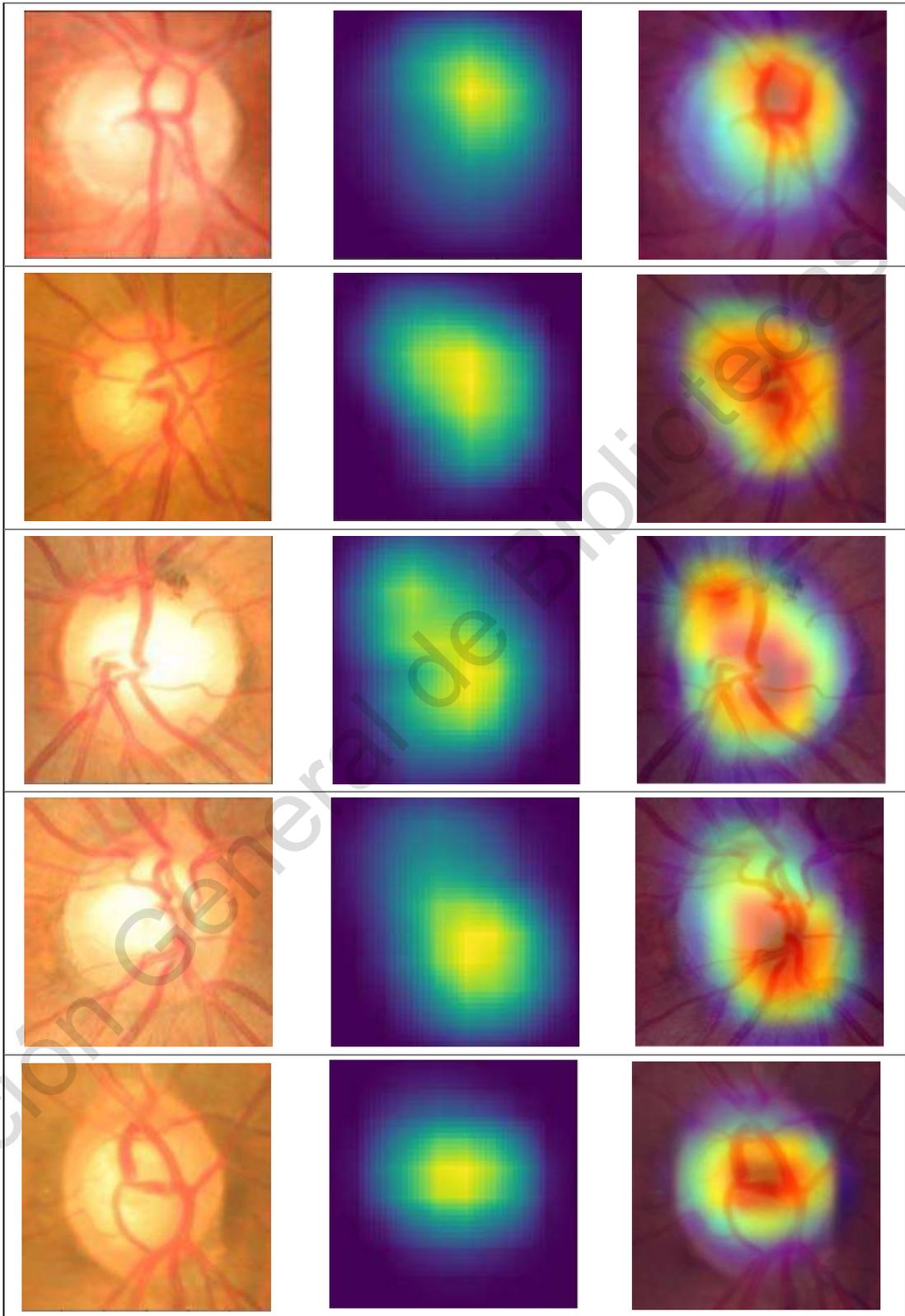






- Mapa de características de algunas de las imágenes





B. Artículo publicado

REVISTA MEXICANA DE Ingeniería Biomédica

SOMIB
Sociedad Mexicana de Ingeniería Biomédica

Ciudad de México a 11 de septiembre del 2021
No. 055/RMIB/2021

**Hiram José Sandoval-Cuellar
Gendry Alfonso-Francia
Miguel Angel Vázquez-Membrillo
Juan Manuel Ramos-Arreguín
Saúl Tovar-Arriaga
Universidad Autónoma de Querétaro/
Instituto Mexicano de Oftalmología
P R E S E N T E**

Por este conducto me permito hacer de su conocimiento que su manuscrito con ID-1188 titulado:

"Image-based Glaucoma Classification Using Fundus Images and Deep Learning"

Sometido para su posible publicación a la Revista Mexicana de Ingeniería Biomédica ha sido **ACEPTADO** en la modalidad de **ARTÍCULO DE INVESTIGACIÓN** por el Comité Editorial de la Revista.

El cual será publicado en el Volumen 42 - Número 3 de la Revista Mexicana de Ingeniería Biomédica. Una vez más, gracias por someter su manuscrito a la Revista Mexicana de Ingeniería Biomédica.

Reciban un cordial saludo.

Dr. César Antonio González Díaz.
Editor en Jefe de la Revista Mexicana de Ingeniería Biomédica
www.rmib.mx

WITCOM 2021 notification for paper 17 Recibidos x



WITCOM 2021 <witcom2021@easychair.org>
para mí ▾

mié, 28 jul. 00:02 ☆ ↶ ⋮

🌐 inglés ▾ > español ▾ [Traducir mensaje](#)

[Desactivar para: inglés](#) x

Dear Hiram Sandoval,

We are happy to inform you that your submission 17 ,Optic disc and optic cup segmentation using polar coordinate and encoder-decoder architecture, was accepted for an online presentation at WITCOM 2021.

Every submission was evaluated by three regular reviewers. Senior PC members then looked at these reviews and initiated a discussion among the reviewers. At the conclusion of the discussion, the senior PC members made an overall recommendation and then wrote a summary meta-review (only if needed). Then a final decision was applied.

For the document, the deadline camera ready-version is required before August 5th, 2021 , 12 PM PST time

Your submission's reviews are at the end of this email

We strongly recommend considering all the comments in the reviews when you prepare your final version (CAMERA READY version).
Please consider the following checklist when you send us your camera -ready version:

- 1.- It is mandatory to follow editorial rules and Conference format. You can find the templates at (in english): <https://drive.google.com/open?id=1Ze4G5R5nmKhxQr10c68tesOp-j6V19t0>
(if you used LaTeX , also the templates are available in Overleaf)
- 2.- Ensure you follow Guidelines for Authors

C. Estancia académica



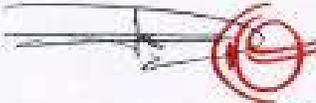
Santiago de Querétaro, Qro., a 17 diciembre del 2020

Dr. Saúl Tovar-Arriaga
Coordinador de la Maestría en Ciencias en Inteligencia Artificial
De la Universidad Autónoma de Querétaro
Presente

Por este medio hacemos constar que el Ing. Hiram Sandoval Cuellar, alumno de la Maestría en Ciencias en Inteligencia Artificial, concluyó su estancia corta de un mes en nuestro Instituto en donde trabajó en el etiquetado de manera adecuada de una base de datos pública, que le ayudará en el desarrollo de su tesis de maestría en la detección automatizada de glaucoma por medio de imágenes RGB.

Esperamos que esta experiencia académica haya sido de utilidad.

Atentamente,



INSTITUTO MEXICANO DE OFTALMOLOGÍA I.A.P.
DIRECCIÓN DE ENSEÑANZA

Dr. Jaime Macías Martínez
Jefe de Enseñanza e Investigación del
Instituto Mexicano de Oftalmología I.A.P.

Circuito Estadio Corregidora S/N Col. Colinas del Cimarrón
C.P. 76099, Querétaro, Qro. México Tels. (442) 229-0776, 229-0778
www.imoiaq.edu.mx

D. Certificado de inglés



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hago **C O N S T A R** que

SANDOVAL CUELLAR HIRAM JOSÉ

Presentó el Examen de Manejo de la Lengua efectuado el día diez de noviembre de dos mil veinte, en el cual obtuvo la siguiente calificación:

8-

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día diecinueve de noviembre de dos mil veinte.



Atentamente,
"Enlazar Culturas por la Palabra"

LIC. LAURA PÉREZ TÉLLEZ

LPT/jals*CL*FLL-C.-1702

SOMOS UAQ
TRABAJAMOS JUNTOS

Campus Aeropuerto, Avdo. Vial Frey y Jiménez Suro S/N, Querétaro, Qro. C.P. 76140
Tel. 442 192 12 00 Dirección Ext. 61010, Secretaría Administrativa Ext. 61300, Rectoría Ext. 61140
Laboratorio Ext. 61070, Centro de Idiomas Ext. 61050, Secretaría Académica Ext. 61100 y Planeación Ext. 61110



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



Campus Aeropuerto, 3 de diciembre de 2020.
FLL/SA/HCA/2020/1075

C. Sandoval Cuellar Hiram José
Expediente 290827
Facultad de Ingeniería
Presente.

Me permito comunicar a usted que en Sesión Ordinaria del H. Consejo Académico de la Facultad de Lenguas y Letras, celebrada el **03 de diciembre de 2020**, se acordó **aprobar** su solicitud de **convalidación** de constancia de Examen de Manejo de la Lengua (8-), expedida por la Facultad de Lenguas y Letras de la UAQ, por el Examen de Comprensión de Textos en Inglés.

Sin más por el momento, me es grato quedar a sus órdenes para cualquier aclaración.

Atentamente,
"Enlazar Culturas por la Palabra"

Dra. Ma. de Lourdes Rico Cruz
Secretaría Académica



SECRETARÍA
ACADÉMICA

C.c. Expediente
Archivo.

MURC/mrp