



Universidad Autónoma de Querétaro
Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

Análisis y detección de patrones de riesgo cardiovascular en
imágenes de retina usando aprendizaje profundo.
Tesis

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

Gendry Alfonso Francia

Dirigido
por:

Dr. Saúl Tovar Arriaga

Dr. Saúl Tovar Arriaga
Presidente

Dr. Jesús Carlos Pedraza Ortega
Secretario

Dr. Edgar Alejandro Rivas Araiza
Vocal

Dr. Marco Antonio Aceves Fernández
Suplente

Dr. Efrén Gorrostieta Hurtado
Suplente

Centro Universitario Querétaro,
QRO mayo de 2020
México.

“Sin compromiso nunca empezaremos. Sin disciplina y consistencia nunca terminaremos.”

Dirección General de Bibliotecas UAQ



DEDICATORIA

A mi madre, Julia, faro y guía de mi vida. A mis familiares y amigos que siempre me han brindado su apoyo. A todos los profesores del claustro del programa de maestría, que han sabido guiarme en este arduo camino.

Dirección General de Bibliotecas UAQ



RECONOCIMIENTO

En primer lugar, al Consejo Nacional de Ciencia y Tecnología (CONACyT), por proveer los fondos necesarios para realizar este trabajo de investigación. A la Universidad Autónoma de Querétaro y en específico a la Facultad de Ingeniería, por dar soporte y facilidades en mis estudios del posgrado. Reconocimiento especial al Dr. Saúl Tovar Arriaga, por su asesoría en la elaboración y desarrollo de esta tesis.

Dirección General de Bibliotecas UAQ



RESUMEN

Las enfermedades cardiovasculares (ECV) son sin dudas la principal causa de defunción en el mundo. Se desea desarrollar un algoritmo que permita explorar y analizar imágenes de retina, para predecir eventos cardiovasculares, ya que en dichas imágenes se puede detectar retinopatías hipertensivas y émbolos de colesterol, además varias características como el calibre de los vasos sanguíneos, bifurcaciones o tortuosidad que pueden reflejar la salud del sistema cardiovascular y predecir un riesgo futuro; además de que el procedimiento no es invasivo, es rápido y barato.

Para el desarrollo del trabajo se pretende utilizar técnicas de inteligencia artificial, específicamente de aprendizaje profundo (Deep Learning), ya que estas permiten computar múltiples capas de información en una red neuronal y aprender los patrones correctos sin tener que hacerlos de manera manual. Para ello se hará uso de las redes neuronales convolutivas, las cuales están optimizadas para producir algoritmos de alta precisión que diagnostican enfermedades a partir de imágenes médicas, con precisiones parecidas a las de un experto humano.

Palabras claves: enfermedades cardiovasculares, imágenes de retina, inteligencia artificial, aprendizaje profundo, red neuronal, red neuronal convolucional.

ABSTRACT

Cardiovascular diseases (CVD) are undoubtedly the leading cause of death in the world. The aim of this study is to develop an algorithm that allows the exploration and analysis of retinal images in order to predict cardiovascular events, since in these images hypertensive retinopathies and cholesterol emboli can be detected, as well as several characteristics such as the caliber of the blood vessels, bifurcations or tortuosity that can reflect the health of the cardiovascular system and predict a future risk. The advantages of this procedure are that it is non-invasive, fast and inexpensive.

For the development of this work we intend to use artificial intelligence techniques, specifically deep learning, since these allow us to compute multiple layers of information in a neural network and learn the correct patterns without having to do it manually. For this purpose, convoluted neuronal networks will be used, which are optimized to produce high precision algorithms that diagnose diseases from medical images, with precision similar to that of a human expert.

Keywords: cardiovascular diseases, retinal imaging, artificial intelligence, deep learning, neuronal network, convolutional neuronal network.

ÍNDICE DE CONTENIDO

DEDICATORIA.....	II
RECONOCIMIENTO.....	III
RESUMEN	IV
ABSTRACT	V
ÍNDICE DE CONTENIDO.....	VI
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABLAS.	XII
1. INTRODUCCIÓN	1
1.1. Inteligencia Artificial.....	1
1.2. Enfermedades cardiovasculares	3
1.3. Justificación del problema	5
1.4. Descripción del problema.....	6
1.5. Objetivo general	8
1.6. Objetivos específicos	9
1.7. Hipótesis	9
1.8. Estrategia de investigación	9
1.9. Métodos científicos.....	9
1.10. Estructura de la tesis	10
2. MARCO TEÓRICO	11
2.1. Revisión médica.....	11
2.1.1. Imágenes de retina	11
2.1.2. La retina y los factores de riesgos cardiovasculares.	12
2.1.3. Segmentación de vasos sanguíneos	13
2.2. Revisión tecnológica.	15

2.2.1.	Red neuronal.....	15
2.2.2.	Red neuronal profunda o Deep learning.....	17
2.2.2.1.	Redes neuronales convolucionales (CNNs).....	20
2.2.2.2.	U-Net.....	22
2.2.2.3.	Redes residuales.....	24
2.3.	Materiales y herramientas.....	26
2.3.1.	Python.....	26
2.3.2.	Pytorch.....	26
2.3.3.	Entorno virtual.....	27
2.3.4.	Anaconda.....	27
2.3.5.	Spyder como IDE.....	28
2.3.6.	Programación paralela.....	29
2.3.7.	Hardware utilizado.....	29
2.3.8.	Fuente de datos.....	29
2.4.	Estado del arte.....	30
2.5.	Propuesta de solución.....	31
3.	METODOLOGÍA.....	32
3.1.	Pre-procesamiento.....	33
3.1.1.	Conversión de imagen RGB a escala de grises.....	34
3.1.2.	Normalización de los datos.....	35
3.1.3.	Ecuilización de histograma a través de contraste limitado adaptativo (CLAHE).....	35
3.1.4.	Ajuste de la luminosa.....	36
3.1.5.	Parches y volteos aleatorios.....	37
3.2.	Arquitectura.....	37

3.2.1.	UNet_1.	38
3.2.2.	Bloque residual.	38
3.2.3.	UNet_2 con bloques residuales.	39
4.	RESULTADOS Y EVALUACIÓN.	41
4.1.	Elementos de configuración.	41
4.1.1.	Función de pérdida.	41
4.1.2.	Inicialización de pesos.	42
4.1.3.	Hiper-parámetros.	42
4.1.3.1.	Adam como optimizador.	42
4.1.3.2.	Tasa de aprendizaje.	43
4.1.3.3.	Otros hiper-parámetros configurados.	44
4.2.	Métricas de evaluación.	44
4.2.1.	Recuperación, precisión y F1-score.	45
4.2.2.	Validación cruzada k-fold.	46
4.2.3.	Índice de similitud estructural (SSIM).	47
4.3.	Experimentación y resultados.	47
5.	CONCLUSIONES.	67
6.	TRABAJOS FUTUROS.	69
7.	REFERENCIAS.	70
8.	ANEXO I.	76
9.	ANEXO II.	80
10.	ANEXO III.	85
11.	ANEXO IV.	86

ÍNDICE DE FIGURAS

Figura 1.1: Línea de tiempo de los avances en la IA dentro de la última década aproximadamente. (Elaboración propia)	2
Figura 1.2: Vasculatura entre corazón y ojos (Heart & Eye, 2015).....	4
Figura 1.3: Imagen de retina con estrechamiento arterial y entrecruzamiento arterio-venar (Modi, 2020).	5
Figura 1.4: Mortalidad por infarto agudo al miocardio según OCDE (“OECD iLibrary OECD Health Statistics,”2015).....	7
Figura 2.1: Imagen de retina y sus principales componentes.	12
Figura 2.2:Gráfica de función sigmoide.....	16
Figura 2.3: Gráfica de función tangente hiperbólica.	16
Figura 2.4:Topología de una red neuronal con una capa oculta.	17
Figura 2.5: Gráfica de función ReLU.	18
Figura 2.6: Gráfica de la función Leaky ReLU.....	19
Figura 2.7: Red neuronal convolucional (Josh Patterson & Adam Gibson, 2017).20	
Figura 2.8: Operación de convolución entre imagen de entrada (I), el filtro aplicado (K) y el resultado obtenido de la operación (“ <i>asymptote - Drawing a convolution with Tikz - TeX - LaTeX Stack Exchange,</i> ” n.d.).....	21
Figura 2.9: Operaciones de pooling, normalmente son dos, como se pueden apreciar en la imagen (“ <i>A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,</i> ” 2018).....	22
Figura 2.10: Arquitectura de red U-Net (Ronneberger, Fischer, & Brox, 2015).	23
Figura 2.11: Tendencia del error, tanto en el entrenamiento como en las pruebas, para dos modelos, uno con 20 capas y otro con 56 capas (He, Zhang, Ren, & Sun, 2015).....	24
Figura 2.12 Arquitectura Res-Net (He et al., 2015).	25
Figura 3.1: Flujo de proceso para la detección de patrones de riesgo cardiovascular en imágenes de retina.....	32
Figura 3.2: Resultado del pre-procesamiento. En la izquierda la imagen original y a la derecha la imagen pre-procesada.	36

Figura 3.3: Bloque residual utilizado en el presente trabajo. Un mapa de características x pasa una secuencia de capas (azul: convolucional, amarillo: normalización por lotes, verde: activación ReLU). Después de este procedimiento, la característica original x se añade con los resultados obtenidos a través de las capas. Finalmente, se aplica una activación ReLU. 39

Figura 3.4: La arquitectura propuesta. Consiste en una U-Net, que le sigue otra red U-Net con bloques residuales. En la primera red U-Net se realiza la extracción de características, mientras que en la segunda red U-Net con bloques residuales se realiza la detección de nuevas características o de aquellas que son ambiguas.. 40

Figura 4.1: Efecto de diferentes valores de tasa de aprendizaje en la búsqueda del óptimo global (Josh Patterson & Adam Gibson, 2017)..... 44

Figura 4.2: Pérdida asociada al entrenamiento de la arquitectura propuesta. 51

Figura 4.3: Accuracy en la fase de entrenamiento. 52

Figura 4.4: F1-Score asociado al entrenamiento..... 52

Figura 4.5: Comportamiento de la precisión respecto a la recuperación en la fase de entrenamiento. 53

Figura 4.6: Accuracy sobre conjunto de imágenes de prueba, en DRIVE..... 54

Figura 4.7: F1-Score sobre conjunto de imágenes de prueba, en DRIVE..... 54

Figura 4.8: Curva de precisión y recuperación en conjunto de imágenes de prueba, en DRIVE. 55

Figura 4.9: Resultados sobre las imágenes de prueba en DRIVE. A imagen original, B imagen pre-procesada, C ground truth, d salida segmentada. 56

Figura 4.10: Pérdida asociada al entrenamiento de la arquitectura propuesta sobre CHASEDB..... 58

Figura 4.11: Accuracy asociado al entrenamiento de la arquitectura propuesta sobre CHASEDB..... 58

Figura 4.12: F1-Score asociado al entrenamiento de la arquitectura propuesta sobre CHASEDB..... 59

Figura 4.13: Comportamiento de la precisión respecto a la recuperación en la fase de entrenamiento sobre CHASEDB. 59

Figura 4.14: Accuracy sobre conjunto de imágenes de prueba, en CHASEDB. ... 60

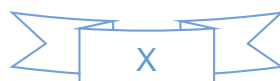


Figura 4.15: F1-Score sobre conjunto de imágenes de prueba, en CHASEDB. ...	61
Figura 4.16: Curva de precisión y recuperación en conjunto de imágenes de prueba, en CHASEDB.	62
Figura 4.17: Resultados sobre las imágenes de prueba en CHASEDB. A imagen original, B imagen pre-procesada, C ground truth, d salida segmentada.	63
Figura 4.18: Valor de SSIM entre el ground truth y el resultado de la U-Net 1.	64
Figura 4.19: Valor de SSIM entre el ground truth y el resultado de la arquitectura propuesta.	64
Figura 4.20: Porción de segmento, donde se pueden apreciar vasos sanguíneos tanto pequeños, medianos como grandes.	65
Figura 4.21: Resultado en imagen de prueba de DRIVE. A imagen original, B Imagen segmentada, C área de lesión, D área de la lesión en la imagen segmentada.	66

Dirección General de Bibliotecas UAQ

ÍNDICE DE TABLAS.

Tabla 1.1: Definiciones de varios autores de inteligencia artificial (Norvig, P. y Russell, 2014).	1
Tabla 2.1: Resumen de artículos para la segmentación de imágenes de retina. Los informes métricos fueron la precisión, la recuperación y el F1-score entre otros, para el conjunto de datos de DRIVE.	30
Tabla 4.1: Resultados de la experimentación donde se combinan la arquitectura del trabajo propuesto en (Khanal & Estrada, 2019), con la arquitectura propuesta en este trabajo.....	48
Tabla 4.2: Resultados con dos redes U-Net encadenadas y bloques residuales en la segunda.....	49
Tabla 4.3: Comparación con otros trabajos del estado del arte sobre la base de datos DRIVE.....	49
Tabla 4.4: Comparación respecto al tiempo de entrenamiento.	51
Tabla 4.5: Comparación con otros trabajos del estado del arte sobre CHASEDB.	57

1.INTRODUCCIÓN

1.1. Inteligencia Artificial

En los últimos años el crecimiento tecnológico ha tenido un avance vertiginoso. Muchas aplicaciones actuales hacen uso de sistemas de reconocimiento del habla, reconocimiento de imágenes, sistemas de diagnóstico médico, predicciones de fallas, detecciones de fraudes, entre muchas otras. Esta revolución tecnológica ha estado marcada principalmente por la Inteligencia Artificial (IA), la cual tiene varias definiciones, pero de manera general se relacionan con actividades que realizan las máquinas e imitan las funciones cognitivas que los humanos asocian con otras mentes humanas, como aprender o resolver problemas.

Por lo tanto, la IA no solo intenta comprender, sino que también se esfuerza en construir entidades inteligentes. En la siguiente imagen se pueden encontrar algunas definiciones de varios autores.

Tabla 1.1: Definiciones de varios autores de inteligencia artificial (Norvig, P. y Russell, 2014).

Sistemas que piensan como humanos	Sistemas que piensan racionalmente
<p>“El nuevo y excitante esfuerzo de hacer que los computadores piensen... máquinas con mentes, en el más amplio sentido literal.” (Haugeland, 1985)</p> <p>“[La automatización de] actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas, aprendizaje...” (Bellman, 1978)</p>	<p>“El estudio de las facultades mentales mediante el uso de modelos computacionales”. (Charniak y McDermott, 1985)</p> <p>“El estudio de los cálculos que hacen posible percibir, razonar y actuar”. (Winston, 1992)</p>
Sistemas que actúan como humanos	Sistemas que actúan racionalmente
<p>“El arte de desarrollar máquinas con capacidad para realizar funciones que cuando son realizadas por personas requieren de inteligencia”. (Kurzweil, 1990)</p> <p>“El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor”. (Rich y Knight, 1991)</p>	<p>“La Inteligencia Computacional es el estudio del diseño de agentes inteligentes”. (Poole et al., 1998)</p> <p>“IA... está relacionada con conductas inteligentes en artefactos”. (Nilsson, 1998)</p>

Históricamente hablando, el término ya tiene más de medio siglo, ya que desde la lejana década de 1950 aparecieron los primeros aportes significativos y conceptualizaciones. Sin embargo, es bueno reseñar que los primeros conceptos se relacionan con la antigua Grecia, específicamente con el filósofo Aristóteles, quien fue el primero en describir un conjunto de reglas que detallan una parte del funcionamiento de la mente para obtener conclusiones racionales. También se reporta, años después a Ctesibio de Alejandría, quien llegó a construir la primera máquina auto-controlada de manera racional, pero sin razonamiento.

Aunque ya la IA tiene sus años no es hasta esta última década donde ha habido una verdadera explosión y avances, debidos en gran medida al aumento de las capacidades de cómputo, así como de los grandes volúmenes de datos. La siguiente figura muestra en una línea de tiempo los principales eventos y descubrimientos que han tenido lugar.

Historia reciente de la IA

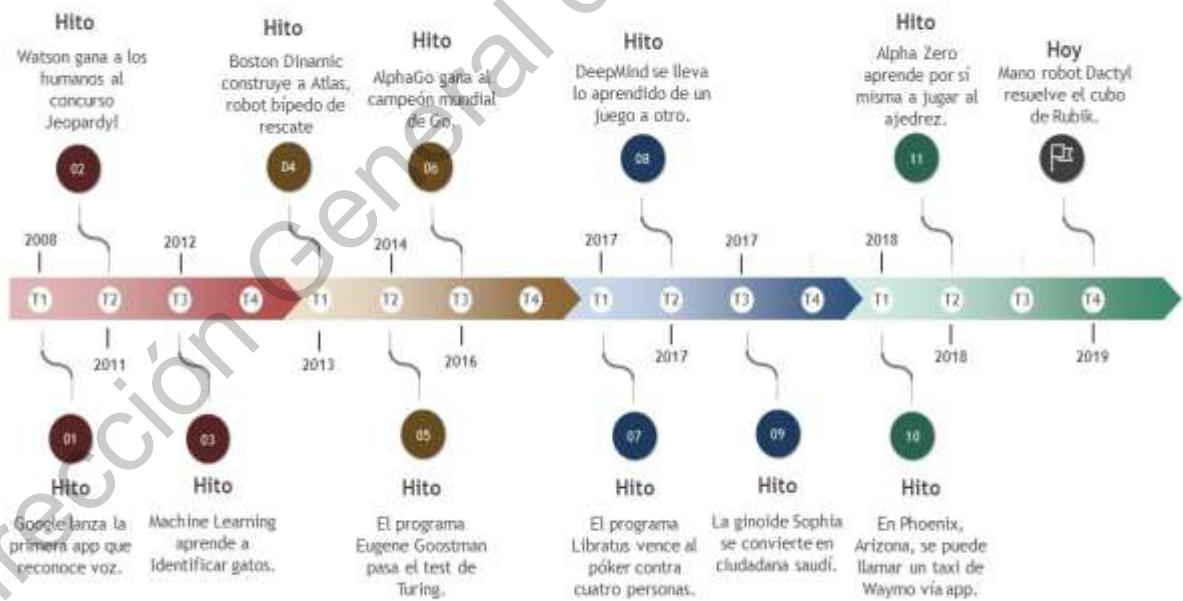


Figura 1.1: Línea de tiempo de los avances en la IA dentro de la última década aproximadamente. (Elaboración propia)

Aunque no se pueda apreciar, todos estos hitos, en la figura 1.1 previamente mostrada, tienen un elemento en común y no es otro que el uso de redes neuronales artificiales, las cuales se pueden definir como un modelo matemático para procesamiento de información, implementado por hardware o por software y que tiene una entidad propia expresada en un conjunto de parámetros internos, arquitecturas y modelos que la diferencian de otras técnicas y que reúne algunas características comunes con las redes neuronales biológicas.

Estas redes neuronales artificiales tienen su origen en el modelo de Perceptrón propuesto por Frank Rosenblatt (Rosenblatt, 1958). Sin embargo, su verdadero potencial se ha alcanzado hasta ahora a través de las redes neuronales profundas o *Deep Learning*, la cual se diferencia de la red neuronal tradicional en que presenta múltiples capas ocultas no lineales capaces de aprender diferentes representaciones de los datos sin la necesidad de supervisión de un humano. En otras palabras, no requieren de un experto, este tipo de red es capaz de aprender la representación de los datos para una tarea posterior, ya sea de clasificación o regresión.

1.2. Enfermedades cardiovasculares

Las enfermedades cardiovasculares (ECV) son un conjunto de patologías interrelacionadas, que incluyen entre otras las enfermedades coronarias, enfermedades cerebrovasculares, enfermedades arteriales periféricas y cardíacas congénitas, constituyendo en general la tercera parte de las muertes a nivel mundial (Stewart, et al., 2017). Entre los factores que provocan las ECV se encuentran la condición de fumador, la hipertensión, diabetes y la obesidad; así como buenos métodos de prevención son los ejercicios físicos y dietas basadas en frutas y vegetales.

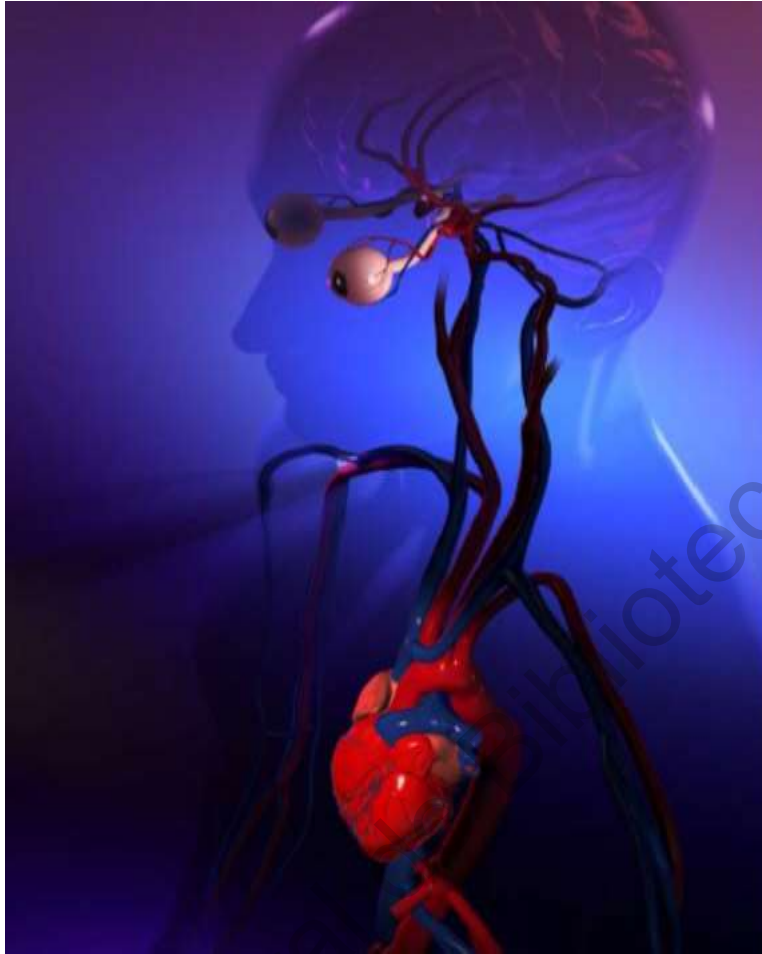


Figura 1.2: Vasculatura entre corazón y ojos (Heart & Eye, 2015).

Estudios recientes han mostrado una relación entre enfermedades oftalmológicas y enfermedades sistémicas, especialmente, las ECV. Los vasos sanguíneos de la retina les brindan a los oftalmólogos información acerca del estado de los vasos sanguíneos del cerebro o en los pequeños vasos sanguíneos del corazón.

Cambios en la micro-vasculatura, como por ejemplo aterosclerosis y cruces arterio-venosos, son efectos de la presión sanguínea y otros factores sistémicos. Eventos isquémicos como émbolos y oclusiones también pueden ser apreciados en las imágenes de retina (Heart & Eye, 2015).

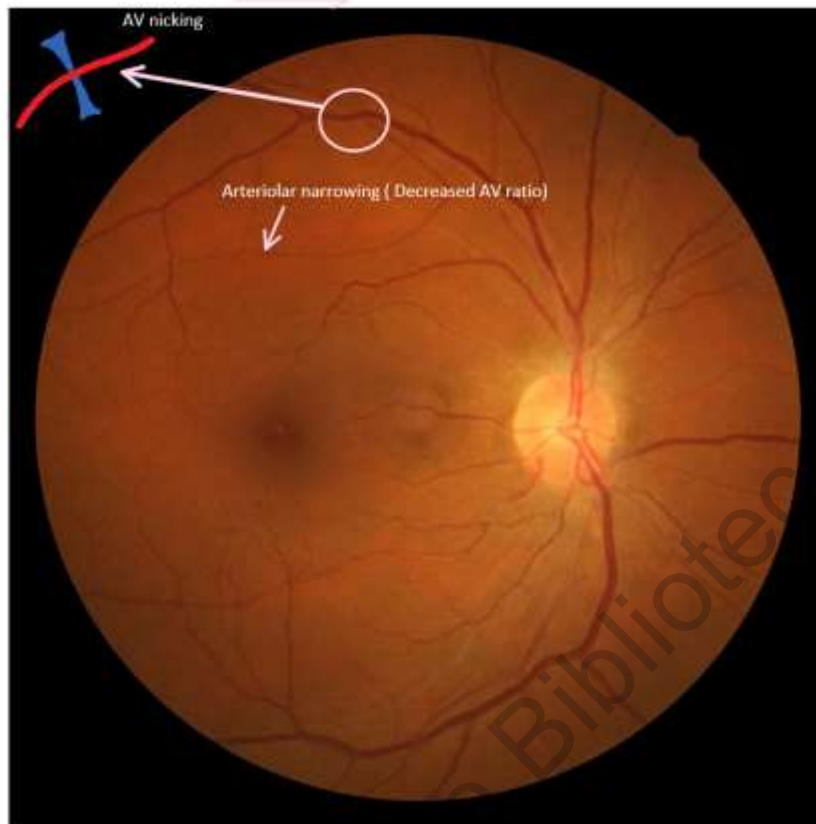


Figura 1.3: Imagen de retina con estrechamiento arterial y entrecruzamiento artereo-venar (Pranav Modi, 2020).

1.3. Justificación del problema

En México, las enfermedades no transmisibles, incluyendo las cardiovasculares constituyen el 77% de las muertes de adultos, de ellas, el 24% de los decesos corresponden a ECV, siendo algunos de los factores de riesgos:

- 17% de la población fumadora.
- 7.2 litros de alcohol puro consumido por persona.
- 22.8% tienen hipertensión, que puede aumentar el riesgo de ataque o insuficiencia cardiaca.
- 32.1% de los adultos de México son obesos (World Heart Federation, 2016).

Evidentemente el país toma acciones, como la vigente desde 2014 donde se introdujo un nuevo impuesto a los alimentos con un contenido energético superior a

275 Kcal por 100 gramos, así como a las bebidas endulzadas con azúcar. La OMS también exhorta a los gobiernos, incluido el de México a combatir el escenario negativo, con estrategias como (World Heart Federation, 2016):

- Políticas de control del tabaquismo.
- Impuesto a alimentos con altos contenidos de grasas.
- Construcción de vías peatonales y carriles para bicicletas, con el fin de promover las actividades físicas.
- Reducir el consumo nocivo de alcohol.

Aun con estos programas, las ECV constituyen la principal causa de muerte, por lo que aplicar otras medidas que contribuyan a mejorar la calidad y esperanza de vida de un paciente es siempre bienvenida. Se propone por tanto una solución no invasiva, porque se basa en la detección de anomalías y patrones de riesgos en imágenes de retinas a través de técnicas de aprendizaje profundo.

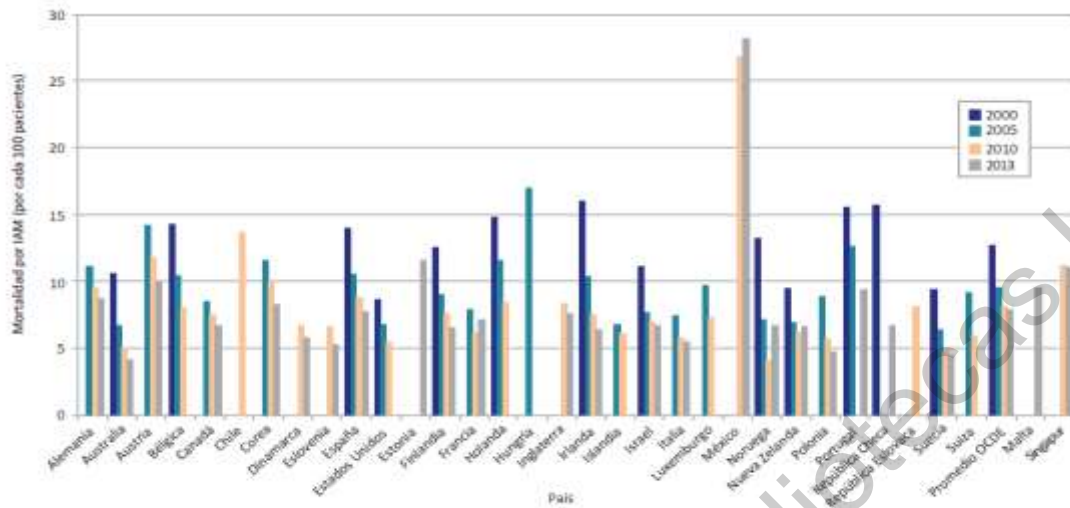
Con el uso de esta técnica se pretende dotar a los modelos predictivos de mejoras, que puedan ser de ayuda para los profesionales de la medicina. Además, su uso, a través de las redes neuronales convolutivas han mostrado resultados prometedores, con una alta precisión en el diagnóstico de enfermedades, comparadas a los expertos, por lo que aplicar algoritmos de IA, resulta factible, incluso barato, independientemente del costo computacional.

1.4. Descripción del problema

Las ECV siguen siendo la causa de muerte número uno en México, siendo la presión arterial, la diabetes mellitus y la obesidad los principales factores de riesgo, y dando un estimado de 300 personas fallecidas diarias.

Estas cifras se explican a través de la Organización para la Cooperación y el Desarrollo Económico (OCDE), donde dentro de los países pertenecientes a dicha organización, México presentó un promedio cuatro veces mayor al resto en la tasa

de mortalidad a 30 días por infarto agudo de miocardio, la cual está en 7.5 muertes por cada 100 admisiones hospitalarias (García, et al., 2017).



Fuente: OECD Health Statistics 2015 <http://dx.doi.org/10.1787/health-data-en>.

Figura 1.4: Mortalidad por infarto agudo al miocardio según OCDE ("OECD iLibrary | OECD Health Statistics,"2015).

Estudios revelan que la mayoría de los hospitales no están equipados para tratar estos casos y que los retrasos en los tratamientos se deben, en parte, al desconocimiento sobre el estado de salud de los pacientes.

Debido a que las ECV incluyen entre otras afectaciones, daños en las arterias, fijarnos sistemáticamente en la evolución de las mismas, nos puede dar un patrón de riesgo en cada persona.

La retina es el único lugar del cuerpo humano donde se pueden apreciar las venas y arterias directamente, lo que nos brinda la oportunidad de apreciar la estructura y patología en vivo de la circulación humana, permitiendo detectar cambios microvasculares relacionados con el desarrollo de enfermedades cardiovasculares. (Ting & Wong, 2018).

Con los avances en las tecnologías relacionadas con la fotografía digital y el procesamiento de imágenes, se ha hecho posible una medición cuantitativa de los distintos cambios que ocurren en los micro-vasos sanguíneos antes mencionados en la retina, por lo que identificar micro-aneurismas, hemorragias aisladas,

estrechamiento arteriolar focal y el cruce arterio-venoso sin duda pueden contribuir a identificar factores de riesgo cardiovascular.

Otros factores de riesgo cardiovascular pueden ser obtenidos a través de imágenes de retina como son la edad de una persona, el género, si es fumador o no, si es diabético y las presiones arteriales (sistólica y diastólica). Estos patrones pueden ser detectados en la retina a través de características como la degeneración de la mácula, daños en la cabeza del nervio óptico, así como cambios en las paredes de arterias y venas. Con respecto a los cambios en la red vascular se procederá con la segmentación de la misma.

En la actualidad el uso de redes profundas ha mostrado avances significados y de resultados comprobados en la detección de patrones en imágenes, pero para su correcto funcionamiento es necesario el uso de gran cantidad de información, en el caso que nos ocupa un amplio volumen de imágenes de retina.

Esta gran cantidad de imágenes es necesaria para entrenar la red de aprendizaje profundo propuesta. Ya una vez obtenida, la misma se prueba con nuevas imágenes etiquetadas, lo que permite comprobar la calidad del proceso de detección de patrones llevado a cabo, arrojando un porcentaje de presencia de factores de riesgo cardiovascular, lo cual puede ser útil como alerta temprana de un evento de este tipo.

Otra situación a tener en cuenta es la capacidad de cómputo, ya que para entrenar una red de aprendizaje profundo son necesario dispositivos con altas capacidades de hardware para lograr un entrenamiento en un tiempo de ejecución aceptable.

1.5. Objetivo general

Desarrollar un algoritmo que permita la identificación de patrones de riesgo cardiovascular en imágenes de retina a través de redes neuronales profundas.

1.6. Objetivos específicos

- Realizar análisis y diseño de un modelo de red neuronal que permita la detección de patrones de riesgo cardiovascular.
- Implementar el procedimiento algorítmico que permita la detección de patrones de riesgo cardiovascular.
- Establecer las técnicas y métricas para probar y validar el modelo propuesto contra otras técnicas de aprendizaje automático u otras técnicas de procesamiento digital de imágenes.

1.7. Hipótesis

Con la implementación de un modelo de red neuronal para la detección de patrones de riesgo cardiovascular, a través de técnicas de aprendizaje profundo, se logrará mayor precisión y fiabilidad en los modelos predictivos de afectaciones cardiovasculares.

1.8. Estrategia de investigación

La estrategia de investigación utilizada es la exploratoria ya que este tipo de investigación se realiza cuando existe una problemática que está afectando la sociedad y no se tiene una idea clara del asunto en cuestión. Su principal objetivo es familiarizar al investigador con el tema objeto de estudio, la situación en que se encuentra y los métodos y técnicas utilizados en su ejecución. La fuente de información más apropiada es la acumulada en la bibliografía existente y la práctica que se obtiene de los sujetos que están vinculados a la problemática que se investiga.

1.9. Métodos científicos

Los métodos científicos utilizados en esta investigación fueron:

- **Histórico-Lógico:** Permite analizar la trayectoria completa del fenómeno de detección de patrones cardiovasculares, que se llevaba a cabo en las imágenes de retina; así como el estudio histórico del mismo que permite ver deficiencias y proponer soluciones acordes a las necesidades.
- **Hipotético-Deductivo:** A partir de la hipótesis y siguiendo la lógica de deducción tomada se llega a nuevos conocimientos y predicciones, que son sometidos a verificaciones.
- **Analítico - Sintético:** Mediante este método se va a analizar toda la teoría recopilada a través de los diferentes medios bibliográficos que pueda servir para desarrollar mejor el diseño del algoritmo, y poder aplicar así estos conocimientos en la práctica de manera que se adquiriera una mayor preparación sobre el tema en cuestión.
- **Modelación:** Método mediante el cual se crean abstracciones con el objetivo de explicar la realidad. El modelo como sustituto del objeto de investigación es semejante a él, existiendo una correspondencia objetiva entre el modelo y el objeto.

1.10. Estructura de la tesis

El presente documento cuenta con 3 capítulos:

Capítulo 1 - Marco Teórico: Se describen los conceptos fundamentales asociados al dominio del problema y el objeto de estudio, haciéndose un análisis de la situación actual y revisando el estado del arte de trabajos similares. Se presenta la fundamentación de las tecnologías utilizadas para el diseño del sistema y las propuestas para su implementación y desarrollo.

Capítulo 2 - Metodología: Se describen los pasos a seguir en la propuesta del modelo a desarrollar, para ello se estudiaron los principales procesos vinculados al objeto de estudio y al campo de acción del trabajo.

Capítulo 3 - Resultados y Evaluación: Se muestran los resultados alcanzados al implementar la metodología y diseños propuestos. Se realiza la evaluación y comparación con respecto a trabajos similares.

2. MARCO TEÓRICO

2.1. Revisión médica

2.1.1. Imágenes de retina

La retina es la capa más interna del globo ocular. Su misión es transformar la luz que recibe en un impulso nervioso que viaja hasta el cerebro a través del nervio óptico, y se convierte en las imágenes que percibimos. Entre los elementos que podemos identificar en ella se encuentran la red vascular, la mácula y fovea y el disco óptico, la figura 2.1 a continuación, muestra la localización de estos elementos. La retina es el único lugar del cuerpo humano donde se puede apreciar en vivo la estructura de las arterias y venas de forma no invasiva y de esta forma monitorear los cambios que puedan ocurrir en la micro-vasculatura relacionada a ECV.

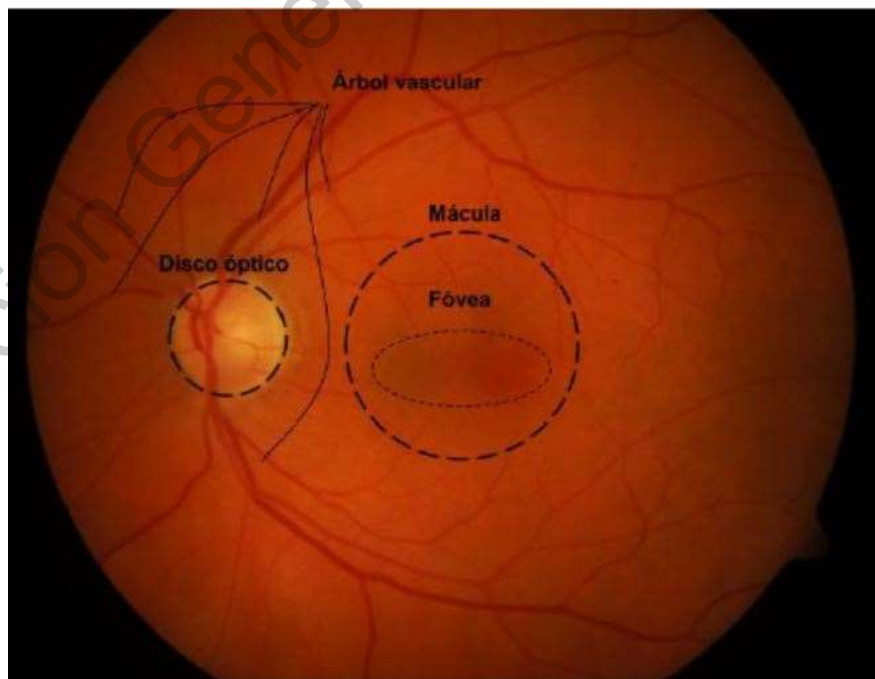


Figura 2.1: Imagen de retina y sus principales componentes (Cristobal Mollá, 2017).

Importantes enfermedades que pueden afectar la vasculatura retiniana, así como el tejido de la misma son la retinopatía diabética, el glaucoma, degeneración macular asociada a la edad y Alzheimer.

Los últimos adelantos en la tecnología, específicamente en la fotografía digital de alta resolución y en procesamiento de imágenes han permitido una medición cuantitativa y reproducible de los distintos cambios que se dan en los vasos de la retina.

Estudios histopatológicos han demostrado que estas manifestaciones retinianas reflejan el daño vascular por envejecimiento, hipertensión y otros procesos, y estudios patológicos también indican que las manifestaciones de la retina se relacionan estrechamente con afecciones micro-vasculares de otros órganos (por ejemplo, en los hipertensos, se estrecha la arteriola retiniana, su media se engrosa y aparece esclerosis). De forma paralela, se han observado cambios escleróticos similares en pequeñas arteriolas dentro del miocardio, que en presencia de hipertensión muestran un estrechamiento luminal como el de la retina. El aumento de la relación entre la media y la luz de las arterias en la grasa subcutánea es un factor independiente predictivo del riesgo de episodios de enfermedad cardiovascular, como el infarto de miocardio (Liew Gerald, 2015).

Por tanto, hay razones anatómicas, fisiológicas y patológicas que indican que los cambios en la micro-vasculatura retiniana pueden ser indicadores útiles de enfermedades estructurales vasculares de la microcirculación coronaria y que la evaluación no invasiva de la retina puede ayudar a estratificar el riesgo de cardiopatía coronaria (Liew Gerald, 2015).

2.1.2. La retina y los factores de riesgos cardiovasculares.

En la retina se pueden encontrar patrones relacionados a ECV, un ejemplo son las hemorragias que constituyen un desorden del ojo en el cual ocurre un sangrado en el tejido de la misma (Wankhade, 2016). Estos pueden estar relacionados a la

retinopatía diabética, que causa la formación de pequeños y frágiles vasos sanguíneos, los cuales pueden ser dañados por altos niveles de azúcar en la sangre, causando potencialmente el crecimiento o creación de nuevos de estos (Laibacher, et al., 2018).

Otra enfermedad relacionada a la vasculatura retiniana es la retinopatía hipertensiva, que consiste en daño a la retina provocada por alta presión sanguínea, que puede derivar en la deformación de los vasos.

Por distintos estudios se sabe que las manifestaciones vasculares retinianas están relacionadas con la hipertensión arterial crónica y los marcadores sistémicos de inflamación y alteración de la función endotelial. El estrechamiento de las arteriolas retinianas se asocia significativamente con la hipertensión arterial (HTA) ambiental y en menor medida, con los valores previos de HTA. Por el contrario, la dilatación de las vénulas retinianas puede ser un indicador de hipoxia cerebral, alteración de la función endotelial, hiperglucemia e inflamación. Así mismo, las lesiones de retinopatía se han asociado con hiperglucemia, hipertensión, alteración endotelial e inflamación.

La hipertensión puede afectar los ojos de diferentes formas, algunos ejemplos son el estrechamiento focal artereolar, el cruzamiento artereovenoso, las hemorragias, los micro-aneurismas y los exudados duros (Wong & Mitchell, 2007). Estas manifestaciones en los vasos sanguíneos pueden dar detalles de distintos procesos en enfermedades vasculares y explicar por qué algunas de estas manifestaciones se relacionan con cardiopatías coronarias clínicas.

2.1.3. Segmentación de vasos sanguíneos

El diagnóstico asistido por computadoras (CAD), por sus siglas en inglés, son esenciales en el análisis de imágenes médicas, porque facilitan la identificación de anormalidades que necesitan tratamientos, además de que brindan soporte a los especialistas clínicos, en el procesamiento de grandes volúmenes de datos en un corto período de tiempo.

Gracias a los avances en la visión por computadora y el aprendizaje automático, específicamente el aprendizaje profundo, es posible tener en la actualidad un mejor entendimiento e incluso descubrir nuevas señales de afectaciones en esos grandes volúmenes de imágenes médicas.

Muchos CAD han sido desarrollados para procesar imágenes de retina, donde se detectan estructuras como los vasos sanguíneos, se segmenta el disco óptico y varias lesiones que predicen enfermedades como la retinopatía diabética (D'Agostino et al., 2008), el glaucoma y la degeneración macular asociada a la edad.

Para muchas de estas enfermedades, es crucial analizar la red vascular de la retina, por lo que lograr la extracción de los vasos sanguíneos puede ser una tarea de suma importancia en la detección de anomalías relacionadas con dichas enfermedades.

La segmentación de los vasos sanguíneos de la retina puede darnos importantes pistas en el diagnóstico de enfermedades oftalmológicas como micro-aneurismas, arteriosclerosis e hipertensión. Entre las propiedades que se pueden extraer de dicha segmentación están el ancho de los vasos, la dimensión fractal, la tortuosidad de los vasos, el radio artereo-venoso, las bifurcaciones asimétricas, entre otras. También es de suma importancia detectar con mucha precisión los pequeños vasos sanguíneos, porque estos nos pueden llevar a detectar enfermedades en sus estados iniciales. Sin embargo, esta es una tarea que requiere mucho tiempo de los especialistas, ya que realizan la misma de forma manual, lo que trae consigo que haya mayores errores humanos, incluso dentro de los especialistas más experimentados. Además, el proceso de segmentación se puede ver afectado por ruidos, como regiones de bajo contraste, luminosidad no uniforme y cruce de vasos entre otras dificultades (Sureshjani, 2017).

En el presente trabajo nos centraremos en la segmentación de vasos sanguíneos en imágenes de retina, ya que esta actividad constituye el punto de partida para la detección de muchas afectaciones oftalmológicas y sus derivados en ECV.

2.2. Revisión tecnológica.

Para darle solución al problema planteado se pretende incursionar en áreas del aprendizaje profundo, una forma de inteligencia artificial que emplea matemáticas avanzadas y sistemas computacionales para revelar información de los datos, en el caso que nos ocupa, imágenes médicas, generalmente con el propósito de predecir y descubrir nuevo conocimiento. Debido a sus reconocidos avances en el área, se trabajará con redes neuronales profundas, que ha generado excelentes resultados en la clasificación y detección de patrones.

2.2.1. Red neuronal.

Una red neuronal es un conjunto de neuronas o unidades con una función de activación sobre los parámetros de entrada multiplicados por sus pesos, además de sumados a un elemento bias, esto representa una combinación lineal de las entradas y sus parámetros, al que se le agrega un elemento de no linealidad, referido a una función de transferencia. La representación matemática es la siguiente:

$$\partial = \sigma(W^T x + b) \quad (1)$$

Donde W es la matriz de pesos, x las unidades de entrada y b el bias, σ representa la función de transferencia que tradicionalmente son las sigmoides cuya ecuación es

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

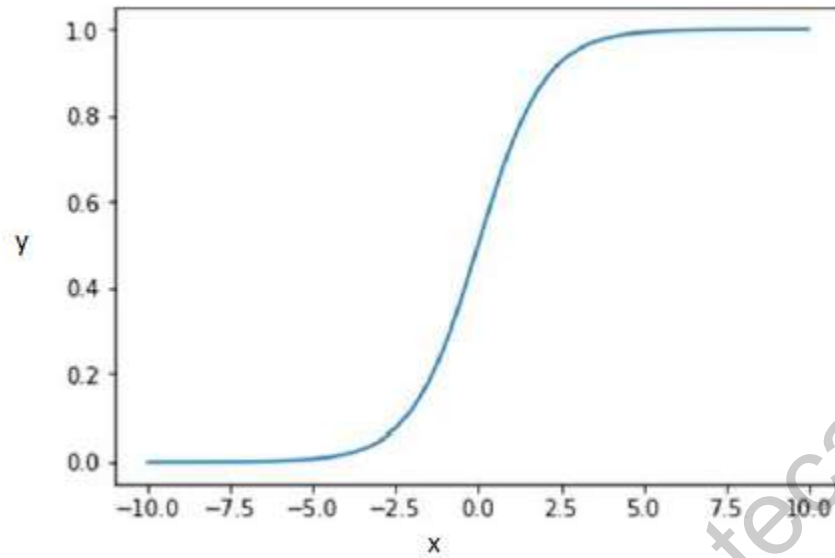


Figura 2.2: Gráfica de función sigmoide.

y la tangente hiperbólica (Bonifacio, et. al., 2006)

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (3)$$

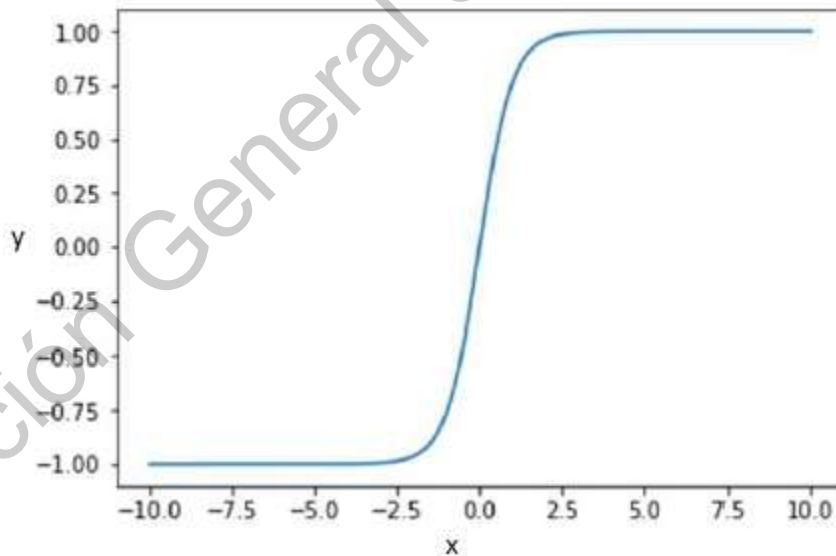


Figura 2.3: Gráfica de función tangente hiperbólica.

La estructura tradicional de las redes neuronales con capas ocultas es la que se muestra en la siguiente figura, donde hay una capa de entrada, una capa oculta y una capa de salida.

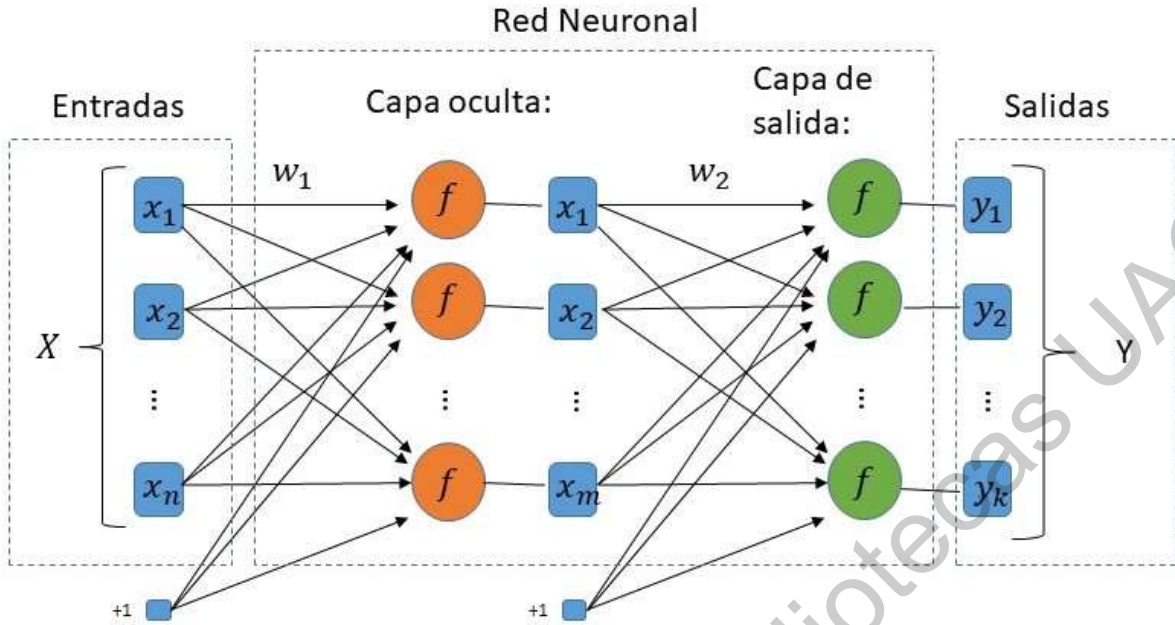


Figura 2.4: Topología de una red neuronal con una capa oculta (Adaptada de referencia).

Donde:

f : función de transferencia que se aplica.

(x_1, \dots, x_n) : vector extendido de la capa de entrada.

(x_1, \dots, x_m) : vector extendido de la capa oculta.

(y_1, \dots, y_k) : vector extendido de la capa de salida, que se aproxima al valor deseado o esperado.

2.2.2. Red neuronal profunda o Deep Learning

Es cuando una red neuronal tiene múltiples capas ocultas, las cuales se encargan de resolver un problema en específico cada una de ellas y van creciendo en complejidad y abstracción.

Deep Learning, el aprendizaje de máquinas más utilizado como método de extracción y transformación de datos, que consiste en una transformación no lineal de una red neuronal de capas interconectadas, ha provocado tremendo interés en la medicina, en particular en las aplicaciones de imágenes como la detección de la retinopatía diabética y enfermedades retinianas relacionadas, tuberculosis,

melanoma y metástasis en los ganglios linfáticos secundaria al cáncer de mama. Tal interés proviene de mejoras significativas en los modelos matemáticos subyacentes y en los datos de la disponibilidad, así como los resultados sólidos obtenidos. De hecho, el aprendizaje profundo ha logrado rendimientos comparables, si no superiores, a los estándares actuales de atención (típicamente, evaluación por expertos humanos) (Ting & Wong, 2018).

Para el trabajo con redes neuronales profundas hasta la actualidad se ha comprobado que la función de activación, unidad lineal rectificadora (*ReLU*, por sus siglas en inglés) ha brindado el mejor desempeño, ya que ha obtenido resultados de menor error comparada con las funciones logísticas antes mencionadas (Goodfellow, et. al., 2015).

$$f(x) = \max(0, x) \quad (4)$$

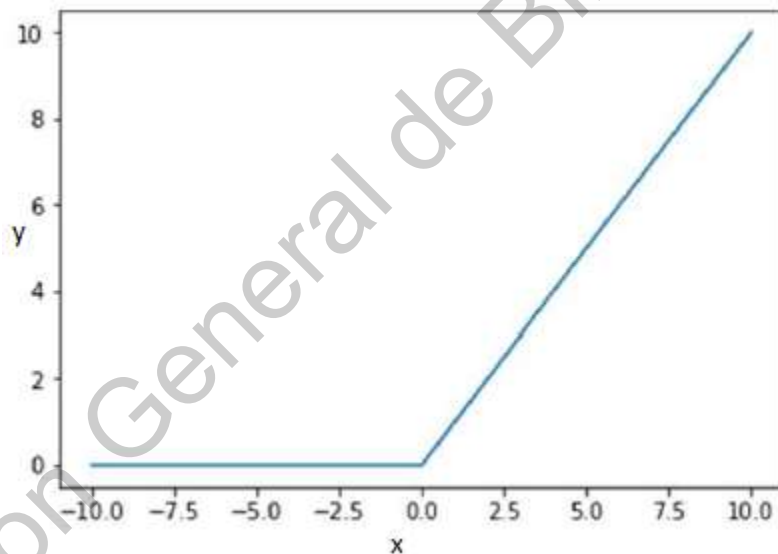


Figura 2.5: Gráfica de función *ReLU* ("A Gentle Introduction to the Rectified Linear Unit (ReLU)," n.d.).

Una variación de esta función de activación, con la cual también se experimentó en la presente investigación es *Leaky ReLU*, en la que se establece un parámetro λ en un rango entre (0, 1) para comprimir la parte negativa en dicho intervalo. Su ecuación queda representada de la siguiente forma:

$$f(x) = \max(x, 0) + \lambda \min(x, 0) \quad (5)$$

La siguiente figura muestra su gráfica:

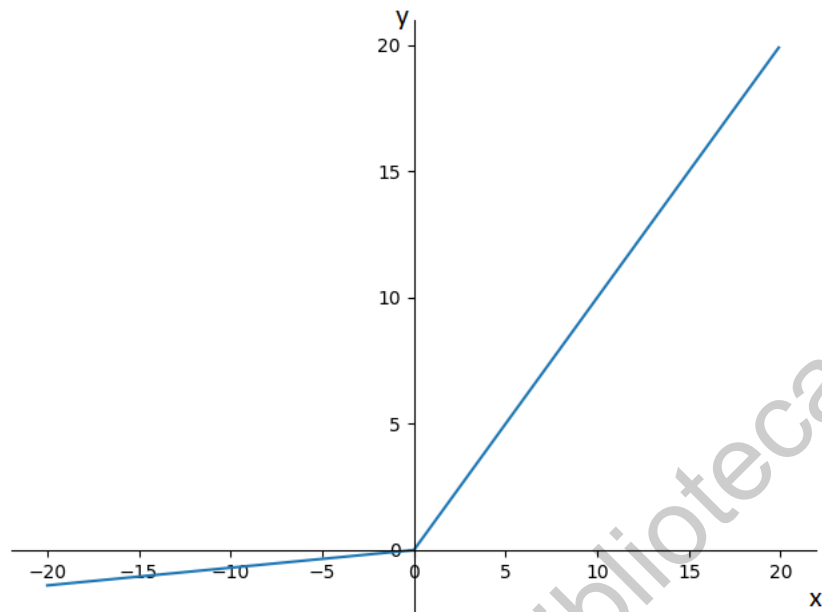


Figura 2.6: Gráfica de la función Leaky ReLU ("Machine Learning Note: ReLU function - Clay-Technology World," n.d.).

El aprendizaje profundo es utilizado en imágenes médicas para clasificación, detección, segmentación, registro entre otras actividades y contribuye de manera activa en regiones anatómicas como el cerebro, el pecho, mamas, abdomen, análisis cardíacos, análisis musculoesqueléticos y en los ojos donde la mayoría de las investigaciones emplean redes neuronales convolutivas, para el análisis de imágenes a color de fondo de ojos y muchas aplicaciones están enfocadas a la segmentación de estructuras anatómicas, segmentación y detección de anomalías en la retina, al diagnóstico de enfermedades en los ojos y a la evaluación de la calidad de las imágenes.

Por muchos años las redes neuronales profundas fueron consideradas difíciles de entrenar eficientemente hasta que en 2006 los investigadores Bengio, Hilton y Salakhudinov mostraron un entrenamiento capa a capa con un enfoque no supervisado, seguido de un ajuste supervisado sobre una red apilada, la cual mostró un buen rendimiento. Actualmente las arquitecturas más populares son las redes neuronales convolutivas y las redes neuronales recurrentes, con muy buenos desempeños, sobre todo la primera, en el procesamiento de imágenes médicas,

este tipo de red será descrita a continuación. Otras arquitecturas también usadas en imágenes médicas son las *auto-encoder*, las máquinas de Boltzmann restringidas, redes neuronales convolutivas *multi-stream* y U-Net (Litjens et al., 2017).

2.2.2.1. Redes neuronales convolucionales (CNNs).

El objetivo de las CNNs es aprender características de orden superior en los datos a través de convoluciones y su uso ha sido bien aceptado en el reconocimiento de objetos, así como en competiciones de clasificación de imágenes.

La eficiencia de las CNN para el reconocimiento de imágenes es la que ha posicionado al Deep Learning como una técnica de vanguardia, las CNNs son buenas para construir y detectar las posiciones, así como las características invariantes a pesar de la rotación en datos de imágenes (Josh Patterson & Adam Gibson, 2017).



Figura 2.7: Red neuronal convolucional (Josh Patterson & Adam Gibson, 2017).

El uso de redes neuronales convolutivas se justifica debido a su modelo, donde la red compuesta por neuronas convolucionales extrae características de las imágenes suministradas, a través de operaciones de convolución usando filtros

cuyos pesos se aprenden automáticamente durante el entrenamiento, además de que reduce su dimensionalidad, comportamiento parecido al de la corteza visual primaria de nuestro cerebro biológico.

Esta capacidad puede ser controlada variando su profundidad y anchura logrando obtener buenas suposiciones sobre la naturaleza de las imágenes (por ejemplo, la localidad de la dependencia de los píxeles). Por lo tanto, en comparación con las redes neuronales de retroalimentación estándar con capas de tamaño similar, las redes neuronales convolutivas tienen muchas menos conexiones y parámetros, siendo más fáciles de entrenar, mientras que su rendimiento teóricamente óptimo probablemente sea sólo ligeramente peor (Krizhevsky, et. al., 2012).

- **Capa convolucional:**

Las computadoras leen las imágenes como píxeles y se expresan como una matriz de $N \times N \times 3$. La capa convolucional utiliza un conjunto de filtros para detectar la presencia de características o patrones específicos presentes en la imagen original de entrada, el filtro se desliza por la imagen y se calcula un producto de puntos para obtener un mapa de activación, posteriormente se aplica la función de activación *ReLU* para romper la linealidad de la imagen.

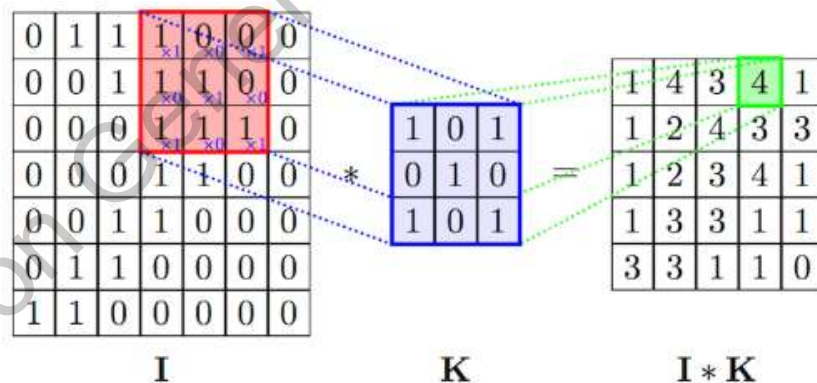


Figura 2.8: Operación de convolución entre imagen de entrada (I), el filtro aplicado (K) y el resultado obtenido de la operación (“asymptote - Drawing a convolution with Tikz - TeX - LaTeX Stack Exchange,” n.d.).

- **Capa pooling:**

Se encarga de reducir la cantidad de parámetros y el cálculo de la red, controlando el sobreajuste y reduciendo progresivamente el tamaño espacial de la red, la técnica utilizada aquí comúnmente es *max-pooling*. Esta capa utiliza

filtros, normalmente de 2x2, para reducir el volumen de la imagen de entrada. Esta operación toma el mayor número de la región del filtro seleccionado, esta operación no afecta la profundidad.

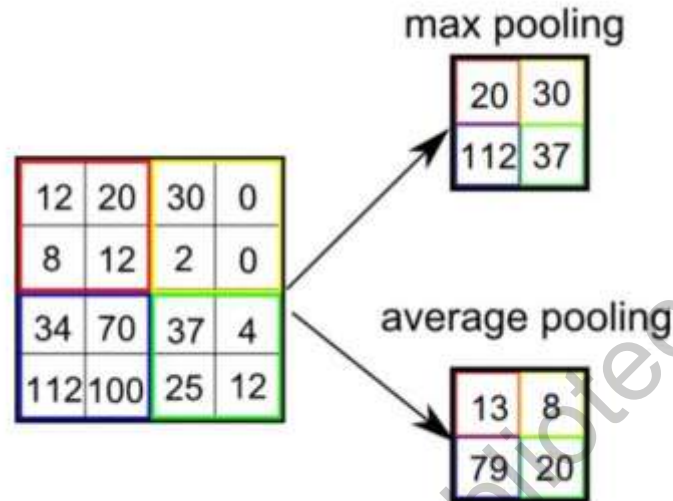


Figura 2.9: Operaciones de pooling, normalmente son dos, como se pueden apreciar en la imagen ("A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," 2018).

- **Capa completamente conectada:**
Esta capa se utiliza para calcular los puntos para cada clase, la cual usaremos como salida de la red. La dimensión del volumen de salida es de $[1 \times 1 \times N]$, donde N es el número de clases de salida que estaremos evaluando.

2.2.2.2. U-Net

Las CNNs trabajan bien en tareas donde una imagen de entrada es convertida en un vector de salida, el cual es usado para una futura clasificación. Sin embargo, en la segmentación de imágenes, no solo necesitamos convertir la imagen en un vector, sino reconstruir la imagen desde este vector, y esta es precisamente la tarea del modelo U-Net.

La arquitectura luce como una U, de ahí el nombre de la misma, que se puede apreciar en la siguiente figura:

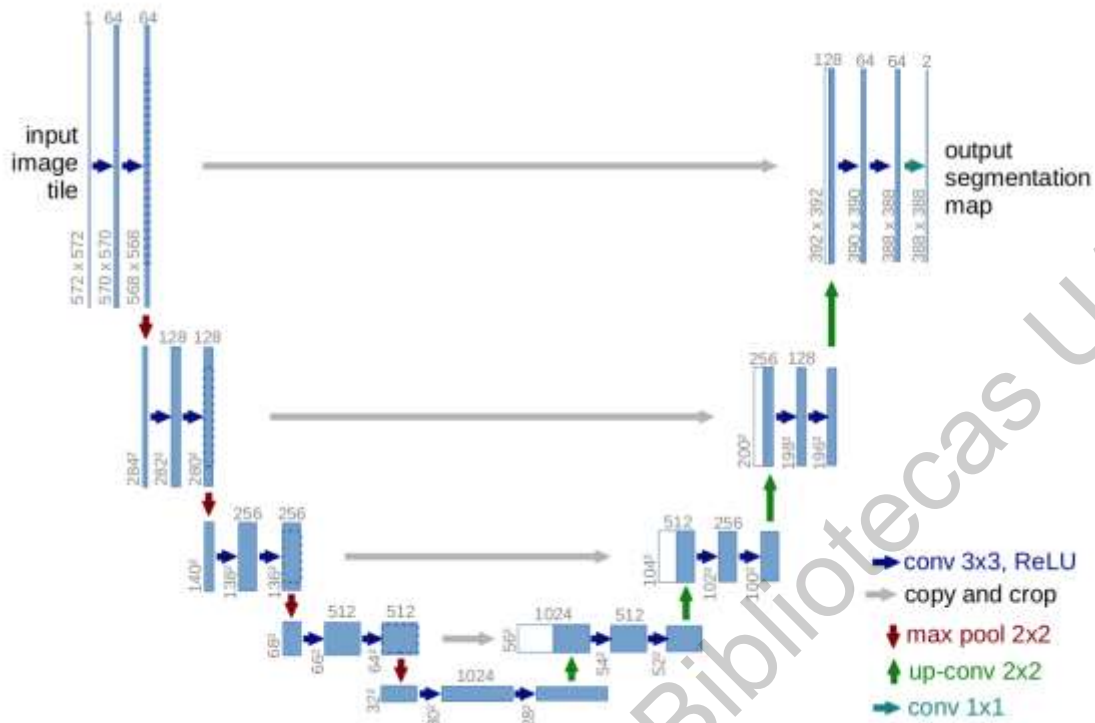


Figura 2.10: Arquitectura de red U-Net (Ronneberger, et. al., 2015).

La misma está compuesta de 3 partes fundamentales:

- **Contracción:**
Está compuesta de varios bloques de contracción; cada uno de ellos le aplica a la entrada que recibe dos capas de convolución de 3x3 seguida de una capa de 2x2 de *max pooling*. De esta forma el número de mapas de características después de cada bloque se duplica, permitiendo aprender efectivamente estructuras más complejas.
- **Bottleneck:**
Es la capa que media entre las capas de contracción y expansión, está compuesta de dos capas de convolución de 3x3 y una capa de 2x2 *upsampling*.
- **Expansión:**
Similar a la capa de contracción, está compuesta de varias capas de expansión, donde también a cada entrada que recibe se le aplica una doble convolución de 3x3, así como una capa de 2x2 *upsampling*.

Como elemento clave de esta arquitectura, ocurre una transmisión de información entre cada nivel de contracción y expansión, de esta forma se asegura que las características aprendidas durante la primera fase, sean incorporadas como conocimiento en la segunda y no ocurra una pérdida de información significativa.

Por último, el resultado obtenido se pasa por una capa convolucional de 3x3 con el número del mapa de características igual al número de clases segmentadas deseadas (Ronneberger, et. al., 2015).

2.2.2.3. Redes residuales.

En el trabajo con redes neuronales profundas se tenía una idea inicial de que mientras más capas más características complejas se detectarían. Sin embargo, en (He, et. al., 2015), se demostró que no siempre se cumple la tendencia antes planteada en referencia a la profundidad de una red. Ver figura a continuación:

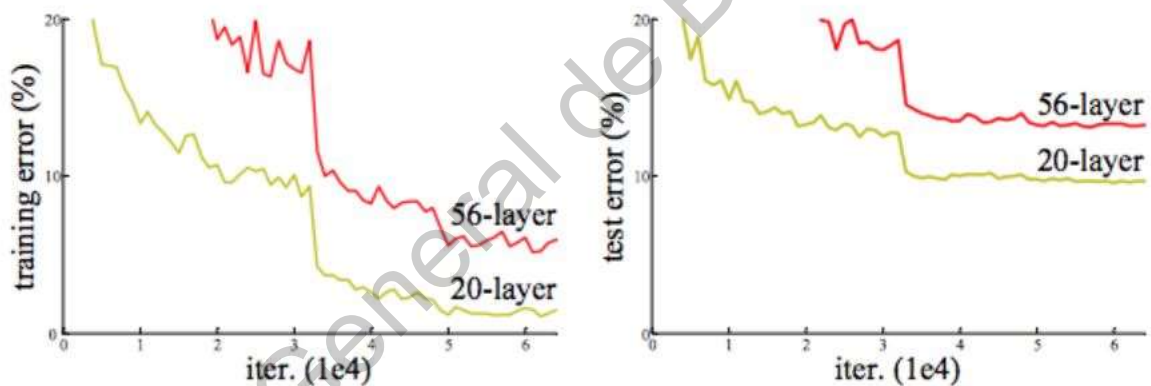


Figura 2.11: Tendencia del error, tanto en el entrenamiento como en las pruebas, para dos modelos, uno con 20 capas y otro con 56 capas (He, et. al., 2015).

Para atenuar esta situación un nuevo enfoque fue propuesto, la introducción de bloques residuales como capas, la siguiente imagen muestra su estructura:

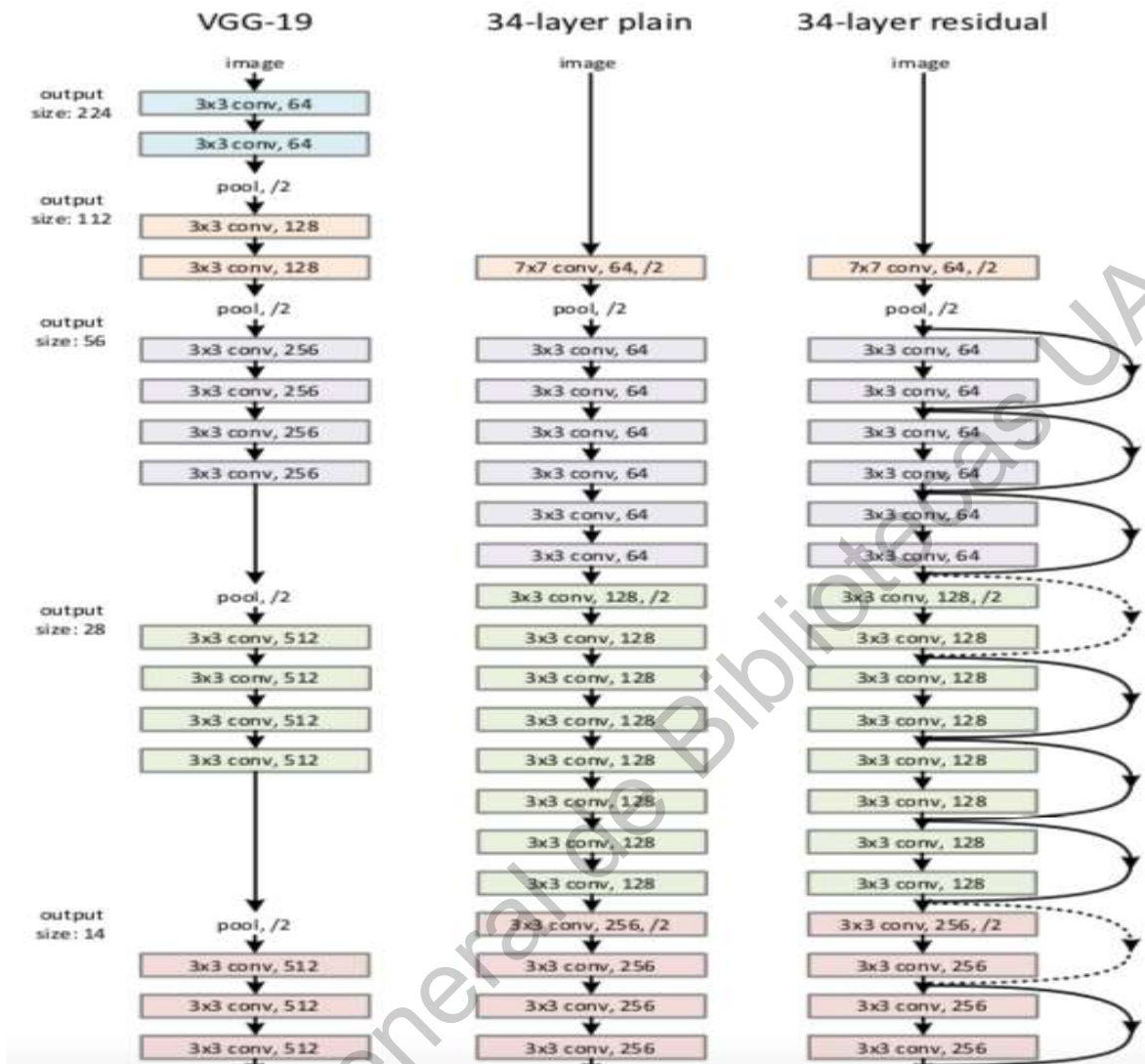


Figura 2.12: Arquitectura Res-Net (He et. al., 2015).

El principal aporte de esta innovación consiste en adicionar una operación de identidad, a través de bloques residuales, al final de las operaciones de convolución que se realizan.

Esta operación, conocida como conexión residual, no añade más parámetros al modelo y facilita la optimización del mismo. En el trabajo aquí propuesto se hará uso de conexiones residuales, lo cual atenuará el problema de degradación que surgen en este tipo de modelos.

2.3. Materiales y herramientas.

2.3.1. Python.

Python es un lenguaje de programación interpretado que busca simplificar al máximo la legibilidad de su código. Es poderoso y fácil de aprender y cuenta con estructuras de datos eficientes y de alto nivel, con un enfoque simple pero efectivo para la programación orientada a objetos. También soporta programación imperativa y en menor medida programación funcional, por lo que se considera un lenguaje de programación multiparadigma (“Python Software Foundation,” 2017).

La elegante sintaxis de Python y su tipado dinámico, junto con su naturaleza interpretada, hacen de éste un lenguaje ideal para scripting y desarrollo rápido de aplicaciones en diversas áreas y sobre la mayoría de las plataformas.

Apareció a principios de la década del 90 y su creador fue Guido Van Rossum. Actualmente es administrado por la Python Software Foundation. Posee una licencia de código abierto, denominada Python Software Foundation License (“Lista de licencias con comentarios - Proyecto GNU - Free Software Foundation,” n.d.), que es compatible con la licencia pública general de GNU a partir de la versión 2.1.1, e incompatible en ciertas versiones anteriores.

Su selección para el desarrollo de la presente investigación se debe a que posee una comunidad muy fuerte, así como una amplia gama de librerías que facilitan su uso.

Se utilizó Python en su versión 3.7.

2.3.2. Pytorch.

PyTorch es una biblioteca de aprendizaje automático de código abierto basada en la biblioteca Torch, utilizada para aplicaciones como la visión por computadora y el procesamiento del lenguaje natural. Está desarrollado principalmente por el laboratorio de investigación de IA de Facebook.

Es un paquete de Python diseñado para realizar cálculos numéricos haciendo uso de la programación de tensores. Además, permite su ejecución en GPU para acelerar los cálculos.

Normalmente PyTorch es usado tanto para sustituir Numpy (una biblioteca de funciones matemáticas de alto nivel para operar con matrices y vectores) y procesar los cálculos en GPU como para la investigación y desarrollo en el campo del aprendizaje automático. Está centrado principalmente en el desarrollo de redes neuronales (“PyTorch,” n.d.)

Se utilizó Pytorch en su versión 1.1.0.

2.3.3. Entorno virtual.

Un ambiente o entorno virtual persigue aislar paquetes y librerías de diferentes versiones y trabajar en cada una de ellas sin afectar otros entornos. Estos entornos se pueden crear, modificar, eliminar e incluso compartir con otras máquinas. Es importante señalar que un entorno virtual no es multiplataforma, por lo que se debe crear uno específico tanto para sistemas operativos Mac OS, como Unix o Windows (“Virtualenv,” n.d.).

2.3.4. Anaconda.

Es un gestor y administrador de paquetes científicos. Es el estándar de la industria para la ciencia de datos de código abierto. Anteriormente conocida como *Anaconda Distribution*, esta herramienta cuenta con el respaldo de una vibrante comunidad de código abierto y cuenta con más de 20 millones de usuarios en todo el mundo. Es la herramienta elegida por los científicos de datos en solitario que desean usar Python o R para proyectos de computación científica. Su propagación en la comunidad de la ciencia de datos se debe a su escalabilidad, seguridad y simplicidad (“Anaconda | The World’s Most Popular Data Science Platform,” n.d.).

Los científicos de datos pueden desarrollar e implementar modelos de inteligencia artificial y aprendizaje automático rápidamente en la producción. Anaconda proporciona las herramientas necesarias para:

- Recopilar datos de archivos, bases de datos y lagos de datos.
- Administrar entornos con Conda (todas las dependencias de paquetes se atienden en el momento de la descarga).
- Compartir, colaborar y reproducir proyectos.
- Implementar proyectos en producción con solo hacer clic en un botón.

2.3.5. Spyder como IDE.

Un entorno de desarrollo integrado (IDE) es un sistema de software para el diseño de aplicaciones que combina herramientas del desarrollador comunes en una sola interfaz gráfica de usuario (GUI). Generalmente, un IDE cuenta con las siguientes características:

- Editor de código fuente: editor de texto que ayuda a escribir el código de software con funciones como el resaltado de la sintaxis con indicaciones visuales, el relleno automático específico del lenguaje y la comprobación de errores a medida que se escribe el código.
- Automatización de compilación local: herramientas que automatizan tareas sencillas y repetibles como parte de la creación de una compilación local del software para su uso por parte del desarrollador, como la compilación del código fuente de la computadora en un código binario, el empaquetado del código binario y la ejecución de pruebas automatizadas.
- Depurador: programa que sirve para probar otros programas y mostrar la ubicación de un error en el código original de forma gráfica.

Spyder es un poderoso entorno científico escrito en Python, para Python, y diseñado por y para científicos, ingenieros y analistas de datos. Ofrece una combinación única de la funcionalidad avanzada de edición, análisis, depuración y creación de perfiles de una herramienta de desarrollo integral con la exploración de datos, ejecución interactiva, inspección profunda y hermosas capacidades de visualización de un paquete científico (“Spyder Website,” n.d.).

2.3.6. Programación paralela.

La computación paralela es una forma de cómputo en la que muchas instrucciones se ejecutan simultáneamente, operando sobre el principio de que problemas grandes, a menudo se pueden dividir en unos más pequeños, que luego son resueltos simultáneamente.

Por lo general se utilizan en programas científicos debido a la complejidad que encierran sus operaciones

2.3.7. Hardware utilizado.

El equipo utilizado durante todas las fases de la investigación fue una PC con un procesador Intel(R) Core(TM) i5-8400, con una velocidad de CPU a 2.80 GHz, con 16 GB de RAM y una tarjeta de video NVIDIA GeForce GTX 1070, con 8 GB de RAM dedicada al video.

2.3.8. Fuente de datos.

Para el desarrollo de este trabajo se han utilizado las bases de datos DRIVE (Staal, et. al., 2004) y CHASE (Owen et. al., 2009) que son gratuitas y han sido utilizadas en concursos en la búsqueda y detección de diferentes patologías que pueden ser identificadas en las imágenes de la retina. Estas imágenes fueron utilizadas en el entrenamiento y prueba de una red de aprendizaje profundo.

- DRIVE: Contiene 40 imágenes en color de la retina, con dimensiones de 565x584. Este conjunto ya está dividido en 20 imágenes para el entrenamiento, las cuales fueron separadas en 15 para entrenar la red neuronal propuesta y 5 para validación, así como otras 20 imágenes para pruebas.
- CHASEDB: Consisten en 28 imágenes de la retina de 14 niños, centradas en el nervio óptico, cada una con una dimensión de 999x960 píxeles.

2.4. Estado del arte.

Numerosos trabajos para la segmentación de imágenes de retina han sido propuestos, algunos de ellos usan técnicas tradicionales de aprendizaje automático. Orlando et. al. (Orlando, et. al., 2017) utilizó un modelo totalmente conectado y un campo aleatorio condicional con los parámetros aprendidos del método, utilizando un clasificador de máquinas de soporte vectorial. Oliveira *et al.* (Oliveira, et. al., 2016) proponen un método no supervisado que combina un filtro emparejado, un filtro de Frangi y un filtro de Gabor Wavelet para mejorar las imágenes.

Actualmente, las principales técnicas de segmentación utilizan las CNNs, por ejemplo, el trabajo de Maninis *et al.* (Maninis et al., 2016), cuya arquitectura se basa en la red VGG (Simonyan & Zisserman, 2014). Dasgupta *et al.* (Dasgupta & Singh, 2017) propusieron una tarea de inferencia multi-etiqueta, combinando una CNN con una predicción estructurada. En (Alom, et. al., 2018), Alom *et al.* introducen bloques residuales y lo complementan con capas convolucionales residuales recurrentes. Sin embargo, Zhuang en (Zhuang, 2018) fue más allá y apiló dos U-Net con bloques residuales, aumentando las rutas para el flujo de información. Otro trabajo importante fue propuesto por (Khanal & Estrada, 2019), que utilizan pesos estocásticos para obtener un buen equilibrio entre los píxeles de fondo y los vasos sanguíneos, utilizando una segunda red reducida, para los píxeles de clasificación ambigua. En la tabla a continuación, se resumen estos trabajos y otros que se basan en sus métodos y en las métricas reportadas.

Tabla 2.1: Resumen de artículos para la segmentación de imágenes de retina. Los informes métricos fueron la precisión, la recuperación y el F1-score entre otros, para el conjunto de datos de DRIVE.

Autor	Método	Métrica
(Oliveira et al., 2016)	Combination of filters (matched, Frangi's & Gabor Wavelet). Deformable models and Fuzzy C-means for segmentation.	Acc = 0.9402

(Orlando et al., 2017)	Fully connected conditional random field with SVM.	F1 = 0.7857
(Maninis et al., 2016)	CNN based on VGG network.	F1 = 0.8210
(Soomro et al., 2017)	CNN.	Recall = 0.746
(Dasgupta & Singh, 2017)	CNN with structured prediction.	F1 = 0.8074
(Alom et al., 2018)	CNN with residual block and recurrent residual convolutional layers.	F1 = 0.8171
(Jin et al., 2019)	Deformable CNN.	F1 = 0.7883
(Khanal & Estrada, 2019)	DNN.	F1 = 0.8259
(Zhuang, 2018)	Chain of multiple U-Nets.	F1 = 0.8202

2.5. Propuesta de solución.

El presente trabajo propone una arquitectura basada en una arquitectura U-Net común y una arquitectura U-Net con bloques residuales, donde ambas redes están conectadas. La primera parte consiste en la extracción de características, mientras que la segunda se centra en la detección de nuevas características o de aquellas que resulten ambiguas, mediante la importación del flujo de información de la red anterior y el uso de bloques residuales. En el siguiente capítulo se abundarán en detalles de la propuesta de solución.

3. METODOLOGÍA

Para el desarrollo del trabajo se pretende trabajar con bases de datos disponibles en la web, las cuales son libres y han sido utilizadas en competencias en la búsqueda y detección de diferentes patologías que pueden ser identificadas en imágenes de fondo de ojo. Con dichas imágenes puede ser entrenada una red de aprendizaje profundo y puesta a prueba la misma.

De manera general se pretende seguir un flujo de trabajo en el cual se obtienen las imágenes de retina, las cuales son sometidas a pre-procesamiento, en este apartado varias técnicas se pueden aplicar como extracción de características, resaltar determinado patrón u omitir otros, se seleccionan las características más relevantes para la detección en sí como una segmentación o determinada área recortada, posteriormente esta imagen pasa a la arquitectura de red neuronal profunda que se diseñe y se corre el algoritmo sobre la red para efectuar la detección de los elementos de interés; en esta parte se entrenan y evalúan las imágenes pasadas en lote, se calcula un error y se ajustan los pesos, todo este proceso es un ciclo sujeto a cambios de arquitecturas hasta lograr los mejores índices de rendimiento y por último se realiza una interpretación de los resultados obtenidos y se aplican. La siguiente imagen muestra el flujo del proceso.

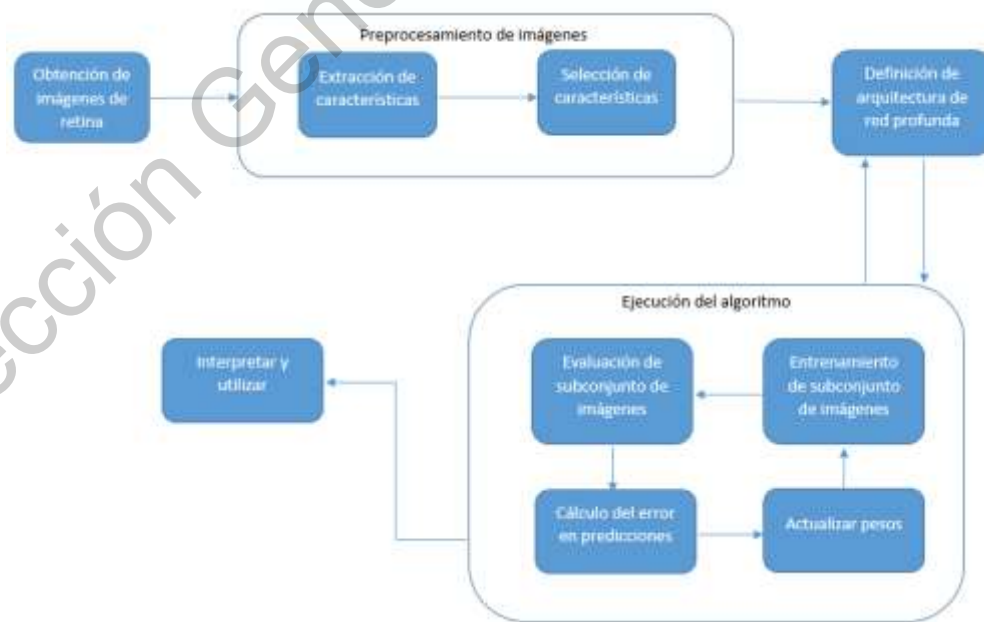


Figura 3.1: Flujo de proceso para la detección de patrones de riesgo cardiovascular en imágenes de retina.

3.1. Pre-procesamiento.

Ocupa un lugar primordial dentro de cualquier tarea de aprendizaje automático, ya que los datos se presentan usualmente en un formato no óptimo, para ser procesados y analizados por el modelo, de ahí la expresión de habla inglesa que expresa “*garbage in, garbage out*”.

Con el fin de hacer nuestros modelos confiables, existen un conjunto de técnicas de pre-procesamiento como la limpieza de datos, que nos permite eliminar ruidos o inconsistencias de nuestros datos, integración de datos que mezcla los mismos desde diferentes fuentes de una forma coherente, transformación de datos y reducción de los mismos. Estas técnicas no son mutuamente excluyentes, o sea, pueden trabajar juntas. Estas técnicas cuando se aplican antes de cualquier modelo, pueden mejorar de manera exponencial, no solo la precisión de los modelos sino incluso disminuir los tiempos de procesamiento (Kamber, 2006).

En resumen, los datos en el mundo real tienden a estar “sucios”, incompletos e inconsistentes, por lo que las diferentes técnicas de pre-procesamiento pueden mejorar la calidad de nuestros datos y de esta forma mejorar la precisión de nuestros modelos. Indudablemente el pre-procesamiento juega un papel fundamental en el descubrimiento de conocimiento desde grandes fuentes de datos, ocupando hasta un 60% del esfuerzo total dentro de este ciclo.

En la presente investigación esta tarea de pre-procesamiento estuvo asociada a técnicas de mejoras de imágenes, las cuales se dividen fundamentalmente en dos grupos (“Lección.4: Métodos de Mejora y Realce de la Imagen,” n.d.):

Métodos de mejora en el dominio de la frecuencia: las cuales se basan en aplicar la transformada de Fourier a una imagen, multiplicar el resultado por un filtro y tomar la transformada de Fourier Inversa como resultado mejorado. Este método incluye los filtros *low-pass*, *high-pass* y homomorfo.

Método de mejora del dominio espacial: este se basa en operaciones de manipulación de píxeles individuales en una imagen. Estas incluyen operaciones aritméticas de puntos y algoritmos de mejora de vecinos. Las operaciones

aritméticas de puntos incluyen la transformación a escala de grises y la ecualización de histograma que garantiza una distribución uniforme de los niveles de grises.

A continuación, se describen los utilizados en este trabajo de tesis.

3.1.1. Conversión de imagen RGB a escala de grises.

RGB

En las tecnologías asociadas a la computación, el espacio de colores RGB es el más utilizado. Sus siglas hacen referencia a los colores “Red”, “Green” y “Blue” del inglés y se trata de un modelo cromático mediante el cual seremos capaces de generar diferentes colores a partir de la síntesis aditiva de la iluminación de estos colores básicos.

Imagen en escala de grises

Constituye la representación de una imagen en la que cada pixel se dibuja usando un valor numérico individual que representa su luminancia, en una escala que se extiende entre el blanco y el negro (“Escala de grises (grayscale) - Diccionario de fotografía y diseño,” n.d.).

La conversión de una imagen a colores en su representación a escala de grises es el primer paso dentro del pre-procesamiento, de esta forma se obtiene un mayor contraste de los vasos sanguíneos respecto al fondo de la imagen original, esto es debido a que la imagen en escala de grises proporciona sólo la información de la luminancia de la imagen de color luego de remover el tono y la saturación (Saleh, et. al., 2011). Para lograr este contraste se aplicó la siguiente ecuación:

$$Img_{gris} = R * 0.299 + G * 0.587 + B * 0.114 \quad (6)$$

Donde R, G y B son los canales correspondientes a la imagen de entrada. En la ecuación antes mostrada el canal G es sobresaltado, debido a que provee el mayor contraste de los vasos sanguíneos (El Abbadi & Al Saadi, 2014).

3.1.2. Normalización de los datos.

Es un tipo de transformación de datos que puede mejorar la precisión y eficiencia de los algoritmos que involucran medidas de distancias. Es particularmente útil en algoritmos de clasificación, específicamente en redes neuronales de tipo *backpropagation*. La normalización de los valores de entrada para cada atributo medido en las tuplas de entrenamiento, contribuirá a acelerar la fase de entrenamiento (Kamber, 2006). Se utiliza para escalar un atributo dentro de un rango determinado.

Normalización min-max

Es una transformación lineal de los datos de entrada, preservando la relación de los valores originales.

$$v' = \frac{v - \min_A}{\max_A - \min_A} \quad (7)$$

Donde v' es el valor normalizado del valor original v y su relación con los valores mínimos y máximos de un atributo A. Para los efectos de esta investigación, los valores de dicho atributo varían entre 0 y 255.

Normalización Z-Score

Está basada en la media y en la desviación estándar de un atributo A, y es útil cuando hay valores atípicos gobernando los valores extremos, mínimo y máximo, de dicho atributo.

$$v' = \frac{v - \bar{A}}{\sigma_A} \quad (8)$$

Donde \bar{A} y σ_A son la media y la desviación estándar respectivamente.

3.1.3. Ecuación de histograma a través de contraste limitado adaptativo (CLAHE).

El tercer paso es aplicar la técnica de CLAHE, ecuación adaptativa de histograma limitado por contraste, por sus siglas en inglés. El mismo constituye una variante de

la ecualización adaptativa de histogramas, que tiene como limitante la amplificación de ruidos en zonas relativamente homogéneas. Con CLAHE se logra limitar esta amplificación. Esta técnica es efectiva para mejorar los detalles en imágenes de retina en escala de grises de una manera uniforme (Zhou, et. al., 2018).

El método divide la imagen en pequeñas regiones, donde el histograma es ecualizado y de esta forma se produce una mejora local del contraste.

3.1.4. Ajuste de la luminosa.

Por último, se realiza un ajuste de la luminosidad a través de correcciones gamma, que es la forma en la que se llama a cierta operación no lineal que se usa para codificar y decodificar luminancia o valores tri-estímulos en sistemas de videos o imagen. Esta operación se hace necesaria porque muchas imágenes presentan un brillo insuficiente o desigual, lo que puede hacer que los detalles de los diagnósticos sean indetectables (Zhou et al., 2018).

El resultado de aplicar las técnicas de pre-procesamiento antes descritas se puede apreciar en la siguiente imagen.



Figura 3.2: Resultado del pre-procesamiento. En la izquierda la imagen original y a la derecha la imagen pre-procesada.

3.1.5. Parches y volteos aleatorios.

Un parche es una pieza pequeña (generalmente rectangular) de una imagen. Por ejemplo, un parche de 8x8 es un parche cuadrado que contiene 64 píxeles de una imagen más grande (de tamaño, digamos, 256x256 píxeles). Debido al tamaño más pequeño, algunos de los algoritmos de procesamiento de imágenes, son más fáciles de operar en parches en lugar de operar en toda la imagen. Estos algoritmos dividen una imagen en varios parches de menor tamaño (de tamaño, digamos, 8x8), operan individualmente en cada uno de estos parches y finalmente colocan todos estos parches en sus respectivas ubicaciones.

Estudios previos han mostrado que usando parches de imágenes más grandes en CNNs arrojan una mejor precisión en la clasificación, ya que las CNNs pueden capturar más información contextual para tomar una mejor decisión.

En la presente investigación se extraen parches de las imágenes de retina, lo que permite disponer de un conjunto de datos mayor y facilitar así el entrenamiento de una red neuronal configurada arbitrariamente. Este procedimiento puede considerarse como un aumento de los datos y se complementa con el volteo horizontal y vertical aleatorio de estos parches en cada iteración (Sekou, et. al., 2019).

3.2. Arquitectura.

En este trabajo se utilizaron dos redes neuronales convolucionales totalmente conectadas, con la peculiaridad de que la segunda utiliza bloques residuales. La estructura original, también conocida como U-Net, fue propuesta por (Ronneberger et al., 2015) y está compuesta por una parte de codificación, donde se captura tanto la información semántica como la contextual a medida que se duplica el número de filtros; así como una segunda parte de decodificación, donde se restaura la información espacial y se fusionan las características obtenidas en los niveles más bajos. A continuación, se detallan los elementos de la arquitectura propuesta.

3.2.1. UNet_1.

Se compone de cuatro niveles en los que se aplica repetidamente una doble convolución con un filtro de 3x3, con un *padding* igual a cero, un *stride* igual a uno y una operación de *max_pooling* con un filtro de 2x2 y una *padding* de dos. La imagen de entrada tenía un canal y un tamaño de 572x572. En el primer nivel el número de canales utilizados se fijó en 24 y el tamaño de la imagen era de 568x568 tras dobles convoluciones. Este proceso se repitió en los tres niveles siguientes y los resultados para cada nivel fueron 48 canales y 280x280, 96 canales y 136x136, 192 canales y 64x64, en el cuello de botella 384 canales y 28x28. La parte de decodificación sigue la estructura anterior, pero en sentido contrario. El número de canales fue el mismo de la parte de contracción, pero el tamaño de la imagen se reduce, los valores de cada nivel fueron 52x52, 100x100, 196x196 y 388x388. También se añaden las características que se encuentran en cada uno de los diferentes niveles de codificación. Después de cada capa de convolución se aplica una capa de normalización por lotes y una función de activación ReLU, lo que mejora significativamente el tiempo de entrenamiento, así como la estabilidad y la capacidad de generalización del modelo (Ioffe & Szegedy, 2015).

3.2.2. Bloque residual.

La idea surge de un problema de degradación que aparece cuando los modelos comienzan a converger, como se demuestra en (He et al., 2015). Para reducir este problema, se propuso el uso de bloques residuales, resultado esperado está compuesto por la siguiente ecuación:

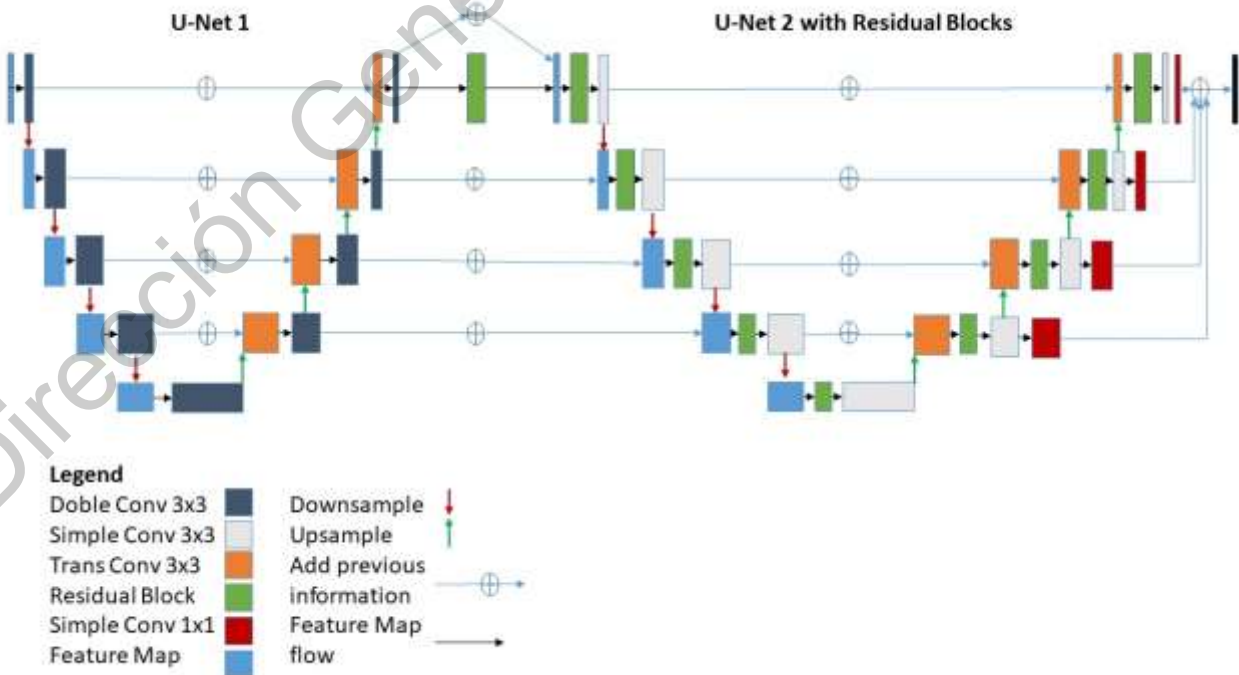
$$FM(x)=F(x)+x \quad (9)$$

Donde $FM(x)$ es el mapa de características esperado al aplicar dos capas convolucionales a las características de entrada, representadas como $F(x)$ y a esta transformación, se añadió la entrada original x . La adición del mapa de características original atenúa el problema de degradación que aparece en los modelos. En la figura a continuación se detalla el proceso utilizado en este trabajo.

Figura 3.3: Bloque residual utilizado en el presente trabajo. Un mapa de características x pasa una secuencia de capas (azul: convolucional, amarillo: normalización por lotes, verde: activación ReLU). Después de este procedimiento, la característica original x se añade con los resultados obtenidos a través de las capas. Finalmente, se aplica una activación ReLU.

3.2.3. UNet_2 con bloques residuales.

La salida de la primera red, así como los mapas de características de cada nivel constituyen la entrada a esta segunda. El número de canales y el tamaño de la imagen en cada nivel se mantuvo igual a la parte de decodificación de la primera mitad de la arquitectura propuesta. La principal diferencia es que en cada nuevo nivel se añade un bloque residual, tanto de contracción como de expansión. Esta vez se utiliza un relleno igual a uno y en la última capa se aplica una convolución de filtro de 1×1 para producir una salida de dos canales, ya que se lleva a cabo una



tarea de clasificación binaria. Véase la figura a continuación, para una representación de toda la arquitectura propuesta.

Figura 3.4: La arquitectura propuesta. Consiste en una U-Net, que le sigue otra red U-Net con bloques residuales. En la primera red U-Net se realiza la extracción de características, mientras que en la segunda red U-Net con bloques residuales se realiza la detección de nuevas características o de aquellas que son ambiguas.

Dirección General de Bibliotecas UAQ

4. RESULTADOS Y EVALUACIÓN.

En este capítulo se busca evaluar el desempeño de la metodología antes propuesta, mostrando los resultados experimentales en la segmentación de la red vascular en imágenes de retina.

4.1. Elementos de configuración.

4.1.1. Función de pérdida.

Las funciones de pérdida cuantifican cuán cerca está una red neuronal dada del ideal hacia el cual es el entrenamiento. La idea es simple, calculamos una métrica basada en el error que observamos en las predicciones de la red, luego agregamos estos errores en todo el conjunto de datos y los promedios y ahora tenemos un único número representativo de cuán cerca está la red neuronal de su ideal.

Encontrar ese estado ideal se puede interpretar como encontrar los parámetros (pesos y bias) que minimicen la pérdida incurrida desde esos errores. Por tal motivo las funciones de pérdidas hacen que el entrenamiento de redes neuronales pueda ser visto como un problema de optimización y en muchos casos no pueden resolverse de forma analítica, pero pueden aproximarse bastante bien a través de algoritmos de optimización de forma iterativa, como el gradiente descendente (Josh et. al., 2017).

El presente trabajo de grado constituye un problema de clasificación con la peculiaridad de la presencia de clases desbalanceadas, ya que la mayoría de los píxeles representan el fondo de la imagen, mientras que la minoría representan a la red vascular de la imagen de retina.

Por esta razón la función de pérdida escogida fue *negative log-likelihood*, la cual se define por la siguiente ecuación:

$$loss(x, y) = -(logy) \quad (10)$$

Esta función maximiza la probabilidad global de los datos, dando un alto valor de pérdida cuando la clasificación es errónea o poco clara y un bajo valor de pérdida

cuando la predicción coincide con la esperada por el modelo. El logaritmo realiza la parte penalizadora, cuanto menor sea la probabilidad, mayor será el logaritmo. Como estas probabilidades tienen valores entre cero y uno, y los logaritmos en ese rango son negativos, se utiliza el signo negativo para convertirlos en valores positivos.

4.1.2. Inicialización de pesos.

La inicialización de los pesos de una red neuronal puede afectar significativamente el entrenamiento de la misma, por lo que dicha inicialización es la llave en el proceso de aprendizaje.

Para manejar el desbalance de clases el atributo del peso fue asignado a cada clase, tanto a la predicción como a la referencia. Aquí se siguió el enfoque de (Khanal & Estrada, 2019), donde se plantea lo siguiente:

$$w_{rand(1,\alpha,s)} \quad (11)$$

Donde el peso w varía aleatoriamente entre uno y el valor de α con un paso igual a s . Esta variación dinámica de los pesos evita que la red caiga en mínimos locales.

4.1.3. Hiper-parámetros.

En el aprendizaje de máquinas y por consecuente, en el aprendizaje profundo estamos en presencia tanto de parámetros del modelo, así como parámetros que modificamos para conseguir un entrenamiento más rápido y mejor. Estos parámetros que modificamos son llamados hiper-parámetros y se ocupan de controlar las funciones de optimización y la selección de modelos durante el entrenamiento, en el proceso de aprendizaje.

4.1.3.1. Adam como optimizador.

El mencionado proceso de ajustar los pesos de una red neuronal para obtener predicciones más y más precisas es conocido como optimización de parámetro, el mismo puede ser visto como un proceso científico, donde formulamos una hipótesis,

la probamos contra la realidad y refinamos o reemplazamos la hipótesis una y otra vez hasta ajustarnos lo más posibles a los escenarios de la realidad.

Estos procesos de optimización permiten a la red neuronal auto-ajustarse en cada iteración y los optimizadores determinan cómo se actualizarán los pesos de cada capa, por lo que son de vital importancia en el rendimiento de una red neuronal.

Adam (*Adaptative moment estimation*) (Kingma & Ba, 2014) es un algoritmo de optimización para la actualizar los pesos de una red de forma iterativa en función de los datos de entrenamiento. Deriva las tasas de aprendizaje de los primeros y segundos momentos de los gradientes.

Entre sus ventajas están:

- Fácil de implementar.
- Computacionalmente eficiente.
- Pequeños requerimientos de memoria.
- Adecuado para problemas con gran número de parámetros.
- Sus hiper-parámetros generalmente requieren poca sintonización.

Este método mantiene una tasa de aprendizaje para cada peso de la red (parámetro) y se adapta por separado a medida que se desarrolla el aprendizaje.

Aplica los beneficios de Algoritmo de gradiente adaptativo (AdaGrad), donde se mantiene una tasa de aprendizaje por parámetro y Propagación cuadrática media (RMSProg), donde también se mantiene una tasa de aprendizaje por parámetro que se adaptan según el promedio de las magnitudes recientes de los gradientes para el peso.

ADAM también utiliza el promedio de los segundos momentos de los gradientes (la varianza no centrada).

4.1.3.2. Tasa de aprendizaje.

Probablemente sea el hiper-parámetro más importante a configurar dentro de una red neuronal, ya que tiene un alto impacto en la estabilidad y eficiencia del tiempo de entrenamiento.

Es común buscar una tasa que tenga un equilibrio entre pasos largos y pasos cortos, ya que desafortunadamente no se puede establecer este valor de forma analítica, por tanto, es un proceso de prueba y error.

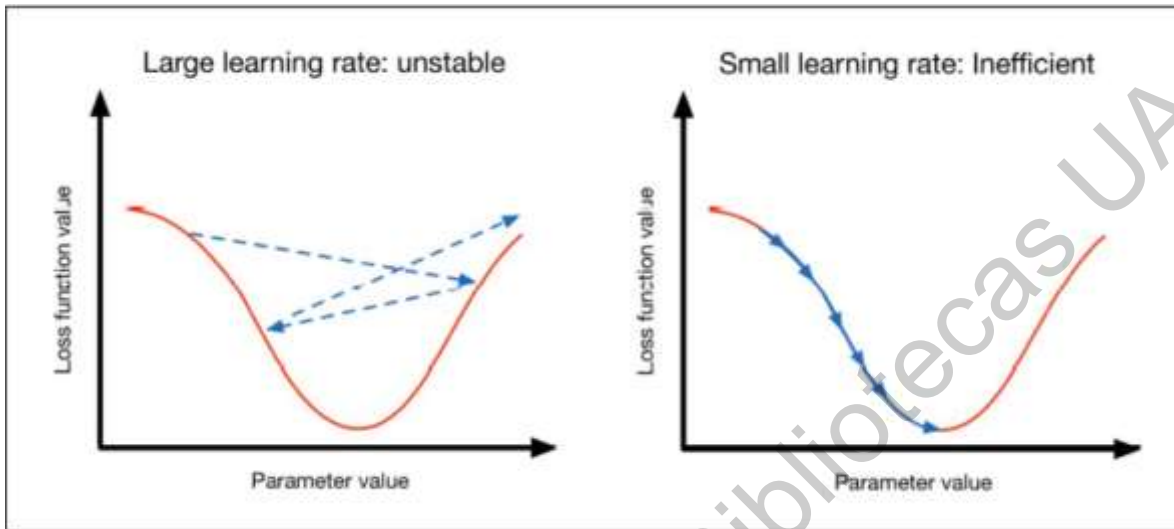


Figura 4.1: Efecto de diferentes valores de tasa de aprendizaje en la búsqueda del óptimo global (Josh et. al., 2017).

Normalmente la tasa de aprendizaje es menor que uno y mayor que 10^{-6} y queremos que empiece a bajar cerca del final del mismo, por lo que se configuró una tasa de aprendizaje que decayera cada 30 iteraciones a una razón de 0.7.

4.1.3.3. Otros hiper-parámetros configurados.

En total, 250 épocas fueron establecidas, con una parada prematura configurada a las 35 iteraciones si no hay cambios en la función de pérdida.

El tamaño de los lotes es también un importante hiper-parámetro, así como la definición del número de muestras que se propagarán a través de la red. En este caso el valor seleccionado fue cuatro porque se requiere menos memoria y típicamente las redes se entrenan más rápido con mini-lotes porque el peso se actualiza después de cada propagación.

4.2. Métricas de evaluación.

Con el presente trabajo estamos en presencia de una tarea de segmentación, que, además, presenta la complejidad de contar con clases desbalanceadas, por lo que

encontrar una métrica adecuada que evalúe el resultado de la tarea de manera efectiva se hace complejo.

Para la evaluación de algoritmos de segmentación se suele utilizar métricas estadísticas y métricas geométricas, según se analicen propiedades relativas a las regiones segmentadas o a su contorno. En nuestro caso particular utilizaremos las relacionadas a las regiones segmentadas, o sea, métricas estadísticas.

4.2.1. Recuperación, precisión y F1-score.

De acuerdo a (Tharwat, 2018), la recuperación, la precisión y la media armónica entre esas dos métricas antes mencionadas, llamada F1-score, son adecuadas para evaluar tareas de segmentación en presencia de clases desbalanceadas. También se hizo referencia a la exactitud (*accuracy*). El cálculo de dichas métricas está basado en cuatro posibles interpretaciones que se describen a continuación.

- Verdadero positivo (TP): La etiqueta de la muestra es positiva y se clasifica como tal.
- Verdadero negativo (TN): la etiqueta de la muestra es negativa y se clasifica como tal.
- Falso positivo (FP): la etiqueta de la muestra es negativa, pero se clasifica como positiva.
- Falso negativo (FN): la etiqueta de la muestra es positiva, pero se clasifica como negativa.

En base a estos valores podemos calcular las métricas antes mencionadas:

- Recuperación (*Recall*): nos dice cuántas muestras relevantes son seleccionadas.

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

- Precisión (*Precision*): nos dice cuántas muestras predichas son relevantes.

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

- F1-Score: es la media armónica entre la recuperación y la precisión.

$$F1_{Score} = \frac{2*TP}{2*TP+FP+FN} \quad (14)$$

- Exactitud (Accuracy): mide cuantas observaciones, tanto positivas como negativas, fueron correctamente clasificadas.

$$ACC = \frac{(TP+TN)}{TP+FP+FN+TN} \quad (15)$$

4.2.2. Validación cruzada k-fold.

La validación cruzada es un procedimiento de remuestreo utilizado para evaluar modelos de aprendizaje automático en una muestra de datos limitada.

El procedimiento tiene un único parámetro llamado k que se refiere al número de grupos en los que se dividirá una muestra de datos determinada. Como tal, el procedimiento a menudo se llama validación cruzada k-fold. Cuando se elige un valor específico para k, se puede usar en lugar de k en la referencia al modelo, como k = 10 convirtiéndose en una validación cruzada de 10 veces.

La validación cruzada se utiliza principalmente en el aprendizaje automático aplicado para estimar la habilidad de un modelo de aprendizaje automático en datos no vistos. Es decir, usar una muestra limitada para estimar cómo se espera que el modelo funcione en general cuando se usa para hacer predicciones sobre datos que no se usaron durante el entrenamiento del modelo.

Es un método popular porque es simple de entender y porque generalmente da como resultado una estimación menos sesgada o menos optimista de la habilidad del modelo respecto a otros métodos, con una división simple de entrenamiento / prueba.

Tanto para el entrenamiento como para la validación y las pruebas, se respetó la configuración por defecto en DRIVE. Sin embargo, CHASEDB utilizó la técnica de validación cruzada utilizando una configuración estratificada de 5 fold, que se generó de forma aleatoria, de manera que en cada fold se dejaron 5 imágenes para las pruebas y las 23 imágenes restantes se dividieron en entrenamiento y validación. El proceso de validación se repitió 5 veces con la misma arquitectura y se entrenó desde cero para cada pliegue. Finalmente, los resultados se promediaron para

producir una única estimación. El valor de 5 para k en k-fold se fijó empíricamente y, según las propuestas de la mayoría de los trabajos del estado del arte, un valor más alto de k habría conducido a un mayor tiempo de entrenamiento.

4.2.3. Índice de similitud estructural (SSIM).

Las métricas del SSIM se utilizaron para hacer comprensible el proceso de segmentación, y debido a la dificultad de diferenciar entre una primera fase de predicción con el empleo de sólo una red de tipo U-Net y una segunda fase con el empleo de toda la arquitectura propuesta donde se hace uso de los bloques residuales.

El índice de similitud estructural (SSIM) analiza la distancia visual, la información de los bordes, entre la referencia y la imagen de prueba. Es una métrica perceptual que cuantifica la degradación de la calidad de la imagen causada por procesos como la compresión de datos (Renieblas, et. al., 2017). El SSIM oscila entre 0 (completamente diferente) y 1 (parches idénticos). Cuanto más alto sea el valor, mejor.

4.3. Experimentación y resultados.

El trabajo aquí descrito, su comportamiento fue cuantificado en términos de la precisión, la recuperación, F1-score y la exactitud o accuracy, constituyendo estos la media de dichos valores a través de todas las imágenes de prueba.

Durante la experimentación se le agregaron los bloques residuales a la arquitectura propuesta en (Khanal & Estrada, 2019), la misma incorpora una versión reducida de la red llamada mini-unet, la cual se encarga de detectar pixeles ambiguos; así como se buscaron diferentes configuraciones para el número de filtros, considerado también un hiper-parámetro, normalmente mientras mayor sea el número de filtros, mayor serán las características detectadas. Los resultados obtenidos se muestran en la siguiente tabla.

Tabla 4.1: Resultados de la experimentación donde se combinan la arquitectura del trabajo propuesto en (Khanal & Estrada, 2019), con la arquitectura propuesta en este trabajo.

Arquitectura	Nro. Filtros U-Net	Nro. Filtros MiniU-Net	Precisión	Recuperación	F1-Score	Accuracy
Propuesta con datos originales	10	128	0.82774	0.8071	0.81729	0.9684
Propuesta con datos preprocesados	10	128	0.82653	0.81392	0.82018	0.96874
Propuesta con datos preprocesados	24	128	0.82095	0.82642	0.82368	0.96901
Propuesta con datos preprocesados	24	256	0.82281	0.82527	0.82404	0.96913

Lo primero que se resalta en esta experimentación es la importancia de realizar pre-procesamiento a los datos. Esta actividad mejora la calidad de la imagen, enfocada en resaltar ciertos detalles, permitiendo a la CNN detectar ciertas características además de que le da a la misma mayor capacidad de generalización. Como se puede observar en la tabla 4.1, el F1-score de los datos originales es el más bajo de todas las variantes, ya que no se le aplicaron técnicas de pre-procesamiento.

En todo momento se buscó un balance entre precisión y recuperación a través de su media armónica, llamada F1-score. Como se puede apreciar en la tabla 4.1 los mejores resultados se alcanzaron cuando el número de filtros en la segunda red, llamada mini-UNet, fue aumentando, esto es debido a que mientras es mayor el número de filtros, mayor es el número de abstracciones que la red es capaz de detectar; sin embargo, el tiempo de procesamiento aumentó considerablemente. Estos valores de filtros no pudieron ser aumentados tampoco debido a falta de recursos computacionales del equipo con el que se trabajó y que se describió previamente.

Por ello se removió la segunda red, para disminuir el número de parámetros totales, ya que esta trabaja con 4 redes de tipo U-Net, dos de ellas con bloques residuales, según la propuesta de la investigación. También se experimentó con la función de

activación *Leaky ReLU*, la cual no aportó una mejora significativa, por lo que se optó por seguir trabajando con la función de activación *ReLU*, estandarizada en la mayoría de los trabajos de segmentación, los resultados se muestran a continuación.

Tabla 4.2: Resultados con dos redes U-Net encadenadas y bloques residuales en la segunda.

Arquitectura	Nro. Filtros U-Net	Precisión	Recuperación	F1-Score	Accuracy
Khanal et. al.	64	0.8284	0.8235	0.8259	0.9693
Propuesta con Leaky ReLU	24	0.80772	0.84081	0.82393	0.96853
Propuesta con ReLU	24	0.83417	0.81609	0.82503	0.96968

En la siguiente tabla se pueden observar comparaciones con trabajos similares, basados en el conjunto de imágenes DRIVE, dichas comparaciones están basadas en las métricas antes descritas.

Tabla 4.3: Comparación con otros trabajos del estado del arte sobre la base de datos DRIVE.

Método	Precisión	Recuperación	F1-Score	Accuracy
(Orlando et al., 2017)	0.7854	0.7897	0.7857	-
(Maninis et al., 2016)		0.8261	0.8210	0.9541
(Dasgupta & Singh, 2017)	-	0.7691	0.8074	0.9533
(Yan, Yang, & Cheng, 2019)	-	0.7631	-	0.9538
U-Net (Alom et al., 2018)	-	0.7537	0.8142	0.9531

Residual U-Net (Alom et al., 2018)	-	0.7726	0.8149	0.9553
Recurrent U-Net (Alom et al., 2018)	-	0.7751	0.8155	0.9556
R2U-Net (Alom et al., 2018)	-	0.7792	0.8171	0.9556
(Zhuang, 2018)	-	0.7856	0.8202	0.9561
(Jin et al., 2019)	0.8529	0.7963	0.8237	0.9566
(Khanal & Estrada, 2019)	0.8284	0.8235	0.8259	0.9693
Propuesta	0.8341	0.8160	0.8250	0.9696

Se puede concluir, a partir de los valores equilibrados de precisión y recuperación obtenidos y mediante el uso de la métrica F1-Score y el valor alto asociado a ella, que estamos clasificando correctamente las venas y los píxeles de fondo de la imagen de entrada. Como se puede ver en la tabla comparativa 4, nuestra propuesta obtuvo el segundo mejor valor de F1-Score, por una ligera diferencia inferior. Por lo tanto, teniendo la mayor *accuracy* entre todos los trabajos, se puede concluir que nuestra clasificación se realizó en la mayoría de los casos, correctamente, y no a expensas de aumentar el número de falsos positivos y falsos negativos. También hay que señalar que en las condiciones técnicas en que se realizaron los experimentos de este estudio, el tiempo de ejecución se redujo en 57 minutos en el conjunto de imágenes DRIVE, y en 10 horas y 38 minutos en CHASEDB, en comparación con (Khanal & Estrada, 2019). Por lo tanto, este estudio logró resultados similares en un período de procesamiento más corto, como se muestra en la Tabla 6.

Tabla 4.4: Comparación respecto al tiempo de entrenamiento.

Arquitectura	DRIVE	CHASEDB
(Khanal & Estrada, 2019)	1:34 horas	19:46 horas
Propuesta	0:37 horas	9:08 horas

Para argumentar el comportamiento de las métricas, se muestran las figuras 23 y 24, asociadas a la pérdida, el accuracy, F1-Score y la razón entre la precisión y la recuperación, todas ellas asociadas a DRIVE.

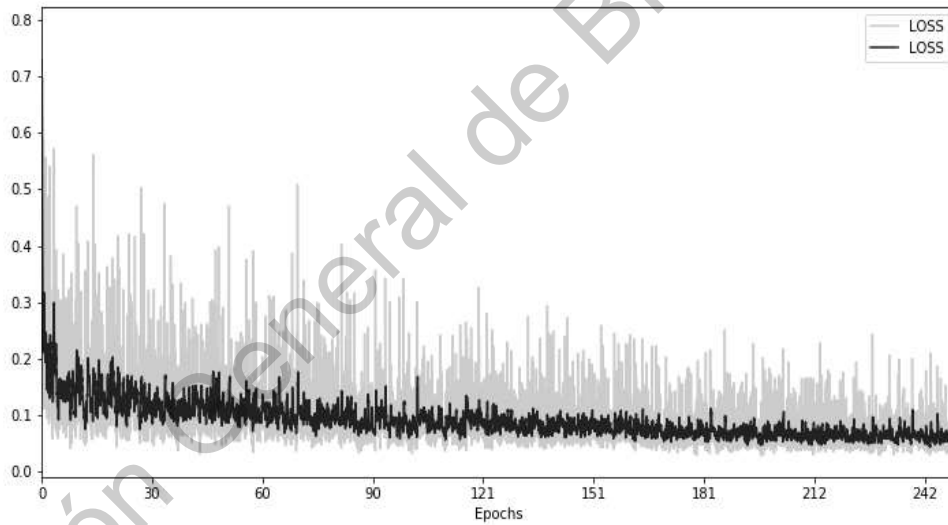


Figura 4.2: Pérdida asociada al entrenamiento de la arquitectura propuesta.

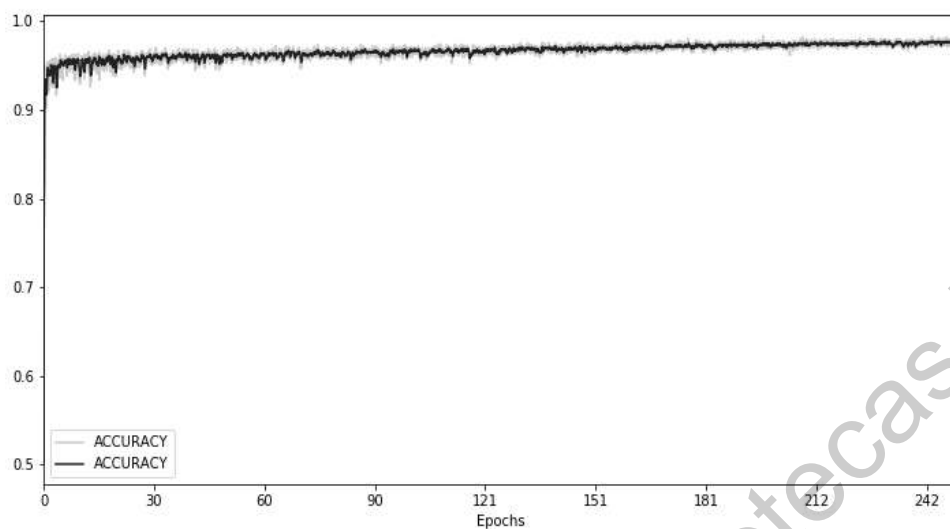


Figura 4.3: Accuracy en la fase de entrenamiento.

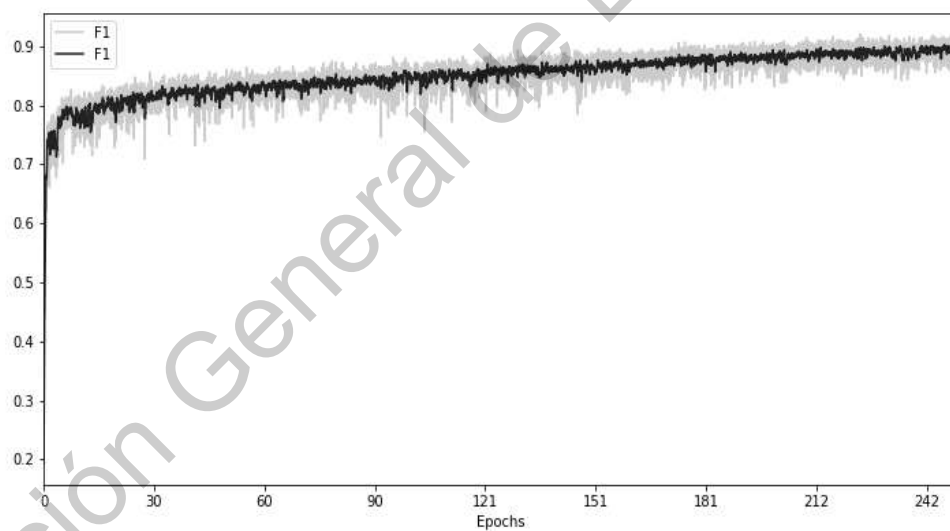


Figura 4.4: F1-Score asociado al entrenamiento.

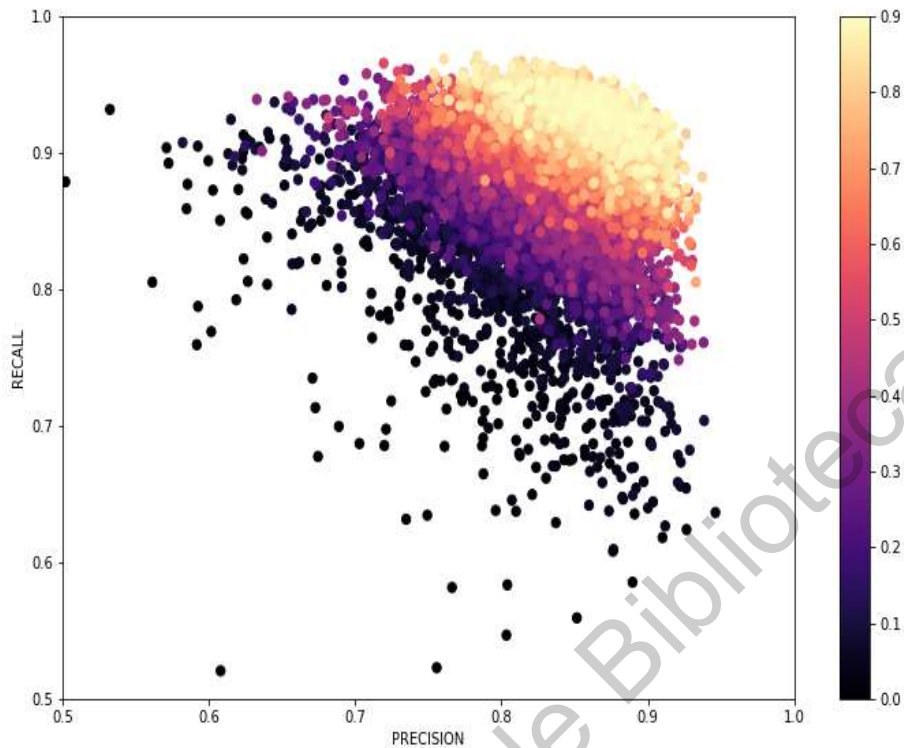


Figura 4.5: Comportamiento de la precisión respecto a la recuperación en la fase de entrenamiento.

Estas imágenes no hacen ver que el resultado del entrenamiento fue consistente a lo largo de todas las épocas establecidas, lográndose un buen balance entre la precisión y la recuperación como muestra la imagen 24. Por lo tanto, la utilización de pesos dinámicos como ha sido propuesto en (Khanal & Estrada, 2019), así como la introducción de la segunda red con bloques residuales, ha permitido una segmentación satisfactoria, logrando un buen manejo de las clases desbalanceadas.

El comportamiento sobre las imágenes de pruebas se muestra a continuación.

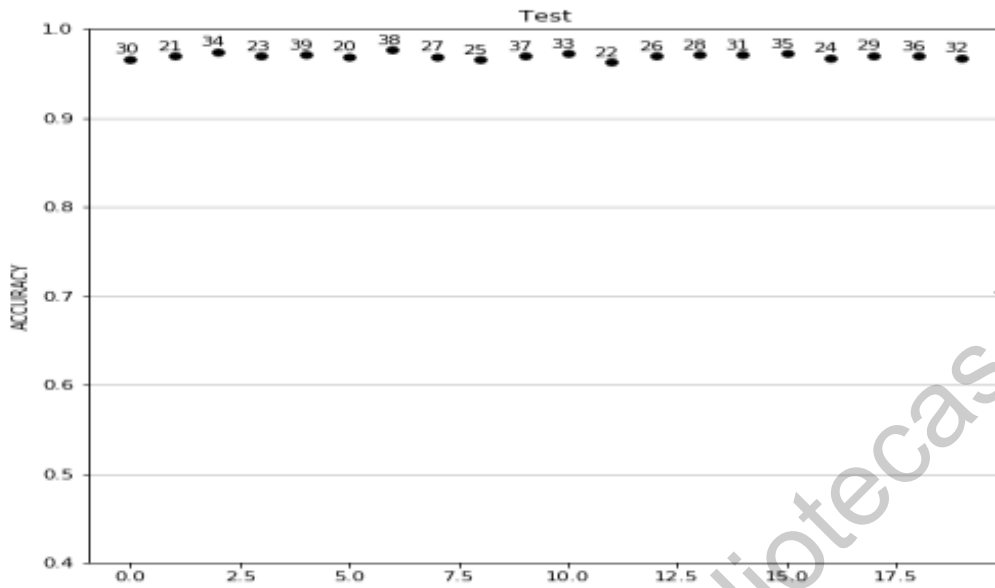


Figura 4.6: Accuracy sobre conjunto de imágenes de prueba, en DRIVE.

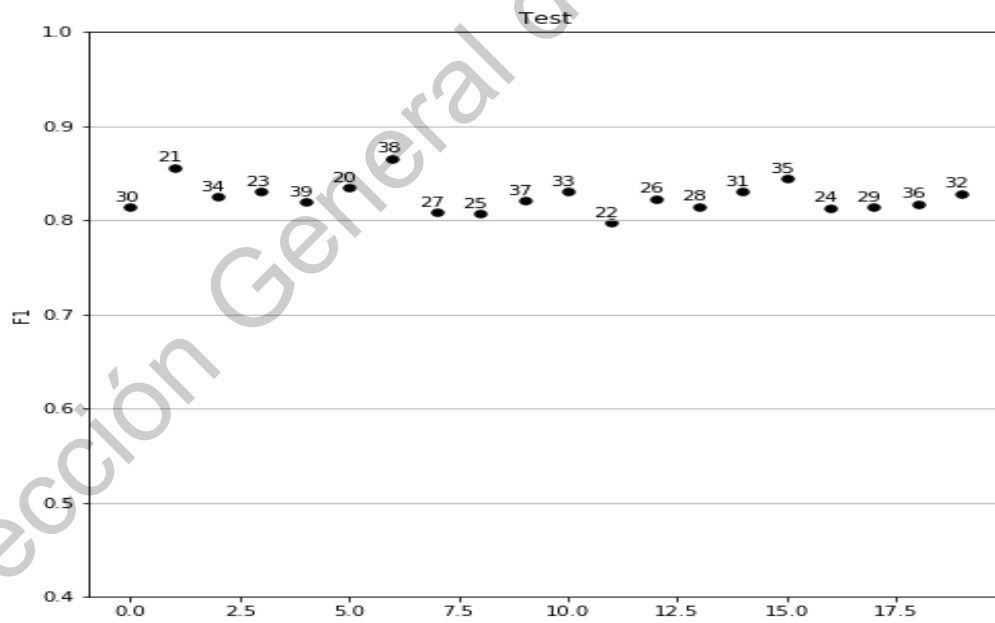


Figura 4.7: F1-Score sobre conjunto de imágenes de prueba, en DRIVE.

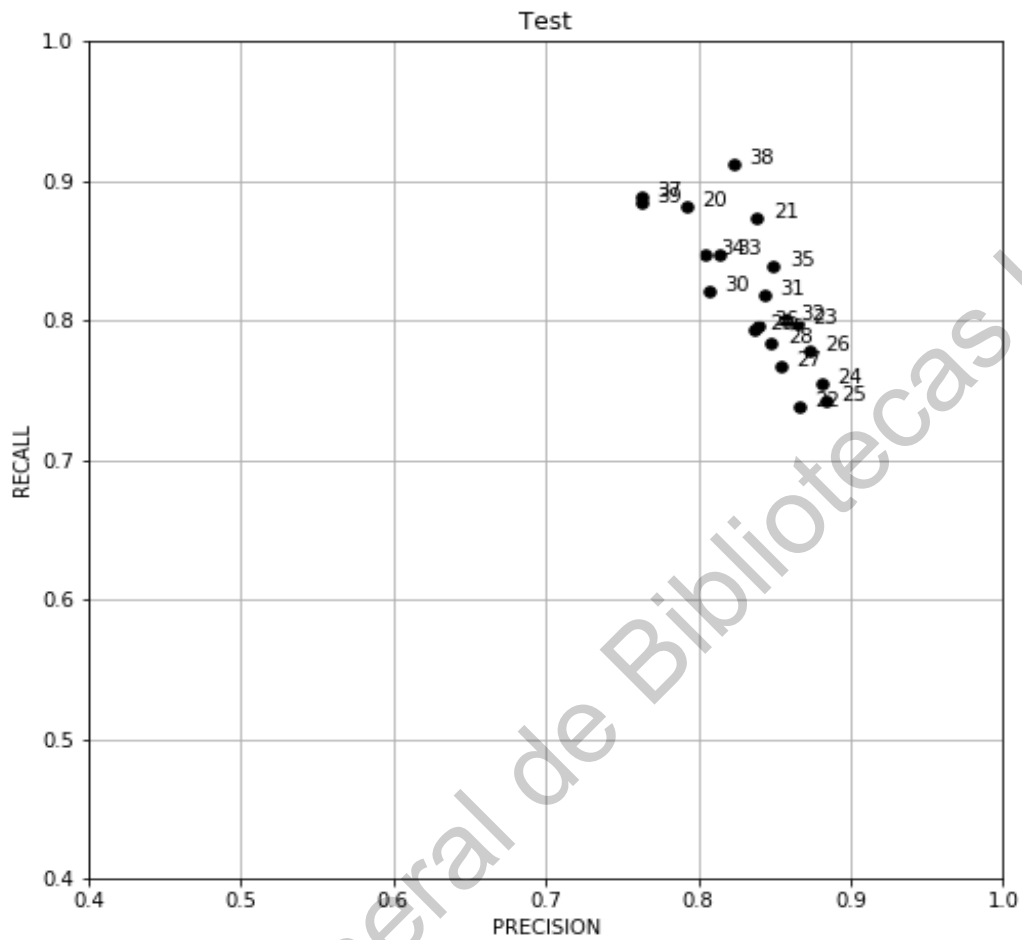


Figura 4.8: Curva de precisión y recuperación en conjunto de imágenes de prueba, en DRIVE.

El resultado global de aplicar la arquitectura propuesta se puede apreciar en las siguientes imágenes.

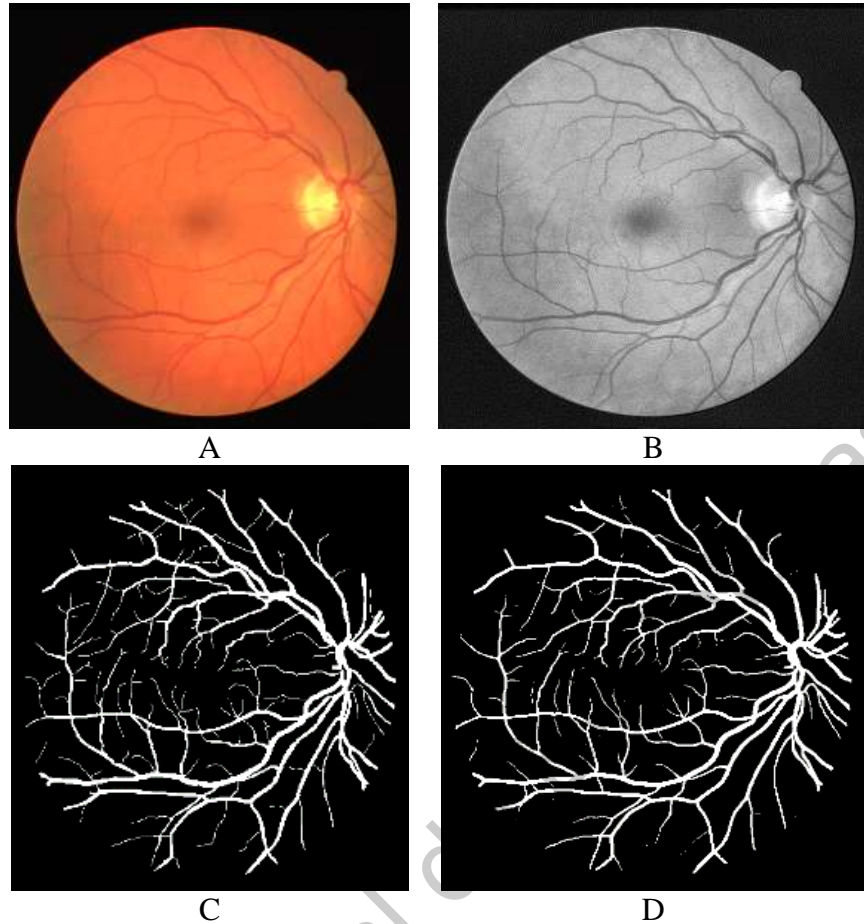


Figura 4.9: Resultados sobre las imágenes de prueba en DRIVE. A imagen original, B imagen pre-procesada, C ground truth, d salida segmentada.

Para ver el resto de las imágenes de la base de datos DRIVE consultar el anexo 1.

El mismo análisis que se ha realizado hasta ahora fue llevado a cabo sobre el conjunto de imágenes de CHASEDB. A continuación, se muestran los resultados empezando con una tabla comparativa de las métricas antes expuestas, en referencia a otros trabajos del estado del arte.

Tabla 4.5: Comparación con otros trabajos del estado del arte sobre CHASEDB.

Método	Precisión	Recuperación	F1-Score	Accuracy
(Orlando et al., 2017)	0.7438	0.7277	0.7332	-
(Yan et al., 2019)	-	0.7641	-	0.9607
U-Net (Alom et al., 2018)	-	0.8288	0.7783	0.9578
Residual U-Net (Alom et al., 2018)	-	0.7726	0.7800	0.9553
Recurrent U-Net (Alom et al., 2018)	-	0.7459	0.7810	0.9622
R2U-Net (Alom et al., 2018)	-	0.7756	0.7928	0.9634
(Zhuang, 2018)	-	0.7978	0.8031	0.9656
(Jin et al., 2019)	0.7630	0.8155	0.7883	0.9610
(Khanal & Estrada, 2019)	0.8550	0.8143	0.8245	0.9759
Proposed	0.8366	0.8258	0.8312	0.9766

Sobre este conjunto de imágenes se puede apreciar que nuestro método propuesto alcanza los mayores valores tanto de F1-Score como de *accuracy*.

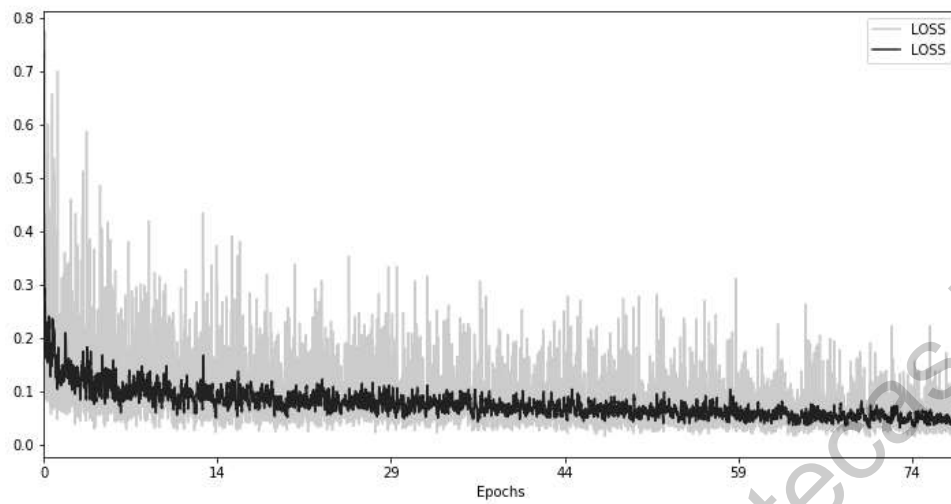


Figura 4.10: Pérdida asociada al entrenamiento de la arquitectura propuesta sobre CHASEDB.

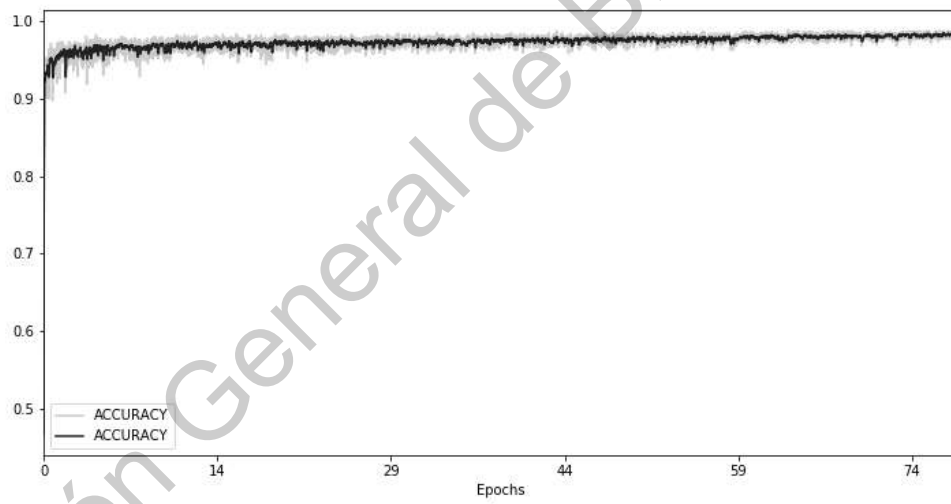


Figura 4.11: Accuracy asociado al entrenamiento de la arquitectura propuesta sobre CHASEDB.

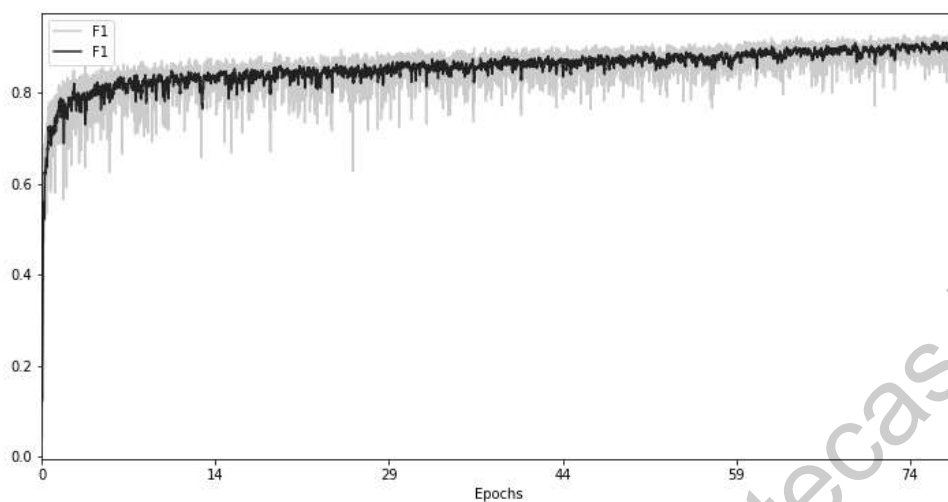


Figura 4.12: F1-Score asociado al entrenamiento de la arquitectura propuesta sobre CHASEDB.

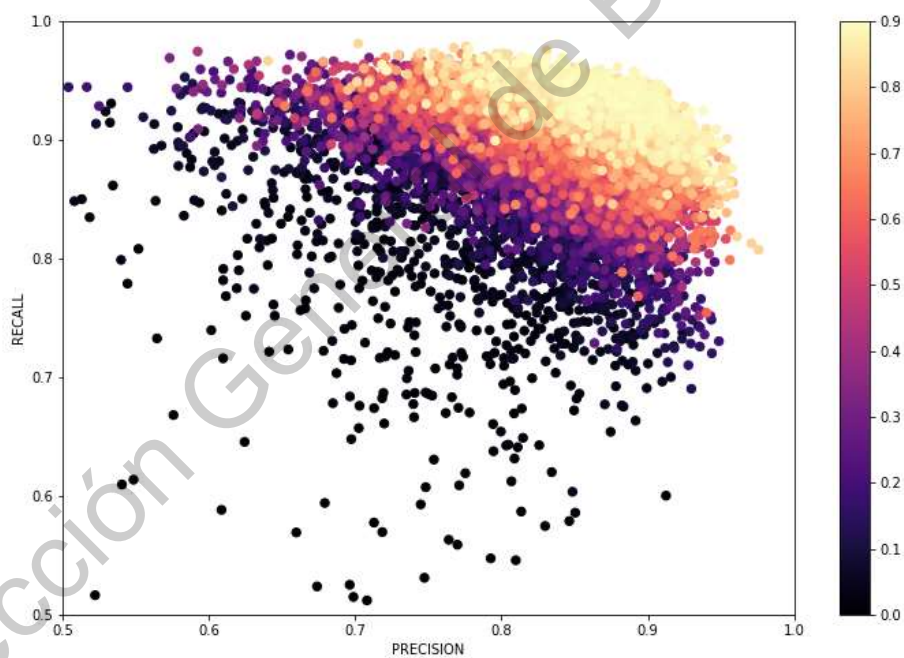


Figura 4.13: Comportamiento de la precisión respecto a la recuperación en la fase de entrenamiento sobre CHASEDB.

A continuación, las imágenes asociadas a la fase de pruebas sobre CHASEDB.

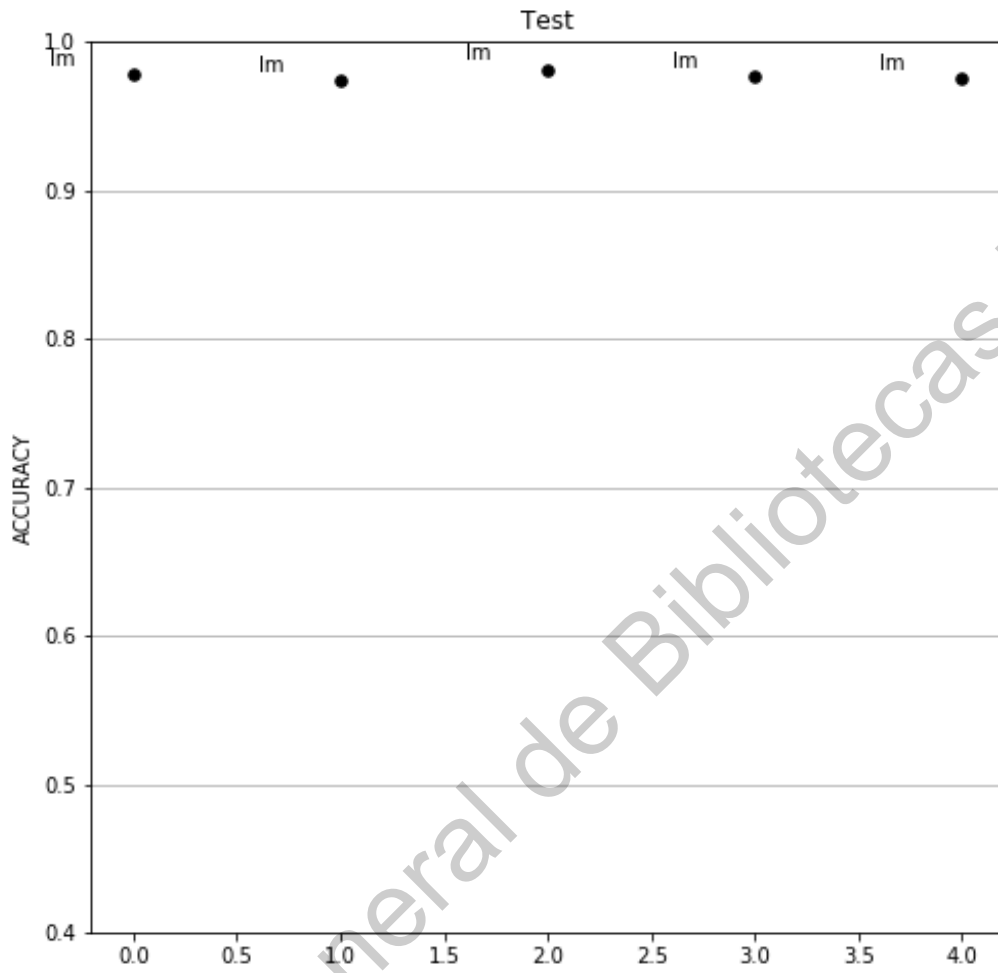


Figura 4.14: Accuracy sobre conjunto de imágenes de prueba, en CHASEDB.

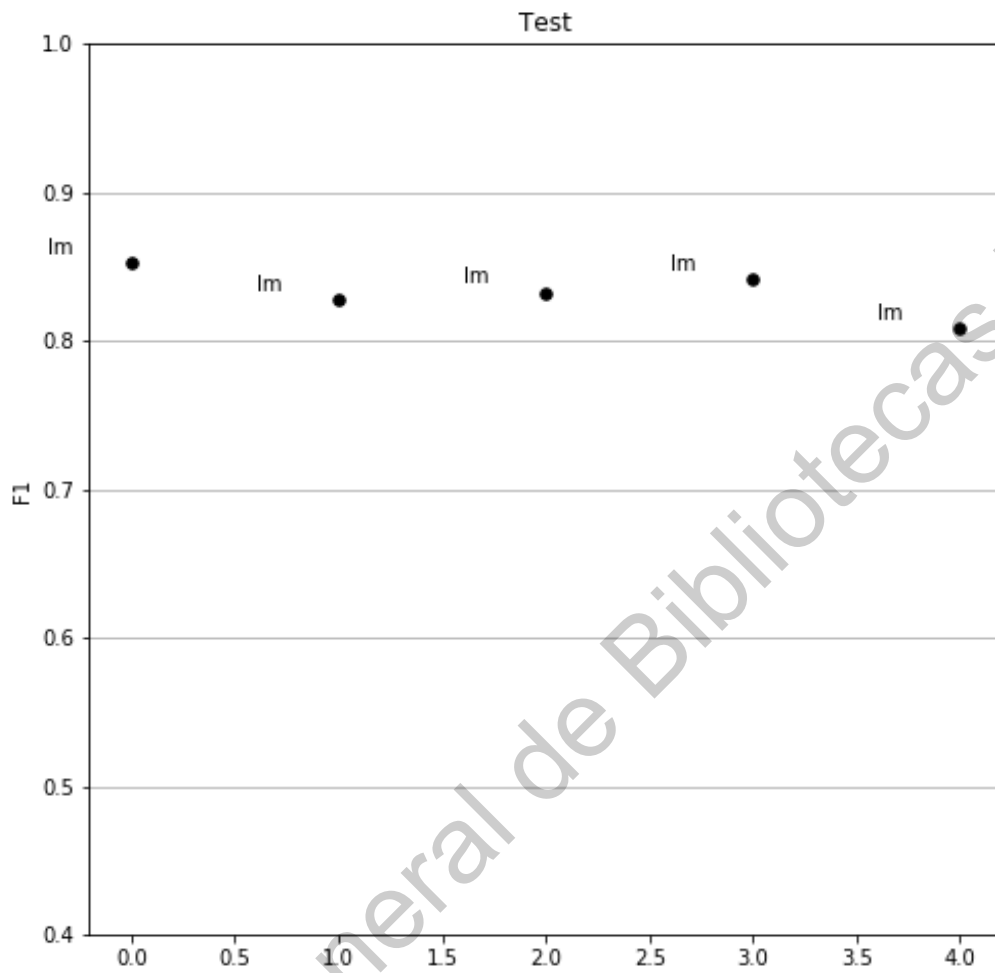


Figura 4.15: F1-Score sobre conjunto de imágenes de prueba, en CHASEDB.

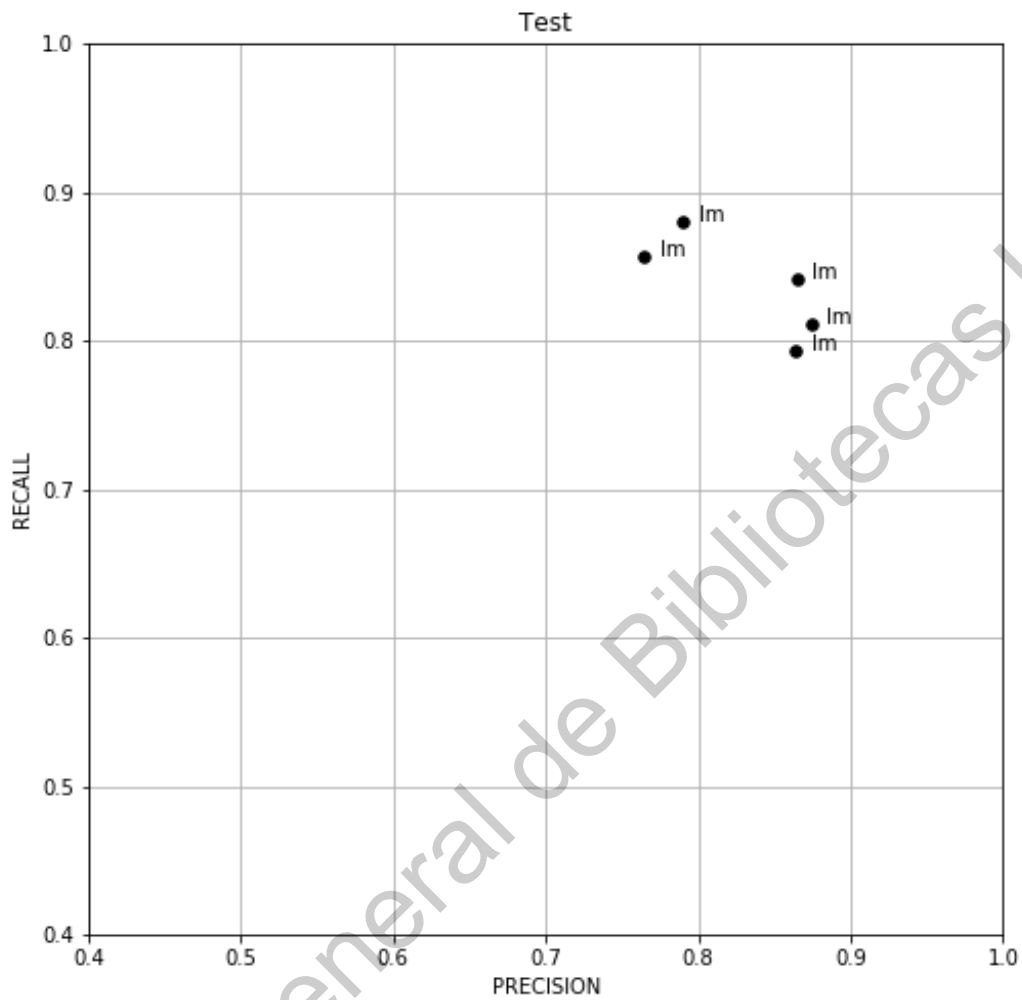


Figura 4.16: Curva de precisión y recuperación en conjunto de imágenes de prueba, en CHASEDB.

El resultado de aplicar la arquitectura sobre las imágenes de CHASEDB se puede apreciar en la siguiente figura.

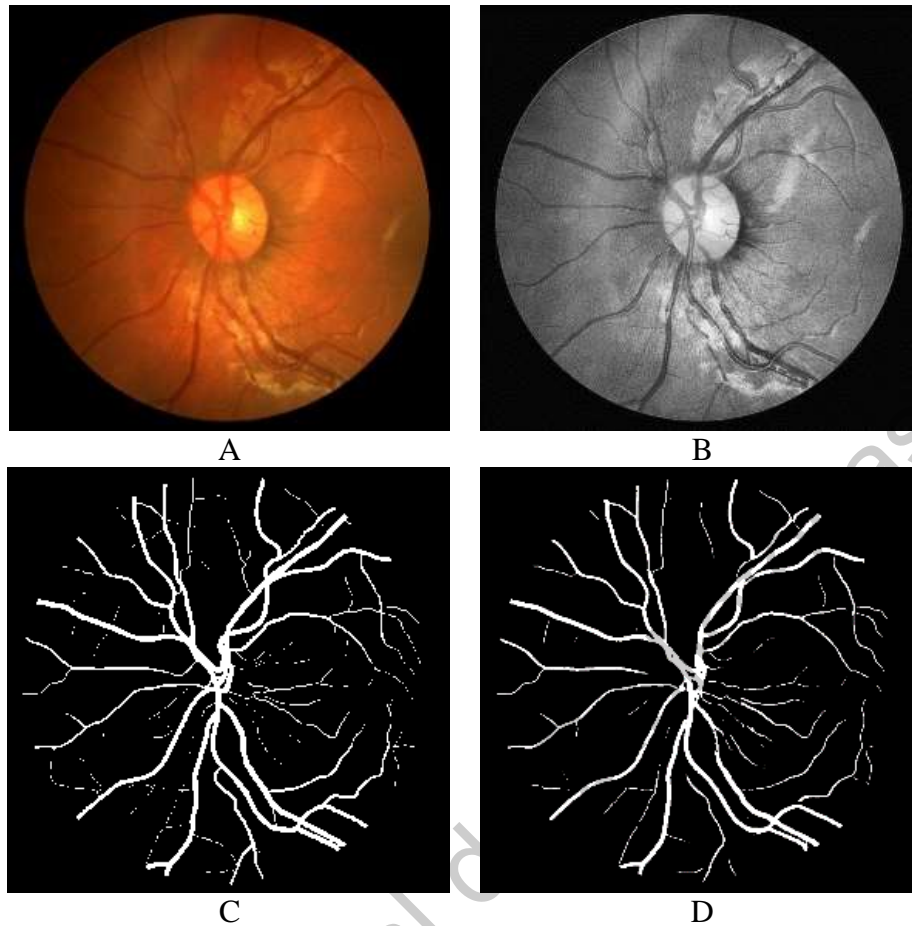


Figura 4.17: Resultados sobre las imágenes de prueba en CHASEDB. A imagen original, B imagen pre-procesada, C ground truth, d salida segmentada.

Para ver el resto de las imágenes de CHASEDB consultar el anexo 2.

Sin embargo, apreciar que tan bien han sido segmentados los vasos sanguíneos de la retina resulta complicado de ver a simple vista, además de que se quiere saber si la arquitectura propuesta en realidad ha aportado a la tarea de segmentación, a pesar de la versión original. Es por ello que se decidió utilizar SSIM como métrica para comparar dos imágenes, el *ground truth* y la salida de la red neuronal. Un valor más alto de este índice es mejor para nuestro propósito. La figura a continuación muestra una comparación entre la primera red U-Net, el resultado que arroja y el *ground truth*.

SSIM: 0.84

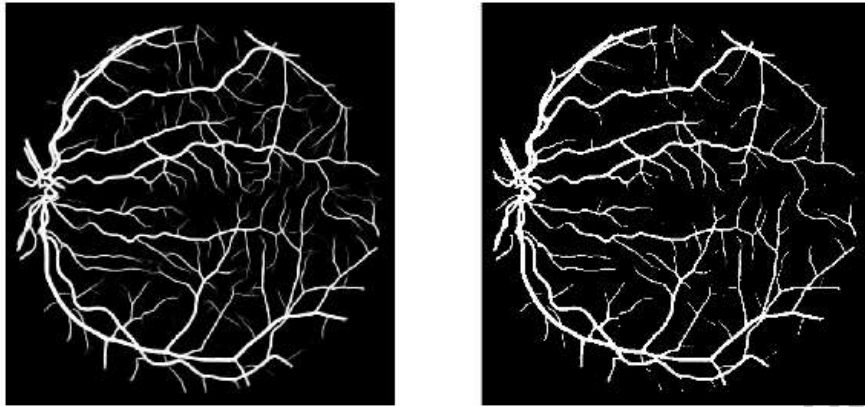


Figura 4.18: Valor de SSIM entre el ground truth y el resultado de la U-Net 1.

La siguiente imagen muestra la misma comparación, pero con toda la arquitectura.

SSIM: 0.85

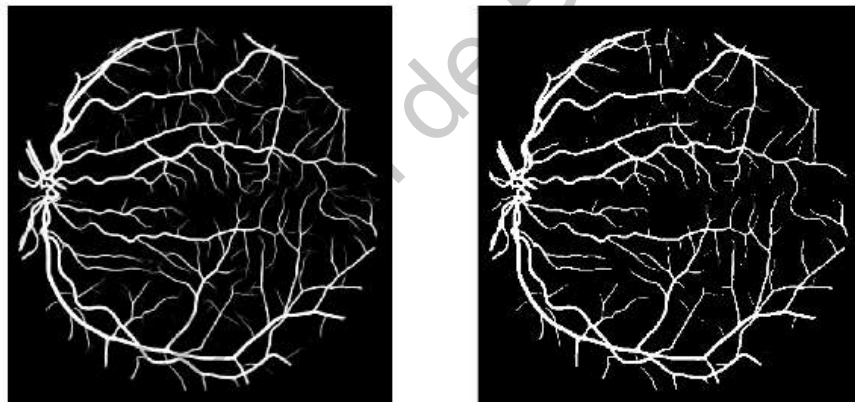


Figura 4.19: Valor de SSIM entre el ground truth y el resultado de la arquitectura propuesta.

En la siguiente imagen se puede apreciar una porción aumentada de la segmentación, donde se pueden ver los pequeños detalles. La muestra fue tomada de una imagen de DRIVE.

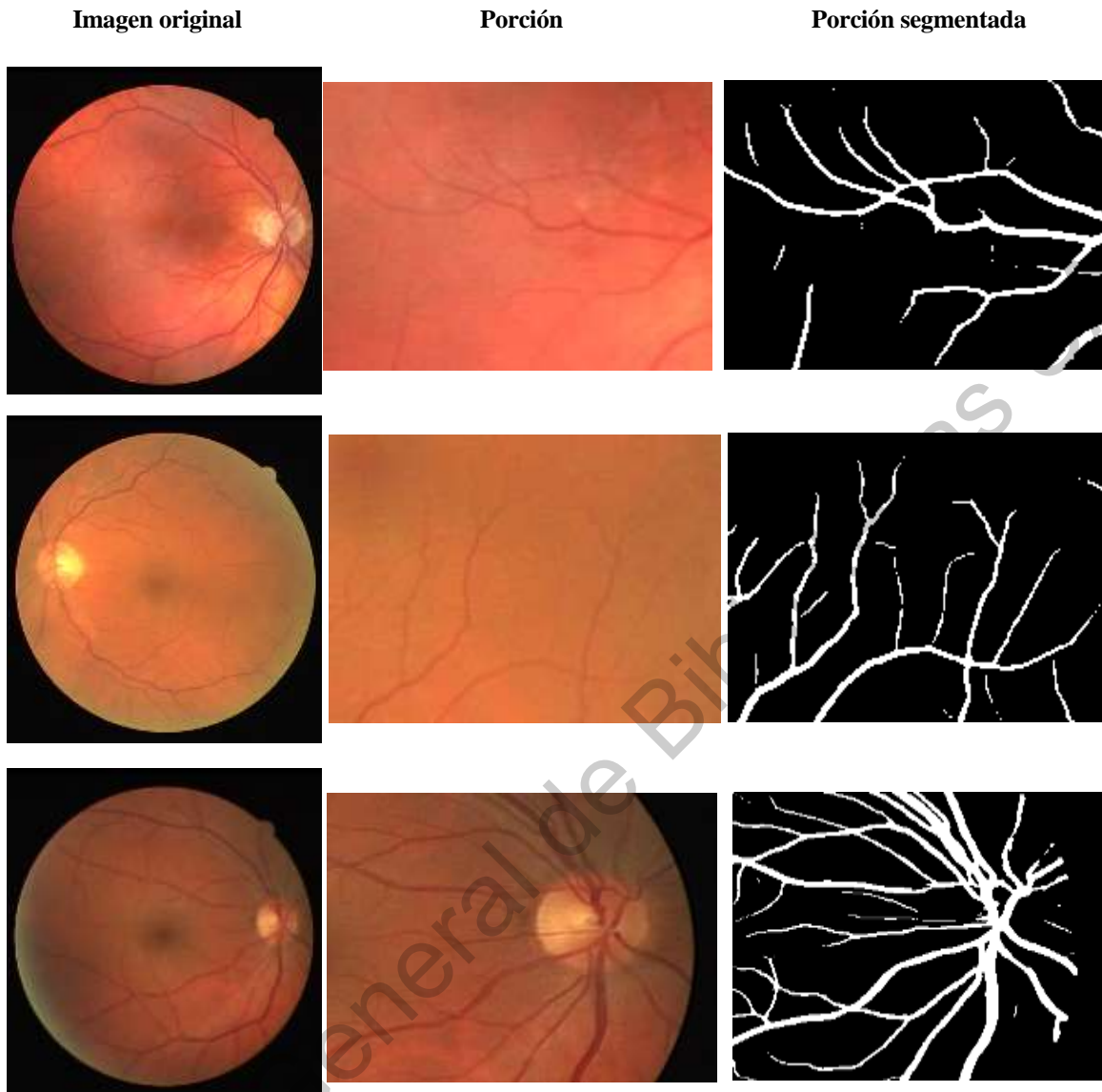


Figura 4.20: Porción de segmento, donde se pueden apreciar vasos sanguíneos tanto pequeños, medianos como grandes.

Otro elemento que se tuvo en cuenta durante la experimentación fue el efecto de las diferentes lesiones que pueden existir cerca de los vasos sanguíneos y su efecto en la segmentación. Las imágenes utilizadas fueron tomadas de las bases de datos DRIVE y CHASE. De las 40 fotografías de DRIVE, 33 no presentaron ningún signo de enfermedad, mientras que los 7 restantes tenían presencia de retinopatía diabética en su estado temprano a medio, así como cambios en el epitelio de los pigmentos, cicatriz pigmentaria en la fovea, o corioidopatía. Ninguna de las imágenes de CHASE presentó enfermedad alguna. En las imágenes con presencia

de alguna enfermedad, la tarea de segmentación se realizó sin ninguna dificultad. Los resultados se pueden observar en la siguiente imagen.

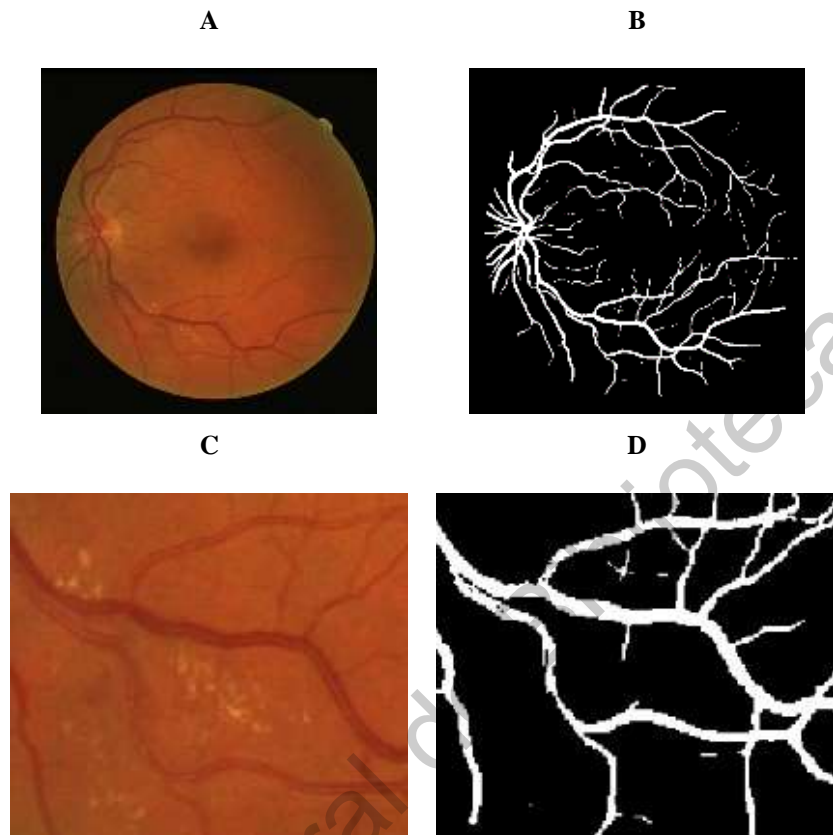


Figura 4.21: Resultado en imagen de prueba de DRIVE. A imagen original, B Imagen segmentada, C área de lesión, D área de la lesión en la imagen segmentada.

Como se puede observar, no hay evidencia de ninguna lesión en la imagen segmentada: Por tanto, se puede inferir que este nivel de lesión no tiene efecto en la segmentación de vasos sanguíneos en la retina.

5. CONCLUSIONES.

La tarea de lograr una segmentación precisa es ardua, y tiene un alto costo de procesamiento en el entrenamiento de los algoritmos. La principal contribución de este trabajo es la creación de un modelo de red neuronal que concatena dos redes de tipo U-Net. La adición de una nueva red U-Net, conectada a la primera, con la peculiaridad de que se le añadieron bloques residuales, atenúa el problema de la degradación. Además, en el diseño se establecen conexiones en todos los niveles, de manera que la información obtenida en cada uno de los niveles anteriores se añade a las nuevas características identificadas. Asimismo, para la salida final, se realiza un acoplamiento de cada nivel de esta última red U-Net con los bloques residuales. Este flujo constante de información permite evitar o minimizar la pérdida natural de información que se produce en la contracción de las imágenes.

Los resultados de este estudio son muy similares a los de los métodos de mayor rendimiento y en varios casos superiores, sin embargo, se obtuvieron con un tiempo de entrenamiento considerablemente menor. Pudimos reducir este tiempo de entrenamiento gracias a la arquitectura propuesta, en la que se unieron dos redes neuronales convolucionales totalmente conectadas, con un diseño codificador-decodificador, con la peculiaridad de añadir bloques residuales en la segunda.

La fase de implementación del modelo general, estuvo marcado por dos etapas fundamentales, en la primera, el pre-procesamiento, se manipuló las imágenes originales con el objetivo de resaltar áreas de interés; la segunda etapa estuvo relacionada con la definición en sí de la arquitectura.

El pre-procesamiento de las imágenes de entrada es muy útil en la tarea de segmentación, que consiste en trabajar con imágenes en escala de grises, normalizarlas, aplicar CLAHE y ajuste de gamma. Debido a las pocas imágenes disponibles, el aumento de los datos también fue significativo, esta vez trabajando con parches de los datos originales y volteándolos al azar.

La implementación de la arquitectura nos permitió establecer la cantidad de bloques convolucionales a utilizar, así como los filtros aplicados en cada reducción. También

se establecieron el número de bloques residuales a implementar, encargados estos de disminuir la degradación parcial de las imágenes. Otros conjuntos de hiperparámetros, vitales en el buen funcionamiento de la red fueron configurados en esta fase de implementación.

El uso de pesos dinámicos redondeó el resultado final, logrando un alto puntaje de F1-Score, 0.8250 en DRIVE y 0.8312 en CHASEDB y valores de *accuracy*, 0.9696 en DRIVE y 0.9766 en CHASEDB, métricas estas utilizadas para la validación de nuestro trabajo en su comparación con otros de la literatura buscados de lo cual se puede concluir que la tarea de segmentación se llevó a cabo con altos valores de confiabilidad y precisión.

Los experimentos se llevaron a cabo en dos conjuntos de datos públicos, DRIVE y CHASEDB. El equipo de investigación espera seguir experimentando, por ejemplo, aumentando el número de filtros para determinar si se pueden detectar nuevos patrones.

6. TRABAJOS FUTUROS.

- Incrementar la capacidad de cómputo para así poder utilizar más filtros en las capas de convolución.
- A partir de imágenes de retina existentes, generar su segmentación manual de manera sintética, para su posterior uso en el entrenamiento.
- Utilizar la propuesta de arquitectura en otras tareas de segmentación y detección de objetos.

Dirección General de Bibliotecas UAQ

7.REFERENCIAS.

A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. (2018). Retrieved June 4, 2020, from <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

A Gentle Introduction to the Rectified Linear Unit (ReLU). (n.d.). Retrieved May 20, 2020, from <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/>

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). *Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation*. Retrieved from <http://arxiv.org/abs/1802.06955>

Anaconda | The World's Most Popular Data Science Platform. (n.d.). Retrieved June 4, 2020, from <https://www.anaconda.com/>

asymptote - Drawing a convolution with Tikz - TeX - LaTeX Stack Exchange. (n.d.). Retrieved June 4, 2020, from <https://tex.stackexchange.com/questions/437007/drawing-a-convolution-with-tikz/437114#437114>

Baxter, R., Hastings, N., Law, a., & Glass, E. J. . (2008). [No Title]. In *Animal Genetics* (Vol. 39).

Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., & Yin, Q. (2016). Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks. *Medical Image Computing and Computer-Assisted Intervention*, 1, 140–148. <https://doi.org/10.1007/978-3-319-46723-8>

Cristobal Mollá. (2017). *Discriminación de patologías retinianas mediante procesado avanzado de imágenes de fondo de ojo y algoritmos de Machine Learning Resumen Resum. 52.*

D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J.

M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation*, 117(6), 743–753. <https://doi.org/10.1161/CIRCULATIONAHA.107.699579>

Dasgupta, A., & Singh, S. (2017). a Fully Convolutional Neural Network Based Structured Prediction. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 248–251.

El Abbadi, N., & Al Saadi, E. (2014). Automatic Early Diagnosis of Diabetic Retinopathy Using Retina Fundus Images. *European Academic Research*, 11(9), 11397–11418.

Escala de grises (grayscale) - Diccionario de fotografía y diseño. (n.d.). Retrieved June 4, 2020, from <https://www.fotonostra.com/glosario/escalagrisesgrayscale.htm>

Goodfellow, I., Bengio, Y., & Courville, A. (2015). Deep Learning. *Nature Methods*, 13(1), 35–35. <https://doi.org/10.1038/nmeth.3707>

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. <https://doi.org/10.3389/fpsyg.2013.00124>

Heart, T. H. E., & Eye, T. H. E. (n.d.). *Seeing the Links*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015*, 1, 448–456.

Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., & Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178, 149–162. <https://doi.org/10.1016/j.knosys.2019.04.025>

Kamber, J. H. and M. (2006). *Data Mining Concepts and Techniques*.

Khanal, A., & Estrada, R. (2019). *Dynamic Deep Networks for Retinal Vessel Segmentation*. (1). Retrieved from <http://arxiv.org/abs/1903.07803>

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. 1–15.

<https://doi.org/http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.183050>

3

Kozai, T., & Niu, G. (2015). Introduction. In *Plant Factory: An Indoor Vertical Farming System for Efficient Quality Food Production*. <https://doi.org/10.1016/B978-0-12-801775-3.00001-9>

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9. <https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007>

Laibacher, T., Weyde, T., & Jalali, S. (2018). *M2U-Net: Effective and Efficient Retinal Vessel Segmentation for Resource-Constrained Environments*. Retrieved from <http://arxiv.org/abs/1811.07738>

Leccion.4: Métodos de Mejora y Realce de la Imagen. (n.d.). Retrieved May 20, 2020, from <http://www6.uniovi.es/vision/intro/node33.html>

Liew Gerald. (n.d.). Con los ojos miremos el corazón. Retrieved June 4, 2020, from <https://www.franjaocular.com/index.php/r-aspectosclnicos/114-con-los-ojos-miremos-el-corazon>

Lista de licencias con comentarios - Proyecto GNU - Free Software Foundation. (n.d.). Retrieved June 4, 2020, from <http://www.gnu.org/licenses/license-list.html#PythonOld>

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciampi, F., Ghafoorian, M., ... Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42(December 2012), 60–88. <https://doi.org/10.1016/j.media.2017.07.005>

Machine Learning Note: ReLU function - Clay-Technology World. (n.d.). Retrieved May 20, 2020, from <https://clay-atlas.com/us/blog/2020/02/03/machine-learning-english-note-relu-function/>

OECD iLibrary | OECD Health Statistics. (n.d.). Retrieved June 4, 2020, from

https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics_health-data-en

Oliveira, W. S., Teixeira, J. V., Ren, T. I., Cavalcanti, G. D. C., & Sijbers, J. (2016). Unsupervised retinal vessel segmentation using combined filters. *PLoS ONE*, 11(2), 1–21. <https://doi.org/10.1371/journal.pone.0149943>

Orlando, J. I., Prokofyeva, E., & Blaschko, M. B. (2017). A Discriminatively Trained Fully Connected Conditional Random Field Model for Blood Vessel Segmentation in Fundus Images. *IEEE Transactions on Biomedical Engineering*, 64(1), 16–27. <https://doi.org/10.1109/TBME.2016.2535311>

Owen, C. G., Rudnicka, A. R., Mullen, R., Barman, S. A., Monekosso, D., Whincup, P. H., ... Paterson, C. (2009). Measuring retinal vessel tortuosity in 10-year-old children: Validation of the computer-assisted image analysis of the retina (caiar) program. *Investigative Ophthalmology and Visual Science*, 50(5), 2004–2010. <https://doi.org/10.1167/iovs.08-3018>

Pranav Modi. (n.d.). Hypertensive Retinopathy Article - StatPearls. Retrieved June 4, 2020, from <https://www.statpearls.com/kb/viewarticle/35600>

Python Software Foundation. (2017). Retrieved June 4, 2020, from <https://www.python.org/psf/>

PyTorch. (n.d.). Retrieved June 4, 2020, from <https://pytorch.org/>

Renieblas, G. P., Nogués, A. T., González, A. M., Gómez-Leon, N., & del Castillo, E. G. (2017). Structural similarity index family for image quality assessment in radiological images. *Journal of Medical Imaging*, 4(3), 035501. <https://doi.org/10.1117/1.jmi.4.3.035501>

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage

and organization in the brain. *Psychological Review*, 65(6), 386–408.
<https://doi.org/10.1037/h0042519>

Saleh, M. D., Eswaran, C., & Mueen, A. (2011). An automated blood vessel segmentation algorithm using histogram equalization and automatic threshold selection. *Journal of Digital Imaging*, 24(4), 564–572.
<https://doi.org/10.1007/s10278-010-9302-9>

Sekou, T. B., Hidane, M., Olivier, J., & Cardot, H. (2019). *From Patch to Image Segmentation using Fully Convolutional Networks -- Application to Retinal Images*. Retrieved from <http://arxiv.org/abs/1904.03892>

Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 1–14. <https://doi.org/10.1016/j.infsof.2008.09.005>

Soomro, T. A., Afifi, A. J., Gao, J., Hellwich, O., Khan, M. A. U., Paul, M., & Zheng, L. (2017). Boosting Sensitivity of a Retinal Vessel Segmentation Algorithm with Convolutional Neural Network. *DICTA 2017 - 2017 International Conference on Digital Image Computing: Techniques and Applications, 2017-Decem*, 1–8.
<https://doi.org/10.1109/DICTA.2017.8227413>

Spyder Website. (n.d.). Retrieved June 4, 2020, from <https://www.spyder-ide.org/>

Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4), 501–509.
<https://doi.org/10.1109/TMI.2004.825627>

Stewart, J., Manmathan, G., & Wilkinson, P. (2017). Primary prevention of cardiovascular disease: A review of contemporary guidance and literature. *JRSM Cardiovascular Disease*, 6, 204800401668721.
<https://doi.org/10.1177/2048004016687211>

Sureshjani, A. (2017). *Contextual and deep learning approaches for retinal image analysis Contextual and Deep Learning Approaches for Retinal Image Analysis*.

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and*

Informatics. <https://doi.org/10.1016/j.aci.2018.08.003>

Ting, D. S. W., & Wong, T. Y. (2018). Eyeing cardiovascular risk factors. *Nature Biomedical Engineering*, 2(3), 140–141. <https://doi.org/10.1038/s41551-018-0210-5>

Virtualenv. (n.d.). Retrieved June 4, 2020, from <https://virtualenv.pypa.io/en/latest/>

Wankhade, M. B. (2016). Analysis of Disease using Retinal Blood Vessels Detection. *International Journal Of Engineering And Computer Science*, 5(12), 19644–19647. <https://doi.org/10.18535/ijecs/v5i12.68>

Wong, T. Y., & Mitchell, P. (2007). Chapter 51 – The Eye in Hypertension. *Comprehensive Hypertension*, 633–640. <https://doi.org/10.1016/B978-0-323-03961-1.50054-4>

World Heart Federation. (2016). Factsheet: Enfermedades Cardiovasculares en México. *World Heart Federation, 2010(Gbd 2010)*, 4–5.

Yan, Z., Yang, X., & Cheng, K. T. (2019). A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 23(4), 1427–1436. <https://doi.org/10.1109/JBHI.2018.2872813>












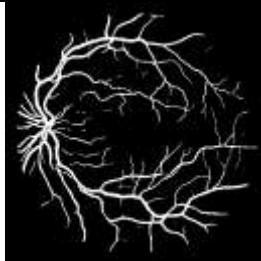


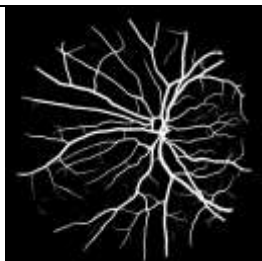
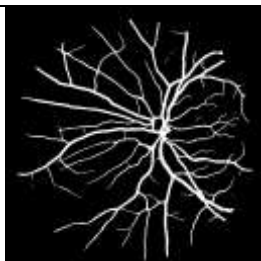




Zadeh, P. D. E. L. A., By, F., & Zadeh, L. A. (n.d.). *No Title*.

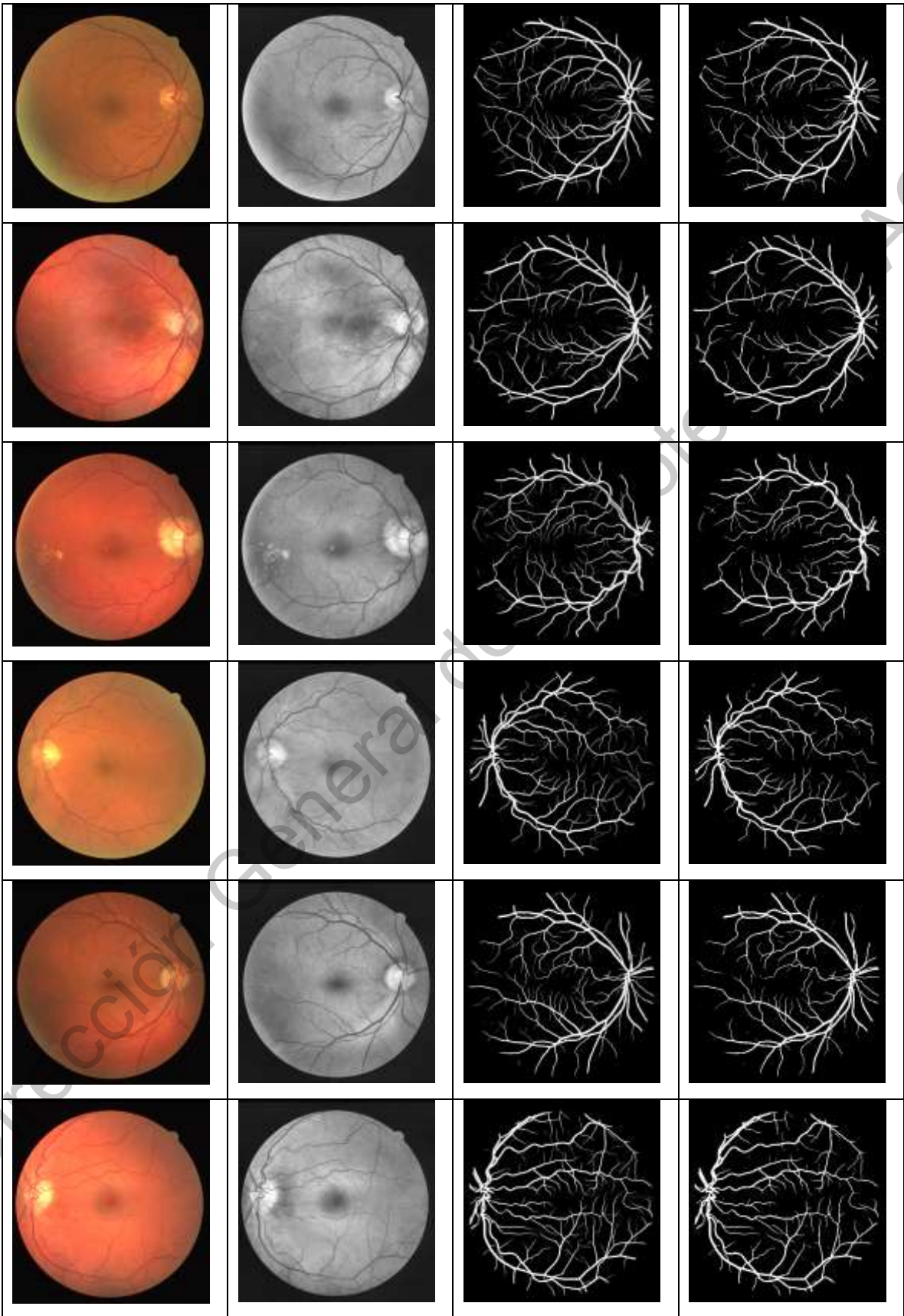
Zhou, M., Jin, K., Wang, S., Ye, J., & Qian, D. (2018). Color Retinal Image Enhancement Based on Luminosity and Contrast Adjustment. *IEEE Transactions on Biomedical Engineering*, 65(3), 521–527. <https://doi.org/10.1109/TBME.2017.2700627>

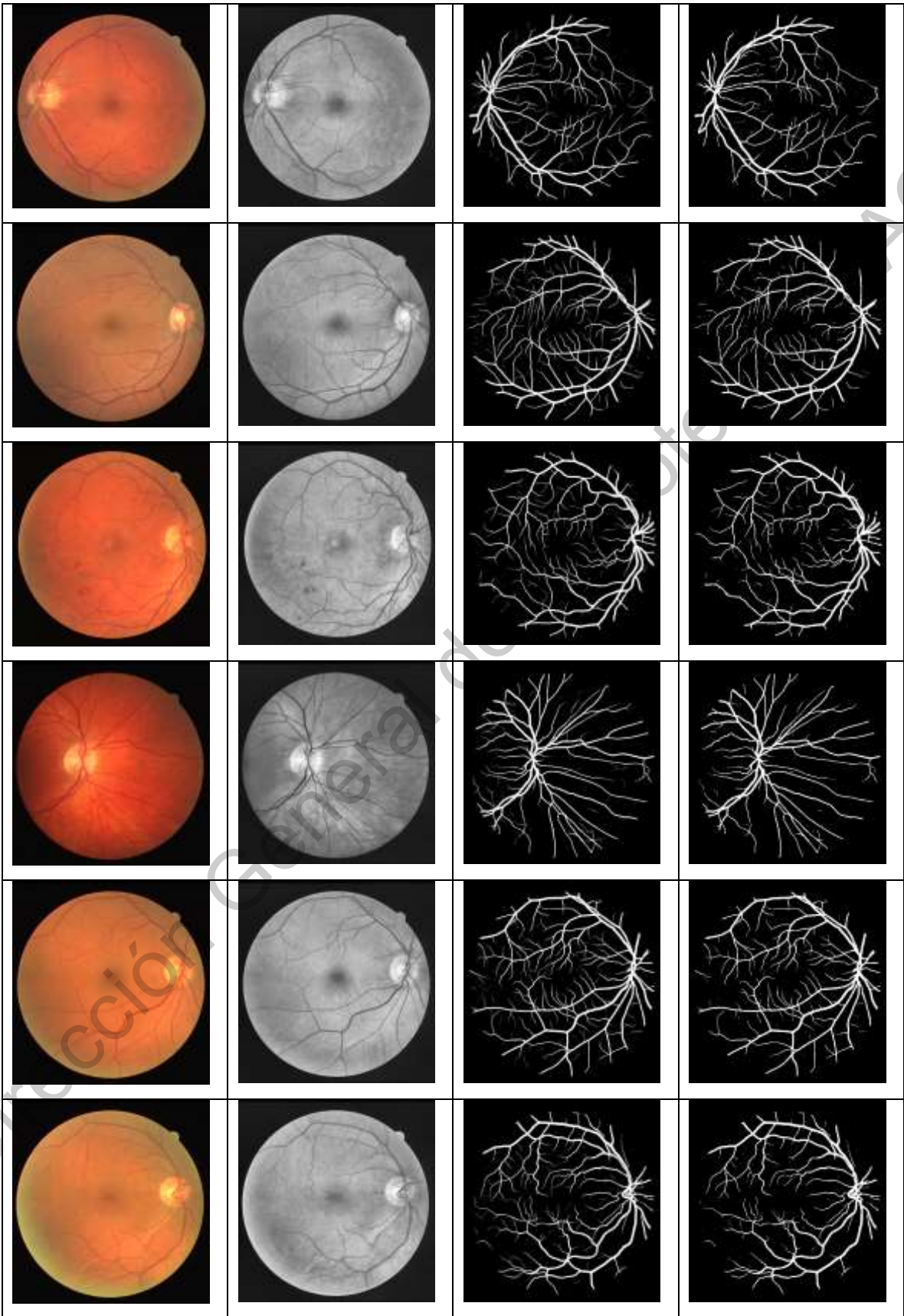
Zhuang, J. (2018). *LadderNet: Multi-path networks based on U-Net for medical image segmentation*. 2–5. Retrieved from <http://arxiv.org/abs/1810.07810>

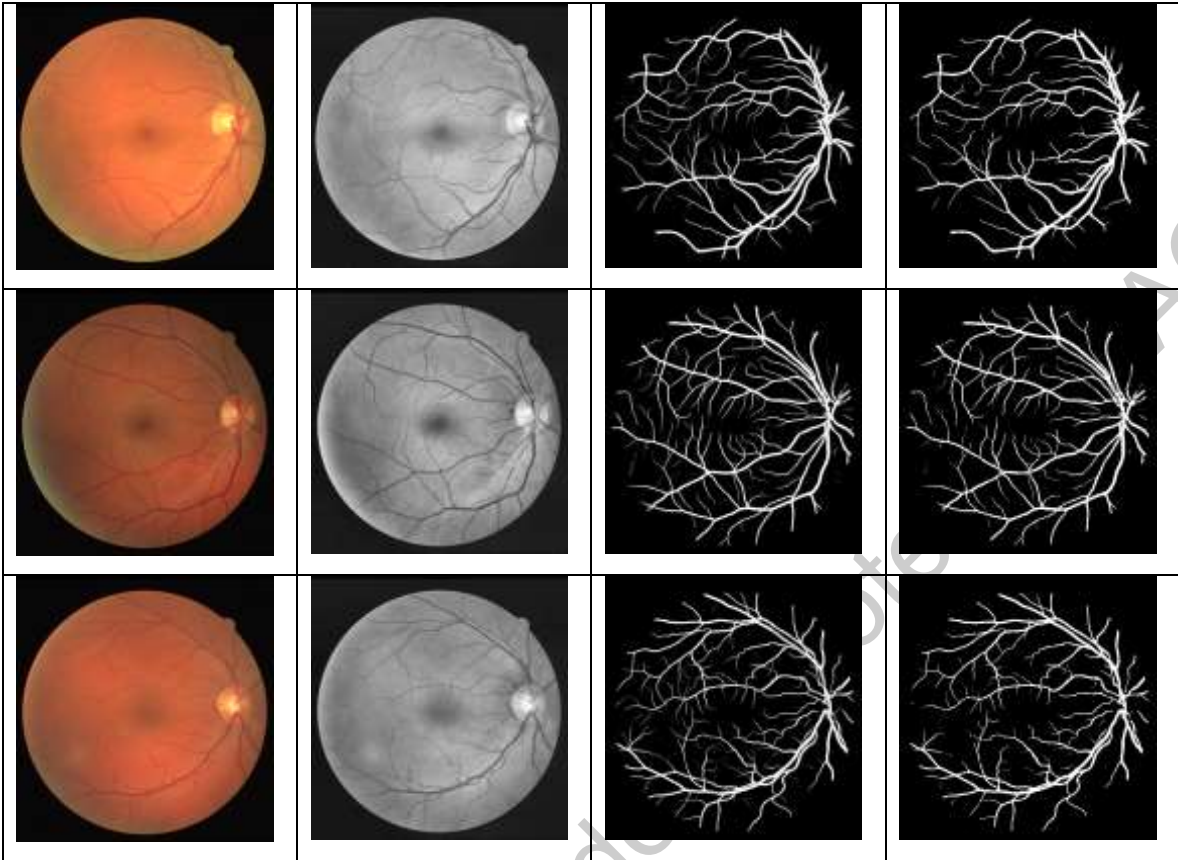
8. ANEXO I.

A continuación, todas las segmentaciones sobre las imágenes de prueba de DRIVE.

Original	Pre-procesada	Ground Truth	Segmentación
			
			
			
			
			














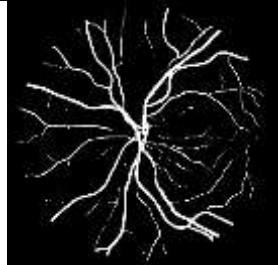



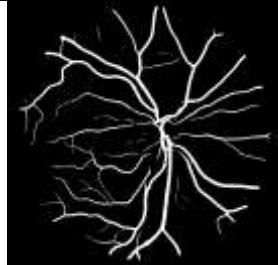


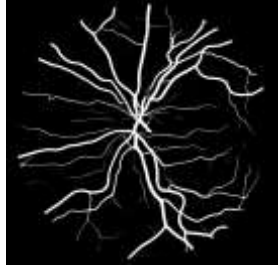
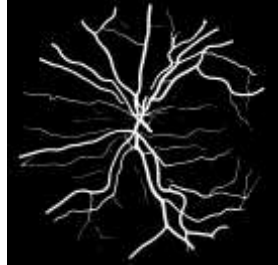


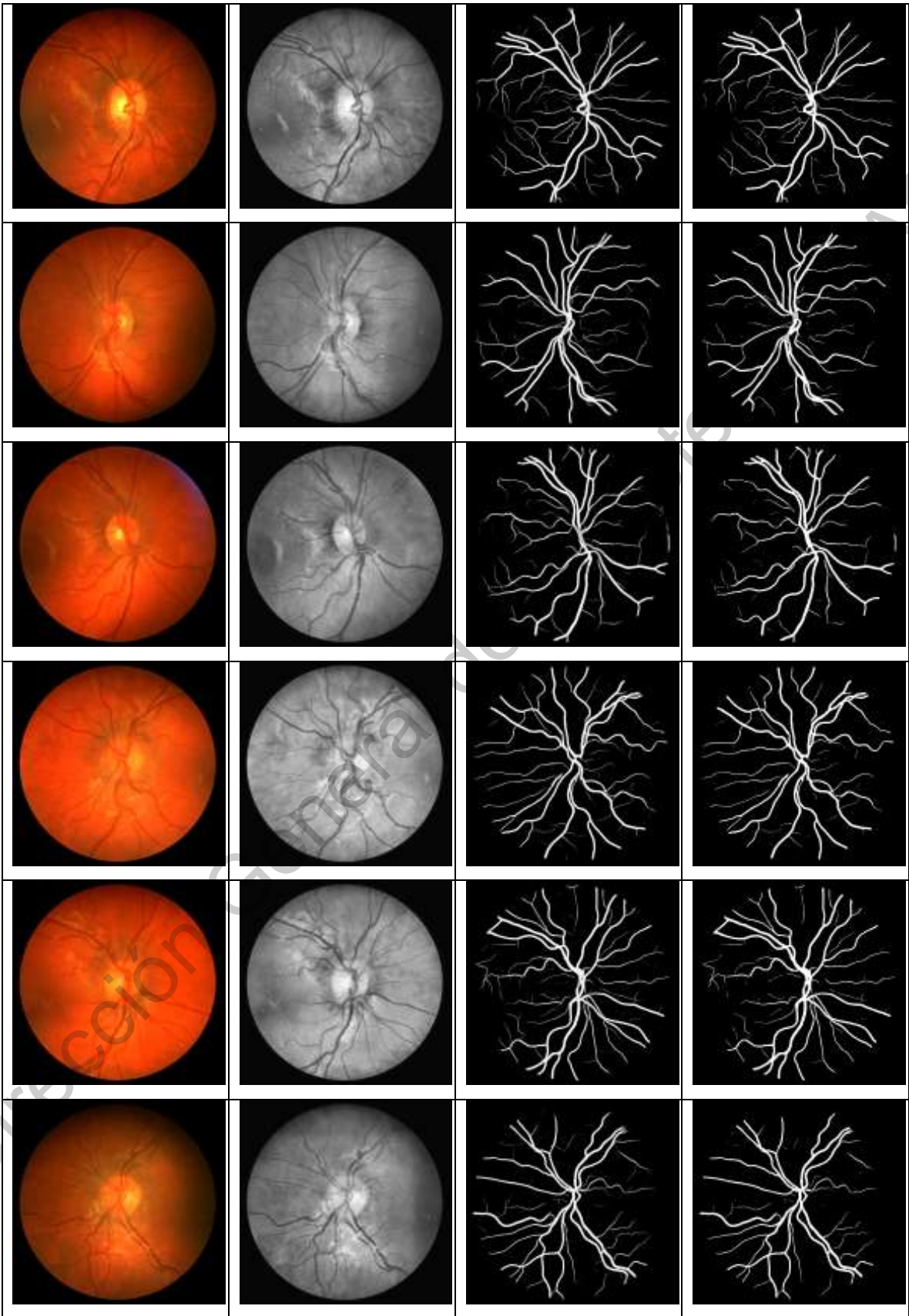


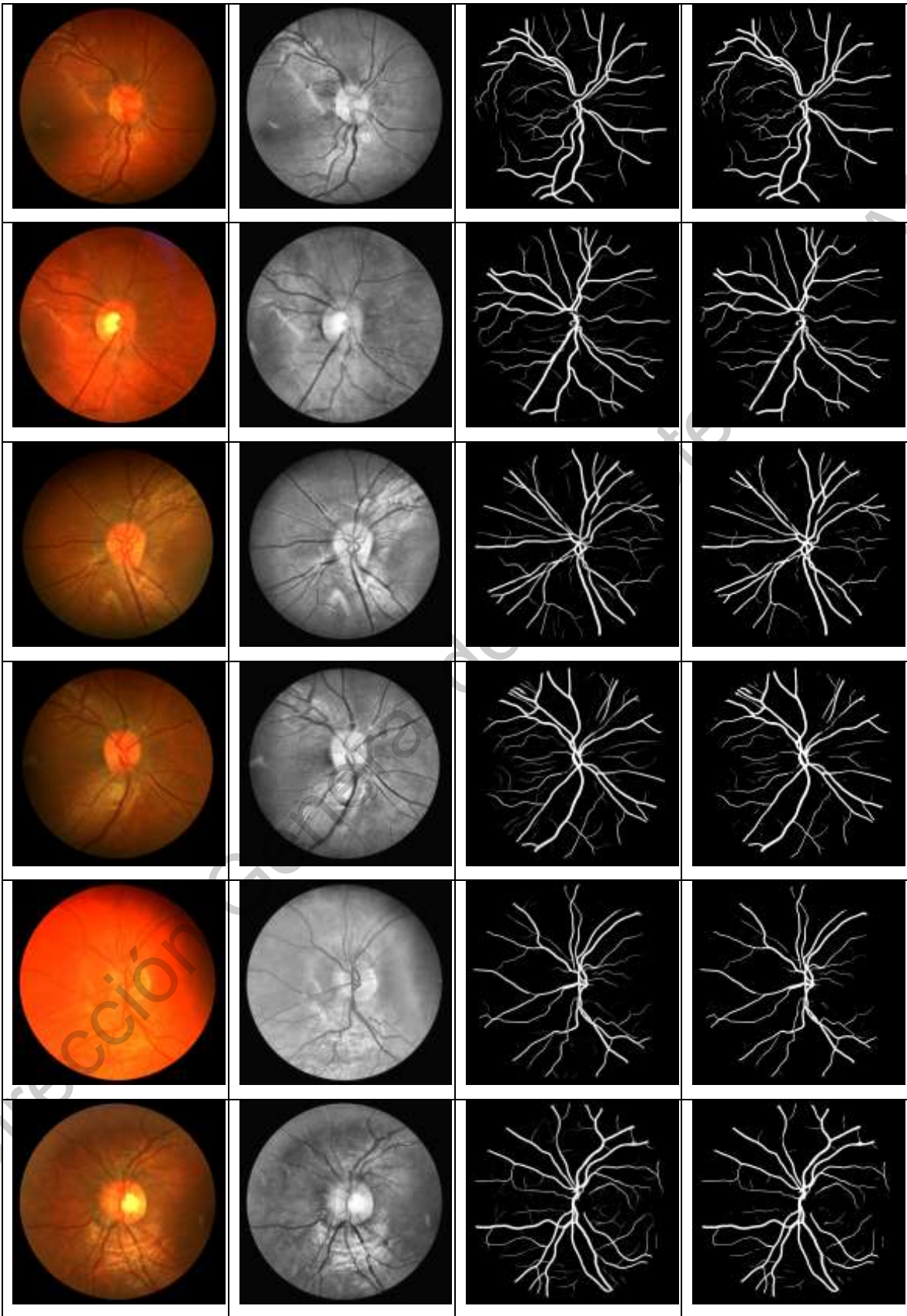
Dirección General de

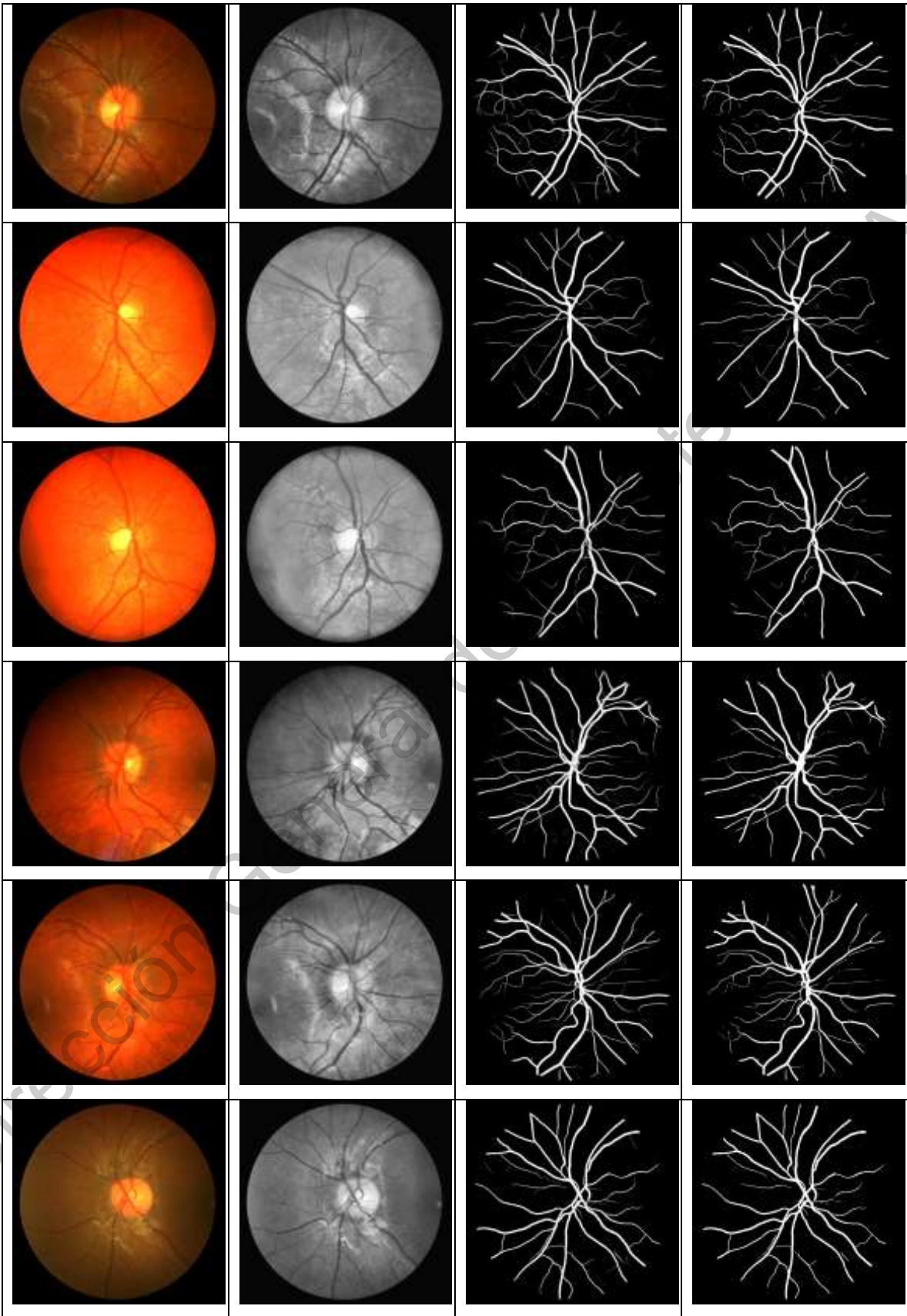
9. ANEXO II.

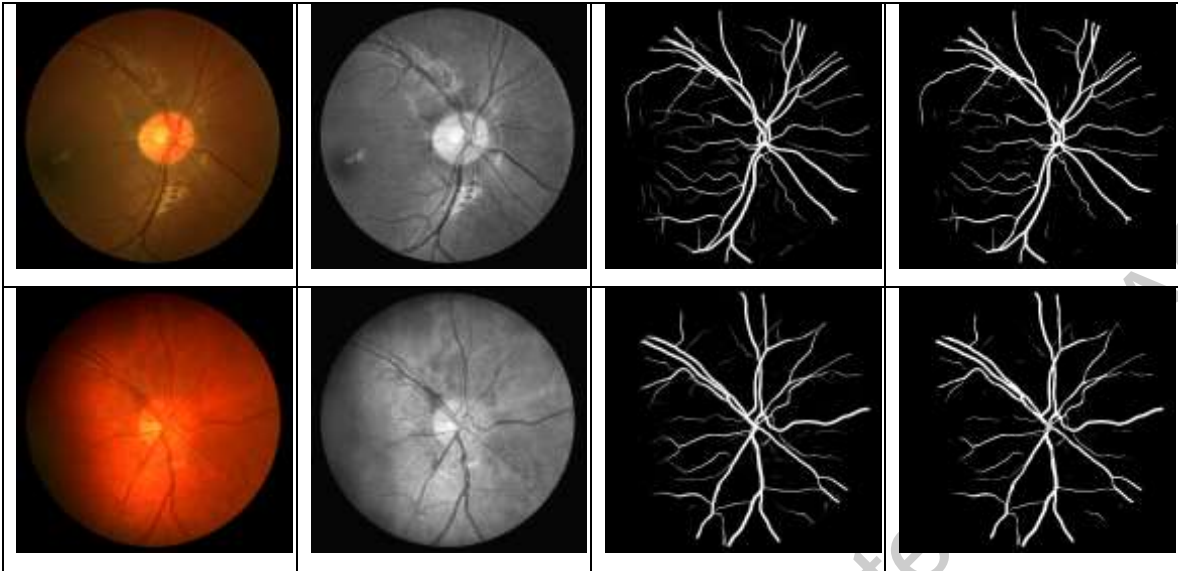
A continuación, todas las segmentaciones sobre las imágenes de prueba de CHASEDB.

Original	Pre-procesada	Ground Truth	Segmentación
			
			
			
			
			









Dirección General de Bibliotecas

10. ANEXO III.

Artículo publicado en revista IEEE Access.

Received February 2, 2020; accepted February 15, 2020; date of publication February 21, 2020; date of current version March 3, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.2975747

Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation

GENDRY ALFONSO FRANCIA¹, (Student Member, IEEE),
CARLOS PEDRAZA, (Senior Member, IEEE), MARCO ACEVES, (Senior Member, IEEE),
AND SAÚL TOVAR-ARRIAGA², (Senior Member, IEEE)

¹Faculty of Engineering, Autonomous University of Querétaro, Santiago de Querétaro 76000, Mexico.

Corresponding author: Saúl Tovar-Arriaga (saul.tovar@uaq.mx)

This work was supported by the National Council of Research of Mexico (CONACYT).

ABSTRACT Retina images are the only non-invasive way of accessing the cardiovascular system, offering us a means of observing patterns such as microaneurysms, hemorrhages and the vasculature structure which can be used to diagnose a variety of diseases. The main goal of this paper is to automate retinal blood vessel segmentation with a good tradeoff between blood vessel classification and training time in the presence of high unbalanced classes. In this work, a novel methodology is proposed using two convolutional neural networks (CNN's), chained to each other. The second CNN has been designed with residual network blocks, which joined to the information flow from the first, give us metrics like recall and F1-Score, which are, in most cases, superior to state of the art in vessel segmentation task. We tested this work on two public datasets for blood vessel segmentation in retinal images showing that this work outperforms many of other contributions by other authors.

INDEX TERMS Retina vessel segmentation, convolutional neural network, U-Net, residual block, F1-Score.

I. INTRODUCTION

Cardiovascular diseases (CVD's) are a whole of interrelated pathologies, including coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic and congenital heart diseases, which are undoubtedly a world leading cause of death [1]. These diseases represent a set of disorders of the heart and blood vessels are most prevalent in developing countries. Habits like smoking, poor diet, sedentary lifestyle, and others are factors that negatively influence CVD's. Therefore, modifications of these factors can reduce their occurrence.

The predictive models which have obtained the most acceptable results in their attempts to predict these events have been the Pooled Cohort equations [2], Framingham [3] and the SCORH project [4]. All of these models try to predict the probability of a CVD, through a follow-up period of 10 years, which avoids predicting events in short periods of time.

The retina is the only place in the human body where veins and arteries can be seen directly. This gives us the opportunity to appreciate their structure, for example the

shape of microvascular changes related to the development of cardiovascular diseases (CVD) [5].

There are other CVD related patterns that can be seen in retina images. One example is the hemorrhage, a disorder of the eye in which bleeding occurs in the light-sensitive tissue on the back wall of the eye [6]. It can be related to diabetic retinopathy, which cause the formation of small fragile blood vessels, which can be easily damaged by high blood sugar level, potentially causing the growth or creation of new blood vessels [7]. Another related disease is hypertensive retinopathy, which consists of damage to the retina from high blood pressure which may result in deformation of the retinal blood vessels. The recognition of this disease may therefore be important in cardiovascular risk stratification of hypertensive patients. Hypertension affect the eyes in several ways, some examples of them are: focal arteriolar narrowing, arteriovenous nicking, hemorrhages, microaneurysms and hard exudates. Related to hypertensive retinopathy, others affection are retinal vein occlusion, artery occlusion, arteriolar emboli, and macroaneurysms [8].

Segmenting blood vessels into retinal images can give us important clues in the diagnosis of ophthalmological diseases such as diabetes, microaneurysms, arteriosclerosis and hypertension. However, this is a task that retina specialists perform

¹The associate editor coordinating the review of this manuscript and approving it for publication was Soon Xin Ng.

11. ANEXO IV.

Constancia de manejo de la lengua inglesa como requisito de egreso.



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS



A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **CONSTAR** que

ALFONSO FRANCIA GENDRY

Presentó el **Examen de Manejo de la Lengua** efectuado el día veintidós de noviembre de dos mil diecinueve, 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 dos de diciembre de dos mil diecinueve.

Atentamente,
"Enlazar Culturas por la Palabra"



LIC. LAURA PÉREZ TÉLLEZ



LPT/evm*CL*FLL-C.-921

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
Campus Aeropuerto, Anillo Vial Fray Junípero Serra S/N, Querétaro, Qro.
E.P. 76140, Tel. 01 (442) 192 12 00 Dirección Ext. 61010,
Secretaría Administrativa 61300, Posgrado 61140, Licenciatura 61070,
Centro de Lenguas 61050, Secretaría Académica 61100 y Planeación 61110

SOMOS UAQ
EDUCAR CRECER CONSOLIDAR