

Tesis de Doctorado

APRENDIZAJE AUTOMÁTICO PARA  
ASISTENCIA AL DIAGNÓSTICO DE  
ENFERMEDADES VISUALES BASADO EN  
IMÁGENES DE FONDO DE OJO

POR  
JOSÉ IGNACIO ORLANDO

Trabajo de Tesis para optar al Título de Doctor en Matemática Computacional e  
Industrial

Directores:  
Dr. Matthew B. Blaschko  
Dr. Mariana del Fresno

Instituto Pladema  
Facultad de Ciencias Exactas  
Universidad Nacional del Centro de la Provincia de Buenos Aires  
Tandil, 2017





---

# Resumen

Las fotografías de fondo de ojo representan una técnica no-invasiva muy empleada por oftalmólogos para detectar retinopatía diabética (RD) y glaucoma, dos de las principales causas mundiales de ceguera prevenible.

En esta tesis se presentan algunas contribuciones al análisis automático de estas imágenes mediante técnicas de aprendizaje automático.

Inicialmente se propone un método de segmentación de vasos sanguíneos basado en el aprendizaje de campos condicionales aleatorios totalmente conectados mediante máquinas de vectores de soporte de salida estructurada. Este enfoque permite obtener representaciones precisas del árbol vascular, que son luego utilizadas en el contexto de dos técnicas para la detección de glaucoma y RD.

Para detectar glaucoma, se plantea transferir redes neuronales convolucionales (CNNs) preentrenadas con datos no médicos. Las imágenes son adaptadas inicialmente mediante métodos de preprocesamiento tradicionales, y los descriptores extraídos se utilizan para entrenar modelos regularizados de regresión logística, obteniendo resultados competitivos con el estado del arte.

Finalmente, se introduce una técnica para detectar lesiones rojas típicas de la RD basada en hibridar descriptores aprendidos utilizando una CNN y otros diseñados manualmente. Posteriormente se utiliza un Bosque Aleatorio entrenado con estos valores para identificar lesiones candidatas, con alta eficacia en diversas bases de datos.



---

# Abstract

Fundus images are a non-invasive imaging modality that is typically used by ophthalmologists to assess the retina. They are widely applied for detecting diabetic retinopathy (DR) and glaucoma, which are leading causes of preventable blindness in the world.

In this thesis we contribute with novel tools for automated fundus image analysis based on machine learning. First, we propose a vessel segmentation method based on learning fully connected conditional random fields using structured output support vector machines. This approach allows to recover accurate segmentations of the retinal vasculature that are afterwards applied in the context of two deep learning based techniques for glaucoma and DR detection.

For automated glaucoma assessment, we propose to transfer convolutional neural networks (CNNs) that were pre-trained using non-medical data. Images are adapted using state of the art preprocessing methods before feeding the CNNs, and the extracted features are used to train regularized logistic regression classifiers, achieving results that are competitive with other methods.

For DR screening, we introduce a red lesion detection approach based on hybridizing deep learned and hand crafted features. A Random Forest classifier is trained on these features to identify true lesion candidates, reporting state of the art performance in benchmark data sets.



---

# Agradecimientos

Escribir los agradecimientos en el medio del lío que supone cerrar una tesis es difícil. Más cuando uno intenta por todos los medios tratar de no caer en lugares comunes, o cuando se intenta nombrar a todxs lxs que pusieron algo de sí para llegar al final de este largo camino. Largo camino, bueno, primer lugar común. Va a ser difícil.

En primer lugar quiero darle las gracias a mis directores, Mariana del Fresno y Matthew B. Blaschko, por haberme acompañado y aconsejado durante el desarrollo de esta tesis. Mariana, valoro enormemente que hayas confiado en mí y me hayas brindado la posibilidad de explorar mi vocación como científico. Matthew, tu colaboración desde lo técnico y lo humano fue crucial para mí, y es en gran parte responsable de que haya podido llegar con éxito a este día. Aprendí muchísimo trabajando con vos, y te estaré siempre agradecido por ello y por ese optimismo y esa buena onda que siempre tuviste para conmigo.

Quiero agradecerle también a los miembros del jurado, a la Dra. Virginia Ballarín (UNMdP), el Dr. Diego Milone (UNR), el Dr. José Massa (UNICEN) y la Dra. Clarisa Sánchez (Radboud University Medical Centre, Holanda), por haber dedicado parte de su valioso tiempo a leer esta tesis y a proponer mejoras y aportes para el documento final.

No puedo obviar tampoco un agradecimiento especial para Alejandro Clause, Pablo Lotito e Ignacio Larrabide, compañeros de Pladema que confiaron en mí y me permitieron trabajar en otros proyectos, derivados de esta tesis en algunos casos, y completamente ajenos a ella en otros. También quiero agradecerle a Claudia Marinelli por ayudarme en el análisis estadístico de los resultados, y a Diego Dalponte, quien confió en mí para acompañarlo en la tarea de dar clases en la universidad,

brindándome la libertad de tomar decisiones junto a él, siempre con la mejor disposición y con muy buena onda, algo invaluable en un mundo tan difícil como el académico.

Quiero agradecerle también a los médicos e investigadores que colaboraron en el desarrollo de esta tesis doctoral: Elena Prokofyeva, Karel van Keer y João Barbosa Breda. Sus aportes desde el punto de vista clínico y epidemiológico fueron invaluableles al momento de idear las soluciones propuestas en este trabajo y de plantear futuros desafíos y aplicaciones.

También quiero agradecerle a las instituciones educativas y científicas de Argentina. Y me quiero extender un poco más aquí. Somos el país más austral del mundo. Estamos al sur, al final del mapa. Vivimos lejos de los grandes centros del conocimiento, de Europa, el Lejano Oriente, Estados Unidos. Y aún así, como científicos, nos damos maña para encontrarle la vuelta a nuestros problemas, para integrarnos con el continente y con el resto del mundo. Lo hicieron las generaciones pasadas, en las peores condiciones posibles, hostigados por las políticas represivas de la dictadura cívico-militar-clerical, primero, y por el rabioso neoliberalismo de los '90, más tarde. Lo hicimos nosotros, también, los más nuevos, ayudados por un país que se decidió a acompañarnos y a protegernos del mercado. Hoy, producto de las profundas transformaciones que sufrió Argentina entre 2003 a 2015, puedo decir con orgullo que soy producto de la educación pública y gratuita de mi país. Fui al jardín, fui a la escuela primaria y a la secundaria, fui a la universidad e hice mi postgrado en aulas que son de todos. Y en la universidad pública, en la UNICEN, descubrí mi verdadera vocación, la de ser científico. No creo que sea casualidad. No es casual. No fue magia: el crecimiento sostenido y la profundización del modelo científico en nuestro país han sido claves en este descubrimiento que muchos de mi generación experimentamos. Hoy, en tiempos en los que un nuevo orden económico parece llevarse puestas tantas conquistas, apelo de corazón a esa voluntad de fierro que la ciencia argentina ha demostrado por años. Vamos a sobrevivir a esto también, porque sobrevivimos a mucho. Y lo haremos en el convencimiento de que lo que hacemos, sirve. Y de que la Patria es el otro.

Me entusiasmé. Vuelvo, mejor.

A mis amigos, Mante, Emma, Mari, Cris, Nacha, Enzo, Adrián, Fernanda, Vicky

y Mai, a todos ellos los abrazo, nomás. Y les digo que los quiero, que los necesito, y que la vida con ellos cerca es siempre mucho más feliz.

Quiero dedicarle unas líneas en estos agradecimientos tan largos ya a mis compañeros de Pladema, en especial a ese archipiélago de gente copada que rodeó mis auriculares en los últimos años. A Delfi, Lover, Mante, Laucha, Lean, Mati, Trapa, Blito, Zeque, Hernán, Sil y Nico, y también al Boro, a Mara, y un poco también a Javier, gracias por el mate, la cotilla y el Slack. De igual forma, un abrazo y otro gracias a los compañeros que conocí en otros lugares del mundo: a los de París, Siddhartha, Aline, Katherina, Eugenio y Puneet, y a los de Leuven, en especial a Maxim y a Amal, dos grandes amigos en cuya compañía la distancia a Argentina siempre tiende a 0.

Gracias también a toda la gente maravillosa que conocí en la Casa Argentina en París, que me hicieron sentir en casa cuando me separaba un océano de mi familia. Fede, Pilu, Maru, Allo y Gus, los quiero con el alma. Y gracias también a Marcelo, director por aquel entonces de la Casa Argentina, quien hizo lo imposible para que todos disfrutáramos al máximo la experiencia de vivir en la Ciudad Luz.

Esto es un lugar común, de vuelta, pero quiero agradecerle a mi familia, obvio. A mis viejos y a mis hermanos, por aceptarme como soy y por siempre apoyarme cuando más los necesito. A mi abuela Norma, fanática número uno de todo lo que hago, principal espectadora de mis experimentos de divulgador. A mis tías, primos, a Nati y Martín, a toda esa familia gigante de Las Flores que ya ni sé como nombrar sin que parezca una lista del colegio.

Y por último, pero también primero, a Sebas, que se sumó con este viaje ya empezado, pero que está ahí cuando lo necesito, que me acompaña y me sostiene cuando parezco aflojar, y al que elijo todos los días para compartir la vida.



---

A mi abuela, Norma. La del corazón más grande del mundo.  
Y a mis abuelos que ya no están. Pero seguro que sí.



---

# Índice general

<b>Agradecimientos</b>	<b>7</b>
<b>1. Introducción</b>	<b>24</b>
1.1. Motivación . . . . .	24
1.2. Aprendizaje automático . . . . .	28
1.3. Contribuciones y organización de la tesis . . . . .	29
<b>2. Antecedentes</b>	<b>33</b>
2.1. El ojo y la retina . . . . .	33
2.2. Enfermedades de la retina . . . . .	36
2.2.1. Retinopatía diabética . . . . .	36
2.2.2. Glaucoma . . . . .	40
2.3. Imagenología de la retina . . . . .	43
2.3.1. Fotografías de fondo de ojo . . . . .	43
2.3.2. Angiografía por fluorescencia . . . . .	46
2.3.3. Tomografía de coherencia óptica . . . . .	48
2.4. Análisis de imágenes de fondo de ojo para diagnóstico asistido por computadora . . . . .	49
2.4.1. Análisis general del estado del arte . . . . .	50
2.4.2. Enfoques basados en caracterización vascular . . . . .	52
2.4.3. Detección de estructuras patológicas . . . . .	54
2.4.4. Caracterización de las imágenes mediante descriptores globales	59
2.4.5. Métodos basados en aprendizaje profundo ( <i>deep learning</i> ) . .	62

---

<b>3. Segmentación de vasos sanguíneos</b>	<b>66</b>
3.1. Motivación . . . . .	68
3.2. Métodos . . . . .	71
3.2.1. Campos condicionales aleatorios (CRFs) para segmentación de imágenes . . . . .	72
3.2.2. Aprendizaje de CRFs mediante SVM de salida estructurada .	76
3.2.3. Descriptores . . . . .	78
3.2.4. Adaptación automática de los modelos a imágenes de diferente resolución . . . . .	80
3.3. Materiales y Evaluación . . . . .	83
3.3.1. Bases de datos utilizadas . . . . .	83
3.3.2. Métricas de evaluación . . . . .	84
3.3.3. Configuración del método . . . . .	85
3.4. Experimentos y resultados . . . . .	88
3.4.1. Resultados . . . . .	88
3.4.2. Comparación con otros métodos . . . . .	94
3.5. Discusión . . . . .	100
3.6. Conclusiones . . . . .	103
<b>4. Transferencia de redes neuronales convolucionales para detección de glaucoma</b>	<b>104</b>
4.1. Motivación . . . . .	105
4.2. Métodos . . . . .	108
4.2.1. Preprocesamiento de las imágenes . . . . .	109
4.2.2. Regresión logística regularizada con normas $\ell_1$ y $\ell_2$ . . . . .	111
4.3. Materiales . . . . .	113
4.3.1. Redes neuronales convolucionales . . . . .	113
4.3.2. Bases de datos utilizadas . . . . .	114
4.4. Resultados . . . . .	116
4.5. Discusión . . . . .	119
<b>5. Detección de lesiones rojas para <i>screening</i> de RD</b>	<b>122</b>
5.1. Motivación . . . . .	124

---

5.2. Métodos . . . . .	127
5.2.1. Detección de lesiones candidatas . . . . .	128
5.2.2. Descriptores basados en CNNs . . . . .	131
5.2.3. Extracción de descriptores diseñados manualmente . . . . .	135
5.2.4. Clasificación de lesiones candidatas usando Bosques Aleatorios . . . . .	139
5.3. Configuración experimental . . . . .	140
5.3.1. Materiales . . . . .	140
5.3.2. Ajuste de parámetros . . . . .	142
5.3.3. Métricas de calidad . . . . .	143
5.4. Resultados . . . . .	144
5.4.1. Evaluación por lesión . . . . .	144
5.4.2. Evaluación por imágenes . . . . .	150
5.4.3. Análisis de los descriptores . . . . .	152
5.5. Discusión . . . . .	154
5.6. Conclusiones . . . . .	159
<b>6. Conclusiones y trabajos futuros</b>	<b>160</b>
6.1. Contribuciones . . . . .	160
6.2. Líneas futuras de trabajo . . . . .	162
6.2.1. Segmentación y caracterización de vasos sanguíneos . . . . .	162
6.2.2. Detección automatizada de glaucoma . . . . .	164
6.2.3. Detección automática de RD . . . . .	165
<b>Bibliografía</b>	<b>167</b>

---

# Índice de figuras

2.1. Representación esquemática del ojo. Fuente: [42]. . . . .	34
2.2. Complicaciones microvasculares causadas por la hiperglucemia. (a) Capilar sano. (b) La hiperglucemia debilita las paredes de los vasos e incrementa el flujo sanguíneo, permitiendo además la fuga de algunos lípidos de la sangre. (c) Las paredes debilitadas de los vasos y el aumento de la presión en las mismas favorece la formación de microaneurismas. (d) La presión elevada en el interior de la microaneurisma provoca eventualmente su ruptura, produciendo hemorragias. . . . .	38
2.3. Etapas principales de la retinopatía diabética. Fuente: [19]. . . . .	38
2.4. Flujo del humor acuoso. En celeste se observa el recorrido del flujo, y en rojo se indica el ángulo iridocorneal. . . . .	40
2.5. Fotografía de fondo de ojo de un paciente sano, extraída del conjunto de imágenes HRF [145]. . . . .	44
2.6. Cámara de fondo de ojo. Fuente: Ralf Roletschek / Wikipedia [171]. . . . .	44
2.7. Volk iNview, un dispositivo para la captura de fotografías de fondo de ojo basado en teléfonos inteligentes. Fuente: [148]. . . . .	45
2.8. Fotografías de fondo de ojo para el diagnóstico de enfermedades visuales. . . . .	45
2.9. Angiografías por fluroescencia. Fases. Fuente: [167]. . . . .	47

2.10. Imagen OCT y capas interiores de la retina, obtenidas utilizando un algoritmo de segmentación automático [111]. (a) Imagen X-Z del volumen de la OCT. (b) Resultados de la segmentación: capa de fibras nerviosas (NFL), capa de células ganglionares (GCL), capa plexiforme interna (IPL), capa nuclear interna (INL), capa plexiforme externa (OPL), capa nuclear externa (ONL), unión del segmento interior y anterior (ISOSJ), fotoreceptores del segmento exterior (OPR), espacio virtual subretiniano (SRVS-de profundidad 0 en pacientes sanos), y epitelio pigmentario retiniano (RPE). Figura extraída de [191]. . . . .	49
2.11. Captura de pantalla de la herramienta SIVA (Singapore I Vessel Assessment). Fuente: [1]. . . . .	52
2.12. Ejemplos de lesiones rojas, como se observan en las imágenes de fondo de ojo. Nótese que algunas de ellas son apenas distinguibles. Fuente: conjunto de entrenamiento de DIARETDB1 [96]. . . . .	55
2.13. Ejemplos de lesiones brillantes, como se observan en las imágenes de fondo de ojo. Fuente: conjunto de entrenamiento de DIARETDB1 [96].	56
2.14. Defectos en la capa de fibras nerviosas retiniana ( <i>Retinal Nerve Fiber Layer</i> , RNFL) producto del glaucoma, según se manifiestan en las fotografías de fondo de ojo. Fuente: [75]. . . . .	58
2.15. Atrofia peripapilar ( <i>Peripapillary Atrophy</i> , PPA) producto del glaucoma. Source: [116]. . . . .	58
2.16. Mapas de anomalías obtenidos utilizando el método de detección de RD basado en <i>deep learning</i> propuesto por Gargeya <i>et al.</i> [69]. Fuente: [69]. . . . .	64
3.1. Una fotografía de fondo de ojo del conjunto de test de DRIVE [141] y su segmentación de los vasos sanguíneos asociada, obtenida manualmente por un experto y utilizando el enfoque propuesto en esta tesis.	68

3.2. Preprocesamiento de las imágenes y ejemplos de descriptores unarios y <i>pairwise</i> . (a) Imagen a color original. (b) Canal verde invertido luego de la expansión del borde. (c) Respuestas al detector de líneas de Nguyen <i>et al.</i> ( $l = 15$ ). (d) Respuestas a la transformada wavelet 2D de Gabor propuesta por Soares <i>et al.</i> ( $a = 3$ ). (e) Imagen invertida luego de corregir la iluminación despereja. (f) Imagen realzada utilizando el método de Zana y Klein ( $l = 9$ ). . . . .	81
3.3. Resultados de segmentación obtenidos en DRIVE. (a) Imagen 04 de DRIVE. (b) Segmentación de referencia ( <i>ground truth</i> ). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF. . . . .	91
3.4. Resultados de segmentación obtenidos en CHASEDB1. (a) Image_05L de CHASEDB1. (b) Segmentación de referencia ( <i>ground truth</i> ). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF. . . . .	92
3.5. Resultados de segmentación obtenidos en un caso patológico serio de STARE. (a) Imagen im0005 de STARE. (b) Segmentación de referencia del primer observador experto ( <i>ground truth</i> ). (c) Segmentación del segundo observador experto. (d) Segmentación obtenida utilizando únicamente los potenciales unarios. (e) Segmentación obtenida utilizando el FC-CRF. . . . .	93
3.6. Resultados de segmentación obtenidos en HRF. (a) Imagen 11.g. (b) Segmentación de referencia ( <i>ground truth</i> ). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF. . . . .	94
3.7. Detección de vasos delgados en condiciones de bajo contraste. (a) Detalle de la imagen 11.g de HRF. (b) Imagen preprocesada. (c) Segmentación de referencia. (d) Segmentación obtenida utilizando únicamente los potenciales unarios. (e) Segmentación obtenida utilizando el FC-CRF. . . . .	95

3.8.	Curvas ROC en DRIVE, CHASEDB1 y HRF, utilizando únicamente los potenciales unarios (UP, línea punteada) o el FC-CRF (línea sólida), y rendimiento del segundo usuario experto (HO). . . . .	98
3.9.	Costo computacional de la inferencia en el FC-CRF en todos los conjuntos de datos utilizados para la evaluación. . . . .	99
4.1.	Representación esquemática del método para transferir CNNs pre-entrenadas sobre datos no médicos a la detección de glaucoma en fotografías de fondo de ojo. . . . .	109
4.2.	Estrategias de preprocesamiento evaluadas. Primer grupo: sin CLAHE. Segundo grupo: con CLAHE. De izquierda a derecha: imagen original, FOV recortado, area peripapilar ( <i>peripapillary area</i> , PPA), y cabeza del nervio óptico ( <i>optic nerve head</i> , ONH). Primera fila: imágenes originales. Segunda fila: imágenes luego de aplicar <i>inpainting</i> de los vasos sanguíneos. . . . .	110
4.3.	Sustracción de vasos. . . . .	112
4.4.	Ejemplo de una muestra de las imágenes naturales del conjunto ImageNet 2012, utilizadas para entrenar tanto OverFeat como VGG-S. . .	113
4.5.	Area bajo las curvas ROC promedio obtenidas a partir de 200 pruebas [58] utilizando los descriptores calculados por OverFeat y clasificadores por regresión logística regularizada con normas $\ell_1$ o $\ell_2$ , respectivamente. Primera fila: utilizando imágenes sin <i>inpainting</i> de vasos sanguíneos. Segunda fila: utilizando imágenes con <i>inpainting</i> de los vasos. La línea azul corresponde al mayor valor de AUC para la fila. .	118
5.1.	Ejemplos de lesiones rojas observadas en fotografías de fondo de ojo del conjunto de imágenes DIARETDB1 [96]. . . . .	124
5.2.	Representación esquemática del método propuesto para la detección de lesiones rojas. . . . .	128
5.3.	Detección de lesiones rojas candidatas. Ver Sección 5.2.1 para una descripción detallada del proceso. . . . .	129
5.4.	Efecto que la expansión del FOV tiene sobre las lesiones candidatas ubicadas de manera cercada al borde del FOV. . . . .	130

---

5.5. Conjunto de entrenamiento de la CNN. Muestra aleatoria de 200 parches para (a) candidatos falsos y (b) lesiones verdaderas. Ver Sección 5.2.2 para detalles acerca de la construcción del conjunto de entrenamiento. . . . .	133
5.6. Descriptor basado en la segmentación de los vasos sanguíneos. (a) <i>I</i> . (b) <i>I</i> con lesiones candidatas superpuestas. (c) Segmentación de los vasos sanguíneos. (d) Segmentación de los vasos sanguíneos tras remover estructuras espurias. (e) Segmentación de los vasos sanguíneos luego de aplicar el <i>closing</i> morfológico. . . . .	137
5.7. Evaluación por lesión en el experimento 1. Curva FROC y valores de CPM obtenidos sobre el conjunto de test de DIARETDB1. . . . .	145
5.8. Evaluación por lesión y para cada tipo de lesión en el conjunto de test de DIARETDB1. . . . .	147
5.9. Evaluación por lesión en el experimento 2. Curva FROC y valores de CPM obtenidos sobre e-ophta. . . . .	148
5.10. Resultados cualitativos. (a) Imagen image015 del conjunto de test de DIARETDB1. (b) Etiquetado de referencia a un nivel de acuerdo entre expertos > 25 % (c) Lesiones rojas detectadas utilizando el método propuesto, umbralando las probabilidades a 0.644, que corresponde a un valor promedio de FPI de 1. (d) Detalle de (c) mostrando lesiones no etiquetas manualmente pero identificadas por el método. . . . .	149
5.11. Evaluación por imagen en MESSIDOR. Curvas ROC para (a) <i>screening</i> de RD (R1 vs. R2, R3 y R4), y (b) detección de pacientes referibles (R1 y R2 vs. R3 y R4). . . . .	150
5.12. Evaluación por imagen sobre e-ophta. Curva ROC curve para <i>screening</i> de RD. . . . .	151
5.13. Filtros aprendidos por la primera capa de la CNN propuesta, obtenidos para cada experimento en la Tabla 5.4. . . . .	153

- 
- 5.14. Visualización  $t$ -SNE de los entornos extraídos de los datos de test de DIARETDB1, mapeados utilizando los descriptores aprendidos con *deep learning*, los diseñados manualmente y el enfoque híbrido. Izquierda: etiquetas de color para cada muestra de test. Derecha: parches alrededor de los candidatos, visualizados utilizando los mapeos de  $t$ -SNE. El detalle para cada tipo de lesión candidata se observa en la Figura 5.15. . . . . 155
- 5.15. Detalles extraídos de la visualización  $t$ -SNE de la Figura 5.14 para diferentes tipos de lesiones rojas candidatas (lesiones reales, estructuras vasculares, suciedad en el lente de la cámara y detecciones falsas en curvas de los vasos en el disco óptico). . . . . 156

---

# Índice de tablas

2.1. Clasificación de las herramientas para el diagnóstico y la detección asistida por computadora de RD y glaucoma. . . . .	51
3.1. Evolución del valor-F1 durante el proceso de selección secuencial de indicadores en DRIVE. . . . .	87
3.2. Evaluación cuantitativa de los resultados obtenidos en DRIVE, STARE, CHASEDB1 y HRF, utilizando únicamente los potenciales unarios (UP) o el CRF totalmente conectado (FC-CRF). . . . .	89
3.3. Comparación de los valores promedios de $Se$ , $Sp$ , $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en DRIVE. . . . .	96
3.4. Comparación de los valores promedios de $Se$ , $Sp$ , $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en STARE. . . . .	97
3.5. Comparación de los valores promedios de $Se$ , $Sp$ , $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en CHASEDB1 y HRF. . . . .	99

---

5.1. Arquitectura de la CNN. Las capas convolucionales ( <i>conv</i> ) indican el ancho, alto y profundidad de cada filtro aprendido. Las capas de <i>pooling</i> ( <i>pool</i> ) incluyen la dimensión de la operación de <i>pooling</i> y el tamaño de paso. El <i>dropout</i> es únicamente aplicado luego de la primera capa convolucional, con una probabilidad de <i>dropout</i> baja. . . . .	135
5.2. Resumen de los descriptores diseñados manualmente ( <i>hand crafted features</i> , HCF) utilizados para complementar la CNN. . . . .	138
5.3. Distribución de los niveles de RD en el conjunto de imágenes MESSIDOR, y criterio diagnóstico empleado. MA = microaneurismas, HE = hemorragias y NV= neovascularizaciones. . . . .	140
5.4. Configuración experimental. $\beta$ es el valor de la pérdida por entropía cruzada balanceada por clase (Ecuación (5.5)). . . . .	142
5.5. Valores de CPM y sensibilidad por lesión para FPI= 1 para los experimentos 1 y 2. . . . .	148
5.6. Comparación de resultados para <i>screening</i> de RD o detección de pacientes referibles sobre MESSIDOR. Los valores de <i>Se</i> corresponden a aquellos obtenidos para $Sp = 50\%$ . . . . .	152

---

# Capítulo 1

## Introducción

En esta tesis se presentan algunas contribuciones metodológicas novedosas en el campo del análisis automatizado de fotografías de fondo de ojo. En particular, las herramientas aquí propuestas se basan en la utilización de técnicas de aprendizaje automático, y son aplicadas al diagnóstico asistido por computadora de dos enfermedades visuales que están entre las principales causas de ceguera prevenible a nivel mundial: la retinopatía diabética y el glaucoma. En la Sección 1.1 se describen en líneas generales las principales motivaciones de este trabajo. En la Sección 1.2, por otro lado, se brindan algunos detalles respecto a los métodos de aprendizaje automático, y cómo este tipo de técnicas pueden aplicarse a la resolución de los problemas abordados en este trabajo. Finalmente, la Sección 1.3 resume las contribuciones realizadas, las publicaciones científicas derivadas de cada una de ellas y la organización general de la tesis.

### 1.1. Motivación

De acuerdo a la Organización Mundial de la Salud (OMS), el número estimado de personas con afecciones visuales en el mundo asciende a 285 millones. De esta cifra, 39 millones son personas ciegas, y 246 millones son personas con visión disminuida [149]. Un 65 % de los individuos con algún tipo de discapacidad visual y un 82 % de los no videntes tienen 50 años o más. Con el envejecimiento natural que hoy por hoy experimenta la población mundial, pues, es de esperar que la prevalencia de este tipo

de afecciones crezca significativamente [149]. En el caso particular de América Latina, la OMS reportó en 2010 que 26 millones de personas tienen la vista reducida, y 3.2 millones son ciegas [54]. Lo más alarmante es, sin embargo, que un 80 % de todos estos casos podría haberse prevenido o tratado mediante alguna intervención médica, de haberse descubierto a tiempo [54]. En Europa se ha observado un comportamiento muy similar a este respecto [164].

Un gran número de estudios de poblaciones a gran escala y de meta-análisis han concluido que la retinopatía diabética (RD) y el glaucoma son dos de las tres causas más comunes de reducción de la capacidad visual y/o de ceguera a nivel mundial, en conjunto con la degeneración macular asociada a la edad [105, 194]. En Europa, por ejemplo, estas enfermedades explican aproximadamente el 30 % de los casos de ceguera [164], mientras que en América Latina la proporción de ciegos a causa de estas afecciones se ha incrementado significativamente en los últimos años [106]. En comparación con los promedios mundiales, Latinoamérica exhibe un mayor número de ciegos a causa de la degeneración macular, la retinopatía diabética y el glaucoma que como consecuencia de vicios de refracción, por ejemplo [106].

Las fotografías de fondo de ojo constituyen una modalidad de imagen médica utilizada extensivamente para la detección de la RD y el glaucoma. Esto se debe en parte a su naturaleza no invasiva y a la simplicidad y bajo costo del proceso de adquisición de las mismas [2]. Sin embargo, el análisis manual de este tipo de imágenes es prohibitivo en el contexto de campañas de *screening* poblacional a gran escala, en las cuales deben estudiarse numerosas imágenes en un período de tiempo relativamente acotado [3]. Por otra parte, los diagnósticos realizados a partir de ellas pueden diferir de un especialista a otro, dependiendo de su experiencia profesional, la calidad de las imágenes, el tiempo dedicado a la tarea o la fatiga que pueda generar el análisis de muchas imágenes sucesivamente. En un esfuerzo por mitigar estos inconvenientes, la comunidad científica internacional se encuentra explorando activamente y desde hace ya varios años la posibilidad de desarrollar métodos para el diagnóstico asistido por computadora de enfermedades visuales. En tal escenario, una computadora recibe como entrada una imagen, y aplica sobre ella una serie de algoritmos para brindar datos adicionales a los médicos. Esto incluye, por ejemplo, el etiquetado de regiones de interés, o una probabilidad que representa el nivel de

riesgo de que el paciente padezca tal o cual afección. Estos datos permiten mejorar la consistencia y precisión de los diagnósticos médicos, y reducen significativamente la variabilidad entre expertos [3].

Como se detallará más adelante en el Capítulo 2, existen numerosas estrategias para el diagnóstico automatizado de la RD y el glaucoma. Sin embargo, estos modelos todavía padecen de numerosos inconvenientes que deben ser resueltos para mejorar su rendimiento en escenarios clínicos reales. A continuación se analizarán algunos de los problemas que las contribuciones presentadas en esta tesis tienen por objetivo resolver:

- En general, los sistemas automatizados para el diagnóstico asistido por computadora a partir de fotografías de fondo de ojo requieren un primer paso que consiste en la segmentación de las vasculatura retiniana [63]. Dado que los vasos sanguíneos son la estructura anatómica más evidente en las imágenes de fondo de ojo, éstos son comúnmente utilizados como referencia para detectar y/o segmentar otras estructuras tales como el disco óptico (que es relevante para el estudio del glaucoma) o patologías tales como las lesiones rojas (que están asociadas al desarrollo de la RD). Además, la distribución morfológica de la vasculatura retiniana provee información valiosa que es estudiada por los profesionales para determinar si los pacientes sufren o no de algunas enfermedades oftalmológicas. Los métodos actuales para la segmentación automática de vasos sanguíneos suelen fallar al momento de lidiar con los vasos más delgados, en general por su imposibilidad para incorporar conocimiento *a priori* acerca de la morfología de las estructuras deseadas. Como consecuencia, estos sistemas automáticos suelen verse negativamente afectados por estas fallas. Contar con mejores algoritmos para la detección precisa de los vasos sanguíneos es esencial para mejorar la efectividad de este tipo de sistemas [63].
- La gran mayoría de los métodos actuales para la detección de glaucoma están basados en la cuantificación de algunas propiedades morfológicas del disco óptico, que es la región anatómica de la retina más afectada por esta enfermedad [75]. El principal inconveniente de este tipo de estrategias, sin embargo, es que requieren delinear de manera precisa el disco óptico y sus componentes internos. Además, estos enfoques ignoran otro tipo de información valiosa y

presente en las imágenes, tal como la distribución general de los vasos sanguíneos. Por el contrario, los métodos basados en aprendizaje profundo (o *deep learning*) son capaces de aprender por sí mismos las características relevantes asociadas con la presencia y/o ausencia de la enfermedad, sin requerir un proceso inicial de ingeniería manual de las mismas [107]. No obstante, esto es posible lograrlo si se dispone de una cantidad suficiente de datos anotados durante la etapa de entrenamiento. En los problemas de análisis de imágenes médicas esto es principalmente dificultoso, teniendo en cuenta que el etiquetado manual de imágenes es tedioso, consume mucho tiempo y debe ser realizado por expertos y profesionales médicos. Así, se vuelve imperioso el desarrollo de estrategias novedosas para el estudio del glaucoma, que sean independientes del proceso de segmentación manual del disco óptico, y que a su vez tengan en cuenta información de otras regiones anatómicas y/o patológicas observables en las imágenes.

- La gran mayoría de los sistemas existentes para la detección automatizada de RD requieren la detección inicial de lesiones rojas. Esto se debe a que este tipo de estructuras patológicas representan el primer signo de existencia de la enfermedad, pero son difíciles de identificar por parte de los observadores expertos sin realizar un análisis intenso y prolongado de las imágenes. Los métodos actuales para la detección automática de lesiones rojas se basan en descriptores diseñados manualmente, que por lo general se ven afectados por su capacidad limitada para generalizar las características propias de las lesiones. Como se mencionó anteriormente, las estrategias basadas en aprendizaje profundo o *deep learning* constituyen una alternativa factible para evitar el diseño manual de descriptores, aunque su aplicabilidad está limitada por la cantidad de datos de entrenamiento disponibles. En particular, el costo de reunir datos para aprender clasificadores que detecten lesiones rojas es extremadamente elevado, dado que requiere etiquetar manualmente estas estructuras. Así, es importante desarrollar nuevas estrategias para la detección de lesiones rojas utilizando métodos basados en *deep learning*, pero considerando el hecho de que éstas deben ser capaces de aprender cómo identificar las estructuras a partir de una cantidad limitada de lesiones anotadas.

## 1.2. Aprendizaje automático

Como se mencionó previamente, esta tesis aplica técnicas de aprendizaje automático para el diagnóstico asistido por computadora de enfermedades visuales. La principal motivación en la utilización de este tipo de enfoques radica en el hecho de que el aprendizaje automático provee un *framework* natural para el desarrollo de esta tarea. La mayoría de los métodos existentes aplica variantes de estos algoritmos en alguna de sus etapas, incluyendo el aprendizaje o extracción de características, la clasificación de imágenes o lesiones, la segmentación, etc [2]. Aunque en las próximas secciones y capítulos se abordarán estas técnicas con un mayor nivel de detalle, esta sección introduce en líneas generales los algoritmos de aprendizaje automático y describe las ideas detrás de su funcionamiento.

Bajo el nombre de aprendizaje automático (o *machine learning*) se agrupan una gran variedad de algoritmos que son capaces de aprender cómo realizar una determinada tarea sin necesidad de ser programados explícitamente para resolverla [77]. Aunque por lo general se lo considera un subcampo de las ciencias de la computación, el aprendizaje automático está sustentado por una variedad de áreas de conocimiento complementarias, incluyendo la estadística, la teoría de la computación y la optimización matemática. La idea general detrás de este tipo de estrategias es la de entrenar un modelo a partir de un conjunto  $S$  dado, formado por muestras (etiquetadas o no), denotadas como vectores  $\mathbf{x}$ . Las mismas son extraídas del dominio particular del problema a resolver, y son utilizadas como ejemplos a seguir por el algoritmo al momento de aprender la tarea específica. El entrenamiento consiste en el aprendizaje de un conjunto de parámetros  $\theta$  a partir de la optimización de una función objetivo definida sobre el conjunto de entrenamiento  $S$ . El modelo obtenido es posteriormente aplicado en tiempo de test para clasificar nuevas muestras.

En esta tesis se utilizarán modelos supervisados tanto para la clasificación de píxeles como de imágenes [77]. El área del aprendizaje automático supervisado está basada en modelos que se aprenden a partir de conjuntos de entrenamiento  $S$  compuestos por pares de la forma  $(\mathbf{x}^{(i)}, y^{(i)})$ ,  $i = \{1, 2, \dots, |S|\}$ , donde  $i$  representa el número de muestra de entrenamiento. Cada par está formado por un vector de descriptores  $\mathbf{x}^{(i)} \in \mathcal{X}$  y su correspondiente etiqueta o *ground truth*  $y^{(i)} \in \mathcal{Y}$ . Los descriptores o *features* (también denominados características o indicadores) representan

a una muestra en función de una serie de medidas cuantitativas. Dependiendo del dominio de aplicación, las etiquetas indican si un píxel pertenece a una región de interés, por ejemplo, o si una imagen corresponde a un paciente sano o que sufre una determinada enfermedad. Finalmente,  $\mathcal{X}$  y  $\mathcal{Y}$  son los espacios de todos los posibles descriptores y etiquetas, respectivamente, que pueden estar asociados a una muestra.

La función de clasificación  $f_\theta : \mathcal{X} \rightarrow \mathcal{Y}$  es aprendida a partir de  $S = \{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(n)}, y^{(n)})\}$  mediante la minimización de lo que se denomina *riesgo empírico*:

$$\mathcal{R}_S = \frac{1}{n} \sum_{i=1}^n \mathcal{L}(y^{(i)}, f_\theta(\mathbf{x}^{(i)})) \quad (1.1)$$

donde  $\theta$  es un conjunto de parámetros de la función  $f$ , y  $\mathcal{L}$  es una función de pérdida que compara una salida estimada  $\hat{y}^{(i)} = f_\theta(\mathbf{x}^{(i)})$  con la etiqueta esperada  $y^{(i)}$ , y que es diseñada de acuerdo al problema a resolver.

El *framework* de aprendizaje automático supervisado puede aplicarse para la clasificación o segmentación de imágenes adaptando la función de pérdida  $\mathcal{L}$  y el conjunto de indicadores  $\mathcal{X}$ , y redefiniendo las etiquetas  $\mathcal{Y}$  a estimar [77]. Esto facilita su traslado y aplicación a tareas tales como la segmentación de vasos sanguíneos o la detección de RD o glaucoma. En los próximos capítulos se brindarán mayores detalles acerca de cómo fueron aplicados estos algoritmos en las contribuciones realizadas.

### 1.3. Contribuciones y organización de la tesis

Esta tesis comienza con una descripción extensiva acerca del conocimiento actual en materia de detección soportada por computadora de la RD y el glaucoma utilizando imágenes de fondo de ojo. Inicialmente se proveen algunos detalles acerca de la anatomía y fisiología del ojo humano, haciendo especial énfasis en la retina y sus componentes anatómicas. Posteriormente se describen estas enfermedades, indicando sus signos más tempranos y cómo son identificados a partir de diferentes modalidades de imagen médica. También se incluye una clasificación general de los principales enfoques aplicados a la detección automática de ambas enfermedades, discutiendo sus ventajas y desventajas.

La primera contribución de esta tesis se describe en el Capítulo 3. La misma

consiste en un método para la segmentación de vasos sanguíneos de la retina basado en el aprendizaje de un campo condicional aleatorio totalmente conectado, utilizando para ello una máquina de vectores de soporte de salida estructurada. En la actualidad existen numerosos enfoques de segmentación basados en la clasificación supervisada de píxeles, aunque los mismos no son capaces de incorporar al proceso de aprendizaje conocimiento *a priori* acerca de la forma de las estructuras deseadas. Para resolver esta dificultad se propone el uso de campos condicionales aleatorios con potenciales de a pares (o *pairwise*) de orden alto, que son capaces de modelar mejor las interacciones entre los píxeles que forman parte de los vasos sanguíneos. Para ello se tienen en cuenta avances recientes en materia de inferencia en campos aleatorios condicionales totalmente conectados [101], integrándolos en un *framework* de aprendizaje de salidas estructuradas que es utilizado ampliamente en el área de visión computacional [150]. El enfoque propuesto es validado utilizando bancos de imágenes de referencia y métricas cuantitativas globales de calidad de la segmentación, obteniendo resultados mejores que los de otros algoritmos existentes [153]. Este capítulo se basa mayoritariamente en las siguientes publicaciones científicas:

- J. I. Orlando y M. del Fresno. Reviewing preprocessing and feature extraction techniques for retinal blood vessel segmentation in fundus images. *Mecánica Computacional*, XXXIII(42):2729–2743, 2014.

Este artículo presenta un análisis extensivo de diferentes algoritmos existentes para el preprocesamiento y la extracción de características, aplicados a la segmentación de vasos sanguíneos en imágenes de fondo de ojo. Las estrategias de caracterización relevadas son evaluadas en términos de su contribución en la mejora de los resultados de una máquina de vectores de soporte.

- J. I. Orlando y M. B. Blaschko. Learning fully-connected CRFs for blood vessel segmentation in retinal images. En P. Golland, C. Barillot, J. Hornegger, y R. Howe, eds., *MICCAI 2014, LNCS*, tomo 8149, págs. 634–641. Springer, 2014.

Este artículo introduce por primera vez la estrategia de segmentación basada en campos condicionales aleatorios totalmente conectados. El algoritmo es evaluado sobre el banco de imágenes DRIVE, y se brinda una comparación con respecto a algunos enfoques del estado del arte, demostrando que la estrategia

presentada alcanza un rendimiento competitivo con respecto al de un segundo observador humano.

- J. I. Orlando, E. Prokofyeva, y M. B. Blaschko. 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, 2017.

Este artículo extiende la publicación anterior con experimentos adicionales sobre otros conjuntos de imágenes tales como STARE, CHASEDB1 y HRF. También incorpora un mecanismo de escalado de parámetros que permite adaptar los descriptores utilizados durante el entrenamiento para su aplicación en imágenes de mayor resolución.

En el Capítulo 4 se introduce una estrategia novedosa para la identificación de glaucoma, basada en la transferencia de redes neuronales convolucionales (conocidas también por su sigla en inglés, CNNs, de *Convolutional Neural Networks*). La falta de disponibilidad de grandes conjuntos de imágenes anotadas de pacientes con y sin glaucoma hace imposible entrenar de cero una CNN para resolver esta tarea. Como alternativa, se propone transferir CNNs que fueron preentrenadas a partir de datos no médicos hacia esta nueva aplicación en particular. La hipótesis de trabajo es que una CNN entrenada con bases de datos de gran tamaño será capaz de extraer características lo suficientemente generales como para que las mismas sean de utilidad en otros dominios. Inicialmente, las imágenes de fondo de ojo se adaptan utilizando técnicas de preprocesamiento tales como *inpainting* [119] de vasos sanguíneos, realce de contraste y acercamientos al área del disco óptico. Posteriormente, se extraen características utilizando una CNN preentrenada, y las mismas se utilizan luego para entrenar un clasificador de regresión logística. Esta estrategia es capaz de alcanzar un rendimiento competitivo con respecto al de otras estrategias evaluadas en imágenes similares. El capítulo está íntegramente basado en la publicación:

- J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Convolutional neural network transfer for automated glaucoma identification. En *12th International Symposium on Medical Information Processing and Analysis*, págs. 101600U–101600U. International Society for Optics and Photonics, 2017.

La tercera contribución de este trabajo es una estrategia para la detección de lesiones rojas para *screening* de RD, descrita en detalle en el Capítulo 5. La misma consiste en extraer de manera inicial un gran número de lesiones candidatas, utilizando para ello un método no supervisado, basado en operaciones morfológicas. Estas estructuras son posteriormente clasificadas en lesiones verdaderas o falsas por medio de un vector híbrido de descriptores, compuesto por características diseñadas manualmente y otras aprendidas utilizando *deep learning*. En particular, se entrena una CNN con una arquitectura de profundidad limitada utilizando parches en torno a lesiones candidatas y sus respectivas anotaciones a nivel de lesión. Los descriptores aprendidos por la red son posteriormente aumentados utilizando características diseñadas manualmente. Esta estrategia conjunta es utilizada para entrenar un Bosque Aleatorio, que provee una probabilidad por cada candidato que indica si el mismo corresponde a una lesión roja auténtica o no. Los resultados obtenidos en una serie de bancos de imágenes estándar muestran que esta estrategia permite mejorar el rendimiento original de las CNNs, superando otros enfoques del estado del arte al evaluárselo tanto a nivel lesión como a nivel imagen. El capítulo se basa en el siguiente artículo:

- J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Learning to detect red lesions in fundus photographs: An ensemble approach based on deep learning. *arXiv preprint arXiv:1706.03008*, 2017.

Finalmente, la tesis concluye en el Capítulo 6, que resume los principales resultados y logros alcanzados, y provee algunas futuras líneas de investigación que pueden derivarse de estas contribuciones originales.

---

# Capítulo 2

## Antecedentes

Este capítulo resume en líneas generales algunos aspectos esenciales del análisis automático de imágenes de fondo de ojo para el diagnóstico de enfermedades de la retina. La Sección 2.1 describe la anatomía del ojo, con especial atención en las características fisiológicas y anatómicas de la retina. En la Sección 2.2 se analizan las dos enfermedades en las cuales se enfoca esta tesis, la retinopatía diabética y el glaucoma. La Sección 2.3 detalla las principales modalidades de imágenes médicas utilizadas para diagnosticar estas enfermedades en la práctica clínica cotidiana. Finalmente, la Sección 2.4 presenta una revisión general del estado del arte en materia de diagnóstico asistido por computadora de afecciones retinianas.

### 2.1. El ojo y la retina

La vista constituye uno de los sentidos más importantes de los seres humanos, y es lograda mediante un sistema biológico complejo que involucra principalmente la interacción de dos órganos: los ojos y el cerebro. Los ojos son responsables de enfocar la luz del ambiente, por un lado, y convertirla en una serie de impulsos eléctricos que son transmitidos hacia el cerebro. Este último, por otro lado, es el encargado de interpretar los impulsos recibidos y convertirlos en imágenes [177].

El ojo (Figura 2.1) es una estructura esférica que se divide principalmente en dos regiones, el segmento anterior y el posterior.

El segmento anterior es visible desde el exterior, y ocupa alrededor de un sexto

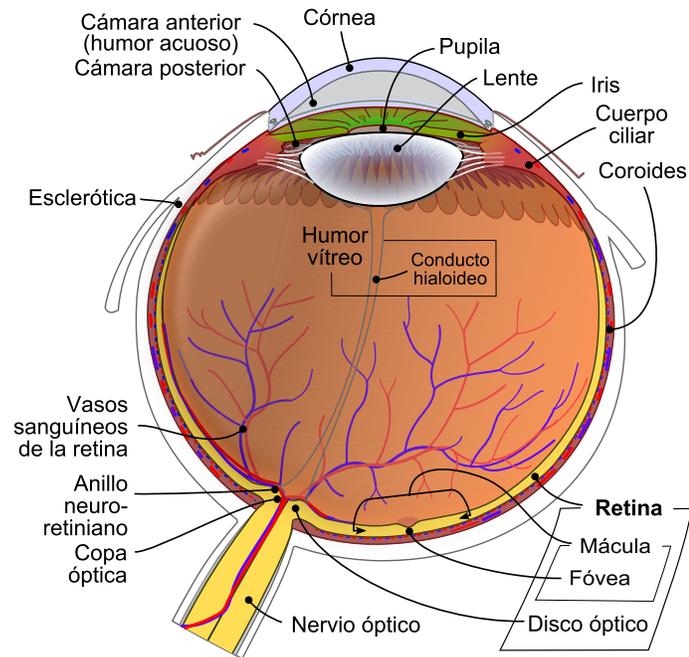


Figura 2.1: Representación esquemática del ojo. Fuente: [42].

del ojo. Está compuesto por la córnea, la pupila, el iris y el lente [95]. La córnea es un tejido transparente localizado en la región exterior del segmento anterior. La pupila, por su parte, es una abertura que se observa como un círculo blanco rodeado por una región coloreada, el iris, el cual está constituido por un tejido muscular blando y liso, que puede contraer y dilatar la pupila. Esta parte controla automáticamente cuánta luz entra en el ojo, dejando pasar poca cuando el ambiente está lo suficientemente iluminado, o más cuando está oscuro [95]. El lente es un disco convexo y transparente que ayuda a enfocar la luz. Entre el iris y el lente y entre la córnea y el iris, existe un fluido transparente conocido como humor acuoso, cuya función principal es mantener la presión interna del ojo para asegurar que el globo ocular conserve su forma esférica. Además, provee de nutrientes a los tejidos no alimentados por vasos sanguíneos, tales como la córnea, la malla trabecular, el lente y la hialoides. El humor acuoso es segregado por el epitelio ciliar (la estructura que soporta al lente), y fluye hacia afuera del ojo a través de un sistema de drenaje cuya componente principal es la malla trabecular [95].

La cámara posterior involucra los cinco sextos restantes del ojo, y está rodeada por una capa de grasa protectora conectada a un conjunto de músculos extrínsecos, que permiten mover el globo ocular [95]. La cámara posterior está compuesta por la hialoides y una pared formada por tres capas diferentes de tejido: una fibrosa, una vascular y una última capa interna. La hialoides es una sustancia gelatinosa que rellena la cavidad posterior y es, al igual que el humor acuoso, responsable de asegurar la forma esférica del ojo. La túnica fibrosa rodea los segmentos anterior y posterior, y está constituida principalmente por tejido conectivo. Se compone de dos partes, la córnea y la esclerótica, que es un recubrimiento de material blanco. La capa vascular corresponde a la coroides, formada por una red de capilares delgados que brinda sangre y oxígeno a las demás capas [95]. Finalmente, la capa más interna es la retina.

La retina es un tejido delgado, parcialmente transparente, compuesto por millones de células fotorreceptoras y neuronas [177]. Existen dos tipos de fotorreceptores: los conos y los bastones. Los conos son capaces de distinguir detalles finos y colores, y están adaptadas para ser expuestas a un mayor nivel de luz que los bastones. Los bastones, por el contrario, únicamente registran grises, y son más relevantes para la visión periférica y la vista en condiciones de baja iluminación. Ambos tipos de células están conectadas con las neuronas ganglionares a través de una capa de células bipolares. Las neuronas ganglionares transmiten las señales eléctricas a través del nervio óptico, para que finalmente sean procesadas por el cerebro [203].

Anatómicamente, la retina está compuesta principalmente por tres estructuras: la mácula, el árbol vascular y el disco óptico [2]. La mácula es una región amarillenta que se especializa en la visión más fina. Tiene aproximadamente 5 mm de diámetro, y está localizada en la parte posterior de la retina. Histológicamente, la mácula se caracteriza por una total ausencia de bastones y por poseer una abundante cantidad de conos. En el centro de la mácula hay una pequeña depresión conocida como fovea, que es más oscura de aspecto y tiene una extensión aproximada de 1.5 mm de diámetro [177]. El disco óptico (también conocido como cabeza del nervio óptico, u *optic nerve head*, ONH) es un área circular localizada en el centro de la retina, en donde el nervio óptico ingresa en ella. Este región tiene un tamaño aproximado de  $1,5 \times 2,5$  mm en personas adultas, y constituye un punto ciego en donde no hay ni

conos ni bastones. Dentro del ONH hay una pequeña excavación fisiológica conocida como la copa óptica (*optic cup*, OC), que se observa como un área pálida dentro del anillo neuroretiniano [177]. Alrededor del 65% de la sangre requerida por la retina es provista a través de la membrana coroides. Sin embargo, el restante 35% es brindado por los vasos sanguíneos de la retina. Estas arterias y venas sumamente delgadas lucen a simple vista como estructuras elongadas que se distribuyen a partir del centro del disco óptico hacia las demás regiones de la retina.

La visión se logra mediante la interacción de todas estas regiones anatómicas [95]. Primero, la luz reflejada por los objetos entra en el ojo a través de la córnea, que enfoca la luz para permitir una visión más clara y detallada. Dependiendo de la iluminación del ambiente, la pupila será dilatada o contraída por el iris, que actúa a modo de esfínter. Luego de atravesar la pupila, la luz penetra a través del lente, que se contrae o expande para enfocar mejor los objetos lejanos o cercanos, respectivamente. Por último, los fotones son proyectados en la retina, donde los conos y bastones convierten las ondas electromagnéticas de la luz en las señales eléctricas que son transmitidas vía las células ganglionares hacia el nervio óptico, para ser procesadas posteriormente por el cerebro [203].

## 2.2. Enfermedades de la retina

Diversas enfermedades oftalmológicas y sistémicas se manifiestan a través de cambios en la retina, algunas de las cuales pueden afectar la vista si no son tratadas a tiempo [2]. Esta tesis está enfocada en brindar herramientas para la detección temprana de dos afecciones específicas, la retinopatía diabética y el glaucoma.

### 2.2.1. Retinopatía diabética

La retinopatía diabética (RD) es una de las consecuencias más comunes de la diabetes. La diabetes es una enfermedad sistémica en la que la glucosa no puede ser debidamente metabolizada a causa de inconvenientes en la producción o utilización de la insulina [131]. Existen dos tipos diferentes de diabetes, la Tipo 1 y la Tipo 2. La diabetes Tipo 1 ocurre cuando el sistema inmunológico ataca y mata por error a un tipo particular de células del páncreas. Como resultado, se liberan pequeñas o

nulas cantidades de insulina en el cuerpo, insuficientes para procesar correctamente la glucosa. Esta condición explica entre un 5 y un 10 % de los casos de diabetes, y se desarrolla por lo general durante la niñez o la adolescencia. La Tipo 2, por el contrario, es causada por un síndrome conocido como insensitividad a la insulina, en la que el cuerpo no puede utilizar correctamente la insulina o no produce la cantidad suficiente. La mayoría de los pacientes con diabetes sufre de este tipo de enfermedad en particular. La principal consecuencia de cualquiera de los tipos de diabetes, como se mencionó anteriormente, es un metabolismo inapropiado de la glucosa. Esto conduce eventualmente a la hiperglucemia, caracterizada por una presencia excesiva de glucosa en el plasma sanguíneo. La hiperglucemia puede provocar una gran variedad de complicaciones vasculares de mantenerse sin tratamiento durante largos períodos de tiempo [2].

Algunos estudios recientes han demostrado que el número de personas que sufre diabetes crecerá de 366 millones (2.8%) en 2011 a 552 millones (4.4 %) hacia 2030 [215]. Además, se prevé que el mayor incremento ocurra en países de ingresos bajos y medios [215]. En Argentina, la prevalencia de diabetes ha crecido de 8.4 % a 9.6 % desde 2005 a 2009 [175], y es una de las principales causas de muerte en el país [6]. Aunque los factores que provocan la diabetes no han podido ser determinados aún, se ha observado que la obesidad, una vida sedentaria o antecedentes genéticos son factores de riesgo de esta enfermedad [131].

La RD es una consecuencia del daño microvascular producido por la diabetes [2], y es una de las principales causas de ceguera a nivel mundial [164]. Alrededor del 5 % de las personas con diabetes Tipo-2 tienen RD, y se espera que el número de pacientes afectados por esta dolencia se vea significativamente incrementado en los próximos años [3]. En Argentina, estudios recientes identificaron a la RD como la segunda causa de ceguera prevenible, totalizando un 16 % de los casos relevados [23].

La Figura 2.2 describe las complicaciones microvasculares asociadas con la RD. La hiperglucemia daña los pericitos, que son células especializadas ubicadas en la superficie de los capilares sanguíneos y que controlan el flujo microvascular [195]. Producto de este trastorno, las paredes de los vasos se debilitan y se vuelven más susceptibles a fugas, mientras el flujo sanguíneo se incrementa (Figura 2.2.1). Este último factor eleva la tensión de cizalladura o *shear stress* de las paredes de los va-

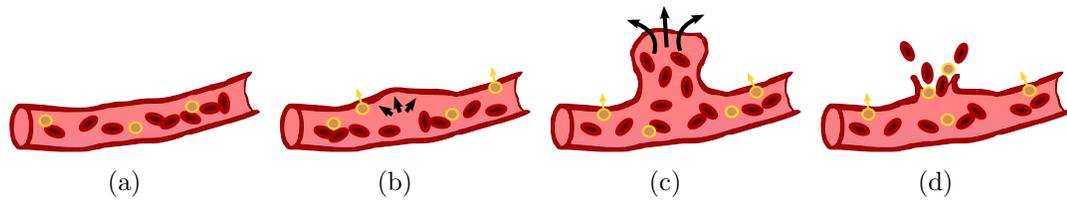


Figura 2.2: Complicaciones microvasculares causadas por la hiperglucemia. (a) Capilar sano. (b) La hiperglucemia debilita las paredes de los vasos e incrementa el flujo sanguíneo, permitiendo además la fuga de algunos lípidos de la sangre. (c) Las paredes debilitadas de los vasos y el aumento de la presión en las mismas favorece la formación de microaneurismas. (d) La presión elevada en el interior de la microaneurisma provoca eventualmente su ruptura, produciendo hemorragias.

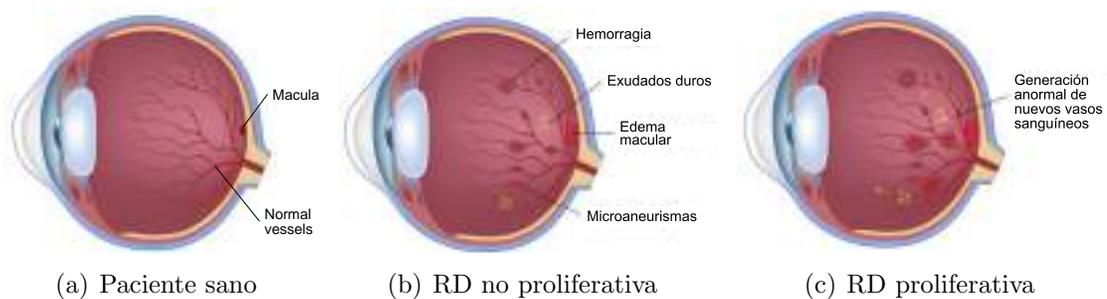


Figura 2.3: Etapas principales de la retinopatía diabética. Fuente: [19].

Los [120], favoreciendo la generación de microaneurismas (MAs) [131] (Figura 2.2.1). Las MAs son deformaciones de las paredes vasculares, en forma de globo, que pueden romperse y producir hemorragias (HEs) en la retina (Figura 2.2.1). El término más comúnmente utilizado para referirse tanto a las MAs como a las primeras HEs pequeñas es el de «lesiones rojas», dado que las mismas pueden visualizarse como pequeños puntos rojos en las fotografías de fondo de ojo [143, 180]. Por otro lado, la permeabilidad de los vasos permitirá eventualmente que algunas proteínas y lípidos presentes en la sangre se fuguen de los vasos e inunden algunas de las capas de la retina. Estas acumulaciones son conocidas como exudados duros, y constituyen, en conjunto con las lesiones rojas, los signos más tempranos de la RD no proliferativa [2].

Las complicaciones microvasculares descritas con anterioridad tienen consecuencias que, eventualmente, afectan el metabolismo retiniano (Figura 2.3). A medida

que el número de lesiones rojas aumenta, algunas regiones de la retina dejan de recibir la cantidad necesaria de sangre, reduciéndose sensiblemente las cantidades de oxígeno y glucosa a las que éstas acceden. Las células requieren de estas sustancias para mantenerse con vida, y el acceso limitado a ellas provoca la muerte del tejido del cual forman parte [229]. Este fenómeno es conocido como isquemia, y el cuerpo reacciona a ella liberando un factor angiogénico que estimula la generación de nuevos vasos para reemplazar los dañados, intentando llevar hacia esas regiones la sangre necesaria para mantener al tejido vivo. Dicho proceso de neovascularización conduce al estado más avanzado de la RD, conocido como RD proliferativa (RDP) [214]. En esta etapa, los nuevos vasos desarrollados son muy débiles y producen nuevas fugas de sangre o lípidos, aumentando el riesgo de perder la visión irreversiblemente (del 25.6 % al 36.9 %) por desprendimiento de retina o por una hemorragia vítrea [2].

Los tratamientos médicos para lidiar con esta enfermedad dependen siempre del estadio de la misma. Por lo general, a los pacientes diabéticos sin RD evidente se les administran drogas para controlar la cantidad de glucosa en sangre [131]. Además, se les recomienda asistir al menos una vez al año al oftalmólogo para estudiar sus ojos y corroborar la inexistencia de lesiones rojas. En estados más avanzados, sin embargo, el tratamiento es más limitado [40]. Si se observan casos severos de neovascularización, por ejemplo, se puede recurrir a la fotocoagulación láser o a la vitrectomía para prevenir la pérdida de la visión [40]. No obstante, ambos tratamientos tienen asociado un riesgo adicional de pérdida de la vista, y no pueden revertir la reducción en la agudeza visual que el paciente ya haya experimentado, producto de una detección tardía de la enfermedad [40]. En otros casos, los pacientes reciben inyecciones intravítreas para prevenir la proliferación de neovasos [40].

La detección temprana de la RD es esencial para prevenir su progreso y reducir el riesgo de pérdida de visión. Sin embargo, la RD es asintomática en sus primeros estadios, con lo cual la única forma de detectarla es estudiando si los signos clínicos de la enfermedad están o no presentes [2]. Como se mencionó anteriormente, las lesiones rojas son el principal indicador de desarrollo de la enfermedad. Por lo general son detectados por los especialistas por medio de una oftalmoscopia, o utilizando alguna modalidad de imagen tal como la fotografía de fondo de ojo (Sección 2.3) [2, 158]. El diagnóstico de la RD basado en este tipo de imágenes se aborda en más detalle

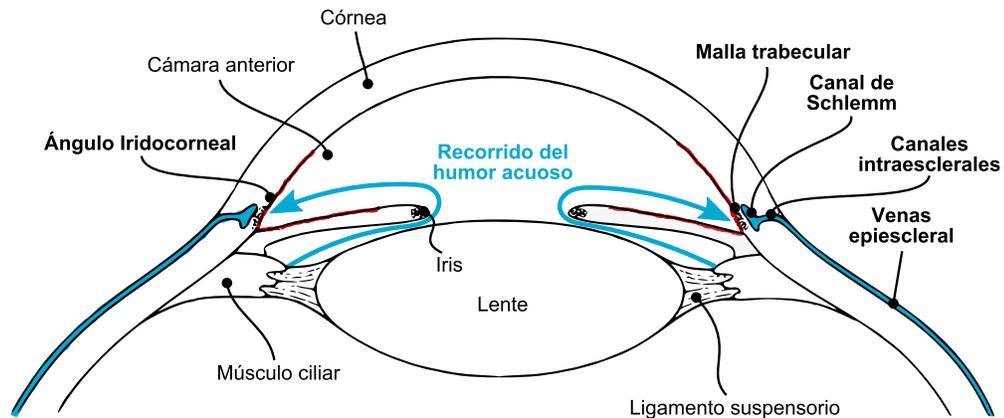


Figura 2.4: Flujo del humor acuoso. En celeste se observa el recorrido del flujo, y en rojo se indica el ángulo iridocorneal.

en la Sección 2.4.

### 2.2.2. Glaucoma

El glaucoma es una enfermedad neurodegenerativa, crónica e irreversible, que constituye una de las principales causas de ceguera prevenible a nivel mundial [194]. En 2013, el número estimado de personas de entre 40 y 80 años afectadas por glaucoma en todo el mundo fue de 64.3 millones, y se espera que trepe a 76 millones hacia 2020, y a 111.8 millones en 2040 [194]. En los países europeos, el glaucoma explica un 20.5% de los casos de ceguera prevenible [164]. En Latinoamérica, por su parte, la prevalencia de ceguera como consecuencia del glaucoma ha continuado en aumento durante la última década [106].

Aunque bajo el término glaucoma se agrupa a una familia de diferentes enfermedades, todas están emparentadas por una misma característica, que es el daño del nervio óptico provocado producto de un incremento desmedido de la presión intraocular. Esto es causado por una acumulación de humor acuoso debido a algún tipo de defecto en el sistema de drenaje del ojo (Figura 2.4). Como se mencionó al momento de describir la anatomía del ojo, el humor acuoso es segregado por el cuerpo ciliar y fluye a través de la pupila hacia la cámara anterior, desde la cual es drenado a través de la malla trabecular. Este tejido esponjoso se ubica en la córnea, cerca del ángulo iridocorneal (que es el ángulo que existe entre la córnea y el iris), y es

responsable de drenar la mayor parte del humor acuoso [41]: el líquido es recolectado por la malla trabecular y drenado vía el canal de Schlemm, a partir del cual fluye por los canales intraesclerales hacia la vena episcleral, que lo incorpora al torrente sanguíneo [41]. En pacientes sanos, la tasa de secreción de humor acuoso es prácticamente equivalente a la de drenaje, con lo que la presión intraocular se mantiene relativamente constante. Sin embargo, si la tasa de secreción es más alta que la de drenaje, el humor acuoso comienza a acumularse en la cámara anterior. Esta región se llenará eventualmente con este líquido, y la cámara posterior también, elevando progresivamente la presión intraocular y comprimiendo la hialoides hacia la retina. Esto puede dañar la capa de fibras nerviosas de la retina, los vasos o el nervio óptico, conduciendo a una pérdida progresiva de la visión que puede resultar, de no tratarse a tiempo, irreversible.

Los factores de riesgo del glaucoma incluyen la edad, la predisposición genética, una elevada presión intraocular, la obesidad, el consumo de corticoides y traumas ocurridos en el ojo [110]. Además, a los pacientes con antecedentes familiares de glaucoma se les recomienda asistir una vez al año al oftalmólogo para controlar la presión interna del ojo y, en caso de detectarse una presión elevada, verificar si el nervio óptico ha sufrido algún tipo de daño [164].

En general, el glaucoma puede clasificarse en dos tipos principales: el glaucoma de ángulo abierto y el de ángulo cerrado. El glaucoma primario de ángulo abierto (*primary open-angle glaucoma*, POAG) es el tipo más común de glaucoma, y ocurre cuando el humor acuoso no puede drenarse lo suficiente debido a algún tipo de degeneración u obstrucción en la malla trabecular [132]. Esto produce un incremento crónico e indoloro de la presión intraocular, que es generalmente asintomático pero puede eventualmente dañar el nervio óptico y producir ceguera. El glaucoma de ángulo cerrado es menos común, y es provocado por un bloqueo repentino del ángulo iridocorneal. Como consecuencia de este trastorno, el humor acuoso no puede abandonar las cámaras anteriores y posteriores. Esta situación es repentina, y conduce a un crecimiento de la presión muy doloroso y rápido, que puede provocar pérdida permanente de la visión en intervalos de tiempo mucho más cortos [213].

El POAG es conocido como el "ladrón silencioso de la vista", y es el tipo de glaucoma más difícil de detectar [164]. Esto se debe a que los pacientes que padecen

esta enfermedad no notan sus efectos hasta que la visión se ve reducida. En general, las primeras manifestaciones de pérdida de la vista como consecuencia del POAG se relacionan con una visión periférica nublada. A medida que la enfermedad progresa, sin embargo, el campo visual se reduce en mayor medida, y los pacientes acaban teniendo una visión cada vez más tubular, hasta perderla por completo [213]. En lo que resta de esta tesis se utilizará el término glaucoma para referirse únicamente al POAG, dado que los casos analizados corresponden a esta patología en particular.

Al menos la mitad de los pacientes con glaucoma continúan sin ser diagnosticados, mientras que la mitad de aquellos que están en tratamiento en realidad son falsos positivos y no padecen glaucoma [164]. Si se tiene en cuenta que el glaucoma es una enfermedad crónica que se sufre de por vida, el desafío principal es poder detectar el mayor número posible de pacientes enfermos [164]. La práctica más usual en las campañas de *screening* poblacional incluye el examen del nervio óptico a través de modalidades de imagen médica, el estudio de potenciales defectos en la capa de fibras nerviosas retinales (*retinal nerve fiber layer*, RNFL) y la medición de la proporción simétrica de copa/disco (o *cup to disc ratio*, CDR) [192]. A pesar de que los defectos en la RNFL son considerados el signo más temprano del glaucoma, el CDR continúa siendo el criterio diagnóstico más utilizado para detectar glaucoma a partir de imágenes de fondo de ojo [75]. En general, este valor se calcula como la relación entre el área o la longitud vertical u horizontal del disco óptico (*optic disc*, OD) y la copa óptica (*optic cup*, OC). Cuando la presión en el ojo es demasiado elevada, las fibras del nervio óptico se dañan y comienzan a desaparecer. Como resultado, el tamaño de la OC aumenta en proporción al OD, provocando un incremento del valor del CDR [75]. La cuantificación manual de este parámetro es comúnmente utilizada como biomarcador de glaucoma, aunque factores tales como el sexo, la edad o la raza influyen significativamente en su valor, y pueden confundirse con casos patológicos [80]. A los pacientes que exhiben un CDR alto se los deriva por lo general a un especialista en glaucoma para realizarse estudios de mayor complejidad que complementen al original mediante el estudio de la capa de fibras nerviosas o la distribución de los vasos sanguíneos, utilizando modalidades de imagen tales como la tomografía de coherencia óptica o la tomografía retinal de Heidelberg [216].

Actualmente no hay una cura para el glaucoma, aunque existen algunos tra-

tamientos para intentar reducir sus efectos o prevenir su progreso. Estos incluyen medicación, terapia láser y/o cirugías. La medicación por lo general intenta reducir la presión intraocular, y se administra directamente en el ojo en forma de gotas. Los tratamientos láser suelen utilizarse para tratar casos graves de POAG, y consisten en disparos de láser aplicados sobre la malla trabecular para estimular el flujo del humor acuoso. Las cirugías son otra alternativa, también, y se aplican con un propósito similar sobre la malla trabecular, abriendo nuevos conductos para facilitar el drenaje [213].

## 2.3. Imagenología de la retina

Las características anatómicas de la retina permiten a los médicos examinarla de manera no invasiva utilizando diferentes tecnologías de imagen [158]. Quizá la manera más sencilla de realizar esta tarea es mediante una oftalmoscopia, que consiste en la observación del interior del ojo por medio de un oftalmoscopio, un dispositivo de bajo costo que consiste en una lente y una fuente de luz [222].

La imagenología de la retina constituye una alternativa robusta a la oftalmoscopia, ya que permite preservar el examen en una imagen digital. Las tres modalidades más empleadas son la fotografía de fondo de ojo, la angiografía por fluorescencia y la tomografía de coherencia óptica (OCT) [2]. Esta tesis está enfocada en la utilización de fotografías de fondo de ojo. Sin embargo, en las siguientes subsecciones se brindarán detalles sobre las otras dos modalidades mencionadas anteriormente.

### 2.3.1. Fotografías de fondo de ojo

Las fotografías de fondo de ojo (Sección 2.5), también conocidas como retinografías, son imágenes proyectivas y a color de la superficie interna del ojo [157]. Esta modalidad de imagen constituye la técnica más económica para el diagnóstico de enfermedades de la retina y sistémicas, incluyendo la RD y el glaucoma [2]. El proceso de captura de estas imágenes es no invasivo y relativamente sencillo de realizar [131]. Las imágenes son capturadas por medio de una cámara de fondo de ojo (Figura 2.6), que consiste principalmente de dos partes: un microscopio especializado y de baja potencia con una fuente de luz, que ilumina la parte interior del ojo, y una cámara

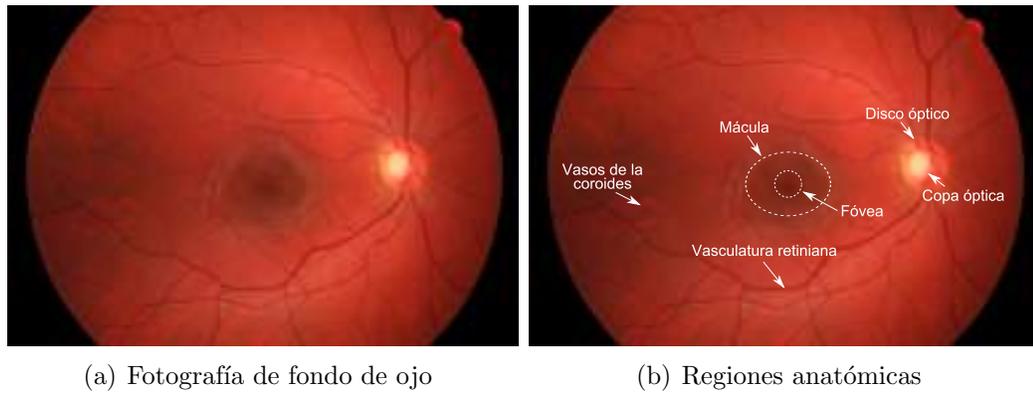


Figura 2.5: Fotografía de fondo de ojo de un paciente sano, extraída del conjunto de imágenes HRF [145].

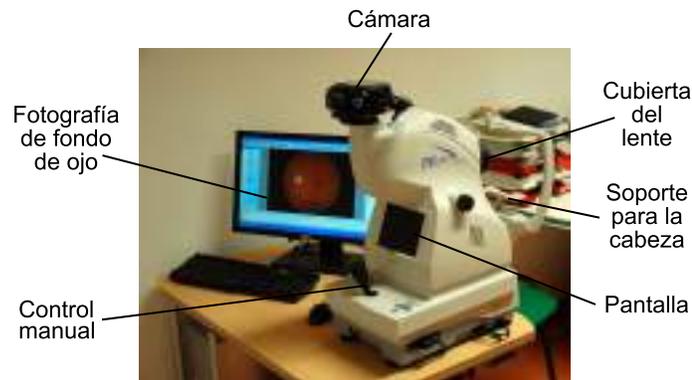


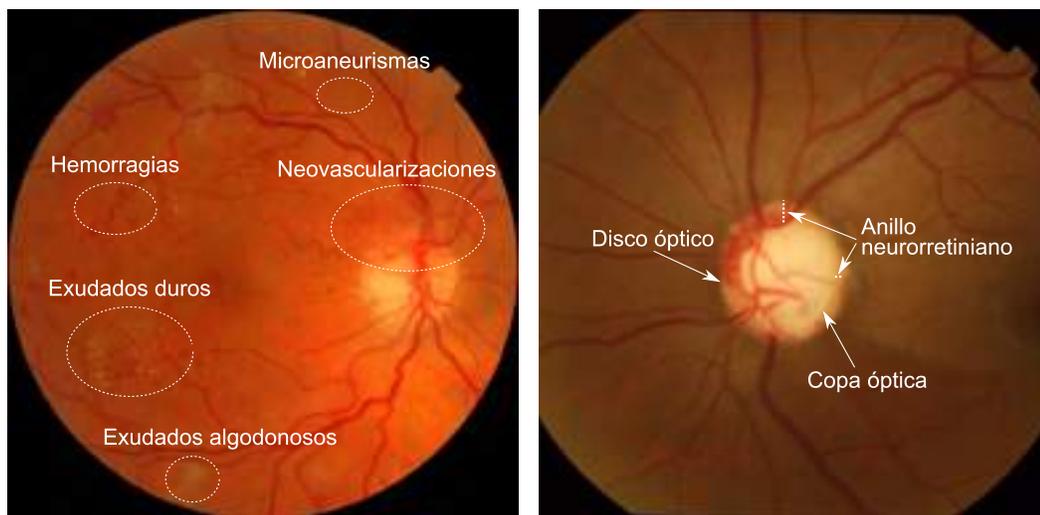
Figura 2.6: Cámara de fondo de ojo. Fuente: Ralf Roletschek / Wikipedia [171].

digital convencional que captura la fotografía y la almacena como un archivo digital. En los últimos años, sin embargo, se han desarrollado dispositivos portátiles capaces de acoplarse a teléfonos inteligentes (Figura 2.7), que han permitido reducir el costo original de adquirir una cámara de fondo de ojo y han mejorado su portabilidad [47].

El protocolo de captura utilizando una cámara de fondo de ojo convencional comienza con el paciente colocando su mentón en un soporte que forma parte del dispositivo. A continuación, el operador alinea la cámara con el iris y la pupila utilizando un *joystick*. Posteriormente, la cámara enfoca e ilumina la cámara posterior, y se brinda una primera visión global de la retina en una pantalla. El paciente fija a continuación su vista en un objetivo (por lo general, una luz verde), y el operador ajusta el foco para obtener una mejor visión global de la retina. Finalmente se ad-



Figura 2.7: Volk iNview, un dispositivo para la captura de fotografías de fondo de ojo basado en teléfonos inteligentes. Fuente: [148].



(a) Retinopatía diabética. Fuente: banco de imágenes MESSIDOR [49]. (b) Glaucoma. Fuente: banco de imágenes DRISHTI-GS1 [187].

Figura 2.8: Fotografías de fondo de ojo para el diagnóstico de enfermedades visuales.

quiere una serie de fotografías de distintas partes de la retina simplemente solicitando al paciente que observe diferentes objetivos y disparando la cámara [2].

La Figura 2.5(b) ilustra los diferentes componentes anatómicos de la retina, según se observan en las fotografías de fondo de ojo. La vasculatura retiniana, el disco óptico, la mácula y la fovea pueden visualizarse de manera sencilla utilizando esta modalidad de imagen. Las Figuras 2.8(a) y 2.8(b) presentan fotografías de fondo de

ojo de pacientes con RD y glaucoma, respectivamente. En el caso correspondiente al paciente con RD se pueden observar algunas de las estructuras patológicas más típicas de esta enfermedad, tales como la presencia de lesiones rojas, los exudados duros, las lesiones algodonosas y las neovascularizaciones. Por otro lado, en la imagen del paciente glaucomatoso se puede observar un crecimiento en el tamaño de la copa óptica en relación con el diámetro del disco óptico. Comparadas con la oftalmoscopia, las imágenes de fondo de ojo permiten una visión más amplia de la retina [158]. Además, es posible combinar diferentes fotografías de distintas regiones utilizando algoritmos automáticos de registración para cubrir un área mayor respecto al campo de visión (*field of view*, FOV) original de la cámara [2].

La principal desventaja de las fotografías de fondo de ojo es que son representaciones proyectivas 2D y no brindan información de profundidad, lo que hace difícil validar la existencia de lesiones en las capas interiores de la retina o la medición del daño real en el nervio óptico. Sin embargo, su bajo costo y rapidez en la captura las hacen extremadamente útiles para el análisis de grandes poblaciones [2, 3].

### 2.3.2. Angiografía por fluorescencia

Las angiografías por fluorescencia (*fluorescein angiographies*, FA), también conocidas como angiografías fluorescentes, permiten el estudio de los vasos sanguíneos de la retina no sólo desde un punto de vista anatómico sino también funcional [2]. La fluorescencia es una propiedad molecular que permite a ciertos compuestos tales como la fluoresceína sódica emitir luz a una longitud de onda larga cuando se los estimula con una longitud de onda corta [170]. Por medio de un agente de contraste fluorescente y una cámara angiográfica especializada, es posible observar tanto la coroides como los vasos sanguíneos de la retina en diferentes intervalos de tiempo, brindando fotografías que describen la circulación sanguínea en estas capas [170].

El protocolo de captura de las imágenes es similar al utilizado para obtener fotografías de fondo de ojo. Inicialmente se le administran al paciente gotas oculares midriáticas para dilatar la pupila. Posteriormente, se le inyecta intravenosamente la fluoresceína sódica, y la cámara angiográfica toma una serie de fotografías que registran el movimiento del agente contraste a lo largo del sistema circulatorio [170]. Estas imágenes se obtienen capturando la luz emitida por el compuesto al ilumi-

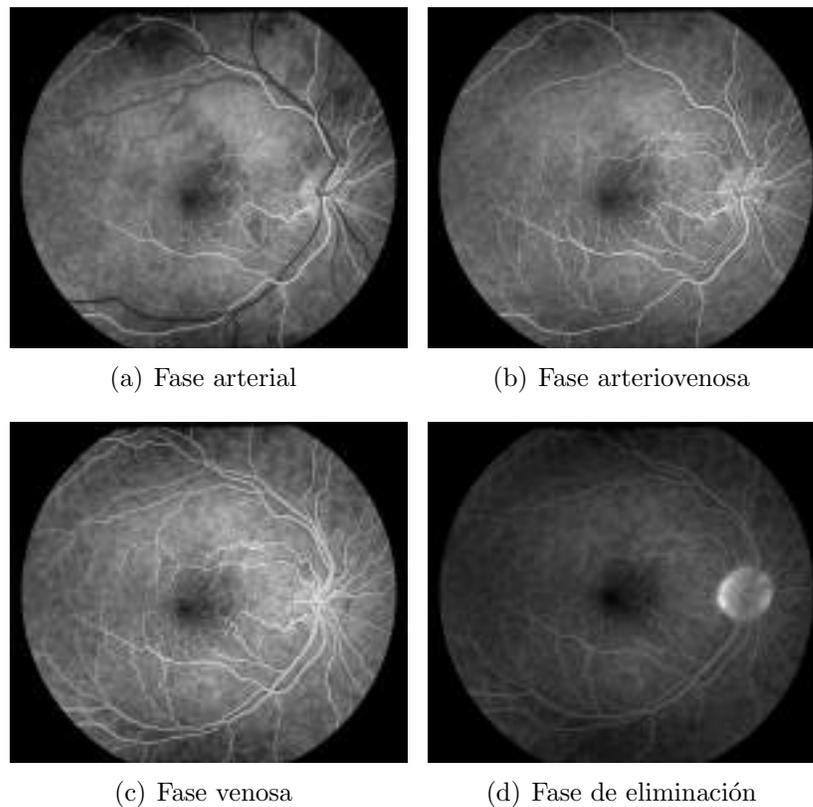


Figura 2.9: Angiografías por fluorescencia. Fases. Fuente: [167].

narse utilizando una luz azul. En general se fotografían cuatro fases, denominadas arterial, arteriovenosa, venosa y de eliminación (Figura 2.9) [170]. Durante la fase arterial, las venas aparecen más oscuras que las arterias. En la arteriovenosa, tanto las arterias como las venas se observan como estructuras brillantes. En la venosa, las arterias pierden algo de brillo pero las venas continúan destacándose. Finalmente, la fase de eliminación, que ocurre entre 5 y 10 minutos después de haber inyectado el contraste, se caracteriza por un contraste menor de los vasos con respecto a las demás estructuras de la retina.

Las FAs son principalmente utilizadas para estudiar pacientes con RD o para estudiar potenciales edemas maculares o degeneraciones maculares asociadas a la edad [170]. Las MAs y los exudados son más fáciles de observar en las FAs que en las fotografías de fondo de ojo, dado que el agente contraste también se acumula o fuga en estas estructuras. Esta propiedad hace a las FAs útiles para auxiliar tratamientos

láser en los que las fuentes de fuga de exudados o sangre deben ser correctamente identificadas de manera inicial. Sin embargo, la inyección del agente contraste hace a este procedimiento invasivo: los pacientes por lo general experimentan náuseas o vómitos como consecuencia del procedimiento, y no pueden realizárselo aquellos con antecedentes de anafilaxis, alergia a la fluoresceína sódica o dificultades en los riñones [170]. Esto hace preferible la utilización de otras modalidades tales como las OCTs o las fotografías de fondo de ojo [2].

### 2.3.3. Tomografía de coherencia óptica

La tomografía de coherencia óptica (OCT) es una técnica no invasiva y sin contacto que provee imágenes volumétricas de los tejidos internos de la retina [191]. Las imágenes se obtienen a partir de un principio conocido como interferometría de baja coherencia, que está basado en una fuente de luz casi infrarroja que penetra en los tejidos retinianos y permite recuperar cortes transversales con una resolución hasta 10 veces mayor que la de los dispositivos basados en ultrasonido [2]. Mediante la combinación de cada uno de estos cortes 2D es posible reconstruir un volumen 3D de la retina, que provee no solo una vista general de su superficie sino también de sus capas internas, incluyendo las fibras nerviosas, el epitelio pigmentario y la coroides [2]. Esto facilita el estudio de distintas anomalías en la capa exterior de la retina, o de patologías más profundas tales como la neovascularización corooidal, el desprendimiento de retina o el daño del nervio óptico producto del glaucoma [191]. También permiten estudiar el espesor de la retina, que es útil para la planificación de cirugías o intervenciones farmacológicas.

La Figura 2.10 presenta una imagen OCT de la región macular de la retina. El procedimiento de captura de la imagen es relativamente simple, aunque toma más tiempo que una fotografía de fondo de ojo. Primero se le administran al paciente gotas midriáticas para dilatar la pupila, y posteriormente ubica el mentón en un pedestal del dispositivo. Posteriormente, el tomógrafo escanea la retina de 10 a 14 minutos utilizando un rayo de luz casi infrarrojo. El dispositivo analiza la demora de tiempo y el cambio de magnitud de este disparo lumínico respecto a cómo es retrodispersado por los tejidos que forman la retina, y transforma estos valores en intensidades de gris [59]. Varios artefactos tales como un medio ocular turbio o movimientos involuntarios

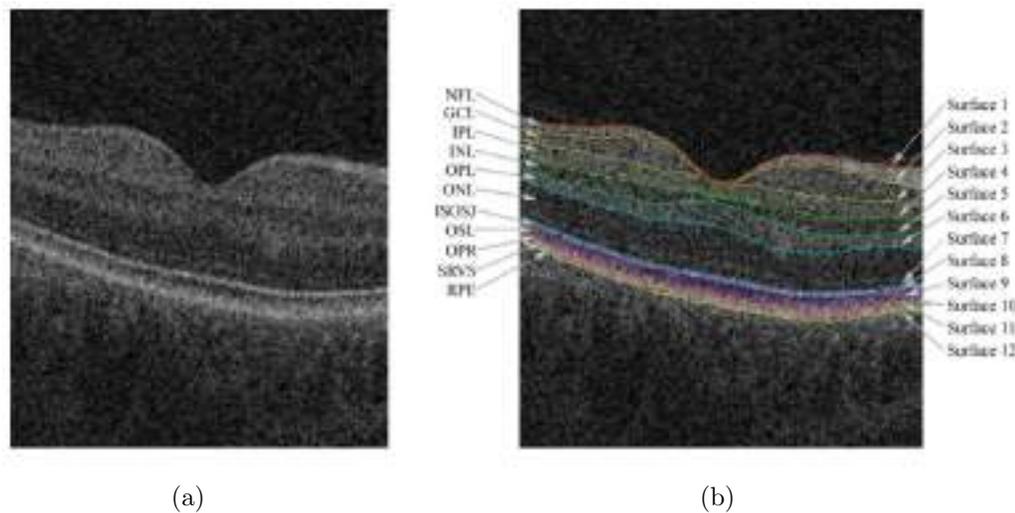


Figura 2.10: Imagen OCT y capas interiores de la retina, obtenidas utilizando un algoritmo de segmentación automático [111]. (a) Imagen X-Z del volumen de la OCT. (b) Resultados de la segmentación: capa de fibras nerviosas (NFL), capa de células ganglionares (GCL), capa plexiforme interna (IPL), capa nuclear interna (INL), capa plexiforme externa (OPL), capa nuclear externa (ONL), unión del segmento interior y anterior (ISOSJ), fotorreceptores del segmento exterior (OPR), espacio virtual subretiniano (SRVS-de profundidad 0 en pacientes sanos), y epitelio pigmentario retiniano (RPE). Figura extraída de [191].

del paciente pueden influir en el análisis manual de estas imágenes por parte de expertos [189]. Además, el ruido *speckle* típico de esta modalidad de imagen dificulta significativamente el procesamiento automático de estas imágenes también [140].

## 2.4. Análisis de imágenes de fondo de ojo para diagnóstico asistido por computadora

En esta sección se describirán en líneas generales los métodos más utilizados en la literatura para el diagnóstico asistido por computadora de RD y glaucoma, basado en fotografías de fondo de ojo. La detección de estas enfermedades en estadios en los que no existen síntomas y solamente se observan unas pocas manifestaciones es clave, teniendo en cuenta que el acceso temprano a terapias médicas es la única forma de prevenir su progreso hacia la ceguera [127].

### 2.4.1. Análisis general del estado del arte

Las fotografías de fondo de ojo constituyen la modalidad de imagen más económica para el *screening* de RD y glaucoma por diversas razones. La más importante es que las mismas pueden obtenerse mediante un proceso de captura relativamente simple, que puede ser realizado por técnicos y no exclusivamente por especialistas en oftalmología. Esto permite minimizar la intervención de los expertos al momento de la captura y favorece la realización de campañas basadas en sistemas de telemedicina, en los que los médicos no necesitan desplazarse hacia el lugar en el que la campaña tiene lugar [113]. Por el contrario, son los técnicos quienes adquieren las imágenes *in situ*, transmitiéndolas posteriormente a través de una red segura hacia instalaciones remotas en las que los expertos realizan el correspondiente diagnóstico. Tales sistemas son conocidos por su efectividad en la reducción de costos, y son utilizados desde hace no muchos años para estudiar poblaciones de riesgo en busca de glaucoma y RD [3, 136]. Más aún, las cámaras de fondo de ojo para teléfonos inteligentes han permitido mejorar la movilidad y usabilidad de los dispositivos de captura, facilitando el acceso a estas campañas y mejorando su efectividad significativamente.

El cuello de botella en cualquier campaña de *screening*, sin embargo, está en el análisis de las imágenes [3]. Si la población a estudiar es significativamente grande, se adquirirá un gran número de imágenes que requerirán de muchos oftalmólogos para su estudio y diagnóstico. Esta tarea es particularmente tediosa y consume mucho tiempo. Además, puede verse influida por el nivel de experticia de los médicos involucrados o por su propia fatiga, incrementando la variabilidad inter e intra observadores [3].

Los métodos para detectar estas enfermedades con asistencia de la computadora son conocidos por reducir estos inconvenientes y mitigar el costo en expertos de las campañas de *screening* [127]. Por medio de sistemas capaces de detectar lesiones o cambios patológicos en las imágenes, se puede generar un orden de los casos desde los más críticos hacia los menos atendibles, permitiendo a los médicos decidir cuáles deben analizarse primero [127]. Además, la información brindada por este tipo de sistemas ayuda a los expertos a estudiar cambios sutiles que son difíciles de detectar a simple vista, tales como variaciones en la distribución vascular o la existencia de lesiones rojas [127].

Tabla 2.1: Clasificación de las herramientas para el diagnóstico y la detección asistida por computadora de RD y glaucoma.

	Propiedades vasculares	Detección de estructuras patológicas	Caracterización por descriptores globales de imagen	Métodos de aprendizaje profundo
<b>RD</b>	Calibre [44, 137] Dimensión fractal [38, 72]	Lesiones rojas (MAs y HEs) [168, 176, 179], lesiones brillantes (exudates) [176, 179], neovascularizaciones [109, 174]	AM-FM multiescala [7], HOS [4], Texturas (GLCMs y RLMs) [5]	CNNs [163], CNNs con visualización de anomalías [69, 165, 212]
<b>Glaucoma</b>	Calibre [11, 13, 89, 129], tortuosidad [218], ángulo de bifurcación [218].	CDR [75, 90], Defectos en la RNFL [133, 146], PPA [116]	Intensidades [26], texturas [26, 130], Coeficientes de la FFT [26, 27], histogramas [26], Coeficientes de la B-spline [27], Wavelets [50].	CNNs [34] (limitadas por la poca disponibilidad de imágenes).

En general, los métodos automáticos para el estudio de enfermedades oftalmológicas pueden agruparse en cuatro categorías, como se indica en la Tabla 2.1. Los enfoques basados en caracterización vascular correlacionan cambios en la distribución general de los vasos sanguíneos de la retina con la existencia de alguna enfermedad. Los métodos basados en la detección de estructuras patológicas, por el contrario, cuantifican la existencia de lesiones asociadas con estas dolencias. Otras estrategias menos tradicionales se basan en el cálculo de descriptores globales a partir de imágenes enfermas y sanas, y utilizan sus valores para discriminar los casos enfermos. Finalmente, los métodos de aprendizaje profundo o *deep learning* utilizan redes neuronales convolucionales (CNNs) para aprender de forma automática las características relevantes de estas enfermedades. La clasificación aquí propuesta no es estricta, teniendo en cuenta que existen diversos algoritmos que ocupan más de una categoría en particular. Un ejemplo es el método para la detección de glaucoma propuesto por Chakravarty *et al.* [32], el cual combina indicadores extraídos globalmente de las imágenes y descriptores obtenidos a partir de la segmentación del nervio óptico para entrenar un clasificador supervisado. No obstante, esta clasificación agrupa a la gran mayoría de las estrategias hoy por hoy disponibles en la literatura científica. Las ideas centrales detrás de cada tipo de enfoque se abordarán con un mayor nivel de detalle en las próximas secciones.

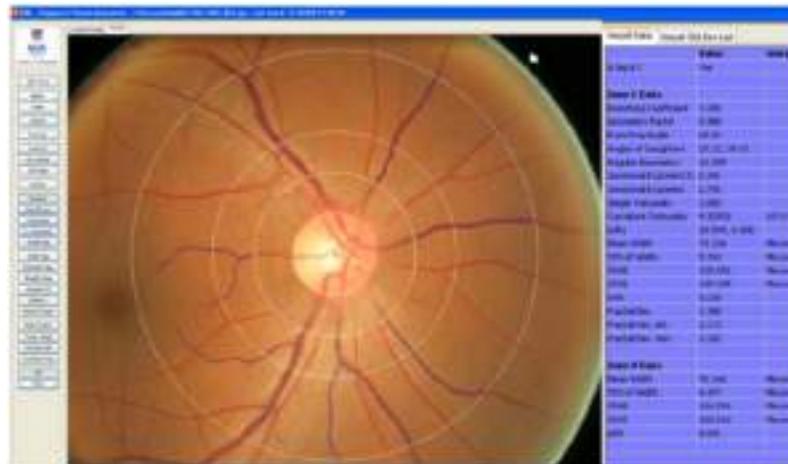


Figura 2.11: Captura de pantalla de la herramienta SIVA (Singapore I Vessel Assessment). Fuente: [1].

### 2.4.2. Enfoques basados en caracterización vascular

En las últimas décadas se han realizado diversos estudios tratando de correlacionar cambios en la distribución y la morfología de los vasos sanguíneos de la retina con la existencia de RD o glaucoma. En general, estos estudios están basados en complejos métodos de segmentación y caracterización vascular, o en el uso de herramientas comerciales tales como SIVA (Singapore I Vessel Assessment, [37], Figura 2.11) o VAMPIRE (Vessel Assessment and Measurement Platform for Images of the REtina, [159]). Estos sistemas primero delimitan los vasos sanguíneos de la retina utilizando un enfoque automático, y luego extraen una serie de descriptores que caracterizan su morfología y/o distribución con respecto a otras estructuras anatómicas tales como la mácula o el disco óptico. En algunos casos, estos indicadores son sólo obtenidos en una pequeña región localizada a una cierta distancia del disco óptico. En otros, se obtienen a partir de toda la estructura vascular segmentada. Por otro lado, algunos enfoques están basados en la caracterización de arterias y venas por separado. En estos casos, se requiere de la clasificación de los vasos en una u otra clase con posterioridad a la segmentación de los mismos.

La vasculatura retiniana puede caracterizarse en términos de diversas propiedades, incluyendo el diámetro de arterias y venas, la cantidad y ángulos de las bifurcaciones, la tortuosidad y área de los diferentes segmentos, etc [9, 122]. La arquitectura

y distribución general del árbol vascular también puede describirse en función de su dimensión fractal [37]. Los objetos fractales son estructuras que se caracterizan por su auto similaridad: es decir, existe un patrón que persiste al analizar el objeto en múltiples escalas [55]. Objetos ramificados tales como los vasos sanguíneos son ejemplos naturales de este tipo de objetos [83].

En general, la habilidad de los métodos basados en la caracterización de los vasos está limitada por la efectividad del enfoque de segmentación utilizado para delinear el árbol vascular [85]. La segmentación de estas estructuras es una tarea muy compleja y difícil de resolver computacionalmente, que continúa siendo un problema abierto en la literatura [63]. Por otra parte, los resultados obtenidos utilizando descriptores vasculares para detectar enfermedades oftalmológicas son significativamente peores a los obtenidos utilizando otras estrategias basadas en descriptores más robustos. Sin embargo, la investigación en este tipo de indicadores tiene un impacto formidable en el entendimiento de las enfermedades y en el desarrollo de nuevas terapias y tratamientos [63].

#### **2.4.2.1. Retinopatía diabética**

El calibre de los vasos y su dimensión fractal han sido identificados como potenciales indicadores de estadios tempranos de la RD, o como biomarcadores capaces de predecir su progresión hacia la RD proliferativa. En [137], se observó un calibre más grande en las vénulas de la retina de pacientes diabéticos con retinopatía en comparación con el de pacientes sin retinopatía evidente. El estudio presentado en [38] reveló que valores altos en la dimensión fractal del árbol vascular están asociados de manera independiente con signos tempranos de RD en pacientes con diabetes tipo I. En [72] se midieron descriptores fractales a partir de imágenes de la retina de pacientes diabéticos para estimar su asociación con complicaciones micro y macrovasculares tales como la RD, la neuropatía y la nefropatía. Como resultado se observó que los pacientes con menor dimensión fractal son más tendientes a padecer RD proliferativa y neuropatía. En [44] se observó, además, que el calibre de los vasos constituye un importante factor de riesgo a tener en cuenta para predecir la progresión hacia RD proliferativa. En el mismo estudio se identificaron resultados similares para los descriptores fractales.

### 2.4.2.2. Glaucoma

Las propiedades de los vasos sanguíneos también han sido analizadas como potenciales caracterizadores de glaucoma, aunque con resultados más contradictorios que en el caso de la RD. En general, únicamente se descubrió la asociación de tres patrones vasculares con el glaucoma: el calibre, el ángulo de bifurcación y la tortuosidad de los vasos. Se ha reportado, por ejemplo, que aquellos pacientes que padecen glaucoma suelen presentar calibres arteriales y venulares significativamente inferiores a los de pacientes sanos [11, 13, 89, 129]. Sin embargo, los datos prospectivos del estudio de Rotterdam no han descubierto asociación alguna entre el calibre vascular y el glaucoma [84]. Otro estudio realizado sobre angiografías por fluorescencia había arribado oportunamente a conclusiones similares [17]. Finalmente, en [218] se observó una asociación entre el glaucoma y la disminución de la tortuosidad arterial y venular y un ángulo de bifurcación venular más angosto.

### 2.4.3. Detección de estructuras patológicas

La gran mayoría de los métodos de diagnóstico asistido por computadora de RD y glaucoma están basados en detectar estructuras patológicas asociadas con los estadios tempranos de estas enfermedades [3, 51]. En la práctica clínica cotidiana, los oftalmólogos buscan algún signo de presencia de MAs o HEs para determinar si los pacientes tienen o no RD [180]. De manera similar, el daño en la RNFL o un incremento en el CDR son potenciales signos de glaucoma [75]. Contar con sistemas capaces de detectar la presencia de este tipo de lesiones es relevante para la asistencia al diagnóstico realizado por los oftalmólogos, ya que los médicos reciben, en este caso, no sólo una imagen de fondo de ojo a analizar sino además un reporte indicando dónde fueron detectadas las estructuras patológicas, con una cierta probabilidad de certeza [176]. Además, la cuantificación automática de la cantidad o severidad de las lesiones puede ayudar a determinar el grado de avance de la enfermedad y detectar los casos más urgentes.

Se han propuesto numerosos métodos basados en reconocimiento de patrones y aprendizaje automático para detectar la presencia de lesiones y cuantificar sus propiedades. La principal ventaja de este tipo de enfoques es la comprensibilidad de sus

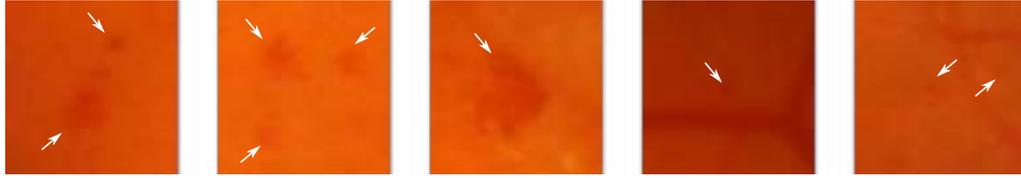


Figura 2.12: Ejemplos de lesiones rojas, como se observan en las imágenes de fondo de ojo. Nótese que algunas de ellas son apenas distinguibles. Fuente: conjunto de entrenamiento de DIARETDB1 [96].

resultados. Por otro lado, y teniendo en cuenta que estos algoritmos reproducen las prácticas clínicas usualmente seguidas por los especialistas, permiten a los oftalmólogos estudiar la validez de los resultados obtenidos y decidir si la salida del sistema es correcta o no [51].

La principal desventaja de estos métodos es que pueden ser seriamente afectados por la calidad de los algoritmos de segmentación subyacentes [24]. El reconocimiento automático de cada patología en particular requiere contar con un método dedicado lo suficientemente robusto como para lidiar con la gran variedad de formas, texturas e intensidades de las lesiones. Este inconveniente ha motivado un esfuerzo cada vez mayor en la literatura por desarrollar estrategias más precisas para la detección de lesiones, o la utilización de enfoques basados en descriptores globales que sean independientes de la segmentación.

Se analizarán a continuación diversos sistemas para el estudio de la RD y el glaucoma basados en la detección de lesiones. Cabe señalar, sin embargo, que la gran variedad de algoritmos de segmentación utilizados para tal fin no serán discutidos. Por otro lado, se brindarán mayores precisiones acerca del estado del arte en materia de detección de lesiones rojas en el Capítulo 5.

#### 2.4.3.1. Retinopatía diabética

Los sistemas para screening automatizado de RD basados en la detección de lesiones utilizan principalmente como guía el hallazgo de lesiones rojas (MAs y HEs pequeñas) y brillantes (exudados y lesiones algodinosas) [57]. El estudio de la RD proliferativa, por otro lado, se basa también en la detección automática de las áreas con neovascularizaciones [174].

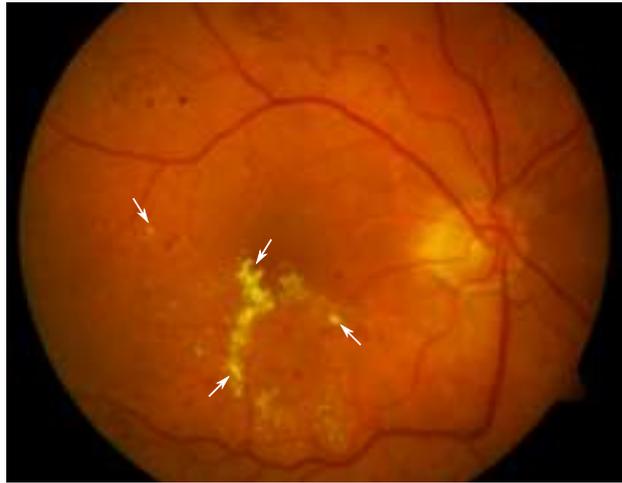


Figura 2.13: Ejemplos de lesiones brillantes, como se observan en las imágenes de fondo de ojo. Fuente: conjunto de entrenamiento de DIARETDB1 [96].

Las lesiones rojas son difíciles de detectar debido a su tamaño pequeño y a que presentan intensidades similares respecto a las de otras estructuras tales como los vasos sanguíneos o las cicatrices producidas por los tratamientos láser [180] (Figura 2.12). Si las imágenes tienen baja calidad o contraste, por otra parte, su hallazgo se vuelve complejo incluso a simple vista. Las lesiones brillantes (Figura 2.13), por el contrario, son más reconocibles, aunque su identificación automática requiere separar inicialmente el disco óptico, que presenta intensidades similares y puede provocar falsos positivos [57]. Por otra parte, la detección de neovascularizaciones requiere contar con enfoques de caracterización de vasos sanguíneos para determinar las regiones en las que ocurren vasos delgados y tortuosos y diferenciarlas de aquellas en las que las estructuras vasculares son normales y no patológicas [174].

Uno de los primeros enfoques para la detección de RD basada en lesiones es el método presentado en [68], que utiliza redes neuronales artificiales para detectar exudados y hemorragias, y emplea esta información posteriormente para estudiar la existencia o no de la enfermedad. Desde entonces se han desarrollado numerosos sistemas de *screening* basados en características similares, aunque siempre intentando mejorar las estrategias de detección y/o segmentación [168]. Otras contribuciones han tenido por objetivo mejorar la estrategia de cuantificación de las lesiones. El método presentado en [139] y evaluado de manera extensiva en [176], por ejemplo,

se basa en la detección de lesiones rojas y brillantes también, aunque las estructuras son cuantificadas en términos de su área y de las probabilidades brindadas por el clasificador. Esta información es utilizada posteriormente para entrenar un clasificador que determina un nivel de riesgo para cada imagen. Un enfoque similar pero basado únicamente en lesiones rojas fue presentado recientemente en [179]. Por otro lado, se han propuesto también algoritmos para la detección automática de RD proliferativa basados en diferentes métodos para la determinación de neovascularizaciones [109, 174].

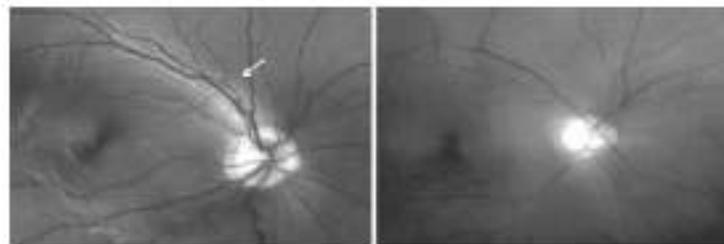
La abundante investigación en este campo ha motivado el traslado de algunos de estos sistemas a ambientes clínicos reales, con un éxito significativo en algunos casos [219]. Sin embargo, existe un consenso generalizado en cuanto a que estos métodos todavía requieren ser mejorados, tanto sea desarrollando mejores métodos para detectar las estructuras involucradas como para integrar información adicional de otros orígenes, tales como las historias clínicas de los pacientes [186]. En [208], por ejemplo, una herramienta denominada IRIS (*Intelligent Retinal Imaging System*) fue comparada con respecto al estudio manual de una cohorte de imágenes realizado previamente por un equipo de expertos. Aunque los resultados obtenidos son prometedores, los autores mencionan explícitamente la necesidad de refinar los algoritmos subyacentes para alcanzar mejores resultados [208].

#### 2.4.3.2. Glaucoma

La gran mayoría de los métodos para diagnóstico asistido por computadora de glaucoma se basan en la detección del crecimiento en el CDR, de defectos en la RNFL o de atrofia peripapilar (*peripapillary atrophy*, PPA) [75]. La RNFL (Figura 2.14) se observa en las imágenes de fondo de ojo como una sucesión de estrías rugosas y brillantes que se distribuyen de manera despareja en los ojos normales (Figura 2.14(a)). En pacientes glaucomatosos, sin embargo, pueden surgir defectos localizados en la RNFL como los ilustrados en la Figura 2.14(a) [88]. Además, la pérdida de células ganglionares generada por la presión intraocular típica del glaucoma puede producir defectos difusos de la RNFL (Figura 2.14(b)). Por otro lado, el CDR se obtiene de manera automática mediante la segmentación tanto del disco óptico como de la copa óptica. Debido a la ausencia de información de profundidad en las fotografías



(a) Defectos localizados en la RNFL. Izquierda: paciente sano. Derecha: paciente glaucomatoso.



(b) Defectos difusos en la RNFL. Izquierda: paciente sano. Derecha: paciente glaucomatoso.

Figura 2.14: Defectos en la capa de fibras nerviosas retiniana (*Retinal Nerve Fiber Layer*, RNFL) producto del glaucoma, según se manifiestan en las fotografías de fondo de ojo. Fuente: [75].

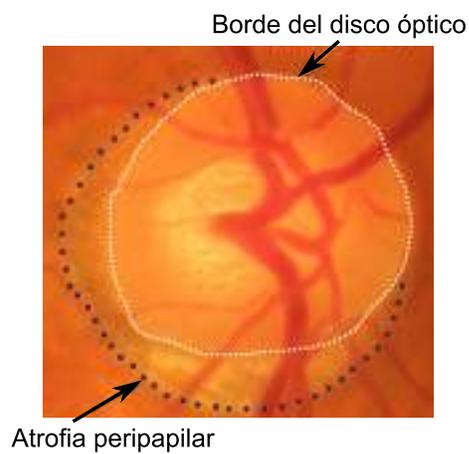


Figura 2.15: Atrofia peripapilar (*Peripapillary Atrophy*, PPA) producto del glaucoma. Source: [116].

de fondo de ojo, la segmentación de la copa óptica es un problema desafiante que aún continúa explorándose. Finalmente, las PPAs (Figura 2.15) se observan como regiones texturadas en torno al disco óptico.

La mayor parte de la literatura en materia de detección de glaucoma basada en lesiones se enfoca en el cálculo del CDR. Existen numerosos trabajos que proponen nuevos métodos para la segmentación del disco y la copa óptica para estudiar el CDR. Aunque cubrir cada una de estas estrategias no es el propósito de esta tesis, el lector interesado puede recurrir a la revisión de la literatura presentada en [75], que resume algunos de los más recientes avances en este campo. En general, la segmentación de la copa óptica requiere recurrir a características complementarias a las intensidades de gris, tales como la curvatura de los vasos, dado que es necesario inferir la profundidad en la excavación para hallar sus bordes. En [90], por ejemplo, se utilizan las curvaturas pronunciadas de los vasos como conocimiento *a priori*. En otros casos, sin embargo, los vasos sanguíneos son sustraídos de las imágenes utilizando técnicas de inpainting digital para que sus intensidades no interfieran en la segmentación [75].

En general, las demás lesiones asociadas a la existencia de glaucoma son mucho menos evidentes que el incremento en el CDR. Así, detectar defectos en la RNFL o PPA utilizando métodos automatizados es significativamente más difícil [75]. En la literatura del área existen, sin embargo, algunos esfuerzos por brindar solución a este inconveniente. El método propuesto en [133], por ejemplo, estima la localización de potenciales defectos en la RNLF mediante propiedades texturales, obtenidas de la imagen luego de remover de ella los vasos sanguíneos utilizando *inpainting*. De manera similar, otras características de textura son utilizadas en [146] para segmentar estos defectos en la RNFL utilizando un campo aleatorio de Markov (o *Markov Random Field*) [100]. Por otro lado, se ha propuesto en [116] un sistema para el estudio de la PPA asociada al glaucoma.

#### 2.4.4. Caracterización de las imágenes mediante descriptores globales

Otra familia de métodos se basa en la caracterización de las imágenes de forma global utilizando descriptores generales de las mismas. Estos indicadores corresponden a métricas cuantitativas que describen la distribución de las intensidades de

la imagen, por ejemplo. Posteriormente, y a partir de características extraídas de imágenes de pacientes sanos y enfermos, se las utiliza para entrenar un clasificador que será luego utilizado para determinar el riesgo de existencia de la enfermedad en una nueva imagen [75, 131]. Este enfoque tiene como ventaja que no requiere la segmentación de ninguna lesión o estructura anatómica. Por el contrario, esta familia de algoritmos asume que las propiedades de las imágenes serán influenciadas por los cambios asociados con las enfermedades, con lo que pueden permitirse ignorar el paso intermedio de la segmentación [75]. Sin embargo, debe aclararse que una importante limitación de esta estrategia es que recae en las utilización de caracterizadores diseñados *ad-hoc*, que requieren de un importante esfuerzo para calibrarse y para mantener su capacidad discriminativa al aplicarse sobre imágenes diferentes a las utilizadas en su diseño [131]. Además, los oftalmólogos no son capaces de validar los descriptores utilizados para generar el indicador de riesgo, sino el indicador de riesgo en sí mismo [75]. Más recientemente, los algoritmos basados en aprendizaje profundo (*deep learning*) (Sección 2.4.5) han demostrado ser capaces de estimar por sí mismos los descriptores a extraer de las imágenes (a partir de grandes volúmenes de datos), sorteando la difícil etapa de calibración manual de los mismos y alcanzando menores errores de generalización.

#### 2.4.4.1. Retinopatía diabética

Existen sólo algunos métodos para detección de RD utilizando descriptores globales, comparados con aquellos basados en lesiones. En [7] se propone una estrategia para la caracterización de pacientes con RD utilizando los métodos de modulación de amplitud y modulación de frecuencia (AM-FM). En lugar de utilizar las imágenes completas, los autores emplean un conjunto de ventanas trazadas en torno a lesiones que son típicas de la RD (tales como las MAs, las HEs, las neovascularizaciones o los exudados) y alrededor de regiones normales (vasos sanguíneos o el fondo de la retina). Mediante el ajuste de una serie de parámetros, se obtiene un vector de descriptores que describen la textura de estos parches. Luego, se reduce la dimensionalidad del vector a partir de un método de análisis de componentes principales (*Principal Component Analysis*, PCA), y se entrena un clasificador por regresión logística que es posteriormente utilizado para asignarle etiquetas a las imágenes. De

manera similar, en [4] se utilizan descriptores obtenidos utilizando espectros de alto orden (*higher order spectra*, HOS) para caracterizar las imágenes, y una máquina de vectores de soporte (SVM) que asigna a cada una de ellas un índice de riesgo de RD, basándose en estos indicadores. En [5] se presentó una estrategia mejorada, que utiliza otros caracterizadores de texturas tales como las matrices de coocurrencia de niveles de gris (*gray level co-occurrence matrices*, GLCMs) y las matrices de *run length* (*run length matrices*, RLMs).

#### 2.4.4.2. Glaucoma

De manera similar a lo que ocurre con la RD, existen pocos enfoques basados en la caracterización global de las imágenes para identificar la presencia de glaucoma. En [26] se utilizan descriptores de intensidades, texturas, coeficientes de la transformada rápida de Fourier (*Fast Fourier Transform*, FFT) y un modelo de histograma para describir imágenes de pacientes con glaucoma. En lugar de utilizar toda la imagen como fuente de información, los autores proponen utilizar la versión sin vasos de la imagen a color original, donde los mismos son removidos utilizando *inpainting*. Además, las imágenes son centradas en torno al disco óptico y reducidas en su resolución. El trabajo analiza diferentes combinaciones de descriptores y clasificadores, y se reportan los resultados obtenidos para la mejor configuración hallada. Basándose en este estudio inicial, se desarrolló un método para el cálculo automático de un índice de riesgo de glaucoma [27]. Este enfoque modifica el método anterior mediante la incorporación de un paso adicional de corrección de iluminación al momento del preprocesamiento. Por otro lado, los descriptores varían, en este caso reemplazando el modelo de histograma por los coeficientes de una *B-spline* [27]. Tras una fase de reducción de dimensionalidad por PCA, se ejecuta una etapa adicional de clasificación basada en dos partes: inicialmente, se obtiene un nivel de riesgo de existencia de glaucoma basado en cada tipo de descriptor y posteriormente se entrena un clasificador separado sobre estas salidas, que es quien asigna el riesgo final. Más recientemente, Dua *et al.* [50] propuso utilizar un conjunto de descriptores de energía basados en Wavelets para caracterizar la textura de las imágenes de pacientes glaucomatosos y diferenciarlos de sujetos sanos. En dicho trabajo se entrenan cuatro clasificadores distintos y se los evalúa utilizando estos descriptores, alcanzando una

alta precisión en un conjunto privado de imágenes. Otra alternativa observada es BRIEF (*Binary Robust Independent Elementary Features*), un enfoque basado en una serie de descriptores de textura propuestos en [130] para la identificación de glaucoma. La misma fue validada en combinación con una SVM, y comparada con otros enfoques de caracterización alternativos.

De acuerdo al relevamiento realizado, solamente un único sistema para la detección de glaucoma basada en características globales de las imágenes, denominado CatEye [125], ha sido validado de manera clínica. De acuerdo a [75], los enfoques que no están basados en segmentación pueden no ser lo suficientemente aptos para ser empleados en la práctica médica diaria, dado que los descriptores empleados no pueden ser validados de forma directa por parte de los oftalmólogos. Sin embargo, los autores plantean la posibilidad de integrarlos con otras estrategias basadas en la detección de lesiones para mejorar su capacidad de detección actual.

#### 2.4.5. Métodos basados en aprendizaje profundo (*deep learning*)

Bajo el término «aprendizaje profundo» o *deep learning* se agrupan una serie de métodos para regresión o clasificación basados en la utilización de redes neuronales profundas (o *deep neural networks*, DNNs) [25]. En particular, *deep learning* se refiere a la tarea de entrenar una red neuronal para resolver una tarea dada. La palabra «profundo» o *deep* se antepone para subrayar el hecho de que este tipo de redes tienen numerosas capas formadas por cientos de miles de unidades de procesamiento, que de alguna forma se asemejan a la forma en la que se organizan las neuronas en el cerebro. Así, dado un conjunto de datos de entrenamiento masivo y compuesto por abundantes cantidades de imágenes y sus respectivas anotaciones, las DNNs son capaces de aprender por sí mismas los descriptores necesarios para resolver una tarea específica. El proceso de aprendizaje en sí mismo consiste en la minimización de una función de costo que penaliza cualquier error cometido en el conjunto de entrenamiento.

Aunque hoy días los métodos basados en *deep learning* son los más utilizados y están siendo ampliamente investigados por la comunidad científica internacional, las redes neuronales han existido desde hace varias décadas [108]. Sin embargo, la ausen-

cia de conjuntos de datos anotados lo suficientemente grandes y de dispositivos de computación de alto desempeño volvía infactible el aprendizaje de redes demasiado profundas. En consecuencia, las redes se caracterizaron durante años por poseer una arquitectura poco profunda, que dificultaba la reducción de su error de generalización y, por ende, su capacidad de escalar hacia otros conjuntos de datos reales diferentes a los empleados para su entrenamiento o validación [107]. Sin embargo, los últimos avances en materia de optimización matemática, unidades de procesamiento gráfico (*Graphical Processing Units*, GPUs) y computación paralela, sumados a la existencia de herramientas para la anotación de imágenes a gran escala tales como Amazon Mechanical Turk [211] han renovado el interés por este tipo de métodos [107]. Desde la publicación del trabajo de Krizhevsky *et al.* [102], las redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs), un tipo particular de DNNs, han sido utilizadas y modificadas para la resolución de una gran variedad de problemas, incluyendo la clasificación de imágenes [181] y la segmentación semántica [182], por ejemplo.

La principal ventaja de las técnica de aprendizaje profundo es su habilidad para aprender automáticamente los mejores descriptores para realizar la tarea de clasificación. Esto permite evitar el proceso de diseñar y seleccionar descriptores, que resulta tedioso y por lo general involucra décadas de investigación específica. En el caso particular de la detección automatizada de enfermedades de la retina, las DNNs son lo suficientemente robustas para evitar el paso de segmentar las lesiones observadas, dado que aprenden cómo discriminar las enfermedades reconociendo por sí mismas a las lesiones como patrones locales [73]. Sin embargo, los descriptores que caracterizan estos procesos patológicos resultan difíciles de validar por parte de los expertos, aunque estudios recientes han introducido algunas estrategias para representarlos visualmente (Figura 2.16).

La alta precisión de las DNNs se alcanza al costo de entrenarlas a partir de enormes conjuntos de datos anotados. Así, su aplicabilidad queda usualmente restringida a dominios en los que hay gran disponibilidad de tales datos [86], algo limitado por la dificultad que supone normalmente la adquisición de grandes volúmenes de imágenes y por el costo de anotarlas, que requiere de usuarios expertos y de dispositivos costosos [22]. Estas características vuelven a las CNNs factibles de ser utilizadas

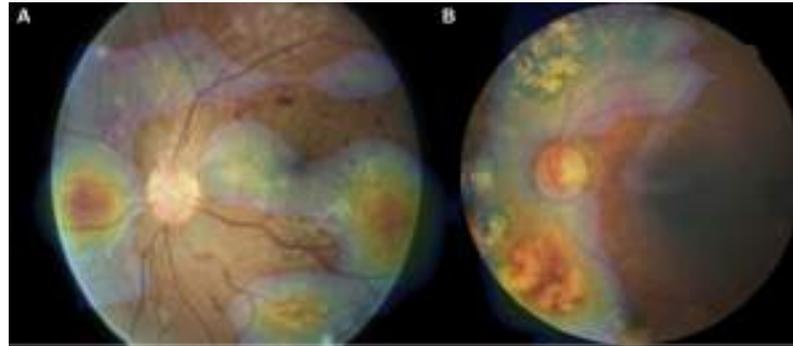


Figura 2.16: Mapas de anomalías obtenidos utilizando el método de detección de RD basado en *deep learning* propuesto por Gargeya *et al.* [69]. Fuente: [69].

únicamente cuando estos datos están disponibles. Por otro lado, el proceso de entrenamiento demanda contar con hardware específico y de alto desempeño, o de lo contrario tomaría mucho tiempo entrenar un modelo o arquitectura apropiada para resolver el problema.

#### 2.4.5.1. Retinopatía diabética

Como se mencionó anteriormente, la utilización de técnicas de *deep learning* está limitada por la disponibilidad de datos suficientes para su entrenamiento. Esta característica volvió infactible durante años la aplicabilidad de DNNs para detección de RD. Sin embargo, la reciente liberación del conjunto de imágenes EyePACS [45] como parte del torneo Kaggle de clasificación de RD [93] ha promovido el desarrollo de este tipo de técnicas. En [163], por ejemplo, se propone una arquitectura para una CNN capaz de diferenciar distintos niveles de RD. Este enfoque fue evaluado sobre un subconjunto de 5000 imágenes de validación extraídas del conjunto original, demostrando su habilidad para discriminar los casos más evidentes. Sin embargo, las imágenes de los pacientes con RD leve no fueron detectados de forma adecuada, probablemente como consecuencia de que las imágenes son reducidas a una dimensión muy pequeña antes de alimentar la red, lo que perjudica su habilidad para identificar MAs.

Debido a que las etiquetas que indican la existencia o no de RD dependen fuertemente del número y tipo de lesiones observadas, se ha realizado un esfuerzo sig-

nificativo no sólo para etiquetar las imágenes de forma automática si no también para indicar las regiones que fueron consideradas relevantes por la CNN al momento de asignar las etiquetas. Por medio de este tipo de enfoques, los oftalmólogos son más capaces de determinar si el grado de RD asociado es correcto o no. Quellec *et al.* [165], for ejemplo, propusieron adaptar las CNNs de forma tal que puedan obtenerse distintos mapas de temperatura representando las áreas principales de activación de la red. Los autores demostraron que, mediante la optimización conjunta de una función que asigna un costo a los errores de clasificación y a los producidos por el mapa de temperatura, se obtienen segmentaciones difusas de las áreas en las que ocurren lesiones brillantes y rojas. De manera similar, el método presentado en [69] y entrenado sobre EyePACS permite recuperar mapas de anormalidades que resaltan las regiones de las imágenes en las que el modelo identificó lesiones (Figura 2.16). De manera alternativa, Wang *et al.* [212] propusieron añadir un mapa de activaciones por regresión a la CNN para recuperar estas áreas relevantes.

#### 2.4.5.2. Glaucoma

El desarrollo de métodos automatizados para la detección de glaucoma utilizando *deep learning* está limitado por el hecho de que no existen grandes conjuntos de datos anotados disponibles de forma pública y accesible. Por el contrario, la gran mayoría de las imágenes disponibles contienen segmentaciones del disco óptico, que son útiles para entrenar métodos para la segmentación de esta estructura y de la copa óptica que pueden o no basarse en técnicas de aprendizaje profundo [118]. Conforme al relevamiento realizado del estado del arte, el único trabajo observado que utiliza CNNs para la detección de glaucoma es el presentado en [34]. En dicho artículo se propone entrenar una CNN a partir de imágenes recortadas en torno al disco óptico, y que, al aplicarse sobre dos conjuntos privados de imágenes de pacientes con y sin glaucoma, permite obtener muy buenos resultados. Debido a que el modelo es entrenado sobre un conjunto de imágenes relativamente pequeño, la arquitectura de la red no es demasiado profunda. Por otra parte, no se introduce estrategia alguna para validar el criterio seguido por la CNN al momento de etiquetar las imágenes, lo que limita la posibilidad de que los expertos puedan evaluar las salidas correspondientes.

---

## Capítulo 3

# Segmentación de vasos sanguíneos

La gran mayoría de los sistemas actuales para el diagnóstico de enfermedades visuales asistido por computadora requiere primero la segmentación automática de la vasculatura retiniana. Esto es motivado por el hecho de que los vasos sanguíneos son, en conjunto con el disco óptico, la parte anatómica de la retina más visible. Además, está demostrado que una gran variedad de enfermedades tanto oftalmológicas como sistémicas se manifiestan a través de cambios en la arquitectura y distribución general de los vasos sanguíneos de la retina.

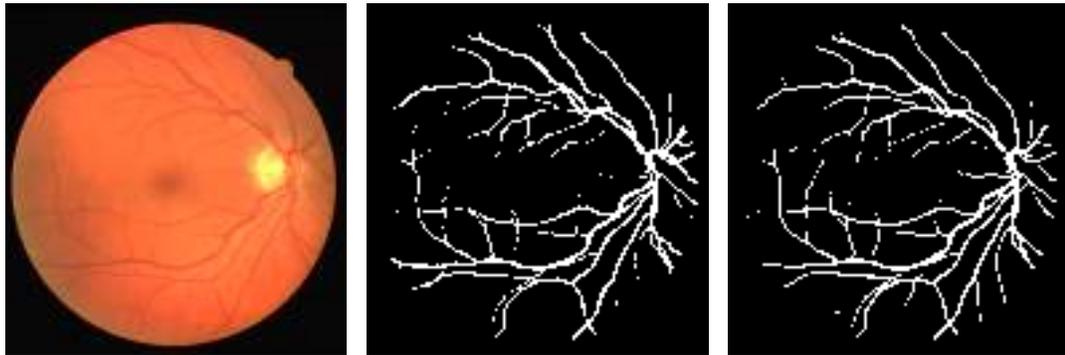
En este capítulo se propone un método novedoso para la segmentación de vasos sanguíneos en fotografías de fondo de ojo, basado en un campo aleatorio condicional totalmente conectado. En general los modelos tradicionales de segmentación utilizan restricciones de forma que fallan cuando deben lidiar con estructuras delgadas y elongadas tales como los vasos sanguíneos. Aquí se propone resolver esta dificultad mediante el uso de un campo aleatorio condicional con potenciales más expresivos, aprovechando avances recientes que permiten la inferencia en campos totalmente conectados casi en tiempo real. Los parámetros del método son aprendidos de manera automática por medio de una máquina de vectores de soporte de salida estructurada, una técnica supervisada que es ampliamente utilizada para predecir salidas estructuradas en numerosas aplicaciones de aprendizaje automático. El método aquí propuesto, entrenado utilizando descriptores básicos y conocidos en el estado del arte, es evaluado tanto cuantitativa como cualitativamente en una serie de bancos de imágenes de acceso público: DRIVE, STARE, CHASEDB1 y HRF. Además, se

provee una comparación cuantitativa con respecto a otras estrategias conocidas. Los resultados muestran que este enfoque supera a otras técnicas al ser evaluados utilizando métricas de calidad estándar. Además, se observó que el modelo totalmente conectado permite distinguir mejor las estructuras deseadas con respecto al modelo más tradicional, basado en vecindarios locales. Los resultados sugieren, por otra parte, que el método puede explotarse para contribuir en otras aplicaciones médicas y biológicas.

El capítulo se organiza de la siguiente forma: en la Sección 3.1 se resumen la importancia de la segmentación de los vasos sanguíneos para el análisis de fotografías de fondo de ojo, y se ilustra más en detalle la necesidad de utilizar modelos más robustos que los actuales. La Sección 3.2 explica en detalle el método propuesto, mientras que la Sección 3.3 provee información acerca de las bases de datos utilizadas y las métricas cuantitativas empleadas en la evaluación experimental. La Sección 3.4 presenta los resultados obtenidos, incluyendo una comparación con otros enfoques recientemente publicados. La Sección 3.5 discute las ventajas del enfoque propuesto y plantea algunas futuras líneas de investigación. Finalmente, la Sección 3.6 concluye el capítulo.

La implementación de este algoritmo, de código abierto, está disponible para su descarga y utilización en <https://github.com/ignaciorldo/fundus-vessel-segmentation-tbme>. El trabajo presentado en el capítulo fue publicado en los siguientes artículos:

- J. I. Orlando y M. del Fresno. Reviewing preprocessing and feature extraction techniques for retinal blood vessel segmentation in fundus images. *Mecánica Computacional*, XXXIII(42):2729–2743, 2014
- J. I. Orlando y M. B. Blaschko. Learning fully-connected CRFs for blood vessel segmentation in retinal images. En P. Golland, C. Barillot, J. Hornegger, y R. Howe, eds., *MICCAI 2014, LNCS*, tomo 8149, págs. 634–641. Springer, 2014
- J. I. Orlando, E. Prokofyeva, y M. B. Blaschko. A discriminatively trained fully connected conditional random field model for blood vessel segmentation



(a) Fotografía de fondo de ojo (b) Segmentación manual (c) Segmentación automática

Figura 3.1: Una fotografía de fondo de ojo del conjunto de test de DRIVE [141] y su segmentación de los vasos sanguíneos asociada, obtenida manualmente por un experto y utilizando el enfoque propuesto en esta tesis.

in fundus images. *IEEE Transactions on Biomedical Engineering*, 64(1):16–27, 2017

### 3.1. Motivación

Los vasos sanguíneos de la retina constituyen una de las referencias anatómicas más relevantes en las fotografías de fondo de ojo (Figura 3.1). En particular, la vasculatura retiniana provee información valiosa que resulta de utilidad para el diagnóstico, detección, tratamiento y evaluación tanto de la RD como del glaucoma [2]. Como se analizó en el Capítulo 2, el análisis de los atributos morfológicos de los vasos sanguíneos permite el estudio y seguimiento de la evolución de estas enfermedades (ver Sección 2.4.2). En otras aplicaciones, el árbol vascular debe ser segmentado previamente para facilitar la detección de lesiones con intensidades similares, tales como las microaneurismas y las hemorragias [52], o para la obtención de segmentaciones más precisas de otras regiones anatómicas tales como el disco o la copa óptica [75]. Las bifurcaciones de los vasos sanguíneos también pueden utilizarse como referencias o *landmarks* en algoritmos de registración de imágenes de la retina [2], lo que resulta particularmente valioso para el análisis en simultáneo de diferentes estudios de un mismo paciente, adquiridos mediante diferentes moda-

lidades de imagen, tales como las fotografías de fondo de ojo o las angiografías por fluoresceína.

Toda herramienta automática para la caracterización o análisis del árbol vascular retiniano requiere inicialmente su segmentación precisa. La práctica actual más corriente es realizar esta segmentación de forma manual, aunque esto representa una tarea tediosa, que consume mucho tiempo y que además requiere ser realizada por usuarios expertos. Por otro lado, las dificultades que puedan surgir como consecuencia de una captura errónea de las imágenes (tales como un contraste inadecuado entre los vasos sanguíneos y el fondo, o una iluminación del fondo despereja), sumadas a la variabilidad en el ancho, brillo y forma de los segmentos vasculares, reducen significativamente la coincidencia entre segmentaciones realizadas por diferentes observadores [46]. Estos hechos motivan el desarrollo de estrategias automáticas para la segmentación de los vasos sanguíneos sin intervención manual [2].

Aunque se han realizado numerosos esfuerzos en el área, esta tarea continúa siendo un área de investigación muy activa debido al potencial impacto que representa la obtención de segmentaciones más precisas [63]. En general, los enfoques existentes pueden clasificarse en dos categorías principales: supervisados y no supervisados. Los métodos supervisados requieren un conjunto de muestras de entrenamiento (en general, un conjunto de descriptores a nivel de píxel y sus correspondientes anotaciones conocidas) para aprender un modelo o un clasificador. Diversos clasificadores han sido considerados en la literatura del área, incluyendo  $k$ -vecinos más cercanos [141], clasificadores bayesianos [188], máquinas de vectores de soporte (SVMs) [220, 225], redes neuronales artificiales [121, 201], árboles de decisión [64, 65], modelos de mezcla de Gaussianas [46], AdaBoost [117], entre otros. Un filtro entrenable, denominado B-COSFIRE, fue recientemente introducido en [20] para resaltar y caracterizar el árbol vascular de la retina. Aunque el método es no supervisado en el sentido de que no requiere el entrenamiento de un clasificador, la estrategia que los autores siguen para el ajuste de sus parámetros está basada en la utilización de datos de entrenamiento. Las redes neuronales convolucionales o CNNs también constituyen un método supervisado, y fueron aplicadas recientemente para la segmentación de los vasos sanguíneos de la retina [67, 118].

Por otro lado, los métodos no supervisados constituyen una familia de enfoques

que son capaces de segmentar la vasculatura retiniana sin necesidad de datos previamente anotados. Sin embargo, estas estrategias usualmente reportan resultados de menor precisión que los métodos supervisados. En general, la mayoría de los enfoques correspondientes a esta familia están basados en la aplicación de umbrales, en técnicas para la detección simultánea de las paredes de los vasos [223] o en enfoques orientados a regiones, tales como crecimiento de regiones [61, 62, 173] o contornos activos [8, 228]. En la gran mayoría de los casos, estos algoritmos se utilizan luego de aplicar alguna técnica de realce de los vasos sanguíneos. Esta tarea se realiza por lo general utilizando operadores morfológicos [226], respuestas a filtros matcheados [31, 94, 145], transformadas Wavelet continuas y complejas [56], entre otros enfoques [190]. El método que se propone en esta tesis es un método supervisado.

Los campos condicionales aleatorios (o *Conditional Random Fields*, CRFs) son ampliamente utilizados para la segmentación de imágenes en numerosas aplicaciones [78, 103, 112]. Sin embargo, y de acuerdo al relevamiento realizado de la literatura del área, nunca fueron aplicados aún para la segmentación de vasos sanguíneos en imágenes de fondo de ojo. Esto quizá se debe a que los potenciales de pares estándar, tales como el modelo de Potts, asignan una prioridad baja a tales estructuras elongadas. Este hecho motivó la introducción, en el marco de esta tesis doctoral, de un método novedoso para la segmentación de los vasos sanguíneos, basado en CRFs totalmente conectados. Estos modelos fueron previamente aplicados en [91] y [92] para la segmentación de hígado y tumores cerebrales en tomografía computada y resonancia magnética, respectivamente, aunque sus implicaciones en la segmentación de estructuras bidimensionales y delgadas no han sido previamente estudiadas. En los experimentos realizados en este trabajo se demostró que la conectividad densa de este tipo de modelos incrementa la capacidad del método para detectar estructuras elongadas, resolviendo la dificultad original de los CRFs basados en vecindarios locales y mejorando significativamente sus resultados. Esta propiedad puede contribuir potencialmente a la mejora de numerosas aplicaciones médicas y biológicas en las que la segmentación de estructuras elongadas es necesaria, incluyendo el fenotipado automático de raíces de plantas [60] o el análisis de neuronas [79].

Como se muestra en [101], los clasificadores locales llevan a problemas de detec-

ción que pueden mitigarse mediante la incorporación en el proceso de aprendizaje de información relativa a la forma de las estructuras deseadas. Los CRFs son capaces de brindar este tipo de información a través de los denominados potenciales de a pares, o *pairwise*. Las máquinas de vectores de soporte de salida estructurada (Structured Output SVMs, SOSVMs) han sido ampliamente utilizadas en la literatura para el aprendizaje de CRFs basados en vecindarios locales [87, 197]. Sin embargo, el entrenamiento de CRFs densos utilizando este tipo de modelos fue evitado previamente dado que los mismos han sido intratables merced a su elevado costo computacional. Esto se debe particularmente a que las SOSVMs requieren realizar múltiples llamadas al algoritmo de inferencia durante el entrenamiento, y a que la inferencia en CRFs de conectividad densa es usualmente muy costosa. En esta tesis, sin embargo, se propone resolver esta dificultad mediante el uso de avances recientes en materia de inferencia eficiente de CRFs totalmente conectados, que permiten obtener segmentaciones binarias de inmediato [101].

En este capítulo se presenta un método para la segmentación de vasos sanguíneos en imágenes de fondo de ojo basado en el aprendizaje automático de CRFs totalmente conectados utilizando SOSVMs. También se presenta una estrategia para la estimación de otros parámetros adicionales del método tales como la configuración de los descriptores utilizados, a los efectos de optimizar su rendimiento durante el entrenamiento. El método es evaluado tanto cuantitativa como cualitativamente en cuatro bases de datos de disponibilidad pública (DRIVE, STARE, CHASEDB1 and HRF) para estudiar el comportamiento del método en diferentes escenarios, incluyendo imágenes de pacientes sanos, con patologías y tomadas a diferentes resoluciones. De acuerdo a los experimentos realizados, este algoritmo supera a las técnicas disponibles y actuales al evaluarse utilizando diferentes medidas de calidad.

## 3.2. Métodos

Esta sección explica en detalle el método propuesto. Inicialmente se describen los campos condicionales aleatorios en sus formulaciones basadas en vecindarios locales y totalmente conectadas (Sección 3.2.1). A continuación, se detalla la estrategia de aprendizaje de dichos modelos por medio de una SOSVM (Sección 3.2.2). Los

descriptores utilizados para evaluar el modelo se explican en la Sección 3.2.3. Finalmente, la Sección 3.2.4 describe un factor de compensación que puede ser utilizado para segmentar imágenes con diferentes resoluciones, sin necesidad de recalibrar los parámetros de los descriptores.

### 3.2.1. Campos condicionales aleatorios (CRFs) para segmentación de imágenes

El problema de la segmentación de imágenes puede modelarse como un problema de minimización de energía en un campo aleatorio condicional (CRF). En la definición original de estos modelos, las imágenes se mapean a grafos en los que cada píxel representa un nodo del mismo, y cada nodo está conectado por un arco a sus vecinos, de acuerdo a determinada regla de conectividad [101, 103, 104]. En los CRFs basados en vecindarios locales, los nodos están conectados con sus 4 píxeles vecinos (norte, sur, este y oeste) [28], mientras que en la versión totalmente conectada cada nodo se asume como conectado a todo otro píxel de la imagen [101].

Sea  $\mathbf{y} = \{y_i\}$  un etiquetado sobre todos los píxeles de la imagen  $I$  en el espacio de etiquetas  $\mathcal{L} = \{-1, 1\}$ , donde 1 está asociado a vasos sanguíneos y -1 a cualquier otra clase. Un campo aleatorio condicional  $(I, \mathbf{y})$  es caracterizado por la distribución de Gibbs:

$$p(\mathbf{y}|I) = \frac{1}{Z(I)} \exp\left(-\sum_{c \in \mathcal{C}_{\mathcal{G}}} \Phi_c(\mathbf{y}_c|I)\right) \quad (3.1)$$

donde  $Z(I)$  es una constante de normalización,  $\mathcal{G}$  es el grafo asociado a  $I$  y  $\mathcal{C}_{\mathcal{G}}$  es un conjunto de cliques en  $\mathcal{G}$ , cada uno induciendo un potencial  $\Phi_c$  [112]. Esta distribución determina la probabilidad condicional de un etiquetado  $\mathbf{y}$  dada la imagen  $I$ . La función de energía de Gibbs puede derivarse a partir de la siguiente función de verosimilitud, o *likelihood*:

$$E(\mathbf{y}|I) = \sum_{c \in \mathcal{C}_{\mathcal{G}}} \Phi_c(\mathbf{y}_c|I) \quad (3.2)$$

Así, el etiquetado maximum a posteriori (MAP) puede obtenerse mediante la mini-

mización de la correspondiente energía:

$$\mathbf{y}^* = \arg \min_{\mathbf{y} \in \mathcal{L}} E(\mathbf{y}|I) \quad (3.3)$$

Como resultado de minimizar  $E(\mathbf{y}|I)$ , se obtiene una segmentación binaria del árbol vascular. Por conveniencia en la notación, se omitirá el condicionamiento durante el resto de este capítulo, y se utilizará  $\psi_c(\mathbf{y}_c)$  para denotar  $\Phi_c(\mathbf{y}_c|I)$ . Por otro lado, se considerarán energía que pueden descomponerse como sumatorias sobre potenciales unarios y de a pares o *pairwise*, a diferencia de otros potenciales de mayor orden [100].

Dado un grafo  $\mathcal{G}$  en  $\mathbf{y}$ , su energía se obtiene mediante la sumatoria de sus potenciales unarios y *pairwise* (denotados  $\psi_u$  y  $\psi_p$ , respectivamente):

$$E(\mathbf{y}) = \sum_i \psi_u(y_i, \mathbf{x}_i) + \sum_{(i,j) \in \mathcal{C}_{\mathcal{G}}} \psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) \quad (3.4)$$

donde  $\mathbf{x}_i$  y  $\mathbf{f}_i$  son los descriptores unarios y pairwise, respectivamente. Los potenciales unarios definen una log-verosimilitud o *log-likelihood* sobre la asignación de etiquetas  $\mathbf{y}$ , generalmente obtenidos mediante un clasificador supervisado [101]. Los potenciales *pairwise* definen una distribución similar pero considerando únicamente las interacciones entre los descriptores de cada píxel y sus etiquetas, de acuerdo a  $\mathcal{C}_{\mathcal{G}}$ , que se determina a partir de la conectividad del grafo.

Los potenciales unarios son comunes tanto para los CRFs basados en vecindarios locales como para los totalmente conectados, y se obtienen mediante:

$$\psi_u(y_i, \mathbf{x}_i) = -\langle \mathbf{w}_{u_{y_i}}, \mathbf{x}_i \rangle - \mathbf{w}_{\beta_{y_i}} \beta \quad (3.5)$$

donde  $\beta$  es una constante de sesgo o *bias*, y  $\mathbf{w}_{u_{y_i}}$  y  $\mathbf{w}_{\beta_{y_i}}$  representan los vectores de pesos para los descriptores y la constante de sesgo, respectivamente, ambos asociados a la etiqueta  $y_i$ . El vector unario  $\mathbf{x}_i$  está dado por una combinación arbitraria de los descriptores extraídos de la imagen.

Los potenciales *pairwise* se definen como una combinación lineal de funciones.

Así, la energía *pairwise* se obtiene de acuerdo a:

$$\psi_p(y_i, y_j, \mathbf{f}_i, \mathbf{f}_j) = \mu(y_i, y_j) \sum_{m=1}^M w_p^{(m)} k^{(m)}(f_i^{(m)}, f_j^{(m)}) \quad (3.6)$$

donde cada  $k^{(m)}$  es una función fija sobre un descriptor arbitrario  $f^{(m)}$ ,  $w_p^{(m)}$  es un peso de la combinación lineal, y  $\mu(y_i, y_j)$  representa una función de compatibilidad de etiquetas. Los núcleos o *kernels* Gaussianos determinan, por medio del descriptor  $f^{(m)}$ , la similaridad entre dos píxeles adyacentes. Dado que la información sobre la vecindad de dos píxeles es provista por la regla de conectividad que sigue el modelo, estos *kernels* dependerán de la formulación del CRF utilizada, con lo cual serán descritos en las próximas subsecciones. Los términos restantes son detallados a continuación.

La función de compatibilidad  $\mu$  penaliza que píxeles con descriptores similares sean asignados a etiquetas diferentes, y están dadas por el modelo de Potts  $\mu(y_i, y_j) = [y_i \neq y_j]$ . Nótese que en esta ecuación se utiliza la notación de corchetes de Iverson, donde  $[\cdot]$  indica 1 si la sentencia entre corchetes es verdadera y 0 en cualquier otro caso.

Los parámetros  $\mathbf{w}_u$ ,  $\mathbf{w}_p^{(m)}$  controlan la relevancia de los descriptores unarios y los núcleos *pairwise* en la función global de energía, respectivamente. Además,  $\mathbf{w}_\beta$  es utilizado para aprender automáticamente el término de sesgo. El ajuste de estos parámetros no es factible de ser realizado manualmente debido a su elevada dimensionalidad, por lo que en este trabajo se propone aprenderlos utilizando una máquina de vectores de soporte de salida estructurada (*Structured Output SVM*), explicada en detalle en la Sección 3.2.2.

### 3.2.1.1. CRFs basados en vecindarios locales

Los CRFs basados en vecindarios locales (LNB-CRFs) están definidos sobre grafos en forma de grilla. Luego, en este tipo de modelos, cada nodo (píxel) se asume como conectado por un arco a sus vecinos de orden 4 (esto es, ubicados al norte, sur, este y oeste del píxel de análisis).

La función de los potenciales *pairwise*, para un  $m$ -ésimo descriptor *pairwise*, se

obtiene a partir de:

$$k^{(m)}\left(f_i^{(m)}, f_j^{(m)}\right) = \frac{|f_i^{(m)} - f_j^{(m)}|}{2\theta_{(m)}^2} \quad (3.7)$$

donde  $\theta_{(m)}$  es un parámetro que controla la relevancia de las disimilaridades entre los descriptores de los píxeles.

La energía de estos modelos basados en grillas pueden ser minimizadas de forma muy eficiente utilizando el algoritmo de corte-mínimo/flujo-máximo (*min-cut/max-flow*) propuesto en [28].

### 3.2.1.2. CRFs totalmente conectados

En la variante totalmente conectada de los CRFs (FC-CRF), cada nodo del grafo se asume vecino de todo otro píxel de la imagen. Mediante la utilización de estos potenciales de orden alto, el método puede tener en cuenta no sólo la información de los vecinos cercanos sino también las interacciones de más largo alcance entre los píxeles. Esta propiedad mejora la precisión de la segmentación, aunque la implementación del proceso de inferencia necesario para su minimización es por lo general muy costosa computacionalmente. Recientemente, sin embargo, Krähenbühl y Koltun [101] introdujeron un método eficiente para inferir esta segmentación, estableciendo la restricción de que los potenciales *pairwise* deben modelarse como una combinación lineal de *kernels* Gaussianos sobre un espacio euclideo de descriptores. Este enfoque, que está basado en la aproximación del campo medio del CRF original, es capaz de producir segmentaciones precisas en pocos segundos.

Los núcleos *pairwise* del modelo totalmente conectado se definen como:

$$k^{(m)}\left(f_i^{(m)}, f_j^{(m)}\right) = \exp\left(-\frac{|\mathbf{p}_i - \mathbf{p}_j|^2}{2\theta_p^2} - \frac{|f_i^{(m)} - f_j^{(m)}|^2}{2\theta_{(m)}^2}\right) \quad (3.8)$$

donde  $\mathbf{p}_i$  y  $\mathbf{p}_j$  son los vectores de coordenadas de los píxeles  $i$  y  $j$ . La ubicación de los píxeles se incluye en los términos *pairwise* para incrementar el efecto que los píxeles más cercanos tienen sobre la energía global con respecto a la influencia de los píxeles más lejanos entre sí. Los parámetros  $\theta_p$  y  $\theta_{(m)}$  de los *kernels* controlan el grado de relevancia de ambos *kernels* en la expresión. Así, cuando  $\theta_p$  crece, se tienen

en cuenta interacciones de más alto rango. Por el contrario, cuando  $\theta_p$  disminuye, únicamente los vecindarios más locales afectan significativamente los resultados. De manera similar, cuando  $\theta_{(m)}$  crece o decrece, se toleran mayores o menores diferencias en el  $m$ -ésimo descriptor, respectivamente.

### 3.2.2. Aprendizaje de CRFs mediante SVM de salida estructurada

El objetivo es aprender un vector de la forma  $\mathbf{w} = (\mathbf{w}_u, \mathbf{w}_\beta, \mathbf{w}_p)$ , donde  $\mathbf{w}_u$ ,  $\mathbf{w}_\beta$  y  $\mathbf{w}_p$  son los pesos para los descriptores unarios, para el término de sesgo y para los núcleos *pairwise*, respectivamente. El vector  $\mathbf{w}$  puede tener una alta dimensionalidad si se consideran muchos descriptores, por lo que su ajuste manual o utilizando técnicas tales como búsqueda secuencial no es factible para brindar soluciones en un tiempo razonable. El aprendizaje supervisado de los potenciales unarios de forma separada con respecto a los potenciales *pairwise* puede considerarse como una alternativa superadora respecto al enfoque anterior, aunque este enfoque ignora la influencia que los potenciales *pairwise* pueden tener sobre la formulación general de la energía, conduciendo a peores resultados que los que pueden obtenerse aprendiendo en simultáneo ambos pesos.

Para evitar los inconvenientes anteriormente mencionados, aquí se propone obtener el vector completo  $\mathbf{w}$  de forma supervisada, utilizando la formulación de la SOSVM con reescalado de margen y 1 variable de holgura (*1-slack SOSVM with margin rescaling*) presentada en [87]. Este enfoque de entrenamiento discriminativo ha demostrado resultados prometedores en la construcción de modelos altamente complejos y precisos en numerosas áreas, incluyendo la detección automática de objetos, la segmentación de imágenes y otras aplicaciones de visión computacional, incluso utilizando grandes conjuntos de datos. Sin embargo, las SOSVMs nunca fueron aplicadas antes para la tarea de aprender FC-CRFs.

Sea  $S = \{(s^{(1)}, y^{(1)}), \dots, (s^{(n)}, y^{(n)})\}$  un conjunto de entrenamiento o *training set*, en donde  $n$  es el número de imágenes de entrenamiento. Cada  $y^{(i)}$  corresponde al etiquetado real o *ground truth* de la  $i$ -ésima imagen en el conjunto de entrenamiento. Cada conjunto  $s^{(i)} = \{x^{(i)}, \beta, f^{(i)}\}$  contiene el conjunto  $x^{(i)}$  de vectores de descriptores unarios, una constante de sesgo  $\beta = 1$ , y el conjunto  $f^{(i)}$  de descriptores *pairwise*

para cada píxel de la imagen.

Los pesos  $\mathbf{w}$  se obtienen resolviendo el problema de minimización:

$$\min_{\mathbf{w}, \xi \geq 0} \frac{1}{2} \|\mathbf{w}\|^2 + C\xi \quad (3.9)$$

sujeto a:

$$\forall (\bar{y}^{(1)}, \dots, \bar{y}^{(n)}) : \sum_{i=1}^n \langle \mathbf{w}, \psi(s^{(i)}, y^{(i)}) - \psi(s^{(i)}, \bar{y}^{(i)}) \rangle \geq \sum_{i=1}^n \Delta(y^{(i)}, \bar{y}^{(i)}) - \xi \quad (3.10)$$

donde  $C$  es una constante de regularización;  $\xi$  es una variable de holgura compartida por todas las restricciones  $\bar{y}^{(i)}$ ;  $\varphi(s, y)$  es una función de mapeo de descriptores que relaciona un conjunto  $s$  dado con un etiquetado  $y$ ; y  $\Delta(y, \bar{y})$  es una función de pérdida que evalúa las diferencias entre el *ground truth*  $y$  y una restricción  $\bar{y}$ . En este trabajo, se define  $\Delta$  como la pérdida de Hamming, dada por:

$$\Delta(y, \bar{y}) = \sum_i [y_i \neq \bar{y}_i] \quad (3.11)$$

donde nuevamente se utiliza la notación de corchetes de Iverson. Esta función penaliza todas las diferencias entre el etiquetado predicho y la segmentación de referencia provista por el especialista.

La función de mapeo de descriptores es definida en este trabajo como:

$$\varphi(s, y) = \left( \sum_k \varphi_u(\mathbf{x}_k, y_k), \sum_k \varphi_\beta(\beta, y_k), \sum_k \sum_{j < k} \varphi_p(y_k, y_j, \mathbf{f}_k, \mathbf{f}_j) \right) \quad (3.12)$$

donde sus componentes representan la suma de una función de mapeo de descriptores unarios, del sesgo y de los descriptores *pairwise*, respectivamente, para todos los píxeles de la imagen. A continuación se brindarán descripciones precisas de  $\varphi_u$ ,  $\varphi_\beta$ , y  $\varphi_p$ .

Primero es necesario definir un vector  $\varphi_y(y_i) \in \{0, 1\}^{|\mathcal{L}|}$  tal que:

$$\varphi_y(y_i) = \begin{cases} (1, 0) & \text{if } y_i = -1 \\ (0, 1) & \text{if } y_i = 1 \end{cases} \quad (3.13)$$

Las funciones de mapeo de descriptores individuales se definen como:

$$\varphi_u(\mathbf{x}_k, y_k) = \mathbf{x}_k \otimes \varphi_y(y_k) \quad (3.14)$$

$$\varphi_\beta(\beta, y_i) = \beta \varphi_y(y_i) \quad (3.15)$$

$$\forall m : [\varphi_p(y_k, y_j, \mathbf{f}_k, \mathbf{f}_j)]_m = \mu(y_i, y_j) k^{(m)}(f_i^{(m)}, f_j^{(m)}) \quad (3.16)$$

donde  $\otimes$  es el producto de Kronecker. La ecuación (3.9) puede resolverse eficientemente utilizando el algoritmo de planos cortantes o *cutting-planes* propuesto en [87].

### 3.2.3. Descriptores

El método propuesto es evaluado utilizando descriptores que están ampliamente difundidos en el área de segmentación de vasos sanguíneos en fotografías de fondo de ojo: las respuestas a detectores de líneas multiescala presentado por Nguyen *et al.* en [138] y a la transformada Wavelet 2D de Gabor [188] son utilizadas para el cálculo de los potenciales unarios, mientras que la imagen con los vasos realizados mediante el método de Zana y Klein [226] fue empleada en los potenciales *pairwise*. En [151] se provee un análisis detallado de muchos otros descriptores útiles para esta tarea.

Todos estos descriptores se extraen a partir de imágenes en escala de gris, obtenidas a partir de una versión invertida de la banda verde de la imagen RGB original, como ocurre a menudo en otros trabajos [20, 121]. Además, y debido a que los descriptores utilizados pueden introducir detecciones erróneas en los bordes del FOV, se replica la estrategia propuesta en [188] para simular una mayor apertura del dispositivo de captura. Por medio de esta técnica, las detecciones falsas que ocurran por fuera del FOV original puede ser fácilmente removidas simplemente multiplicando la imagen resultante por la máscara del FOV original. La Figura 3.2(b) muestra un ejemplo del preprocesamiento de una imagen dada.

Los detectores de línea de Nguyen *et al.* explotan la propiedad de que los vasos sanguíneos aparecen como estructuras elongadas en las imágenes. La intensidad promedio es calculada a partir de una línea de longitud  $l$  que atraviesa cada

píxel objetivo  $\mathbf{P}$ , para diferentes ángulos de orientación  $\alpha$ . La línea con la mayor intensidad promedio  $L_l(\mathbf{P})$  es seleccionada respecto a todas las orientaciones consideradas, y el nivel de linearidad asociado al píxel se calcula tomando la diferencia  $S_l(\mathbf{P}) = L_l(\mathbf{P}) - N_s(\mathbf{P})$ , con  $N_s(\mathbf{P})$  equivalente al promedio de las intensidades en un vecindario cuadrado centrado en  $\mathbf{P}$  y de lado  $s$ . En la Figura 3.2(c) se presenta un ejemplo de la respuesta obtenida para este filtro, utilizando  $l = 15$ . En la versión original de este descriptor, se obtiene un único mapa de linearidades tomando diferentes longitudes de  $l$  y sumándolos a la banda de verde invertida de la imagen original. Posteriormente, este mapa es umbralado para obtener una segmentación binaria. En este trabajo se consideran tanto  $S_l$  como la imagen invertida de manera separada, dado que nuestro algoritmo es capaz de aprender los mejores pesos para combinar estos descriptores. Así, en lugar de tener un único valor para cada píxel, se tiene un vector de características compuesto por las respuestas a cada valor de  $l$  y la imagen  $I$ .

Las wavelets 2D de Gabor tienen la capacidad de detectar estructuras orientadas y pueden ajustarse a frecuencias específicas. Esta propiedad es especialmente útil para realzar la vasculatura retiniana, teniendo en cuenta que los vasos sanguíneos aparecen en diferentes tamaños y orientaciones. Este descriptor es calculado en este trabajo exactamente como se reportan Soares *et al.* en [188], para diferentes escalas  $a$ . Las respuestas de la imagen a esta wavelet, tomadas para diferentes valores de  $a$ , se incluyen como descriptores. En la Figura 3.2(d) se ilustra el ejemplo de una respuesta obtenida para  $a = 3$ .

La técnica de Zana y Klein para el realce de vasos aprovecha el hecho de que los vasos son lineales, conexos y con una curvatura que varía suavemente a lo largo de su eje central [226]. Para su cálculo, inicialmente se reduce el ruido de la imagen utilizando una operación morfológica conocida como *opening* por reconstrucción, que consiste en aplicar diferentes *openings* utilizando elementos estructurantes de longitud  $l$  para diferentes ángulos. Posteriormente, se aplican múltiples operaciones morfológicas *top-hat* utilizando idénticos elementos estructurantes, y se toma la suma de las respuestas correspondientes para cada ángulo de orientación dado. Esta transformación reduce el efecto de potenciales píxeles ruidosos brillantes y mejora el contraste de todos los componentes lineales. Aquellas estructuras cuya curvatura es

linealmente coherente son detectadas a continuación por medio de una evaluación de la curvatura cruzada, aplicando un operador LoG (Gaussiano del Laplaciano) para distintas ventanas de tamaño  $7 \times 7$  píxeles y con un desvío estándar 1.75. Finalmente, se aplica un filtro de alternancia compuesto por la aplicación sucesiva de un *opening* morfológico, un *closing* y un *opening* para reducir detecciones falsas a patrones no lineales en regiones delgadas brillantes y oscuras, y en zonas del fondo en las que aparezcan estructuras lineales. En las tres operaciones últimas se utiliza el mismo elemento estructurante de longitud  $l$ . En diferentes experimentos realizados se observó que este descriptor es altamente sensible a la iluminación despareja del fondo, lo que degrada su habilidad para caracterizar los vasos sanguíneos de forma efectiva. A los efectos de mejorar su calidad se incorporó en este trabajo un preprocesamiento adicional, consistente en estimar la iluminación del fondo y sustraerla del canal verde de la imagen original. El fondo es obtenido mediante la convolución de la banda verde con un filtro de mediana, cuyo tamaño es lo suficientemente grande como para asegurar que la versión suavizada no contenga estructuras visibles tales como los vasos sanguíneos. Este enfoque ha sido aplicado en muchas oportunidades en la literatura [121, 126], y un ejemplo de la imagen resultante de su aplicación se ilustra en la Figura. 3.2(e). El descriptor de Zana y Klein se presenta en la Figura 3.2(f).

Todos los descriptores son normalizados de manera independiente a media 0 y desvío 1, utilizando la media y el desvío estándar de cada descriptor calculado para cada imagen [188].

### 3.2.4. Adaptación automática de los modelos a imágenes de diferente resolución

Aunque los pesos para los potenciales unarios y *pairwise* son ajustados durante el proceso de aprendizaje automático, el rendimiento del método propuesto siempre dependerá de la capacidad de los descriptores para caracterizar efectivamente las estructuras vasculares. En general, los descriptores son sensibles a sus propios parámetros, que en general están asociados a propiedades de los vasos tales como su calibre, que está a su vez relacionado con la resolución de las imágenes. Las respuestas a la transformada wavelet 2D de Gabor, por ejemplo, dependen de la escala  $a$  utilizada. De manera similar, los detectores de línea de Nguyen *et al.* y la estrategia

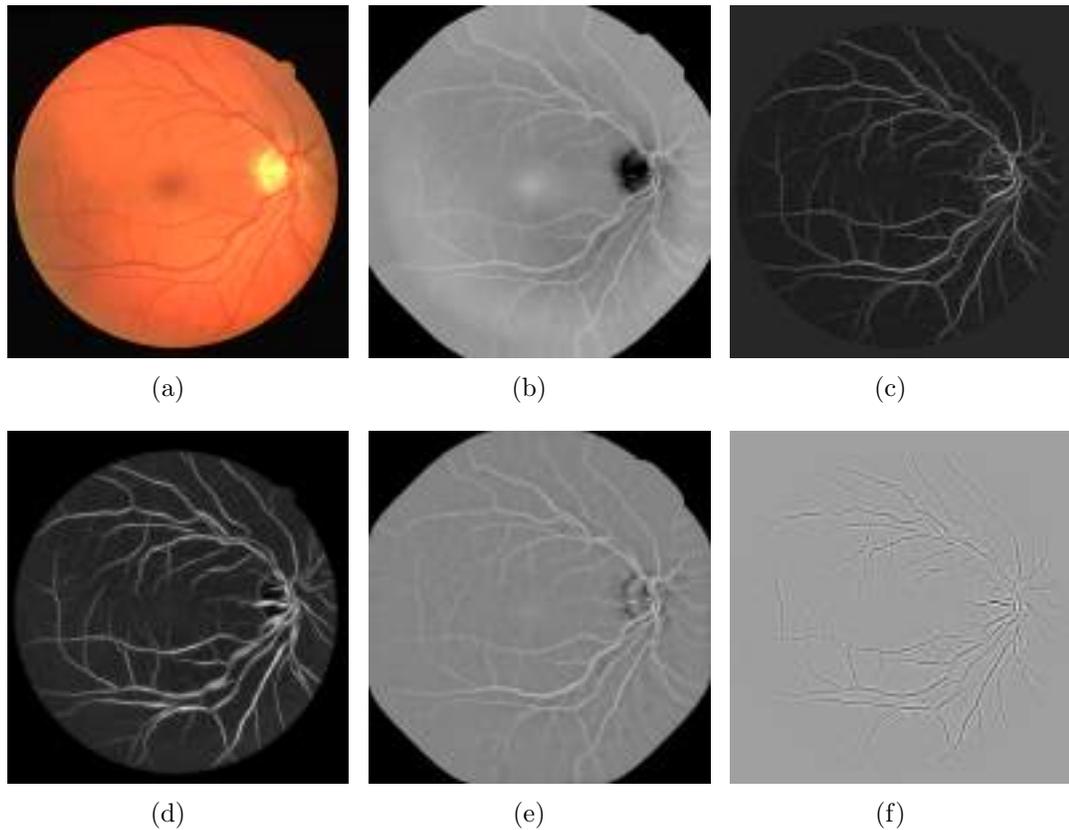


Figura 3.2: Preprocesamiento de las imágenes y ejemplos de descriptores unarios y *pairwise*. (a) Imagen a color original. (b) Canal verde invertido luego de la expansión del borde. (c) Respuestas al detector de líneas de Nguyen *et al.* ( $l = 15$ ). (d) Respuestas a la transformada wavelet 2D de Gabor propuesta por Soares *et al.* ( $a = 3$ ). (e) Imagen invertida luego de corregir la iluminación despareja. (f) Imagen realizada utilizando el método de Zana y Klein ( $l = 9$ ).

de realce de Zana y Klein dependen de la longitud  $l$  de los detectores o del elemento estructurante, respectivamente. La mayoría de los parámetros de estos descriptores fueron originalmente diseñados para su utilización en imágenes de baja resolución, como las del dataset DRIVE [193]. Cuando se aplican tales indicadores en imágenes de mayor resolución, la performance de los mismos se ve significativamente reducida si el proceso de extracción de características no es escalado proporcionalmente. Otros parámetros tales como los ángulos para calcular las respuestas a los descriptores en diferentes orientaciones, por el contrario, no son influidos por cambios en la

resolución de las imágenes.

Un comportamiento similar puede esperarse de los parámetros de la etapa de preprocesamiento, tales como el tamaño del filtro de mediana utilizado para estimar el fondo o el tamaño de la apertura simulada al expandir el borde del FOV. El parámetro  $\theta_p$  utilizado en los potenciales *pairwise* del FC-CRF también está influido por la resolución de las imágenes, dado que controla las interacciones de a pares según la distancia relativa de cada píxel.

El ajuste apropiado de estos parámetros es relevante cuando se aplica el algoritmo propuesto sobre imágenes de diferentes resoluciones, por lo que contar con una estrategia automática para su calibración es de gran utilidad. Utilizar métodos de búsqueda secuencial sobre las imágenes etiquetadas del conjunto de entrenamiento es computacionalmente prohibitivo debido a la alta dimensionalidad del espacio de búsqueda, dada por el número de parámetros y sus combinaciones. Como una alternativa al diseño o ajuste manual de estos parámetros, algunos autores proponen derivarlos a partir de su relación con el calibre del vaso de interés [20] o el tamaño de la cabeza del nervio óptico [160]. Sin embargo, ambas estrategias requieren conocer a priori el tamaño de estas estructuras, algo que es difícil teniendo en cuenta que estos tamaños varían de una imagen a otra. A continuación se propone una estrategia diferente para adaptar automáticamente los parámetros de los descriptores y del modelo aquí presentado para trabajar en imágenes de múltiples resoluciones.

En lugar de ajustar las configuraciones a la resolución de cada conjunto de datos, se propone un enfoque simple basado en estimar la mejor configuración de parámetros en un único *data set*, y luego adaptar tales parámetros multiplicándolos por un factor de compensación  $\rho = \frac{\mathcal{X}_{\text{new}}}{\mathcal{X}_{\text{training}}}$ , donde  $\mathcal{X}_{\text{training}}$  representa el ancho promedio del FOV en las imágenes utilizadas para configurar las escalas, y  $\mathcal{X}_{\text{new}}$  es el ancho promedio del FOV en las nuevas imágenes. Dado que los cambios en la resolución de la imagen tendrán un efecto en el número de píxeles asociados a la región del FOV, este enfoque aproxima la variación en el cálculo de los descriptores que pueda generarse por el cambio en la resolución.

### 3.3. Materiales y Evaluación

Esta sección describe los conjuntos de datos y las métricas utilizadas para evaluar el modelo propuesto. Además, se proveen detalles acerca de la estrategia utilizada para estimar los parámetros  $\theta_p$  and  $\theta_{(m)}$  del CRF, y el parámetro  $C$  de la SOSVM (Ecuación (3.9)).

#### 3.3.1. Bases de datos utilizadas

Los experimentos fueron realizados utilizando DRIVE [193], STARE [82], CHASEDB1 [64] y HRF [144, 145], cuatro bases de datos estándar de imágenes de fondo de ojo, disponibles y ampliamente utilizadas para la evaluación de algoritmos de segmentación de vasos sanguíneos.

DRIVE<sup>1</sup> incluye 40 fotografías de fondo de ojo a color (7 de ellas con patologías) con una apertura de FOV de 45°, 8 bits por canal de color y a una resolución de  $565 \times 584$  píxeles. La base de datos está dividida en conjuntos de entrenamiento y test, ambos conteniendo 20 imágenes cada uno. En el conjunto de test se proveen dos segmentaciones manuales diferentes para cada imagen, mientras que en el de entrenamiento solamente hay disponible una.

STARE<sup>2</sup> contiene 20 imágenes, 10 de las cuales poseen patologías visibles, capturadas con una apertura de FOV de 35°, con 8 bits por canal de color y a una resolución de  $700 \times 605$  píxeles. Dos observadores segmentaron manualmente todas las imágenes, aunque el segundo delineó un mayor número de vasos delgados que el primero [46]. A pesar de la variabilidad en la metodología de marcado de las imágenes de ambos observadores, la performance de los algoritmos es normalmente evaluada utilizando las segmentaciones del primer observador como *ground truth* [20, 121]. Las máscaras del FOV no están disponibles en el *data set* original, con lo cual se utilizaron las construidas por Marin *et al.*<sup>3</sup> [121]. El conjunto no está dividido en entrenamiento y test, y no se ha observado una metodología consistente en la literatura relevada. Luego, a los efectos de poder comparar el algoritmo propuesto con un mayor número de trabajos de la literatura, la evaluación en STARE se realizó

<sup>1</sup><http://www.isi.uu.nl/Research/Databases/DRIVE/>

<sup>2</sup><http://www.ces.clemson.edu/~ahoover/stare/probing/index.html>

<sup>3</sup><http://www.uhu.es/retinopathy/muestras2.php>

utilizando validación cruzada dejando uno afuera (*leave-one-out cross-validation*).

CHASEDB1<sup>4</sup> contiene las imágenes de cada ojo de 14 niños, sumando un total de 28 imágenes. Las fotografías fueron capturadas con un ángulo de apertura del FOV de 30°, utilizando 8 bits por canal de color a 1280 × 960 píxeles de resolución, todas centradas en el disco óptico. Al igual que en STARE y DRIVE, se proveen dos segmentaciones por imagen. Las máscaras del FOV fueron obtenidas utilizando un enfoque inspirado en el propuesto en [20], dado que originalmente las mismas no están provistas en el conjunto de datos. Las 28 imágenes están divididas en entrenamiento y test, con 8 y 20 imágenes cada conjunto, respectivamente [64]. Las primeras 20 imágenes son utilizadas para test, y las últimas 8 para entrenamiento.

HRF<sup>5</sup> contiene 15 imágenes de pacientes sanos, 15 de pacientes con retinopatía diabética y 15 imágenes de pacientes glaucomatosos. Las imágenes se obtuvieron con una apertura de FOV de 60° y una resolución en píxeles de 3304 × 2336. Solamente una segmentación de referencia es provista para cada imagen, generada por un grupo de expertos. De acuerdo al relevamiento de la literatura realizado, el conjunto no se ha utilizado aún para evaluar algoritmos de segmentación supervisada, con lo cual se decidió en este caso construir un conjunto de entrenamiento con las primeras 5 imágenes de cada subconjunto, y testear el método entrenado sobre las imágenes restantes. Para reducir el costo computacional de los experimentos, las imágenes y sus etiquetas fueron redimensionadas a la mitad de la resolución original, y los resultados fueron posteriormente reescalados a la dimensión original para poder compararse con las segmentaciones manuales originales.

### 3.3.2. Métricas de evaluación

Los resultados son analizados cuantitativamente mediante la comparación de las segmentaciones obtenidas con respecto a los etiquetados manuales provistos en cada conjunto de datos. Siete métricas diferentes se obtuvieron para cada conjunto, todas ellas en términos del número de verdaderos positivos (*true positives*,  $TP$ ), verdaderos negativos (*true negatives*,  $TN$ ), falsos positivos (*false positives*,  $FP$ ) y falsos negativos (*false negatives*,  $FN$ ), considerando únicamente los píxeles dentro de la máscara

<sup>4</sup><http://blogs.kingston.ac.uk/retinal/chasedb1/>

<sup>5</sup><https://www5.cs.fau.de/research/data/fundus-images/>

del FOV:

$$Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}, \quad Pr = \frac{TP}{TP + FP}$$

$$F1 = \frac{2 \cdot Pr \cdot Re}{Pr + Re}, \quad G = \sqrt{Se \times Sp}$$

$$MCC = \frac{TP/N - S \times P}{\sqrt{P \times S \times (1 - S) \times (1 - P)}}$$

donde  $N = TP + TN + FP + FN$  es el número total de píxeles en la imagen,  $S = (TP + FN)/N$  y  $P = (TP + FP)/N$ . La sensibilidad (o *sensitivity*,  $Se$ , también conocida como exhaustividad o *recall*,  $Re$ ) mide la capacidad del método para detectar apropiadamente los vasos sanguíneos, mientras que la especificidad (o *specificity*,  $Sp$ ) es un indicador de la capacidad del mismo para distinguir cualquier otra estructura no vascular. La  $Sp$  es afectada cuando existen clases desbalanceadas, debido a la baja influencia del término correspondiente a los falsos positivos en el denominador de la fracción. Por el contrario, la  $Se$  no presenta este inconveniente, dado que solamente involucra píxeles correspondientes a la clase vaso. Sin embargo, aunque se desea conseguir un valor de  $Se$ , también debe analizarse en combinación con la  $Sp$ , teniendo en cuenta que la  $Se$  puede ser trivialmente maximizada etiquetando todos los píxeles de la imagen como pertenecientes a vasos sanguíneos. La precisión (o *precision*,  $Pr$ ) cuantifica la tasa de píxeles clasificados como vasos que fueron correctamente identificados. La exactitud (o *accuracy*,  $Acc$ ) no se incluye dado que es muy sensible a distribuciones desbalanceadas en las que el número de píxeles pertenecientes a la clase positiva es muy inferior a la negativa [114]. Se incluyen, por el contrario, otras métricas globales más robustas, tales como el coeficiente de correlación de Matthews (o *Matthews Correlation Coefficient*, MCC), el Valor-F1 (o *F1-score*, F1) y la media G (o *G-mean*, G), que no son afectadas por el desbalance en el conjunto de datos.

### 3.3.3. Configuración del método

Los parámetros para el cálculo de los descriptores unarios fueron inicializados de acuerdo a los artículos en los que fueron introducidos por primera vez, todos ellos basados en el conjunto de entrenamiento de DRIVE [138, 188]. Las respuestas

a la transformada wavelet 2D de Gabor fueron obtenidas para valores de escala  $a = 2, 3, 4, 5$ , y las respuestas a los detectores de líneas se analizaron para  $l = 1$  hasta 15, con incrementos  $l_0 = 2$ . Debido a que Zana y Klein utilizaron datos diferentes para estimar la longitud del elemento estructurante, se seleccionó un valor  $l = 9$ , que es consistente con el calibre promedio de los vasos sanguíneos en el conjunto DRIVE, según se reporta en [10]. El tamaño de las ventanas utilizadas para preprocesar la imagen para obtener este indicador fue fijado en 35 píxeles, y el borde del FOV fue expandido en 50 píxeles. Para otros conjuntos diferentes a DRIVE se utilizó el factor de compensación  $\rho$  descrito en la Sección 3.2.4. En el caso del detector de líneas de Nguyen, el incremento  $l_0$  también fue multiplicado con el propósito de reducir la dimensionalidad del vector de características al evaluarse en imágenes de mayor resolución. Los ángulos fueron fijados en los valores reportados en las referencias originales [138, 188, 226].

Para estimar el parámetro  $C$  de la SOSVM, cada conjunto de entrenamiento se separó en dos nuevos subconjuntos, denominados *entrenamiento\** y *validación*, cada uno conteniendo el 70 % y el 30 % de las imágenes de entrenamiento, respectivamente [77]. El conjunto *entrenamiento\** se utilizó para entrenar el modelo, y el de *validación* se empleó para estimar el mejor valor del parámetro de regularización  $C$ . Así, siempre se realiza una etapa de ajuste del modelo en la que la SOSVM es entrenada para diferentes valores de  $C$ . El rendimiento de cada modelo entrenado se evalúa luego sobre el conjunto de validación. Se evalúan siempre valores de  $C \in \{(10^i)/c\}$ , con  $i \in \{-2, \dots, 0, \dots, 3\}$  y  $c$  igual al número total de píxeles en el FOV. La configuración seleccionada es aquella que maximiza el valor-F1 promedio en los datos de validación. Esa configuración del CRF aprendida es posteriormente evaluada sobre los datos de test una única vez. Esta organización experimental previene la utilización de los datos de test durante el ajuste de parámetro, permitiendo obtener un estimador no sesgado del error de test [77]. La estrategia de validación cruzada fue utilizada únicamente sobre STARE debido a su tamaño limitado y a la ausencia de una partición por defecto en datos de entrenamiento y test.

Para el ajuste de  $\theta_p$  se utilizó un enfoque similar: para un valor fijo de  $C$ , se exploraron valores de  $\theta_p$  de 1 a 15 con incrementos de 2, y se eligió aquel que maximizó el valor-F1 en el conjunto de validación. Esta búsqueda, sin embargo, se aplicó

Tabla 3.1: Evolución del valor-F1 durante el proceso de selección secuencial de indicadores en DRIVE.

Descriptor	Potenciales unarios			Potenciales <i>pairwise</i>	
	Iter. 1	Iter. 2	Iter. 3	Iter. 1	Iter. 2
Detector de líneas de Nguyen	0.6898	<b>0.7256</b>	-	0.7535	0.6985
Transformada wavelet 2D de Gabor	<b>0.6967</b>	-	-	0.7423	0.7437
Imagen realzada por Zana y Klein	0.6378	0.7043	0.7129	<b>0.7546</b>	-

únicamente sobre el conjunto de validación de DRIVE. Al momento de evaluar sobre los demás conjuntos de imágenes, se empleó  $\theta_p = 5$  multiplicado por el factor de compensación  $\rho$ .

Para identificar qué combinación de descriptores es la más adecuada para los potenciales unarios y *pairwise* se utilizó un algoritmo de selección secuencial [115], utilizando los conjuntos *entrenamiento\** y *validation* de DRIVE. La Tabla 3.1 ilustra el progreso del valor-F1 promedio obtenido en el conjunto de validación para cada configuración de descriptores en cada iteración. Los valores resaltados indican que el conjunto de indicadores fue elegido en esa iteración. Se observó, en general, que la transformada wavelet 2D de Gabor contribuye a detectar las estructuras más anchas pero que al mismo tiempo introduce un gran número de falsos positivos. Al incorporar las respuestas al detector de líneas de Nguyen *et al.*, la cantidad de falsos positivos se reduce sensiblemente y los vasos más delgados son detectados con más facilidad. En cuanto a los potenciales *pairwise*, el indicador que demostró un mejor rendimiento es la imagen realzada con la estrategia de Zana y Klein. Esta configuración fue utilizada en todos los demás experimentos realizados posteriormente.

Una estrategia posible para la estimación del parámetro de escala de una función de base radial es tomar la mediana de todas las distancias de a pares sobre una muestra aleatoria de píxeles [178]. Dado que parte de cada kernel *pairwise* consiste en una función de base radial (Ecuación 3.6), este enfoque fue aplicado para la estimación del parámetro  $\theta_{(m)}$ . Este estimador es robusto dado que posee una varianza baja cuando se lo calcula sobre diferentes muestras aleatorias [178]. Sin embargo, pequeñas perturbaciones en  $\theta_{(m)}$  pueden afectar los resultados debido a la exponenciación en el término *pairwise* y al número de interacciones tenidas en cuenta por el modelo totalmente conectado. En consecuencia, se estimó  $\theta_{(m)}$  como la mediana de las medianas obtenidas para 50 muestras aleatorias diferentes de pares de la for-

ma  $(f_i^{(m)}, f_j^{(m)})$ , extraídas de los conjuntos de entrenamiento de cada conjunto de imágenes.

## 3.4. Experimentos y resultados

En esta sección se presentan los resultados obtenidos en los experimentos realizados. El prototipo del método propuesto se implementó utilizando MATLAB R2013a, con funciones MEX para crear interfaces con las implementaciones C++ del LNB-CRF y el FC-CRF. En la Sección 3.4.1 se resumen y analizan los resultados, mientras que en la Sección 3.4.2 se incluye una comparación detallada con respecto a otros métodos publicados.

### 3.4.1. Resultados

La Tabla 3.2 presenta la evaluación cuantitativa de los resultados obtenidos en los experimentos realizados empleando los potenciales unarios y el modelo propuesto en este trabajo. Las segmentaciones binarias se obtuvieron minimizando las energías correspondientes utilizando la estrategia basada en la aproximación del campo promedio propuesta en [101].

Los resultados obtenidos con el modelo LNB-CRF basado en vecindarios locales no se incluyen en la tabla dado que son exactamente iguales a los obtenidos haciendo uso únicamente los potenciales unarios. Esto es consecuencia de que, bajo esta configuración, la SOSVM asigna un valor extremadamente pequeño a  $\mathbf{w}_p$ , que es el parámetro que pesa los potenciales locales, y al mismo tiempo no modifica los demás pesos, que están asociados a los potenciales unarios. Esto permite concluir que la conectividad de grilla no provee información valiosa en el contexto de la detección de estructuras elongadas tales como los vasos sanguíneos, en línea con la hipótesis inicial presentada en este capítulo. Por el contrario, la SOSVM asigna valores no nulos al parámetro  $\mathbf{w}_p$  cuando se la utiliza para entrenar el FC-CRF, y también modifica los pesos asociados a los descriptores unarios y el término de sesgo, lo que implica que los potenciales *pairwise* influyen sobre los demás parámetros y contribuyen sustancialmente en la función de predicción. Para evaluar el nivel de significancia estadística de tal influencia se realizaron diferentes tests de hipótesis (en particular,

Tabla 3.2: Evaluación cuantitativa de los resultados obtenidos en DRIVE, STARE, CHASEDB1 y HRF, utilizando únicamente los potenciales unarios (UP) o el CRF totalmente conectado (FC-CRF).

Base de datos	Método	Se	Sp	Pr	F1	G	MCC
DRIVE	UP	0.7079	<b>0.9802</b>	<b>0.8394</b>	0.7661	0.8324	0.7401
	FC-CRF	<b>0.7897</b>	0.9684	0.7854	<b>0.7857</b>	<b>0.8741</b>	<b>0.7556</b>
STARE	UP	<b>0.7692</b>	0.9675	0.7445	0.7517	0.8618	0.7252
	FC-CRF	0.7680	<b>0.9738</b>	<b>0.7740</b>	<b>0.7644</b>	<b>0.8628</b>	<b>0.7417</b>
CHASEDB1	UP	0.7110	0.9707	0.7386	0.7209	0.8304	0.6919
	FC-CRF	<b>0.7277</b>	<b>0.9712</b>	<b>0.7438</b>	<b>0.7332</b>	<b>0.8403</b>	<b>0.7046</b>
HRF	UP	0.7315	<b>0.9680</b>	<b>0.7012</b>	0.7127	0.8413	0.6851
	FC-CRF	<b>0.7874</b>	0.9584	0.6630	<b>0.7158</b>	<b>0.8686</b>	<b>0.6897</b>

se utilizó el test del signo de Wilcoxon de una cola a derecha) sobre los valores de calidad obtenidos utilizando solamente los potenciales unarios y empleado el FC-CRF. No se realizaron tests de hipótesis sobre los resultados en STARE debido a que las segmentaciones en este conjunto fueron obtenidos por validación cruzada. Teniendo en cuenta que el parámetro  $C$  fue ajustado de acuerdo a los conjuntos de validación muestreados para cada conjunto de entrenamiento, no es posible asumir que todos los resultados hayan sido obtenidos utilizando la misma configuración. Finalmente, además de los resultados cuantitativos se proveen ejemplos de las segmentaciones obtenidas en cada conjunto de imágenes, a los efectos de analizar cualitativamente los cambios introducidos por los potenciales *pairwise*.

En alguno de los conjuntos de datos, el FC-CRF contribuye a una mejora estadísticamente significativa de los resultados al evaluarlos en términos del valor-F1 (DRIVE:  $p \approx 4 \times 10^{-5}$ ; CHASEDB1:  $p \approx 1 \times 10^{-3}$ ), la media-G (DRIVE:  $p \approx 4 \times 10^{-5}$ ; CHASEDB1:  $p \approx 8 \times 10^{-5}$ ; HRF:  $p \approx 9 \times 10^{-7}$ ) y el MCC (DRIVE:  $p \approx 4 \times 10^{-5}$ ; CHASEDB1:  $p \approx 4 \times 10^{-3}$ ; HRF:  $p \approx 2 \times 10^{-2}$ ). El valor-F1 y el MCC promedio son mejorados en STARE y HRF, como se indica en la Tabla 3.2. En el caso de HRF, la mejora en el valor-F1 es inferior a la obtenida en los demás conjuntos de datos. Por otro lado, la media-G también se ve incrementada en STARE.

Al evaluar el algoritmo propuesto en DRIVE, se observa que los potenciales *pairwise* mejoran la  $Se$  media ( $p \approx 4 \times 10^{-5}$ ) y apenas reducen la  $Sp$  promedio ( $p \approx 4 \times 10^{-5}$ ). Eso se debe a que el FC-CRF introduce una cierta fracción de falsos

positivos, como se observa en la reducción de la  $Pr$  promedio ( $p \approx 4 \times 10^{-5}$ ). En todos los casos, sin embargo, la fracción de mejora en  $Se$  es mayor que la reducción observada en la  $Sp$  y la  $Pr$ , lo que se traduce en una mejora de la media-G, el valor-F1 y el MCC. La Figura 3.3 presenta algunas segmentaciones de ejemplo obtenidas en DRIVE. En ellas es posible notar que el modelo FC-CRF incorpora un gran número de vasos delgados y mejora significativamente la conectividad de la estructura vascular.

En STARE, la media-G es mejorada al utilizar el enfoque denso propuesto en este capítulo, aunque no de manera significativa. Al descomponer esta métrica en sus términos respectivos, es posible observar que la  $Se$  es apenas reducida en promedio, mientras que tanto la  $Sp$  como la  $Pr$  evidencian una sensible mejora. Esto se asocia a una reducción en el número de falsos positivos en la segmentación resultante. La Figura 3.5 presenta un ejemplo de una imagen patológica extrema en la que esta situación es claramente observable. El FC-CRF contribuye a reducir el número de falsos positivos en la hemorragia interna del disco óptico. Aún cuando el segundo observador experto identificó vasos en la región, el primer observador (que es considerado como referencia) no marcó nada allí, lo que afecta directamente a la  $Se$ . Otras estructuras delgadas también son integradas al árbol vascular cuando se utiliza el FC-CRF, y es posible observar también que los potenciales unarios sobreestiman el calibre de algunos de los vasos mayores.

En CHASEDB1, los resultados son mejorados por el FC-CRF en términos tanto de la  $Se$  ( $p \approx 7 \times 10^{-5}$ ) como de la  $Sp$ , aunque el crecimiento de esta última métrica no es significativo estadísticamente. En consecuencia, la media-G es también incrementada. La  $Pr$  promedio también es mejorada por el FC-CRF, lo que es explicado por una reducción en el número de falsos positivos. La Figura 3.4 ilustra este comportamiento para una imagen de ejemplo. También es posible observar que los potenciales unarios sobreestiman el calibre de los vasos más delgados, una característica que es sensiblemente mitigada al incorporar los potenciales *pairwise*.

En HRF se observa un comportamiento diferente (Figura 3.6), ya que la  $Se$  se incrementa significativamente ( $p \approx 9 \times 10^{-7}$ ) pero la  $Sp$  ( $p \approx 9 \times 10^{-7}$ ) es disminuida. La  $Pr$  también disminuye ( $p \approx 9 \times 10^{-7}$ ), lo que significa que el número de falsos positivos se incrementa. Cualitativamente, sin embargo, se observa que el FC-CRF

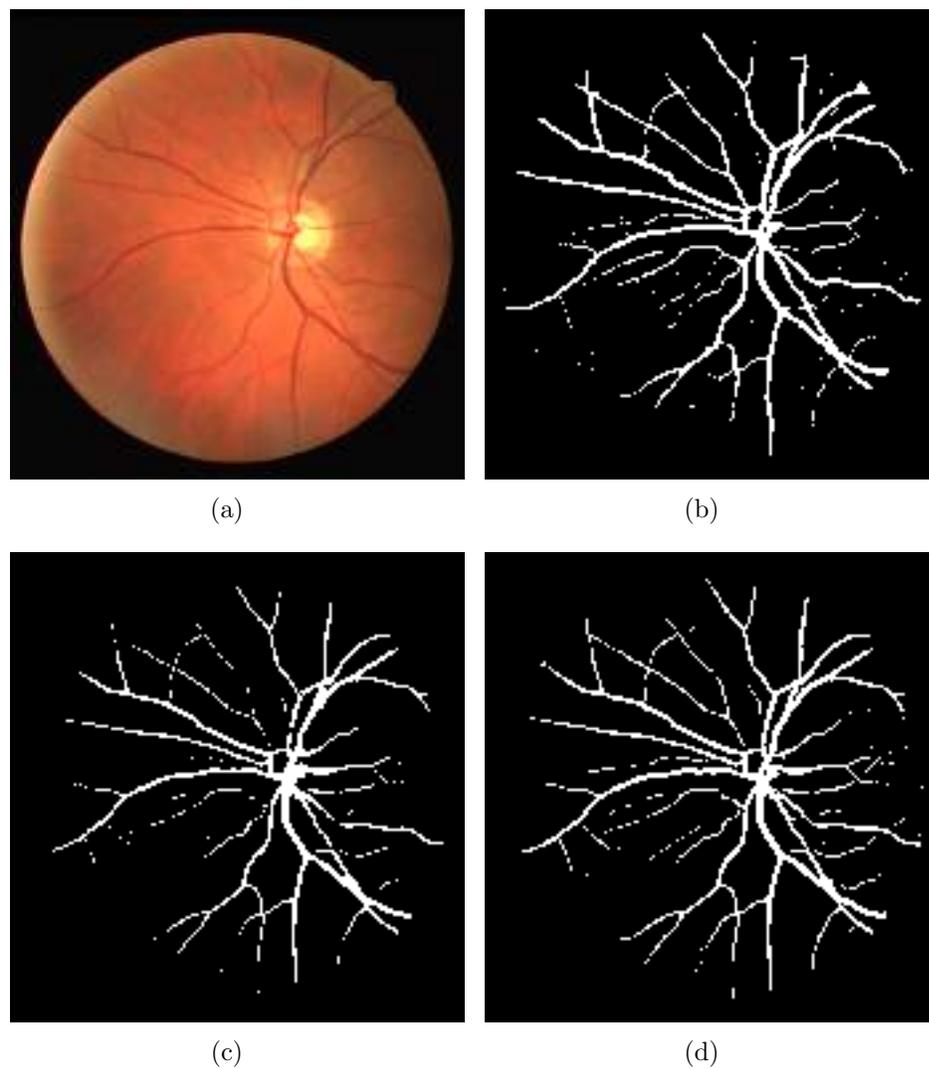


Figura 3.3: Resultados de segmentación obtenidos en DRIVE. (a) Imagen 04 de DRIVE. (b) Segmentación de referencia (*ground truth*). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF.

detecta un gran número de vasos delgados que eran ignorados al utilizar únicamente los potenciales unarios, como se ve en la Figura 3.7. Sin embargo, es posible observar que algunos de los capilares más delgados permanecen ignorados. Mejorar los descriptores utilizados puede ayudar a incorporar tales estructuras. A pesar de esto último, los potenciales *pairwise* demuestran su habilidad para unir detecciones ais-

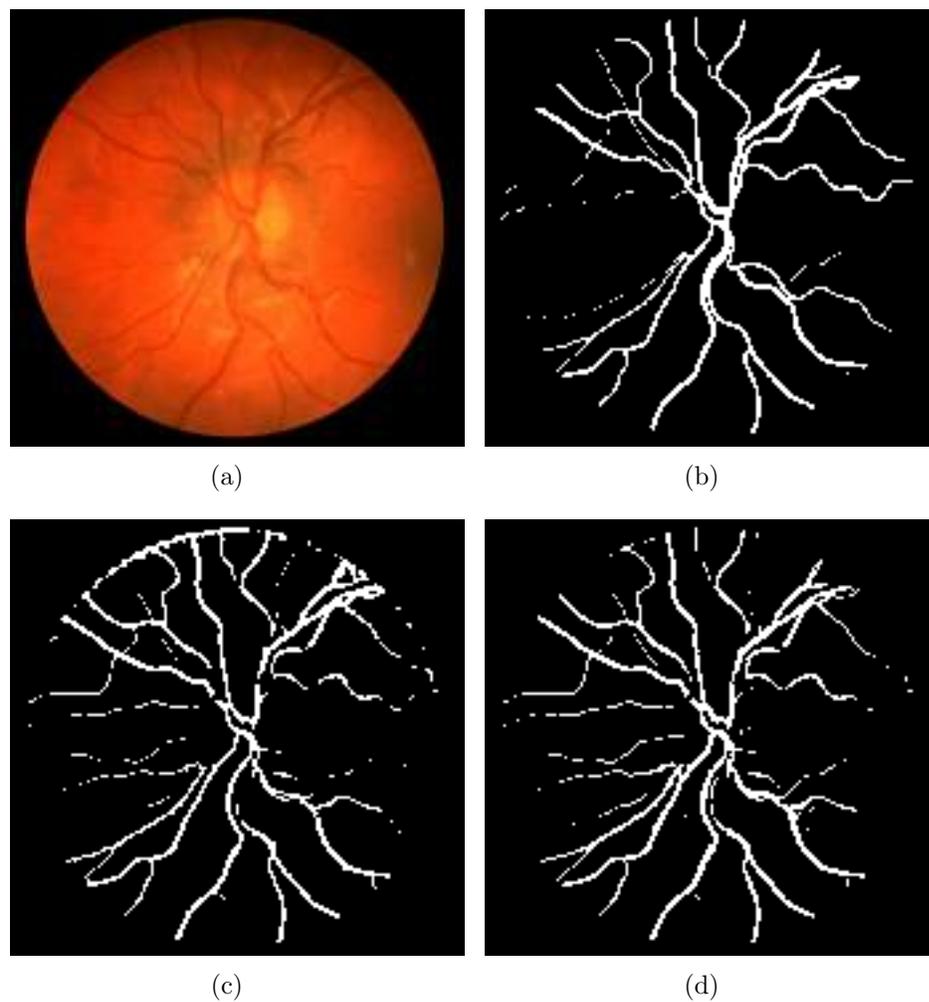


Figura 3.4: Resultados de segmentación obtenidos en CHASEDB1. (a) Image\_05L de CHASEDB1. (b) Segmentación de referencia (*ground truth*). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF.

ladas, lo que resulta en un árbol vascular mucho más conexo, y en un crecimiento en el valor de la media-G.

La Figura 3.8 presenta una comparación entre las curvas ROC obtenidas utilizando únicamente los potenciales unario y empleado el FC-CRF. Además, se incorporan los valores de  $Se$  y  $Sp$  obtenidos por el segundo experto en cada conjunto de datos (siempre que ambos valores estén disponibles). La curva obtenida en STARE no se

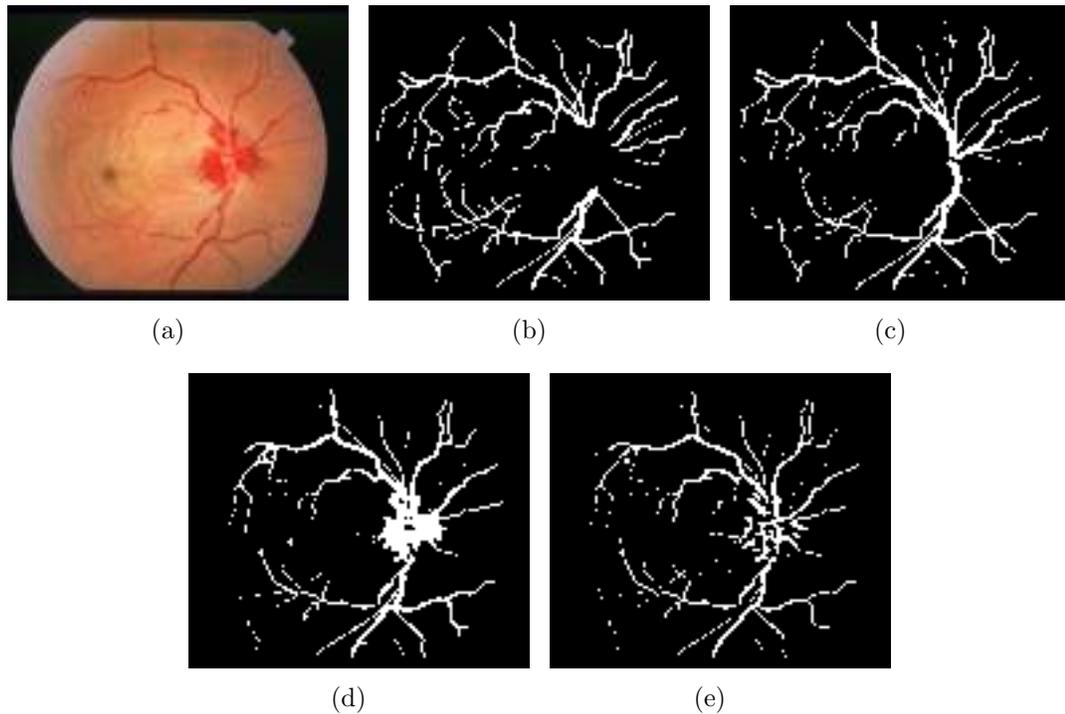


Figura 3.5: Resultados de segmentación obtenidos en un caso patológico serio de STARE. (a) Imagen im0005 de STARE. (b) Segmentación de referencia del primer observador experto (*ground truth*). (c) Segmentación del segundo observador experto. (d) Segmentación obtenida utilizando únicamente los potenciales unarios. (e) Segmentación obtenida utilizando el FC-CRF.

incluye debido a que los resultados en este conjunto de imágenes se obtuvieron por validación cruzada, y la segmentación de cada imagen fue hecha utilizando un modelo diferente. Los resultados en DRIVE muestran que el FC-CRF tiene un mejor rendimiento que los potenciales unarios, y también demuestran estar cuantitativamente cercanos al rendimiento de un segundo observador humano. Cuando se evalúa en HRF, sin embargo, y en línea con el análisis hecho anteriormente, los potenciales unarios resultan en una mejor curva ROC que el FC-CRF. Las áreas bajo cada una de las curvas, como puede visualizarse en la Figura 3.8, están en línea con estas conclusiones.

Finalmente, el costo computacional del prototipo implementado (no paralelizado ni a nivel de CPU ni de GPU) para la inferencia en el FC-CRF se evaluó en una CPU Intel(R) Xeon(R) CPU E5-2690 0 a 2.90GHz con 64 GB of RAM. A tal efecto, se

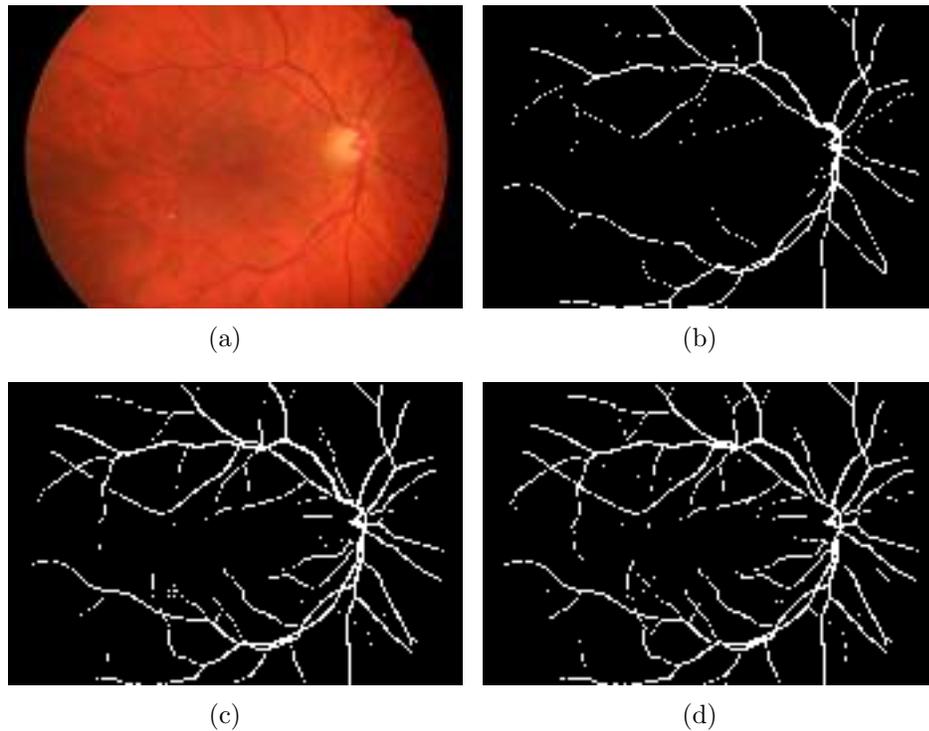


Figura 3.6: Resultados de segmentación obtenidos en HRF. (a) Imagen 11\_g. (b) Segmentación de referencia (*ground truth*). (c) Segmentación obtenida utilizando únicamente los potenciales unarios. (d) Segmentación obtenida utilizando el FC-CRF.

midió el tiempo promedio de aplicar el FC-CRF en cada conjunto de test. Como se ve en la Figura 3.9, aunque el costo computacional crece en relación con la resolución de las imágenes, el tiempo se mantiene suficientemente bajo como para que sea factible aplicar este método en un ambiente clínico.

### 3.4.2. Comparación con otros métodos

Se incluye una comparación de los resultados obtenidos utilizando el enfoque propuesto con respecto a los reportados por otros métodos del estado del arte, evaluados en DRIVE (Tabla 3.3), STARE (Tabla 3.4), CHASEDB1 y HRF (Tabla 3.5). Aunque el método propuesto es supervisado, también se incluyen en la comparación otros enfoques no supervisados. Aquellos algoritmos que obtuvieron las segmentacio-

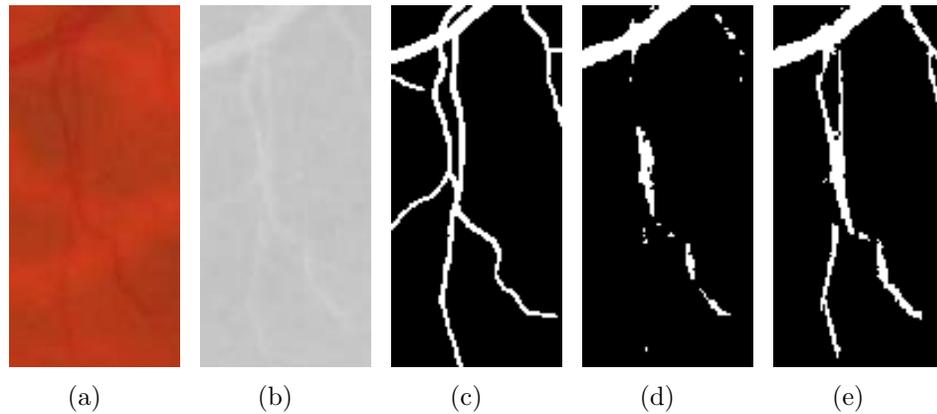


Figura 3.7: Detección de vasos delgados en condiciones de bajo contraste. (a) Detalle de la imagen 11\_g de HRF. (b) Imagen preprocesada. (c) Segmentación de referencia. (d) Segmentación obtenida utilizando únicamente los potenciales unarios. (e) Segmentación obtenida utilizando el FC-CRF.

nes binarias finales utilizando parámetros estimados utilizado datos de test no fueron incluidos en las tablas de DRIVE y STARE, dado que, bajo estas circunstancias, tales métodos pueden haber subestimado el verdadero error de test [74, Section 6].<sup>6</sup> Los resultados obtenidos de manera similar pero en CHASEDB1 se incluyeron en la Tabla 3.5 dado que son los únicos trabajos evaluados en estos conjuntos. Sin embargo, se los identificó con un asterisco para resaltar esta salvedad.

El método presentado reportó el valor-F1 y el MCC más altos en DRIVE al compararse con otras estrategias supervisadas y no supervisadas. La  $Se$  es también la más alta, para un valor de  $Sp$  relativamente aceptable. Como se mencionó anteriormente, la  $Sp$  cuantifica la capacidad del método de distinguir la clase no-vaso, y se ve por lo general afectada cuando las segmentaciones tienen un alto número de falsos positivos. Sin embargo, la  $Pr$  es mayor que la del método de Fathi *et al.* [56], que ha reportado una mayor  $Sp$  pero una  $Se$  inferior.

La comparación sobre STARE es difícil si se tiene que en cuenta que la mayoría

<sup>6</sup>En [20], los resultados reportados en el conjunto de test de DRIVE, en STARE y en CHASEDB1 corresponden a segmentaciones binarias obtenidas umbralando las respuestas al filtro B-COSFIRE. Sin embargo, el umbral fue elegido maximizando el MCC promedio en cada conjunto de test. De forma similar, en [64, 65, 121, 172] las probabilidades brindadas por diferentes clasificadores son umbraladas utilizando el parámetro que maximiza la  $Acc$  promedio en los datos de test (de DRIVE, STARE y CHASEDB1 en el caso de [64, 65, 172], y en DRIVE y STARE en [121]).

Tabla 3.3: Comparación de los valores promedios de  $Se$ ,  $Sp$ ,  $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en DRIVE.

Métodos	DRIVE					
	Se	Sp	Pr	F1	G	MCC
<b>FC-CRF</b>	<b>0.7897</b>	0.9684	0.7854	<b>0.7857</b>	<b>0.8741</b>	<b>0.7556</b>
Segundo observador experto	0.7760	0.9730	0.8066	0.7881	0.8689	0.7601
<b>Supervisados</b>						
Cheng <i>et al.</i> [36]	0.7252	0.9798	-	-	0.8429	-
Dai <i>et al.</i> [46]	0.7359	0.9720	-	-	0.8458	-
Niemeijer <i>et al.</i> [141]	0.6793	0.9725	-	-	0.8128	-
Lupascu <i>et al.</i> [117]	0.6728	<b>0.9874</b>	-	-	0.8151	-
Orlando and Blaschko [150]	0.7850	0.9670	0.7770	0.7810	0.8713	0.7482
Soares <i>et al.</i> [188]	0.7283	0.9788	-	-	0.8443	-
Xu and Luo [220]	0.7760	-	-	-	-	-
You <i>et al.</i> [225]	0.7410	0.9751	-	-	0.8500	-
Vega <i>et al.</i> [201]	0.7444	0.9600	-	0.6884	0.8454	0.6617
<b>No supervisados</b>						
Amin <i>et al.</i> [12]	0.6608	0.9799	-	-	0.8047	-
Al-Diri <i>et al.</i> [8]	0.7282	0.9551	-	-	0.8340	-
Bankhead <i>et al.</i> [21]	0.7027	0.9717	-	-	0.8263	-
Budai <i>et al.</i> [30]	0.6440	0.9870	-	-	0.7973	-
Chakraborti <i>et al.</i> [31]	0.7205	0.9579	-	-	0.8308	-
Espona <i>et al.</i> [53]	0.6634	0.9682	-	-	0.8014	-
Espona <i>et al.</i> [53]	0.7436	0.9615	-	-	0.8456	-
Fathi and Naghsh-Nilchi [56]	0.7768	0.9759	0.7559	0.7669	0.8707	-
Fraz <i>et al.</i> [61]	0.7152	0.9768	<b>0.8205</b>	0.7642	0.8358	0.7333
Fraz <i>et al.</i> [62]	0.7302	0.9742	0.8112	0.7686	0.8434	0.7359
Martínez <i>et al.</i> [123]	0.7246	0.9655	-	-	0.8364	-
Miri <i>et al.</i> [128]	0.7352	0.9795	-	-	0.8486	-
Odstrcilik <i>et al.</i> [145]	0.7060	0.9693	-	-	0.8272	-
Palomera <i>et al.</i> [156]	0.6440	0.9670	-	-	0.7891	-
Roychowdhury <i>et al.</i> [173]	0.7390	0.9780	-	-	0.8501	-
Vlachos and Dermatas [204]	0.7468	0.9551	-	-	0.8446	-
Wang <i>et al.</i> [210]	0.7520	0.9800	-	-	0.8585	-
Yin <i>et al.</i> [223]	0.6522	0.9710	-	-	0.7958	-
Zhang <i>et al.</i> [227]	0.7120	0.9724	-	-	0.8321	-
Zhao <i>et al.</i> [228]	0.7420	0.9820	-	-	0.8536	-

Tabla 3.4: Comparación de los valores promedios de  $Se$ ,  $Sp$ ,  $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en STARE.

Métodos	STARE					
	Se	Sp	Pr	F1	G	MCC
<b>FC-CRF</b>	0.7680	0.9738	<b>0.7740</b>	<b>0.7644</b>	0.8628	<b>0.7417</b>
Segundo observador experto	0.8951	0.9387	0.6424	0.7401	0.9166	0.7225
<b>Supervisado</b>						
Cheng <i>et al.</i> [36]	0.7813	<b>0.9843</b>	-	-	0.8769	-
Dai <i>et al.</i> [46]	0.7769	0.9550	-	-	0.8614	-
Soares <i>et al.</i> [188]	0.7200	0.9750	-	-	0.8379	-
You <i>et al.</i> [225]	0.7260	0.9751	-	-	0.8414	-
Vega <i>et al.</i> [201]	0.7019	0.9671	-	0.6082	0.8239	0.5927
<b>No supervisado</b>						
Al-Diri <i>et al.</i> [8]	0.7521	0.9681	-	-	0.8533	-
Budai <i>et al.</i> [30]	0.5800	0.9820	-	-	0.7547	-
Chakraborti <i>et al.</i> [31]	0.6786	0.9586	-	-	0.8065	-
Fathi and Naghsh-Nilchi [56]	<b>0.8061</b>	0.9717	0.7027	0.7509	<b>0.8850</b>	-
Fraz <i>et al.</i> [61]	0.7409	0.9665	0.7363	0.7386	0.8462	0.7003
Fraz <i>et al.</i> [62]	0.7318	0.9660	0.7294	0.7306	0.8408	0.6908
Martínez <i>et al.</i> [123]	0.7506	0.9569	-	-	0.8475	-
Palomera <i>et al.</i> [156]	0.7790	0.9409	-	-	0.8561	-
Odstrcilik <i>et al.</i> [145]	0.7847	0.9512	-	-	0.8639	-
Roychowdhury <i>et al.</i> [173]	0.7320	0.9840	-	-	0.8487	-
Wang <i>et al.</i> [210]	0.7800	0.9780	-	-	0.8734	-
Yin <i>et al.</i> [223]	0.7248	0.9666	-	-	0.8370	-
Zhang <i>et al.</i> [227]	0.7177	0.9753	-	-	0.8366	-
Zhao <i>et al.</i> [228]	0.7800	<b>0.9780</b>	-	-	0.8734	-

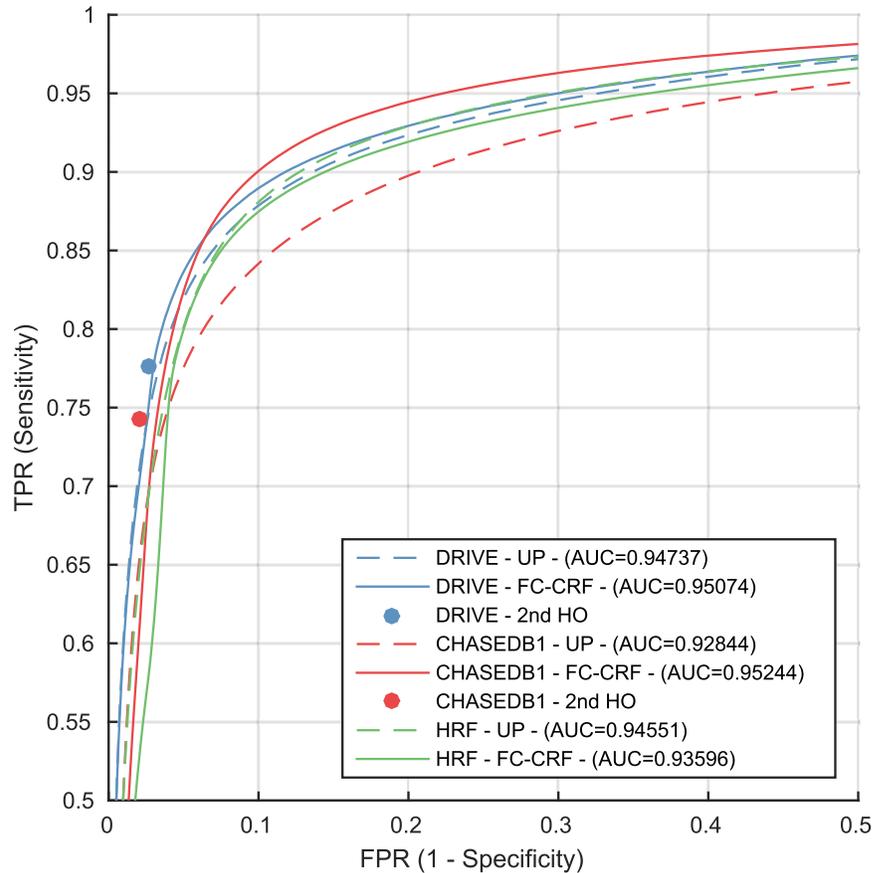


Figura 3.8: Curvas ROC en DRIVE, CHASEDB1 y HRF, utilizando únicamente los potenciales unarios (UP, línea punteada) o el FC-CRF (línea sólida), y rendimiento del segundo usuario experto (HO).

de los enfoques existentes en la literatura realizan sus análisis utilizando sus propias estrategias de entrenamiento y test.<sup>7</sup> Debe tenerse en cuenta, entonces, que los métodos supervisados listados en la comparación presentada en la Tabla 3.4 fueron entrenados de forma miscelanea. De la evaluación en este conjunto se desprende que el método propuesto en este capítulo reporta el valor-F1 promedio y el MCC medio más alto, lo que indica un mejor rendimiento global del algoritmo. Además, el FC-CRF supera otras estrategias al evaluarse en términos de la  $Pr$ , lo que implica que

<sup>7</sup>Algunos métodos fueron entrenados en la primera mitad de STARE y testeados en la segunda [12] o incluso sobre el conjunto completo [20], o fueron entrenados en una muestra aleatoria de píxeles extraídos de STARE [188], etc.

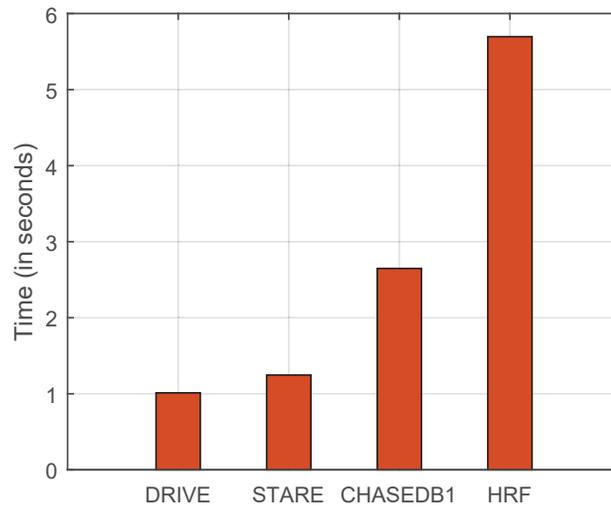


Figura 3.9: Costo computacional de la inferencia en el FC-CRF en todos los conjuntos de datos utilizados para la evaluación.

Tabla 3.5: Comparación de los valores promedios de  $Se$ ,  $Sp$ ,  $Pr$ , valor-F1, media-G y MCC del método propuesto con respecto a otros algoritmos de segmentación de vasos sanguíneos existentes, y al segundo observador experto, al evaluar en CHASEDB1 y HRF.

CHASEDB1	Se	Sp	Pr	F1	G	MCC
<b>FC-CRF</b>	0.7277	0.9712	0.7438	0.7332	0.8407	0.7046
Segundo observador experto	0.7425	0.9793	0.8090	0.7686	0.8527	0.7475
Fraz <i>et al.</i> [65]*	0.7259	0.9770	0.7732	0.7488	0.8421	-
HRF	Se	Sp	Pr	F1	G	MCC
<b>FC-CRF</b>	<b>0.7874</b>	0.9584	0.6630	0.7158	<b>0.8687</b>	0.6897
Odstrcilik <i>et al.</i> [145]	0.7794	<b>0.9650</b>	<b>0.6950</b>	<b>0.7324</b>	0.8672	<b>0.7065</b>

el número de falsos positivos es inferior al introducido por otros algoritmos.

Al evaluar sobre CHASEDB1, es posible observar que el FC-CRF alcanza una mejor  $Se$  con respecto a otras estrategias. El valor-F1 es superado por Fraz *et al.* [65], aunque los resultados reportados en dicho trabajo fueron obtenidos utilizando parámetros estimados utilizando las etiquetas del conjunto de test, y los parámetros de los descriptores que empleó fueron ajustados para este conjunto de imágenes en particular. Por el contrario, el método basado en FC-CRF fue entrenado utilizando parámetros adaptados mediante el factor de compensación.

De acuerdo al relevamiento realizado, únicamente métodos supervisados fueron testeados sobre HRF [30, 145] al momento de realizar este trabajo. Los resultados obtenidos en [145] sobre los datos de test fueron calculados a partir de las segmentaciones binarias provistas por los autores. En general, es posible ver que el FC-CRF brinda mayores valores de  $Se$  que el método propuesto en [145], aunque para menores  $Sp$  y  $Pr$ . Esto permite inducir que el FC-CRF obtiene un mayor número de falsos positivos que las otras estrategias. Sin embargo, las métricas globales más representativas (el valor-F1 y el MCC) son competitivas.

### 3.5. Discusión

La combinación del CRF totalmente conectado y la estrategia de aprendizaje propuesta demostraron, en la validación experimental, explotar más eficientemente las interacciones entre los descriptores de los píxeles que el enfoque basado en vecindarios locales. Este último, en particular, no fue capaz de mejorar los resultados con respecto a la versión basada únicamente en los potenciales unarios. Esto se debe a que la SOSVM asignó un peso nulo al término *pairwise*. Los test de hipótesis realizados sobre los resultados obtenidos para el FC-CRF y los potenciales unarios en diferentes conjuntos de imágenes (detallados en la Sección 3.4.1) demostraron que los potenciales densos empleados en este trabajo introducen mejoras estadísticamente significativas en numerosas métricas. Las curvas ROC presentadas en la Figura 3.8, por otro lado, permiten concluir que el FC-CRF alcanza resultados que son cercanos cuantitativamente a los de un segundo observador experto. Tales propiedades son consecuencia de la contribución de los potenciales *pairwise* de alto orden, que son capaces de reconstruir mejor los vasos en condiciones de bajo contraste (Figura 3.7) con una mínima sobrecarga de tiempo computacional (Figura 3.9). Aunque el modelo no incluye de manera explícita un término que aprenda características direccionales, la combinación de los términos del kernel *pairwise* (Ecuación (3.8)) que analizan la disimilaridad entre los descriptores y la ubicación de los píxeles proveen una forma de penalizar interacciones entre píxeles muy diferentes o muy lejanos, respectivamente. Así, si los descriptores utilizados en los potenciales *pairwise* son lo suficientemente robustos, entonces el modelo asignará baja energía a los etiquetas de estructuras fi-

lamentosas, y penalizará otras estructuras no elongadas. Por otro lado, cabe señalar que los descriptores analizados en este trabajo alcanzaron mejores resultados que utilizar únicamente los potenciales unarios, como se observa en la Tabla 3.1. Por el contrario, el modelo basado en vecindarios locales no fue capaz de tomar ventaja de los descriptores utilizados en sus potenciales *pairwise*, como es evidenciado por la falta de mejora con respecto a los potenciales unarios. Basándose en esta propiedad, es posible concluir que los potenciales densos son capaces de caracterizar de mejor manera los vasos sanguíneos. Otras aplicaciones biológicas o médicas podrían beneficiarse del uso de este enfoque para segmentar otras estructuras elongadas y tubulares tales como capilares, neuronas o raíces de plantas. En lo que respecta a esta última aplicación, se ha presentado un método no supervisado basado en FC-CRFs, que fue capaz de alcanzar resultados prometedores en la segmentación de raíces de *Arabidopsis thaliana* [152].

La extensa comparación con respecto a otros métodos del estado del arte también demostró que el enfoque propuesto es capaz de obtener buenos resultados al evaluarse con numerosas métricas de calidad. Además, constituye una estrategia totalmente automática para segmentar que alcanza los mejores resultados de la literatura al compararse en función de métricas globales como el valor-F1, la media-G y el MCC. Esto se debe, en parte, a que los estudios anteriores se han enfocado exclusivamente en utilizar la precisión (Acc) como medida de rendimiento, siendo que la misma ignora el hecho de que el número de píxeles ocupado por los vasos sanguíneos constituye una fracción pequeña de los píxeles totales de las imágenes. Como resultado, los métodos alternativos presentan valores inferiores de valor-F1, media-G y MCC, que son particularmente importantes dado que reflejan una mejor estimación de la precisión para detectar los vasos, que es el objetivo principal de cualquier enfoque de segmentación del árbol vascular para análisis de fotografías de fondo de ojo.

Como en el caso de otras estrategias supervisadas—tales como los modelos de mezcla de gaussianas [188] o las SVMs [220]—el rendimiento del método se ve afectado por la habilidad general de los descriptores para caracterizar la vasculatura retiniana. En los experimentos realizados en este trabajo se utilizaron descriptores ampliamente difundidos en el estado del arte, dado que el propósito del mismo fue estimar la contribución del modelo totalmente conectado en la mejora de los resul-

tados originales. La mayoría de los descriptores presentados en la literatura fueron designados para trabajar sobre imágenes de baja resolución como las de DRIVE y STARE. Teniendo en cuenta que el diseño de estos descriptores involucra el ajuste de diferentes parámetros que pueden verse afectados por la resolución de las imágenes, es de esperar una reducción en su rendimiento al aplicarlos sobre imágenes de mayor tamaño que las utilizadas para su configuración inicial. Una alternativa para mitigar este efecto podría ser el ajuste de cada descriptor a cada resolución específica, aunque esta búsqueda es costosa en términos de tiempo computacional. Más recientemente, Vostatek *et al.* [206] propuso una metodología para ajustar automáticamente los parámetros de los descriptores a diferentes resoluciones de imagen, basándose en regresión lineal. Sin embargo, dicha estrategia aún requiere de optimizar configuraciones primero, de forma tal que la recta de estimación luego se ajuste de acuerdo a estas medidas. Por el contrario, en este trabajo se propone utilizar una técnica más simple, basada en aplicar un factor de compensación  $\rho$  a cada parámetro para escalarlo antes del cálculo del descriptor. Utilizando un enfoque básico como este, el método presentado en este trabajo es capaz de compensar parcialmente los cambios en la resolución de las imágenes, superando a otras estrategias de segmentación conocidas.

Aunque el algoritmo presentado alcanza en general un rendimiento más que aceptable, se han observado algunos errores producto del reflejo central brillante de las arterias principales en imágenes de alta resolución como las de CHASEDB1 y HRF. Esto puede deberse a la limitada capacidad de los descriptores utilizados tanto en los potenciales unarios como en los *pairwise* para lidiar con este tipo de estructura, que no fue tomada en cuenta en el diseño original de los mismos. Un mejor desarrollo de descriptores o la utilización de algoritmos de aprendizaje profundo permitiría dar con mejores caracterizaciones tanto en este como en otros escenarios igual de complejos, como es el caso de la presencia de cambios patológicos serios. En la actualidad, la mayoría de los indicadores trabajan bien sobre imágenes de baja resolución, y muchos menos trabajos han abordado la caracterización de vasos en imágenes más grandes. Esto último abre una línea de investigación más que prometedora, entendiendo que muchos de estos escenarios complejos son más evidentes en resoluciones altas. Por otro lado, la evaluación de otros tipos de potenciales de mayor orden, combinados con

este enfoque de aprendizaje automático, podrían potencialmente mejorar los resultados, dado que permitirían capturar otros tipos de interacciones entre los píxeles. Sin embargo, debe tenerse en cuenta que estos enfoques suelen involucrar un compromiso entre rendimiento y costo computacional.

### 3.6. Conclusiones

En este capítulo se presentó un algoritmo de segmentación basado en el aprendizaje discriminativo de un CRF totalmente conectado, con el propósito de detectar vasos sanguíneos en imágenes de fondo de ojo. Por medio de descriptores extraídos a partir de las imágenes y de potenciales *pairwise* totalmente conectados, este enfoque es capaz de reconstruir la vasculatura retiniana de manera más precisa que utilizando enfoques basados en vecindarios locales o únicamente los potenciales unarios. La efectividad de la estrategia presentada se evidencia en una mejora general en el valor-F1, la media-G y el coeficiente de correlación de Matthew (tres métricas cuantitativas que son ampliamente utilizadas en problemas en los que el número de muestras positivas y negativas están desbalanceadas) obtenida en una serie de bancos de imágenes estándar. Las curvas ROC también muestran que los resultados obtenidos con el FC-CRF son comparables con los obtenidos por un segundo observador experto. La habilidad de los potenciales densos para reconstruir estructuras elongadas puede beneficiar otras aplicaciones médicas y biológicas.

---

## Capítulo 4

# Transferencia de redes neuronales convolucionales para detección de glaucoma

La gran mayoría de los sistemas actuales para la detección automatizada de glaucoma en imágenes de fondo de ojo requiere la utilización de descriptores basados en segmentaciones. Estas características, sin embargo, suelen verse influidas por la calidad de los algoritmos de segmentación utilizados. Las redes neuronales convolucionales (conocidas por su sigla en inglés, CNNs, de *Convolutional Neural Networks*) constituyen una herramienta muy útil para la resolución de problemas de clasificación de imágenes, dado que son capaces de aprender características altamente discriminativas a partir de las intensidades crudas de los píxeles. Sin embargo, su aplicabilidad a problemas de análisis de imágenes médicas está limitada por la no disponibilidad de grandes conjuntos de datos anotados, que son necesarios para su entrenamiento. En este capítulo se analiza la viabilidad de utilizar CNNs para la detección automática de glaucoma, preentrenadas con datos no médicos. Se utilizan para ello dos redes diferentes (OverFeat y VGG-S), que son aplicadas sobre imágenes de fondo de ojo para generar vectores de descriptores. Se evalúa, además, la contribución a la mejora de la capacidad discriminativa de estos descriptores provista por diversas técnicas de preprocesamientos tales como *inpainting* de vasos sanguíneos, ecualización de histograma adaptativa con contraste limitado (*contrast-limited adap-*

tive histogram equalization, CLAHE) o recortes en torno a la cabeza del nervio óptico. Todas estas estrategias son combinadas con modelos de regresión logística regularizados utilizando las normas  $\ell_1$  y  $\ell_2$ . Los resultados obtenidos en el banco de imágenes DRISHTI-GS1, evaluados en términos del área bajo la curva ROC promedio, sugieren que este enfoque es viable y ofrecen evidencia significativa acerca de la importancia de elegir una estrategia de preprocesamiento adecuada para lograr transferir estas redes preentrenadas, sobre todo en situaciones en las que la cantidad de datos disponibles dificulta el reajuste de la red al problema en particular.

Este capítulo se organiza en la siguiente manera. La Sección 4.1 detalla las motivaciones de transferir redes neuronales para la detección de glaucoma, incluyendo una revisión del estado del arte actual para la realización de esta tarea. En la Sección 4.2 se abordan las particularidades de la estrategia desarrollada, mientras que la Sección 4.3 resumen las diferentes arquitecturas de CNNs empleadas y los datos utilizados en los experimentos. La Sección 4.4 presenta los resultados obtenidos. Finalmente, la Sección 4.5 concluye el artículo.

La implementación open source del enfoque aquí propuesto está disponible en <https://github.com/ignaciorlando/overfeat-glaucoma>. Por otra parte, el trabajo presentado en este capítulo fue oportunamente publicado en:

- J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Convolutional neural network transfer for automated glaucoma identification. En *12th International Symposium on Medical Information Processing and Analysis*, págs. 101600U–101600U. International Society for Optics and Photonics, 2017

## 4.1. Motivación

Las estrategias más recientes para la detección automática de glaucoma se basan en analizar propiedades del disco óptico [90, 98], de los vasos sanguíneos [97, 218], o la capa de fibras nerviosas [133, 146], para lo cual es esencial detectar y segmentar previamente estas regiones. Sin embargo, las características basadas en segmentaciones suelen verse significativamente afectadas por la precisión de los métodos de segmentación utilizados para recuperar estas regiones [75]. En consecuencia, existe un especial interés en desarrollar nuevas estrategias basadas en características globa-

les de las imágenes. Como se mencionó anteriormente en el Capítulo 2, tales enfoques están basados en transformaciones de las intensidades de las imágenes [27], análisis de textura [130], o combinaciones de descriptores derivados de las imágenes con información clínica y genética [221]. Una limitación importante de estas estrategias, sin embargo, es que utilizan descriptores diseñados de manera manual, caracterizadas por requerir el ajuste de un gran número de parámetros [32], o bien emplean datos que son muy difíciles de obtener en ciertos contextos clínicos (como es el caso de la información genética).

Las redes neuronales convolucionales (CNNs) son lo suficientemente robustas como para aprender características menos evidentes de las imágenes, usualmente ignoradas por los observadores pero relevantes para la realización de la tarea de clasificación. Esto permite evitar pasos intermedios tales como la segmentación de las estructuras afectadas por la enfermedad o el diseño y selección manual de los descriptores [107]. Sin embargo, esta ventaja es lograda al costo de entrenar las redes sobre grandes conjuntos de datos anotados [86]. Así, su aplicabilidad queda por lo general restringida a dominios en los que hay gran disponibilidad de datos etiquetados, algo normalmente difícil en la imagenología médica, donde la adquisición y etiquetado de las imágenes requiere dispositivos de captura costosos y expertos entrenados [22]. De acuerdo al relevamiento de la literatura realizados, las CNNs fueron utilizadas una única vez para detección de glaucoma, en el trabajo de Chen *et al.* [35]. El método requiere segmentar manualmente el nervio óptico, primero, y luego preprocesar la imagen resultante. Posteriormente, se entrena una CNN en una versión aumentada de un gran banco de imágenes privadas. Mediante este enfoque, los autores fueron capaces de aprender características relevantes del disco óptico, factibles de ser utilizadas para la identificación de pacientes con glaucoma.

Como se mencionó previamente, entrenar una CNN requiere enormes volúmenes de datos para lograr aprender los descriptores necesarios para la resolución del problema. Como una alternativa a esta costosa etapa de entrenamiento, diferentes autores han propuesto recientemente pre-entrenar redes sobre datos de otro contexto, diferentes a los requeridos por el problema puntual, y aplicarlas posteriormente a la tarea que se busca resolver [22, 39, 183, 199]. Puntualmente, este enfoque consiste en entrenar las redes utilizando grandes volúmenes de imágenes anotadas correspon-

dientes a un dominio diferente, y posteriormente transferirlas al problema puntual. Esta transferencia se basa en alimentar a la red con las nuevas imágenes, y recobrar los vectores de características que la CNN extrae a partir de ellas. Estos vectores se utilizan posteriormente para entrenar un clasificador que está exclusivamente dedicado a resolver la tarea deseada. Por medio de este enfoque es posible aplicar un único modelo a la resolución de diferentes problemas de visión computacional, partiendo de la hipótesis de que, si los datos utilizados para el entrenamiento fueron lo suficientemente abundantes y diversos, la red va a haber aprendido a distinguir características que son útiles en diferentes contextos [184]. Esto permite que el clasificador dedicado pueda entrenarse con un conjunto de datos mucho más pequeño que el necesario para entrenar la red desde cero. A pesar de que esta estrategia fue explorada con éxito para numerosas aplicaciones de visión computacional [183], sólo algunos estudios describen la aplicación de este concepto en problemas de clasificación de imágenes médicas [22, 39, 184, 199], y ninguno para el análisis de fotografías de fondo de ojo.

En este capítulo se analiza la viabilidad de realizar detección automática de glaucoma en imágenes de fondo de ojo utilizando características extraídas por CNNs pre-entrenadas sobre datos no médicos. Se explora, además, qué tanto contribuye en la capacidad discriminativa de los descriptores obtenidos la aplicación previa de diferentes niveles de recorte en torno al área del disco óptico, combinados con estrategias conocidas de preprocesamiento (tales como *inpainting* de vasos sanguíneos o mejora de contraste). Así, en lugar de utilizar descriptores diseñados de forma manual, se propone utilizar métodos de preprocesamiento más simples para transferir la red pre-entrenada a este otro problema de clasificación. Las características obtenidas son utilizadas para el entrenamiento de clasificadores por regresión logística regularizados con normas  $\ell_1$  y  $\ell_2$ , los cuales proveen un valor de verosimilitud que indica la presencia o ausencia de signos de glaucoma en cada imagen. Los resultados alcanzados en una base de datos de imágenes de fondo de ojo de acceso abierto permiten observar que este enfoque es capaz de discriminar sobresalientemente pacientes glaucomatosos de pacientes sanos.

## 4.2. Métodos

En este capítulo se explorará cómo transferir CNN pre-entrenadas sobre datos no médicos a la tarea de detectar glaucoma sobre fotografías de fondo de ojo, sin necesidad de reentrenar la red utilizando grandes cantidades de imágenes. En general, este problema es conocido como transferencia de aprendizaje [224], y puede resolverse utilizando dos estrategias posibles. Una de ellas consiste en alimentar con nuevas imágenes a la red pre-entrenada, extrayendo la salida de la primera capa totalmente conectada y utilizándola como vector de descriptores [183]. Posteriormente, este vector es utilizado para entrenar un nuevo clasificador dedicado explícitamente a la resolución de la nueva tarea. La segunda estrategia consiste no solamente en reemplazar y reentrenar al clasificador ubicado en la capa final de la CNN utilizando el nuevo conjunto de datos, si no también reajustar los pesos de la red pre-entrenada continuada el proceso de optimización de la función de pérdida. Dependiendo de la nueva tarea, es posible reajustar todas las capas de la red o mantener fijas las primeras, restringiendo entonces el ajuste a algunas de las capas de más alto nivel [184]. Esta estrategia se basa en el hecho de que los descriptores aprendidos por las primeras capas de una CNN usualmente son lo suficientemente generales como para aplicarse exitosamente en otros contextos (tales como detectores de borde, colores o texturas). Por otro lado, las capas ubicadas en el interior profundo de las redes se vuelven progresivamente más específicas respecto a los detalles de las clases contenidas en el conjunto de imágenes original. No obstante, el refinamiento de una CNN requiere contar con suficientes muestras para evitar el sobreajuste del modelo a los datos de entrenamiento. Esto vuelve al enfoque infactible para la detección de glaucoma, dado que las bases de datos disponibles en la actualidad cuentan con un número reducido de imágenes.

El enfoque alternativo basado en utilizar los descriptores de salida de la red pre-entrenada (conocidos como *CNN codes*) y en entrenar un clasificador separado basándose en ellos se ve usualmente afectado si las imágenes utilizadas para entrenar la red difieren significativamente de las del problema a resolver. Para reducir el efecto de este inconveniente, en este trabajo se propone analizar el uso de diferentes métodos de preprocesamiento antes de alimentar la red con las imágenes. Si la estrategia de preprocesamiento es capaz de resaltar las características de las imágenes que están

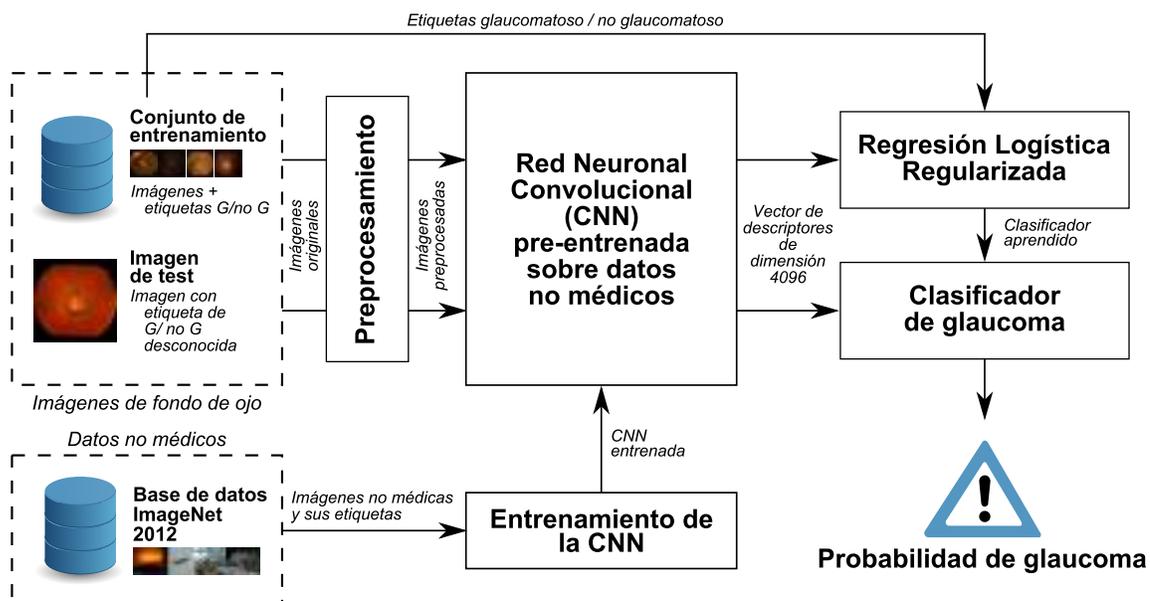


Figura 4.1: Representación esquemática del método para transferir CNNs pre-entrenadas sobre datos no médicos a la detección de glaucoma en fotografías de fondo de ojo.

más relacionadas con la presencia de glaucoma, es de esperar que esto contribuya en el cálculo de mejores descriptores. En la Figura 4.1 se presenta una representación esquemática del algoritmo propuesto. En las próximas subsecciones se describirán los diferentes métodos de preprocesamiento analizados y el clasificador utilizado.

### 4.2.1. Preprocesamiento de las imágenes

Las fotografías de fondo de ojo son preprocesadas utilizando técnicas conocidas antes de extraer los *CNN codes*. La Figura 4.2 muestra ejemplos de todas las operaciones consideradas. En particular, se evaluó la contribución de recortar la imagen en torno al disco óptico, aplicar estrategias de mejora de contraste (CLAHE), e incluir o remover los vasos sanguíneos de las mismas.

#### 4.2.1.1. Acercamientos al ONH

El glaucoma se manifiesta especialmente en la región del ONH [27]. Dado que la resolución y tamaño de las imágenes por lo general es reducida antes de alimentar a

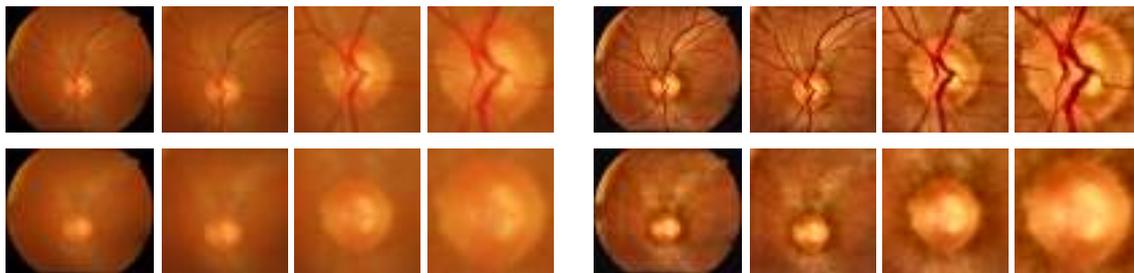


Figura 4.2: Estrategias de preprocesamiento evaluadas. Primer grupo: sin CLAHE. Segundo grupo: con CLAHE. De izquierda a derecha: imagen original, FOV recortado, area peripapilar (*peripapillary area*, PPA), y cabeza del nervio óptico (*optic nerve head*, ONH). Primera fila: imágenes originales. Segunda fila: imágenes luego de aplicar *inpainting* de los vasos sanguíneos.

la red, una hipótesis posible es que la CNN no va a ser capaz de capturar suficiente información acerca de la existencia de la enfermedad debido al tamaño resultante del ONH. O, por el contrario, podría suponerse que la CNN podrá caracterizar mejor al glaucoma si captura información de otras estructuras diferentes de la retina, como el árbol vascular. Para validar estas hipótesis se realizaron diversos experimentos considerando diferentes acercamientos y zooms en las fotografías de fondo de ojo originales, centradas todas en el ONH. En particular, se utilizaron 4 acercamientos diferentes (Figura 4.2): la imagen original tal cual es provista en el conjunto de imágenes de entrada, incluyendo las regiones vacías por fuera del campo visual (o *field of view*, FOV); un recorte cuadrado del area interna del FOV; el area peripapilar (o *peripapillary area*, PPA), que incluye al ONH y el área que lo rodea; y el ONH en sí mismo.

#### 4.2.1.2. Mejora de contraste mediante CLAHE

La ecualización adaptativa de histograma limitada por contraste (o *Contrast-Limited Adaptive Histogram Equalization*, CLAHE) es una operación de incremento de contraste que es ampliamente utilizada en análisis de imágenes de fondo de ojo dada su capacidad para mejorar el contraste y la iluminación de la retina [20] (Figura 4.2). Además, se ha demostrado previamente que esta técnica es capaz de mejorar la caracterización de las fibras nerviosas retinales [99, 147]. En los experimentos realizados se evaluó la contribución de utilizar o no CLAHE sobre las imágenes.

### 4.2.1.3. Sustracción de los vasos sanguíneos

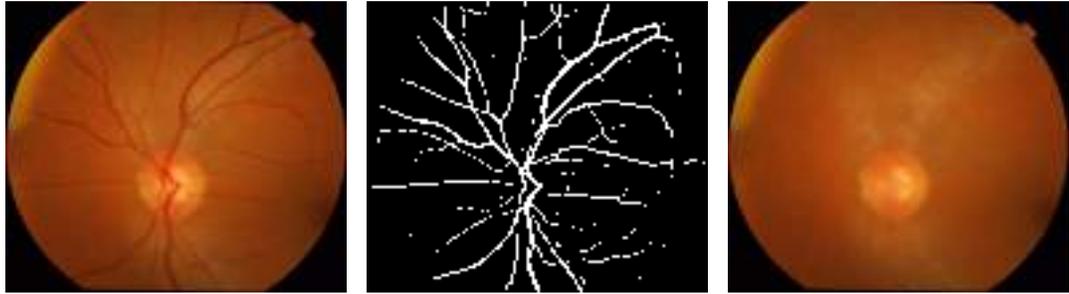
Aunque existen algunas publicaciones indicando cierta correlación entre el glaucoma y algunos cambios en las estructuras vasculares (ver Sección 2.4.2.2), otros autores sugieren removerlos antes de extraer descriptores [27, 75, 124]. Basándose en ambas asunciones, se realizaron experimentos alimentando la CNN con imágenes con y sin *inpainting* de los vasos sanguíneos.

El árbol vascular fue segmentado utilizando una versión ligeramente modificada del método presentado en el Capítulo 3. En particular, los descriptores a nivel de píxel necesarios para entrenar y ejecutar el algoritmo se extrajeron a partir de la banda verde de las imágenes RGB luego de aplicar CLAHE, en lugar de a partir de la banda verde como en el trabajo original. Además, los descriptores de entrada utilizados para los potenciales unarios fueron las respuestas a los detectores de línea de Nguyen *et al.* [138] y a los filtros B-COSFIRE [20], mientras que se mantuvo el uso del algoritmo para resaltar vasos de Zana y Klein [226] para los potenciales *pairwise*. Empíricamente se observó que esta combinación de características permite alcanzar mejores representaciones binarias del árbol vascular en el conjunto de imágenes empleado.

Los vasos fueron removidos de la imagen a color original aplicando un algoritmo de interpolación hacia adentro, en cada una de las bandas de color: las nuevas intensidades, dentro de las regiones ocupadas por los vasos, fueron suavemente interpoladas de afuera hacia adentro a partir de los píxeles ubicados en los bordes exteriores de la máscara binaria del árbol vascular. Como esta estrategia se ve por lo general afectada si la segmentación subestima el calibre real de los vasos (dado que el *inpainting* iterativo podría reintroducir las intensidades originales de los vasos en la imagen en lugar de removerlos), las máscaras vasculares fueron dilatadas morfológicamente utilizando discos de tamaño 3 píxeles. La Figura 4.3 ilustra una segmentación del árbol vascular y la imagen resultante tras removerla utilizando este algoritmo.

### 4.2.2. Regresión logística regularizada con normas $\ell_1$ y $\ell_2$

La detección de glaucoma puede modelarse como un problema de clasificación binaria que puede resolverse a través de un enfoque supervisado de aprendizaje au-



(a) Imagen DrishtiGS\_041. (b) Segmentación de los vasos sanguíneos (c) Imagen luego de aplicar *inpainting*

Figura 4.3: Sustracción de vasos.

tomático, en este caso regresión logística regularizada [77]. Sea  $S$  un conjunto de datos de entrenamiento compuesto por  $n$  instancias de entrenamiento de la forma  $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$ . Cada  $\mathbf{x}_i \in \mathbb{R}^d$  es un vector de descriptores  $d$ -dimensional obtenido a partir de la  $i$ -ésima imagen utilizando una CNN, y  $y_i \in \{-1, +1\}$  representa su correspondiente etiqueta de clase, siendo  $+1$  glaucomatoso y  $-1$  no-glaucomatoso. La regresión logística puede escribirse en términos de minimizar un riesgo regularizado con una pérdida logística. La función objetivo a minimizar está dada por:

$$\hat{\beta} = \arg \min_{\beta} \lambda \Omega(\beta) + \sum_{i=1}^n \log(1 + e^{-y_i \langle \beta, \mathbf{x}_i \rangle}) \quad (4.1)$$

donde  $\beta$  es un vector de coeficientes de la función lineal discriminante,  $\lambda \geq 0$  es un parámetro escalar que controla el grado de regularización provisto por el regularizador  $\Omega : \mathbb{R}^d \mapsto \mathbb{R}_+$ , y  $\langle \cdot, \cdot \rangle$  es el producto canónico interno en  $\mathbb{R}^d$ . Se utilizaron dos instancias de regularizadores conocidos para  $\Omega$ , la norma  $\ell_1 = \|\cdot\|_1$  y la norma  $\ell_2 = \|\cdot\|_2$ .  $\ell_1$  impone que la suma de los valores absolutos de los parámetros  $\beta$  se mantenga pequeña, lo que genera coeficientes ralos (o *sparse*) [135]. Esta propiedad permite utilizar este regularizador para realizar de manera simultanea tanto el aprendizaje del modelo como la selección automática de los descriptores más relevantes, cuando el tamaño  $n$  del conjunto de entrenamiento es pequeño y la dimensionalidad  $d$  de los vectores de características, como es el caso en este trabajo. Por el contrario, la regularización por norma  $\ell_2$  puede no ser efectiva cuando sólo algunos parámetros son relevantes [135]. Ambas normas fueron validadas experimentalmente para evaluar si



Figura 4.4: Ejemplo de una muestra de las imágenes naturales del conjunto ImageNet 2012, utilizadas para entrenar tanto OverFeat como VGG-S.

los vectores de características extraídos por la CNN son discriminativos en sí mismos, o sí solamente algunos de sus componentes son relevantes para la identificación de glaucoma. La verosimilitud o *likelihood* de que una imagen corresponda a un paciente con glaucoma o no es obtenida mediante la función discriminante  $\hat{y} = \langle \beta, \mathbf{x}_i \rangle$ . Para la minimización de la función objetivo se utilizó una implementación existente de regresión logística regularizada con norma de  $k$ -soportes [18, 185], asignando  $k = 1$  para la norma  $\ell_1$  y  $k = d$  para la norma  $\ell_2$ .

## 4.3. Materiales

### 4.3.1. Redes neuronales convolucionales

Se utilizaron dos CNNs diferentes en los experimentos realizados, OverFeat [181]<sup>1</sup> y VGG-S [33]<sup>2</sup>. Ambas redes fueron inicialmente entrenadas utilizando el conjunto de entrenamiento de *ImageNet 2012* (Figura 4.4), constituido por 1.2 millones de imágenes naturales, no médicas y a color, etiquetadas a partir de un conjunto de 1000 clases distintas. VGG-S, por otro lado, reportaba, al año 2015, la mayor performance en la competencia PASCAL-VOC.

La arquitectura de la red OverFeat está inspirada en el trabajo de Krizhevsky *et*

<sup>1</sup><http://cilvr.nyu.edu/doku.php?id=software:overfeat:start>

<sup>2</sup><http://www.vlfeat.org/matconvnet/pretrained/>

*al.* [102], con algunos cambios menores que son descritos por Sermanet *et al.* [181]. En particular, en este trabajo se utilizó la versión más precisa de su extractor de descriptores, compuesta por 6 capas convolucionales, con tamaños de filtros que van de  $7 \times 7$  a  $3 \times 3$  y conteniendo de 96 a 1024 *kernels* distintos. La red incorpora además técnicas clásicas como *max-pooling*, ReLUs (*rectification non-linearities*) y *dropout*, descritas en detalle en el trabajo de referencia. La arquitectura de VGG-S, introducida por Chatfield *et al.* [33], es relativamente similar. La misma consiste en 5 capas convolucionales en lugar de en 6, con tamaños de filtros similares a los utilizados por OverFeat pero utilizando 512 *kernels* en las últimas 3 capas convolucionales. Las últimas tres capas de ambas CNNs son totalmente conectadas, y son las responsables de asignar el nivel de verosimilitud de cada una de las clases. Teniendo en cuenta que el propósito de este trabajo es utilizar las redes como extractoras de descriptores, la salida de la primera capa totalmente conectada es empleada para extraer vectores de características de dimensionalidad  $d = 4096$  a partir de las imágenes de entrada. Cabe destacar que las imágenes de entrada deben ser reescaladas a  $231 \times 231$  o  $224 \times 224$  píxeles antes de alimentar OverFeat o VGG-S, respectivamente.

### 4.3.2. Bases de datos utilizadas

Los experimentos fueron llevados a cabo utilizando el conjunto de imágenes DRISHTI-GS1 [187]. El mismo consiste de 101 fotografías de fondo de ojo, tomadas con dilatación, centradas en el ONH, con  $30^\circ$  de FOV, a una resolución de  $2896 \times 1944$  píxeles y guardadas en formato PNG sin compresión. Todas las imágenes fueron originalmente capturadas en el Aravind Eye Hospital de la India. La selección de los pacientes con glaucoma fue realizada por médicos y basándose en estudios clínicos realizados a los pacientes durante el examen. Los sujetos seleccionados tenían edades de entre 40 y 80 años al momento de capturar las imágenes, y la proporción de hombres y mujeres fue aproximadamente la misma. Aquellos pacientes que concurren a la institución a realizarse estudios de refracción de rutina y no fueron diagnosticados como glaucomatosos fueron seleccionados para representar la clase sana. El diagnóstico asignado a cada paciente y, por consiguiente, a cada imagen (como normal o como glaucomatoso) fue obtenido a partir de la opinión de 4 expertos en glaucoma, que se basaron únicamente en las características de las imágenes.

La etiqueta final fue obtenida a partir de la opinión de la mayoría (esto es, 3 de los 4 expertos).

Para la remoción o *inpainting* de los vasos sanguíneos, los árboles vasculares fueron inicialmente segmentados utilizando un modelo entrenado sobre el conjunto de entrenamiento de DRIVE [141] y calibrado utilizando su conjunto correspondiente de test. Tras una etapa de ajuste del modelo realizada sobre este último banco de imágenes, los parámetros fueron fijados en  $C = 10^2/c$ , con  $c$  igual al número de píxeles de entrenamiento, y  $\theta_p = 5$ . Los parámetros de los extractores de descriptores utilizados para caracterizar los píxeles durante la segmentación fueron fijados en los valores reportados originalmente en sus referencias originales. Debido a que los descriptores fueron originalmente diseñados para trabajar en DRIVE y que dependen fuertemente de la resolución de las imágenes (que difiere significativamente entre los conjuntos de datos utilizados), las imágenes de DRISHTI-GS1 fueron reducidas en su resolución utilizando un factor de 2.96 antes de extraer los vasos. De esta forma, el calibre aproximado de los mismos logra asemejarse en mayor medida al de los vasos en las imágenes de DRIVE. Esto puede realizarse sin que afecte los resultados finales, ya que las imágenes deben reducirse aún más antes de introducirse a las CNNs para extraer sus características.

Cabe señalar que las imágenes de DRISHTI-GS1 no proveen máscaras que señalen el FOV, lo que resulta necesario para realizar la extracción de los vasos sanguíneos. Por tal motivo, se generaron las mismas utilizando un enfoque basado en umbralar el plano de luminosidad de las imágenes RGB tras convertirlas al espacio de color CIELab, utilizando un umbral igual a 0.01[20]. Si la máscara binaria resultante correspondía solo a valores no nulos (lo que implica que toda la imagen es considerada como FOV), se utilizó una estrategia alternativa consistente en estimar el FOV como la suma de todas las bandas de color, umbraladas a 150. Para suavizar los bordes y reducir cualquier ruido potencial que pueda generarse durante este proceso, la máscara fue preprocesada con un filtro de mediana, utilizando ventanas cuadradas de lado 5 píxeles, y sólo la mayor componente conectada fue preservada.

## 4.4. Resultados

Todas las estrategias de preprocesamiento presentadas en la Sección 4.2.1 fueron analizadas en combinación con la red OverFeat con el objetivo de determinar cuál de todas es la más apropiada para mejorar la capacidad discriminativa de los descriptores. Debido al tamaño reducido del conjunto de imágenes disponible, cada configuración fue evaluada unas 200 veces de la siguiente forma. Primero, el conjunto de 101 imágenes de DRISHTI-GS1 fue dividido aleatoriamente en un conjunto de entrenamiento (70 %) y otro de test (30 %), como suele sugerirse en la literatura [77]. Un 30 % de las imágenes de entrenamiento fue seleccionado aleatoriamente y separado como un conjunto de validación, que fue utilizado para determinar el valor óptimo del parámetro de regularización  $\lambda = 10^i$  del clasificador por regresión logística, con  $i \in \{-5, -4, \dots, 6\}$ . El area bajo la curva ROC (AUC) fue utilizado como un indicador cuantitativo de la performance del modelo durante el ajuste del modelo. Por otro lado, se emplearon dos técnicas diferentes para aumentar el tamaño de los conjuntos de entrenamiento utilizados: las imágenes fueron rotadas utilizando dos incrementos de ángulos distintos ( $90^\circ$  y  $45^\circ$ ) hasta alcanzar los  $360^\circ$ , y fueron espejadas horizontalmente, incrementando así el tamaño de los conjuntos originales en un factor de 8 y 16, respectivamente. El rendimiento de cada configuración fue cuantificado en términos del promedio de las areas bajo las curvas ROC [58]. Los resultados se presentan en la Figura 4.5 y son analizados en detalle a continuación.

Si los vasos sanguíneos no son removidos, es posible observar que los códigos provistos por OverFeat alcanzan los mejores resultados cuando las imágenes son recortadas en la región del FOV, independientemente del regularizador utilizado. Por el contrario, si los vasos son sustraídos de las imágenes, entonces ambas normas reportan la mejor performance cuando se utilizan acercamientos en torno a la PPA. En cualquiera de estos casos no es posible determinar cuál es la mejor estrategia de aumentación del conjunto de entrenamiento, dado que éste varía de una configuración a otra. En el caso de las imágenes recortadas en el FOV y sin *inpainting* de los vasos sanguíneos, se obtiene el mejor valor de AUC (0.7626) cuando el conjunto de entrenamiento se aumenta utilizando rotaciones a  $90^\circ$ ; al utilizar *inpainting*, el valor más alto (0.7212) se obtiene sin utilizar aumentación. La mejora de contraste mediante CLAHE no es recomendable para la gran mayoría de las configuraciones utilizadas,

dato que disminuye significativamente los valores de AUC. Dicha estrategia solamente demuestra una contribución positiva cuando se remueven los vasos, en los FOVs y ONHs recortados, y cuando se emplea la norma  $\ell_1$  en las imágenes originales. La norma  $\ell_2$  se comporta de manera similar o incluso mejor que la  $\ell_1$  para casi todas las configuraciones evaluadas. En lo que respecta a los vasos sanguíneos, se observa que su sustracción no ofrece ninguna mejora en los resultados. Más aún, la mayoría de los valores de AUC disminuyen cuando esta operación es tomada en cuenta. La única excepción se observa cuando las imágenes se enfocan en el ONH y se aplica CLAHE. Se realizó un experimento adicional, también, no incluido en la figura, en el que los clasificadores fueron entrenados con códigos de OverFeat extraídos a partir de las segmentaciones binarias de los vasos, aunque los valores de AUC obtenidos no difirieron casi respecto de los obtenidos por un comportamiento completamente aleatorio (el valor más alto correspondió a 0.5303, obtenido en la versión aumentada con rotaciones a  $90^\circ$ ).

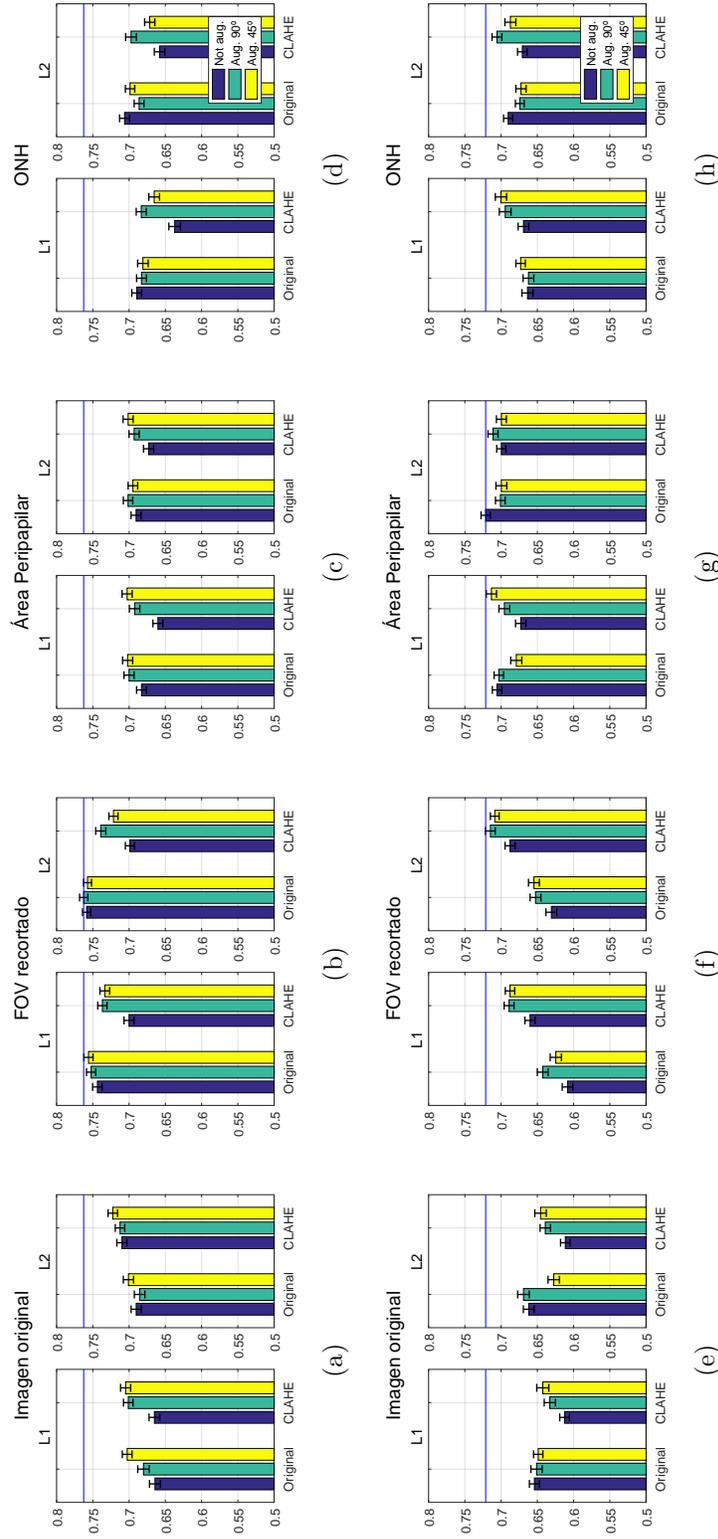


Figura 4.5: Area bajo las curvas ROC promedio obtenidas a partir de 200 pruebas [58] utilizando los descriptores calculados por OverFeat y clasificadores por regresión logística regularizada con normas  $\ell_1$  o  $\ell_2$ , respectivamente. Primera fila: utilizando imágenes sin *inpainting* de vasos sanguíneos. Segunda fila: utilizando imágenes con *inpainting* de los vasos. La línea azul corresponde al mayor valor de AUC para la fila.

Se realizó, además, un experimento adicional utilizando VGG-S, sólo que alimentando la red utilizando las imágenes preprocesadas con la mejor estrategia observada al utilizar OverFeat (esto es, las imágenes RGB recortadas en torno al FOV, sin aplicar CLAHE ni *inpainting*, aumentadas con rotaciones a  $90^\circ$ ; e imágenes RGB recortadas en torno al área peripapilar y sin aplicar CLAHE, removiendo los vasos y sin aplicar aumentación alguna). El objetivo de este análisis fue determinar si pequeños cambios en la arquitectura de la red pueden afectar la performance de la red transferida. En ambos experimentos se utilizó regresión logística regularizada con norma  $\ell_2$  debido a que esta configuración demostró ser capaz de obtener mejores resultados al evaluar las técnicas de preprocesamiento. Se observó que los descriptores obtenidos con OverFeat reportaron un mejor rendimiento que los de VGG-S tanto en imágenes con (AUC= 0,7212 vs. AUC= 0,6655, respectivamente) como sin *inpainting* de los vasos sanguíneos (AUC= 0,7626 vs. AUC= 0,7180, respectivamente).

Cabe señalar que no es factible realizar una comparación exhaustiva respecto a otros enfoques del estado del arte debido a que la gran mayoría de los trabajos existentes evalúan sus resultados en conjuntos de imágenes privados. De acuerdo al relevamiento realizado, únicamente el trabajo de Chakrabarty *et al.* [32] fue evaluado sobre DRISHTI-GS1, reportando un valor de AUC de 0.78. Tal enfoque está basado en un *framework* de clasificación que fusiona características obtenidas a partir de las imágenes completas y de sus segmentaciones, utilizando para ello un método de co-entrenamiento basado en aprendizaje semi-supervisado. La principal diferencia de su enfoque respecto al aquí presentado es que éste último emplea únicamente descriptores extraídos de versiones recortadas de las imágenes de fondo de ojo originales, y no integra ningún tipo de información sobre las características de las segmentaciones del ONH. Además, el enfoque presentado en [32] es pre-entrenado sobre un conjunto de fotografías de fondo de ojo mucho mayor que el empleado aquí.

## 4.5. Discusión

En este capítulo se analizó la capacidad de transferir redes neuronales convolucionales (CNNs) pre-entrenadas sobre datos no médicos, a la detección de glaucoma en fotografías de fondo de ojo. Se estudió particularmente la contribución de una

gran variedad de técnicas de preprocesamiento del estado del arte (incluyendo recortes en torno al ONH) a la mejora de la capacidad discriminativa de los descriptores extraídos utilizando dos arquitecturas de CNNs distintas. Cada enfoque fue evaluado en combinación con modelos de regresión logística, regularizados a través de dos normas diferentes. El rendimiento de los recortes del FOV fue mayor que el de otros tipos de acercamientos al combinarse con OverFeat, lo que indica que esta red podría ser capaz de recobrar información valiosa a partir de regiones diferentes al ONH. Esta característica se ve evidenciada también en los bajos valores de AUC obtenidos al emplearse *inpainting* del árbol vascular. Estos resultados contradicen otros estudios previos, que se han enfocado en extraer descriptores únicamente del área peripapilar o del ONH luego de remover los vasos sanguíneos [26, 27]. Los valores de AUC reportados utilizando solamente el árbol vascular, sin embargo, sugieren que éstos no son útiles por sí mismos si no que deben ser combinados con información proveniente de otras áreas de la retina para alcanzar diagnósticos más certeros. Mejorar contraste utilizando CLAHE, una técnica ampliamente utilizada por diversos algoritmos para el procesado de imágenes de fondo de ojo, incrementó los resultados únicamente en algunas configuraciones. En la mayoría de los casos en los que los vasos fueron conservados, la técnica no resultó efectiva. Sin embargo, cuando los vasos fueron eliminados, CLAHE benefició los resultados reportados en imágenes recortadas en torno al FOV y en el ONH, y también en el PPA cuando se utilizó regularización  $\ell_1$ . Esto puede asociarse a un incremento en la diferencia entre la copa óptica y el resto del ONH, aunque esta mejora no permite alcanzar resultados competitivos respecto a otras configuraciones. La regularización por norma  $\ell_1$ , como se mencionó previamente, constituye una herramienta útil para realizar simultáneamente tanto la tarea de clasificación como la de selección de características, dado que favorece el aprendizaje de vectores de coeficientes raros. Esta propiedad no es compartida por la norma  $\ell_2$ , que, por el contrario, tiende a utilizar todos los coeficientes. Como la norma  $\ell_2$  obtuvo un rendimiento equivalente o superior al de la norma  $\ell_1$  en la mayoría de los experimentos, podría conjeturarse que todas las componentes de los vectores de características son útiles para detectar pacientes con glaucoma. Sin embargo, vale la pena mencionar que la mejora obtenida al combinar CLAHE con la regularización  $\ell_1$  también puede asociarse a que algunos coeficientes son más beneficiados por contar

con un mayor contraste en las imágenes. Por otro lado, se observó que, al menos en el conjunto de imágenes utilizado, los descriptores obtenidos empleando VGG-S no mejoran los resultados reportados por los de OverFeat.

En conclusión, es posible observar que la utilización de diferentes estrategias de preprocesamiento, incluso siendo éstas simples y conocidas, permiten mejorar los resultados obtenidos por transferencia de redes neuronales convolucionales, particularmente en conjuntos de datos pequeños en los que el reajuste de los pesos de la red por reentrenamiento no es factible. En términos del rendimiento general del método propuesto, se alcanzó un área bajo la curva ROC competitiva respecto a la reportada por otras estrategias existentes [32], con la evidente ventaja, además, de que no es necesario segmentar ni la copa ni el disco óptico en el enfoque aquí presentado. No obstante, resulta imprescindible continuar la evaluación del algoritmo sobre conjuntos de datos mayores, para validar la capacidad de generalización del mismo.

---

## Capítulo 5

# Detección de lesiones rojas para *screening* de RD

En el Capítulo 2 se mencionó que el signo clínico más temprano de la retinopatía diabética (RD) son las lesiones rojas. Bajo este término se agrupan en a las microaneurismas (MAs) y a las hemorragias (HEs). En la práctica clínica diaria, estas lesiones son detectadas de manera manual por los oftalmólogos, utilizando para ello fotografías de fondo de ojo. Sin embargo, esta tarea resulta tediosa y requiere de mucho tiempo de análisis, principalmente debido al tamaño pequeño de las lesiones y a su falta de contraste respecto a otras estructuras de la retina. Los métodos para diagnóstico asistido por computadora de RD basados en detección de lesiones rojas están siendo ampliamente explorados debido a su capacidad de mejorar tanto la precisión como la consistencia de los diagnósticos realizados por los especialistas. Más aún, este tipo de sistemas proveen de información sumamente útil a los oftalmólogos, que pueden validar o corroborar fácilmente los resultados reportados por estas herramientas. Se han propuesto numerosos métodos para la detección de lesiones rojas en la literatura, la gran mayoría de ellos basados primero en caracterizar un conjunto de lesiones candidatas utilizando descriptores diseñados manualmente, y posteriormente en clasificarlos en detecciones reales o falsas. Los enfoques basados en aprendizaje profundo o *deep learning*, por el contrario, son menos abundantes en este dominio debido al enorme costo que supone anotar las lesiones manualmente.

En este capítulo se propone un método novedoso para la detección de lesiones

rojas basado en combinar *deep learning* con el conocimiento actual del dominio. En particular, se propone aumentar los descriptores aprendidos por una red neuronal convolucional (CNN) utilizando otros que fueron diseñados manualmente. Este vector integral de características es posteriormente utilizado para identificar lesiones candidatas reales mediante un clasificador conocido como Bosque Aleatorio (o *Random Forest*, RF). Experimentalmente se observó que la combinación de ambas fuentes de información permite mejorar significativamente los resultados obtenidos con respecto a utilizar cada enfoque de forma separada. Más aún, el método propuesto reportó la mejor performance en los conjuntos de imágenes DIARETDB1 [96] y e-ophtha [48] en una evaluación por lesión, y la mejor para detectar pacientes con signos tempranos (*screening*) y referibles en MESSIDOR [49] comparados con un segundo observador experto. El capítulo provee además un análisis extensivo acerca de la complementariedad de los descriptores aprendidos por la CNN con respecto a los diseñados manualmente, con el propósito de estudiar su contribución al proceso de discriminación de lesiones reales y falsamente detectadas. Los resultados subrayan la importancia de integrar ambos enfoques para mejorar los resultados obtenidos por las CNNs cuando son entrenadas con datos anotados a nivel de lesión.

El capítulo se organiza de la siguiente manera. La Sección 5.1 brinda detalles acerca del problema a resolver y de la literatura actual en el área. En la Sección 5.2 se describe el método en detalle, explicando la estrategia de detección de lesiones candidatas empleada, la arquitectura de la CNN escogida, los descriptores utilizados para complementarla, y el clasificador por RF. En la Sección 5.3 se detalla el diseño experimental, y la Sección 5.4 presenta los resultados obtenidos. La Sección 5.5 analiza estos resultados en detalle y el nivel de complementariedad de ambos conjuntos de descriptores. Finalmente, la Sección 5.5 concluye el capítulo.

La implementación del método propuesto está disponible para su uso en <https://github.com/ignaciorlando/red-lesion-detection>. Por otro lado, el trabajo presentado en este capítulo está descrito en el *preprint*:

- J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Learning to detect red lesions in fundus photographs: An ensemble approach based on deep learning. *arXiv preprint arXiv:1706.03008*, 2017.

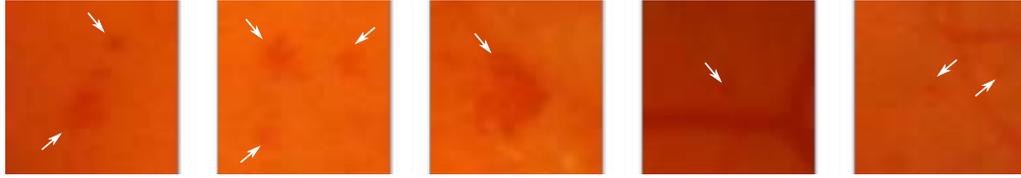


Figura 5.1: Ejemplos de lesiones rojas observadas en fotografías de fondo de ojo del conjunto de imágenes DIARETDB1 [96].

## 5.1. Motivación

Aunque las fotografías de fondo de ojo constituyen actualmente la técnica de imagenología médica no-invasiva más económica y útil para el diagnóstico de la RD, el diagnóstico manual a partir de ellas requiere un esfuerzo intenso por parte de los especialistas [131]. Esto se debe principalmente a que detectar las lesiones rojas típicas de la RD es sumamente dificultoso, dado que las mismas aparecen como pequeños puntos rojos que suelen ser apenas distinguibles y demasiado diminutos como para descubrirlos en una inspección rápida (Figura 5.1). Las HEs grandes, por el contrario, son más evidentes y menos difíciles de visualizar.

Los métodos automáticos para diagnóstico asistido por computadora permiten reducir el tiempo, costo y esfuerzo de detectar pacientes con RD: su capacidad para tratar varias imágenes en poco tiempo asegura un análisis más eficiente de grandes poblaciones de pacientes [176]. También permiten reducir la variabilidad entre expertos, usualmente alta debido al tamaño pequeño de las lesiones y a su forma irregular [3]. Este tipo de sistemas por lo general cuentan con un módulo particular para la detección automática de lesiones rojas. En general, el problema de hallar este tipo de estructuras es tratado mediante un enfoque en dos etapas, basado primero en determinar una serie de potenciales lesiones candidatas, y luego refinar el conjunto resultante empleando un clasificador que ha sido previamente entrenado con descriptores diseñados manualmente [142, 143, 180, 207].

Las redes neuronales convolucionales (*Convolutional Neural Networks*, CNNs) han surgido recientemente como un *framework* sumamente útil par resolver una gran variedad de problemas de visión computacional y análisis de imágenes médicas [102, 202, 230]. Tales métodos son capaces de aprender características automáticamente a partir de un conjunto de imágenes lo suficientemente grande, sin necesidad

de recurrir al diseño manual de los filtros. Las CNNs han superado a los enfoques basados en la tradicional ingeniería manual de descriptores en una gran variedad de aplicaciones [169]. No obstante, su capacidad discriminativa se ve usualmente afectada por la cantidad de datos de entrenamiento con los que disponen: las arquitecturas más profundas de redes neuronales convolucionales son capaces de aprender descriptores más discriminativos, aunque al costo de contar con mayores conjuntos de datos para prevenir el sobreajuste a los datos de entrenamiento y para asegurar un adecuado error de generalización [71]. El etiquetado de grandes volúmenes de datos a nivel de imagen (esto es, una etiqueta para cada imagen del conjunto) puede realizarse de manera relativamente económica [196]. Sin embargo, etiquetarlas a nivel de lesión es costoso, tanto en tiempo como en recursos económicos, ya que requiere la intervención de observados expertos que deben inspeccionar áreas pequeñas de las imágenes para identificar cada estructura patológica, y delinearlas de la forma más precisa posible. Este hecho suele influir significativamente en la performance de los algoritmos basados en *deep learning*, que deben entrenarse con datos anotados a nivel lesión.

En este estudio se propone aprovechar las bondades tanto de los enfoques basados en la ingeniería manual de descriptores como en las características aprendidas por una CNN. En particular, se propone aprender un primer conjunto de descriptores utilizando una CNN con una arquitectura de una profundidad relativa, y posteriormente aumentar su habilidad original para caracterizar las lesiones utilizando características diseñadas manualmente. Estos vectores híbridos son utilizados para entrenar un Bosque Aleatorio (RF) que es luego aplicado sobre cualquier nueva imagen dada para remover falsos candidatos y discriminar cuáles de ellos constituyen efectivamente lesiones de interés. Los experimentos realizados permitieron observar que ambos enfoques son altamente complementarios, y que los descriptores aprendidos automáticamente se ven notoriamente beneficiados cuando se les incorpora información del dominio.

La detección de lesiones rojas en fotografías de fondo de ojo ha sido explorada durante años por la literatura científica, aunque la mayoría de los métodos existentes están basados en detectar MAs o HEs separadamente, y no ambos tipos de estructuras de manera simultánea [142, 180, 200]. Además, los enfoques existentes se basan

exclusivamente en descriptores diseñados manualmente. Esto se debe principalmente a que los métodos basados en *deep learning* deben ser entrenados a partir de grandes conjuntos de datos anotados a nivel lesión. Esta característica tiene implicancias directas en por qué los enfoques basados en CNNs han sido ignorados para resolver este problema. Una excepción es el método para detección de HEs de van Grinsven *et al.* [200]. Dicho algoritmo está enfocado en hallar HEs a diferentes escalas, que son por lo general más evidentes que las MAs. El enfoque propuesto en este capítulo, por el contrario, es utilizado para detectar tanto MAs como HEs pequeñas en forma simultánea, un problema aún más difícil dado que estas lesiones son menos distinguibles para los oftalmólogos.

Como se mencionó anteriormente, aquí se presenta un algoritmo integrado que mejora la capacidad discriminativa de una CNN mediante la incorporación de otros descriptores basados en el conocimiento del dominio. En la literatura se han observado muy pocos esfuerzos por analizar la viabilidad de combinar ambas fuentes de información. Annunziata *et al.* [15], por ejemplo, proponen inicializar un enfoque de codificación rala convolucional (o *convolutional sparse coding*) con filtros diseñados manualmente. El propósito final de los autores es acelerar el proceso de aprendizaje, por un lado, y por otro mejorar el poder discriminativo original de los filtros empleados. Este enfoque fue aplicado exitosamente para detectar estructuras curvilíneas tales como neuronas o vasos sanguíneos de la retina, que son más fáciles de delinear manualmente que las lesiones rojas. En el trabajo de Venkataramani *et al.* [202] se observó que los descriptores del estado del arte permiten mejorar la performance de las características de una CNN pre-entrenada sobre datos no médicos, cuando se los aplica para detectar los riñones en imágenes de ultrasonido. La principal diferencia de tal enfoque con el aquí propuesto es que la CNN utilizada en este trabajo está entrenada sobre un conjunto de datos específico del problema a resolver, mientras que la red utilizada en [202] está reajustada para su problema particular pero inicializada con los pesos aprendidos a partir de imágenes naturales.

De la revisión de la literatura realizada se identificaron dos métodos principales que se asemejan en algún sentido al aquí propuesto, aunque se basan en diferentes arquitecturas de CNNs y están aplicados a problemas diametralmente distintos. En [230] se introdujo un método para identificar referencias en imágenes 3D de tomo-

grafía computada combinando la salida de una CNN dedicada con descriptores de Haar. Su componente basada en *deep learning* se divide en dos etapas: una inicial, basada en una arquitectura delgada con una única capa oculta, permite recobrar un conjunto de puntos de referencia candidatos; mientras que la segunda, constituida por tres capas ocultas y entrenada con restricciones de *sparsity*, es utilizada para extraer un vector de descriptores que se combina con los de Haar para entrenar un clasificador. Con el propósito de ahorrar la mayor cantidad de datos posibles para utilizarlos en el entrenamiento tanto de la CNN como del clasificador, el método que propone este capítulo no se basa en una etapa de detección de candidatos supervisada. Por el contrario, utiliza una combinación de operaciones morfológicas y técnicas de procesamiento de imagen para recuperar un conjunto de lesiones potenciales, sin utilizar datos de entrenamiento. Esto permite entrenar una red con una profundidad mayor en la etapa subsiguiente, dedicada únicamente a clasificar estos candidatos, que es más capaz de capturar características relevantes a partir de las ventanas de entrenamiento empleadas.

El método para detección de mitosis en imágenes histopatológicas propuesto en [209] también se asemeja al aquí propuesto. El mismo también emplea una etapa no supervisada para detectar áreas que puedan constituir mitosis, y utiliza un clasificador RF entrenado con descriptores diseñados manualmente para asignar una probabilidad a cada candidato. En paralelo se entrena una CNN con dos capas convolucionales y una totalmente conectada, utilizando para ello ventanas extraídas en torno a los candidatos. La CNN permite, entonces, obtener una probabilidad adicional, que se combina con la del RF por medio del ajuste manual de dos parámetros. En este trabajo se decidió utilizar una estrategia alternativa, basada en emplear ambos vectores de características de manera simultánea para entrenar el RF, dado que esto permite aprovechar mejor las interacciones entre los descriptores aprendidos por la CNN y los diseñados manualmente.

## 5.2. Métodos

En este capítulo se propone aprender modelos discriminativos para la detección de lesiones rojas mediante la combinación de descriptores aprendidos con *deep learning*

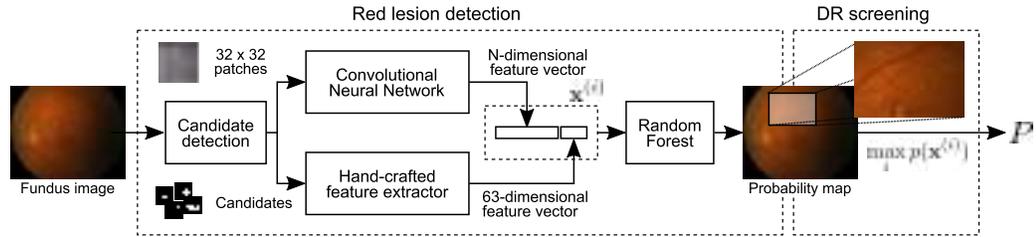


Figura 5.2: Representación esquemática del método propuesto para la detección de lesiones rojas.

y otros diseñados manualmente. En la Figura 5.2 se presenta un esquema representando el método propuesto. Inicialmente se utiliza un enfoque no supervisado para detectar lesiones candidatas, basado en operaciones morfológicas (Sección 5.2.1). Posteriormente se entrena una CNN a partir de un conjunto de ventanas regulares centradas en cada candidato, con el objetivo de aprender un primer vector de descriptores (Sección 5.2.2). Estas características son posteriormente aumentadas utilizando un conjunto de 63 nuevas características diseñadas manualmente, con el objetivo de mejorar su habilidad original para distinguir a las lesiones reales (Sección 5.2.3). Toda esta información es empleada para entrenar un Bosque Aleatorio o *Random Forest* (RF), que el luego aplicado para refinar el conjunto de candidatos y discriminar entre las lesiones verdaderas y los falsos positivos introducidos por la etapa inicial de detección de candidatos (Sección 5.2.4). Debido a que la presencia de lesiones rojas es el primer indicador de RD, la probabilidad máxima de todas las asignadas a las lesiones candidatas es asignada a la imagen para representar el riesgo de que la misma corresponda a un paciente con RD, como se propone en [180] y [16].

### 5.2.1. Detección de lesiones candidatas

La Figura 5.3 ilustra la estrategia utilizada para detectar lesiones candidatas. Inicialmente se extrae la banda verde  $G$  de la imagen RGB original  $I$ , dado que es la que permite una mejor discriminación visual de las lesiones rojas. Para evitar artefactos en los bordes del FOV que puedan ocultar lesiones potenciales (Figura 5.4(b)), se simula un ángulo de apertura del FOV mayor, de  $\frac{3}{30}\mathcal{X}$  píxeles, utilizando el enfoque propuesto en [188]. Este algoritmo se aplica sobre  $G$ , y utilizando un parámetro  $\mathcal{X}$  que se corresponde con el ancho en píxeles del FOV. Debido a que el propósito de este

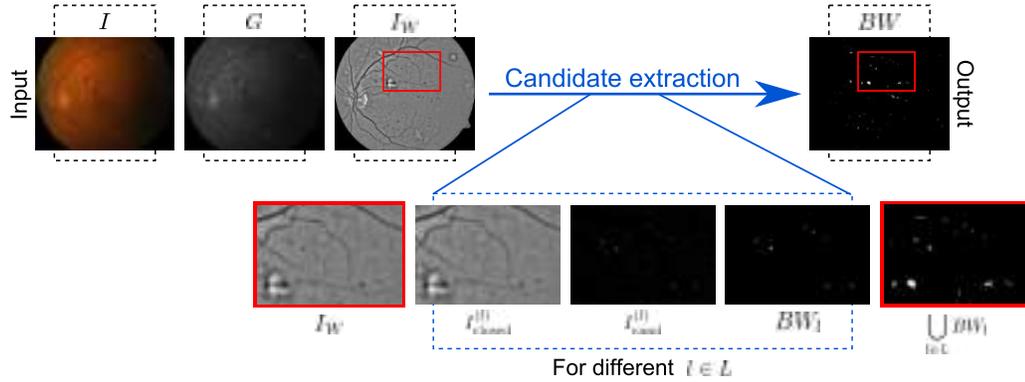


Figura 5.3: Detección de lesiones rojas candidatas. Ver Sección 5.2.1 para una descripción detallada del proceso.

trabajo es proponer un algoritmo lo suficientemente general como para operar sobre imágenes con diferentes resoluciones, todos los parámetros relevantes del algoritmo son expresados en función de  $\mathcal{X}$ .

Debido a que una iluminación despareja del fondo de la retina puede ocultar potenciales lesiones rojas que ocurran sobre las áreas más oscuras de las imágenes, se aplica sobre  $G$  una transformación  $r$ -polinomial de las intensidades de sus píxeles:

$$I_W(i, j) = \begin{cases} \frac{\frac{1}{2}(u_{\max} - u_{\min})}{(\mu_W(i, j) - \min(G))^r}, & G(i, j) \leq \mu_W(i, j) \\ \frac{-\frac{1}{2}(u_{\max} - u_{\min})}{(\mu_W(i, j) - \max(G))^r}, & G(i, j) > \mu_W(i, j) \end{cases} \quad (5.1)$$

con  $r = 2$ ,  $u_{\min} = 0$  y  $u_{\max} = 1$ , respectivamente, y donde  $\mu_W$  es la intensidad promedio local en vecindarios cuadrados de lado  $W$ , calculados para cada píxel  $(i, j)$  [207]. Se observó experimentalmente que utilizar un  $W = 25$  permite mejorar suficientemente el contraste en imágenes con 536 píxeles de resolución horizontal como las del conjunto de imágenes DRIVE [141], de manera que este parámetro es ajustado automáticamente utilizando  $W = \frac{25}{536}\mathcal{X}$ . La Figura 5.4 ilustra cómo la expansión del borde del FOV y la subsecuente transformación de las intensidades permite mejorar el contraste de las lesiones menos evidentes y ubicadas cerca del borde del FOV.

Sobre  $I_W$  se utiliza posteriormente un filtro Gaussiano con  $\sigma = \frac{5}{536}\mathcal{X}$  para reducir el ruido que potencialmente haya podido introducirse como consecuencia del proceso de mejora de contraste, resultando en una nueva imagen  $I'_W$ . Posteriormente se apli-

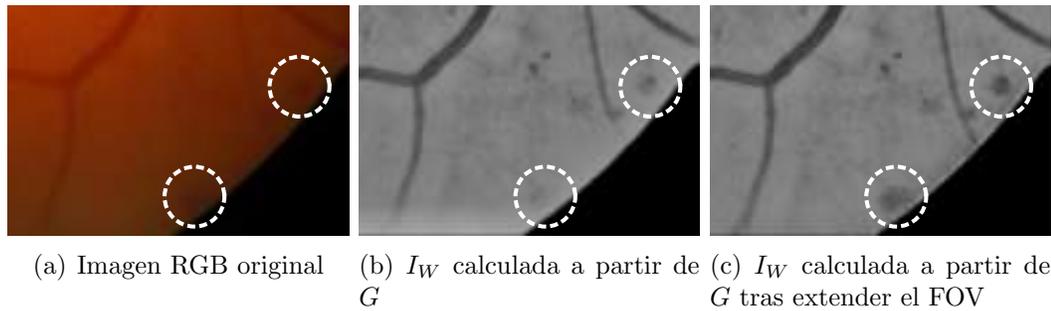


Figura 5.4: Efecto que la expansión del FOV tiene sobre las lesiones candidatas ubicadas de manera cercada al borde del FOV.

can diferentes *closings* morfológicos sobre  $I'_W$  utilizando elementos estructurantes de lado  $l \in L$ , orientados en ángulos  $\theta$  de 0 a  $180^\circ$ , con incrementos de  $15^\circ$ . El conjunto de escalas relevantes  $L$  es un parámetro fijo que es también ajustado automáticamente en términos de  $\mathcal{X}$ , como se detalla más adelante, en la Sección 5.3.2. Tomando la mínima respuesta a la operación morfológica para todos los ángulos considerados es que se obtiene una nueva imagen  $I_{\text{closed}}^{(l)}$  en la que las lesiones con tamaño inferior a  $l$  fueron prácticamente eliminadas visualmente, y todas las demás estructuras continúan intactas [207]. Luego, es posible obtener un mapa de respuestas a la existencia de las lesiones aplicando:

$$I_{\text{cand}}^{(l)} = I_{\text{closed}}^{(l)} - I_W. \quad (5.2)$$

Sobre este mapa  $I_{\text{cand}}^{(l)}$  es posible aplicar una operación de umbralado, para recobrar componentes conectadas binarias de un cierto tamaño, que representen lesiones candidatas. El umbral es determinado de forma automática de manera tal que se recobre un máximo de  $K = 120$  candidatos. Para lograrlo, se exploran una serie de umbrales  $t_s$  desde  $\text{mín}(I_{\text{cand}}^{(l)})$  a  $\text{máx}(I_{\text{cand}}^{(l)})$ , con incrementos de 0.002, hasta que el número de componentes conectadas en los mapas binarios resultantes sea menor o igual a  $K$ . Para brindar soporte a aquellos casos en los que no se detecten lesiones o no sea posible hallar menos de  $K$  candidatos, se establecen experimentalmente una

cota mínima  $t_l$  y una máxima  $t_u$  tales que:

$$t_K = \begin{cases} t_l, & \forall t_s : \text{CC}(I_{\text{cand}}^{(l)} > t_s) < K \\ t_k, & \text{CC}(I_{\text{cand}}^{(l)} > t_s) \leq K \\ t_u, & \forall t_s : \text{CC}(I_{\text{cand}}^{(l)} > t_s) > K \end{cases} \quad (5.3)$$

donde CC es una función que cuenta el número de componentes conectadas en el mapa binario resultante de aplicar el umbralado. Una vez que se ajusta  $t_K$ , se obtiene finalmente el mapa de candidatos aplicando  $BW_l = I_{\text{cand}}^{(l)} > t_K$  [207]. Esta operación es repetida secuencialmente para diferentes valores de  $l \in L$  a los efectos de capturar potenciales lesiones a diferentes escalas. Finalmente, el mapa de candidatos  $BW$  se obtiene como  $BW = \bigcup_{l \in L} BW_l$ . Teniendo en cuenta que  $BW$  todavía puede incluir pequeñas estructuras que correspondan a ruido en lugar de a lesiones reales, todas las estructuras conectadas en  $BW$  con menos de  $px$  píxeles son descartadas. El procedimiento para el ajuste automático de los parámetros  $K$  y  $px$  y las diferentes escalas en  $L$  se describe en la Sección 5.3.2.

La Figura 5.5 presenta una muestra aleatoria de algunas potenciales lesiones recobradas por el método de selección de candidatos propuesto, para una imagen cualquiera del conjunto de entrenamiento de DIARETDB1. Es posible observar que la gran mayoría de las muestras que son falsos positivos corresponden a bifurcaciones, cruces, ensanchamientos o segmentos de los vasos sanguíneos, o bien a cicatrices producto de tratamientos con fotocoagulación láser o pequeñas manchas de polvo en el dispositivo de captura (señaladas también en [180]). Esta característica subraya la importancia de incorporar una etapa adicional de clasificación que permita refinar el conjunto de candidatos y remover aquellos que correspondan a falsos positivos.

### 5.2.2. Descriptores basados en CNNs

A partir del conjunto de lesiones candidatas recobradas utilizando el método descrito en la sección anterior, se entrena una CNN con el objetivo de aprender características relevantes que permitan diferenciar las lesiones auténticas de las falsas. Para ello, se aplica inicialmente un preprocesamiento sobre cada banda de color de

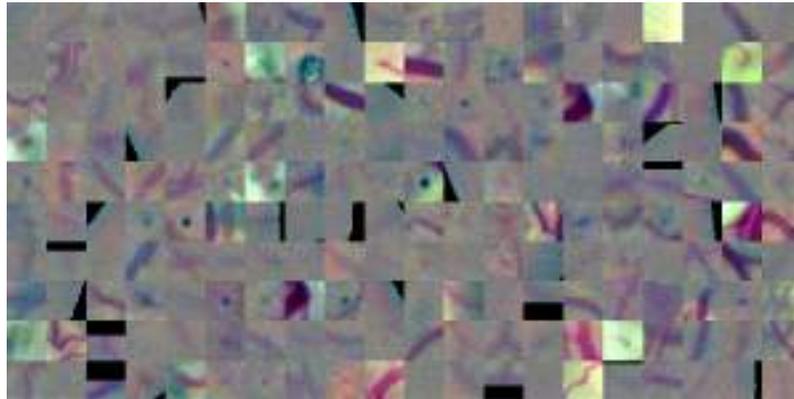
la imagen  $I$ , ecualizándolas utilizando la estrategia propuesta en [200]:

$$I_{ce}(i, j; \sigma) = \alpha \cdot I(i, j) + \tau \cdot \text{Gaussiano}(i, j; \sigma) * I(i, j) + \gamma \quad (5.4)$$

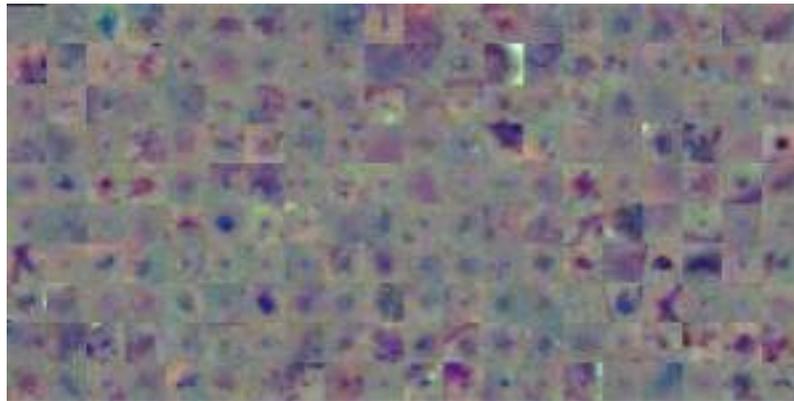
donde  $*$  es una operación de convolución, el filtro Gaussiano tiene un desvío  $\sigma = \frac{\lambda}{30}$ , y  $\alpha = 4$ ,  $\tau = -4$  y  $\gamma = 128$  son parámetros seteados como en [200]. Empíricamente se observó que esta operación no solamente reduce dramáticamente el número de épocas necesarias para entrenar la red si no que también mejora significativamente su capacidad discriminativa. Como se explica en la Sección 5.2.1, también se simula en este caso un FOV más amplio para prevenir cualquier efectos indeseado en sus bordes.

Para entrenar la CNN se construye un conjunto de entrenamiento  $S = \{(X^{(i)}, y^{(i)})\}$ ,  $i = 1, \dots, n$  en el que cada muestra  $X^{(i)}$  es un parche o ventana en torno al centro de cada lesión roja candidata, extraído a partir de  $I_{ce}$  (Figura 5.5). El tamaño del parche se fija como el doble de la longitud del eje mayor del candidato, con el objetivo de recobrar no solamente el candidato en sí mismo si no también el área que lo rodea. Esto permite a la CNN capturar tanto características internas de la lesión como información acerca de su morfología general, sus bordes y el contexto en el que se encuentra. Para reducir el número de parámetros de la red, todos los parches fueron normalizados al tamaño regular de  $32 \times 32$  píxeles. Por otro lado, todas las muestras son centradas en la media del conjunto de entrenamiento. La etiqueta  $y^{(i)} \in \{0, 1\}$  asociada a cada candidato es asignada de acuerdo al etiquetado de referencia a nivel lesión que se provee en el conjunto de imágenes: si el candidato se superpone con una región etiquetada como lesión, entonces la ventana se asume como una lesión roja verdadera ( $y = 1$ ); en caso contrario, se la asume como un falso positivo ( $y = 0$ ).

La CNN es entrenada de cero a partir de una versión del conjunto de entrenamiento aumentada 8 veces, reflejando las imágenes horizontalmente y rotándolas a  $90^\circ$ ,  $180^\circ$  y  $270^\circ$ . La arquitectura de la red se presenta en la Tabla 5.1. La misma consiste en 4 capas convolucionales y 1 capa totalmente conectada, con 128 unidades. Esta capa es utilizada para recuperar un vector de dimensión 128 de características aprendidas mediante *deep learning*. La CNN fue diseñada utilizando como modelo base la arquitectura original de LeNet, introduciendo cambios conforme los mismos contribuían a reducir el error empírico en un conjunto de validación extraído aleato-



(a) Lesions falsas (candidatos que son falsos positivos)



(b) Lesiones verdaderas (true positive candidates)

Figura 5.5: Conjunto de entrenamiento de la CNN. Muestra aleatoria de 200 parches para (a) candidatos falsos y (b) lesiones verdaderas. Ver Sección 5.2.2 para detalles acerca de la construcción del conjunto de entrenamiento.

riamente a partir del conjunto de entrenamiento. También se incorporó *dropout* luego de la primera capa convolucional. La utilización de arquitecturas más profundas no demostró mejorar los resultados empíricamente.

Dependiendo del número de imágenes con un nivel avanzado de RD presentes en el conjunto de entrenamiento y también del tamaño de las lesiones que el método se enfoca en detectar, el problema de clasificación puede estar desbalanceado en un mayor o menor grado. Así, la proporción de lesiones verdaderas puede ser significativamente menor al número de candidatos falsos. Si este desbalance crece dramáticamente, se ha observado que la pérdida basada en la entropía cruzada (típi-

camente utilizada para guiar el entrenamiento de las redes) se ve afectada, forzando la recuperación de menos lesiones reales [118]. Para evitarlo, se utiliza la pérdida por entropía cruzada balanceada por clase, dada por:

$$\mathcal{L}_\beta(\mathbf{W}) = -\beta \sum_{i \in Y_+} \log P(y^{(i)} | X^{(i)}; \mathbf{W}) - (1 - \beta) \sum_{i \in Y_-} \log P(y^{(i)} | X^{(i)}; \mathbf{W})$$

donde  $\mathbf{W}$  corresponde a los pesos que son parámetro de la red;  $P$  es la probabilidad obtenida de aplicar la función sigmoidea a las activaciones de la capa totalmente conectada;  $Y_+$  e  $Y_-$  son los subconjuntos de muestras realmente positivas y de falsos positivos, respectivamente; y  $\beta = |Y_-| / (|Y_+| + |Y_-|)$  es la proporción de muestras negativas vs. positivas en  $S$ . Esta función de costo fue optimizada utilizando el método de optimización basado en gradiente descendente estocástico, con un tamaño de lote o *batch* de 100 muestras.

Aplicar *dropout* luego de todas las capas convolucionales (algo tradicionalmente realizado en el estado del arte) fue analizado como una alternativa a la arquitectura finalmente elegida, aunque se observó que esta configuración no mejoró los resultados sobre un conjunto de validación. Más aún, aplicar esta otra estrategia demostró incrementar significativamente el tiempo de entrenamiento requerido para ajustar la red. Por el contrario, utilizar *dropout* luego de la primera capa convolucional con una probabilidad alta de mantener los pesos mejora ligeramente los resultados. Esta estrategia de regularización se complementa, además, con un decaimiento de los pesos o *weight decay* de  $5 \times 10^{-4}$ , para penalizar valores de  $\mathbf{W}$  altos durante el entrenamiento. La combinación de ambas estrategias demostró experimentalmente ser suficiente para controlar el sobreajuste u *overfitting*. La tasa de aprendizaje o *learning rate* fue inicializada en  $\eta = 0,05$ , y dividida por un factor de 2 toda vez que la mejora relativa en el valor de  $\mathcal{L}_\beta(\mathbf{W})$  sea menor al 1 % de la pérdida promedio de las últimas 10 épocas. Finalmente, todos los pesos de la red neuronal fueron inicializados aleatoriamente utilizando una distribución Gaussiana con desvío estándar 0.05, con la sola excepción de la capa convolucional, en la que el desvío estándar utilizado fue de 0.01.

Tabla 5.1: Arquitectura de la CNN. Las capas convolucionales (conv) indican el ancho, alto y profundidad de cada filtro aprendido. Las capas de *pooling* (pool) incluyen la dimensión de la operación de *pooling* y el tamaño de paso. El *dropout* es únicamente aplicado luego de la primera capa convolucional, con una probabilidad de *dropout* baja.

Bloque	Capas	Tamaño de filtro	Tamaño de la salida
1	conv	$5 \times 5 \times 3$	32
	maxpool	$3 \times 3$ - paso = 2	
	dropout	$p = 0,01$	
2	conv	$5 \times 5 \times 32$	32
	avgpool	$3 \times 3$ - paso = 2	
3	conv	$5 \times 5 \times 32$	64
	avgpool	$3 \times 3$ - paso = 2	
4	conv	$4 \times 4 \times 64$	128
5	totalmente conectada	128	128
6	$\mathcal{L}_\beta(\mathbf{W})$	128	2

### 5.2.3. Extraction de descriptores diseñados manualmente

Para complementar las características aprendidas por la CNN, se incorpora un vector de 63 descriptores diseñados manualmente (*hand-crafted features*, HCF), todos ellos extraídos para cada lesión candidata. Algunos de estos descriptores fueron explorados previamente en la literatura [142, 143, 180], mientras que otros son introducidos por primera vez aquí para mejorar los pre-existentes. En general, los descriptores pueden dividirse en dos categorías: basados en intensidades, por un lado, y basados en la morfología del candidato, por otro (Tabla 5.2).

Los descriptores basados en intensidades explotan las propiedades visuales de las áreas ocupadas por los candidatos, y son extraídas a partir de diferentes versiones de la imagen a color  $I$ , obtenidas a partir de aplicar distintas estrategias de preprocesamiento. En particular, se extrajeron descriptores que ya han sido aplicados en otros trabajos de la literatura [142, 180], pero a partir de las siguientes imágenes derivadas:

- Bandas de color roja, verde y azul ( $R$ ,  $G$  y  $B$ , respectivamente).
- Banda de color verde  $G$  luego de aplicar la estrategia de corrección de iluminación ( $I_W$ , obtenida como se detalla en la Sección 5.2.1).

- Bandas de color e  $I_W$  luego de aplicar CLAHE para mejora de contraste ( $R_c, G_c, B_c, I_{Wc}$ ).
- Bandas de color tras aplicar ecualización de color ( $R_{ce}, G_{ce}, B_{ce}$ ).
- $I_{SC}$ , que corresponde a la diferencia entre la banda verde  $G$  y un fondo estimado  $I_{BG}$ , obtenido utilizando un filtro de mediana con ventanas cuadradas de lado  $\frac{25}{536}\mathcal{X}$ .
- $I_{match}$ . Esta imagen es obtenida calculando inicialmente  $I_{lesion}$ , que corresponde a una versión libre de vasos de la imagen  $I_{SC}$ , obtenida aplicando la estrategia de *inpainting* descrita en [154] y en la Sección 4.2.1.3. Finalmente, la diferencia entre la intensidad de cada píxel  $(i, j)$  en  $I_{lesion}$  y su vecindario de  $11 \times 11$  píxeles se asigna a  $I_{match}(i, j)$ .
- $I_{cand} = \max_l I_{cand}^{(l)}$ , correspondiente a la máxima respuesta al mapa de *scores* utilizado para la detección de los candidatos, presentado en la Sección 5.2.1, tomado a partir de  $I_W$  pero restringiendo el tamaño del elemento estructurante a las longitudes  $l \in \{5, 7, \dots, 15\}$ .

Los descriptores de morfología permiten caracterizar la estructura y forma de los candidatos. Las lesiones rojas por lo general son relativamente circulares, con área y perímetro pequeños, y aproximadamente la misma longitud de eje principal y menor. Estos estadísticos, incluyendo la compacidad, excentricidad y su relación de aspecto, también se incluyen como parte del vector de descriptores.

Por otro lado, se analizó también la viabilidad de utilizar la segmentación de la vasculatura retiniana como una fuente potencial de información. Como se observa en la Figura 5.5(a), la mayoría de los candidatos que son falsos positivos están localizados en regiones vasculares. Para facilitar su remoción, inicialmente se obtiene, para cada imagen, una primera representación de los vasos sanguíneos, utilizando el método presentado en el Capítulo 3, y postprocesando la salida mediante la remoción de cada componente conectada espuria con menos de  $\frac{100}{536}\mathcal{X}$  píxeles [154]. A continuación se aplica un *closing* morfológico con discos de radio 2 para rellenar cualquier hueco en la segmentación que pueda haberse generado como consecuencia del reflejo central de las arterias. Luego, se utiliza como descriptor a la cantidad de

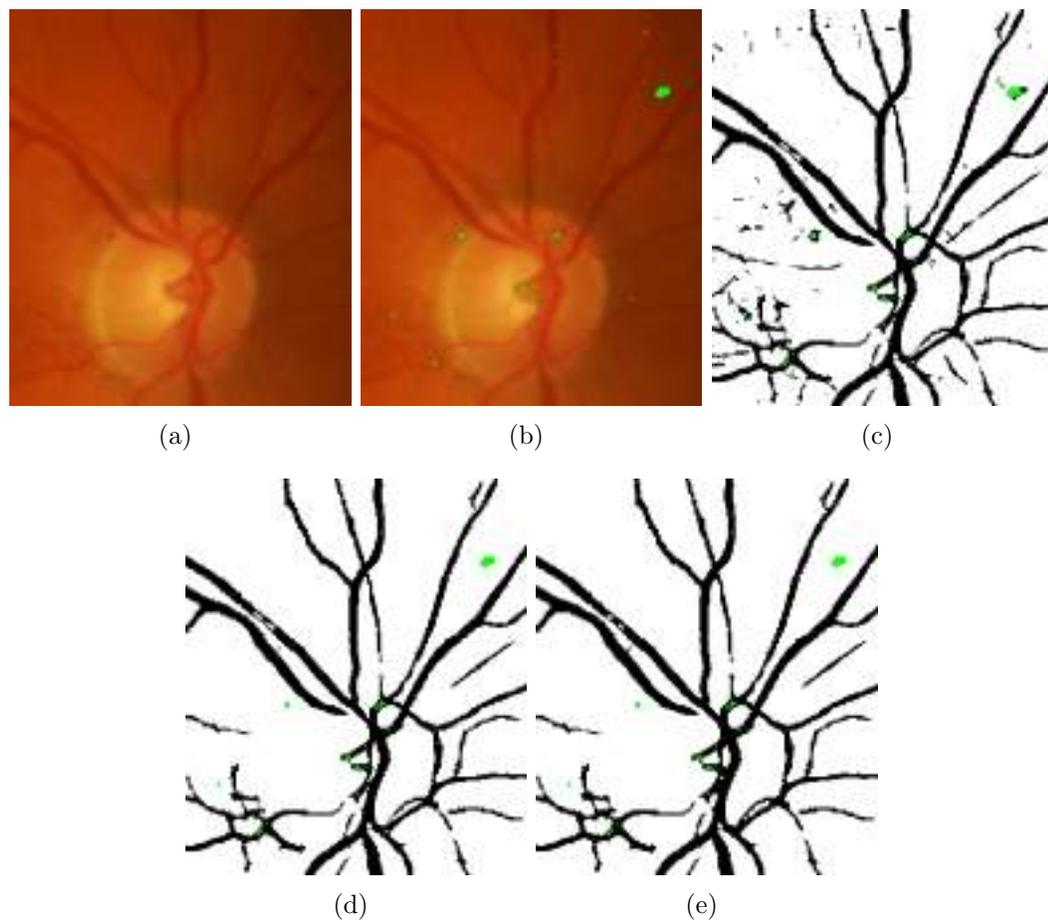


Figura 5.6: Descriptor basado en la segmentación de los vasos sanguíneos. (a)  $I$ . (b)  $I$  con lesiones candidatas superpuestas. (c) Segmentación de los vasos sanguíneos. (d) Segmentación de los vasos sanguíneos tras remover estructuras espurias. (e) Segmentación de los vasos sanguíneos luego de aplicar el *closing* morfológico.

píxeles en la región candidata que se superponen con la segmentación de los vasos sanguíneos, dividida por el total de píxeles del candidato. La Figura 5.6 ilustra el proceso de cálculo de este indicador. Puede observarse que la gran mayoría de los falsos positivos localizados dentro del disco óptico se superponen con la máscara de los vasos, y pueden ser fácilmente removidos utilizando este descriptor.

Tabla 5.2: Resumen de los descriptores diseñados manualmente (*hand crafted features*, HCF) utilizados para complementar la CNN.

Descriptor (dimensionalidad)	Extraído de:
Promedio de las intensidades en el candidato. (13)	$R, G, B, I_W, R_c, G_c, B_c, I_{W_c}, R_{ce}, G_{ce}, B_{ce}, I_{SC}, I_{top-hat}$
Suma de las intensidades del candidato. (12)	$R, G, B, I_W, R_c, G_c, B_c, I_{W_c}, R_{ce}, G_{ce}, B_{ce}, I_{SC}$
Desvío estándar de las intensidades del candidato. (12)	$R, G, B, I_W, R_c, G_c, B_c, I_{W_c}, R_{ce}, G_{ce}, B_{ce}, I_{SC}$
Contraste: diferencia entre la intensidad media del candidato y la intensidad media de la región dilatada. (12)	$R, G, B, I_W, R_c, G_c, B_c, I_{W_c}, R_{ce}, G_{ce}, B_{ce}, I_{SC}$
Intensidad total normalizada: diferencia entre la intensidad total y la intensidad media en el candidato en $I_{BG}$ , dividida por el desvío estándar del candidato en $I_{BG}$ . (3)	$G, I_{SC}, I_W$
Intensidad media normalizada: diferencia entre la intensidad media del candidato en $I_W$ y la intensidad media del candidato en $I_{BG}$ , dividida por el desvío estándar del candidato en $I_{BG}$ . (1)	$I_W$
Intensidad mínima dentro del candidato. (1)	$I_{match}$
Área: cantidad total de píxeles en la región candidata. (1)	$BW$
Perímetro: cantidad de píxeles en el borde del candidato. (1)	$BW$
Relación de aspecto: cociente entre las longitudes de los ejes mayor y menor. (1)	$BW$
Circularidad = $4\pi \text{Area} / \text{Perimeter}^2$ . (1)	$BW$
Compacidad = $\sqrt{(\sum_{j=1}^n d_j - \tilde{d}) / n}$ , donde $d_j$ es la distancia del centroide del objeto a su $j$ -ésimo píxel del borde y $\tilde{d}$ es la media de todas las distancias del centroide a todos los píxeles del borde. $n$ corresponde al perímetro del objeto. (1)	$BW$
Eje mayor de la elipse que tiene los mismos momentos centrales de segundo orden normalizados que la región. (1)	$BW$
Eje menor de la elipse que tiene los mismos momentos centrales de segundo orden normalizados que la región. (1)	$BW$
Excentricidad de la elipse que tiene los mismos momentos de segundo orden que la región candidata. La excentricidad es el cociente entre la distancia entre el foco de la elipse y la longitud de su eje mayor. (1)	$BW$
Proporción de píxeles de la región candidata que están incluidos en la segmentación binaria de los vasos sanguíneos, obtenida como se explica en la Sección 4.2.1.3. (1)	Segmentación de los vasos sanguíneos

### 5.2.4. Clasificación de lesiones candidatas usando Bosques Aleatorios

Un Bosque Aleatorio o *Random Forest* (RF) es un clasificador basado en el ensamble de árboles de decisión, ampliamente utilizado en la literatura por sus capacidades para clasificar y realizar selección de descriptores de manera simultánea [29, 114]. También se caracteriza por su robustez frente al sobreajuste u *overfitting* (algo de vital importancia cuando los conjuntos de entrenamiento disponibles son pequeños) y por su buen desempeño para lidiar con datos desbalanceados y ruidosos, con descriptores de dimensión alta. En este trabajo se utiliza este clasificador para refinar el conjunto de candidatos, empleando para su entrenamiento el vector híbrido de descriptores propuestos. En todos los experimentos realizados, los descriptores fueron estandarizados a media nula y varianza unitaria.

Un RF es una combinación de  $T$  árboles de decisión. Estos árboles son aprendidos a partir de  $T$  muestras que se seleccionan aleatoriamente, con reposición, a partir del conjunto de entrenamiento  $S$ . Cada nodo en un árbol corresponde a una partición realizada utilizando el mejor descriptor de un subconjunto de  $m = \sqrt{d}$  descriptores escogidos también de forma aleatoria, con  $d$  igual a la dimensionalidad total de los vectores de características. La calidad del corte está dada por la reducción en el índice de Gini que produce la partición en sí misma [29]. Dado un vector de descriptores  $\mathbf{x}^{(j)}$ , el RF evalúa la probabilidad condicional  $p_i(c|\mathbf{x}^{(j)})$ , donde  $c \in \{-1, 1\}$  es una clase—con -1 no lesión y 1 una lesión verdadera— e  $i$  es el índice del árbol en el bosque aleatorio. La probabilidad final es, entonces, calculada mediante la repetición de este proceso para cada árbol  $0 < i \leq T$ , y promediando las respuestas a todas ellas:

$$p(c|\mathbf{x}^{(j)}) = \frac{1}{T} \sum_i^T p_i(c|\mathbf{x}^{(j)}) \quad (5.5)$$

A los efectos de determinar la probabilidad  $P$  de que imagen  $I$  corresponda a un paciente con RD o no, se aplicó el mismo procedimiento utilizado en [180]:

$$P(I) = \underset{j}{\text{máx}} p(c = 1|\mathbf{x}^{(j)}), \quad (5.6)$$

lo que significa que, para una imagen  $I$  con  $m$  lesiones candidatas, la probabilidad

Tabla 5.3: Distribución de los niveles de RD en el conjunto de imágenes MESSIDOR, y criterio diagnóstico empleado. MA = microaneurismas, HE = hemorragias y NV= neovascularizaciones.

Nivel	Criterio	Núm. imágenes
R0	$(N_{MA} = 0) \text{ AND } (N_{HE} = 0)$	546
R1	$(0 < N_{MA} \leq 5) \text{ AND } (N_{HE} = 0)$	153
R2	$(5 < N_{MA} < 15) \text{ AND } (0 < N_{HE} < 5) \text{ AND } (N_{NV} = 0)$	247
R3	$(N_{MA} \geq 15) \text{ OR } (N_{HE} \geq 5) \text{ OR } (N_{NV} > 0)$	254

de que corresponda a un paciente con RD está asociada a la máxima certeza del clasificador de haber observado una lesión real ( $c = 1$ ).

## 5.3. Configuración experimental

### 5.3.1. Materiales

Se efectuaron experimentos sobre 3 conjuntos de imágenes disponibles y de acceso público: DIARETDB1<sup>1</sup> [96], e-ophtha<sup>2</sup> [48], y MESSIDOR<sup>3</sup> [49]

DIARETDB1 y e-ophtha fueron utilizados para efectuar una evaluación por lesión, dado que proveen anotaciones a nivel lesión. MESSIDOR, por el contrario, cuenta con etiquetas a nivel de imagen que indican el grado de la RD, asignadas utilizando el criterio descrito en la Tabla 5.3. Por tal motivo, este conjunto de imágenes fue utilizado para evaluar el comportamiento del algoritmo propuesto como una herramienta para *screening* de RD, a nivel imagen y utilizando las probabilidades  $P$  de presencia/ausencia de RD. También se utilizó e-ophtha para este propósito, generando anotaciones a nivel de imagen utilizando a tal efecto la cantidad de lesiones rojas presentes en las segmentaciones de referencia. Así, cualquier imagen con al menos una lesión roja fueron etiquetadas como RD. Por otro lado, el conjunto de imágenes de entrenamiento de ROC<sup>4</sup> [142] (consistente en 50 fotografías de fondo de ojo tomadas a diferentes resoluciones) fue utilizado para aumentar el conjunto de entrenamiento de DIARETDB1 para detectar lesiones rojas pequeñas en e-ophtha.

<sup>1</sup><http://www.it.lut.fi/project/imageret/diaretddb1/>

<sup>2</sup><http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html>

<sup>3</sup><http://messidor.crihan.fr>.

<sup>4</sup><http://webeye.ophth.uiowa.edu/ROC/>

En la Tabla 5.4 se brindan detalles adicionales acerca de la organización de los experimentos realizados.

DIARETDB1 consiste en 89 imágenes de fondo de ojo a color, obtenidas bajo diferentes configuraciones de captura [96]. 84 de estas imágenes contienen signos de RD leve o pre-proliferativa, mientras que las restantes 5 son consideradas normales. El conjunto completo se divide en un conjunto de entrenamiento y otro de test, formados por 28 y 61 imágenes, respectivamente. Cuatro expertos delinearon las regiones donde se encuentran las MAs y las HEs, y se provee un mapa de consensos para cada tipo de lesión. La práctica estándar es evaluar los métodos para detectar MAs o HEs a un  $\geq 75\%$  de coincidencia entre las anotaciones de los expertos [96], un nivel por demás conservativo si se tiene en cuenta que el nivel de acuerdo entre ellos es muy pequeño. Para la detección de lesiones rojas, sin embargo, en [180] se propone utilizar como referencia a la unión de los mapas de consenso tanto para MAs como para HEs luego de umbralar cada uno a un nivel de coincidencia  $> 25\%$ . En la evaluación realizada se utiliza este último enfoque.

e-ophta [48] es una base de datos generada a partir de una red de telemedicina para la detección de RD, e incluye las segmentaciones manuales de MAs y HEs pequeñas. El conjunto está formado por 148 imágenes con lesiones rojas diminutas, y 233 sin ningún signo visible de RD. Para obtener las etiquetas a nivel de imagen que indican la presencia o ausencia de la RD, cualquier imagen con alguna lesión roja fue etiquetada como RD.

Finalmente, MESSIDOR [49] cuenta con 1200 imágenes de fondo de ojo adquiridas por 3 instituciones oftalmológicas en Francia. Las imágenes fueron originalmente capturadas a diferentes resoluciones, y etiquetadas como pertenecientes a cuatro niveles de RD, siendo R0 la categoría utilizada para pacientes sanos y R3 para los pacientes con patología más severa. A partir de estas etiquetas es posible establecer dos problemas de clasificación diferentes: *screening* o detección de RD, que corresponde a distinguir R0 de las restantes categorías R1, R2 y R3 [16, 180]; y la detección de pacientes referibles (o *need for referral*), que corresponde a R0 y R1 vs. R2 y R3 [162, 176]. El método propuesto fue evaluado para ambas configuraciones.

Debido a que estos conjuntos de imágenes no incluyen máscaras del FOV, que son necesarias para los procesamientos realizados, las mismas fueron generadas au-

Tabla 5.4: Configuración experimental.  $\beta$  es el valor de la pérdida por entropía cruzada balanceada por clase (Ecuación (5.5)).

ID del exp.	Detección	Conjunto de entrenamiento	Etiquetas de referencia	Lesiones reales	No lesiones	$\beta$	Evaluación por lesión	Evaluación por imagen
1	Lesiones rojas de múltiples tamaños	Conjunto de entrenamiento de DIARETDB1 (28 imágenes)	MA > 25 % HE > 25 %	1059 (27%)	2905 (73%)	$\beta = 0,5$	Conjunto de test de DIARETDB1	MESSIDOR
2	Lesiones rojas pequeñas	Conjuntos de entrenamiento de DIARETDB1 y ROC (78 imágenes)	MA > 75 % para DIARETDB1 y MA en ROC	407 (4%)	10282 (96%)	$\beta \sim 0,96$	e-ophta	e-ophta

tomáticamente umbralando el plano de luminosidad de la versión en el espacio color CIELab de las imágenes originales RGB. Los umbrales utilizados fueron de 0.15 para DIARETDB1, e-ophta y MESSIDOR, y de 0.26 para ROC [154]. En los casos en los que la máscara resultante fue tal que toda la imagen se estimó como FOV, se aplicó un enfoque alternativo en el que se sumaron las tres bandas de color RGB y se umbraló la imagen resultante a un valor ajustado empíricamente de 150. Para suavizar los bordes y reducir ruido, todas las máscaras fueron postprocesadas con un filtro de mediana utilizando ventanas cuadradas de lado 5, preservando únicamente la mayor componente conectada. Nótese que este proceso es equivalente al descrito en la Sección 4.3.2. En principio, estas máscaras podrían estar disponibles a partir de la cámara de fondo de ojo, y el proceso de replicar esta información directamente a partir de las imágenes es necesario aunque no crítico para este trabajo en particular. Las máscaras para todos los conjuntos de datos utilizados en este capítulo están disponibles para su descarga en el sitio web del proyecto (Section 5.6).

### 5.3.2. Ajuste de parámetros

La efectividad del método para la detección de lesiones candidatas descrito en la Sección 5.2.1 está dada por el ajuste apropiado de tres parámetros:  $L$ , que corresponde al conjunto de escalas utilizado para estudiar las potenciales lesiones;  $K$ , que es el número de candidatos recuperados para una escala dada; y  $px$ , o el área mínima en píxeles que un candidato debe tener para no ser considerado como ruido. En

los experimentos realizados, estos parámetros fueron ajustados experimentalmente utilizando el conjunto de imágenes de entrenamiento de DIARETDB1, resultando en los valores  $L = \{3, 6, 9, \dots, 60\}$ ,  $K = 120$  y  $px = 5$ . La máxima escala para  $L$  fue adaptada en los demás conjunto de datos utilizando el factor de escala  $\frac{x}{1425}$ , donde 1425 es el ancho promedio de las imágenes en DIARETDB1. Esto permite extraer un conjunto de candidatos con un tamaño proporcional a la resolución de cada imagen. Los parámetros de la CNN (en particular, la probabilidad de *dropout*  $1 - p$  y el tamaño  $N$  de la capa totalmente conectada) fueron designados de acuerdo a la performance observada en un conjunto de validación extraído aleatoriamente de cada conjunto de entrenamiento. Los parámetros que maximizaron el área bajo la curva de precision/recall ( $N = 128$  y  $p = 0,99$ ) sobre este conjunto fueron posteriormente empleados para la evaluación en los conjuntos de test. El número de árboles  $T \in \{100, 110, \dots, 200\}$  para el RF fue fijado en el valor que minimizó el *out-of-bag error* (OOB) en el conjunto de entrenamiento [29]. Cabe señalar, finalmente, que la máxima cantidad de árboles posible se mantuvo fija en un valor bajo para reducir el costo computacional del entrenamiento y la predicción. De cualquier forma, se realizaron experimentos incorporando hasta 2000 árboles al modelo, sin obtener mejoras en en la reducción del OOB.

### 5.3.3. Métricas de calidad

Para evaluar a nivel lesión al método propuesto se utilizaron curvas FROC (*free-response ROC*). Estas gráficas han sido empleadas en numerosos trabajos de la literatura para estimar la performance general de diferentes algoritmos aplicados a la detección de lesiones rojas. En particular, las mismas representan la sensibilidad por lesión vs. el promedio de falsos positivos por imagen (FPI) obtenidos en un conjunto de test y para diferentes umbrales aplicados sobre las probabilidades asignadas a los candidatos. Así, las curvas FROC proveen una representación gráfica de la habilidad del modelo para detectar lesiones reales en todas las imágenes utilizadas en los experimentos. Por otro lado, y a los efectos de contar con un indicador numérico del rendimiento general del método, se utilizó la misma métrica del *Retinopathy Online Challenge* [142], denominada *Competition Metric* (CPM). Esta medida se calcula como el promedio de las sensibilidades por lesión para valores de FPI de referencia

$\in \{1/8, 1/4, 1/2, 1, 2, 4, 8\}$ . Finalmente, vale la pena señalar una vez más que la evaluación sobre DIARETDB1 se realizó siguiendo exactamente el mismo protocolo de evaluación utilizado en [180], detallado en la Sección 5.3.1.

Para la evaluación a nivel de imagen se utilizaron curvas ROC estándar, en las que se representan en una misma gráfica tanto la sensibilidad ( $Se = \frac{TP}{FN+TP}$ ) y  $1 -$  especificidad ( $Sp = \frac{TN}{FP+TN}$ ) para los diferentes valores de probabilidad obtenidos como indica la Ecuación 5.6. Por otro lado, se estudió la  $Se$  para  $Sp = 50\%$ , que constituye una métrica de comparación estándar para los sistemas de *screening* de RD [176].

## 5.4. Resultados

### 5.4.1. Evaluación por lesión

Para la evaluación por lesión del método propuesto se realizaron dos experimentos diferentes, detallados en la Tabla 5.4. Para comparar los resultados obtenidos se utilizaron curvas FROC, y la significancia estadística de la diferencia entre los valores de sensibilidad por lesión se estimó mediante tests del signo de Wilcoxon. En particular, estos tests fueron realizados utilizando 100 valores de sensibilidad por cada curva, obtenidos para valores de FPI espaciados logaritmicamente en el intervalo  $[\frac{1}{8}, \dots, 8]$ , que corresponde a una versión más densa de los valores de FPI de referencia utilizados para calcular la CPM [142].

El experimento 1 evalúa la capacidad del modelo para detectar MAs y HEs simultáneamente a múltiples escalas, siguiendo el protocolo utilizado en [180] (Figura 5.7). Los resultados de [180], incluidos en la figura, fueron provistos por los autores y obtenidos utilizando la misma configuración de entrenamiento y test. Los tests de hipótesis muestran una mejora estadísticamente significativa en los valores de sensibilidad por lesión al utilizar el enfoque combinado, respecto a utilizar cada representación de forma separada ( $p < 2 \times 10^{-18}$  y  $p < 4 \times 10^{-17}$  para las probabilidades dadas por la CNN y los descriptores diseñados manualmente, respectivamente). Además, el método híbrido reportó mejores resultados que los obtenidos por [180] ( $p < 2 \times 10^{-18}$ ).

Como DIARETDB1 incluye etiquetas separadas tanto para MAs como para HEs,

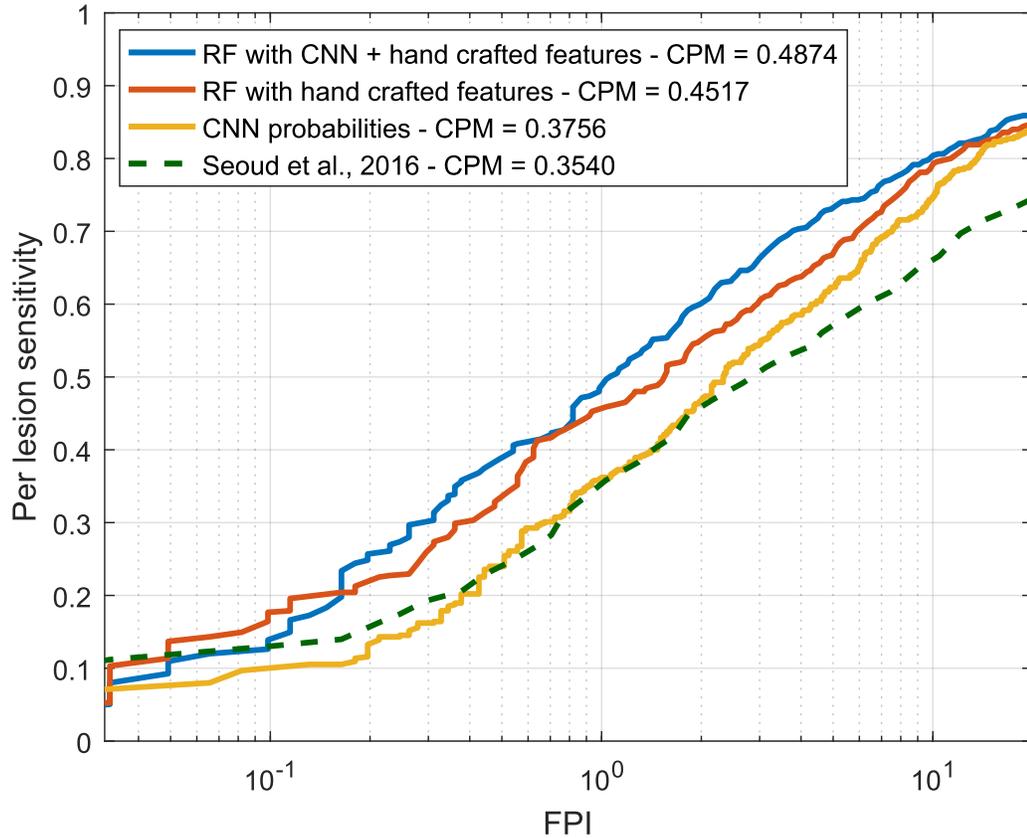


Figura 5.7: Evaluación por lesión en el experimento 1. Curva FROC y valores de CPM obtenidos sobre el conjunto de test de DIARETDB1.

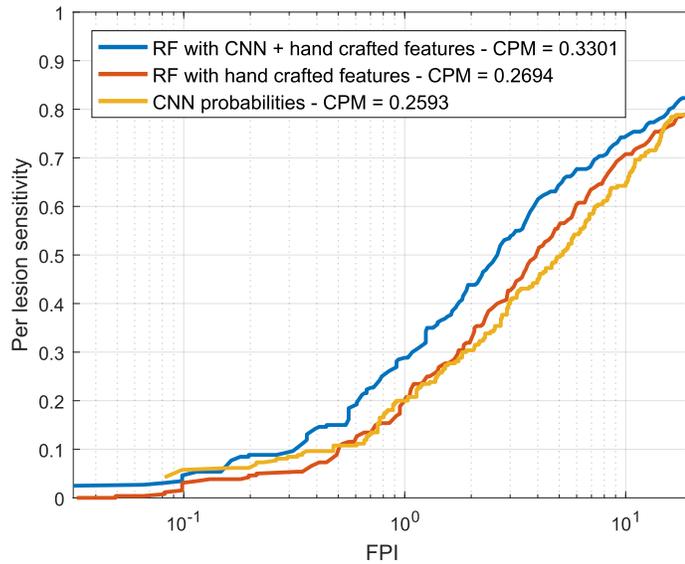
es posible estudiar cuantitativamente la precisión del método para detectar cada tipo de lesión. La Figura 5.8 ilustra las curvas FROC y los valores de CPM obtenidos por los modelos aprendidos en el experimento 1, al analizar MAs y HEs de manera separada. El enfoque combinado alcanza mayores valores de sensibilidad por lesión para detectar MA, comparado con el uso de cada enfoque por separado ( $p < 2 \times 10^{-18}$  y  $p < 3 \times 10^{-17}$  para los descriptores manuales y la CNN, respectivamente), con una mejora notoria en el valor de FPI=1, que es considerado clínicamente relevante en la literatura (0.2885 vs. 0.202 y 0.2 para el enfoque combinado, la CNN, y las características diseñadas manualmente, respectivamente). Por otra parte, cabe señalar que las diferencias entre el enfoque ajustado manualmente y las probabilidades de la CNN no son estadísticamente significativas. Al evaluar la capacidad del sistema propuesto para detectar HEs en el conjunto de test de DIARETDB1, es posible observar

que las sensibilidades por lesión son mayores que las reportadas para detección de MAs. Más aún, el enfoque basado en descriptores diseñados manualmente es capaz de alcanzar mejores valores de sensibilidad que el método combinado ( $p < 5 \times 10^{-5}$ ) para esta tarea específica. Para el valor de FPI= 1 clínicamente relevante, sin embargo, el algoritmo combinado reporta una sensibilidad por lesión algo mayor que la de los descriptores diseñados manualmente (0.4907 vs. 0.4724).

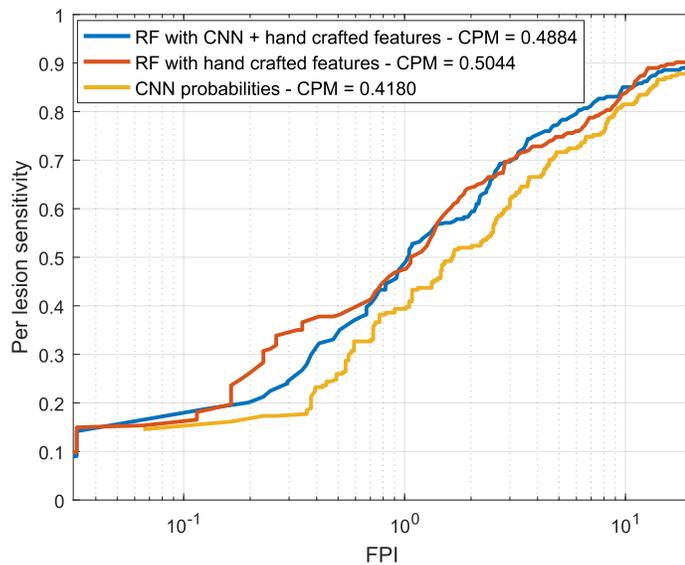
El experimento 2 fue llevado a cabo sobre e-ophtha para estimar la capacidad del método para detectar MAs y HEs más pequeñas, todas de manera simultánea. En este caso se utilizó para el entrenamiento una combinación de los conjuntos de entrenamiento de DIARETDB1 (con las etiquetas para MAs asignadas para un nivel de acuerdo entre expertos  $\geq 75\%$ ) y de ROC. Fue necesario aumentar el número de muestras con las extraídas de ROC dado que DIARETDB1 cuenta únicamente con 182 MAs al nivel de acuerdo mencionado. De acuerdo al relevamiento bibliográfico realizado, el único método para detección de lesiones evaluado sobre e-ophtha corresponde al presentado en [217], aunque su análisis es realizado sobre una muestra pequeña de 74 imágenes con lesiones en lugar de utilizando el conjunto completo. Por el contrario, en este experimento se evaluó una configuración más desafiante al emplear e-ophtha completo, incluyendo también las 233 imágenes del conjunto sin señales visibles de RD. La Figura 5.9 presenta las curvas FROC obtenidas para cada enfoque. Como se observó en el experimento anterior, las pruebas del signo de Wilcoxon demostraron una mejora estadísticamente significativa en los valores de sensibilidad por lesión al utilizar el vector híbrido de descriptores, con respecto a las probabilidades de la CNN y a los descriptores diseñados manualmente ( $p < 2 \times 10^{-18}$  y  $p < 2 \times 10^{-9}$ , respectivamente).

La Tabla 5.5 resume los valores de CPM obtenidos para cada experimento y combinación de descriptores, incluyendo además los reportados por dos métodos del estado del arte recientemente publicados. También se incluyen los valores de sensibilidad por lesión cuando FPI= 1, considerada como una cantidad clínicamente relevante de falsos positivos [142].

Finalmente, la Figura 5.10 ilustra algunos resultados cualitativos para una imagen seleccionada aleatoriamente del conjunto de test de DIARETDB1. Los círculos verdes corresponden a lesiones detectadas de acuerdo al etiquetado de referencia



(a) Microaneurismas



(b) Hemorragias

Figura 5.8: Evaluación por lesión y para cada tipo de lesión en el conjunto de test de DIARETDB1.

provisto por el conjunto de datos, mientras que los círculos amarillos corresponden a lesiones detectadas por el método propuesto pero no identificadas como tales en el *ground truth*. Finalmente, los círculos rojos corresponden a las lesiones anotadas

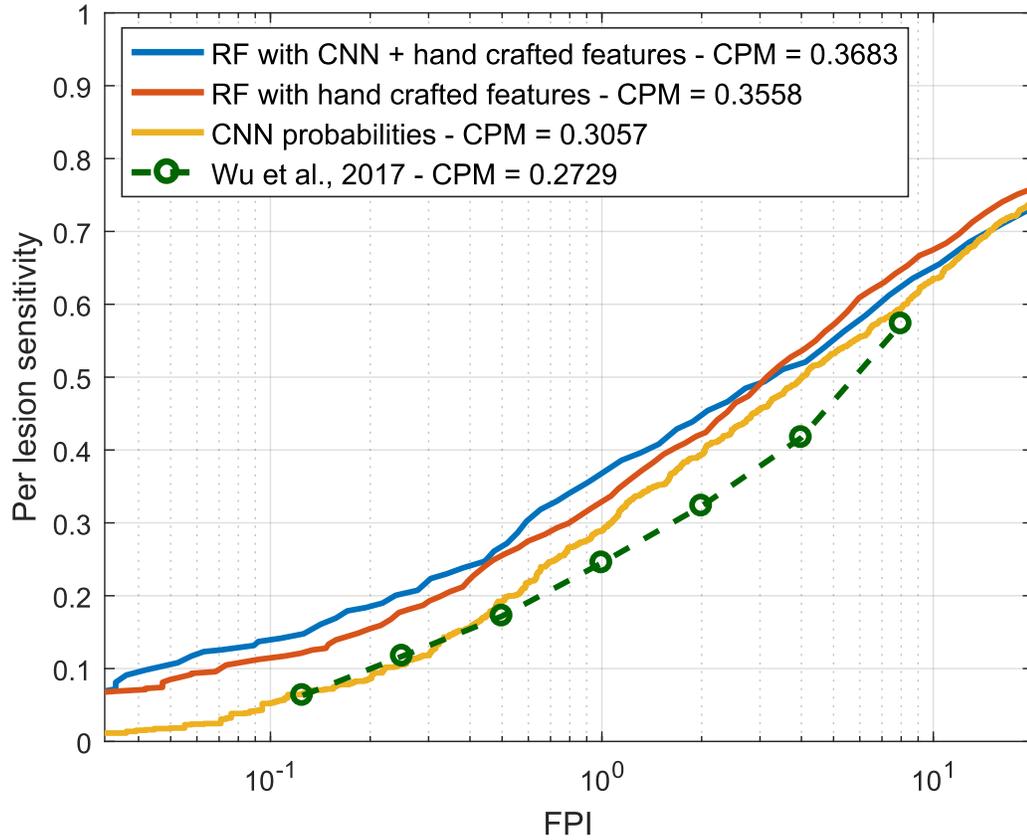


Figura 5.9: Evaluación por lesión en el experimento 2. Curva FROC y valores de CPM obtenidos sobre e-optha.

Tabla 5.5: Valores de CPM y sensibilidad por lesión para FPI= 1 para los experimentos 1 y 2.

Método	Experimento 1		Experimento 2	
	CPM	$Se$	CPM	$Se$
Seoud <i>et al.</i> , 2015 [180]	0.3540	0.3462	-	-
Wu <i>et al.</i> , 2017 [217]	-	-	0.2729	0.2450
Probabilidades de la CNN	0.3756	0.3621	0.3057	0.2894
RF entrenado con HCF	0.4517	0.4601	0.3558	0.3291
<b>RF entrenado con CNN + HCF</b>	<b>0.4874</b>	<b>0.4883</b>	<b>0.3683</b>	<b>0.3680</b>

como reales en el etiquetado de referencia pero ignoradas por el método. Puede observarse cualitativamente que muchos de los círculos amarillos rodean candidatos que corresponden a microaneurismas o hemorragias que no fueron detectadas durante el

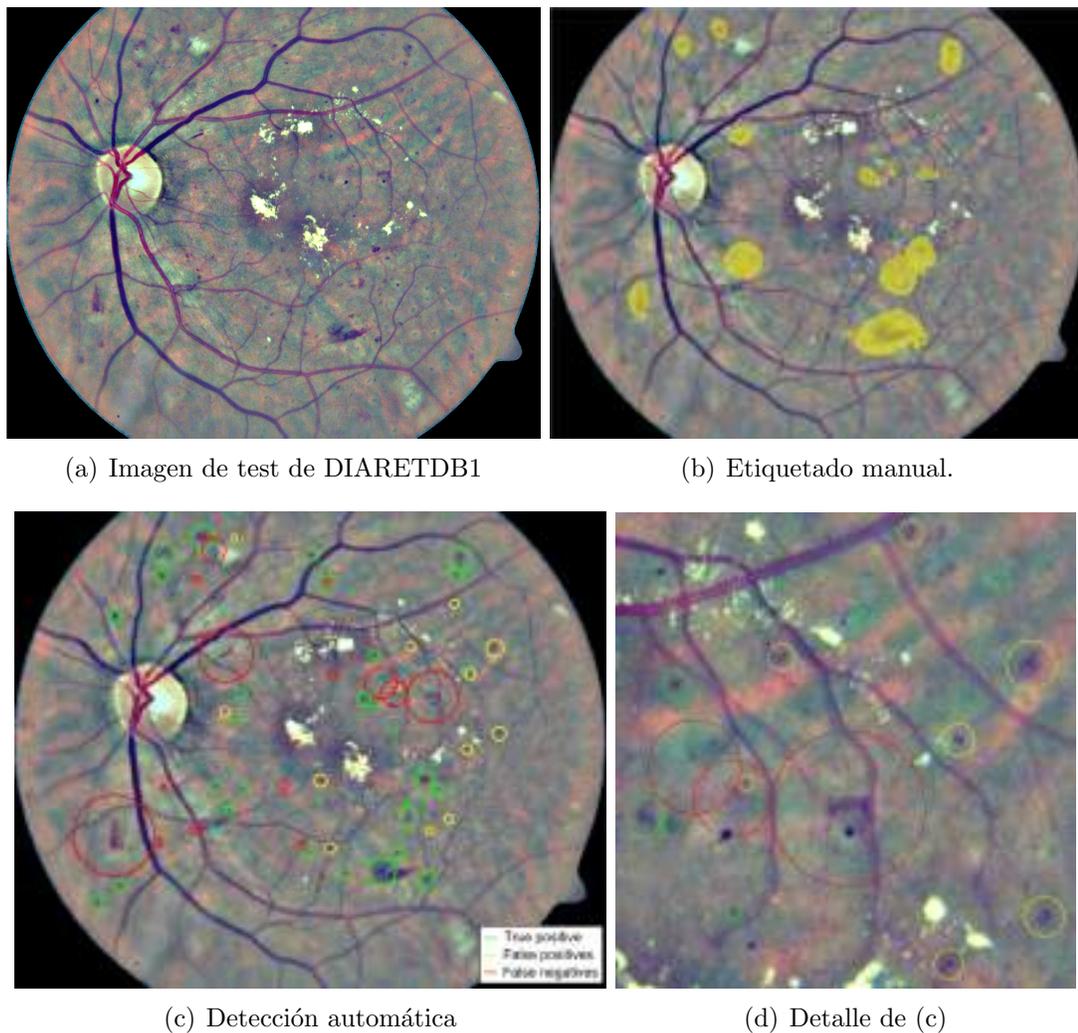
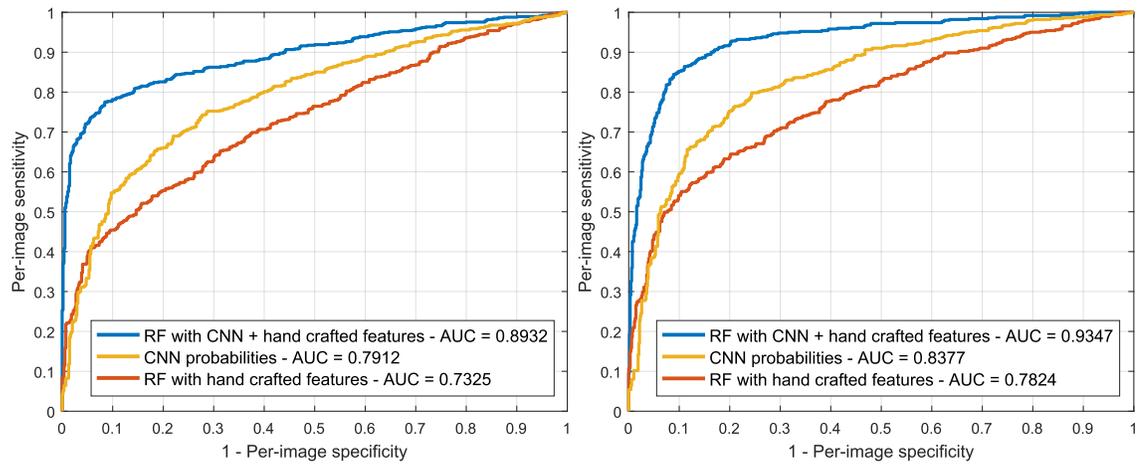


Figura 5.10: Resultados cualitativos. (a) Imagen image015 del conjunto de test de DIARETDB1. (b) Etiquetado de referencia a un nivel de acuerdo entre expertos  $> 25\%$  (c) Lesiones rojas detectadas utilizando el método propuesto, umbralando las probabilidades a 0.644, que corresponde a un valor promedio de FPI de 1. (d) Detalle de (c) mostrando lesiones no etiquetadas manualmente pero identificadas por el método.

etiquetado manual debido a su apariencia sutil en las imágenes RGB originales.

(a) Performance para *screening* de RD

(b) Performance para detectar pacientes que deben ser referidos

Figura 5.11: Evaluación por imagen en MESSIDOR. Curvas ROC para (a) *screening* de RD (R1 vs. R2, R3 y R4), y (b) detección de pacientes referibles (R1 y R2 vs. R3 y R4).

## 5.4.2. Evaluación por imágenes

Sobre el conjunto de datos MESSIDOR se realizaron dos experimentos diferentes para estimar la performance del método por imagen, uno enfocado en detectar pacientes con RD (*screening*), y otro basado en detectar aquellos que necesitan ser referidos inmediatamente a un especialista (*need for referral*). En ambos casos, se utilizó el modelo aprendido en el experimento 1.

La Figura 5.11(a) ilustra las curvas ROC para *screening* de RD en MESSIDOR, obtenidas utilizando la representación híbrida propuesta en este trabajo y cada uno de los enfoques por separado. Los resultados de la CNN fueron obtenidos utilizando la red como clasificador. Para estudiar la significancia estadística en las diferencias entre los valores de AUC se realizaron una serie de pruebas  $U$  de Mann-Whitney  $U$  ( $\alpha = 0,05$ ) [76]. Los descriptores aprendidos por la CNN (AUC = 0,7912) reportaron una performance significativamente mejor ( $p < 1 \times 10^{-3}$ ) que la de los descriptores diseñados manualmente (AUC = 0,7325) para esta tarea específica, mientras que la combinación de ambas fuentes de información resultó en un valor de AUC de 0.8932 sustancialmente mayor ( $p < 1 \times 10^{-6}$ ). La Figura 5.11(b) muestra un comportamiento

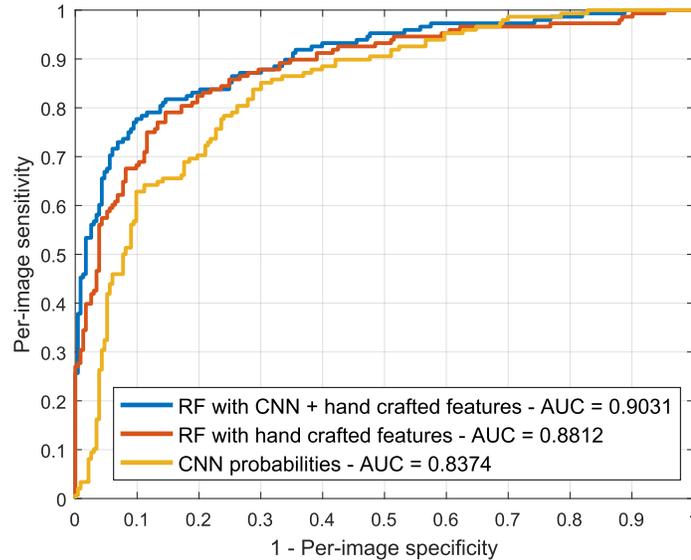


Figura 5.12: Evaluación por imagen sobre e-ophta. Curva ROC curve para *screening* de RD.

análogo en la detección de pacientes que deben ser referidos al especialista, con la CNN reportando también mejores resultados que los descriptores manuales ( $p < 2 \times 10^{-3}$ ), y el enfoque combinado superando ambas técnicas individuales ( $p < 1 \times 10^{-6}$ ).

El enfoque combinado presenta un comportamiento análogo al evaluarse sobre e-ophta para *screening* de RD, como se ilustra en la Figura 5.12. El valor de AUC reportado (0.9031) es significativamente mayor que el obtenido por la CNN (AUC = 0.8374,  $p < 5 \times 10^{-3}$ ) y por el RF entrenado con descriptores ajustados manualmente (AUC = 0.8812). Estos últimos reportaron mejores resultados que la CNN para *screening* en este conjunto de imágenes, aunque la diferencia no es estadísticamente significativa de acuerdo a la prueba  $U$  de Mann-Whitney.

La Tabla 5.6 presenta una comparación con respecto a otras estrategias del estado del arte. También se incluyen los resultados obtenidos por dos observadores expertos, tal y como se reportan en [176]. Los resultados del método base presentado en [180] corresponden a los obtenidos utilizando el conjunto de entrenamiento de DIARETDB1. Los demás métodos incluidos están o bien basados en detectar lesiones rojas o bien complementan esta información con otras funcionalidades tales como el estudio de la calidad de las imágenes o la detección de exudados y neovascularizaciones.

Tabla 5.6: Comparación de resultados para *screening* de RD o detección de pacientes referibles sobre MESSIDOR. Los valores de *Se* corresponden a aquellos obtenidos para  $Sp = 50\%$ .

Método	Screening		Need for referral	
	AUC	Se	AUC	Se
<i>Experto A</i> [176]	0.9220	0.9450	0.9400	0.9820
<i>Experto B</i> [176]	0.8650	0.9120	0.9200	0.9760
Antal y Hajdu, 2012 [16]	0.8750	-	-	-
Costa <i>et al.</i> , 2016 [43]	0.8700	-	-	-
Giancardo <i>et al.</i> , 2013 [70]	0.8540	-	-	-
Nandy <i>et al.</i> , 2016 [134]	-	-	0.9210	-
Pires <i>et al.</i> , 2015 [161]	-	-	0.8630	-
Sánchez <i>et al.</i> , 2011 [176]	0.8760	<b>0.9220</b>	0.9100	0.9440
Seoud <i>et al.</i> , 2016 [180] (DIARETDB1)	0.844	-	-	-
Vo y Verma, 2016 [205] (I)	0.8620	-	0.8910	-
Vo y Verma, 2016 [205] (II)	0.8700	-	0.8870	-
<b>HCF</b>	0.7325	0.7645	0.7824	0.8283
<b>CNN</b>	0.7912	0.8471	0.8377	0.9102
<b>HCF + CNN</b>	<b>0.8932</b>	0.9109	<b>0.9347</b>	<b>0.9721</b>

### 5.4.3. Análisis de los descriptores

La Figura 5.13 representa los 32 filtros de  $5 \times 5 \times 3$  aprendidos por la primera capa de la CNN. La misma es incluida a los efectos de estudiar la apariencia visual de los descriptores aprendidos por la red. En la Figura 5.13(a) es posible observar que los filtros aprendidos en el experimento 1 son mayormente descriptores de propiedades de color de las lesiones. Esta característica se asocia al hecho de que el conjunto de entrenamiento utilizado en este caso contiene no solamente MAs pequeñas si no también HEs de tamaño medio, que pueden ser descriptos más fácilmente en términos de su homogeneidad interna que de sus bordes, que varían significativamente de una a otra. Otros filtros aprendidos son capaces de capturar estructuras púrpura y elipsoidales, correspondientes a lesiones reales como las ilustradas en la Figura 5.5(b). Este último tipo de filtro es más común en la primera capa de la CNN aprendida en el experimento 2 (Figura 5.13(b)), que puede estar asociada con lesiones reales de tamaño más pequeño como las observadas en el conjunto de entrenamiento construido

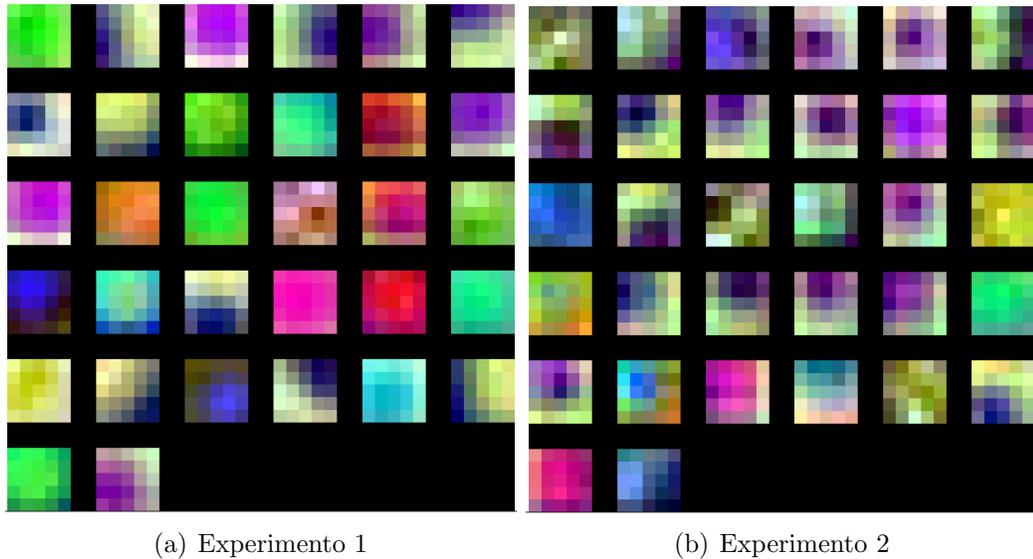


Figura 5.13: Filtros aprendidos por la primera capa de la CNN propuesta, obtenidos para cada experimento en la Tabla 5.4.

ensamblando ROC y las MAs de DIARETDB1.

El método de encastre  $t$  estocástico distribuido, o *t-distributed stochastic neighbor embedding* ( $t$ -SNE) es una técnica introducida recientemente para estudiar visualmente la capacidad discriminativa de diferentes vectores multidimensionales de descriptores para clasificar distintas entradas [198]. En esta sección se incluyen este tipo de visualizaciones para estudiar la complementariedad de cada enfoque de caracterización individual, y para al mismo tiempo validar cualitativamente cómo la integración de ambos enfoques contribuye a mejorar cada algoritmo en particular. La Figura 5.14 muestra los mapeos obtenidos utilizando  $t$ -SNE sobre muestras del conjunto de test de DIARETDB1, para cada enfoque de caracterización por separado y para el método combinado. Los descriptores de la CNN corresponden a aquellos aprendidos en el experimento 1. La figura también incluye una representación visual de la organización de los diferentes entornos en el espacio de encastre. En general, es posible observar que el enfoque combinado agrupa a la mayoría de los candidatos reales en un único vecintario. Por el contrario, las estrategias de caracterización individuales no son capaces de identificar un único clúster si no dos (en el caso de los descriptores aprendidos con *deep learning*) o más (utilizando descriptores diseñados

manualmente). En la Figura 5.15 se ilustran algunas regiones particulares de los encastres. Esto permite visualizar mejor algunos escenarios particulares tales como las ventanas en torno a lesiones rojas verdaderas, en torno a falsos positivos localizados sobre estructuras vasculares, los artefactos debidos a suciedad en el lente (propios de DIARETDB1) y las detecciones falsas en el disco óptico. En general, se observa que los descriptores de las CNN son capaces de caracterizar mejor la orientación y apariencia visual de las lesiones verdaderas, mientras que los descriptores diseñados manualmente permiten detectar lesiones menos obvias en condiciones de bajo contraste. La habilidad para discriminar orientaciones que tienen las características aprendidas por la CNN son más evidentes al momento de lidiar con estructuras vasculares. El enfoque manual, por el contrario, sólo es capaz de capturar el tamaño general de los vasos y sus intensidades. Al combinar ambas estrategias, las principales ventajas de cada enfoque se preservan. La robustez ante la presencia de artefactos es evidente tanto para el enfoque basado en *deep learning* como para el método tradicional, dado que los falsos positivos están agrupados en clústeres separados de las lesiones reales. Finalmente, vale la pena señalar que se observa un comportamiento similar al lidiar con candidatos falsos en el área del disco óptico.

## 5.5. Discusión

En general, la integración del enfoque tradicional basado en características ajustadas manualmente y las aprendidas utilizando *deep learning* mejoró significativamente los resultados obtenidos al compararse con los alcanzados por cada enfoque por separado. En la evaluación por lesión, el algoritmo combinado alcanzó un valor de CPM consistentemente mayor tanto en los conjuntos de test tanto de e-optha como de DIARETDB1, además de una sensibilidad por lesión mayor para FPI=1, que corresponde a una cantidad clínicamente relevante de falsos positivos [142]. Estos valores son también más altos que los reportados por dos métodos base recientemente publicados y evaluados sobre los mismos datos. Se observó también un comportamiento similar al evaluar el método utilizando un protocolo por imagen. El enfoque combinado mejoró la performance obtenida por cada estrategia de caracterización por separado, lo que implica que la integración de ambas fuentes de información

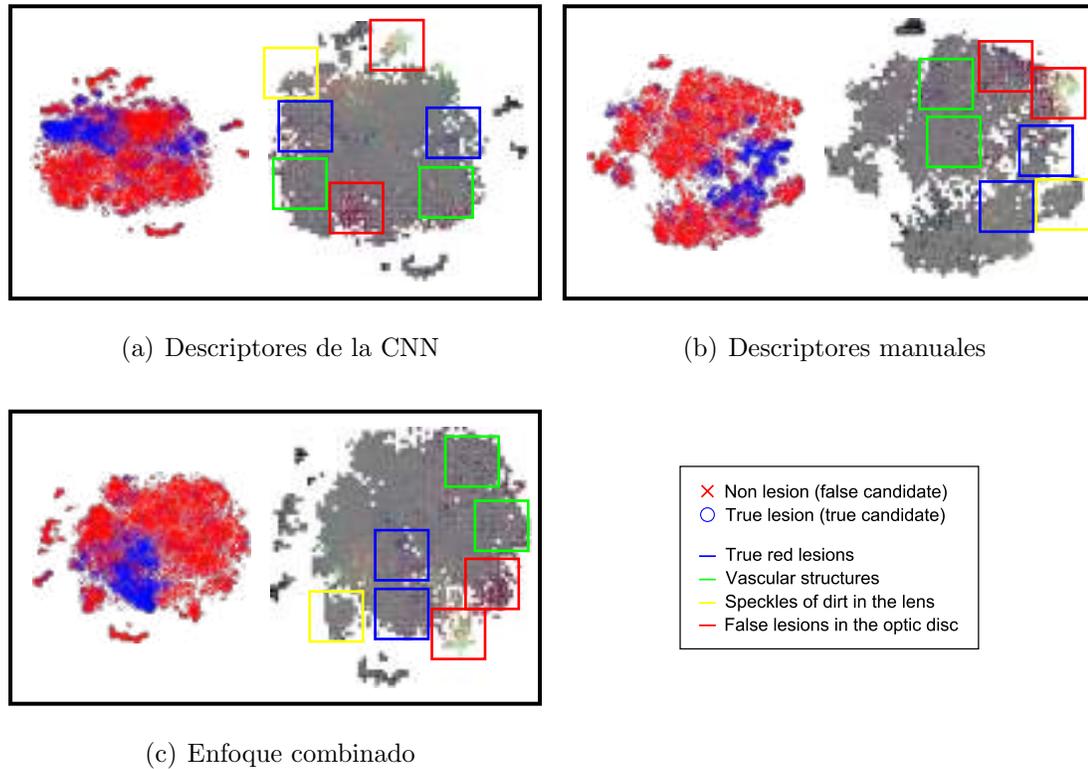


Figura 5.14: Visualización  $t$ -SNE de los entornos extraídos de los datos de test de DIARETDB1, mapeados utilizando los descriptores aprendidos con *deep learning*, los diseñados manualmente y el enfoque híbrido. Izquierda: etiquetas de color para cada muestra de test. Derecha: parches alrededor de los candidatos, visualizados utilizando los mapeos de  $t$ -SNE. El detalle para cada tipo de lesión candidata se observa en la Figura 5.15.

tiene implicancias positivas en la caracterización de las lesiones candidatas y, en consecuencia, permite una detección más precisa de las lesiones individuales. Esta hipótesis también es evidenciada en el análisis de los mapeos  $t$ -SNE obtenidos para cada método (Figuras 5.14 y 5.15), en donde es posible ver que el enfoque combinado aprovecha las ventajas de la CNN para caracterizar detalles finos tales como la orientación de las lesiones, mientras que los descriptores diseñados manualmente permiten reconocer mejor las lesiones con bajo contraste.

Los resultados obtenidos en la evaluación por imagen también demostraron que la estrategia propuesta es capaz de alcanzar valores de AUC más altos que otros enfo-

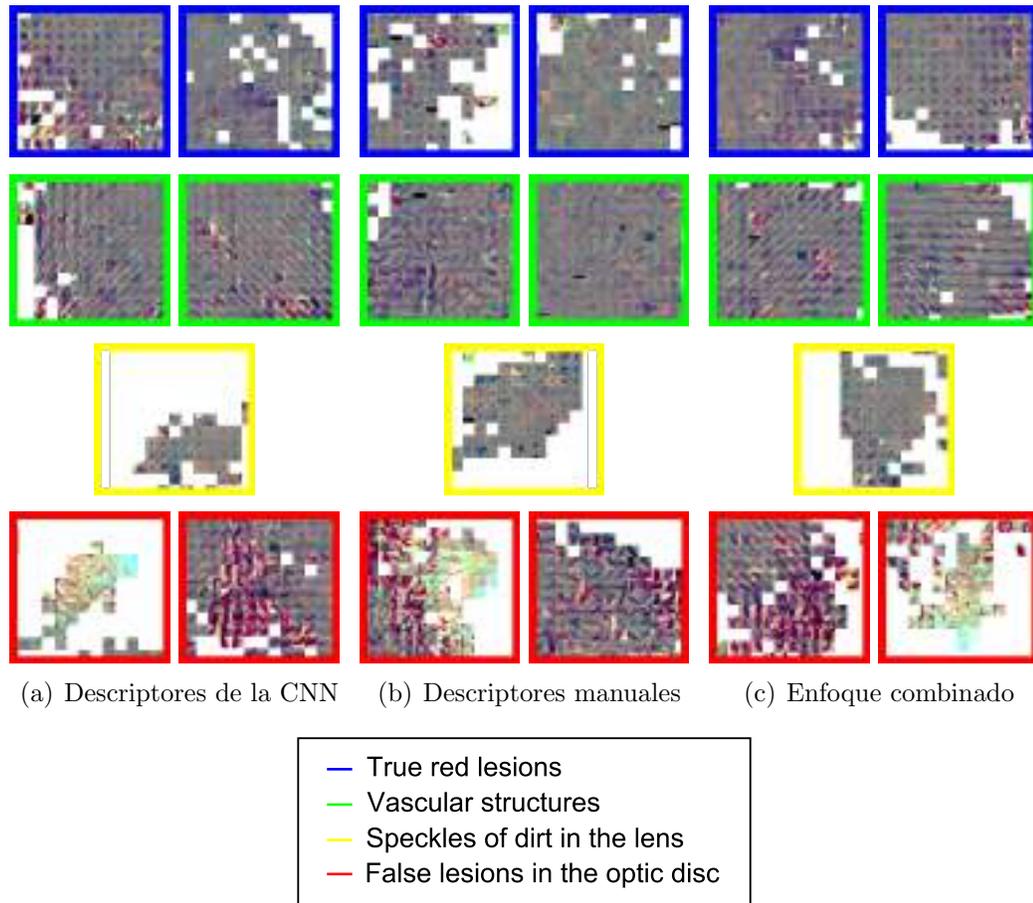


Figura 5.15: Detalles extraídos de la visualización  $t$ -SNE de la Figura 5.14 para diferentes tipos de lesiones rojas candidatas (lesiones reales, estructuras vasculares, suciedad en el lente de la cámara y detecciones falsas en curvas de los vasos en el disco óptico).

ques para *screening* de RD o para la detección de pacientes referibles. Cabe señalar, además, que los métodos incluidos en la Tabla 5.6 están basados no solamente en la detección de lesiones rojas [16, 70, 180] sino también en otras funcionalidades adicionales que van desde el estudio de la calidad de las imágenes [176] y la presencia o ausencia de otras estructuras patológicas como exudados o neovascularizaciones [43, 161, 176]. Comparado con respecto a todos estos otros enfoques, el método aquí propuesto alcanza mayores valores de AUC. Más aún, obtiene mejores resultados que el método para *grading* de RD presentado en [205], el cual se basa en CNNs reajus-

tadas a partir de un conjunto de 50.000 imágenes de entrenamiento con etiquetas de imagen. Por otro lado, se obtuvo una performance a la del método de Quellec *et al.* [166] similar para *screening* de RD, el cual reporta un AUC= 0,893 al evaluarse en MESSIDOR. Dicho enfoque utiliza múltiples imágenes por paciente, información de contexto y registros clínicos para aprender reglas de diagnóstico a partir de un conjunto de 12.000 exámenes. El método presentado en este capítulo, por el contrario, es capaz de alcanzar un AUC algo mayor sin incluir ningún tipo de información clínica adicional que la extraída a partir de las imágenes. Además, se obtuvo un valor de *Se* competitivo con respecto al del experto B [176], indicando que el enfoque presentado puede igualar la capacidad de un observador humano para detectar RD o pacientes referibles. Así, este sistema para la detección automática de lesiones rojas podría integrarse en una plataforma para *screening* de RD para mejorar su capacidad para detectar pacientes que padecen esta enfermedad. Por otra parte, también es posible incorporar otros módulos para detectar otras estructuras patológicas distintas para mejorar los resultados reportados.

Es importante destacar que todas las etapas del método propuesto tienen parámetros que son ajustados automáticamente para cada resolución de imagen. Sus valores, reportados oportunamente en la Sección 5.3.2, fueron escogidos empíricamente utilizando datos diferentes a los utilizados para la evaluación final, y proporcionalmente ajustados en los experimentos siguientes para compensar eventuales cambios en la resolución de las imágenes. Este enfoque simple provee una invariancia aproximada respecto a cambios de escala que es útil para facilitar la adaptabilidad del método para ser aplicado sobre imágenes obtenidas utilizando diferentes cámaras.

Al estudiar cada enfoque de caracterización individualmente, es posible observar en el experimento 1 que tanto el RF entrenado con los descriptores diseñados manualmente como la CNN alcanzaron mayores sensibilidades por lesión que las reportadas por el método descrito en [180] ( $p < 2 \times 10^{-18}$  y  $p < 2 \times 10^{-4}$ , respectivamente). Esto probablemente se debe al hecho de que el método propuesto difiere en la extracción de candidatos respecto al utilizado por el enfoque alternativo. Además, [180] elimina las lesiones candidatas que ocurren dentro de un área estimada en torno al centro del disco óptico, que se determina utilizando un método automático. En consecuencia, si el diámetro del disco óptico es accidentalmente sobreestimado por tal estrategia,

algunos candidatos que ocurran sobre regiones válidas serán suprimidos y no será posible recuperarlos luego durante la etapa de clasificación. Como se ve en las Figuras 5.14 y 5.15, el enfoque integrado es capaz de discriminar los candidatos dentro del disco óptico y en las estructuras vasculares. Así, en lugar de recurrir a una fase rígida de eliminación basada en la segmentación del disco óptico, es factible dejar al clasificador para que decida si el candidato es realmente una lesión o es un falso positivo en una región anatómica. Este enfoque eleva la máxima sensibilidad por lesión alcanzable, permitiendo entrenar el clasificador sobre una mayor cantidad de falsos positivos y obtener una mejor sensibilidad en tiempo de test. Una observación similar puede realizarse sobre los resultados del experimento 2, en el que los descriptores diseñados manualmente y el enfoque basado en *deep learning* obtuvieron mayores sensibilidades por lesión que las reportadas en [217]. Debe señalarse, además, que el método en [217] fue entrenado utilizando la primera mitad de las imágenes patológicas de e-ophtha y evaluadas sobre la segunda mitad, en lugar de entrenado en un conjunto de imágenes separado y evaluado sobre el conjunto completo, como en el caso de este trabajo. Además, vale la pena indicar que las imágenes de pacientes sanos dentro de e-ophtha también fueron incluidas durante el proceso de evaluación, de forma tal que sea posible obtener una estimación más precisa de la performance del método en un escenario clínico real.

En la evaluación a nivel de imagen, es posible observar que los enfoques individuales entrenados en el experimento 1 no son capaces de alcanzar valores de AUC mayores que los reportados por [180] (Tabla 5.6). Esto puede deberse a que, como indican los autores, su método es más preciso para detectar HEs circulares y MAs que HEs con otras formas. Las imágenes de MESSIDOR fueron originalmente etiquetadas como pertenecientes a las categorías R0 y R1 teniendo en cuenta la cantidad de MAs (Tabla 5.3) [49]. Así, ser más preciso en la detección de MAs resultará en una mejor habilidad para distinguir las etapas más tempranas de la RD. Cuando se utilizan individualmente los descriptores ajustados manualmente o los aprendidos por la CNN, ambos métodos son menos precisos para detectar MAs pero mejores para discriminar otras HEs. Este argumento está soportado por la evidencia presentada en la Figura 5.8, en la que es posible ver que los valores de sensibilidad por lesión obtenidos para detección de MA son menores que los reportados para HEs. Más aún,

se observó que, para detección de MAs, la CNN tiene un rendimiento igual o superior al del RF entrenado sobre los descriptores diseñados manualmente, para valores bajos de FPI. Esto explica el comportamiento observado en la Figura 5.11, donde las probabilidades de la CNN alcanzan un mayor valor de AUC para *screening* de RD y para detección de pacientes referibles. No obstante, la combinación de ambos enfoques con el RF mejora consistentemente sus rendimientos individuales, alcanzando una mejor caracterización de las MAs (observada en las mejoras reportadas en la Figura 5.8(a)) y, en consecuencia, una mejor discriminación de los pacientes con RD.

## 5.6. Conclusiones

En este capítulo se propuso un método novedoso para la detección de lesiones rojas en fotografías de fondo de ojo basado en un vector híbrido de descriptores diseñados manualmente y de otros aprendidos por *deep learning*. Se entrenó para ello una CNN utilizando entornos alrededor de lesiones candidatas, con el objetivo de aprender sus características automáticamente, y se complementó a los descriptores resultantes con otros ad-hoc, basados en el conocimiento del dominio. Los resultados obtenidos en una serie de conjuntos de datos estándar demostraron empíricamente que el sistema resultante es capaz de alcanzar un nuevo estado del arte en el dominio, y que la combinación de ambas fuentes de información provee mejoras estadísticamente significativas en comparación con el uso de cada algoritmo de forma separada. Un comportamiento similar fue observado al evaluar el algoritmo propuesto como una herramienta para detección de RD o de pacientes referibles, reportando valores de AUC más altos que los obtenidos por otros enfoques basados no solamente en detectar lesiones rojas sino también en analizar otras patologías tales como las lesiones brillantes o las neovascularizaciones, o incluso aprendiendo clasificadores utilizando información clínica adicional. Considerando el alto costo de etiquetar manualmente fotografías de fondo de ojo a nivel lesión, el método aquí propuesto representa una alternativa robusta para mejorar la performance de otros enfoques basados en *deep learning*.

---

# Capítulo 6

## Conclusiones y trabajos futuros

### 6.1. Contribuciones

En esta tesis se presentaron diversas contribuciones al análisis automatizado de fotografías de fondo de ojo, aplicadas al diagnóstico asistido por computadora de RD y glaucoma. Estas dos enfermedades constituyen las principales causas de ceguera prevenible a nivel mundial, y su detección utilizando imágenes de fondo de ojo es valiosa teniendo en cuenta el bajo costo y la naturaleza no invasiva de esta modalidad de imagen.

En el Capítulo 3 se presentó un método novedoso para la segmentación de vasos sanguíneos de la retina basado en un campo condicional aleatorio totalmente conectado (*fully connected conditional random field model*, FC-CRF), entrenado de manera discriminativa utilizando una máquina de vectores de soporte de salida estructurada (*structured output support vector machine*, SOSVM). En general, es difícil incorporar información referente a la morfología de las estructuras deseadas en algoritmos de segmentación supervisada. Por otra parte, las estrategias comúnmente utilizadas, tales como los modelos de Potts o variación total, usualmente fallan al momento de detectar estructuras elongadas. En el capítulo se demostró empíricamente que los potenciales apareados (o *pairwise*) totalmente conectados de los FC-CRF son más expresivos, permitiendo recuperar representaciones más precisas de las vasculatura retiniana, que incluyen vasos delgados que son ignorados normalmente por los enfoques convencionales basados en vecindarios locales. El modelo totalmente conectado

fue entrenado utilizando una SOSVM, algo que era infactible previamente debido a la alta complejidad temporal de la inferencia de CRFs en grafos totalmente conectados. Mediante la utilización del modelo de energía y el enfoque de inferencia propuestos en [101], se demostró la capacidad de entrenar el FC-CRF y aprender simultáneamente los pesos para los potenciales unarios y *pairwise*. Se realizó una evaluación extensiva y una comparación detallada con otros trabajos, utilizando conjuntos estándar de imágenes tales como DRIVE, STARE, CHASEDB1 y HRF, que demostraron que el método aquí propuesto supera a diversos enfoques existentes y alcanza los mejores resultados del estado del arte al medirse en términos del coeficiente de correlación de Matthews, el valor-F y la media-G. Además, el método fue utilizado también en las demás contribuciones de esta tesis, demostrando su habilidad para tratar con imágenes no solamente de pacientes sanos si no también de pacientes afectados por RD y flaucoma, requiriendo un mínimo de postprocesamiento para remover artefactos que ocurren en las lesiones.

Se introdujo, por otra parte, una estrategia para la detección de glaucoma basada en la transferencia de redes neuronales convolucionales (CNNs), descrita detalladamente en el Capítulo 4. Debido a que existen pocos conjuntos de datos para el diseño de métodos para la detección de glaucoma, con apenas unas pocas imágenes cada uno, todavía no es factible entrenar CNNs desde cero para resolver esta tarea. Con el objetivo de mitigar los efectos de este inconveniente, se propuso caracterizar a los pacientes con y sin glaucoma utilizando descriptores obtenidos a partir de una CNN preentrenada sobre datos no médicos. Esta red fue alimentada con imágenes ajustadas utilizando métodos de preprocesamiento tradicionales en la literatura. Los experimentos realizados demostraron un rendimiento competitivo con respecto a otras estrategias evaluadas sobre el mismo conjunto de imágenes, DRISHTI-GS1. Por otro lado, se observó empíricamente que remover la vasculatura retiniana utilizando inpainting digital reduce la calidad de los resultados, lo que en principio puede indicar que los vasos sanguíneos proveen información valiosa para el diagnóstico de esta enfermedad.

Finalmente, en el Capítulo 5 se presentó un enfoque híbrido para *screening* de RD basado en la detección de lesiones rojas. Los métodos actuales para diagnóstico asistido por computadora de la RD por lo general requieren de la detección de

microaneurismas (MAs) o hemorragias (HEs) pequeñas que, a pesar de ser el signo más temprano de la enfermedad, requieren un enorme esfuerzo por parte de los especialistas para ser detectadas. En particular, se observó que los enfoques actuales se basan en dos etapas: una inicial de detección de candidatos y otra posterior en la que los falsos positivos son removidos por medio de un clasificador supervisado, entrenado previamente utilizando descriptores diseñados manualmente. Se propuso mejorar este enfoque preexistente incorporando una serie de descriptores aprendidos por *deep learning*, y obtenidos utilizando una CNN entrenada de cero sobre parches en torno a lesiones candidatas. El vector híbrido de descriptores obtenido mediante la combinación de las características tradicionalmente utilizadas y las diseñadas en este trabajo fue posteriormente utilizado para entrenar un Bosque Aleatorio o Random Forest, conocido por su buen desempeño en problemas de clase sesgada y con datos de dimensiones muy altas. La validación experimental sobre una serie de conjuntos de imágenes estándar como DIARETDB1, e-ophtha y MESSIDOR demostró que el enfoque propuesto es capaz de alcanzar un rendimiento consistentemente más alto que el de las estrategias actuales, llegando a obtener incluso una mejor performance para detección de RD que la de un segundo observador experto.

## 6.2. Líneas futuras de trabajo

Diversos trabajos futuros pueden derivarse a partir de las contribuciones realizadas en esta tesis. Algunos de ellos se discuten en las siguientes subsecciones.

### 6.2.1. Segmentación y caracterización de vasos sanguíneos

El método de segmentación de vasos sanguíneos presentado en esta tesis permite alcanzar representaciones precisas de los vasos sanguíneos en imágenes de pacientes sanos, glaucomatosos o en los primeros estadios de la RD. Al evaluarse sobre sujetos con RD avanzada, sin embargo, ocurre un mayor número de falsos positivos en áreas donde existen hemorragias de gran tamaño o exudados. Este inconveniente fue parcialmente resuelto mediante la aplicación de técnicas de postprocesamiento como las descritas en el Capítulo 5, aunque esta estrategia fuerza en algunos casos la pérdida de vasos sanguíneos más delgados. Los artefactos que aparecen en las segmentacio-

nes se deben a la capacidad discriminativa de los descriptores utilizados, que suelen verse afectados por la presencia de lesiones oscuras o por los patrones elongados que eventualmente pueden ocurrir entre lesiones brillantes, por ejemplo. Una opción para mitigar este inconveniente puede ser segmentar de manera inicial estas estructuras, utilizar técnicas de *inpainting* digital para suprimirlas de las imágenes, y luego aplicar el algoritmo de segmentación de vasos sanguíneos. Este enfoque fue utilizado en [14], por ejemplo, en donde la supresión de exudados mejora la capacidad discriminativa de descriptores basados en el Hessiano. Este enfoque puede aplicarse en el contexto del enfoque de segmentación propuesto en esta tesis, y puede contribuir a la eliminación de estas detecciones falsas. Otra alternativa es entrenar redes neuronales convolucionales y utilizar la respuesta a los filtros aprendidos por la red como entrada para el FC-CRF. Esta idea fue tenida en cuenta recientemente por algunos autores, motivadas por los resultados obtenidos en este trabajo. En [66, 67], Fu *et al.* propusieron entrenar de manera simultánea un FC-CRF y una CNN, alcanzando un mejor rendimiento en DRIVE y STARE en presencia de estructuras patológicas.

Otra dificultad relacionada con el diseño manual de los descriptores se relaciona con el esfuerzo intensivo que supone su adaptación a otras resoluciones de imagen. Esta dificultad podría resolverse utilizando un enfoque diferente al empleado en este trabajo para el escalado de los parámetros de los descriptores, como el propuesto recientemente en [206]. Por medio de un enfoque simple de aprendizaje automático tal como es la regresión lineal, es posible diseñar diferentes curvas que permiten escalar los parámetros de mejor manera que utilizando un factor de escala. Los enfoques basados en *deep learning* también sufren inconvenientes similares en cuanto a la parametrización de sus entradas, lo que explica en parte por qué estas técnicas usualmente se evalúan en imágenes de baja resolución como las de DRIVE y STARE. Si se tiene en cuenta que la mayor de las dificultades al momento de segmentar vasos sanguíneos se relaciona con la detección de los vasos más delgados (que son más evidentes en resoluciones altas), es necesario comenzar a enfocarse en la utilización de conjuntos de imágenes de mayor tamaño, como las de HRF o las del recientemente introducido DRHAGIS [81].

Un problema todavía pendiente de resolver de manera precisa es el de la clasificación automática de los vasos sanguíneos de la retina en arterias y venas (A/V),

requerida para la caracterización de procesos vasculares asociados a la presencia de diferentes enfermedades [2]. Esta tarea es compleja y difícil de abordar debido a que las diferencias visuales entre ambas clases de capilares no son evidentes. Se han propuesto numerosos métodos para la caracterización de tales diferencias por medio de descriptores diseñados manualmente. Una alternativa para brindar solución a este problema es la de combinar estos indicadores con otros filtros para la detección de vasos sanguíneos, y realizar de manera simultánea tanto la segmentación como la clasificación de los vasos en arterias y venas. Teniendo en cuenta que tanto las SOSVMs como los FC-CRFs son lo suficientemente generales como para ser utilizados en clasificación binaria y multiclase, el método de segmentación presentado en esta tesis puede adaptarse fácilmente a esta nueva tarea. La principal desventaja de este enfoque es, nuevamente, el carácter artesanal de los descriptores existentes para clasificación A/V, que puede no ser suficiente para discriminar estas dos clases. Esta dificultad puede mitigarse parcialmente mediante la utilización de descriptores aprendidos mediante *deep learning*.

### 6.2.2. Detección automatizada de glaucoma

La detección de glaucoma a partir de fotografías de fondo de ojo es extremadamente requerida aún, debido al gran número de pacientes no diagnosticados que existen [164]. Actualmente, existen diferentes iniciativas que proponen aplicar técnicas de aprendizaje profundo para detectar la enfermedad a partir de tomografías de coherencia óptica (OCTs), siendo quizá la más notoria de ellas la que surge de la colaboración entre Google DeepMind y el Moorfields Eye Hospital<sup>1</sup>. Sin embargo, se observa un interés menor en la aplicación de esta familia de métodos sobre fotografías de fondo de ojo, algo que debe considerarse teniendo en cuenta el bajo costo de los dispositivos de captura y su aplicabilidad en campañas de *screening* de poblaciones. Su principal complicación se relaciona con la falta de conjuntos de imágenes a gran escala para el entrenamiento de modelos basados en *deep learning*. Los conjuntos de datos disponibles en la actualidad están particularmente diseñados para la segmentación del disco óptico, por ejemplo, o han sido etiquetados basándose únicamente

---

<sup>1</sup><https://www.theguardian.com/technology/2016/jul/05/google-deepmind-nhs-machine-learning-blindness>

en la medición de la proporción copa-disco, que, aunque útil, no es suficiente para realizar un diagnóstico de glaucoma preciso [75].

El método propuesto en esta tesis para detección de glaucoma ha demostrado un rendimiento parejo con respecto al obtenido por otros enfoques ajustados manualmente, utilizando simplemente descriptores aprendidos por una CNN sin utilizar datos médicos. De existir bases de datos con una mayor cantidad de imágenes, una continuación inmediata de este trabajo podría ser intentar reajustar los pesos de la red mediante la inicialización de la CNN con los valores de la red preentrenada, y continuar a partir de allí el proceso de optimización, pero sobre los datos nuevos. Otra estrategia potencial podría ser intentar entrenar de cero una CNN utilizando conjunto de fotografías de fondo de ojo con otras enfermedades, como las pertenecientes al campeonato Kaggle para detección de RD<sup>2</sup>, y luego reajustar estos pesos para detección de glaucoma. Esta configuración no sólo permitiría acelerar la convergencia de la red, si no además permitiría aprender filtros más asociados con las propiedades de las imágenes de fondo de ojo que los aprendidos a partir de imágenes naturales.

Finalmente, un trabajo interesante para derivar de la contribución aquí propuesta es el desarrollo de un método combinado para detección de glaucoma, basado en información de múltiples fuentes. La probabilidad dada por el clasificador por regresión logística empleado en el Capítulo 4 puede utilizarse como parte de un nuevo vector de descriptores compuesto por otros parámetros interesantes tales como la presión intraocular o el registro médico del paciente, por ejemplo. Mediante una estrategia de estas características sería posible capturar patrones representativos en los datos que puede ser útiles para alcanzar diagnósticos más precisos.

### 6.2.3. Detección automática de RD

El método para detección de lesiones rojas propuesto en el Capítulo 5 tiene un impacto directo en el desarrollo de futuras herramientas para el diagnóstico asistido por computadora de RD, ya que ha demostrado su habilidad para reconocer este tipo de lesiones con una precisión muy significativa. Sin embargo, aún queda lugar para incorporar mejoras. La principal de ellas es la incorporación de otras técnicas

---

<sup>2</sup><https://kaggle.com/c/diabetic-retinopathy-detection>

similares pero para la detección de lesiones brillantes tales como exudados o lesiones algodonosas, que son también relevantes en el contexto clínico.

Para detectar RD automáticamente, se observó que el método presentado en esta tesis es capaz de alcanzar también un alto rendimiento. La estratificación automática de los diferentes estadios de la enfermedad, sin embargo, resulta más dificultosa, dado que es necesario no solamente detectar MAs y HEs sino también otras lesiones tales como exudados o neovascularizaciones. Una mejora potencial al método propuesto es entrenar una CNN utilizando el conjunto de imágenes de Kaggle (compuesto por unas 90.000 fotografías de fondo de ojo), e integrar su salida con la salida del algoritmo presentado en esta tesis. Recientemente, diferentes autores han conseguido resultados prometedores sobre estas imágenes [73]. Sin embargo, y debido a que las CNNs por lo general trabajan sobre imágenes cuyo tamaño debe ser previamente reducido, éstas suelen ser capaces de identificar lesiones grandes pero no MAs o HEs pequeñas, que, como se observó en el Capítulo 5, son sutiles y difíciles de distinguir a simple vista. Esta característica refuerza la idea de que integrar ambos enfoques permitiría en principio mejorar los resultados actuales.

Alternativamente pueden plantearse nuevos descriptores basados en la distribución de las probabilidades que el clasificador asocia a las lesiones rojas candidatas, y utilizarlos para detectar o estratificar casos de RD. En el conjunto de imágenes MESSIDOR, por ejemplo, las etiquetas fueron originalmente asignadas contando la cantidad de MAs y HEs, y validando la presencia y/o ausencia de neovascularizaciones. Otros descriptores tales como histogramas representando la distribución de las salidas del Bosque Aleatorio, por ejemplo, podrían reducir el efecto que algunos *outliers* obtenidos durante la etapa de detección de lesiones rojas pueda introducir en el diagnóstico automático. Además, podría emplearse un enfoque similar al utilizado en el Capítulo 4 para detección de glaucoma, y utilizarlo para caracterizar casos leves o pre-proliferativos, en los que la presencia de HEs de gran tamaño afecta la distribución global de las intensidades en las imágenes y, en consecuencia, en la salida de la CNN preentrenada. Finalmente, técnicas de caracterización vascular tales como el análisis de su dimensión fractal podrían también integrarse para detectar los casos proliferativos.

---

# Bibliografía

- [1] A\*. SIVA brochure. <https://www.etpl.sg/qq1/slot/u94/Software%20to%20License/SIVA/SIVA%20ebrochure.pdf>. Accessed: 2017-03-23.
- [2] M. D. Abràmoff, M. K. Garvin, y M. Sonka. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208, 2010.
- [3] M. D. Abràmoff y M. Niemeijer. Mass screening of diabetic retinopathy using automated methods. En *Teleophthalmology in Preventive Medicine*, págs. 41–50. Springer, 2015.
- [4] R. Acharya, C. K. Chua, E. Ng, W. Yu, y C. Chee. Application of higher order spectra for the identification of diabetes retinopathy stages. *Journal of Medical Systems*, 32(6):481–488, 2008.
- [5] U. R. Acharya, E. Y.-K. Ng, J.-H. Tan, S. V. Sree, y K.-H. Ng. An integrated index for the identification of diabetic retinopathy stages using texture parameters. *Journal of Medical Systems*, 36(3):2011–2020, 2012.
- [6] M. Agudelo-Botero y C. A. Dávila-Cervantes. Carga de la mortalidad por diabetes mellitus en América latina 2000-2011: los casos de Argentina, Chile, Colombia y México. *Gaceta Sanitaria*, 29(3):172–177, 2015.
- [7] C. Agurto, S. Murillo, V. Murray, M. Pattichis, S. Russell, M. Abramoff, y P. Soliz. Detection and phenotyping of retinal disease using AM-FM processing for feature extraction. En *42nd Asilomar Conference on Signals, Systems and Computers*, págs. 659–663. IEEE, 2008.

- 
- [8] B. Al-Diri, A. Hunter, y D. Steel. An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical Imaging*, 28(9):1488–1497, 2009.
- [9] B. Al-Diri, A. Hunter, D. Steel, y M. Habib. Automated analysis of retinal vascular network connectivity. *Computerized Medical Imaging and Graphics*, 34(6):462–470, 2010.
- [10] M. Al-Rawi, M. Qutaishat, y M. Arrar. An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine*, 37(2):262–267, 2007.
- [11] N. Amerasinghe, T. Aung, N. Cheung, C. W. Fong, J. J. Wang, P. Mitchell, S.-M. Saw, y T. Y. Wong. Evidence of retinal vascular narrowing in glaucomatous eyes in an Asian population. *Investigative Ophthalmology & Visual Science*, 49(12):5397–5402, 2008.
- [12] M. A. Amin y H. Yan. High speed detection of retinal blood vessels in fundus image using phase congruency. *Soft Computing*, 15(6):1217–1230, 2011.
- [13] M. M. Angelica, A. Sanseau, y C. Argento. Arterial narrowing as a predictive factor in glaucoma. *International Ophthalmology*, 23(4):271–274, 2001.
- [14] R. Annunziata, A. Garzelli, L. Ballerini, A. Mecocci, y E. Trucco. Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1129–1138, 2016.
- [15] R. Annunziata y E. Trucco. Accelerating convolutional sparse coding for curvilinear structures segmentation by refining SCIRD-TS filter banks. *IEEE Transactions on Medical Imaging*, 35(11):2381–2392, 2016.
- [16] B. Antal y A. Hajdu. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *IEEE Transactions on Biomedical Engineering*, 59(6):1720–1726, 2012.

- [17] O. Arend, A. Remky, N. Plange, B. Martin, y A. Harris. Capillary density and retinal diameter measurements and their impact on altered retinal circulation in glaucoma: a digital fluorescein angiographic study. *British Journal of Ophthalmology*, 86(4):429–433, 2002.
- [18] A. Argyriou, R. Foygel, y N. Srebro. Sparse prediction with the  $k$ -support norm. En *Advances in Neural Information Processing Systems*, págs. 1457–1465. 2012.
- [19] E. Arlington. <http://www.eyeglasses-arlington.com/diabetic-retinopathy.html>. <http://www.eyeglasses-arlington.com/diabetic-retinopathy.html>. Accessed: 2017-03-23.
- [20] G. Azzopardi, N. Strisciuglio, M. Vento, y N. Petkov. Trainable COSFIRE filters for vessel delineation with application to retinal images. *Medical Image Analysis*, 19(1):46–57, 2015.
- [21] P. Bankhead, C. N. Scholfield, J. G. McGeown, y T. M. Curtis. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLOS ONE*, 7(3):e32435, 2012.
- [22] Y. Bar, I. Diamant, L. Wolf, y H. Greenspan. Deep learning with non-medical training used for chest pathology identification. En *SPIE Medical Imaging*, págs. 94140V–94140V. International Society for Optics and Photonics, 2015.
- [23] R. Barrenechea, I. d. la Fuente, R. G. Plaza, N. Flores, L. Segovia, Z. Villagómez, E. E. Camarero, L. C. Zepeda-Romero, V. C. Lansingh, H. Limburg, et al. Encuesta nacional de ceguera y deficiencia visual evitable en Argentina, 2013. *Pan American Journal of Public Health*, 37(1), 2015.
- [24] E. J. Bekkers, J. Zhang, R. Duits, y B. M. ter Haar Romeny. Curvature based biomarkers for diabetic retinopathy via exponential curve fits in SE(2). En E. Trucco, X. Chen, Garvin, M. K., J. J. Liu, y X. Y. Frank, eds., *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA*. 2015.

- [25] Y. Bengio, I. J. Goodfellow, y A. Courville. Deep learning. *Nature*, 521:436–444, 2015.
- [26] R. Bock, J. Meier, G. Michelson, L. Nyúl, y J. Hornegger. Classifying glaucoma with image-based features from fundus photographs. *Pattern Recognition*, págs. 355–364, 2007.
- [27] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, y G. Michelson. Glaucoma risk index: automated glaucoma detection from color fundus images. *Medical Image Analysis*, 14(3):471–481, 2010.
- [28] Y. Boykov y V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.
- [29] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [30] A. Budai et al. Robust vessel segmentation in fundus images. *International Journal of Biomedical Imaging*, 2013, 2013.
- [31] T. Chakraborti, D. K. Jha, A. S. Chowdhury, y X. Jiang. A self-adaptive matched filter for retinal blood vessel detection. *Machine Vision and Applications*, 26(1):55–68, 2014.
- [32] A. Chakravarty y J. Sivaswamy. Glaucoma classification with a fusion of segmentation and image-based features. En *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, tomo 1, págs. 689–692. IEEE, 2016.
- [33] K. Chatfield, K. Simonyan, A. Vedaldi, y A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. En *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014. doi: <http://dx.doi.org/10.5244/C.28.6>.
- [34] X. Chen, Y. Xu, D. W. K. Wong, T. Y. Wong, y J. Liu. Glaucoma detection based on deep convolutional neural network. En *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, págs. 715–718. IEEE, 2015.

- [35] X. Chen, Y. Xu, S. Yan, D. W. K. Wong, T. Y. Wong, y J. Liu. Automatic feature learning for glaucoma detection based on deep learning. En *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, págs. 669–677. Springer, 2015.
- [36] E. Cheng, L. Du, Y. Wu, Y. J. Zhu, V. Megalooikonomou, y H. Ling. Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features. *Machine Vision and Applications*, 25(7):1779–1792, 2014.
- [37] C. Y.-I. Cheung, E. Lamoureux, M. K. Ikram, M. B. Sasongko, J. Ding, Y. Zheng, P. Mitchell, J. J. Wang, y T. Y. Wong. Retinal vascular geometry in asian persons with diabetes and retinopathy. *Journal of Diabetes Science and Technology*, 6(3):595–605, 2012.
- [38] N. Cheung, K. C. Donaghue, G. Liew, S. L. Rogers, J. J. Wang, S.-W. Lim, A. J. Jenkins, W. Hsu, M. L. Lee, y T. Y. Wong. Quantitative assessment of early diabetic retinopathy using fractal analysis. *Diabetes Care*, 32(1):106–110, 2009.
- [39] F. Ciompi, B. de Hoop, S. J. van Riel, K. Chung, E. T. Scholten, M. Oudkerk, P. A. de Jong, M. Prokop, y B. van Ginneken. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Medical Image Analysis*, 26(1):195–202, 2015.
- [40] T. A. Ciulla, A. G. Amador, y B. Zinman. Diabetic retinopathy and diabetic macular edema. *Diabetes Care*, 26(9):2653–2664, 2003.
- [41] V. L. Clark y J. A. Kruse. Clinical methods: the history, physical, and laboratory examinations. *JAMA*, 264(21):2808–2809, 1990.
- [42] W. Commons. Schematic diagram of the human eye: Wikimedia commons. [https://commons.wikimedia.org/wiki/File:Schematic\\_diagram\\_of\\_the\\_human\\_eye\\_en.svg](https://commons.wikimedia.org/wiki/File:Schematic_diagram_of_the_human_eye_en.svg). Accessed: 2017-03-23.
- [43] J. Costa, I. Sousa, y F. Soares. Smartphone-based decision support system for elimination of pathology-free images in diabetic retinopathy screening. En

- International Conference on IoT Technologies for HealthCare*, págs. 83–88. Springer, 2016.
- [44] R. Crosby-Nwaobi, L. Z. Heng, y S. Sivaprasad. Retinal vascular calibre, geometry and progression of diabetic retinopathy in type 2 diabetes mellitus. *International Journal of Ophthalmology*, 228(2):84–92, 2011.
- [45] J. Cuadros y G. Bresnick. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *Journal of Diabetes Science and Technology*, 3(3):509–516, 2009.
- [46] P. Dai, H. Luo, H. Sheng, Y. Zhao, L. Li, J. Wu, Y. Zhao, y K. Suzuki. A new approach to segment both main and peripheral retinal vessels based on gray-voting and gaussian mixture model. *PLOS ONE*, 10(6):e0127748, 2015.
- [47] D. C. DeBuc. The role of retinal imaging and portable screening devices in tele-ophthalmology applications for diabetic retinopathy management. *Current Diabetes Reports*, 16(12):132, 2016.
- [48] E. Decencière, G. Cazuguel, X. Zhang, G. Thibault, J.-C. Klein, F. Meyer, B. Marcotegui, G. Quellec, M. Lamard, R. Danno, et al. Teleophtha: Machine learning and image processing methods for teleophthalmology. *IRBM*, 34(2):196–203, 2013.
- [49] E. Decencière, X. Zhang, G. Cazuguel, B. Laÿ, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014.
- [50] S. Dua, U. R. Acharya, P. Chowriappa, y S. V. Sree. Wavelet-based energy features for glaucomatous image classification. *IEEE Transactions on Information Technology in Biomedicine*, 16(1):80–87, 2012.
- [51] B. Dupas, T. Walter, A. Erginay, R. Ordonez, N. Deb-Joardar, P. Gain, J.-C. Klein, y P. Massin. Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a

- computer-assisted diagnostic system for grading diabetic retinopathy. *Diabetes & Metabolism*, 36(3):213–220, 2010.
- [52] M. Esmaeili et al. A new curvelet transform based method for extraction of red lesions in digital color retinal images. En *17th IEEE International Conference on Image Processing (ICIP)*, págs. 4093–4096. IEEE, 2010.
- [53] L. Espona, M. J. Carreira, M. Penedo, y M. Ortega. Retinal vessel tree segmentation using a deformable contour model. En *19th International Conference on Pattern Recognition*, págs. 1–4. IEEE, 2008.
- [54] C. F. Etienne. Reducing avoidable blindness and visual impairment in the region of the Americas. *Revista Panamericana de Salud Pública*, 37(1):1–3, 2015.
- [55] K. Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- [56] A. Fathi y A. R. Naghsh-Nilchi. Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. *Biomedical Signal Processing and Control*, 8(1):71–80, 2013.
- [57] O. Faust, R. Acharya, E. Y.-K. Ng, K.-H. Ng, y J. S. Suri. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. *Journal of Medical Systems*, 36(1):145–157, 2012.
- [58] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [59] A. Fercher y E. Roth. Ophthalmic laser interferometry. *SPIE Milestone Series*, 165:242–245, 2001.
- [60] F. Fiorani y U. Schurr. Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, 64:267–291, 2013.
- [61] M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, C. Owen, A. Rudnicka, y S. Barman. Retinal vessel extraction using first-order derivative of gaussian and

- morphological processing. En *Advances in Visual Computing*, págs. 410–420. Springer, 2011.
- [62] M. M. Fraz, A. Basit, y S. Barman. Application of morphological bit planes in retinal blood vessel extraction. *Journal of Digital Imaging*, 26(2):274–286, 2013.
- [63] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, y S. A. Barman. Blood vessel segmentation methodologies in retinal images—a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407–433, 2012.
- [64] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, y S. A. Barman. An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548, 2012.
- [65] M. M. Fraz, A. R. Rudnicka, C. G. Owen, y S. A. Barman. Delineation of blood vessels in pediatric retinal images using decision trees-based ensemble classification. *International Journal of Computer Assisted Radiology and Surgery*, 9(5):795–811, 2014.
- [66] H. Fu, Y. Xu, S. Lin, D. W. K. Wong, y J. Liu. DeepVessel: Retinal vessel segmentation via deep learning and conditional random field. En *International Conference on Medical Image Computing and Computer-Assisted Intervention*, págs. 132–139. Springer, 2016.
- [67] H. Fu, Y. Xu, D. W. K. Wong, y J. Liu. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. En *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, págs. 698–701. IEEE, 2016.
- [68] G. Gardner, D. Keating, T. Williamson, y A. Elliott. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British Journal of Ophthalmology*, 80(11):940–944, 1996.

- [69] R. Gargeya y T. Leng. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 2017.
- [70] L. Giancardo, T. P. Karnowski, K. W. Tobin, F. Meriaudeau, y E. Chaum. Validation of microaneurysm-based diabetic retinopathy screening across retina fundus datasets. En *IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS)*, págs. 125–130. IEEE, 2013.
- [71] I. Goodfellow, Y. Bengio, y A. Courville. *Deep learning*. MIT Press, 2016.
- [72] J. Grauslund, A. Green, R. Kawasaki, L. Hodgson, A. K. Sjølie, y T. Y. Wong. Retinal vascular fractals and microvascular and macrovascular complications in type 1 diabetes. *Ophthalmology*, 117(7):1400–1405, 2010.
- [73] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, 2016.
- [74] I. Guyon y A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.
- [75] M. S. Haleem, L. Han, J. van Hemert, y B. Li. Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review. *Computerized Medical Imaging and Graphics*, 37(7):581–596, 2013.
- [76] J. A. Hanley y B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, 1982.
- [77] T. Hastie, R. Tibshirani, y J. Friedman. *The elements of statistical learning*, tomo 1. Springer, 2009.
- [78] X. He, R. S. Zemel, y M. Á. Carreira-Perpiñán. Multiscale conditional random fields for image labeling. En *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, tomo 2, págs. II–695. IEEE, 2004.

- [79] M. Helmstaedter. Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nature Methods*, 10(6):501–507, 2013.
- [80] E. M. Hoffmann, L. M. Zangwill, J. G. Crowston, y R. N. Weinreb. Optic disk size and glaucoma. *Survey of Ophthalmology*, 52(1):32–49, 2007.
- [81] S. Holm, G. Russell, V. Nourrit, y N. McLoughlin. DR HAGIS: a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. *Journal of Medical Imaging*, 4(1):014503–014503, 2017.
- [82] A. Hoover, V. Kouznetsova, y M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, 2000.
- [83] F. Huang, B. Dashtbozorg, J. Zhang, E. Bekkers, S. Abbasi-Sureshjani, T. T. Berendschot, y B. M. ter Haar Romeny. Reliability of using retinal vascular fractal dimension as a biomarker in the diabetic retinopathy detection. *Journal of Ophthalmology*, 2016, 2016.
- [84] M. K. Ikram, S. de Voogd, R. C. Wolfs, A. Hofman, M. M. Breteler, L. D. Hubbard, y P. T. de Jong. Retinal vessel diameters and incident open-angle glaucoma and optic disc changes: the Rotterdam study. *Investigative Ophthalmology & Visual Science*, 46(4):1182–1187, 2005.
- [85] M. K. Ikram, Y. T. Ong, C. Y. Cheung, y T. Y. Wong. Retinal vascular caliber measurements: clinical significance, current knowledge and future perspectives. *Ophthalmologica*, 229(3):125–136, 2012.
- [86] M. Jaderberg, A. Vedaldi, y A. Zisserman. Speeding up convolutional neural networks with low rank expansions. *arXiv preprint arXiv:1405.3866*, 2014.
- [87] T. Joachims et al. Cutting-plane training of structural SVMs. *Machine Learning*, 77(1):27–59, 2009.
- [88] J. B. Jonas, W. M. Budde, y S. Panda-Jonas. Ophthalmoscopic evaluation of the optic nerve head. *Survey of Ophthalmology*, 43(4):293–320, 1999.

- [89] J. B. Jonas, X. N. Nguyen, y G. Naumann. Parapapillary retinal vessel diameter in normal and glaucoma eyes. I. morphometric data. *Investigative Ophthalmology & Visual Science*, 30(7):1599–1603, 1989.
- [90] G. D. Joshi, J. Sivaswamy, y S. Krishnadas. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Transactions on Medical Imaging*, 30(6):1192–1205, 2011.
- [91] S. Kadoury, N. Abi-Jaoudeh, y P. A. Valdes. Higher-order CRF tumor segmentation with discriminant manifold potentials. En *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, págs. 719–726. Springer, 2013.
- [92] S. Kadoury et al. Metastatic liver tumour segmentation from discriminant Grassmannian manifolds. *Physics in Medicine and Biology*, 60(16):6459, 2015.
- [93] Kaggle. Diabetic Retinopathy Detection. <https://www.kaggle.com/c/diabetic-retinopathy-detection>, 2015. [Online; accessed 10-April-2013].
- [94] G. B. Kande, P. V. Subbaiah, y T. S. Savithri. Unsupervised fuzzy based vessel segmentation in pathological digital fundus images. *Journal of Medical Systems*, 34(5):849–858, 2010.
- [95] P. L. Kaufman, F. H. Adler, L. A. Levin, y A. Alm. *Adler's Physiology of the Eye*. Elsevier Health Sciences, 2011.
- [96] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, y J. Pietilä. The DIARETDB1 diabetic retinopathy database and evaluation protocol. En *Proceedings of the British Machine Vision Conference*. 2007.
- [97] V. Koh, C. Y.-l. Cheung, Y. Zheng, T. Y. Wong, W. Wong, y T. Aung. Relationship of retinal vascular tortuosity with the neuroretinal rim: The Singapore Malay eye study. *Investigative Ophthalmology & Visual Science*, 51(7):3736–3741, 2010.

- [98] T. Köhler, R. Bock, J. Hornegger, y G. Michelson. Computer-aided diagnostics and pattern recognition: Automated glaucoma detection. En *Teleophthalmology in Preventive Medicine*, págs. 93–104. Springer, 2015.
- [99] R. Kolar, R. P. Tornow, R. Laemmer, J. Odstrcilik, M. A. Mayer, J. Gazarek, J. Jan, T. Kubena, y P. Cernosek. Analysis of visual appearance of retinal nerve fibers in high resolution fundus images: a study on normal subjects. *Computational and Mathematical Methods in Medicine*, 2013, 2013.
- [100] N. Komodakis et al. MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):531–552, 2011.
- [101] P. Krähenbühl y V. Koltun. Efficient inference in fully connected CRFs with Gaussian edge potentials. En *Advances in Neural Information Processing Systems*, págs. 109–117. 2012.
- [102] A. Krizhevsky, I. Sutskever, y G. E. Hinton. Imagenet classification with deep convolutional neural networks. En *Advances in Neural Information Processing Systems*, págs. 1097–1105. 2012.
- [103] S. Kumar y M. Hebert. Discriminative random fields. *International Journal of Computer Vision*, 68(2):179–201, 2006. ISSN 0920-5691. doi: 10.1007/s11263-006-7007-9.
- [104] J. D. Lafferty et al. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. En *Proceedings of the Eighteenth International Conference on Machine Learning*, págs. 282–289. Morgan Kaufmann Publishers Inc., 2001.
- [105] J. L. Leasher, R. R. Bourne, S. R. Flaxman, J. B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, R. A. White, T. Y. Wong, et al. Global estimates on the number of people blind or visually impaired by diabetic retinopathy: A meta-analysis from 1990 to 2010. *Diabetes Care*, 39(9):1643–1649, 2016.
- [106] J. L. Leasher, V. Lansingh, S. R. Flaxman, J. B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, J. C. Silva, R. A. White, et al. Prevalence and causes

- of vision loss in Latin America and the Caribbean: 1990–2010. *British Journal of Ophthalmology*, págs. bjophthalmol–2013, 2014.
- [107] Y. LeCun, Y. Bengio, y G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [108] Y. LeCun, L. Bottou, Y. Bengio, y P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [109] J. Lee, B. C. Y. Zee, y Q. Li. Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy. *PLOS ONE*, 8(12):e75699, 2013.
- [110] M. C. Leske, A. Connell, S.-Y. Wu, L. G. Hyman, y A. P. Schachat. Risk factors for open-angle glaucoma: the Barbados Eye Study. *Archives of Ophthalmology*, 113(7):918–924, 1995.
- [111] K. Li, X. Wu, D. Z. Chen, y M. Sonka. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):119–134, 2006.
- [112] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 3rd ed<sup>ón</sup>., 2009.
- [113] D. Y. Lin, M. S. Blumenkranz, R. J. Brothers, D. M. Grosvenor, y T. D. D. S. Group. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *American Journal of Ophthalmology*, 134(2):204–213, 2002.
- [114] L. Lo Vercio, M. del Fresno, y I. Larrabide. Detection of morphological structures for vessel wall segmentation in IVUS using random forests. En *12th International Symposium on Medical Information Processing and Analysis*, págs. 1016012–1016012. International Society for Optics and Photonics, 2017.

- [115] L. Lo Vercio, J. I. Orlando, M. del Fresno, y I. Larrabide. Assessment of image features for vessel wall segmentation in intravascular ultrasound images. *International Journal of Computer Assisted Radiology and Surgery*, 11(8):1397–1407, 2016.
- [116] C.-K. Lu, T. B. Tang, A. Laude, B. Dhillon, y A. F. Murray. Parapapillary atrophy and optic disc region assessment (PANDORA): retinal imaging tool for assessment of the optic disc and parapapillary atrophy. *Journal of Biomedical Optics*, 17(10):1060101–1060108, 2012.
- [117] C. A. Lupascu, D. Tegolo, y E. Trucco. FABC: retinal vessel segmentation using AdaBoost. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1267–1274, 2010.
- [118] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, y L. Van Gool. Deep retinal image understanding. En *International Conference on Medical Image Computing and Computer-Assisted Intervention*, págs. 140–148. Springer, 2016.
- [119] H. Manterola, L. Lo Vercio, y M. del Fresno. Reducing artifacts impact on IVUS automatic segmentation via inpainting. *Mecánica Computacional*, 33:2703–2716, 2014.
- [120] H. Manterola, M. del Fresno, y I. Larrabide. An analysis of mechanical and computational properties for noninvasive vascular elastography. En *11th International Symposium on Medical Information Processing and Analysis (SIPAIM 2015)*, págs. 968113–968113. International Society for Optics and Photonics, 2015.
- [121] D. Marín, A. Aquino, M. E. Gegúndez-Arias, y J. M. Bravo. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Transactions on Medical Imaging*, 30(1):146–158, 2011.
- [122] M. E. Martínez-Perez, A. Highes, A. V. Stanton, S. A. Thorn, N. Chapman, A. A. Bharath, y K. H. Parker. Retinal vascular tree morphology: a

- semi-automatic quantification. *IEEE Transactions on Biomedical Engineering*, 49(8):912–917, 2002.
- [123] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, y K. H. Parker. Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis*, 11(1):47–61, 2007.
- [124] J. Meier, R. Bock, G. Michelson, L. G. Nyúl, y J. Hornegger. Effects of pre-processing eye fundus images on appearance based glaucoma classification. En *Computer Analysis of Images and Patterns*, págs. 165–172. Springer, 2007.
- [125] J. Meier, R. Bock, L. G. Nyúl, y G. Michelson. Eye fundus image processing system for automated glaucoma classification. En *Proceedings of the 52nd Internationales Wissenschaftliches Kolloquium*. Technische Universität Ilmenau, 2007.
- [126] A. M. Mendonca y A. Campilho. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, 25(9), 2006.
- [127] G. Michelson et al. *Teleophthalmology in Preventive Medicine*. Springer, 2015.
- [128] M. S. Miri y A. Mahloojifar. Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction. *IEEE Transactions on Biomedical Engineering*, 58(5):1183–1192, 2011.
- [129] P. Mitchell, H. Leung, J. J. Wang, E. Rochtchina, A. J. Lee, T. Y. Wong, y R. Klein. Retinal vessel diameter and open-angle glaucoma: the Blue Mountains Eye Study. *Ophthalmology*, 112(2):245–250, 2005.
- [130] S. Mohammad y D. T. Morris. Texture analysis for glaucoma classification. En *International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, págs. 98–103. IEEE, 2015.
- [131] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, y A. Laude. Computer-aided diagnosis of diabetic retinopathy: A review. *Computers in Biology and Medicine*, 43(12):2136–2155, 2013.

- [132] M. Mozaffarieh, M. C. Grieshaber, y J. Flammer. Oxygen and blood flow: players in the pathogenesis of glaucoma. *Molecular Vision*, 14:224–233, 2008.
- [133] C. Muramatsu, Y. Hayashi, A. Sawada, Y. Hatanaka, T. Hara, T. Yamamoto, y H. Fujita. Detection of retinal nerve fiber layer defects on retinal fundus images for early diagnosis of glaucoma. *Journal of Biomedical Optics*, 15(1):016021–016021, 2010.
- [134] J. Nandy, W. Hsu, y M. L. Lee. An incremental feature extraction framework for referable diabetic retinopathy detection. En *IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, págs. 908–912. IEEE, 2016.
- [135] A. Y. Ng. Feature selection, L1 vs. L2 regularization, and rotational invariance. En *Proceedings of the International Conference on Machine Learning*, pág. 78. 2004.
- [136] H. V. Nguyen, G. S. W. Tan, R. J. Tapp, S. Mital, D. S. W. Ting, H. T. Wong, C. S. Tan, A. Laude, E. S. Tai, N. C. Tan, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in singapore. *Ophthalmology*, 123(12):2571–2580, 2016.
- [137] T. T. Nguyen, J. J. Wang, A. R. Sharrett, F. A. Islam, R. Klein, B. E. Klein, M. F. Cotch, y T. Y. Wong. Relationship of retinal vascular caliber with diabetes and retinopathy. *Diabetes Care*, 31(3):544–549, 2008.
- [138] U. T. Nguyen et al. An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognition*, 46(3):703–715, 2013.
- [139] M. Niemeijer, M. D. Abramoff, y B. Van Ginneken. Information fusion for diabetic retinopathy cad in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 28(5):775–785, 2009.
- [140] M. Niemeijer, M. K. Garvin, B. van Ginneken, M. Sonka, y M. D. Abramoff. Vessel segmentation in 3D spectral OCT scans of the retina. En *Medical Imaging*, págs. 69141R–69141R. International Society for Optics and Photonics, 2008.

- [141] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M. D. Abramoff, et al. Comparative study of retinal vessel segmentation methods on a new publicly available database. En *Medical Imaging 2004*, págs. 648–656. International Society for Optics and Photonics, 2004.
- [142] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Transactions on Medical Imaging*, 29(1):185–195, 2010.
- [143] M. Niemeijer, B. Van Ginneken, J. Staal, M. S. Suttorp-Schulten, y M. D. Abramoff. Automatic detection of red lesions in digital color fundus photographs. *IEEE Transaction on Medical Imaging*, 24(5):584–592, 2005.
- [144] J. Odstrčilík, J. Jan, J. Gazárek, y R. Kolář. Improvement of vessel segmentation by matched filtering in colour retinal images. En *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, págs. 327–330. Springer, 2009.
- [145] J. Odstrčilík, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, y E. Angelopoulou. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. *IET Image Processing*, 7(4):373–383, 2013.
- [146] J. Odstrcilik, R. Kolar, J. Jan, J. Gazarek, Z. Kuna, y M. Vodakova. Analysis of retinal nerve fiber layer via markov random fields in color fundus images. En *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*, págs. 504–507. IEEE, 2012.
- [147] J. Odstrcilik, R. Kolar, R.-P. Tornow, J. Jan, A. Budai, M. Mayer, M. Vodakova, R. Laemmer, M. Lamos, Z. Kuna, et al. Thickness related textural properties of retinal nerve fiber layer in color fundus images. *Computerized Medical Imaging and Graphics*, 38(6):508–516, 2014.
- [148] V. Optical. Volk Optical - Ophthalmic Imaging Designer & Manufacturer — Lenses — Surgical — Cameras. <https://volk.com/>. Accessed: 2017-03-23.

- 
- [149] W. H. Organization et al. Global data on visual impairments 2010. *URL: <http://www.who.int/blindness/GLOBALDATAFINALforweb.pdf> [accessed 2013-02-28]/[WebCite Cache]*, 2012.
- [150] J. I. Orlando y M. B. Blaschko. Learning fully-connected CRFs for blood vessel segmentation in retinal images. En P. Golland, C. Barillot, J. Hornegger, y R. Howe, eds., *MICCAI 2014, LNCS*, tomo 8149, págs. 634–641. Springer, 2014.
- [151] J. I. Orlando y M. del Fresno. Reviewing preprocessing and feature extraction techniques for retinal blood vessel segmentation in fundus images. *Mecánica Computacional*, XXXIII(42):2729–2743, 2014.
- [152] J. I. Orlando, H. L. Manterola, E. Ferrante, y F. Ariel. Arabidopsis roots segmentation based on morphological operations and CRFs. *arXiv preprint arXiv:1704.07793*, 2014.
- [153] J. I. Orlando, E. Prokofyeva, y M. B. Blaschko. 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, 2017.
- [154] J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Convolutional neural network transfer for automated glaucoma identification. En *12th International Symposium on Medical Information Processing and Analysis*, págs. 101600U–101600U. International Society for Optics and Photonics, 2017.
- [155] J. I. Orlando, E. Prokofyeva, M. del Fresno, y M. B. Blaschko. Learning to detect red lesions in fundus photographs: An ensemble approach based on deep learning. *arXiv preprint arXiv:1706.03008*, 2017.
- [156] M. A. Palomera-Pérez, M. E. Martínez-Pérez, H. Benítez-Pérez, y J. L. Ortega-Arjona. Parallel multiscale feature extraction and region growing: application in retinal blood vessel detection. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):500–506, 2010.

- [157] N. Panwar, P. Huang, J. Lee, P. A. Keane, T. S. Chuan, A. Richhariya, S. Teoh, T. H. Lim, y R. Agrawal. Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare. *Telemedicine and e-Health*, 22(3):198–208, 2016.
- [158] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, y I. J. Constable. Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1):99–127, 2006.
- [159] A. Perez-Rovira, T. MacGillivray, E. Trucco, K. Chin, K. Zutis, C. Lupascu, D. Tegolo, A. Giachetti, P. Wilson, A. Doney, et al. VAMPIRE: vessel assessment and measurement platform for images of the retina. En *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*, págs. 3391–3394. IEEE, 2011.
- [160] A. Perez-Rovira, K. Zutis, J. Hubschman, y E. Trucco. Improving vessel segmentation in ultra-wide field-of-view retinal fluorescein angiograms. En *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, págs. 2614–2617. IEEE, 2011.
- [161] R. Pires, S. Avila, H. Jelinek, J. Wainer, E. Valle, y A. Rocha. Beyond lesion-based diabetic retinopathy: a direct approach for referral. *IEEE Journal of Biomedical and Health Informatics*, 2015.
- [162] R. Pires, H. F. Jelinek, J. Wainer, S. Goldenstein, E. Valle, y A. Rocha. Assessing the need for referral in automatic diabetic retinopathy detection. *IEEE Transactions on Biomedical Engineering*, 60(12):3391–3398, 2013.
- [163] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, y Y. Zheng. Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90:200–205, 2016.
- [164] E. Prokofyeva y E. Zrenner. Epidemiology of major eye diseases leading to blindness in Europe: A literature review. *Ophthalmic Research*, 47(4):171–188, 2012. doi: 10.1159/000329603.

- [165] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, y M. Lamard. Deep image mining for diabetic retinopathy screening. *arXiv preprint arXiv:1610.07086*, 2016.
- [166] G. Quellec, M. Lamard, A. Erginay, A. Chabouis, P. Massin, B. Cochener, y G. Cazuguel. Automatic detection of referral patients due to retinal pathologies through data mining. *Medical Image Analysis*, 29:47–64, 2016.
- [167] R. Ranjan. Fundus Fluoroscein Angiography. <https://es.slideshare.net/rashmiranjan589/fundus-fluoroscein-angiography/7?smtNoRedir=1>. Accessed: 2017-03-23.
- [168] S. Ravishankar, A. Jain, y A. Mittal. Automated feature extraction for early detection of diabetic retinopathy in fundus images. En *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, págs. 210–217. IEEE, 2009.
- [169] A. S. Razavian, H. Azizpour, J. Sullivan, y S. Carlsson. CNN features off-the-shelf: An astounding baseline for recognition. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, págs. 512–519. IEEE, 2014.
- [170] G. Richard, G. Soubrane, y L. Yannuzzi. Fluorescein angiography: textbook and atlas (2nd rev. and expanded ed.). 1998.
- [171] R. Roletschek. Wikimedia Commons: Fundus kamera by Ralf Roletschek / Wikipedia. <https://commons.wikimedia.org/wiki/File:2010-12-07-funduskamera-by-RalfR-02.jpg>. Accessed: 2017-03-23.
- [172] S. Roychowdhury et al. Blood vessel segmentation of fundus images by major vessel extraction and sub-image classification. *IEEE Journal of Biomedical and Health Informatics*, PP(99):1–1, 2014.
- [173] S. Roychowdhury, D. D. Koozekanani, y K. K. Parhi. Iterative vessel segmentation of fundus images. *IEEE Transactions on Biomedical Engineering*, 62(7):1738–1749, 2015.

- [174] S. Roychowdhury, D. D. Koozekanani, y K. K. Parhi. Automated detection of neovascularization for proliferative diabetic retinopathy screening. En *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, págs. 1300–1303. IEEE, 2016.
- [175] A. Rubinstein, L. Gutierrez, A. Beratarrechea, y V. E. Irazola. Increased prevalence of diabetes in Argentina is due to easier health care access rather than to an actual increase in prevalence. *PLOS ONE*, 9(4):e92245, 2014.
- [176] C. I. Sánchez, M. Niemeijer, A. V. Dumitrescu, M. S. Suttorp-Schulten, M. D. Abràmoff, y B. van Ginneken. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Investigative Ophthalmology & Visual Science*, 52(7):4866–4871, 2011.
- [177] P. J. Savino y H. V. Danesh-Meyer. *Color Atlas and Synopsis of Clinical Ophthalmology—Wills Eye Institute—Neuro-Ophthalmology*. Lippincott Williams & Wilkins, 2012.
- [178] B. Schölkopf. *Support Vector Learning*. Tesis Doctoral, Oldenbourg Verlag, Munich, 1997.
- [179] L. Seoud, J. Chelbi, y F. Cheriet. Automatic grading of diabetic retinopathy on a public database. En *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA. MICCAI*, 2015.
- [180] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet, y J. P. Langlois. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Transactions on Medical Imaging*, 35(4):1116–1126, 2016.
- [181] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, y Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013.
- [182] M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, N. Paragios, y I. Kokkinos. Sub-cortical brain structure segmentation using F-CNN's. En *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, págs. 269–272. IEEE, 2016.

- [183] A. Sharif Razavian, H. Azizpour, J. Sullivan, y S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, págs. 806–813. 2014.
- [184] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, y R. M. Summers. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, 2016.
- [185] H. Sidahmed, E. Prokofyeva, y M. B. Blaschko. Discovering predictors of mental health service utilization with  $k$ -support regularized logistic regression. *Information Sciences*, 329:937–949, 2016. doi: <http://dx.doi.org/10.1016/j.ins.2015.03.069>.
- [186] D. A. Sim, P. A. Keane, A. Tufail, C. A. Egan, L. P. Aiello, y P. S. Silva. Automated retinal image analysis for diabetic retinopathy in telemedicine. *Current Diabetes Reports*, 15(3):1–9, 2015.
- [187] J. Sivaswamy, S. Krishnadas, A. Chakravarty, G. Joshi, A. S. Tabish, et al. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, 2(1), 2015.
- [188] J. V. Soares et al. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9), 2006.
- [189] O. P. Society. Retinal OCT imaging. <http://www.opsweb.org/?page=RetinalOCT>, 2017. Accessed: 2017-03-27.
- [190] A. Soltanipour et al. Vessel centerlines extraction from fundus fluorescein angiogram based on hessian analysis of directional curvelet subbands. En *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, págs. 1070–1074. IEEE, 2013.
- [191] M. Sonka y M. D. Abràmoff. Quantitative analysis of retinal OCT. 2016.

- [192] G. Spaeth. Appearances of the optic disc in glaucoma: a pathogenetic classification. En *Trans Acad Ophthalmol Symposium on Glaucoma*. CV Mosby, St. Louis. 1981.
- [193] J. Staal et al. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.
- [194] Y.-C. Tham, X. Li, T. Y. Wong, H. A. Quigley, T. Aung, y C.-Y. Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014.
- [195] A. Trost, S. Lange, F. Schroedl, D. Bruckner, K. A. Motloch, B. Bogner, A. Kaser-Eichberger, C. Strohmaier, C. Runge, L. Aigner, et al. Brain and retinal pericytes: origin, function and role. *Frontiers in Cellular Neuroscience*, 10, 2016.
- [196] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. P. Hubschman, B. Al-Diri, C. Y. Cheung, D. Wong, M. Abramoff, et al. Validating retinal fundus image analysis algorithms: Issues and a proposal. *Investigative Ophthalmology & Visual Science*, 54(5):3546–3559, 2013.
- [197] I. Tsochantaridis et al. Large margin methods for structured and interdependent output variables. En *Journal of Machine Learning Research*, págs. 1453–1484. 2005.
- [198] L. Van Der Maaten. Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, 15(1):3221–3245, 2014.
- [199] B. van Ginneken, A. A. Setio, C. Jacobs, y F. Ciompi. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. En *IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, págs. 286–289. IEEE, 2015.
- [200] M. J. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, y C. I. Sánchez. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Transactions on Medical Imaging*, 35(5):1273–1284, 2016.

- [201] R. Vega et al. Retinal vessel extraction using lattice neural networks with dendritic processing. *Computers in Biology and Medicine*, 58:20–30, 2015.
- [202] R. Venkataramani, S. Thiruvankadam, P. Annangi, N. Babu, y V. Vaidya. Understanding the mechanisms of deep transfer learning for medical images. En *LABELS and DLMIA, at Medical Image Computing and Computer-Assisted Intervention-MICCAI*, tomo 10008, pág. 188. Springer, 2016.
- [203] J. A. Vilensky, W. Robertson, y C. A. Suarez-Quian. *The Clinical Anatomy of the Cranial Nerves: The Nerves of.ºn Old Olympus Towering Top*. John Wiley & Sons, 2015.
- [204] M. Vlachos y E. Dermatas. Multi-scale retinal vessel segmentation using line tracking. *Computerized Medical Imaging and Graphics*, 34(3):213–227, 2010.
- [205] H. H. Vo y A. Verma. New deep neural nets for fine-grained diabetic retinopathy recognition on hybrid color space. En *2016 IEEE International Symposium on Multimedia (ISM)*, págs. 209–215. IEEE, 2016.
- [206] P. Vostatek, E. Claridge, H. Uusitalo, M. Hauta-Kasari, P. Fält, y L. Lensu. Performance comparison of publicly available retinal blood vessel segmentation methods. *Computerized Medical Imaging and Graphics*, 55:2–12, 2017.
- [207] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, y J.-C. Klein. Automatic detection of microaneurysms in color fundus images. *Medical Image Analysis*, 11(6):555–566, 2007.
- [208] O. B. Walton, R. B. Garoon, C. Y. Weng, J. Gross, A. K. Young, K. A. Camero, H. Jin, P. E. Carvounis, R. E. Coffee, y Y. I. Chu. Evaluation of automated teleretinal screening program for diabetic retinopathy. *JAMA Ophthalmology*, 134(2):204–209, 2016.
- [209] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, y A. Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003–034003, 2014.

- [210] L. Wang, A. Bhalerao, y R. Wilson. Analysis of retinal vasculature using a multiresolution hermite model. *IEEE Transactions on Medical Imaging*, 26(2):137–152, 2007.
- [211] X. Wang, L. I. Mudie, M. Baskaran, C.-Y. Cheng, W. L. Alward, D. S. Friedman, y C. J. Brady. Crowdsourcing to evaluate fundus photographs for the presence of glaucoma. *Journal of Glaucoma*, 2017.
- [212] Z. Wang y J. Yang. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *arXiv preprint arXiv:1703.10757*, 2017.
- [213] R. N. Weinreb, T. Aung, y F. A. Medeiros. The pathophysiology and treatment of glaucoma: a review. *JAMA*, 311(18):1901–1911, 2014.
- [214] R. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, y S. Barman. Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. *Computer Methods and Programs in Biomedicine*, 114(3):247–261, 2014.
- [215] D. R. Whiting, L. Guariguata, C. Weil, y J. Shaw. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Research and Clinical Practice*, 94(3):311–321, 2011.
- [216] K. Willekens, S. Bataillie, I. Sarens, S. Odent, L. Abegão Pinto, E. Vandewalle, K. Van Keer, y I. Stalmans. Funduscopy versus HRT III Confocal Scanner Vertical Cup-Disc Ratio Assessment in Normal Tension and Primary Open Angle Glaucoma (The Leuven Eye Study). *Ophthalmic Research*, 57(2):100–106, 2016.
- [217] B. Wu, W. Zhu, F. Shi, S. Zhu, y X. Chen. Automatic detection of microaneurysms in retinal fundus images. *Computerized Medical Imaging and Graphics*, 55:106–112, 2017.
- [218] R. Wu, C. Y.-L. Cheung, S. M. Saw, P. Mitchell, T. Aung, y T. Y. Wong. Retinal vascular geometry and glaucoma: the singapore malay eye study. *Ophthalmology*, 120(1):77–83, 2013.

- [219] D. Xiao y Y. Kanagasingam. Screening of the retina in diabetes patients by morphological means. En *Teleophthalmology in Preventive Medicine*, págs. 15–26. Springer, 2015.
- [220] L. Xu y S. Luo. A novel method for blood vessel detection from retinal images. *Biomedical Engineering Online*, 9(1):14, 2010.
- [221] Y. Xu, L. Duan, D. W. K. Wong, T. Y. Wong, y J. Liu. Glaucoma detection by learning from multiple informatics domains. En E. Trucco, X. Chen, Garvin, M. K., J. J. Liu, y X. Y. Frank, eds., *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA*. 2015.
- [222] L. A. Yannuzzi, M. D. Ober, J. S. Slakter, R. F. Spaide, Y. L. Fisher, R. W. Flower, y R. Rosen. Ophthalmic fundus imaging: today and beyond. *American Journal of Ophthalmology*, 137(3):511–524, 2004.
- [223] Y. Yin et al. Automatic segmentation and measurement of vasculature in retinal fundus images using probabilistic formulation. *Computational and Mathematical Methods in Medicine*, 2013, 2013.
- [224] J. Yosinski, J. Clune, Y. Bengio, y H. Lipson. How transferable are features in deep neural networks? En *Advances in Neural Information Processing Systems*, págs. 3320–3328. 2014.
- [225] X. You et al. Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition*, 44(10), 2011.
- [226] F. Zana y J.-C. Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, 10(7):1010–1019, 2001.
- [227] B. Zhang, L. Zhang, L. Zhang, y F. Karray. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Computers in Biology and Medicine*, 40(4):438–445, 2010.

- 
- [228] Y. Zhao et al. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retina images. *IEEE Transactions on Medical Imaging*, 2015.
- [229] L. Zheng, B. Gong, D. A. Hatala, y T. S. Kern. Retinal ischemia and reperfusion causes capillary degeneration: similarities to diabetes. *Investigative Ophthalmology & Visual Science*, 48(1):361–367, 2007.
- [230] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, y D. Comaniciu. 3D deep learning for efficient and robust landmark detection in volumetric data. En *Medical Image Computing and Computer-Assisted Intervention-MICCAI*, págs. 565–572. 2015.