



UNIVERSIDAD DE VALLADOLID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE MÁSTER

MÁSTER EN INVESTIGACIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN Y LAS  
COMUNICACIONES

**Evaluación automática de la calidad  
en retinografías para la ayuda al  
diagnóstico de la Retinopatía Diabética**

Autor:

**D. Jorge Jiménez García**

Tutora:

**Dra. D<sup>a</sup> María García Gadañón**

Valladolid, Septiembre de 2018



---

**TÍTULO:** Evaluación automática de la calidad en retinografías para la ayuda al diagnóstico de la Retinopatía Diabética

**AUTOR:** D. Jorge Jiménez García

**TUTORA:** Dra. D<sup>a</sup>. María García Gadañón

**DEPARTAMENTO:** Departamento de Teoría de la Señal y Comunicaciones e Ingeniería Telemática

---

**TRIBUNAL**

---

**PRESIDENTE:** Dr. D. Roberto Hornero Sánchez

**SECRETARIO:** Dr. D. Mario Martínez Zarzuela

**VOCAL:** Dr. D. Jesús Poza Crespo

**SUPLENTE:** Dr. D. Miguel López-Coronado  
Sánchez-Fortún

**SUPLENTE:** Dr. D. Carlos Gómez Peña

**SUPLENTE:** Dr. D. Francisco Javier Díaz Pernas

---

---

**FECHA:** Septiembre de 2018

**CALIFICACIÓN:**

---



*Research is what I am doing  
when I don't know what I am doing  
(Wernher von Braun, 1957)*



## Agradecimientos

En primer lugar, quiero dar las gracias a todos los miembros del Grupo de Ingeniería Biomédica de la Universidad de Valladolid. Muchas gracias por vuestra ayuda, por vuestros consejos, por vuestras lecciones, por los buenos ratos que pasamos cada día. En fin, gracias por todo. Quiero hacer mención especial a Rober. Muchas gracias por soportarme, por resolver gran parte de mis dudas y por facilitarme tanto el trabajo. También quiero dar las gracias a mi tutora María García por guiarme con infinita paciencia en la realización de este trabajo. Finalmente, quiero agradecer a Roberto Hornero una vez más la confianza que siempre ha depositado en mí.

Creo que este año ha pasado demasiado deprisa. A pesar de las dudas iniciales, haber realizado el MUITIC ha sido una fantástica experiencia. Como hemos sido pocos alumnos, me puedo permitir el lujo de dar las gracias a Luis, a Javi, a Adrián y a Jaime por tantos ratos agradables de trabajo y ocio. Me gustaría expresar que ha sido un honor formar equipo con Adrián y Jaime.

Por último, quiero agradecer a mis padres José Luis y Agustina la fe inagotable que tienen en mí, su apoyo y su comprensión. Son unos auténticos luchadores que dan todo por mi hermano Álvaro y por mí.



## Resumen

La Diabetes Mellitus es una enfermedad cuya prevalencia supone un grave problema de salud pública. Entre sus complicaciones destaca la Retinopatía Diabética (RD), que se ha convertido en una de las principales causas de pérdida de visión en países desarrollados. La retinografía es una prueba diagnóstica que consiste en la captura de imágenes en color de la retina mediante un instrumento óptico llamado retinógrafo. Esta técnica permite una visión exacta de la retina y es útil para el diagnóstico y el seguimiento de enfermedades que la afectan, como RD, Glaucoma y Degeneración Macular Asociada a la Edad. En los programas de cribado de la RD, la adquisición de retinografías aptas para ser analizadas es básico. De hecho, los sistemas automáticos de análisis de retinografías para la ayuda al diagnóstico pueden fallar si las imágenes no tienen calidad suficiente. El objetivo de este Trabajo Fin de Máster fue el desarrollo de un sistema de evaluación automática de la calidad en retinografías con el fin de determinar si las imágenes son aptas para ser analizadas. De esta manera, se puede mejorar la eficacia de sistemas de cribado y de ayuda al diagnóstico de diversas enfermedades, entre ellas la RD.

Este TFM se centró en la extracción, selección y clasificación de características globales que cuantifican la calidad de las retinografías para tratar de determinar si son aptas para ser analizadas por sistemas automáticos. Para ello, se aplicaron técnicas de análisis de imágenes basadas en métodos de evaluación de calidad de imagen sin referencia (*No Reference Image Quality Assessment*, NR-IQA). En concreto, se aplicaron los métodos de propósito general *Natural Image Quality Evaluator* (NIQE) y *Spatial and Spectral Entropy-based Quality* (SSEQ). Además, se analizaron las imágenes mediante la Transformada *Wavelet* Continua y la luminosidad del fondo de las imágenes a partir del espacio de color HSV. Después, se seleccionaron las características más relevantes mediante el algoritmo *Fast Correlation-Based Filter* (FCBF). Para la clasificación, se empleó una red neuronal artificial de tipo Perceptrón Multicapa. Para comprobar el funcionamiento del método se utilizó una base de datos de 2107 retinografías proporcionada por el Instituto Universitario de Oftalmobiología Aplicada de la Universidad de Valladolid y el Hospital Clínico Universitario de Valladolid. Los expertos determinaron si las imágenes tenían calidad suficiente para ser analizadas según criterios objetivos de claridad. La base de datos se dividió en dos conjuntos independientes para el desarrollo del método: entrenamiento (1053 imágenes) y test (1054 imágenes).

El método se evaluó sobre el conjunto de test, consiguiéndose una precisión del 92.13%, con una sensibilidad del 93,26% y un valor predictivo positivo del 97.46%. Los resultados logrados fueron similares a los obtenidos en la literatura, si bien las bases de datos empleadas difieren entre estudios. Estos valores de sensibilidad y valor predictivo positivo hacen que el método propuesto sea adecuado para seleccionar las retinografías de buena calidad como paso previo a su análisis automático. En este sentido, se pretende incluir este

método como una de las etapas de un algoritmo de análisis de retinografías para la ayuda al diagnóstico y cribado de la RD.

## Palabras Clave

Análisis de retinografías; Evaluación de la calidad en imágenes sin referencia; Perceptrón Multicapa; Retina; Transformada *wavelet*.

## Abstract

Diabetes Mellitus prevalence has become a severe public health issue. Among its complications, Diabetic Retinopathy (DR) is nowadays one of the main causes of blindness in developed countries. Fundus or retinal imaging is a diagnostic procedure that consists in acquiring color pictures of the retina with an optical system called fundus camera. This procedure allows an exact view of the retina and is useful in the diagnosis and treatment of diseases that alter it, such as DR, Glaucoma, or Age-Related Macular Degeneration. When dealing with DR screening programs, acquisition of suitable fundus images is a basic step. In fact, automatic computer-aided diagnostic systems tend to fail if images do not exhibit enough quality. The objective of this Master Thesis was to develop an automatic retinal image quality assessment method aimed at determining whether retinal images have enough quality to be analyzed. Consequently, the effectiveness of automated screening programs and computer-aided diagnostic systems for retinal pathologies, including DR, could be improved.

This Master Thesis was focused on the extraction, selection and classification of global image quality features in fundus images to determine whether they are suitable to be analyzed by automatic systems. For this purpose, image analysis techniques based on No Reference Image Quality Assessment (NR-IQA) methods were used. More precisely, general purpose Natural Image Quality Evaluator (NIQE) and Spatial and Spectral Entropy-based Quality (SSEQ) methods were applied. In addition, images were analyzed using the Continuous Wavelet Transform and the background luminosity in the HSV color space. Then, feature selection was performed in order to keep the most relevant features using the Fast Correlation-Based Filter (FCBF) algorithm. Classification was carried out using a Multilayer Perceptron Neural Network. In order to assess the performance of the proposed method, a database with 2107 retinal images, provided by the Instituto Universitario de Oftalmobiología Aplicada of the University of Valladolid and the Hospital Clínico Universitario of Valladolid was used. Expert ophthalmologists decided whether the images had enough quality to be analyzed following objective clarity criteria. The database was divided into two independent subsets for method development: training set (1053 images) and test set (1054 images).

The performance of the method was evaluated on the test set. Our results reached 92.13% accuracy, with a sensitivity of 93.26% and a positive predictive value of 97.46%. These results were similar to those found in the literature, but the databases differ among studies. Our results show that the proposed method could be used as a first step in automatic retinal image analysis methods. In this sense, we intend to include the proposed quality assessment algorithm as the first stage of a retinal image analysis method aimed for screening and computer-aided diagnosis of DR.

## Keywords

Multilayer Perceptron; No Reference Image Quality Assessment; Retina; Retinal Imaging; Wavelet Transform.

# Índice General

---

Capítulo 1. Introducción.....	1
1.1 Imagen médica. Retinografías.....	1
1.2 Diabetes Mellitus y Retinopatía Diabética .....	7
1.2.1 Diabetes Mellitus .....	7
1.2.2 Retinopatía Diabética.....	9
1.3 Hipótesis de trabajo .....	11
1.4 Objetivos .....	12
1.5 Metodología .....	13
1.6 Estructura del documento .....	13
Capítulo 2. Revisión del estado de la técnica.....	15
2.1 Introducción .....	15
2.2 Métodos de evaluación de la calidad.....	15
2.2.1 Métodos estructurales.....	16
2.2.2 Métodos genéricos .....	16
2.2.3 Métodos híbridos.....	17
2.2.4 Métodos basados en <i>Deep Learning</i> .....	18
Capítulo 3. Materiales y métodos .....	21
3.1 Introducción .....	21
3.2 Base de datos de retinografías.....	21
3.3 Preprocesado .....	23
3.3.1 Obtención de la máscara FOV.....	24
3.3.2 Extensión del FOV .....	25
3.4 Extracción de características .....	26
3.4.1 Características basadas en naturalidad .....	27
3.4.2 Características basadas en la entropía espacial y la entropía espectral .....	31
3.4.3 Características basadas en la transformada <i>wavelet</i> y la varianza local .....	33
3.4.4 Características basadas en la luminosidad del fondo retiniano .....	38
3.4.5 Conjunto de características extraídas.....	41

3.5	Selección de características .....	41
3.6	Clasificación de características.....	45
3.6.1	Balanceo de clases mediante sobremuestreo .....	46
3.6.2	Fase de entrenamiento de la red MLP.....	47
3.6.3	Fase de test de la red MLP .....	50
Capítulo 4.	Resultados.....	51
4.1	Introducción .....	51
4.2	Características seleccionadas por el algoritmo FCBF .....	51
4.3	Evaluación de los resultados .....	52
4.4	Resultados obtenidos.....	53
4.4.1	Fase de entrenamiento.....	53
4.4.2	Fase de test .....	55
Capítulo 5.	Discusión .....	57
5.1	Introducción .....	57
5.2	Preprocesado .....	57
5.3	Extracción de características .....	58
5.4	Selección de características .....	67
5.5	Clasificación mediante MLP .....	68
5.6	Comparativa con estudios previos.....	72
Capítulo 6.	Conclusiones y líneas futuras.....	75
6.1	Introducción .....	75
6.2	Grado de cumplimiento de los objetivos .....	75
6.3	Contribuciones originales.....	76
6.4	Conclusiones.....	77
6.5	Limitaciones y líneas futuras de investigación.....	78
Bibliografía .....		81
Listado de acrónimos .....		87

## Índice de Figuras

<b>Figura 1.1:</b> Corte del globo ocular con algunas estructuras del ojo humano señaladas (Fuente: National Eye Institute, National Institutes of Health). .....	2
<b>Figura 1.2:</b> Corte del globo ocular donde se representa la situación de los conos y los bastones (Fuente: Blue Cone Monochromacy Families Foundation). .....	3
<b>Figura 1.3:</b> Retinografía donde se puede observar la mácula (centro), la papila (izquierda) y los vasos sanguíneos. ....	6
<b>Figura 1.4:</b> Diferencia entre la observación del fondo de ojo con la pupila no dilatada y la pupila dilatada (Fuente: National Eye Institute, National Institutes of Health). ....	7
<b>Figura 1.5:</b> Retinografía en la que se aprecian lesiones características de la RD (MA, HE, EX y CW). ....	10
<b>Figura 1.6:</b> Metodología del estudio. ....	13
<b>Figura 3.1:</b> Retinógrafo no midriático TOPCON TRC-NW400 (Topcon, n.d.). ....	22
<b>Figura 3.2:</b> Ejemplo de obtención de una máscara del FOV circular; (a) Imagen original en la que se ha superpuesto la diagonal a través de la que se calcula el perfil; (b) Gráfica del perfil $p(d)$ ; (c) Derivada del perfil $p(d)$ (d) Máscara del FOV obtenida. ....	25
<b>Figura 3.3:</b> Resultado del algoritmo de extensión del FOV. (a) Imagen original; (b) Imagen preprocesada. ....	27
<b>Figura 3.4:</b> Respuestas de la CWT para diferentes escalas. (a) Retinografía preprocesada en color; (b) Respuesta para $s=4$ ; (c) Respuesta para $s=16$ ; (d) Respuesta para $s=64$ . ....	35
<b>Figura 3.5:</b> Desviación típica local de la CWT para diferentes escalas. (a) Retinografía preprocesada en color; (b) Resultado para $s=4$ ; (c) Resultado para $s=16$ ; (d) Resultado para $s=64$ . ....	36
<b>Figura 3.6:</b> Distribuciones promedio de la desviación típica local para diferentes escalas. (a) Distribuciones para $s=4$ ; (b) Distribuciones para $s=8$ ; (c) Distribuciones para $s=16$ ; (d) Distribuciones para $s=32$ . ....	37
<b>Figura 3.7:</b> Extracción del fondo empleando el canal V. (a) Retinografía de buena calidad; (b) Luminosidad del fondo de la imagen (a); (c) Retinografía de mala calidad; (d) Luminosidad del fondo de la imagen (b). ....	39
<b>Figura 3.8:</b> Distribución promedio de los valores de luminosidad en el canal V, $F(x,y)$ : (a) Función de densidad $p(F)$ ; (b) Función de distribución acumulada $P(F)$ . ....	40
<b>Figura 3.9:</b> Selección de características mediante FCBF: (a) Matriz de coeficientes de correlación de las características extraídas; (b) Matriz de coeficientes de correlación de las características seleccionadas. ....	45
<b>Figura 3.10:</b> Criterio de parada temprana (early stopping). Adaptado de Bishop (1995). ....	49
<b>Figura 4.1:</b> Valores de Acc estimados para los diferentes parámetros de regularización $v$ . Cada curva muestra la evolución de Acc según aumenta $n_{oc}$ . (a) Parámetros $v$ entre 0.1 y 0.5; (b) Parámetros $v$ entre 0.6 y 0.9. ....	54
<b>Figura 4.2:</b> Curva ROC obtenida para el conjunto de entrenamiento. ....	55

<b>Figura 4.3:</b> (a) Evaluación de Acc, Se y Sp frente a los umbrales; (b) Se, VPP y F-Score frente a los umbrales.....	56
<b>Figura 5.1:</b> Representación mediante boxplots de los valores obtenidos con el método NIQE.....	58
<b>Figura 5.2:</b> Representación mediante boxplots de las características extraídas mediante SSEQ (escala 1). (a): SpacEn <sub>MED</sub> ; (b) SpacEn <sub>ASIM</sub> ; (c) SpecEn <sub>MED</sub> ; (d) SpecEn <sub>ASIM</sub> .....	59
<b>Figura 5.3:</b> Representación mediante boxplots de las características extraídas mediante SSEQ (escala 2). (a): SpacEn <sub>MED</sub> ; (b) SpacEn <sub>ASIM</sub> ; (c) SpecEn <sub>MED</sub> ; (d) SpecEn <sub>ASIM</sub> .....	60
<b>Figura 5.4:</b> Representación mediante boxplots de las características extraídas mediante SSEQ (escala 3). (a): SpacEn <sub>MED</sub> ; (b) SpacEn <sub>ASIM</sub> ; (c) SpecEn <sub>MED</sub> ; (d) SpecEn <sub>ASIM</sub> .....	61
<b>Figura 5.5:</b> Representación mediante boxplots de CWT_Ent para diferentes escalas. (a) s=2; (b) s=4; (c) s=8; (d) s=16; (e) s=32; (f) s=64.....	62
<b>Figura 5.6:</b> Representación mediante boxplots de CWT_Med para diferentes escalas. (a) s=2; (b) s=4; (c) s=8; (d) s=16; (e) s=32; (f) s=64.....	63
<b>Figura 5.7:</b> Representación mediante boxplots de CWT_Desv para diferentes escalas. (a) s=2; (b) s=4; (c) s=8; (d) s=16; (e) s=32; (f) s=64.....	64
<b>Figura 5.8:</b> Representación mediante boxplots de las características extraídas a partir de la luminosidad del fondo. (a) Lum <sub>1</sub> ; (b) Lum <sub>5</sub> ; (c) Lum <sub>10</sub> ; (d) Lum <sub>15</sub> ; (e) Lum <sub>20</sub> . ....	65
<b>Figura 5.9:</b> Representación mediante boxplots de las características extraídas a partir de la luminosidad del fondo. (a) Lum <sub>5-1</sub> ; (b) Lum <sub>10-5</sub> ; (c) Lum <sub>15-10</sub> ; (d) Lum <sub>20-15</sub> . ....	66
<b>Figura 5.10:</b> Imágenes de mala calidad clasificadas erróneamente por el método de evaluación automática (a) Papila y parte de los vasos borrosos; (b) Vasos borrosos en algunas zonas de la imagen; (c) Gran diferencia de contraste entre zonas central y periférica; (d) Imagen parcialmente oscurecida.....	70
<b>Figura 5.11:</b> Imágenes de buena calidad clasificadas erróneamente por el método de evaluación automática (a) Imagen borrosa debido al movimiento en la captura; (b) Contraste de los vasos escaso, posible subjetividad en la anotación; (c), (d) Imágenes con calidad aceptable en general. ....	71

## Índice de Tablas

<b>Tabla 2.1:</b> Métodos de evaluación de la calidad en retinografías.....	19
<b>Tabla 4.1:</b> Características seleccionadas por el algoritmo FCBF. ....	51
<b>Tabla 4.2:</b> Posibles resultados del clasificador al comparar la salida proporcionada por el algoritmo con las anotaciones sobre la calidad de las imágenes de la BD.....	52
<b>Tabla 4.3:</b> Resultados de la red MLP en el conjunto de test. ....	56
<b>Tabla 4.4:</b> Resultados de la red MLP en el conjunto de entrenamiento.....	56
<b>Tabla 5.1:</b> Comparación de los resultados logrados en estudios previos. ....	73



# Capítulo 1. Introducción

---

Este Trabajo Fin de Máster (TFM) está centrado en el análisis de retinografías, una clase de imagen médica en la que se aprecia la retina junto a las estructuras oculares que la conforman. Las retinografías se utilizan en el diagnóstico de diversas patologías que afectan a la retina, y que pueden causar la pérdida de la visión. En este capítulo se contextualiza el TFM, se definen conceptos básicos en los que está enmarcado y se establecen la hipótesis de trabajo y los objetivos del TFM. Finalmente se describe la metodología empleada y la organización del documento.

## 1.1 IMAGEN MÉDICA. RETINOGRAFÍAS

La ingeniería biomédica es una rama interdisciplinar de la ingeniería que trata de intercambiar conocimientos y metodologías entre ingeniería y medicina para resolver problemas médicos. Mientras la ingeniería aporta conocimientos sobre ciencias exactas y tecnología, la medicina aporta conocimientos sobre ciencias biológicas y de la salud (Mompín Poblet, 1988; Bronzino, 2006). Combinando ambas disciplinas, se consiguen mejoras en el diagnóstico, tratamiento, seguimiento y prevención de las enfermedades. La ingeniería biomédica es una disciplina orientada a la investigación, que trata de aplicar técnicas novedosas para mejorar el funcionamiento de los sistemas de salud (Mompín Poblet, 1988; Bronzino, 2006).

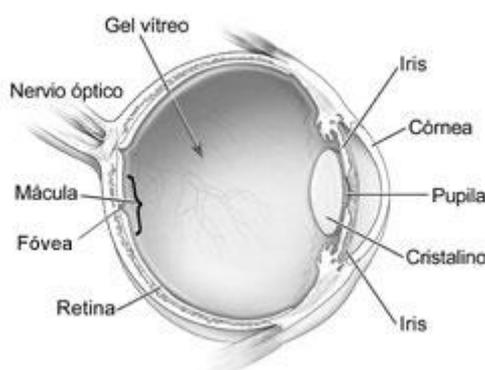
Como subdisciplina de la ingeniería biomédica, la imagen médica trata de aplicar técnicas de obtención y análisis de imágenes de tejidos y órganos del cuerpo humano. Las imágenes obtenidas con estas técnicas son útiles en el entorno médico para la prevención, tratamiento, seguimiento y control de enfermedades (World Health Organization, 2017). En muchos casos, el diagnóstico de enfermedades únicamente puede realizarse mediante estas técnicas. Además, el empleo de imágenes médicas permite mejorar el diagnóstico de enfermedades y realizar valoraciones más objetivas por parte de los especialistas de la medicina (World Health Organization, 2017).

El análisis automático de imágenes médicas complementa las técnicas de imagen médica aplicando métodos de procesamiento digital de imágenes y técnicas de reconocimiento de patrones, visión por ordenador, etc. para analizar su contenido (Dhawan, 2011; Toennies, 2012). La automatización del análisis de imágenes médicas es primordial en muchas áreas de la medicina. La falta de especialistas capaces de analizar las imágenes obtenidas, tarea que requiere mucho tiempo y esfuerzo, hace necesario desarrollar técnicas de ayuda al diagnóstico (Rangayyan, 2005; Toennies, 2012). Además de incrementar la eficacia en el diagnóstico y seguimiento de enfermedades, estas técnicas permiten reducir la subjetividad de los observadores humanos (Toennies, 2012).

En el campo de la oftalmología, se han aplicado numerosas técnicas de obtención de imágenes médicas (Yolcu *et al.*, 2014). Su uso es primordial para el diagnóstico de algunas patologías que afectan al ojo humano, y que en muchas ocasiones provocan visión reducida y pérdida de visión (Ng *et al.*, 2014). El examen de la retina mediante técnicas de imagen médica tiene gran relevancia y es uno de los métodos diagnósticos más empleados por oftalmólogos especialistas en retina y vítreo (Ng *et al.*, 2014).

Dentro de la gran variedad de técnicas de obtención de imágenes médicas, se encuentran las modalidades de imagen médica centradas en el diagnóstico de enfermedades que afectan a la retina. La retina es un tejido de varias capas situado en la parte posterior interna del ojo, en la que se proyectan las imágenes que pasan a través de la córnea, el cristalino y el vítreo (Figura 1.1) (Abràmoff *et al.*, 2010; Ng *et al.*, 2014). Este tejido se compone de células sensibles a la luz, encargadas de convertir las imágenes que se forman sobre ella en señales neuronales que viajan hacia el córtex visual del cerebro. Es por esto por lo que se considera la retina como una extensión del cerebro que permite la visión (Abràmoff *et al.*, 2010).

La retina está formada por varias capas, y se divide en dos partes: la retina neurosensorial y el epitelio pigmentario. La primera parte se encuentra más próxima al vítreo, y contiene las capas formadas por células fotosensibles, células bipolares y células ganglionares, que transportan los potenciales de acción hacia el nervio óptico, situado en la zona central de la retina (Alio *et al.*, 1993; Graue, 1995; García, 2008). Por lo tanto, en la retina neurosensorial se concentran todas las células encargadas de convertir las imágenes en señales neuronales. Por su parte, el epitelio pigmentario es la capa más profunda de la retina, sobre la que se asienta la retina neurosensorial. Es de color rojizo, y como las capas de la retina neurosensorial son transparentes, la retina se observa de este color. Justo por debajo de la retina se encuentra la coroides, una capa muy vascularizada que proporciona nutrientes a la retina. La separación entre el epitelio pigmentario y la coroides es la membrana de Bruch,



**Figura 1.1:** Corte del globo ocular con algunas estructuras del ojo humano señaladas (Fuente: National Eye Institute, National Institutes of Health).

que permite el intercambio de sustancias entre ambas (Alio *et al.*, 1993; Graue, 1995; Thibodeau and Patton, 2000; Grisan, 2005; García, 2008).

Las células fotosensibles de la retina se dividen en conos y bastones. Estas células se distribuyen en torno a la zona de la retina donde se enfocan las imágenes y justo por encima del epitelio pigmentario, formando la capa más profunda de la retina neurosensorial (Figura 1.2) (Abràmoff *et al.*, 2010; Mookiah *et al.*, 2015). Los bastones son más numerosos, y se distribuyen en torno al centro de la retina. Son células muy sensibles a la luz y no son capaces de distinguir los colores. Son los encargados de la visión nocturna y periférica, y a la vez son ciegos en la visión diurna. En condiciones de visión diurna, los conos son los encargados de la agudeza visual y del reconocimiento de los colores. Los conos se concentran en el centro de la retina, en una región llamada mácula (Mustafi *et al.*, 2009). La mácula es una excavación en el centro de la retina en la que se concentran los conos, y se reduce la concentración de bastones. La mácula es la zona más importante de la retina, ya que el deterioro de la retina en esta zona conduce a la pérdida de la visión (Mookiah *et al.*, 2015). La fovea es el centro de la mácula, lugar donde la concentración de conos es máxima y la de bastones se anula. Según el color que son capaces de detectar, los conos se dividen en conos rojos, conos verdes y conos azules (Mustafi *et al.*, 2009).

Los conos y bastones establecen conexiones sinápticas con las células bipolares, y éstas con las células ganglionares. Los axones de las células ganglionares convergen en el disco óptico o papila formando la cabeza del nervio óptico (Mookiah *et al.*, 2015). El nervio óptico recoge las señales neuronales, que abandonan el ojo viajando hacia el córtex visual del cerebro. En la papila no hay conos ni bastones, por lo que se forma un punto ciego. Además, de la papila parte una red de vasos sanguíneos que nutre a la retina neurosensorial. Estos

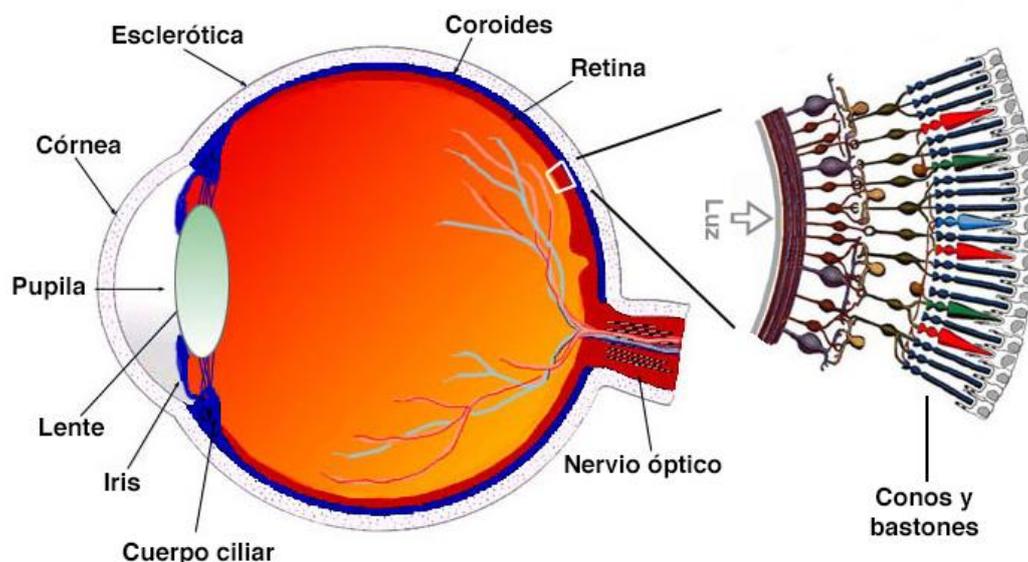


Figura 1.2: Corte del globo ocular donde se representa la situación de los conos y los bastones (Fuente: Blue Cone Monochromacy Families Foundation).

vasos sanguíneos se sitúan en las capas superiores de la retina, y junto con los vasos coroideos aporta los nutrientes necesarios para los diferentes tejidos que conforman la retina (Abràmoff *et al.*, 2010). La mácula, la papila y los vasos sanguíneos retinianos son las principales estructuras oculares que se pueden apreciar en imágenes de la retina (Abràmoff *et al.*, 2010).

Algunas estructuras del ojo humano como la córnea, el cristalino y el vítreo deben ser transparentes para que se formen las imágenes sobre la retina. Por eso, la retina se puede observar desde el exterior del ojo a través de la pupila. Mediante diversas técnicas, los oftalmólogos pueden obtener imágenes del fondo de ojo o la retina para observar anomalías que pueden revelar algunas enfermedades que afectan a la retina y merman la capacidad de visión del paciente (Mookiah *et al.*, 2015; Besenczi *et al.*, 2016). Entre las técnicas de obtención de imágenes de la retina, las más utilizadas son las siguientes (Corcostegui, 1983; Webb *et al.*, 1987; Gómez-Ulla, 1994; Yannuzzi *et al.*, 2004; García, 2008; Aguirre-Díaz, 2014; Mookiah *et al.*, 2015; Nentwich and Ulbig, 2015):

**Oftalmoscopia.** Es una técnica para observar el fondo de ojo de manera sencilla e inmediata, pero no se puede obtener una imagen de la retina. Se puede observar el fondo retiniano mediante oftalmoscopio directo e indirecto.

- El oftalmoscopio directo es un instrumento óptico con el que se puede examinar la retina utilizando una luz que se proyecta sobre la misma a través de un prisma. La luz se refleja en la retina y el observador puede ver la imagen proyectada en un orificio encima del prisma. El observador puede examinar la retina viendo una imagen monocular y amplificada. Es un instrumento sencillo, barato y portátil, aunque tiene poca capacidad de visualización de la retina periférica.
- El oftalmoscopio indirecto es un instrumento que permite visualizar la retina mediante una potente fuente de luz que se hace coincidir con el eje de mirada. El observador puede ver la retina con diferentes aumentos gracias a unas lentes convexas de diferentes dioptrías. La técnica indirecta permite ver la parte posterior interna del ojo con mayor amplitud y claridad que el oftalmoscopio directo. Las ventajas que aporta sobre la oftalmoscopia directa la convierten en la técnica más empleada para realizar una primera observación de la retina en las consultas de oftalmología.

**Angiografía fluoresceínica.** Es una técnica que requiere la inyección de un contraste (fluoresceína sódica) en el antebrazo del paciente. Tras la propagación de esta sustancia por el sistema circulatorio, se visualiza el fondo de ojo a través de un oftalmoscopio que permita visualizar la fluorescencia de los vasos retinianos mediante filtros especiales. Es posible que se tomen fotografías para su posterior análisis. La fluoresceína sódica actúa como un colorante de la sangre, y se utiliza luz azul para iluminar la retina. De esta manera, los vasos se muestran más brillantes que el fondo del ojo y con mayor contraste. Tras la inyección, la fluoresceína llega antes a los vasos coroideos que a los vasos retinianos. Esta técnica se emplea desde finales de los años 60 en el estudio de patologías vasculares que afectan a la

retina, así como anomalías en el epitelio pigmentario. También se puede emplear indocianina verde como agente de contraste, aportando más información en caso de que exista neovascularización coroidea.

**Oftalmoscopia de láser de barrido (*Scanning Laser Ophthalmoscope, SLO*).** Es una técnica con la que se obtienen imágenes de la retina de alta resolución empleando un haz de luz láser, recogiendo los reflejos producidos por la iluminación de puntos de la retina de manera secuencial. Esta técnica no requiere la dilatación de la pupila. El sistema realiza barridos de un estrecho haz de luz láser a gran velocidad sobre un área retiniana. De esta manera, se simula una iluminación homogénea de pequeñas zonas de la retina. Tras recoger los reflejos del láser, se forma una imagen de alta calidad y resolución de la retina. El sistema es robusto frente a opacidades de estructuras oculares como el cristalino, la córnea o el vítreo, evitando artefactos. Mediante SLO se pueden obtener imágenes con detalle de las fibras nerviosas de la retina y de manera instantánea. Se emplea habitualmente para realizar oftalmoscopia con múltiples longitudes de onda, estudios de la hemodinámica retiniana, densitometrías retinianas o la evaluar la topografía de la papila. Los sistemas SLO más actuales permiten obtener imágenes de campo de visión ultra amplio sin necesidad de dilatar la pupila, llegando a los 200° de apertura.

**Tomografía de coherencia óptica (*Optical Coherence Tomography, OCT*).** Mediante esta técnica se pueden obtener imágenes de cortes histológicos de la retina. Surgida en los años 90, la OCT hace posible el examen detallado de zonas de la retina como la mácula y el nervio óptico mediante tomografías. La captura de imágenes no ocasiona molestias al paciente, pues el instrumento no está en contacto con el ojo. Este instrumento supone un gran avance en la visualización de la retina, pues permite examinar en vivo la papila, la mácula, la retina posterior y su relación con el vítreo y la coroides. Estas capacidades permiten que la OCT sea especialmente útil en el diagnóstico de enfermedades maculares. Se pueden encontrar equipos que combinan las tecnologías SLO y OCT, conocidos como oftalmoscopios OCT. Estos equipos producen simultáneamente imágenes de la superficie retiniana y tomografías en las que se pueden analizar las capas más profundas de la retina.

**Autofluorescencia.** Esta técnica aprovecha la emisión estimulada de luz de las moléculas presentes en el epitelio pigmentario, principalmente debida a la lipofuscina. Las moléculas emiten esta luz debido a que contienen pigmentos autofluorescentes acumulados en el epitelio pigmentario, que aparecen por la degradación incompleta de los segmentos exteriores de los fotorreceptores del ojo. En una imagen de autofluorescencia, la intensidad de esta es mayor cuanto mayor es la cantidad y distribución de la lipofuscina. La imagen de autofluorescencia revela cambios metabólicos en el epitelio pigmentario provocados por el daño oxidativo, y se utiliza como un complemento de las técnicas de imagen clásicas. Junto con la retinografía, se emplea en la detección de daños relacionados con la degeneración macular asociada a la edad (DMAE).

**Retinografía.** La imagen de fondo de ojo o retinografía es una imagen en color de la retina adquirida con una cámara fotográfica especial. La adquisición de retinografías es un procedimiento no invasivo e indoloro que permite examinar la circulación sanguínea de la retina prescindiendo de inyecciones de fluoresceína, como en el caso de la angiografía. De esta manera, se evitan los efectos secundarios de las mismas (Abràmoff *et al.*, 2010; Bernardes *et al.*, 2011). Las imágenes digitales adquiridas con las cámaras de fondo de ojo tienen gran detalle, y permiten la visión exacta de la retina, la papila, la mácula y los vasos sanguíneos (Figura 1.3) (Besenczi *et al.*, 2016). Además, las imágenes se pueden almacenar en formato digital y pueden ser analizadas por un oftalmólogo posteriormente (Bernardes *et al.*, 2011).

El instrumento habitual para capturar las retinografías se llama retinógrafo o cámara de fondo de ojo. Este instrumento está conectado a un ordenador que incorpora un *software* para la visualización de imágenes digitales. La digitalización y almacenamiento hace posible que las retinografías se puedan analizar de manera automática empleando técnicas de procesado digital de imagen. Existen diversas técnicas útiles en la detección de patologías que afectan a la retina, pero la retinografía es uno de los métodos más aceptados para el cribado de enfermedades que provocan cambios en la retina y pueden conducir a la pérdida de visión (Fong *et al.*, 2003; Besenczi *et al.*, 2016).

Los retinógrafos pueden ser de dos tipos: midriáticos y no midriáticos. La midriasis se produce con el aumento del diámetro de la pupila. Se produce de manera natural cuando el entorno se encuentra en condiciones de baja iluminación. Sin embargo, el oftalmólogo puede

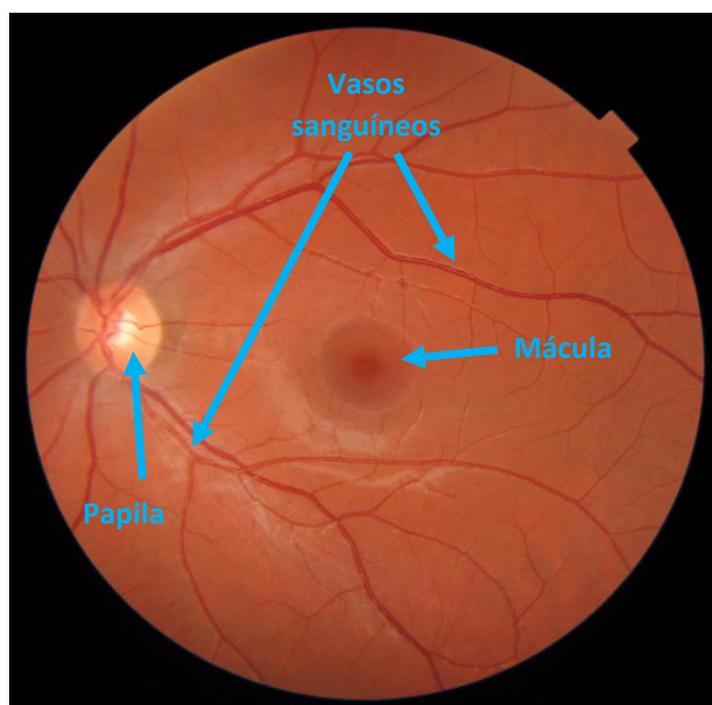


Figura 1.3: Retinografía donde se puede observar la mácula (centro), la papila (izquierda) y los vasos sanguíneos.

provocar la midriasis aplicando un colirio midriático llamado tropicamida (Martínez Rubio *et al.*, 2012). En la Figura 1.4 se observan los efectos de la midriasis junto con el área retiniana que se puede observar en cada caso. Los retinógrafos midriáticos necesitan que se provoque una midriasis en el paciente para la captura de imágenes, mientras que esto no es necesario en los retinógrafos no midriáticos (Pinies, 2005). La utilización de los retinógrafos no midriáticos no ocasiona molestias al paciente provocadas por la midriasis. En general, estas molestias son elevada sensibilidad a la luz y visión borrosa durante un intervalo de tiempo (Gilmartin *et al.*, 1995). Sin embargo, con los retinógrafos midriáticos se pueden obtener imágenes con mayor campo de visión (*Field of View, FOV*). El FOV se expresa en grados, y tiene un rango típico de 15° a 60° (García, 2008). Cuanto mayor es el ángulo de FOV, mayor superficie retiniana se representa en la imagen, pero acarrea una mayor dilatación de la pupila (Hackel and Saine, 2005).

## 1.2 DIABETES MELLITUS Y RETINOPATÍA DIABÉTICA

### 1.2.1 Diabetes Mellitus

La Diabetes Mellitus (DM) es un conjunto de trastornos metabólicos por los que el cuerpo no es capaz de regular el nivel de glucosa en la sangre. Esto se debe a que la producción de insulina por parte del páncreas es insuficiente o a que el cuerpo la rechaza. La insulina es una hormona que interviene en la asimilación por parte de las células de la glucosa

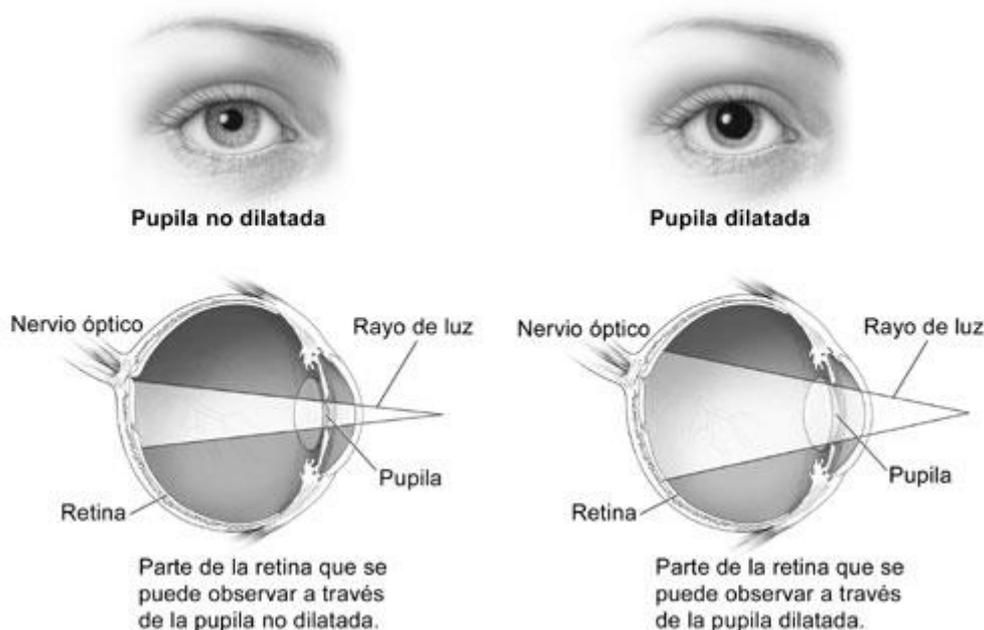


Figura 1.4: Diferencia entre la observación del fondo de ojo con la pupila no dilatada y la pupila dilatada (Fuente: National Eye Institute, National Institutes of Health).

proveniente de los alimentos. Las células  $\beta$  del páncreas producen la insulina necesaria para un metabolismo adecuado (Tebar and Escobar, 2009). Las personas que padecen DM producen una cantidad insuficiente de insulina en unos casos, y en otros casos la insulina que produce es rechazada por el cuerpo al no ser adecuada. En todo caso, la incapacidad de asimilar la glucosa conduce a una hiperglucemia crónica en la sangre. El excesivo nivel de glucosa en sangre produce daños en diferentes órganos del cuerpo humano como el corazón, los riñones, los ojos, y también en los vasos sanguíneos y los nervios (American Diabetes Association, 2009).

Existen síntomas de hiperglucemia provocada por la DM. Los más comunes son poliuria (producción excesiva de orina), polidipsia (sed excesiva), pérdida de peso, polifagia (aumento de la necesidad de comer) y visión borrosa (Tebar and Escobar, 2009). En un plazo largo de tiempo, la DM puede provocar complicaciones severas como retinopatía, nefropatía, neuropatía, riesgo de úlceras en los pies, amputaciones y neuropatías autónomas que causan síntomas gastrointestinales y cardiovasculares, además de disfunción sexual (Tebar and Escobar, 2009). Los casos de DM se pueden clasificar en dos tipos principalmente (American Diabetes Association, 2009; Tebar and Escobar, 2009):

- DM Tipo 1: En este caso se produce una reacción autoinmune que destruye las células  $\beta$ . Como consecuencia, el páncreas es incapaz de producir insulina suficiente. La tasa de destrucción de células  $\beta$  es variable, siendo mayor en niños y jóvenes y menor en adultos. Los pacientes con DM Tipo 1 necesitan inyecciones de insulina diarias para restablecer niveles aceptables de insulina y así poder regular su nivel de glucosa en sangre. La DM Tipo 1 se conoce como insulino dependiente o juvenil, y afecta a entre el 5% y el 10% de las personas que sufren DM.
- DM Tipo 2: En la DM Tipo 2 no se produce la destrucción autoinmune de las células  $\beta$ , por lo que los pacientes con DM Tipo 2 no necesitan en principio inyecciones de insulina para sobrevivir. En este caso, se produce principalmente una resistencia a la insulina. Los pacientes con DM Tipo 2 padecen hiperglucemia crónica a la vez que sus niveles de insulina son adecuados o incluso elevados. Esto se debe a que la producción de insulina es anormal, y no consigue compensar la resistencia a la insulina por parte del cuerpo. La obesidad propicia la resistencia a la insulina y está presente en la mayoría de los casos de DM Tipo 2. En estos casos, la reducción de peso o los tratamientos farmacológicos pueden mejorar la resistencia a la insulina. Sin embargo, habitualmente no se restablece la normalidad. La hiperglucemia se desarrolla gradualmente, y el paciente no nota los síntomas típicos de la DM hasta una fase avanzada. Por ello, la DM Tipo 2 no se diagnostica fácilmente en sus primeros años, y muchos de los pacientes de tienen un riesgo elevado de desarrollar complicaciones microvasculares y macrovasculares. La DM Tipo

---

2 se conoce como no insulino dependiente o adulta y es la clase más prevalente, afecta a entre el 90% y el 95% de las personas que sufren DM.

Las estimaciones de la Federación Internacional de Diabetes (FID) cuantifican en 425 millones las personas que sufrían DM en 2017, y en 628 millones las personas que la sufrirán en 2045 (Federación Internacional de Diabetes, 2017). La DM se ha convertido en una enfermedad muy importante para los sistemas de salud, ya que sus complicaciones afectan gravemente la calidad de vida de los pacientes e incrementan los costes sanitarios destinados a su diagnóstico y tratamiento (Tebar and Escobar, 2009; Federación Internacional de Diabetes, 2017).

### 1.2.2 Retinopatía Diabética

La Retinopatía Diabética (RD) es una complicación microvascular común y específica de la DM. Aproximadamente, la tercera parte de la población que padece DM sufre las consecuencias de la RD. Se puede evitar su aparición si se controla adecuadamente la DM desde fases tempranas, ya que el desarrollo de la RD viene motivado por la hiperglucemia, la hipertensión y la duración de la diabetes (Cheung *et al.*, 2010; Yau *et al.*, 2012). El elevado número de personas con DM hace que las complicaciones visuales de la RD sean la principal causa de ceguera evitable en la población adulta, y una de las principales causas de ceguera en países desarrollados (Nentwich and Ulbig, 2015). No obstante, la hipertensión también es un factor patogenético de muchas de las causas de pérdida de visión como RD, Glaucoma y DMAE (Wong and Mitchell, 2007).

En los pacientes con DM, la presencia durante años de hiperglucemia e hipertensión provoca cambios bioquímicos y fisiológicos que conducen a daños microvasculares y disfunción retinal provocada por la neurodegeneración (Cheung *et al.*, 2010). La RD afecta principalmente a la retina neurosensorial, y se puede considerar una neuropatía que afecta a la retina mediante cambios microvasculares en los vasos sanguíneos de la retina (Alghadyan, 2011). Estos cambios microvasculares provocan un aumento de la presión en los capilares que conduce a dilataciones de sus paredes llamadas microaneurismas (MA). El aumento del calibre de las arterias revela el desarrollo temprano de la RD. A medida que la RD avanza, se pueden producir hemorragias (HE) en los vasos sanguíneos. También se produce una elevada permeabilidad de los capilares y los vasos sanguíneos que provocan fugas de líquidos que están presentes en la sangre en forma de edemas y exudados duros (EX) (Cheung *et al.*, 2010). Debido a la hipertensión se produce también isquemia de las capas nerviosas de la retina, que provoca la aparición de exudados algodonosos (*Cotton wool spots*, CW) (Wong and Mitchell, 2007).

En las retinografías, los MA se aprecian como pequeños puntos de color rojo oscuro situados en zonas cercanas a las venas perimaculares o alrededor de las arterias temporales. A medida que la RD progresa, los MA crecen en número y extensión (Fong *et al.*, 2004; García, 2008; Nentwich and Ulbig, 2015). Las HE se producen principalmente por la rotura de los MA en fases más avanzadas. Son más extensas que los MA, suelen ser redondos en zonas

intermedias de la retina y tienen formas más indefinidas en zonas cercanas a los vasos principales (García, 2008; Alghadyan, 2011; Nentwich and Ulbig, 2015). Los EX son depósitos de material lipídico filtrado por los vasos sanguíneos. En las retinografías se observan como pequeños nódulos compactos con un color amarillento y con bordes bien definidos. Su tamaño y forma pueden variar. Asimismo, pueden aparecer aislados o formando conglomerados (García, 2008; Alghadyan, 2011).

La RD se desarrolla de una manera progresiva. Las primeras fases se asocian a la Retinopatía Diabética no proliferativa (RDNP), en la que predominan cambios vasculares y la presencia de algunos MA, EX y HE (Fong *et al.*, 2004). La Retinopatía Diabética proliferativa (RDP) es una fase más avanzada, en la que además aparecen complicaciones más graves como neovascularización o hemorragia de vítreo (Fong *et al.*, 2004). Las HE y EX producen isquemia en la retina, y la retina responde mediante la neovascularización. Consiste en la creación de nuevos vasos sanguíneos que puedan sustituir a los vasos dañados previamente, pero estos vasos son frágiles, se rompen con facilidad y provocan la hemorragia de vítreo (García, 2008; Cheung *et al.*, 2010). La hemorragia de vítreo provoca que este se vuelva opaco y puede provocar pérdida de visión. La neovascularización también provoca el crecimiento de nuevas fibras nerviosas que alteran la retina, llegando a provocar el desprendimiento de retina y la pérdida irreversible de la visión (García, 2008; Cheung *et al.*, 2010; Alghadyan, 2011).

La clasificación de la severidad de la RD fue propuesta por el *Early Treatment Diabetic Retinopathy Study* (ETDRS). Según la escala propuesta, la RD se clasifica en seis niveles de gravedad (Fong *et al.*, 2004; Cheung *et al.*, 2010):

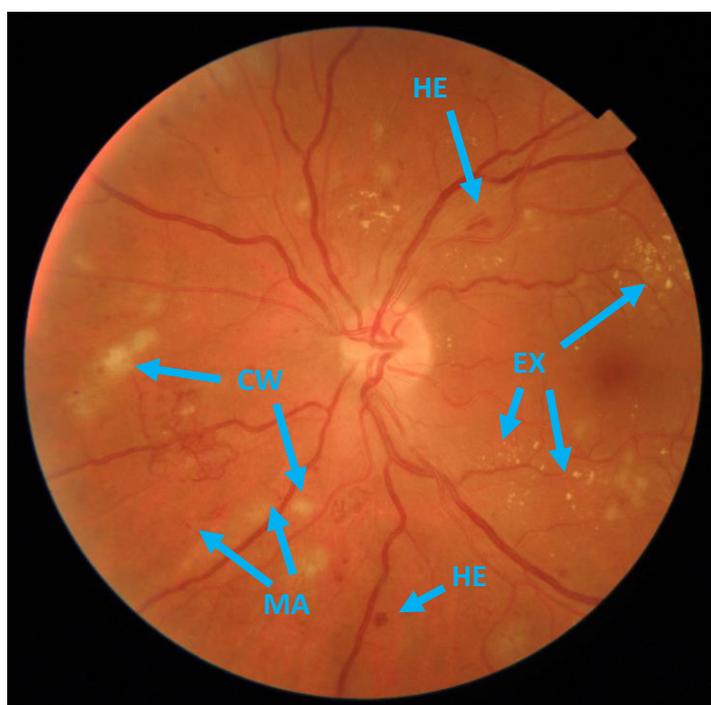


Figura 1.5: Retinografía en la que se aprecian lesiones características de la RD (MA, HE, EX y CW).

- 1) Sin evidencias de retinopatía: No se aprecian marcadores de la RD.
- 2) RDNP temprana: Se aprecian únicamente algunos MA.
- 3) RDNP moderada: Aparte de varios MA se aprecian más signos de RD, pero no es suficiente como para considerarla severa.
- 4) RDNP severa: Se aprecian más de 20 HE en cada uno de los cuadrantes y anomalías microvasculares en cualquier cuadrante. Ningún signo de RDP.
- 5) RDP: Se aprecia neovascularización o hemorragia de vítreo.
- 6) Edema macular diabético (EMD): Se aprecian EX en la zona macular, engrosamiento de la retina debido al edema de mácula.

Según los últimos estudios realizados por la FID, en 2015 más de 93 millones de personas sufrían algún tipo de daño ocular provocado por la DM, y más de una de cada tres desarrollará RD (International Diabetes Federation and The Fred Hollows Foundation, 2015). De estos 93 millones de personas, 17 millones padecían RDP, 21 millones padecían EMD, y 28 millones sufrían RD con una severidad que compromete la visión (International Diabetes Federation and The Fred Hollows Foundation, 2015).

### 1.3 HIPÓTESIS DE TRABAJO

Según estimaciones de la FID, la prevalencia de la DM es del 8.8% de la población adulta en 2017, y para el año 2045 las previsiones son del 9.9% de la población adulta (Federación Internacional de Diabetes, 2017). La población adulta que padece DM en los países desarrollados está expuesta a una gran variedad de complicaciones de las que no es siquiera consciente. Estas complicaciones propician la muerte prematura y la merma de las condiciones de vida de los pacientes y, además, elevan de manera considerable los gastos sanitarios. En los últimos 10 años, el gasto sanitario asociado a la DM se ha triplicado en todo el mundo (Federación Internacional de Diabetes, 2017). La principal complicación ocular de la DM es la RD, pero también puede desencadenar cataratas o glaucoma. Resulta necesario establecer programas de cribado de la RD debido a su posible desarrollo en la población diabética. Los exámenes de retina periódicos son esenciales para el diagnóstico temprano de la RD, y el acceso a los servicios de oftalmología para el diagnóstico y control de la RD de manera eficaz es clave (Federación Internacional de Diabetes, 2017). La adquisición de retinografías de entre 2 y 7 campos según el protocolo establecido por ETDRS se ha establecido como prueba de referencia para la detección de la RD (Pinies, 2005). La experiencia del fotógrafo y la colaboración del paciente influyen en el proceso de adquisición de las imágenes (Pinies, 2005). El análisis de retinografías por parte de los oftalmólogos es una tarea compleja, y la efectividad de dichos programas se reduce si no disponen de las herramientas y el personal adecuados (Besenczi *et al.*, 2016; Mansour, 2017). En ocasiones, las retinografías capturadas durante un examen oftalmológico no se pueden analizar debido a una adquisición defectuosa o a la presencia de artefactos. Como consecuencia, es imposible verificar si existen signos de patología en estas retinografías, que no tienen la calidad suficiente como para ser analizadas (Fleming *et al.*, 2012).

Un sistema de análisis automático de retinografías puede ser una herramienta útil para la ayuda al diagnóstico de la RD. Un sistema de este tipo permitiría analizar de manera automática las imágenes para detectar signos de RD, y así facilitar el diagnóstico de la enfermedad. Para obtener un diagnóstico correcto por parte de este sistema, las retinografías procesadas deberían tener calidad suficiente como para ser analizadas automáticamente. Es posible implementar un método de evaluación automática de la calidad de las retinografías como paso previo a la localización automática de signos de patología. El método propuesto en este TFM contribuiría al desarrollo de un sistema automático de análisis de retinografías para el diagnóstico y cribado de enfermedades oculares. Esto permitiría el análisis automático de un número elevado de imágenes y reduciría la carga de trabajo de los oftalmólogos.

En este TFM se ha desarrollado un método de evaluación automática de la calidad en retinografías. Para llevar a cabo esta tarea, se han estudiado diversas alternativas para la medida de la calidad en este tipo de imágenes. Posteriormente, se ha tratado de desarrollar un método de evaluación de la calidad preciso y efectivo que automatice la tarea de seleccionar las retinografías aptas para su procesamiento posterior. Para ello, se han estudiado diversas técnicas de procesamiento digital de imágenes.

#### 1.4 OBJETIVOS

El principal objetivo de este TFM fue el **desarrollo de un método automático para evaluar la calidad en retinografías**. Este método se plantea como un primer paso en un sistema más general para la ayuda al diagnóstico de la RD. Para conseguir este objetivo general, se han planteado los siguientes objetivos específicos:

- 1) Estudiar el estado de la técnica, realizando búsquedas bibliográficas relacionadas con el análisis de retinografías. De esta manera se pretende comprender los conceptos necesarios para la realización de este trabajo e identificar las técnicas aplicadas en estudios previos.
- 2) Obtener una colección amplia de retinografías cuya calidad tiene que haber sido anotada de forma manual por especialistas. Será necesario también revisar las anotaciones realizadas por los oftalmólogos sobre las imágenes para crear la base de datos (BD) final para el desarrollo del TFM.
- 3) Seleccionar y desarrollar métodos de evaluación de la calidad de las retinografías basados en técnicas de procesamiento digital de imágenes. Para ello se empleará el *software* MATLAB®, aprovechando las capacidades del *toolbox* de procesamiento de imágenes (*Image Processing Toolbox*, IPT).
- 4) Obtener los resultados de los métodos desarrollado sobre la BD de retinografías.
- 5) Analizar los resultados obtenidos por el método de evaluación de la calidad.
- 6) Extraer las conclusiones del estudio realizado.
- 7) Difundir los resultados en foros y publicaciones de ámbito científico.

## 1.5 METODOLOGÍA

La metodología empleada para llevar a cabo este TFM consta de 7 actividades que se detallan a continuación (Figura 1.6):

- 1) Búsqueda de información necesaria para comprender el problema a resolver. Principalmente se consultaron libros, revistas científicas y médicas, trabajos académicos y páginas *web*.
- 2) Revisión de la BD empleada, identificando las características más relevantes de las imágenes de buena calidad y de calidad insuficiente. Comparación con los criterios de evaluación empleados en estudios previos.
- 3) Búsqueda de métodos de evaluación de la calidad en retinografías presentes en la literatura. Estos métodos fueron publicados principalmente en revistas científicas y en libros de actas de congresos.
- 4) Implementación de los métodos seleccionados. Para llevar a cabo esta tarea, se utilizaron las funcionalidades de MATLAB® y su *toolbox* IPT.
- 5) Procesado de las imágenes y obtención de resultados.
- 6) Análisis de resultados y extracción de conclusiones.
- 7) Documentación del trabajo desarrollado.

## 1.6 ESTRUCTURA DEL DOCUMENTO

El primer capítulo de este TFM es una breve introducción a los conceptos básicos sobre los que se asienta este TFM. En él se han descrito la hipótesis, los objetivos y la metodología de trabajo. Los capítulos restantes se organizan de la siguiente manera:

- El capítulo 2 es una revisión del estado de la técnica sobre evaluación de la calidad en retinografías. Se revisaron algunas de las técnicas más relevantes y novedosas aplicadas a esta tarea.
- En el capítulo 3 se explican los materiales y métodos utilizados en este TFM. En primer lugar, se describe la BD de retinografías empleada. A continuación, se exponen tanto los métodos de análisis de imágenes como de clasificación aplicados.



Figura 1.6: Metodología del estudio.

- En el capítulo 4 se presentan los resultados del algoritmo de evaluación de la calidad aplicado en la BD empleada para este TFM.
- En el capítulo 5 se discuten los resultados obtenidos en este TFM. Además, se comparan estos resultados con los obtenidos en otros estudios.
- En el capítulo 6 se exponen las contribuciones y conclusiones sobre el trabajo realizado. Además, se describen las limitaciones y líneas futuras del trabajo.

## Capítulo 2. Revisión del estado de la técnica

---

### 2.1 INTRODUCCIÓN

En los sistemas de análisis automático de retinografías, la evaluación de la calidad se situaría como el primer paso del algoritmo (Abràmoff *et al.*, 2010). Resulta esencial verificar si una imagen tiene calidad suficiente para ser analizada. Si en la imagen no se pueden apreciar claramente las estructuras oculares y posibles lesiones, no es posible determinar la presencia y el grado de severidad de la RD (Fleming *et al.*, 2012). En este TFM se ha realizado una revisión del estado de la técnica sobre los métodos presentes en la literatura para la tarea de evaluación de la calidad en retinografías. El desarrollo de este tipo de métodos no ha recibido una gran atención comparado con otras técnicas (Paulus *et al.*, 2010). Sin embargo, en los últimos años se han presentado varios métodos que abordan esta tarea.

En general, para que una retinografía sea graduable se deben cumplir dos condiciones (Fleming *et al.*, 2006): que tenga una claridad suficiente y que el campo fotografiado sea el correcto. Este TFM se ha centrado en la evaluación de la primera condición, que implica que en la imagen se aprecien claramente las estructuras oculares y las posibles lesiones. Muchas propuestas asumen que las imágenes obtenidas siguen un protocolo de captura, por lo que no sería necesario evaluar la segunda condición (Fasih *et al.*, 2014). La causa más común de pérdida de claridad en la imagen se debe a un enfoque inadecuado, que provoca que la imagen se vea borrosa. También la iluminación insuficiente o no uniforme suelen provocar pérdida de calidad (Bartling *et al.*, 2009; Fleming *et al.*, 2012). Se han establecido ciertas reglas para valorar si las retinografías son graduables (Scanlon *et al.*, 2003). Sin embargo, la tarea de determinar si una imagen es apta para ser analizada es altamente subjetiva si la valoración la realiza un observador humano.

### 2.2 MÉTODOS DE EVALUACIÓN DE LA CALIDAD

Según el enfoque empleado, los métodos de evaluación de la calidad en retinografías se pueden dividir en dos grupos principalmente (Paulus *et al.*, 2010; Sevik *et al.*, 2014; Wang *et al.*, 2016; Welikala *et al.*, 2016): métodos estructurales o basados en segmentación y métodos genéricos o basados en el histograma. Los métodos estructurales son métodos que utilizan la segmentación de estructuras oculares, para evaluar la calidad de la imagen. Por otra parte, los métodos genéricos recogen características de la imagen como el contraste o la iluminación (Pires Dias *et al.*, 2014). Es necesario mencionar que algunos autores han combinado ambos tipos de métodos, desarrollando así métodos híbridos (Paulus *et al.*, 2010). Finalmente, en los últimos años han surgido implementaciones basadas en *Deep Learning* y

redes neuronales convolucionales (*Convolutional Neural Network, CNN*) (Costa *et al.*, 2017; Saha *et al.*, 2018).

### 2.2.1 Métodos estructurales

La red de vasos sanguíneos es la estructura ocular más utilizada por los métodos estructurales para analizar la calidad de la imagen. La idea principal es que si los vasos sanguíneos se pueden segmentar en su mayoría sin problema, la imagen puede considerarse de buena calidad (Welikala *et al.*, 2016). En el primer método estructural conocido se utilizó un método de segmentación de vasos sanguíneos, y tras la segmentación, se midió el área de la red vascular para así obtener el indicador de calidad (Usher *et al.*, 2003). Otros autores se centraron en la detección de pequeños vasos alrededor de la mácula debido a la importancia de esta zona de la imagen (Fleming *et al.*, 2006; Hunter *et al.*, 2011). Niemeijer *et al.* (2006) presentaron un método de agrupación de estructuras de la imagen (*Image Structure Clustering, ISC*). Mediante la respuesta a varios filtros basados en derivadas de gaussianas, los píxeles se agrupan en 5 *clusters* representativos de las estructuras de la imagen. Se clasificaron características extraídas de estos *clusters* utilizando un clasificador *Support Vector Machine* (SVM). Basándose en este trabajo, otros autores (Giancardo *et al.*, 2008) definieron varias zonas del FOV para evaluar la segmentación de los vasos en varias zonas de la imagen. Métodos más recientes evaluaron la calidad a partir de la segmentación de la red vascular extrayendo características de área, fragmentación y complejidad (Welikala *et al.*, 2016).

Otros trabajos se apoyaron en la segmentación de los vasos sanguíneos para localizar las regiones de la imagen por las que discurren, y extraer características de dichas regiones. Fleming *et al.* (2012) localizaron estas zonas relevantes a partir de un método de realce de vasos. Después, extrajeron características a partir de estas regiones, que combinaron con características extraídas de la segmentación de los vasos. Posteriormente, utilizaron un clasificador SVM. Trabajos similares se basaron en la detección de los vasos para seleccionar pequeñas regiones de las retinografías sobre los que se obtuvieron indicadores de calidad (Kohler *et al.*, 2013). Yu *et al.* (2012) extrajeron características de textura y de nitidez local que combinaron con las características extraídas de la red vascular.

### 2.2.2 Métodos genéricos

Los métodos genéricos se basan en evaluar características globales de la imagen. Por lo tanto, emplean un enfoque más general en el que se lleva a cabo un análisis de la luminosidad, nitidez, contraste, etc. (Pires Dias *et al.*, 2014). Se trata, principalmente, de técnicas de análisis de histogramas, tanto de las imágenes sin procesar como del resultado de diversos métodos de procesamiento de imagen.

Los primeros métodos genéricos se basaron en el cálculo de histogramas de la imagen y midiendo su similitud con un histograma de referencia (Lee and Wang, 1999). Ideas muy similares, incluyendo el histograma de la imagen obtenida mediante un método de detección de bordes fueron empleados por Lalonde *et al.* (2001). También se emplearon medidas de

nitidez e iluminación local para medir la calidad global de la imagen (Bartling *et al.*, 2009). Otras características globales que se han explorado para estudiar la iluminación y la nitidez en retinografías son la media, varianza, asimetría, kurtosis, entropía y textura (Davis *et al.*, 2009; Fasih *et al.*, 2014). En trabajos más recientes se han extraído características de contraste basadas en la detección de zonas borrosas mediante las técnicas *Cumulative Probability of Blur Detection* (CPBD) y *Just Noticeable Blur* (JNB) (Kohler *et al.*, 2013; Fasih *et al.*, 2014; Wang *et al.*, 2016). Finalmente, también se han propuesto medidas de contraste obtenidas a través del histograma del ancho de los bordes obtenidos con el operador de Sobel (Lin *et al.*, 2017).

En la literatura se pueden encontrar, asimismo, métodos de clasificación basados en características de color y enfoque. Pires Dias *et al.* (2014) desarrollaron un completo método que utiliza características de color, enfoque, contraste e iluminación. Las características de color se extrajeron con la técnica *histogram backprojection*, definiendo mapas de colores adecuados e inadecuados en retinografías. Para la clasificación se empleó una red neuronal perceptrón multicapa (*Multilayer perceptron*, MLP) (Pires Dias *et al.*, 2014). También se detectaron distorsiones en las imágenes empleando características basadas en el sistema visual humano (*Human Visual System*, HVS) (Wang *et al.*, 2016). Se cuantificaron tres características: distorsión de color e iluminación, distorsión por borrosidad y distorsión de contraste. Finalmente, se compararon los clasificadores árbol de decisión y SVM (Wang *et al.*, 2016).

Propuestas más recientes incluyen el uso de transformada *wavelet*. Algunos trabajos (Abdel-Hamid *et al.*, 2016) emplearon la transformada *wavelet* sin diezmado (*Undecimated Wavelet Transform*, UWT) para analizar la iluminación, la homogeneidad, la saturación y la nitidez de la imagen. Para cada nivel de descomposición se obtuvieron la entropía, la media y el rango intercuartil de los coeficientes de la sub-banda de detalle como descriptores de nitidez (Abdel-Hamid *et al.*, 2016). En trabajos posteriores, los mismos autores propusieron un índice de calidad basado en estas características (Abdel-Hamid *et al.*, 2017).

### 2.2.3 Métodos híbridos

Los métodos híbridos combinan técnicas genéricas con técnicas estructurales para evaluar la calidad de las retinografías. En general, los argumentos que se esgrimen para optar por este tipo de algoritmos se basan en que los enfoques analizados anteriormente son incompletos por sí solos (Paulus *et al.*, 2010; Sevik *et al.*, 2014). Sin embargo, los investigadores que se han decantado por estos algoritmos no han sido numerosos. Frecuentemente se clasifican como métodos híbridos aquellos métodos más completos que han tenido en cuenta la localización de las estructuras oculares para detectar fotografías mal adquiridas (fotografías de la cara externa del ojo, pestañas, etc.) debidas a errores de captura, habitualmente denominadas *outliers* (Sevik *et al.*, 2014).

Hasta donde tenemos conocimiento, el primer método híbrido de evaluación de la calidad en retinografías fue presentado por Paulus *et al.* (2010). En él se definen criterios

tanto estructurales como genéricos para evaluar la calidad de las imágenes. Los métodos desarrollados estaban, por una parte, inspirados en el método ISC presentado anteriormente (Niemeijer *et al.*, 2006). Por otra parte, incluyeron también medidas de textura como la homogeneidad y el contraste. Finalmente, se empleó un clasificador SVM (Paulus *et al.*, 2010).

Otros autores han combinado la segmentación de las estructuras oculares con la evaluación de la calidad mediante medidas genéricas (Sevik *et al.*, 2014). Tras segmentar los vasos sanguíneos, la papila y la fovea, se extrajeron características de forma, textura e intensidad tanto de las imágenes como de las estructuras segmentadas. Con este completo planteamiento, se consiguieron distinguir retinografías de buena calidad, de mala calidad e incluso imágenes que no son retinografías (Sevik *et al.*, 2014).

Recientemente se han combinado métodos genéricos con estructurales utilizando una nueva característica de naturalidad junto con características de iluminación y la segmentación de la papila (Shao *et al.*, 2018). La característica de naturalidad empleada se basó en el algoritmo de evaluación de calidad *Natural Image Quality Evaluator* (NIQE). Las características extraídas se compararon los clasificadores de árbol de decisión, SVM y aprendizaje de diccionario (Mittal *et al.*, 2013; Shao *et al.*, 2018).

#### 2.2.4 Métodos basados en *Deep Learning*

Recientemente han surgido algunas implementaciones basadas en *Deep Learning*, que basan su funcionamiento en arquitecturas CNN. Esta técnica difiere de las técnicas tradicionales de *machine learning* en que no aprende con características extraídas por el desarrollador, sino que aprende características de manera autónoma (Saha *et al.*, 2018). Para desarrollar estos sistemas, se entrenan CNNs con extensas bases de datos de retinografías que disponen de anotaciones de calidad. La red neuronal aprende de manera automática las características de las imágenes y las clasifica (Costa *et al.*, 2017). El sistema “EyeQual” presentado por Costa *et al.* (2017) clasifica regiones de la imagen de manera automática y determina su calidad. Se ha desarrollado mediante una técnica de aprendizaje que reduce el número de ejemplos necesarios para el aprendizaje. Utiliza una CNN para evaluar la calidad de cada región y, finalmente, se ponderan las respuestas para determinar la calidad final de cada imagen (Costa *et al.*, 2017). También se ha utilizado la CNN “AlexNet” para la clasificación de retinografías (Saha *et al.*, 2018).

En la Tabla 2.1 se recogen las técnicas mencionadas en este capítulo, junto con una breve descripción de las mismas.

Tipo de método	Método	Descripción breve
Estructural	Usher <i>et al.</i> (2003)	Área de vasos segmentados
	Fleming <i>et al.</i> (2006)	Vasos perimaculares
	Hunter <i>et al.</i> (2009)	Vasos perimaculares
	Niemeijer <i>et al.</i> (2006)	Agrupación de píxeles en estructuras (ISC), histogramas RGB
	Giancardo <i>et al.</i> (2008)	Densidad local de vasos segmentados
	Welikala <i>et al.</i> (2016)	Área, fragmentación y complejidad de red vascular
	Fleming <i>et al.</i> (2012)	Regiones de relevancia y segmentación
	Yu <i>et al.</i> (2012)	Textura y nitidez de vasos segmentados
	Kohler <i>et al.</i> (2013)	Textura y nitidez de vasos y CPBD
Genérico	Lee y Wang (1999)	Histograma de referencia
	Lalonde <i>et al.</i> (2001)	Histograma de bordes y referencia
	Bartling <i>et al.</i> (2009)	Nitidez e iluminación local
	Davis <i>et al.</i> (2009)	Nitidez, iluminación y textura
	Fasih <i>et al.</i> (2014)	Nitidez, iluminación, textura y CPBD
	Pires <i>et al.</i> (2014)	<i>Histogram backprojection</i> , enfoque, contraste, iluminación
	Wang <i>et al.</i> (2016)	Características HVS
	Abdel-Hamid <i>et al.</i> (2016, 2017)	Transformada <i>wavelet</i> , índice de calidad
Híbrido	Paulus <i>et al.</i> (2010)	Adaptación de ISC, homogeneidad y contraste
	Sevik <i>et al.</i> (2014)	Segmentación estructuras, forma, textura e intensidad
	Shao <i>et al.</i> (2018)	NIQE, iluminación, segmentación de la papila
Deep Learning	Costa <i>et al.</i> (2017)	CNN sobre parches, ponderación de respuestas
	Saha <i>et al.</i> (2018)	CNN "AlexNet" adaptada

Tabla 2.1: Métodos de evaluación de la calidad en retinografías.



## Capítulo 3. Materiales y métodos

---

### 3.1 INTRODUCCIÓN

En este capítulo se describe la BD de retinografías empleada en este TFM, así como los métodos aplicados para resolver la tarea de evaluación de la calidad. Como se ha comprobado en el capítulo anterior, en la literatura se pueden encontrar métodos de evaluación de la calidad en retinografías que emplean técnicas variadas y planteamientos diversos. El objetivo de este TFM fue el desarrollo de un método de evaluación de la calidad basado en técnicas sencillas y efectivas, y cuyos resultados puedan ser comparables a los logrados en estudios previos.

En primer lugar, se eligió qué tipo de técnica resultaba adecuada para llevar a cabo este TFM. Se optó por implementar un método de tipo genérico, ya que se han obtenido buenos resultados en los estudios más recientes. A continuación, se implementó una etapa de preprocesado para adaptar las imágenes y prevenir efectos de borde indeseados al aplicar los métodos de procesado utilizados en fases posteriores. Posteriormente, se extrajeron características representativas de la calidad de las imágenes mediante técnicas de análisis basadas en procesado digital de imágenes. Se empleó una técnica de selección de características para elegir las más relevantes, que se introdujeron como entrada a un clasificador. La BD empleada fue proporcionada por el Instituto Universitario de Oftalmobiología Aplicada de la Universidad de Valladolid (IOBA) y el Hospital Clínico Universitario de Valladolid (HCUV). Para cada una de las imágenes, se revisaron las anotaciones realizadas por los expertos sobre la calidad de las imágenes según una serie de criterios fijados previamente.

### 3.2 BASE DE DATOS DE RETINOGRAFÍAS

Los materiales utilizados en este TFM incluyen una BD de retinografías con anotaciones de su calidad. Se utilizó la BD creada para el proyecto RTC-2015-3467-1 titulado "Análisis Automático de Imágenes de Fondo de Ojo como implementación a los sistemas de cribado de la retinopatía diabética" (A2IFO). Este proyecto está financiado por el Ministerio de Economía y Competitividad y cofinanciado por la Unión Europea con cargo al fondo FEDER dentro del objetivo temático "Promover el desarrollo tecnológico, la innovación y una investigación de calidad". Las imágenes fueron proporcionadas por el IOBA el HCUV. Oftalmólogos expertos realizaron anotaciones sobre la calidad de las imágenes, la patología que presentaban y, en caso de presentar RD, las lesiones y la severidad de la patología. Esta BD no dispone de segmentaciones de las estructuras oculares ni de las máscaras que delimitan el FOV de cada imagen.

El retinógrafo utilizado para capturar las retinografías fue el modelo TOPCON TRC-NW400, un retinógrafo no midriático de diseño compacto (Figura 3.1). Es completamente automático (alineamiento, foco y disparo) y tiene un FOV de 45° (Topcon, n.d.). La resolución con la que se capturaron las imágenes que componen la BD es de 1956×1934 píxels. Se almacenaron las imágenes en color utilizando el formato JPEG de 24 bits. Por lo tanto, cada píxel se representó con 8 bits para cada uno de los tres canales de color RGB (*Red, Green, Blue*). El FOV es circular y de 1850 píxels de diámetro aproximadamente (Romero-Oraá, 2016).

El protocolo de captura de imágenes utilizado ha sido el protocolo de dos campos del programa de cribado de la RD en Reino Unido. El *National Service Framework for Diabetes* del Reino Unido adoptó este protocolo en 2003 al considerar que, bajo midriasis, se trata un protocolo útil y aceptable tanto para los profesionales sanitarios como para los pacientes (Department of Health, 2002). El protocolo de dos campos se basa en la captura de retinografías en dos campos de la retina: uno centrado en mácula y con la papila en el lado nasal y otro centrado en papila y con la mácula en el lado temporal de la fotografía (Boucher *et al.*, 2003; Scanlon *et al.*, 2003). Cumple con los requisitos exigidos por el Instituto Nacional para la Salud y el Cuidado de la Excelencia (*National Institute for Health and Care Excellence, NICE*) para su utilización en un programa de cribado (Boucher *et al.*, 2003).

La BD cuenta con un total de 2342 imágenes de pacientes sanos y con algún tipo de patología. Para este trabajo se seleccionaron 2107 imágenes tras descartar retinografías duplicadas (perteneían al mismo campo y al mismo ojo que otras) y retinografías de otros campos que no se correspondían con el protocolo de captura o se habían obtenido con un FOV distinto a 45°. En general, se capturaron cuatro imágenes por cada paciente: dos del ojo izquierdo y dos del ojo derecho. Un oftalmólogo anotó todas las imágenes, indicando para



Figura 3.1: Retinógrafo no midriático TOPCON TRC-NW400 (Topcon, n.d.).

cada imagen si tiene la calidad suficiente para ser analizada y, en caso afirmativo, si en la imagen se apreciaba patología. En caso de apreciar signos de RD, se indicaron las lesiones que aparecían. Las patologías presentes en las imágenes son diversas, aunque predominan las retinografías de sujetos con algún grado de RD.

Para evitar en mayor medida la subjetividad entre observadores humanos, se establecieron unos criterios para determinar si una imagen tiene calidad suficiente para ser evaluada. En caso de duda, se consultó la decisión con otros expertos. Los expertos que revisaron la calidad de las imágenes siguieron estos mismos criterios, y la decisión final sobre la calidad surgió del consenso entre varios observadores en estos casos dudosos. Se consideró que una imagen tiene calidad suficiente para ser analizada siempre y cuando cumpla con los siguientes requisitos:

- Los bordes de la papila se muestran bien definidos.
- La red vascular aparece bien definida, especialmente los vasos principales.
- El parénquima retiniano (fondo de la retina) es perfectamente visible.
- La mácula es distinguible.

En casos extremos se consideró que se podían cumplir solo 3 de estas 4 condiciones. Siguiendo estos criterios se clasificaron 1810 imágenes de buena calidad y 297 de calidad insuficiente. Se dividieron todas las imágenes de manera aleatoria en dos conjuntos independientes: entrenamiento y test. El conjunto de entrenamiento estaba formado por 1053 imágenes (905 de buena calidad y 148 de mala calidad), mientras que al conjunto de test se destinaron 1054 imágenes (905 de buena calidad y 149 de mala calidad). Además, se aseguró que todas las imágenes de un mismo paciente quedaran encuadradas en el mismo conjunto. De esta manera se aseguró que las muestras empleadas en los conjuntos de entrenamiento y test son completamente independientes.

### 3.3 PREPROCESADO

En el análisis de retinografías los algoritmos de preprocesado se han utilizado principalmente para mejora de las características de las imágenes, aumentando así la precisión de los métodos aplicados posteriormente. Los métodos de preprocesado se han aplicado típicamente para realizar tareas de corrección de iluminación, normalización y realce (Winder *et al.*, 2009; Besenczi *et al.*, 2016).

En este TFM, la etapa de preprocesado se centró en corregir los efectos de borde que producen algunos métodos de procesado de imagen en los bordes del FOV de las retinografías. Para realizar esta tarea, el algoritmo determina en primer lugar la región circular que ocupa el FOV de manera automática. Después, utiliza la máscara obtenida para llevar a cabo un procedimiento de extensión de los bordes del FOV hacia los extremos de la imagen. De esta manera se evitaban respuestas espurias en zonas cercanas a los límites del FOV al eliminar el borde circular de las imágenes.

### 3.3.1 Obtención de la máscara FOV

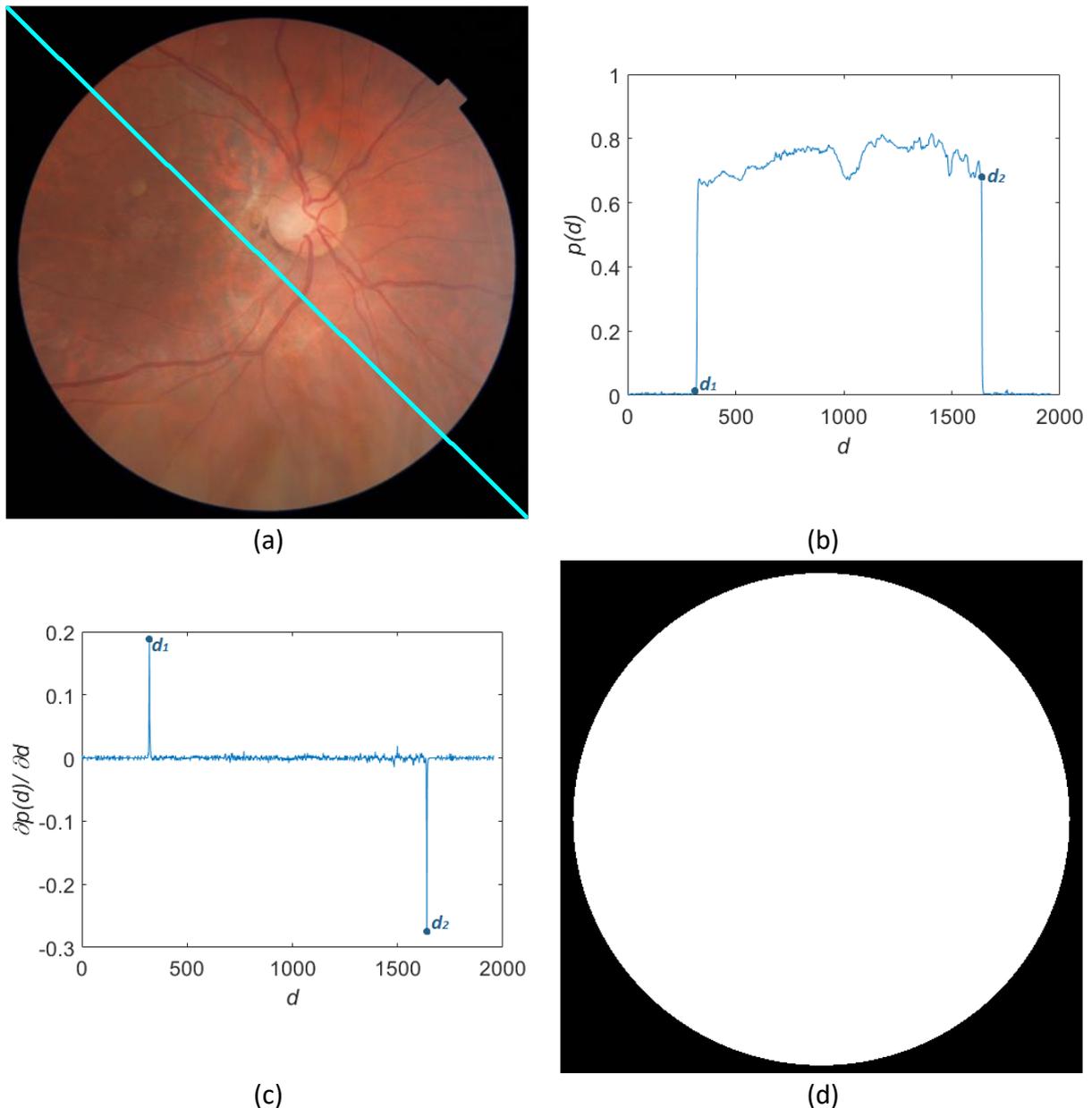
La máscara del FOV se suele utilizar para determinar los píxels de la imagen que muestran la retina. Habitualmente, solo se consideran los píxels dentro de la zona de interés para realizar determinados cálculos y obtener resultados (Fraz *et al.*, 2012). La BD utilizada propia no disponen de máscaras para la FOV, por lo que éstas se obtuvieron mediante un método automático que determina la circunferencia que lo encierra. Este método obtiene una máscara circular a partir del centro y el radio de la circunferencia que lo contiene (García, 2008). El diámetro del FOV también se utilizó para calcular de forma automática algunos parámetros de los métodos de procesado posteriores.

En primer lugar, se aplicó un filtro de mediana sobre el canal  $R$  para eliminar ruido a la vez que se conservan los bordes. Así se evitaron errores provocados por el ruido impulsivo. (Gonzalez and Woods, 2008). Para averiguar el radio, se determina el perfil de intensidad  $p(d)$  del canal  $R$  a lo largo de la línea diagonal que cruza desde la esquina superior izquierda hacia la esquina inferior derecha de la imagen. En la Figura 3.2 (a) se muestra la línea diagonal a través de la que se calcula el perfil  $p(d)$  (Figura 3.2 (b)). En el perfil hay dos cambios bruscos en el nivel de intensidad, uno de valores bajos a valores altos y otro de valores altos a bajos. Para hallar los puntos donde se producen las variaciones, se calcula la derivada del perfil  $p(d)$ . Los puntos donde se produce el máximo ( $d_1$ ) y el mínimo ( $d_2$ ) de la derivada pertenecen al borde del FOV (Figura 3.2 (c)), y se calculan del siguiente modo (García, 2008):

$$d_1 = \operatorname{argmax}_d \left[ \frac{\partial p(d)}{\partial d} \right] \quad (3.1)$$

$$d_2 = \operatorname{argmin}_d \left[ \frac{\partial p(d)}{\partial d} \right] \quad (3.2)$$

La distancia entre  $d_1$  y  $d_2$  es una estimación del diámetro del FOV, a partir del cual se obtiene el radio. Tras calcular una estimación del radio, se determina el centro del FOV. Para ello, primero se aplica el operador de bordes de Canny para obtener los bordes más acusados de la imagen, entre los que se encuentra el borde del FOV (Gonzalez and Woods, 2008). Sobre la imagen de bordes se aplica la transformada de Hough circular para hallar el centro de la circunferencia que delimita el FOV. Se fijó el radio calculado anteriormente y se restringió la búsqueda del centro a una región cuadrada de 20 píxels de lado en el centro de la imagen (García, 2008). Una vez determinados el radio y el centro, se obtiene una máscara con forma



**Figura 3.2:** Ejemplo de obtención de una máscara del FOV circular; (a) Imagen original en la que se ha superpuesto la diagonal a través de la que se calcula el perfil; (b) Gráfica del perfil  $p(d)$ ; (c) Derivada del perfil  $p(d)$  (d) Máscara del FOV obtenida.

de circunferencia que será utilizada como máscara del FOV (García, 2008). En la Figura 3.2 (d) se muestra el resultado de esta técnica.

### 3.3.2 Extensión del FOV

En esta etapa se aplicó un algoritmo que reduce la influencia del borde del FOV sobre los métodos de procesamiento aplicados en etapas posteriores (Soares *et al.*, 2006). Este método se diseñó para reducir respuestas espurias en los bordes del FOV debido al elevado contraste entre el fondo retiniano (en general, de color rojizo) y la región que queda fuera de la apertura

de la cámara (zona de color negro en el borde de la imagen). Su funcionamiento es el siguiente (Soares *et al.*, 2006):

Partiendo de la imagen y la máscara del FOV originales, para cada iteración:

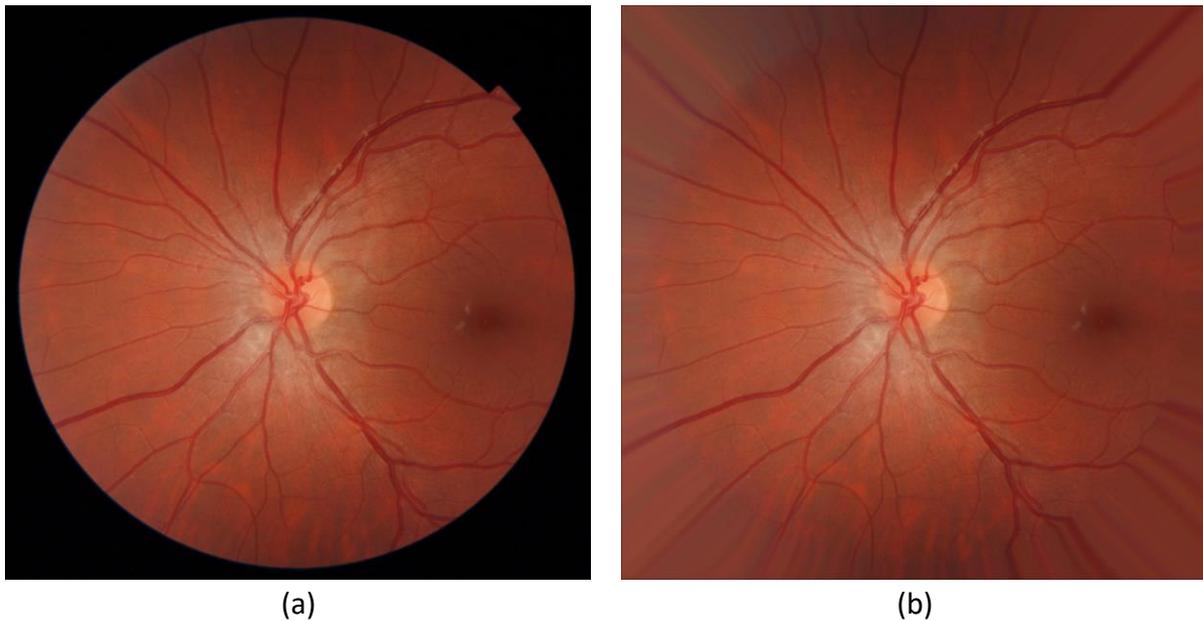
- 1) Se extiende el borde de la máscara 1 píxel hacia fuera mediante el operador de dilatación. Para ello se empleó un elemento estructurante circular de 1 píxel de radio (Gonzalez and Woods, 2008). De esta manera, se amplía la FOV incluyendo nuevos píxels que pasan a considerarse como pertenecientes al FOV. Éstos se pueden identificar fácilmente mediante el operador lógico XOR de la máscara antes y después de aplicar el operador de dilatación. Si la FOV extendida coincide con la FOV de la anterior iteración, el operador XOR es cero en todos los píxels de la imagen (ningún píxel se ha añadido a la FOV). Esto indica que se ha llegado a los extremos de la imagen y el algoritmo se detiene.
- 2) Los píxels de la imagen original que pasan a considerarse dentro del FOV se sustituyen por el promedio de los píxels vecinos (utilizando vecindad de 8 píxels) que estaban en el FOV en la anterior iteración. En la primera iteración se utilizan la imagen y la FOV originales.
- 3) Tras rellenar los píxels del borde, se establece el FOV extendido como el FOV de la imagen.

Se repite de manera iterativa hasta que la apertura de la cámara ha sido completamente eliminada. Al extender de forma circular se llega antes a los bordes de la imagen que a las esquinas. Cuando se llega al borde de la imagen, al aplicar de nuevo el operador de dilatación en el paso 1, los nuevos píxels de la FOV ampliada forman arcos de circunferencia en vez de circunferencias completas. Como se indica en el paso 2, si tras aplicar el operador de dilatación la FOV no ha crecido, el algoritmo se detiene. Esta situación coincide con la situación de que todos los píxels de la máscara extendida son 1.

El resultado de esta técnica se muestra en la Figura 3.3. Se puede observar en la imagen preprocesada que la zona negra exterior a la apertura de la cámara ha sido sustituida por colores similares a los colores del fondo retiniano en el borde de dicha apertura. También se observa que la imagen en el interior de la apertura permanece inalterada. De esta manera se consiguió reducir el contraste entre el interior y el exterior de la apertura de la cámara.

### 3.4 EXTRACCIÓN DE CARACTERÍSTICAS

Tras la etapa de preprocesado, en la etapa de extracción de características se aplicaron algoritmos de análisis de imagen con el objetivo de obtener métricas útiles para evaluar la calidad de la imagen. Se emplearon cuatro algoritmos de análisis para obtener distintos descriptores de calidad de las imágenes. Dos de ellos son algoritmos genéricos de evaluación de calidad sin referencia (*No Reference Image Quality Assessment*, NR-IQA). Éstos métodos se denominan sin referencia o ciegos porque las imágenes a evaluar no se pueden comparan



**Figura 3.3:** Resultado del algoritmo de extensión del FOV. (a) Imagen original; (b) Imagen preprocesada.

con sus versiones no distorsionadas (imágenes de referencia) (Moorthy and Bovik, 2010). En nuestro caso, únicamente se dispone de la imagen capturada y no de la imagen de referencia, por lo que estos métodos resultan adecuados. Los métodos NR-IQA están basados en el enfoque de propósito general *Natural Scenes Statistics* (NSS). La hipótesis de este planteamiento es que las imágenes o escenas naturales, asociadas a imágenes de buena calidad, tienen unas características concretas. Describiendo las imágenes con dichas características, las imágenes naturales ocupan un reducido subespacio dentro de un espacio donde se encuadran todas las imágenes (Moorthy and Bovik, 2010; Mittal *et al.*, 2012, 2013). Hallar dichas características y demostrar que son capaces de diferenciar las imágenes naturales de las distorsionadas es el objetivo de los métodos NR-IQA basados en NSS. Los algoritmos de propósito general aplicados fueron NIQE y SSEQ (*Spatial and Spectral Entropy-based Quality*) (Mittal *et al.*, 2013; Liu *et al.*, 2014).

Siguiendo los planteamientos anteriores, se desarrollaron además otros dos métodos de evaluación de la calidad más centrados en las características de las retinografías. Con el primero de ellos se evaluó la nitidez de las imágenes en base a la transformada *wavelet* continua (*Continuous Wavelet Transform*, CWT). Con el segundo, se evaluó la luminosidad de una retinografía extrayendo el fondo de la misma.

#### 3.4.1 Características basadas en naturalidad

El algoritmo NIQE (Mittal *et al.*, 2013) es un método NR-IQA que cuantifica las desviaciones estadísticas que se producen al evaluar características de las imágenes distorsionadas con respecto de las características de las imágenes naturales. Su principal ventaja es que no requiere el conocimiento a priori sobre las distorsiones que éstas hayan podido sufrir con respecto a las imágenes naturales (Mittal *et al.*, 2013).

El algoritmo NIQE permite obtener un modelo a partir de un conjunto de imágenes naturales (en este caso, retinografías de buena calidad) recopilando características basadas en la distribución de las intensidades de los píxeles a lo largo de pequeñas regiones de las imágenes (Mittal *et al.*, 2013). Dichas características se ajustan a un modelo gaussiano multivariante (*Multivariate Gaussian Model*, MVG) que se utiliza como referencia para evaluar las imágenes. Con los datos obtenidos de las regiones de cada imagen, se creó un modelo ajustándolos a una función de densidad de probabilidad MVG (Mittal *et al.*, 2013):

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{v})^T \Sigma^{-1}(\mathbf{x}-\mathbf{v})} \quad (3.3)$$

Un modelo MVG queda definido con el vector de medias  $\mathbf{v}$  y la matriz de covarianzas  $\Sigma$  de los datos. Se creó un modelo MVG a partir de las imágenes de buena calidad del conjunto de entrenamiento, que se quedó definido mediante  $\mathbf{v}_M$  y  $\Sigma_M$ . Este modelo se tomó como referencia para obtener la calidad de una imagen. El índice NIQE de una imagen se calcula mediante la distancia (Mittal *et al.*, 2013):

$$D_{NIQE} = \sqrt{(\mathbf{v}_M - \mathbf{v}_I)^T \left( \frac{\Sigma_M + \Sigma_I}{2} \right)^{-1} (\mathbf{v}_M - \mathbf{v}_I)} \quad (3.4)$$

Donde  $\mathbf{v}_I$  y  $\Sigma_I$  representan el modelo MVG creado a partir de la imagen a evaluar. Para evaluar la calidad de una imagen se deben obtener dos modelos MVG:  $(\mathbf{v}_M, \Sigma_M)$  y  $(\mathbf{v}_I, \Sigma_I)$ , y el indicador de calidad NIQE es la distancia entre ellos (Mittal *et al.*, 2013). Por lo tanto, cuanto menor sea esta distancia, mejor calidad tiene la imagen. Esta distancia fue una de las características empleadas en este TFM para evaluar la calidad de las retinografías.

La obtención de los dos modelos MVG descritos anteriormente se lleva a cabo de la misma manera. La única diferencia es que para obtener el modelo MVG de referencia  $(\mathbf{v}_M, \Sigma_M)$  se emplean regiones extraídas de una colección de imágenes previamente definidas (imágenes de buena calidad del conjunto de entrenamiento en nuestro caso), mientras que para obtener el modelo MVG  $(\mathbf{v}_I, \Sigma_I)$  sólo se emplean regiones extraídas de la imagen a evaluar. La imagen de entrada,  $I(x,y)$ , fue el canal de luminosidad del modelo YCbCr, adecuado para el análisis de la calidad en imágenes comprimidas en formato JPEG (Mittal *et al.*, 2013). El algoritmo de obtención de los modelos MVG constaba de varias etapas. En primer lugar, se normalizó la imagen  $I(x,y)$  para obtener una imagen normalizada  $\hat{I}(x,y)$  restando a la imagen original la media local,  $\mu(x,y)$ , y dividiendo el resultado entre la desviación típica local,  $\sigma(x,y)$  (Mittal *et al.*, 2013):

$$\hat{I}(x,y) = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + \varepsilon} \quad (3.5)$$

La constante positiva  $\epsilon \neq 0$  evita que se anule el denominador. La media y la desviación típica local se obtuvieron mediante un filtro gaussiano  $w(k,l)$  (Mittal *et al.*, 2013):

$$\mu(x, y) = \sum_{k=-K}^K \sum_{l=-L}^L w(k, l) \cdot I(x - k, y - l) \quad (3.6)$$

$$\sigma(x, y) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w(k, l) \cdot [I(x - k, y - l) - \mu(x, y)]^2} \quad (3.7)$$

El filtro  $w(k,l)$   $\{k = -K, \dots, K; l = -L, \dots, L\}$  ( $K = L = 3$ ) es una función de pesos obtenida muestreando la función gaussiana 2D con simetría circular y de desviación típica 1 (Mittal *et al.*, 2013). Después, se divide la imagen en regiones de tamaño  $P \times P$  píxels y se seleccionan únicamente una porción del total de regiones creadas. La razón para seleccionar únicamente algunas regiones es que en las imágenes existen zonas más nítidas que otras y las valoraciones de calidad normalmente se realizan a partir de las zonas más nítidas, asumiendo que una parte de la imagen está desenfocada. De esta manera, se extraen características únicamente de las zonas más relevantes para el observador. Para seleccionar las regiones de interés, se calcula inicialmente la nitidez local de cada región,  $\delta(n)$ , como (Mittal *et al.*, 2013):

$$\delta(n) = \sum_{x=1}^P \sum_{y=1}^P \sigma(x, y) \quad (3.8)$$

Posteriormente, se fijó un umbral de contraste  $T$  (entre 0 y 1) y se seleccionaron las regiones que cumplían (Mittal *et al.*, 2013):

$$\delta(n) \geq T \cdot \max_n \delta(n) \quad (3.9)$$

Tras seleccionar las regiones que superan el umbral de contraste, se pasó a caracterizar cada una de ellas. Se asume que la distribución de los coeficientes de  $\hat{I}(x,y)$  dentro de cada una de las regiones seleccionadas sigue la distribución gaussiana generalizada (*Gaussian Generalized Distribution, GGD*) con media nula (Mittal *et al.*, 2013):

$$f(x; \alpha, \beta) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-\left(\frac{|x|}{\beta}\right)^\alpha} \quad (3.10)$$

La función  $\Gamma(a)$  se define como (Mittal *et al.*, 2013):

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \quad (3.11)$$

Se estimaron los parámetros  $\alpha$  y  $\beta$  de la GGD que ajusta a los coeficientes  $\hat{l}(x,y)$  en cada región. Posteriormente, se calcularon 4 productos, correspondientes a productos entre coeficientes adyacentes en 4 direcciones (horizontal, vertical y dos diagonales):  $\hat{l}(x,y) \cdot \hat{l}(x,y+1)$ ,  $\hat{l}(x,y) \cdot \hat{l}(x+1,y)$ ,  $\hat{l}(x,y) \cdot \hat{l}(x+1,y+1)$  y  $\hat{l}(x,y) \cdot \hat{l}(x+1,y-1)$ . Las distribuciones de estos 4 productos siguen distribuciones gaussianas generalizadas asimétricas (*Asymmetric Generalized Gaussian Distribution*, AGGD) de moda nula (Mittal *et al.*, 2013):

$$f(x; \gamma, \beta_l, \beta_r) = \begin{cases} \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} e^{-\left(\frac{-x}{\beta_l}\right)^\gamma}, & x \leq 0 \\ \frac{\gamma}{(\beta_l + \beta_r)\Gamma(\frac{1}{\gamma})} e^{-\left(\frac{x}{\beta_r}\right)^\gamma}, & x \geq 0 \end{cases} \quad (3.12)$$

Se estimaron los parámetros  $\gamma$ ,  $\beta_l$  y  $\beta_r$  de cada AGGD que ajustan a cada uno de los cuatro productos en cada región. Además, se calculó la media de la AGGD como (Mittal *et al.*, 2013):

$$\eta = (\beta_l - \beta_r) \frac{\Gamma(\frac{2}{\gamma})}{\Gamma(\frac{1}{\gamma})} \quad (3.13)$$

Por lo tanto, se extrajeron 4 parámetros de cada AGGD ( $\gamma$ ,  $\beta_l$ ,  $\beta_r$  y  $\eta$ ). Como se consideraron 4 productos de coeficientes adyacentes, se obtuvieron 16 parámetros de las AGGD. Añadiendo los 2 parámetros obtenidos de la GGD de los coeficientes normalizados, se obtuvieron 18 parámetros (Mittal *et al.*, 2013) en el proceso. Se repitió el procedimiento escalando las imágenes a la mitad de su resolución para realizar un análisis multi-escala. De este modo, se obtuvieron un total de 36 parámetros (Mittal *et al.*, 2013) basados en la naturalidad para cada imagen.

Como se puede observar, el algoritmo NIQE requiere el ajuste de dos parámetros: el tamaño de bloque  $P$  y el umbral de contraste  $T$ . Se eligieron los valores  $P=64$  y  $T=0.1$  de manera empírica, ya que producía mejor capacidad discriminatoria entre imágenes de buena calidad y mala calidad en el conjunto de entrenamiento. En general, estos parámetros no son críticos, los resultados no variaron de forma significativa al considerar otros valores de  $P$ . Por otra parte, se obtuvieron los mejores resultados para valores de  $T$  cercanos a 0.

### 3.4.2 Características basadas en la entropía espacial y la entropía espectral

En este TFM; se empleó un segundo método (SSEQ) para obtener descriptores relativos a la calidad de las imágenes. El método SSEQ (Liu *et al.*, 2014) es un algoritmo NR-IQA de propósito general que obtiene descriptores de la calidad de las imágenes a partir del cálculo de entropías locales en pequeñas regiones de la imagen. Emplea el enfoque NSS al establecer la hipótesis de que la entropía local de las imágenes naturales varía con respecto a la entropía local de las imágenes distorsionadas (Liu *et al.*, 2014). Estas variaciones de entropía local pueden apreciarse tanto en el dominio espacial como en el dominio de la frecuencia. En el método SSEQ también se establece que la entropía calculada a nivel global no revela las diferencias producidas por la distorsión, puesto que no recoge información sobre las variaciones a lo largo de la imagen (Liu *et al.*, 2014). En este caso, el algoritmo se aplicó sobre el canal verde de RGB, frecuentemente empleado en análisis de retinografías porque es el canal con el mejor contraste entre el fondo retiniano y las estructuras oculares (Fraz *et al.*, 2012).

Las entropías locales se obtienen de cada región aplicando la entropía de Shannon tanto en el dominio espacial como en el dominio espectral. Para cada imagen, se calcularon ambas entropías a lo largo de todas las regiones de la imagen, no se seleccionó un subconjunto de estas como en el caso anterior. A continuación, se extrajeron dos parámetros de la distribución de los datos: la media y la asimetría. Al igual que en el caso anterior, el algoritmo consta de varias etapas que se describen a continuación (Liu *et al.*, 2014).

En primer lugar, se obtuvieron tres imágenes de partida, la original y dos imágenes escaladas a resoluciones inferiores a la original. Esto permite llevar a cabo análisis multi-escala de la imagen empleando tres representaciones: alta resolución (escala 1, 1956×1934 píxels), media resolución (escala 2, 978×967 píxels) y baja resolución (escala 3, 652×645 píxels). Estas resoluciones se obtuvieron tras aplicar los factores de escala 1, 1/2 y 1/3 respectivamente (Liu *et al.*, 2014). Posteriormente, la imagen a evaluar se dividió en regiones de tamaño  $M \times M$  píxels. La entropía espacial (*SpacEn*) de los píxels de una región se calculó como la entropía de Shannon en dicha región (Liu *et al.*, 2014):

$$SpacEn = - \sum_i p(i) \cdot \log_2 p(i) \quad (3.14)$$

El parámetro  $p(i)$  representa la frecuencia relativa del nivel de intensidad  $i$ , por lo que se asemeja a la probabilidad de que el nivel de intensidad  $i$  aparezca en un píxel de la región.

Para calcular la entropía espectral ( $SpecEn$ ) de una región, se calculó previamente una representación del espectro de frecuencia de la misma. Para ello, se utilizó la transformada discreta del coseno (*Discrete Cosine Transform*, DCT) 2D de la región  $c(x,y)$  (Liu *et al.*, 2014):

$$C(k, l) = \frac{2}{M} \sum_{x=1}^M \sum_{y=1}^M c(x, y) \cdot \cos \left[ \frac{(2x + 1)k\pi}{2M} \right] \cdot \cos \left[ \frac{(2y + 1)l\pi}{2M} \right] \quad (3.15)$$

Donde  $x$  y  $y$  son las coordenadas de cada uno de los píxeles de la región, mientras que  $k$  y  $l$  son los índices de la frecuencia discreta horizontal y vertical respectivamente, en el dominio de la DCT 2D (Lim, 1990). Se puede obtener una densidad espectral de potencia normalizada calculando el cuadrado de los coeficientes de la DCT 2D y dividiendo entre la suma de ellos, excluyendo la componente continua  $C(0,0)$  (Liu *et al.*, 2014):

$$P(k, l) = \frac{[C(k, l)]^2}{\sum_k \sum_l [C(k, l)]^2}, (k, l) \neq (0,0) \quad (3.16)$$

La  $SpecEn$  de la región se calculó como la entropía de Shannon de la distribución  $P(k,l)$  (Liu *et al.*, 2014):

$$SpecEn = - \sum_k \sum_l P(k, l) \cdot \log_2 P(k, l) \quad (3.17)$$

Una vez que se obtuvieron las  $SpacEn$  y las  $SpecEn$  de todos los bloques de la imagen, se seleccionaron los datos entre los percentiles 15% y 85% de estos parámetros para cada imagen (Liu *et al.*, 2014). De esta manera, se analizaron los datos más cercanos a la media de la distribución, evitando los valores extremos e incrementando la robustez (Liu *et al.*, 2014). Finalmente, se calcularon la media y la asimetría de los datos de  $SpacEn$  y  $SpecEn$  seleccionados (Liu *et al.*, 2014):

$$SpacEn_{MED} = \frac{1}{N} \sum_{i=1}^N SpacEn_{15-85}(i) \quad (3.18)$$

$$SpacEn_{ASIM} = \frac{\frac{1}{N} \sum_{i=1}^N [SpacEn_{15-85}(i) - \overline{SpacEn_{15-85}}]^3}{\left( \sqrt{\frac{1}{N} \sum_{i=1}^N [SpacEn_{15-85}(i) - \overline{SpacEn_{15-85}}]^2} \right)^3} \quad (3.19)$$

$$SpecEn_{MED} = \frac{1}{N} \sum_{i=1}^N SpecEn_{15-85}(i) \quad (3.20)$$

$$SpecEn_{ASIM} = \frac{\frac{1}{N} \sum_{i=1}^N [SpecEn_{15-85}(i) - \overline{SpecEn_{15-85}}]^3}{\left( \sqrt{\frac{1}{N} \sum_{i=1}^N [SpecEn_{15-85}(i) - \overline{SpecEn_{15-85}}]^2} \right)^3} \quad (3.21)$$

De esta manera, se obtuvieron 12 características (2 medias y 2 asimetrías por cada una de las 3 escalas analizadas).

El único parámetro que se debe fijar en este método es el tamaño de bloque. Se eligió el tamaño de bloque  $M=32$  píxeles de manera empírica, al proporcionar mejor diferenciación entre las retinografías de buena calidad y las de calidad insuficiente. Los resultados obtenidos dependen en gran medida de la elección de  $M$ , por lo que resulta un parámetro importante. Se realizaron pruebas con los valores de  $M = 8, 16, 32$ , y  $64$ , siendo  $M=32$  el caso en el que se obtuvieron los mejores resultados.

### 3.4.3 Características basadas en la transformada *wavelet* y la varianza local

La CWT 2D es un método de procesamiento ampliamente utilizado en análisis y segmentación de imágenes (Antoine *et al.*, 1993). Permite la descomposición de una imagen en representaciones más sencillas atendiendo al factor de escala utilizado. Entre las aplicaciones más típicas de la CWT 2D se encuentran la detección de contornos, el reconocimiento de patrones o el filtrado. La CWT 2D de una imagen  $I$  se define como (Antoine *et al.*, 1993):

$$T_{\Psi}(I, \mathbf{b}, \theta, s) = C_{\Psi}^{-\frac{1}{2}} \cdot \frac{1}{s} \int \Psi^* \left( \frac{r_{-\theta}(\mathbf{x} - \mathbf{b})}{s} \right) I(\mathbf{x}) d^2\mathbf{x}, \quad (3.22)$$

donde  $\mathbf{x}=(x,y)$  son las coordenadas de la imagen  $I$ ,  $r_{-\theta}$  es el operador de giro de la *wavelet* en un ángulo  $\theta$ ,  $C_{\Psi}$  es una constante de normalización,  $\mathbf{b}$  es el vector de desplazamiento y  $s$  es el parámetro de escala. La función  $\Psi^*$  es la *wavelet* madre conjugada, que puede ser diferente para cada aplicación (Antoine and Murenzi, 1996). En este TFM se escogió la *wavelet* Sombrero Mexicano o *Mexican Hat*. Tiene la siguiente expresión (Antoine and Murenzi, 1996):

$$\Psi_{MH}(x, y, s) = \frac{1}{\pi \cdot s^4} \left( 1 - \frac{x^2 + y^2}{2s^2} \right) e^{-\frac{x^2 + y^2}{2s^2}} \quad (3.23)$$

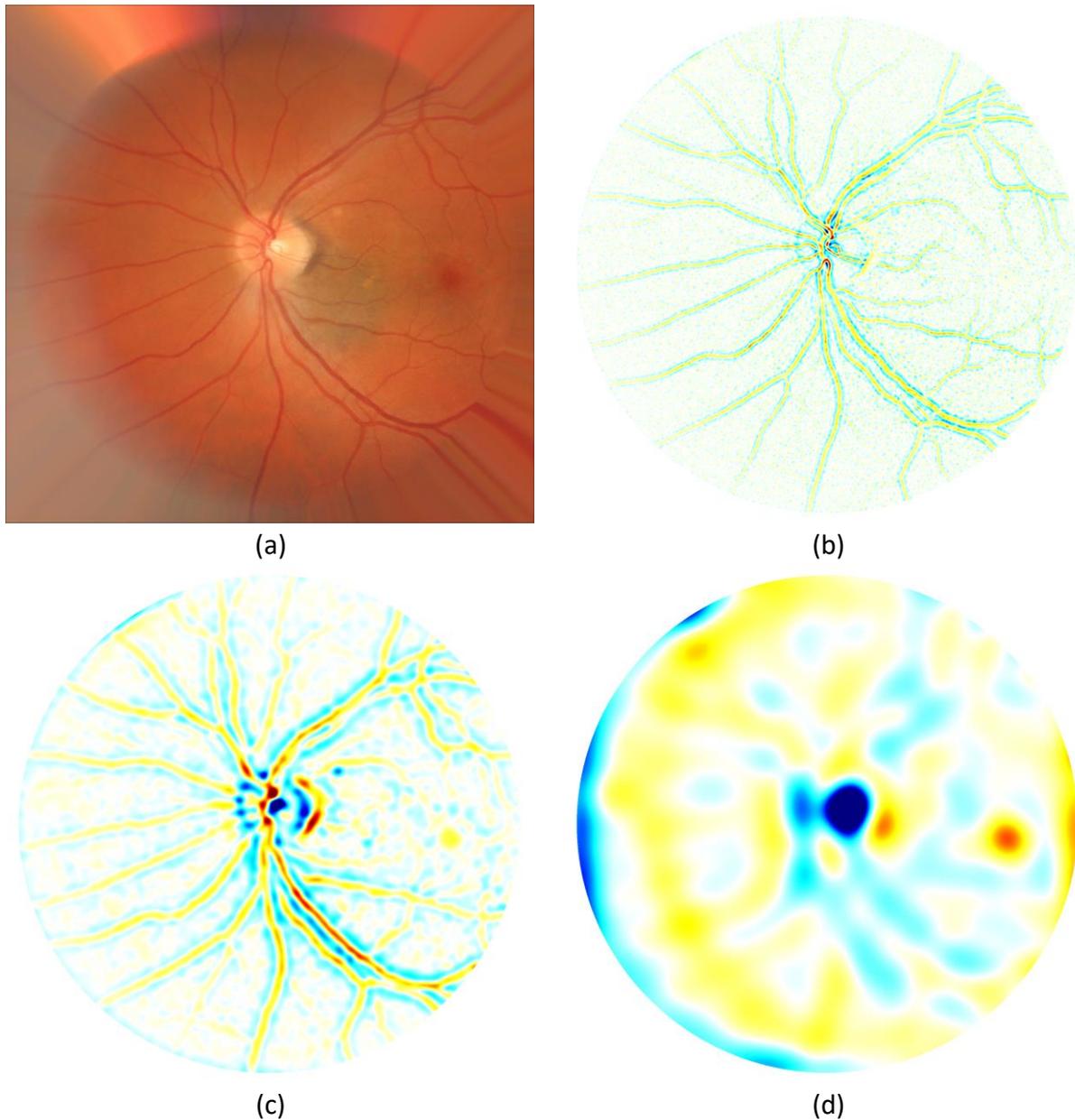
Esta *wavelet* es la implementación del operador Laplaciano de Gaussiana (*Laplacian of Gaussian*, LoG) u operador Marr-Hildreth como *wavelet* y se ha utilizado ampliamente como detector de bordes (Antoine and Murenzi, 1996; Rangayyan, 2005). La *wavelet Mexican Hat* permite detectar los contornos de las estructuras de la imagen en múltiples escalas. Es una *wavelet* isotrópica (invariante a la rotación,  $\theta$ ), por lo que detecta los contornos en cualquier dirección (Antoine and Murenzi, 1996; Rangayyan, 2005). Según la escala  $s$ , la CWT responde a diferentes bordes de las estructuras oculares presentes en las retinografías, como los vasos sanguíneos o la papila (Soares *et al.*, 2006; Fathi and Naghsh-Nilchi, 2013). Cuanto más elevada sea la respuesta de la CWT, mayor es el contraste de las estructuras con el fondo de la imagen y, por lo tanto, mejor es la calidad de la imagen. Por el contrario, la ausencia de objetos con bordes marcados reduciría o incluso anularía la respuesta de la CWT. Además, el empleo de la *wavelet Mexican Hat* reduce el ruido a la vez que conserva los bordes con respecto a la imagen original (Rangayyan, 2005), por lo que se analizaron representaciones menos ruidosas de la imagen.

Se aplicó la CWT sobre el canal verde de las imágenes y se analizó la respuesta para las escalas  $s = 2, 4, 8, 16, 32$  y  $64$ . Estos tamaños se eligieron para detectar los contornos de las diferentes estructuras oculares (Niemeijer *et al.*, 2006). En la Figura 3.4 se observa el resultado de aplicar esta técnica sobre una imagen de ejemplo. Los colores oscuros en la Figura 3.4(b-d) representan respuestas elevadas, tanto positivas (rojo) como negativas (azul). Posteriormente, se calculó la entropía de Shannon de la respuesta para cada escala ( $CWT\_Ent$ ). Para evitar desviaciones en la medida, solo se tuvieron en cuenta para el cálculo los píxels dentro de la máscara FOV (Abdel-Hamid *et al.*, 2016):

$$CWT\_Ent(s) = - \sum_{T_{\Psi}(s)} p(T_{\Psi}) \cdot \log_2 p(T_{\Psi}) \quad (3.24)$$

Un valor de entropía elevado para una escala  $s$  indicaría una mayor irregularidad en la respuesta de la CWT, es decir, se correspondería con la presencia de bordes fuertes de tamaño similar a  $s$  en la imagen. La entropía de la CWT para cada valor de  $s$  considerado se tomó como característica para evaluar la calidad de la imagen. Por lo tanto, de la CWT se obtuvieron 6 características adicionales para evaluar la calidad de las retinografías.

Por otra parte, para analizar las respuestas de la CWT con mayor profundidad se empleó un procedimiento basado en el cálculo de la varianza local (Aja-Fernandez *et al.*, 2006).



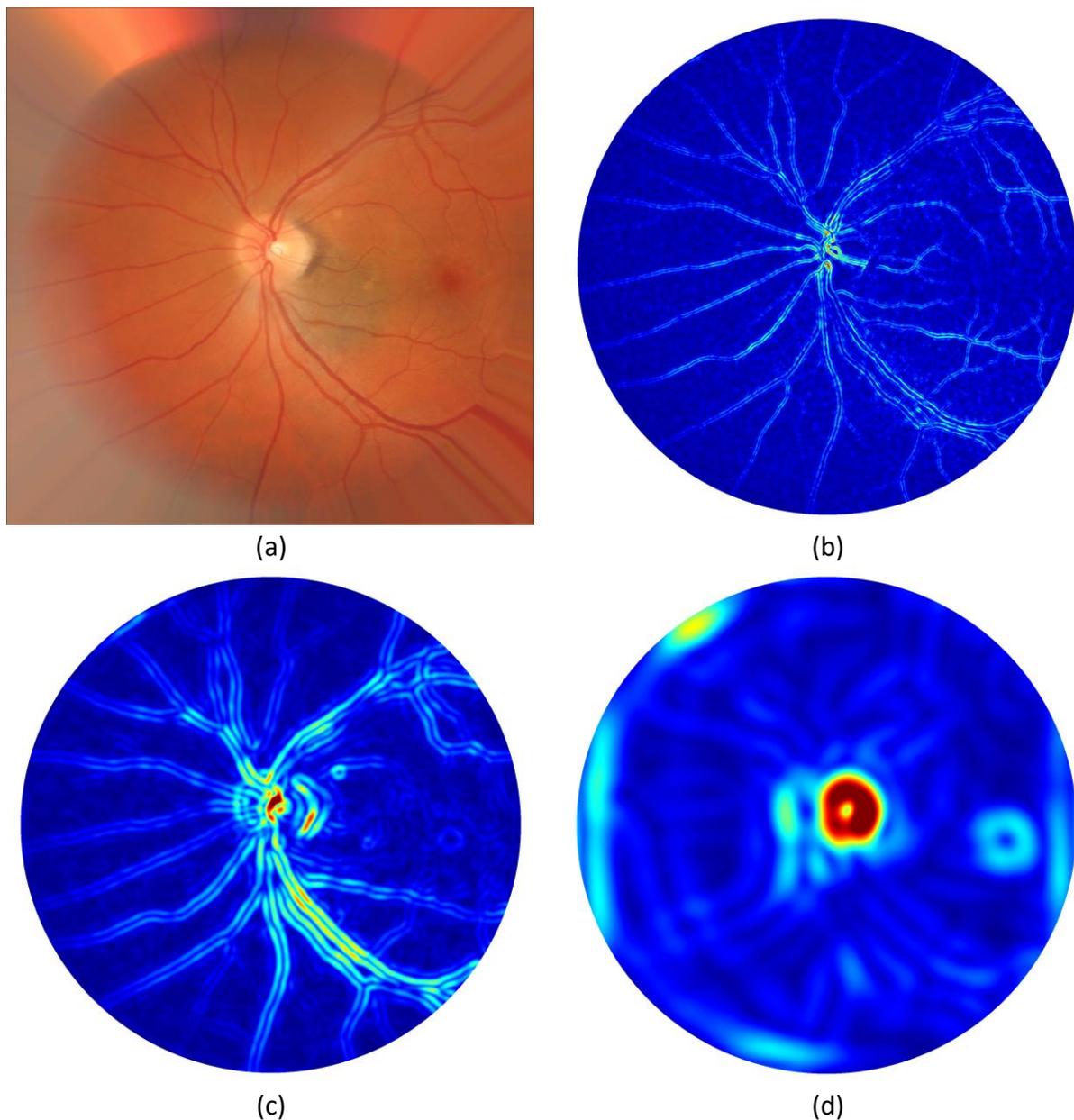
**Figura 3.4:** Respuestas de la CWT para diferentes escalas. (a) Retinografía preprocesada en color; (b) Respuesta para  $s=4$ ; (c) Respuesta para  $s=16$ ; (d) Respuesta para  $s=64$ .

Se filtró la respuesta de la CWT para cada escala con un filtro de desviación típica (Aja-Fernandez *et al.*, 2006). La respuesta de este filtro  $\sigma(x,y,s)$  es la desviación típica de los píxeles en una vecindad  $N(x,y)$  del píxel considerado  $(x,y)$  (Aja-Fernandez *et al.*, 2006):

$$\mu(x, y, s) = \sum_{(i,j) \in N(x,y)} T_{\Psi}(I, x - i, y - j, s) \quad (3.25)$$

$$\sigma(x, y, s) = \sqrt{\sum_{(i,j) \in N(x,y)} [T_{\Psi}(I, x - i, y - j, s) - \mu(x, y, s)]^2} \quad (3.26)$$

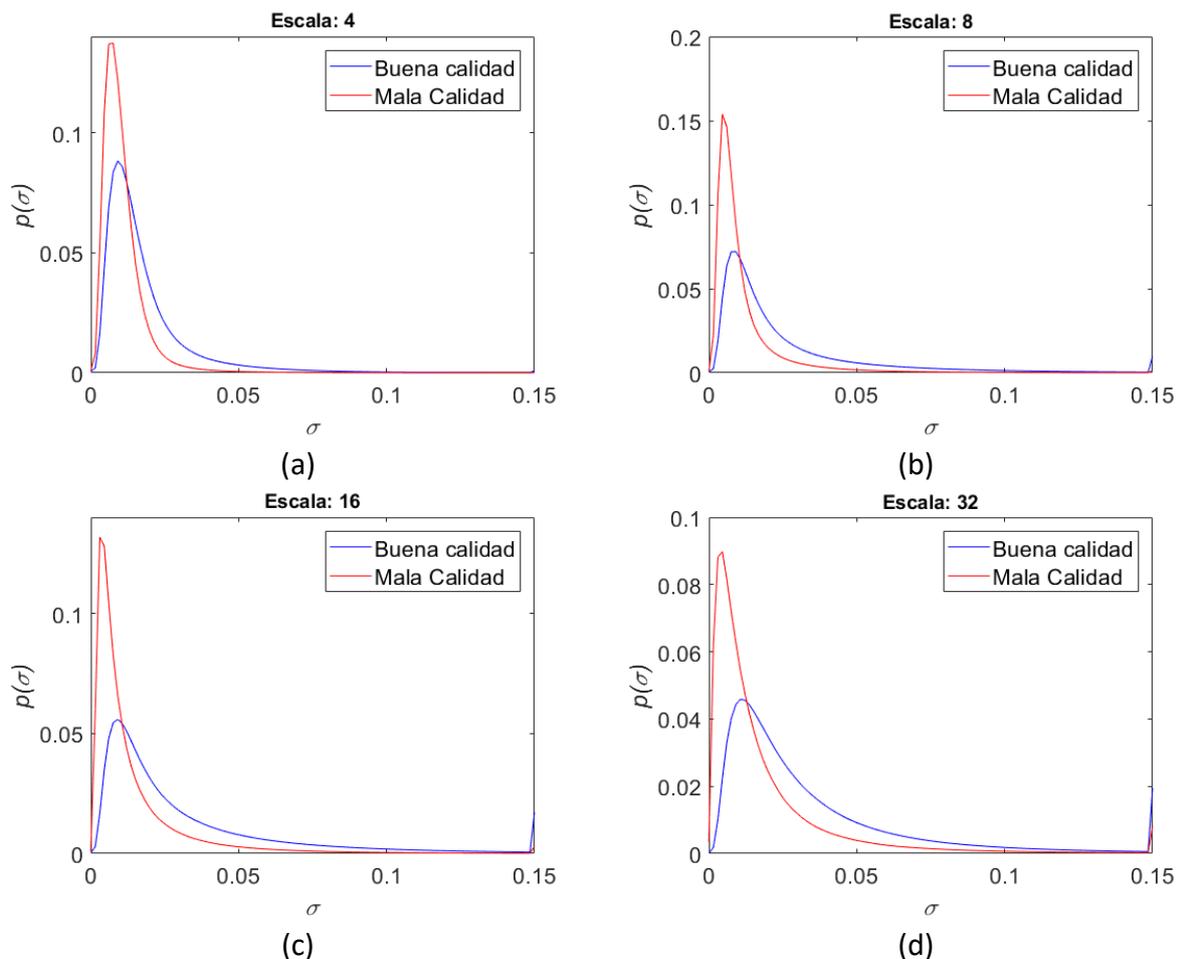
En la Figura 3.5 se observa el resultado obtenido para las representaciones mostradas en la Figura 3.4. En la Figura 3.5(b-d), los tonos azulados representan desviación típica local baja, mientras que los tonos más rojizos representan desviación típica local más elevada. El tamaño de  $N(x,y)$  debe elegirse conforme al tamaño de los objetos que se quieren analizar



**Figura 3.5:** Desviación típica local de la CWT para diferentes escalas. (a) Retinografía preprocesada en color; (b) Resultado para  $s=4$ ; (c) Resultado para  $s=16$ ; (d) Resultado para  $s=64$ .

(Aja-Fernandez *et al.*, 2006). Se eligió  $N(x,y)$  con forma circular y de radio  $r = s$  para cada escala. Se utilizó este tamaño del filtro de desviación típica porque las respuestas de la CWT son imágenes con unas variaciones más suaves cuanto mayor es el parámetro  $s$ . Para poder cuantificar estas variaciones, es necesario emplear un filtro de tamaño mayor a medida que aumenta  $s$ .

Analizando la respuesta del filtro, se observó que las desviaciones típicas locales normalmente siguen una distribución tipo Rayleigh para todas las escalas. Esto se puede observar en la Figura 3.6, donde se muestran las distribuciones promedio de  $\sigma(x,y,s)$  para las imágenes de buena y mala calidad del conjunto de entrenamiento. Como se puede apreciar, para las imágenes de mala calidad la distribución decae rápidamente conforme se incrementa  $\sigma$ . Por el contrario, para las imágenes de buena calidad la distribución decae más lentamente a medida que aumenta  $\sigma$ . Por ello, para caracterizar la distribución de  $\sigma$  se calcularon la media ( $CWT\_Med$ ) y la desviación típica ( $CWT\_Desv$ ) de la respuesta del filtro para cada escala. Se añadieron así 12 características (2 características por cada escala) para evaluar la calidad de las retinografías.



**Figura 3.6:** Distribuciones promedio de la desviación típica local para diferentes escalas. (a) Distribuciones para  $s=4$ ; (b) Distribuciones para  $s=8$ ; (c) Distribuciones para  $s=16$ ; (d) Distribuciones para  $s=32$ .

$$CWT\_Med(s) = \sum_{(x,y) \in FOV} \sigma(x, y, s) \quad (3.27)$$

$$CWT\_Desv(s) = \sqrt{\sum_{(x,y) \in FOV} [\sigma(x, y, s) - CWT\_Med(s)]^2} \quad (3.28)$$

### 3.4.4 Características basadas en la luminosidad del fondo retiniano

En muchos casos, las retinografías no tienen calidad suficiente para ser analizadas debido a que la imagen no se ha capturado con un nivel de luz adecuado. En estos casos, las retinografías aparecen parcial o totalmente oscurecidas, lo que imposibilita su análisis (Zhou *et al.*, 2017; Mitra *et al.*, 2018). En muchos casos, la calidad de la retinografía es aceptable con una luminosidad baja, pero en el caso de la detección de la RD, una luminosidad baja puede ocultar lesiones u otros signos de patología (Saha *et al.*, 2018).

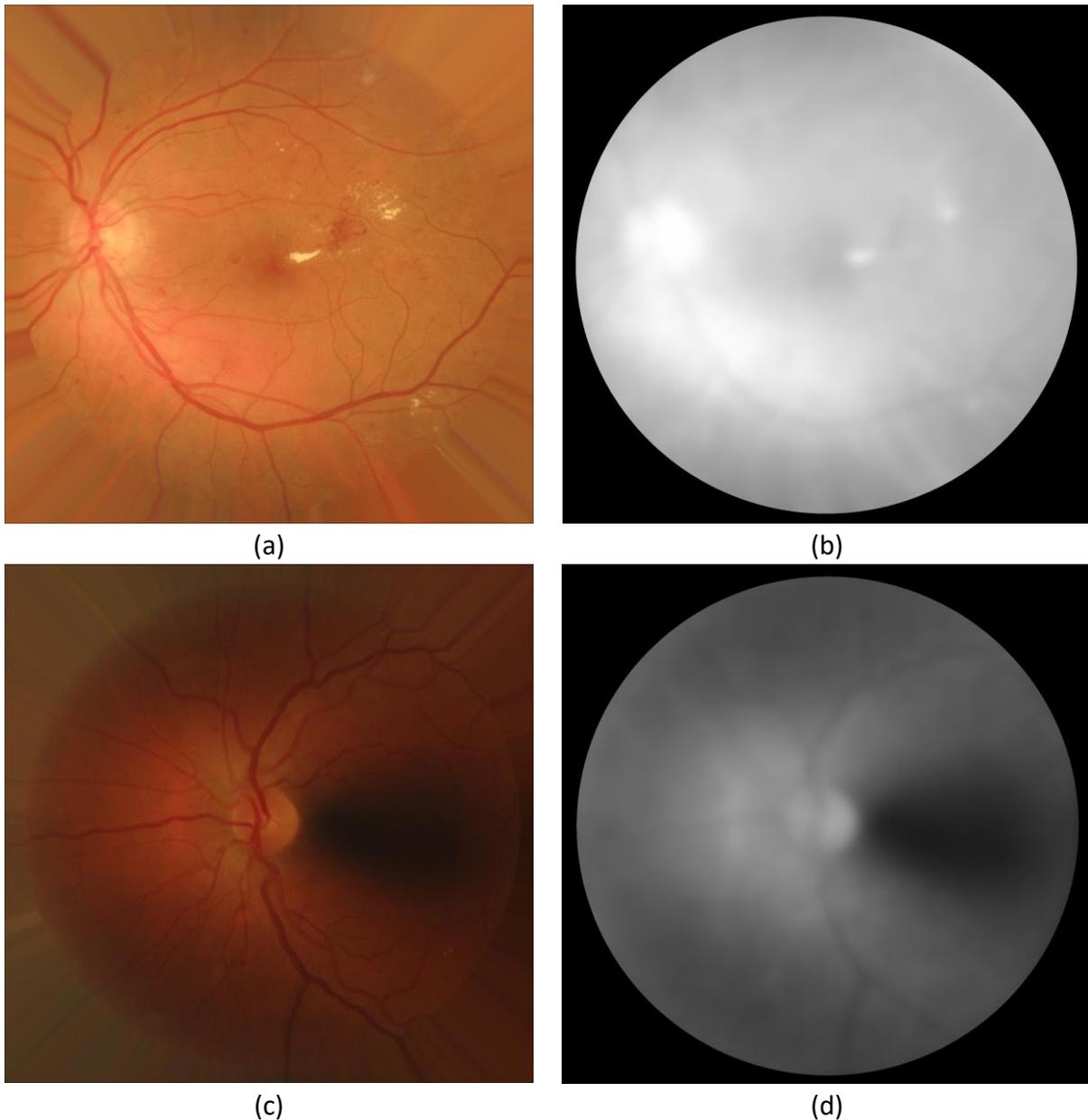
Por ello, en este TFM se analizó también la luminosidad del fondo de las imágenes como indicador de calidad. En este sentido, es posible utilizar distintas representaciones en escala de grises de la imagen. Es ampliamente conocido que el canal verde del modelo de color RGB es el más empleado, ya que es el canal en el que las estructuras oculares y las lesiones se muestran con mayor contraste (Soares *et al.*, 2006; Fraz *et al.*, 2012). Sin embargo, para evaluar la luminosidad de la imagen es más adecuado emplear otros espacios de color como YCbCr, L\*a\*b\* o HSV (*Hue Saturation Value*), que recogen mejor la percepción de la luminosidad y el color respecto al espacio de color RGB (Zhou *et al.*, 2017). Todos estos espacios se analizaron en el TFM y, finalmente, se escogió el espacio de color HSV porque representa mejor las diferencias entre las zonas más claras y oscuras de la imagen (Zhou *et al.*, 2017).

La técnica empleada se basó en extraer el fondo de la imagen y, posteriormente, analizar su histograma. En primer lugar, se extrajo el canal V del espacio de color HSV, que representa la luminosidad de la imagen (Cuevas *et al.*, 2010). La representación del canal V es una imagen que aparece habitualmente oscurecida en la mácula y en el trazado de algunos vasos sanguíneos, que aparecen con poco contraste. Por el contrario, el parénquima retiniano y la papila aparecen con una luminosidad alta. En las imágenes con una iluminación pobre, la mácula habitualmente no se distingue debido a que gran parte del parénquima retiniano aparece oscurecido. Para obtener el fondo  $F(x,y)$  de la imagen V, se aplicaron un filtro de medianas y, posteriormente, un filtro gaussiano (Marín *et al.*, 2011):

$$V_{MED}(x, y) = \underset{(i,j) \in N(x,y)}{\text{med}} V(i, j) \quad (3.29)$$

$$F(x, y) = \sum_{i=-8\sigma}^{8\sigma} \sum_{j=-8\sigma}^{8\sigma} g(i, j) \cdot V_{MED}(x - i, y - j) \quad (3.30)$$

El resultado de esta técnica se muestra en la Figura 3.7. Para la imagen de buena calidad (fila superior), el fondo extraído del canal V se observa con elevada luminosidad y uniforme, incluso saturado en determinadas zonas. Por el contrario, en la imagen de calidad insuficiente (fila inferior) se observan zonas oscurecidas, y la luminosidad no es uniforme a lo

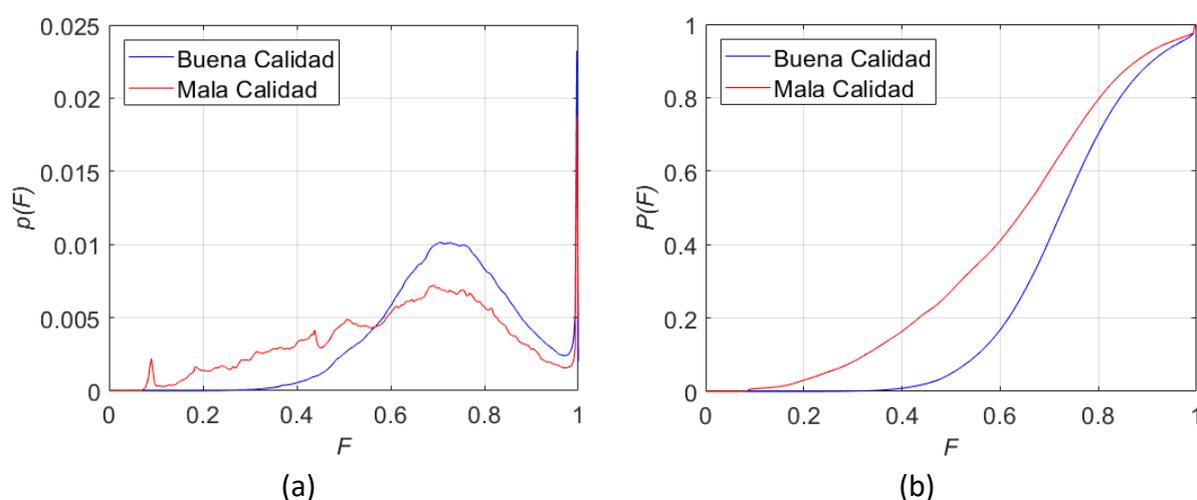


**Figura 3.7:** Extracción del fondo empleando el canal V. (a) Retinografía de buena calidad; (b) Luminosidad del fondo de la imagen (a); (c) Retinografía de mala calidad; (d) Luminosidad del fondo de la imagen (b).

largo de la imagen. El filtro de mediana elimina el ruido impulsivo tomando para cada píxel la mediana de los valores de  $V$  en una vecindad cuadrada  $N(x,y)$  de tamaño  $5 \times 5$  píxels del punto  $(x,y)$  considerado (Gonzalez and Woods, 2008). Por su parte, el filtro gaussiano elimina las variaciones rápidas de  $V$ , que representan vasos sanguíneos o lesiones de pequeño tamaño. El parámetro  $\sigma$  del filtro gaussiano se fijó de manera empírica en 19 píxels, que se corresponde con la anchura de los principales vasos sanguíneos visibles en las retinografías. La imagen resultante es una representación del canal  $V$  en el que se han eliminado el ruido y detalles pequeños conservando el fondo.

Para tener en cuenta estas diferencias, se analizaron las distribuciones de la luminosidad en  $F(x,y)$  para imágenes de buena y mala calidad. Se observó que  $F(x,y)$  toma valores más elevados para las imágenes de buena calidad que para las imágenes de mala calidad. En la Figura 3.8 (a) se muestra la distribución de  $F(x,y)$  como una función de densidad  $p(F)$ . La función de distribución acumulativa de  $F(x,y)$ ,  $P(F)$  se muestra en la Figura 3.8 (b). El eje de abscisas representa los valores de  $F(x,y)$ , y el eje de ordenadas representa la fracción de píxels con intensidad menor o igual a los valores de luminosidad. El percentil es el valor de intensidad para el que un porcentaje del total de píxels tienen intensidad menor o igual que dicho valor.

En la Figura 3.8(b) se observa una gran diferencia entre imágenes de buena y mala calidad en los valores de luminosidad correspondientes a los percentiles más bajos, entre 0% y 20%. Por ello, para caracterizar la luminosidad de las imágenes, se tomaron como características los percentiles de luminosidad 1% ( $Lum_1$ ), 5% ( $Lum_5$ ), 10% ( $Lum_{10}$ ), 15% ( $Lum_{15}$ ) y 20% ( $Lum_{20}$ ). Además, se calcularon diferencias entre los percentiles de luminosidad consecutivos para representar el rango de valores de luminosidad en los que oscilan las zonas más oscuras de la imagen. En concreto, se calcularon:



**Figura 3.8:** Distribución promedio de los valores de luminosidad en el canal  $V$ ,  $F(x,y)$ : (a) Función de densidad  $p(F)$ ; (b) Función de distribución acumulada  $P(F)$ .

- $Lum_{5-1} = Lum_5 - Lum_1$
- $Lum_{5-10} = Lum_{10} - Lum_5$
- $Lum_{10-15} = Lum_{15} - Lum_{10}$
- $Lum_{15-20} = Lum_{20} - Lum_{15}$

De este modo, se extrajeron un total de 9 características relacionadas con la luminosidad para cada imagen.

### 3.4.5 Conjunto de características extraídas

Como se ha comentado en los apartados previos, en la fase de extracción de características se aplicaron diversos métodos para extraer descriptores de la calidad de las retinografías. Tras esta fase, cada imagen queda descrita por un conjunto de 40 características:

- Característica 1: índice de calidad  $D\_NIQE$ .
- Características 2-13:  $SpacEn_{MED}$ ,  $SpacEn_{ASIM}$ ,  $SpecEn_{MED}$  y  $SpecEn_{ASIM}$  obtenidas mediante SSEQ.
- Características 14-25: Respuesta al filtro de desviación típica de la CWT en 6 escalas distintas ( $CWT\_Med$  (14-19) y  $CWT\_Desv$  (20-25)).
- Características 26-31: Entropía de la CWT en 6 escalas ( $CWT\_Ent$ ).
- Características 32-36: Percentiles de la luminosidad del fondo a partir del canal  $V$  ( $Lum_1$ ,  $Lum_5$ ,  $Lum_{10}$ ,  $Lum_{15}$  y  $Lum_{20}$ ).
- Características 37-40: Rangos de valores de luminosidad del fondo a partir del canal  $V$  ( $Lum_{5-1}$ ,  $Lum_{10-5}$ ,  $Lum_{15-10}$  y  $Lum_{20-15}$ ).

## 3.5 SELECCIÓN DE CARACTERÍSTICAS

Tras la fase de extracción de características, se incluyó una fase de selección de características con el objetivo de reducir el conjunto de características obtenido para lograr una representación más compacta y explicativa de los datos extraídos. Mediante técnicas de selección de características se logran tres objetivos (Guyon and Elisseeff, 2003):

- 1) Mejorar la capacidad de predicción de los clasificadores.
- 2) Obtener clasificadores más ligeros y rápidos.
- 3) Mejorar la comprensión sobre los procesos de extracción de características.

Las variables definidas en la fase de extracción de características para describir la calidad de las imágenes pueden ser relevantes, es decir, pueden ser capaces de predecir la calidad de las imágenes satisfactoriamente. Sin embargo, recoger una gran cantidad de características no asegura obtener buenos resultados en la clasificación. Es posible que las características recogidas sean redundantes, por lo que no aportan más información cuando

se combinan (Guyon and Elisseeff, 2003). También es posible que algunas características no sean relevantes para resolver el problema en cuestión, por lo que afectan negativamente a la capacidad de predicción (Guyon and Elisseeff, 2003). Finalmente, un gran número de características puede provocar el sobreajuste del clasificador, lo que disminuye la capacidad de predicción (Guyon and Elisseeff, 2003; Saeys *et al.*, 2007). Las técnicas de selección de características proporcionan un subconjunto óptimo para caracterizar un determinado modelo (Saeys *et al.*, 2007). Es importante destacar que las técnicas de selección de características no aplican ninguna transformación sobre los datos, sino que solo escogen un subconjunto reducido del conjunto de características (Saeys *et al.*, 2007).

Los métodos de selección de características se pueden clasificar en tres categorías (Saeys *et al.*, 2007): métodos de filtro, métodos *wrapper* y métodos embebidos.

- Los métodos de filtro buscan las características más relevantes, excluyendo las irrelevantes. Son métodos independientes del clasificador, por lo que se pueden combinar con cualquier clasificador. Calculan una puntuación de relevancia de cada una de las características individualmente y las que no superan un umbral son eliminadas. Como medidas de relevancia se pueden emplear test de hipótesis, tanto paramétricos como no paramétricos. Además, los métodos filtro multivariados analizan las dependencias entre variables, por lo que pueden eliminar características redundantes (Saeys *et al.*, 2007).
- Los métodos *wrapper* actúan junto con el algoritmo de clasificación. Con estos métodos, se proponen diferentes subconjuntos de características al clasificador y se evalúa la capacidad de predicción del mismo con cada subconjunto. Al estar íntimamente ligados al clasificador elegido, los métodos *wrapper* pueden no ser adecuados cuando se analizan clasificadores distintos (Saeys *et al.*, 2007).
- Los métodos embebidos también se utilizan junto con un algoritmo de clasificación, aunque en este caso la selección de características se realiza conjuntamente con la construcción del modelo de clasificación. Al igual que los métodos *wrapper*, son dependientes del clasificador. Sin embargo, tienen la ventaja de tener menor complejidad computacional (Saeys *et al.*, 2007).

Una manera sencilla de detectar pares de características redundantes es calcular el coeficiente de correlación lineal,  $r$ , entre ellas (Guyon and Elisseeff, 2003; Yu and Liu, 2004):

$$r_{X_m, X_n} = \frac{cov(X_m, X_n)}{\sqrt{var(X_m) \cdot var(X_n)}} = \frac{\sum_k (x_{m,k} - \bar{x}_m)(x_{n,k} - \bar{x}_n)}{\sqrt{\sum_k (x_{m,k} - \bar{x}_m)^2 \sum_k (x_{n,k} - \bar{x}_n)^2}} \quad (3.31)$$

Los valores de  $r$  varían entre -1 y 1. Los valores extremos indican correlación elevada, mientras que los valores más cercanos a 0 indican ausencia de relación lineal entre variables

(Guyon and Elisseeff, 2003). En el caso de que dos características estén muy correladas ( $r \approx \pm 1$ ), esto puede indicar que una de ellas es redundante y que sería conveniente eliminarla (Guyon and Elisseeff, 2003). En este TFM, se buscaron redundancias entre las características extraídas calculando el coeficiente  $r$  entre pares de características. Después, se aplicó la técnica de selección de características *Fast Correlation-Based Filter* (FCBF) para obtener un subconjunto de características relevantes minimizando la redundancia entre ellas (Yu and Liu, 2004).

La técnica de selección de características FCBF es un método de filtro de características (Yu and Liu, 2004). El enfoque empleado en esta técnica se basa en seleccionar, en primer lugar, un subconjunto con las características más relevantes. Después, se eliminan de este subconjunto aquellas que son redundantes con respecto a alguna otra característica con mayor relevancia. El subconjunto final contiene únicamente características relevantes e independientes entre sí (Yu and Liu, 2004).

Para determinar la relevancia y la redundancia de las características, FCBF utiliza medidas de correlación. La relevancia se define como la correlación entre una característica  $X_i$  y la variable a predecir a partir de ella  $C$  (Yu and Liu, 2004). En este trabajo,  $C$  es la anotación sobre la calidad de la imagen. La redundancia entre dos características se define en el algoritmo FCBF como la correlación entre una característica  $X_m$  y otra característica  $X_n$  (Yu and Liu, 2004). Para calcular esta correlación se podría utilizar el coeficiente de correlación lineal,  $r_{X_m, X_n}$ , definido según (3.31). Sin embargo,  $r_{X_m, X_n}$  tiene la limitación de que revela dependencia lineal entre características, pero no es capaz de revelar relaciones no lineales. Por ello, FCBF utiliza una medida de correlación no lineal basada en la entropía de las características (Yu and Liu, 2004):

$$H(X_m) = - \sum_{x_m \in X_m} p(x_m) \cdot \log_2 p(x_m) \quad (3.32)$$

Teniendo en cuenta esta definición, la entropía de una característica después de observar otra característica se define como (Yu and Liu, 2004):

$$H(X_i|X_j) = - \sum_{x_j \in X_j} p(x_j) \sum_{x_i \in X_i} p(x_i|x_j) \cdot \log_2 p(x_i|x_j) \quad (3.33)$$

A partir de estas dos medidas, se define la ganancia de información (*Information Gain*, IG) como (Yu and Liu, 2004):

$$IG(X_i|X_j) = H(X_i) - H(X_i|X_j) \quad (3.34)$$

La magnitud de  $IG$  está sesgada hacia características con mayor entropía. Por ello es conveniente calcular una nueva medida, denominada incertidumbre simétrica (*Symmetrical Uncertainty*,  $SU$ ), que se obtiene normalizando  $IG$  (Yu and Liu, 2004):

$$SU(X_i|X_j) = 2 \frac{IG(X_i|X_j)}{H(X_i) + H(X_j)} \quad (3.35)$$

FCBF utiliza  $SU$  como medida de correlación. Un valor de  $SU=1$  indica que es totalmente posible predecir el valor de una característica conociendo el valor de otra, mientras que  $SU=0$  indica independencia total entre características (Yu and Liu, 2004).

En FCBF se utilizan dos tipos de correlaciones (Yu and Liu, 2004):

- Correlación entre la clase (variable a predecir) y una característica,  $SU(X_m|C)$ .
- Correlación entre pares de características,  $SU(X_m|X_n)$

En la primera fase del algoritmo FCBF se calcula  $SU(X_m|C)$  para cada una de las características para obtener la relevancia de las mismas. Se aplica después un umbral para descartar aquellas que no superen un determinado valor de  $SU(X_m|C)$ , que se consideran características irrelevantes. En este TFM se eligió un valor de  $SU$  de 0.05 de manera empírica. Las características que superen el umbral pasan a la segunda fase. En esta segunda fase se descartan las características redundantes. Una característica  $X_m$  se considera redundante respecto a  $X_n$  si se cumplen dos condiciones (Yu and Liu, 2004):

$$SU(X_n|C) \geq SU(X_m|C) \quad (3.36)$$

$$SU(X_m|X_n) \geq SU(X_m|C) \quad (3.37)$$

En resumen, en el algoritmo FCBF, la característica  $X_m$  se elimina si tiene menor relevancia que la característica  $X_n$  y, además, la correlación entre  $X_m$  y  $X_n$  es mayor que la relevancia de  $X_m$  (Yu and Liu, 2004). Por ello, para implementar este algoritmo de manera eficiente, tras la primera fase se ordenan las características en orden decreciente de

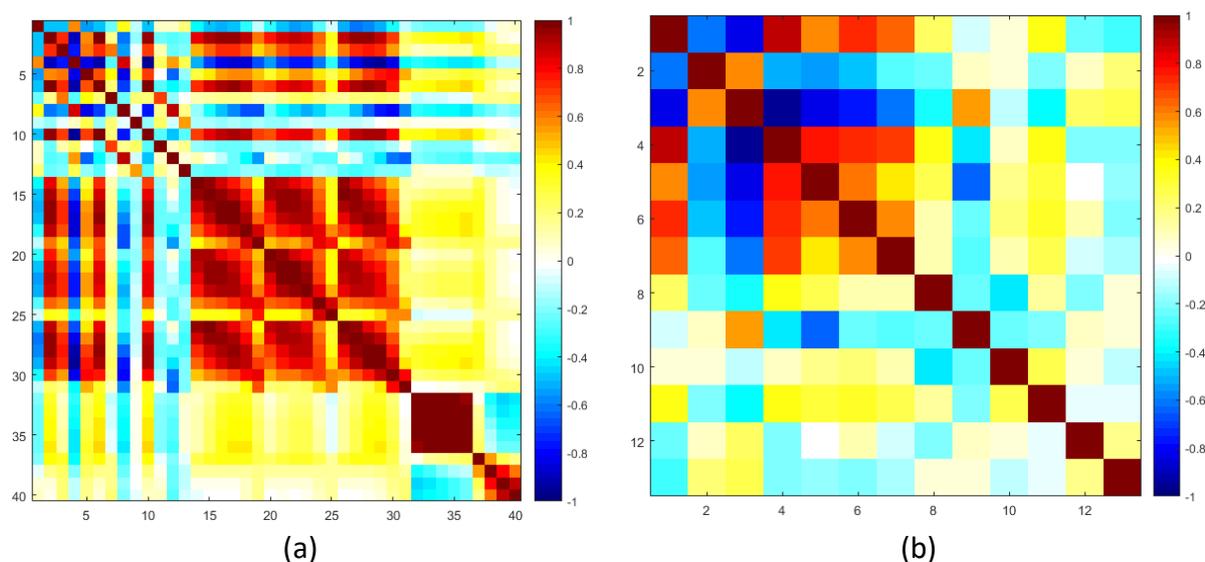
relevancia. En la segunda fase se recorren las características en este orden, y se eliminan las características redundantes, por lo que no es necesario evaluar la primera condición.

En este TFM se aplicó el algoritmo FCBF sobre el conjunto de 40 características extraídas previamente. En la Figura 3.9 (a) se muestra gráficamente la matriz de coeficientes de correlación  $r_{x_m, x_n}$  calculados según (3.31). Se observa que existen muchas características que presentan una correlación elevada entre sí (colores rojos y azules oscuros), por lo que son redundantes. Por ello, tiene sentido aplicar un método de selección de características con el objetivo de reducir esta redundancia y facilitar la etapa de clasificación posterior. Aplicando el algoritmo FCBF sobre las 40 características extraídas para las imágenes del conjunto de entrenamiento, se obtuvo un conjunto reducido de 13 características relevantes y no redundantes entre sí. En la Figura 3.9 (b) se muestra gráficamente la matriz de coeficientes de correlación tras aplicar FCBF. Se puede observar cómo, en este caso, el número de características redundantes se ha reducido notablemente.

El subconjunto de 13 características seleccionadas se utilizó como entrada al algoritmo de clasificación de características con el que se obtendrá la calidad de las imágenes.

### 3.6 CLASIFICACIÓN DE CARACTERÍSTICAS

Tras las fases de extracción y selección de características, en la fase de clasificación se trata de determinar si una retinografía tiene calidad suficiente para ser analizada en base a las características extraídas de la misma. Para esta tarea de reconocimiento de patrones se empleó un clasificador supervisado (Jain *et al.*, 2000), puesto que se dispone de las anotaciones sobre la calidad de todas las imágenes de la BD empleada en este TFM, tal y como se describe en la sección 3.2.



**Figura 3.9:** Selección de características mediante FCBF: (a) Matriz de coeficientes de correlación de las características extraídas; (b) Matriz de coeficientes de correlación de las características seleccionadas.

El algoritmo de clasificación empleado en este TFM fue una Red Neuronal Artificial (RNA) de tipo perceptrón multicapa (*Multilayer Perceptron*, MLP). Las RNA se pueden entender como sistemas de procesamiento masivamente paralelos que constan de un elevado número de unidades de procesamiento o neuronas interconectadas (Jain *et al.*, 2000). Tienen la capacidad de aprender relaciones entrada-salida complejas y no lineales, adaptándose a las características de los datos de entrada. Las RNA más utilizadas en reconocimiento de patrones son las redes *feed-forward*, entre las que se encuentran las redes MLP y las redes de función de base radial (*Radial Basis Function*, RBF) (Jain *et al.*, 2000). La red MLP es una RNA organizada en capas donde las conexiones entre capas son unidireccionales hacia delante, sin bucles de realimentación. Cada una de las neuronas recibe como entrada una combinación lineal de las salidas de la capa anterior y su salida es una función no lineal de la entrada (Bishop, 1995). Este tipo de red se ha utilizado ampliamente como algoritmo de clasificación (Jain *et al.*, 2000), incluido en el procesado de retinografías (García *et al.*, 2009; Marín *et al.*, 2011; Pires Dias *et al.*, 2014).

### 3.6.1 Balanceo de clases mediante sobremuestreo

En muchos problemas de reconocimiento de patrones existe un problema de desbalanceo en los conjuntos de datos empleados: se dispone de muchos más ejemplos de una clase que de otra. No obstante, los clasificadores habitualmente asumen que existe un número similar de instancias de cada una de las clases (Krawczyk, 2016). Para afrontar este problema se han diseñado métodos a nivel de dato y a nivel de algoritmo que tratan de corregir el sesgo que puede tener el clasificador hacia la clase mayoritaria (Krawczyk, 2016). Los primeros modifican el conjunto de datos de entrenamiento, haciéndolo más apto para utilizar un clasificador estándar. Dentro de ellos se pueden encontrar métodos de sobremuestreo de la clase minoritaria y de submuestreo de la clase mayoritaria. Por otro lado, los métodos a nivel de algoritmo tratan de modificar el clasificador penalizando los errores cometidos sobre instancias de la clase minoritaria (Krawczyk, 2016). Los más extendidos se basan en modificar el clasificador asignando un coste elevado al clasificar muestras de la clase minoritaria erróneamente (Krawczyk, 2016).

Como se describió en el apartado 3.2, el conjunto de entrenamiento de la BD empleada en este TFM consta de 905 retinografías de buena calidad y 148 retinografías de calidad insuficiente. Por lo tanto, el conjunto de entrenamiento se encuentra muy desbalanceado. Para solventar este problema y mejorar la etapa de clasificación, en este TFM se empleó el método de sobremuestreo *Synthetic Minority Oversampling Technique* (SMOTE) (Chawla *et al.*, 2002) para equilibrar el número de muestras de imágenes de mala calidad con el número de muestras de las imágenes de buena calidad. Este algoritmo crea nuevas instancias (vectores de características) a partir de las instancias ya existentes a partir de combinaciones lineales en el espacio de características (Chawla *et al.*, 2002).

El algoritmo SMOTE consta de varias etapas (Chawla *et al.*, 2002). En primer lugar, se seleccionan únicamente las instancias de la clase minoritaria. Los datos de la clase mayoritaria

permanecen inalterados. A continuación, para cada instancia de la clase minoritaria, se han de encontrar las  $k$  instancias de la clase minoritaria más cercanas en el espacio de características utilizando la distancia euclídea entre 2 muestras (Chawla *et al.*, 2002):

$$d_{\mathbf{u},\mathbf{v}} = \sqrt{\sum_{i=1}^{13} (u_i - v_i)^2} \quad (3.38)$$

Posteriormente, por cada una de las  $k$  instancias vecinas, se ha de generar una instancia sintética situada en la línea que une la instancia con su vecina. Para ello, se obtiene la diferencia entre ambas muestras y, a continuación, se obtiene la muestra sintética  $\mathbf{w}$  como la suma de la muestra real  $\mathbf{u}$  y la diferencia entre  $\mathbf{u}$  y  $\mathbf{v}$  multiplicada por un número aleatorio  $c$  entre 0 y 1 (Chawla *et al.*, 2002):

$$\mathbf{w} = \mathbf{u} + c \cdot (\mathbf{v} - \mathbf{u}) \quad (3.39)$$

Así, se obtienen  $k$  muestras sintéticas por cada muestra real de la clase minoritaria. Para el correcto funcionamiento del algoritmo SMOTE, es necesario fijar el parámetro  $k$ . En este TFM se eligió  $k = 5$  para obtener 740 muestras sintéticas de la clase “mala calidad”. De este modo, el número de instancias de la clase minoritaria ascendió a 888 frente a las 905 instancias de la clase “buena calidad”.

### 3.6.2 Fase de entrenamiento de la red MLP

Tras obtener las muestras sintéticas con SMOTE, se utilizaron las 13 características seleccionadas para las 1793 muestras de entrenamiento (905 de la clase “buena calidad” y 888 de la clase “mala calidad”) para entrenar la red MLP.

En este TFM se ha implementado una red MLP de tres capas: una capa de entrada, una capa oculta y una capa de salida. Las redes MLP de tres capas son capaces de proporcionar una frontera de decisión arbitraria y suave si las neuronas de la red tienen función de activación sigmoideal. El único requisito es que la capa oculta tenga un número suficiente de neuronas (Bishop, 1995).

Puesto que el número de características seleccionadas asciende a 13, la red MLP creada tenía 13 neuronas en la capa de entrada. Por otra parte, como se pretende distinguir únicamente entre dos clases (“buena calidad” y “mala calidad”), es suficiente con tener una neurona en la capa de salida (Bishop, 1995). El número de neuronas de la capa oculta,  $n_{oc}$ , se configuró de forma experimental. En concreto, se realizaron pruebas variando  $n_{oc}$  entre 1 y 100. La función de activación de las neuronas de la capa oculta elegida fue “tangente de

sigmoide”, mientras que para la capa de salida la función elegida fue “logaritmo de sigmoide”. Estas funciones tienen ventajas para el entrenamiento de la red, y es necesario emplearlas para crear una frontera de decisión arbitraria que separe ambas clases (Bishop, 1995). Las funciones de activación sigmoideas tienen la ventaja de ser continuamente diferenciables. Además, la función “tangente de sigmoide” es capaz de acelerar el proceso de entrenamiento de la red MLP si se elige como función de activación en la capa oculta (Bishop, 1995). No obstante, en la capa de salida se eligió la función de activación “logaritmo de sigmoide” porque proporciona una salida en el rango de valores de 0 a 1, que se asemeja a una probabilidad a posteriori (Bishop, 1995). Estas funciones de activación no son las únicas que pueden aprovechar estas ventajas, aunque son las más comunes (Bishop, 1995).

El entrenamiento de la red MLP consiste en minimizar el error de clasificación producido. Al utilizar una red MLP para clasificación, las salidas habitualmente se modelan como probabilidades a posteriori (Bishop, 1995). La función de error empleada fue *Cross Entropy* o Entropía Cruzada, que resulta adecuada en problemas de clasificación binaria. Para cada salida  $y_n$  de la red MLP, el error con respecto de la salida deseada  $t_n$  es (Bishop, 1995):

$$E = -\frac{1}{N} \sum_n [t_n \ln y_n + (1 - t_n) \ln(1 - y_n)] \quad (3.40)$$

En este caso, la salida deseada corresponde con la anotación de calidad de la imagen:  $t_n = 1$  si la imagen es de buena calidad y  $t_n = 0$  si la imagen es de mala calidad. Si todas las instancias están perfectamente clasificadas por la red MLP ( $y_n = t_n$  para todo  $n$ ), el error es 0. El algoritmo de optimización del error empleado fue el algoritmo *Scaled Conjugate Gradient*. Es un algoritmo de descenso por gradiente rápido y eficiente comparado con otros algoritmos de descenso por gradiente (Bishop, 1995).

Adicionalmente, es necesario tener en cuenta que las redes MLP son algoritmos de clasificación que tienden al sobreajuste, es decir, en la fase de entrenamiento tienden a memorizar los datos de entrenamiento (Bishop, 1995; Jain *et al.*, 2000). Como consecuencia, las fronteras de decisión se ajustan a los datos de entrenamiento, y no resultan útiles para clasificar los datos de test. Para evitar este sobreajuste, en la fase de entrenamiento se emplearon técnicas de parada temprana o *early stopping* y regularización (Bishop, 1995).

La técnica *early stopping* trata de detener la fase de entrenamiento tras un determinado número de iteraciones del algoritmo. Ciertamente, a medida que se suceden las iteraciones del algoritmo de entrenamiento, se va convergiendo hacia el mínimo de la función de error (Bishop, 1995). Sin embargo, si se evalúa el error cometido por la red a medida que avanzan las iteraciones en un conjunto de validación independiente, suele ocurrir que dicho error decrece hasta un mínimo y luego comienza a aumentar. Esto indica que la red MLP comienza a sobreajustarse a los datos de entrenamiento y deja de ser generalizable (Bishop, 1995). En la Figura 3.10 se ilustra este comportamiento. La técnica de parada temprana

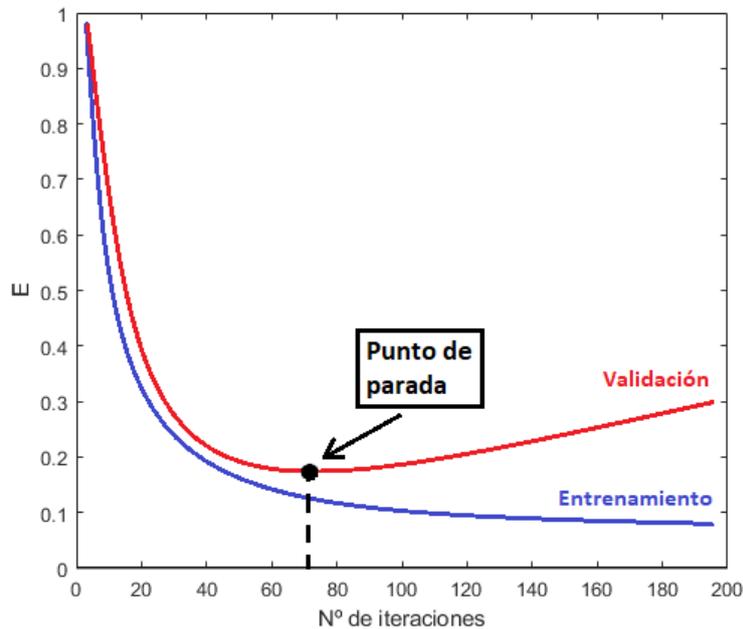


Figura 3.10: Criterio de parada temprana (*early stopping*). Adaptado de Bishop (1995).

detiene el entrenamiento cuando se ha alcanzado el mínimo error en el conjunto de validación (Bishop, 1995).

Por su parte, la regularización trata de evitar la construcción de fronteras de decisión complejas que se ajustan en exceso a los datos de entrenamiento. Las fronteras de decisión complejas se asocian con pesos elevados en la combinación lineal de las entradas de cada neurona. Si se introduce un parámetro de regularización,  $\nu$ , que penalice el uso de pesos elevados a la entrada de las neuronas, se consiguen fronteras de decisión más suaves. No obstante, hay que tener en cuenta que si el parámetro  $\nu$  es demasiado elevado, la frontera de decisión es demasiado simple como para separar los datos de manera eficaz (Bishop, 1995). Según la implementación de la red MLP empleada utilizando el *Neural Network Toolbox* de MATLAB®, el parámetro  $\nu$  se ajusta en el rango (0,1). El error se modifica con respecto a la ecuación (3.40) (Nielsen, 2015):

$$E' = (1 - \nu)E + \nu\Omega \quad (3.41)$$

$$\Omega = \frac{1}{2} \sum_i w_i^2 \quad (3.42)$$

De esta manera, los pesos de la red MLP  $w_i$  influyen en la función de error, dificultando la construcción de fronteras de decisión complejas (Bishop, 1995). El valor del parámetro de regularización  $\nu$  se determinó de manera experimental. En la implementación de la red MLP mediante el *Neural Network Toolbox* de MATLAB®,  $\nu$  puede tomar valores entre 0 y 1. Empleando  $\nu=0$  se anularía el término de penalización  $\Omega$ , mientras que con  $\nu=1$  se anularía el error cometido. Por lo tanto,  $\nu$  debe fijarse en un valor intermedio. Se realizaron pruebas con valores de  $\nu$  entre 0.1 y 0.9 en incrementos de 0.1.

En este TFM se tuvieron en cuenta ambos criterios durante la fase de entrenamiento para evitar el sobreajuste y lograr un clasificador MLP generalizable. En el conjunto de entrenamiento se definió un subconjunto de validación escogiendo muestras aleatorias. Sin embargo, debido a que anteriormente se incluyeron muestras sintéticas mediante SMOTE, no se puede asegurar que dicho subconjunto sea independiente del resto de datos de entrenamiento. Por lo tanto, cabe la posibilidad de que la red MLP se sobreajuste a los datos de entrenamiento incluso empleando *early stopping*.

Puesto que se eligió una arquitectura de la red MLP con una única neurona en la última capa, a la salida de la red MLP se obtiene un único valor entre 0 y 1. Este valor se puede interpretar como una probabilidad a posteriori (Pires Dias *et al.*, 2014). Si la salida supera un determinado umbral, se puede considerar que la imagen es de calidad suficiente para ser analizada. Por el contrario, si la salida de la red MLP está por debajo del umbral, la imagen se consideraría de calidad insuficiente. En este TFM, este umbral se fijó de forma experimental con el objetivo de minimizar la tasa de fallos de la red MLP, como se describe en el Capítulo 4.

### 3.6.3 Fase de test de la red MLP

Tras ajustar todos los parámetros explicados anteriormente, se obtuvo la red MLP entrenada. Esta red se empleó para evaluar la calidad de las imágenes del conjunto de test. Este conjunto tenía 1054 imágenes, de las cuales 905 eran de buena calidad y 149 no tenían calidad suficiente. Como se indicó en el apartado 3.2, estas imágenes son independientes de las utilizadas en el conjunto de entrenamiento. Los resultados obtenidos en este estudio sobre el conjunto de test se exponen en el Capítulo 4.

## Capítulo 4. Resultados

---

### 4.1 INTRODUCCIÓN

En este capítulo se presentan los resultados obtenidos en la evaluación de la calidad en retinografías empleando el método desarrollado en este TFM. Tras entrenar la red MLP con los ejemplos procedentes del conjunto de entrenamiento, se evaluaron los resultados obtenidos sobre el conjunto de test. Para evaluar los resultados, se compararon las salidas producidas por la red MLP con las anotaciones realizadas en la BD empleada. Así, los resultados se evaluaron de manera cuantitativa estimando la tasa de fallos de la red junto a otras medidas que determinan la capacidad de clasificación del método propuesto.

### 4.2 CARACTERÍSTICAS SELECCIONADAS POR EL ALGORITMO FCBF

En la fase de selección de características se redujo el número de características, obteniendo una representación compacta de los datos extraídos evitando características redundantes. Como se indicó en el apartado 3.5, se aplicó el algoritmo FCBF con el objetivo principal de reducir la redundancia de las características. Como se indicó en el apartado 3.5.1, se empleó  $SU$  como medida de correlación para cuantificar la relevancia (correlación entre una característica y la salida deseada) y la redundancia de las características obtenidas (correlación entre pares de características). Los resultados obtenidos se muestran en la Tabla 4.1. Las características seleccionadas por FCBF fueron 13 de las 40 que se extrajeron. Para lograr este resultado, las características se ordenaron en orden decreciente de  $SU$ , se eliminaron las que no superaban el umbral  $SU=0.05$  y finalmente se eliminaron las características redundantes según las ecuaciones (3.36) y (3.37).

Característica	$SU$	Característica	$SU$
$CWT\_Ent$ ( $s=4$ )	0.4182	$D\_NIQE$	0.3574
$SpecEn_{MED}$ (escala 1)	0.3378	$CWT\_Ent$ ( $s=16$ )	0.3281
$SpecEn_{ASIM}$ (escala 1)	0.2380	$SpacEn_{ASIM}$ (escala 1)	0.2325
$CWT\_Desv$ ( $s = 32$ )	0.1896	$Lum_5$	0.1659
$SpecEn_{MED}$ (escala 3)	0.1235	$Lum_{15-10}$	0.1046
$Lum_{5-1}$	0.0850	$SpacEn_{ASIM}$ (escala 3)	0.0752
$SpecEn_{ASIM}$ (escala 3)	0.0686		

Tabla 4.1: Características seleccionadas por el algoritmo FCBF.

### 4.3 EVALUACIÓN DE LOS RESULTADOS

Para obtener los resultados finales del estudio se compararon las salidas obtenidas empleando el método automático desarrollado con las anotaciones que los expertos realizaron sobre la BD. De esta comparación se puede obtener si el resultado del algoritmo automático para una determinada imagen produce un verdadero positivo (VP), falso positivo (FP), verdadero negativo (VN) o falso negativo (FN). Los FP y los FN cuantifican el número de fallos del clasificador, mientras que los VP y los VN son un recuento de los aciertos. La matriz de confusión (MC) es una matriz que contiene los VP, FP, FN y VN. Todas estas métricas se recogen en la Tabla 4.1 (Fawcett, 2006).

Después de obtener la MC, se calcularon la sensibilidad ( $Se$ ), la especificidad ( $Sp$ ), la precisión ( $Acc$ ), y el valor predictivos positivo ( $VPP$ ) (Fawcett, 2006):

- **Se**: es la fracción de imágenes de buena calidad clasificadas correctamente.

$$Se = \frac{VP}{VP + FN} \quad (4.1)$$

- **Sp**: es la fracción de imágenes de mala calidad clasificadas correctamente.

$$Sp = \frac{VN}{VN + FP} \quad (4.2)$$

- **Acc**: es la fracción de imágenes de buena o mala calidad clasificadas correctamente.

$$Acc = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.3)$$

		Resultado del algoritmo automático	
		Buena calidad	Mala calidad
Anotaciones de la BD	Buena Calidad	VP	FN
	Mala Calidad	FP	VN

**Tabla 4.2:** Posibles resultados del clasificador al comparar la salida proporcionada por el algoritmo con las anotaciones sobre la calidad de las imágenes de la BD.

- **VPP**: es la probabilidad de que una imagen clasificada en la clase “buena calidad” sea realmente una imagen de buena calidad.

$$VPP = \frac{VP}{VP + FP} \quad (4.4)$$

Otra medida útil para evaluar los resultados del clasificador es la *F-Score* (*F1*). Esta medida se obtiene a partir de *VPP* (denominada *Precision* en este contexto) y *Se* (denominada *Recall* en este contexto). La medida *F1* se define como la media armónica entre *VPP* (*Precision*) y *Se* (*Recall*) (Saito and Rehmsmeier, 2015):

$$F1 = 2 \frac{VPP \cdot Se}{VPP + Se} \quad (4.5)$$

Dependiendo del umbral aplicado a la salida de la red MLP, los valores de estas métricas serán diferentes. Por lo tanto, se debe encontrar el umbral para el que se obtienen los mejores resultados. Para ello, se pueden emplear curvas *Receiver Operating Characteristic* (ROC). Una curva ROC es una curva paramétrica en el plano XY que representa  $1-Sp$  frente a *Se* (Zweig and Campbell, 1993). El punto ( $1-Sp=0, Se=1$ ) representa la clasificación ideal, en la que  $FP = 0$  y  $FN = 0$  a la vez que *Se*, *Sp*, *Acc*, *VPP* y *F1* alcanzan su valor máximo (i.e., 1). Cuanto más próxima esté la curva ROC al punto de clasificación ideal, mejores serán los resultados de clasificación para un determinado problema (Zweig and Campbell, 1993; Park *et al.*, 2004; Fawcett, 2006). Una forma de resumir la información que proporcionan las curvas ROC es utilizar la medida del área bajo la curva ROC (*Area Under ROC Curve*, *AUC*). Esta es el área encerrada entre la curva ROC y el eje de abscisas (Park *et al.*, 2004). El *AUC* es útil para comprobar la robustez de un método de clasificación en diferentes puntos (por ejemplo, los que se corresponden a diferentes valores del umbral de salida) (Park *et al.*, 2004). Cuanto más se aproxime a 1, mejor efectividad presentará el algoritmo de clasificación (Zweig and Campbell, 1993; Park *et al.*, 2004; Fawcett, 2006).

## 4.4 RESULTADOS OBTENIDOS

### 4.4.1 Fase de entrenamiento

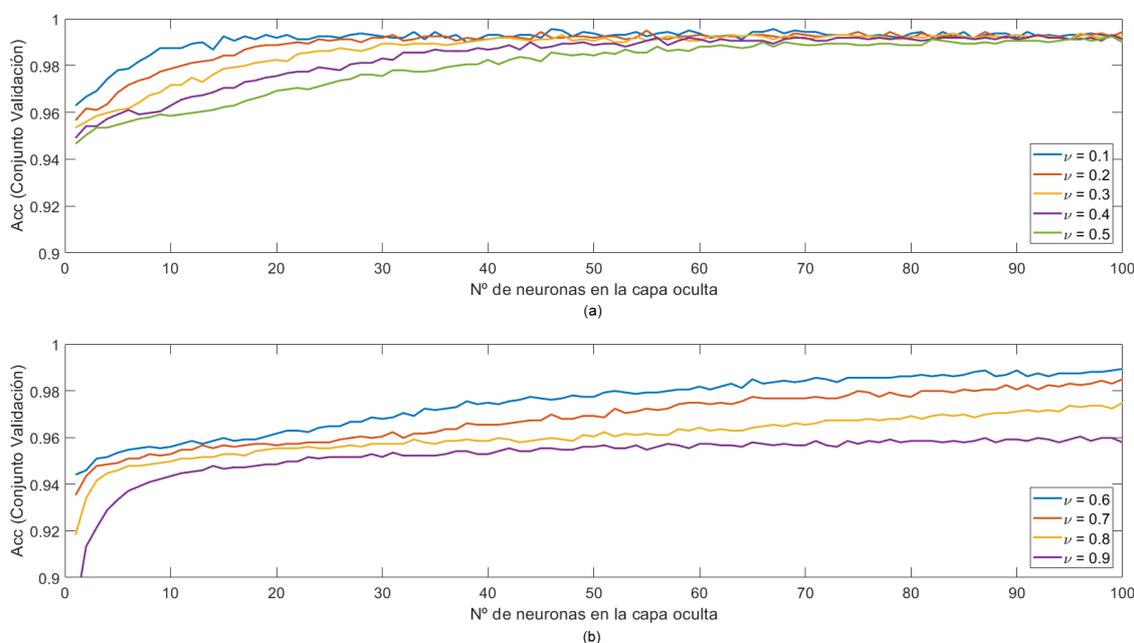
En la fase de entrenamiento fue necesario determinar los valores adecuados de  $n_{oc}$  y  $v$  para la red MLP. Estos parámetros se determinaron de manera empírica, obteniendo *Acc* para cada uno de los valores considerados. Se realizaron pruebas para los valores de  $n_{oc}$  comprendidos entre 1 y 100 neuronas en la capa oculta, y para los valores de  $v$  entre 0.1 y 0.9 en incrementos de 0.1 según se explicó en el apartado 3.6.2.

Para obtener una estimación de la precisión que logra la red MLP, se empleó validación cruzada con 10 iteraciones (*10-fold cross validation*). Esta técnica divide los datos de entrenamiento en 10 particiones de tamaño aproximadamente igual. En cada iteración, se emplearon 9 particiones como conjunto de entrenamiento y una partición como conjunto de validación. El proceso se repite 10 veces, empleando cada vez una partición distinta como conjunto de validación. Finalmente, se calcularon los resultados promedio obtenidos sobre el conjunto de validación en las 10 iteraciones (Bishop, 1995).

En cada una de las 10 iteraciones, para obtener  $Acc$  es necesario aplicar un umbral a la salida de la red MLP. En este TFM se eligió este umbral como el umbral que maximiza  $Acc$ , es decir, se eligió el umbral que minimiza el número de fallos del clasificador:

$$T_{ACC} = \underset{th}{\operatorname{argmax}}[Acc(th)] \quad (4.6)$$

Se aplicó  $T_{ACC}$  sobre la salida de la red MLP para los datos del conjunto de validación, obteniéndose un valor de  $Acc$ . Se repitió 10 veces como se ha descrito, y se obtuvo una estimación de  $Acc$  para cada valor de  $n_{OC}$  y  $\nu$ . En la Figura 4.1 se muestran las variaciones de la  $Acc$  estimada al variar  $n_{OC}$  y  $\nu$ . Se observó que en general la estimación de  $Acc$  decrece a medida que crece  $\nu$ , por lo que los valores de  $\nu = 0.1$  o  $\nu = 0.2$  parecen los más adecuados.



**Figura 4.1:** Valores de  $Acc$  estimados para los diferentes parámetros de regularización  $\nu$ . Cada curva muestra la evolución de  $Acc$  según aumenta  $n_{OC}$ . (a) Parámetros  $\nu$  entre 0.1 y 0.5; (b) Parámetros  $\nu$  entre 0.6 y 0.9.

Finalmente se eligió  $\nu = 0.2$  para evitar el sobreajuste en mayor medida. En cuanto al número de neuronas necesarias en la capa oculta, se observó tras fijar  $\nu = 0.2$  que  $Acc$  no aumenta sustancialmente a partir de  $n_{OC} = 30$  neuronas (Figura 4.1). Por lo tanto, se entrenó la red MLP con  $\nu = 0.2$  y  $n_{OC} = 30$ .

Tras la elección de todos los parámetros de la red MLP, se volvió a entrenar la red con esta configuración utilizando los datos del conjunto de entrenamiento para obtener  $Se$ ,  $Sp$ ,  $Acc$ ,  $VPP$  y  $F1$ , una curva ROC y su  $AUC$ . La curva ROC obtenida se muestra en la Figura 4.2. El  $AUC$  logrado es de 0.9992. Es necesario volver a aplicar un umbral a la salida de la red MLP entrenada para obtener los resultados, ya que estos varían según el umbral aplicado. En la Figura 4.3 se muestran las gráficas de  $Se$ ,  $Sp$ ,  $Acc$ ,  $VPP$  y  $F1$  frente a los umbrales. De nuevo, se calculó el umbral  $T_{ACC}$  según (4.6), obteniéndose  $T_{ACC}=0.4997$ . Tras aplicar  $T_{ACC}$  se obtuvieron los resultados finales sobre el conjunto de entrenamiento, resumidos en la Tabla 4.3.

#### 4.4.2 Fase de test

Una vez elegidos los parámetros de la red MLP ( $n_{OC}$ ,  $\nu$  y  $T_{ACC}$ ), se aplicó el algoritmo completo sobre las imágenes del conjunto de test. Dicho conjunto contenía 1054 imágenes (905 de buena calidad y 149 de mala calidad). Los resultados obtenidos en la fase de test se muestran en la Tabla 4.4.

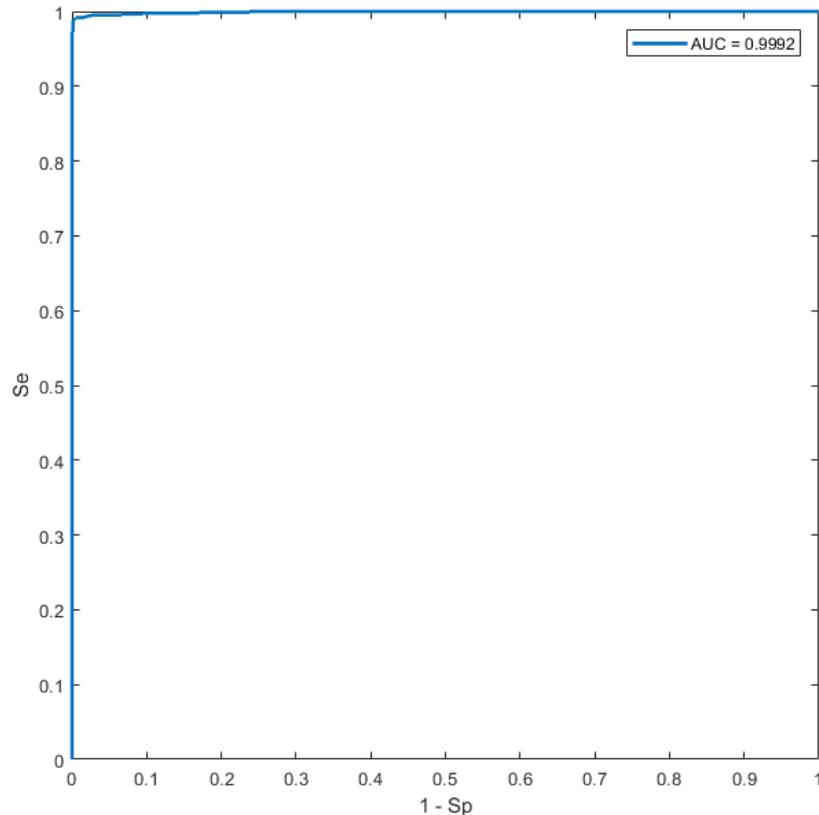


Figura 4.2: Curva ROC obtenida para el conjunto de entrenamiento.

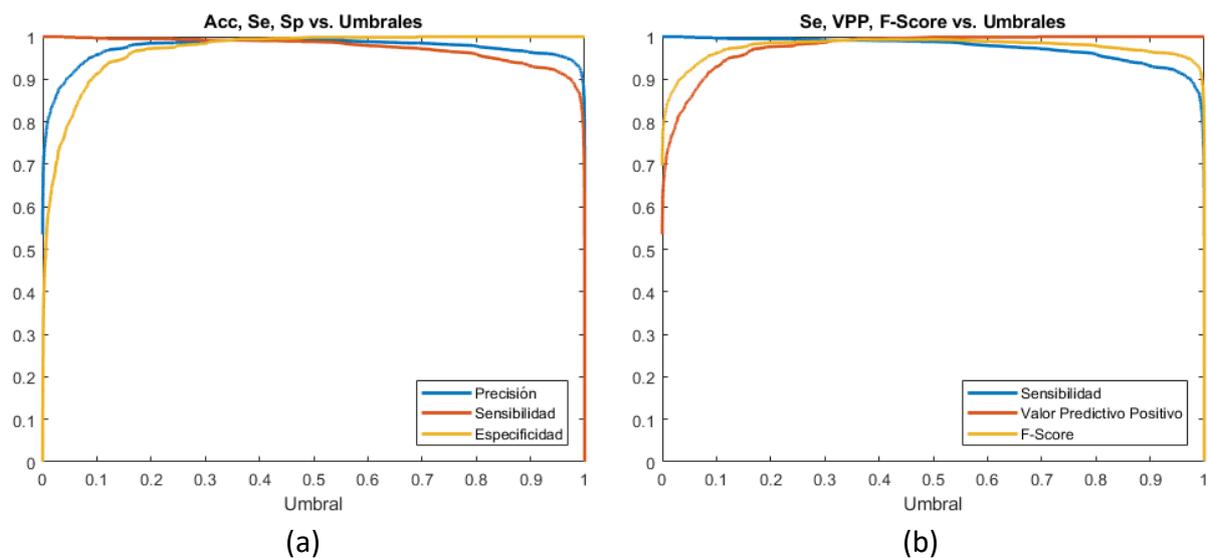


Figura 4.3: (a) Evaluación de  $Acc$ ,  $Se$  y  $Sp$  frente a los umbrales; (b)  $Se$ ,  $VPP$  y  $F$ -Score frente a los umbrales.

$T_{Acc}$	$Se$	$Sp$	$Acc$	$VPP$	$F1$	$AUC$
<b>0.4997</b>	98.82%	99.86%	99.31%	99.88%	0.9935	0.9992

Tabla 4.4: Resultados de la red MLP en el conjunto de entrenamiento.

$T_{Acc}$	$Se$	$Sp$	$Acc$	$VPP$	$F1$
<b>0.4997</b>	93.26%	85.23%	92.13%	97.46%	0.9531

Tabla 4.3: Resultados de la red MLP en el conjunto de test.

## Capítulo 5. Discusión

---

### 5.1 INTRODUCCIÓN

A lo largo de los capítulos anteriores se han descrito el método desarrollado en este TFM para la evaluación de la calidad en retinografías y los resultados obtenidos con el mismo sobre una BD de retinografías proporcionada por el IOBA y el HCUV. Las retinografías de esta BD provenían tanto de sujetos sanos como de pacientes con diversos tipos de patología ocular. En este capítulo se analizan los resultados de una manera más crítica, mostrando las ventajas e inconvenientes del método desarrollado. Además, se comentan los aspectos positivos y las limitaciones de cada una de las fases de procesado. Finalmente, se comparan el método desarrollado y los los resultados obtenidos con otros métodos publicados previamente.

### 5.2 PREPROCESADO

En este método se incluyó una fase de preprocesado con el objetivo de reducir fallos en etapas posteriores del algoritmo. Se empleó para ello un algoritmo de detección de la región circular del FOV, junto con un algoritmo de extensión del mismo que eliminaba los efectos de borde. Este algoritmo se ha aplicado previamente en métodos que utilizan la CWT en retinografías (Soares *et al.*, 2006; Fathi and Naghsh-Nilchi, 2013). Sin embargo, como consecuencia de incluir este algoritmo como parte del preprocesado, en este TFM se introdujeron algunas modificaciones sobre el algoritmo original. En estudios previos, el algoritmo de extensión del FOV se aplicó sobre el canal verde de las retinografías (Soares *et al.*, 2006; Fathi and Naghsh-Nilchi, 2013). En este TFM, este algoritmo se aplicó sobre la imagen en color utilizando el espacio de color RGB. Ciertamente, algunos de los algoritmos aplicados en las fases posteriores del método desarrollado en este TFM requieren convertir las imágenes a espacios de color como YCbCr y HSV. Para ello, es imprescindible que las imágenes de salida de la etapa de preprocesado sean imágenes en color. Por otra parte, el algoritmo de extensión de la FOV se empleó para eliminar el borde negro de la imagen en su totalidad, no sólo parcialmente como en estudios previos (Soares *et al.*, 2006; Fathi and Naghsh-Nilchi, 2013). Los métodos NIQE y SSEQ tienen en cuenta todos los píxeles de la imagen para la evaluación de su calidad. Por ello, es necesario eliminar completamente el borde negro de las retinografías para evitar que esos píxeles que no forman parte del fondo retiniano influyan en el cálculo de los parámetros de calidad.

Los resultados obtenidos con esta técnica de preprocesado son satisfactorios, ya que evitaron que se detectasen bordes fuertes en los límites del FOV con los algoritmos desarrollados para la extracción de características. Por una parte, el método basado en la CWT no produjo falsas detecciones de bordes. Asimismo, el método de análisis de la

luminosidad no incluyó los bordes del FOV como zonas oscuras (ver las Figuras 3.4 y 3.7). La principal desventaja del algoritmo de extensión del FOV es que requiere un elevado tiempo de cálculo en comparación con otras etapas del método desarrollado. Al tratarse de un método iterativo, el número de iteraciones necesario para una cierta imagen influye directamente en el tiempo de procesado. Además, el algoritmo se aplicó sobre los tres canales de color, con lo que este problema ocurre en cada uno de estos canales. El algoritmo de extensión del FOV implementado en este TFM emplea aproximadamente 1 minuto por imagen y es la parte del preprocesado que más influye en el tiempo total de cálculo para el método desarrollado.

### 5.3 EXTRACCIÓN DE CARACTERÍSTICAS

En la fase de extracción de características se emplearon técnicas de análisis de la calidad global de las imágenes. Se emplearon dos métodos de propósito general con el objetivo de probar su efectividad en retinografías. El método NIQE, por sí solo, proporcionó buenos resultados al diferenciar entre retinografías de buena y mala calidad. Esto se puede visualizar en la Figura 5.1, donde se representan mediante *boxplots* los valores del índice NIQE ( $D_{NIQE}$ ) proporcionados por este método. Se puede observar que para las imágenes de buena calidad los valores de  $D_{NIQE}$  son considerablemente menores que los valores obtenidos para imágenes de mala calidad. Esto muestra que esta medida puede ser adecuada para diferenciar las imágenes de buena y mala calidad. Es necesario tener en cuenta que  $D_{NIQE}$  es una medida de distancia a un modelo ideal de imágenes, por lo que obtener correctamente dicho modelo es imprescindible. En este TFM, se obtuvo el modelo NIQE sobre las imágenes de buena calidad del conjunto de entrenamiento. La obtención del modelo MVG requiere un gran tiempo de procesado, alrededor de unas 6 horas. Sin embargo, una vez obtenido este modelo, el cálculo de  $D_{NIQE}$  para una imagen se realizó en menos de un segundo. Por ello, este paso del método desarrollado no influye sustancialmente en su carga computacional.

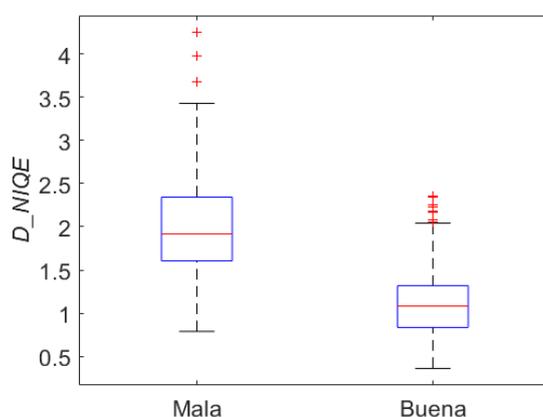
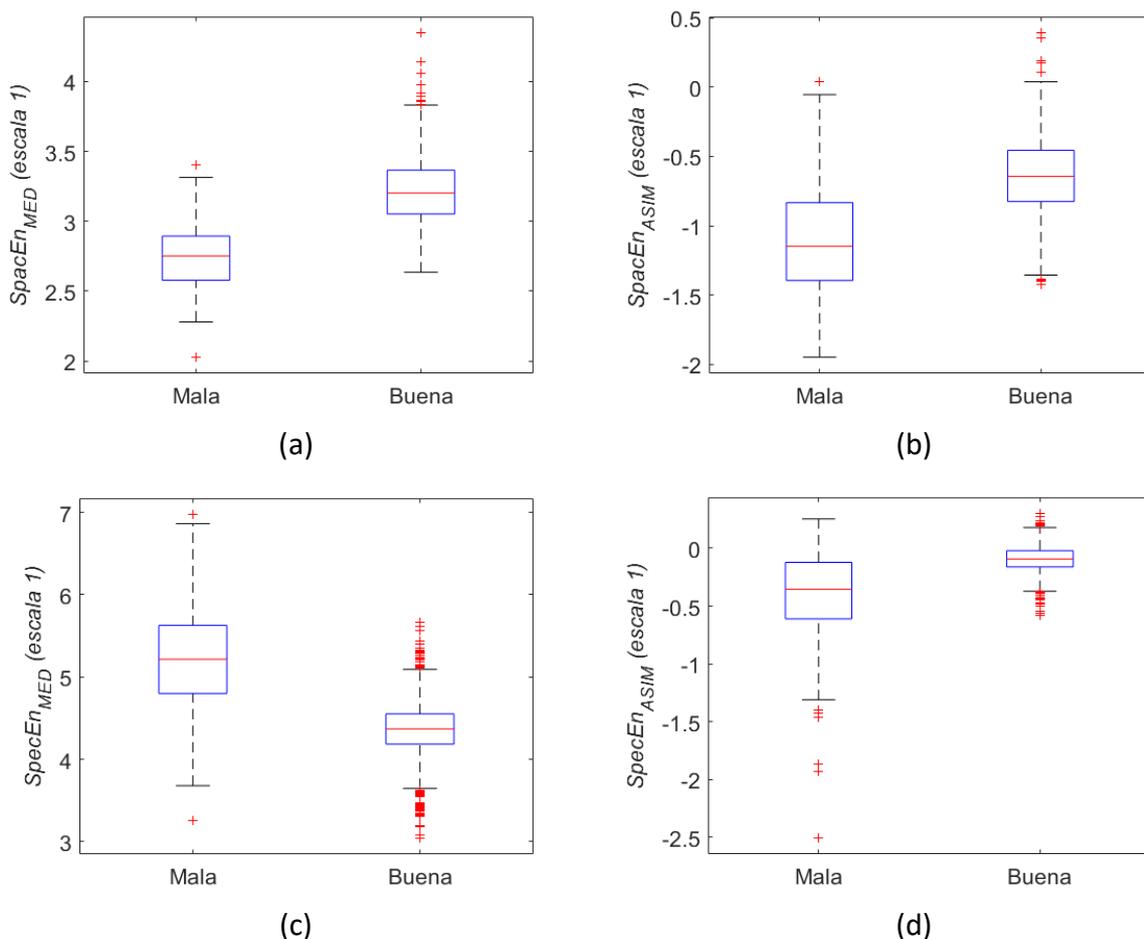
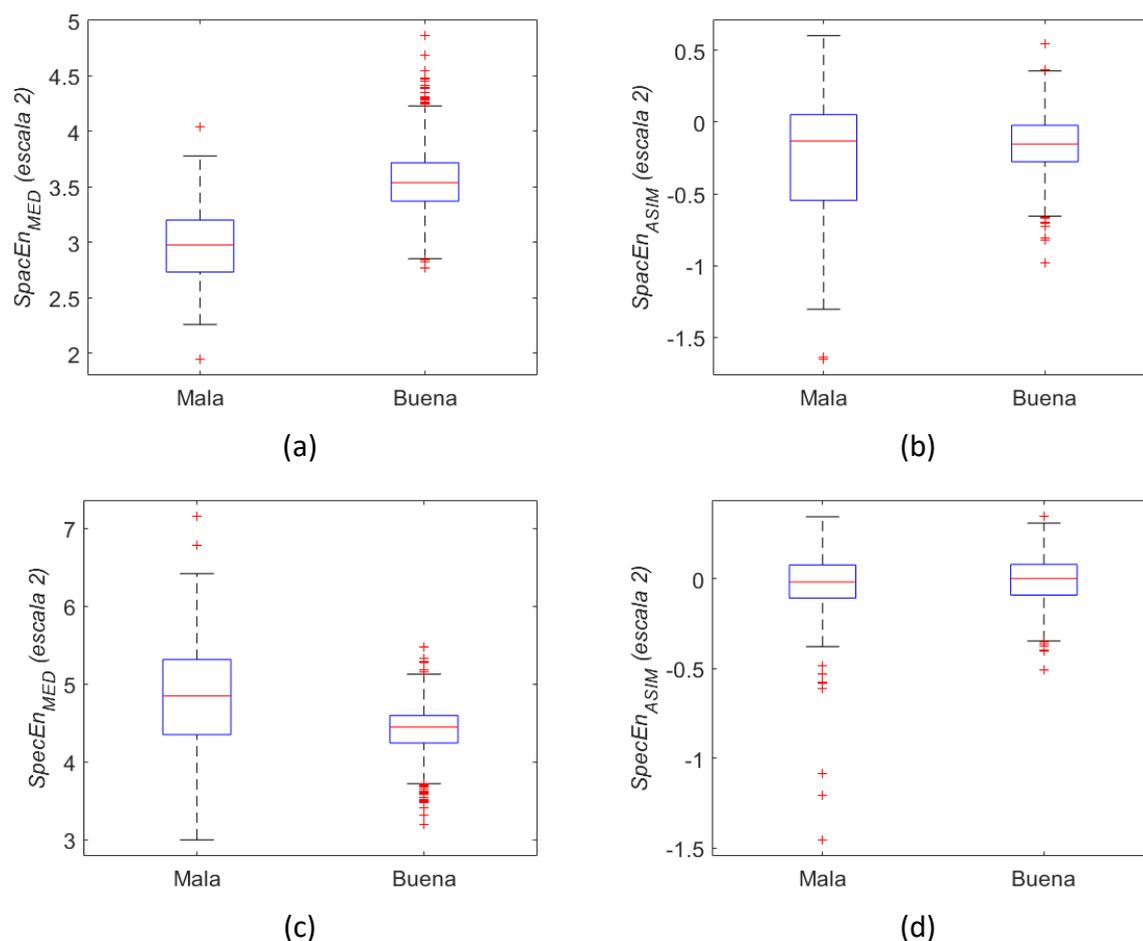


Figura 5.1: Representación mediante *boxplots* de los valores obtenidos con el método NIQE.

Otra de las técnicas aplicadas para la extracción de características fue SSEQ. A partir de ella, se calcularon la media y asimetría de  $SpacEn$  y  $SpecEn$  en tres escalas distintas, obteniéndose un total de 12 características. Al igual que en el caso anterior, es posible hacer un estudio de los *boxplots* obtenidos con cada medida para analizar el poder discriminatorio de estas características (Figuras 5.2, 5.3 y 5.4). Se pueden observar notables diferencias entre las imágenes de buena calidad y las de calidad insuficiente observando algunas características. En concreto,  $SpacEn_{MED}$  (escalas 1, 2 y 3) reveló grandes diferencias entre imágenes de buena y mala calidad. También se consiguió buena separación de clases con  $SpacEn_{ASIM}$  (escala 1). Las medidas de  $SpecEn_{MED}$  (escalas 1 y 2) también muestran diferencias entre ambas clases. Finalmente,  $SpecEn_{ASIM}$  (escalas 1 y 3) muestra leves diferencias entre las clases de buena y mala calidad, aunque no son tan claras como en los casos anteriores. Como se puede comprobar, la mayoría de las características obtenidas pueden ser útiles para determinar la calidad de las imágenes. El tiempo de cálculo de las características extraídas a partir del algoritmo SSEQ sobre las imágenes está en torno a 3 segundos por cada una de las imágenes. Por lo tanto, este paso del método tampoco influye demasiado en la carga computacional total.



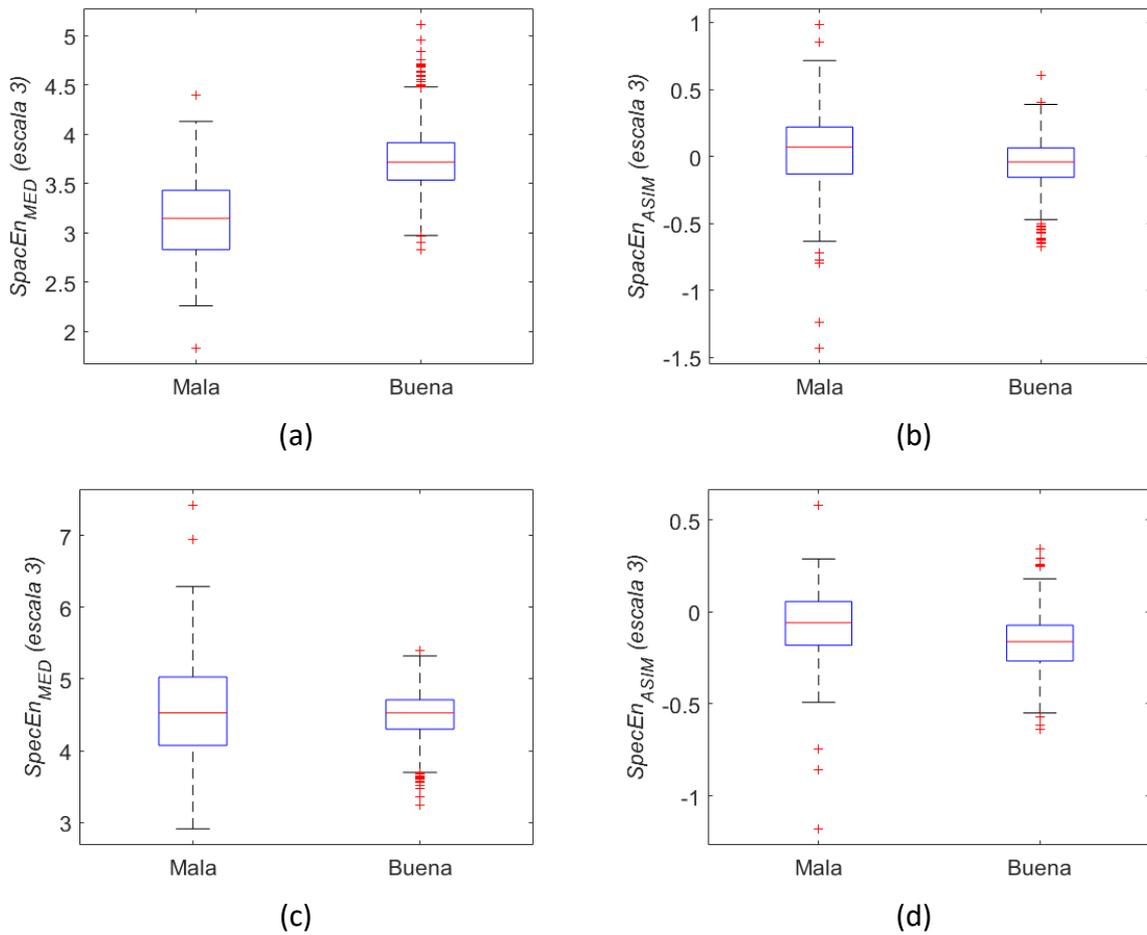
**Figura 5.2:** Representación mediante *boxplots* de las características extraídas mediante SSEQ (escala 1). (a):  $SpacEn_{MED}$ ; (b)  $SpacEn_{ASIM}$ ; (c)  $SpecEn_{MED}$ ; (d)  $SpecEn_{ASIM}$



**Figura 5.3:** Representación mediante *boxplots* de las características extraídas mediante SSEQ (escala 2). (a):  $SpacEn_{MED}$ ; (b)  $SpacEn_{ASIM}$ ; (c)  $SpecEn_{MED}$ ; (d)  $SpecEn_{ASIM}$

Un tercer grupo de características estudiadas en este TFM se extrajo a partir de la CWT 2D y medidas de entropía ( $CWT\_Ent$  en las escalas  $s = 2, 4, 8, 16, 32$  y  $64$ ). También en este caso se obtuvieron los *boxplots* correspondientes (Figura 5.5). Se puede observar que para cualquier valor de  $s$  la separación entre clases es buena, por lo que es adecuado incluir estas características en el método de evaluación de la calidad.

También se analizó la desviación típica local de la CWT 2D ( $CWT\_Med$  y  $CWT\_Desv$  para las escalas  $s = 2, 4, 8, 16, 32, 64$ ). Los resultados obtenidos fueron muy similares a los logrados con  $CWT\_Ent$ . Los *boxplots* resultantes se muestran en las Figuras 5.6 y 5.7. Las diferencias entre clases fueron claras en todas las escalas, especialmente con  $s = 4$  y  $s = 8$ . Estas escalas coinciden con el tamaño típico de los vasos sanguíneos, como se observa en el ejemplo de la Figura 3.4, en la que se puede comprobar que los bordes de la red vascular quedan realzados en estas escalas. Las imágenes de buena calidad muestran las estructuras oculares con mayor contraste y, por lo tanto, los valores de  $CWT\_Ent$ ,  $CWT\_Med$  como  $CWT\_Desv$  que se obtienen para ellas son más elevados. Por otra parte, la elección de la *wavelet* "Mexican Hat" resultó



**Figura 5.4:** Representación mediante *boxplots* de las características extraídas mediante SSEQ (escala 3). (a):  $SpacEn_{MED}$ ; (b)  $SpacEn_{ASIM}$ ; (c)  $SpecEn_{MED}$ ; (d)  $SpecEn_{ASIM}$

adecuada para llevar a cabo el realce de los bordes de los vasos y la papila a la vez que atenúa el ruido. Este resultado concuerda con los criterios de evaluación definidos en el apartado 3.2.1. En general, si los bordes de las estructuras oculares se aprecian claramente, la imagen puede considerarse de buena calidad. Los resultados obtenidos concuerdan con los logrados en estudios previos que emplearon *wavelets* (Bartling *et al.*, 2009; Abdel-Hamid *et al.*, 2016, 2017). Las características basadas en la CWT resultaron ser las más relevantes, aunque también redundantes entre sí.

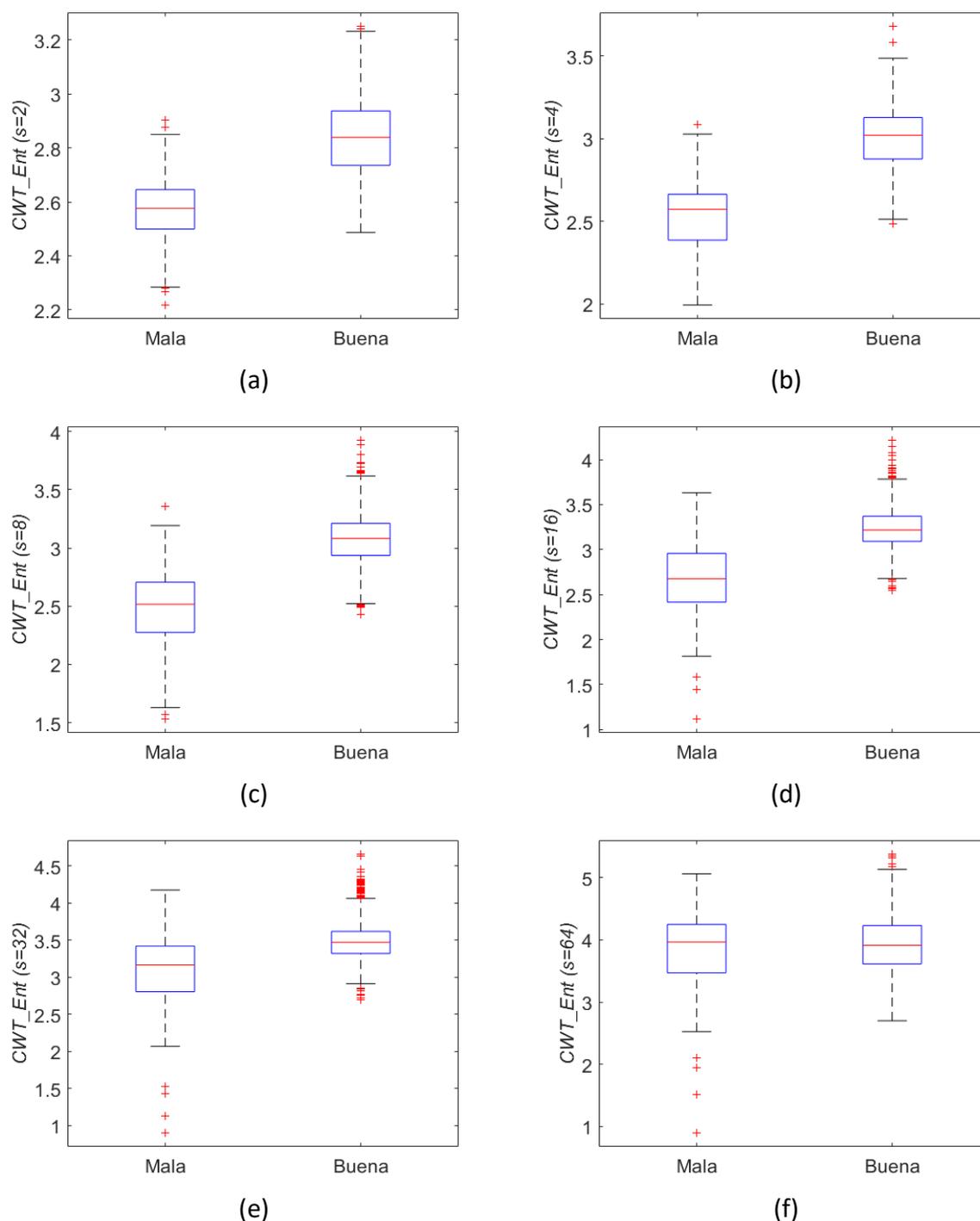
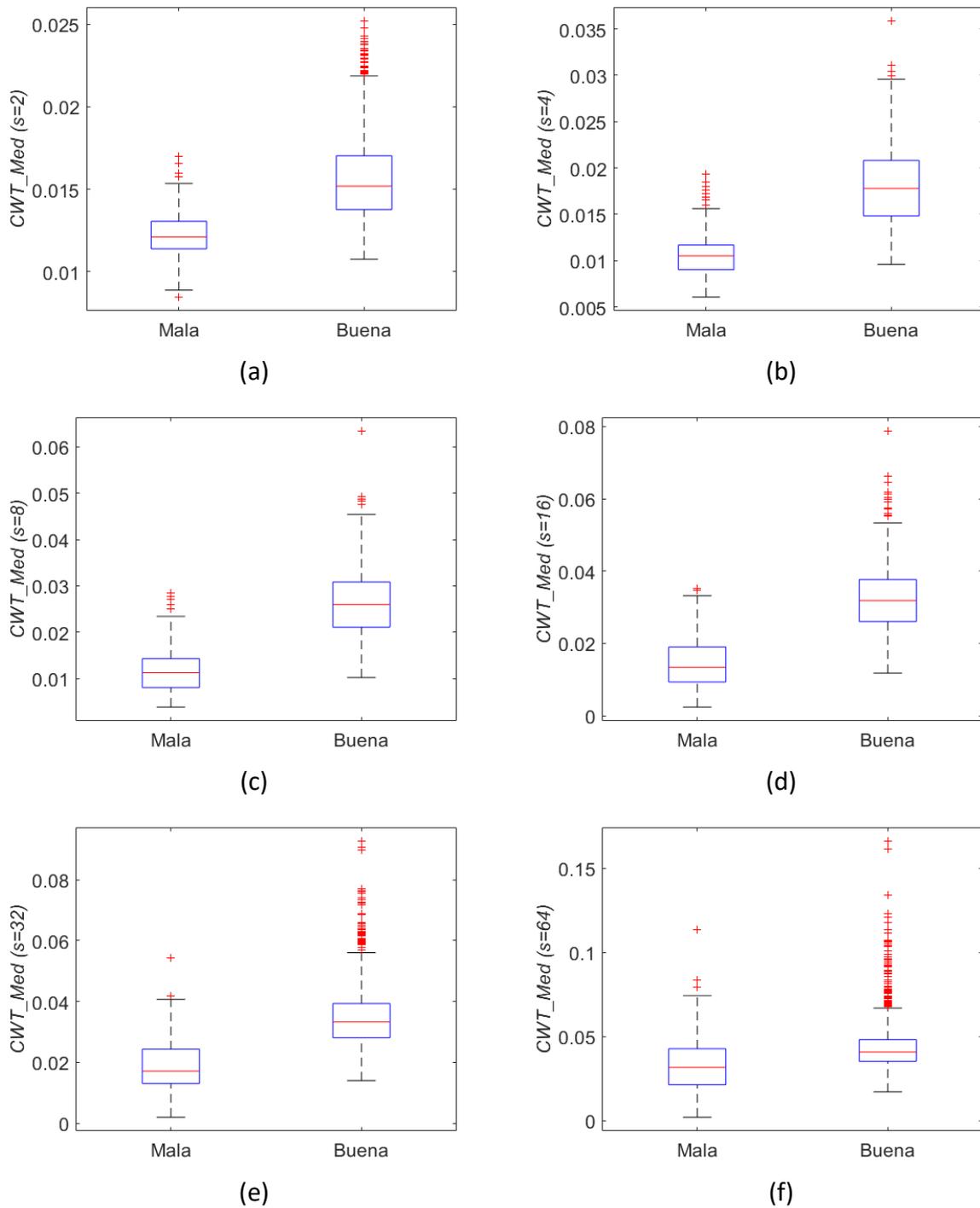


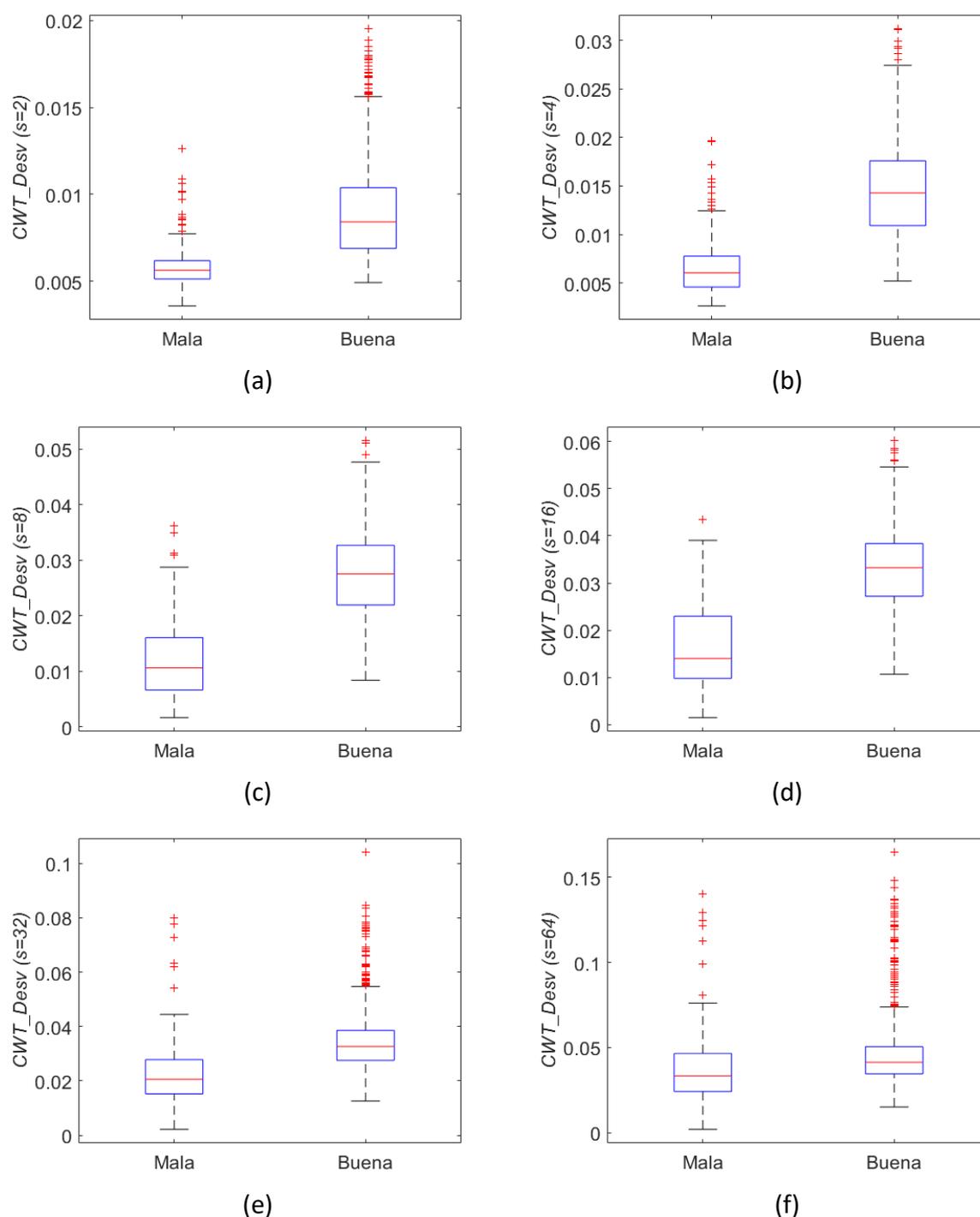
Figura 5.5: Representación mediante *boxplots* de  $CWT\_Ent$  para diferentes escalas. (a)  $s=2$ ; (b)  $s=4$ ; (c)  $s=8$ ; (d)  $s=16$ ; (e)  $s=32$ ; (f)  $s=64$ .

El método basado en la CWT requirió más tiempo que los otros métodos en obtener el conjunto completo de 18 características, alrededor de 15 segundos por cada imagen. Por lo tanto, es el método que más coste computacional tiene en términos de tiempo de ejecución.



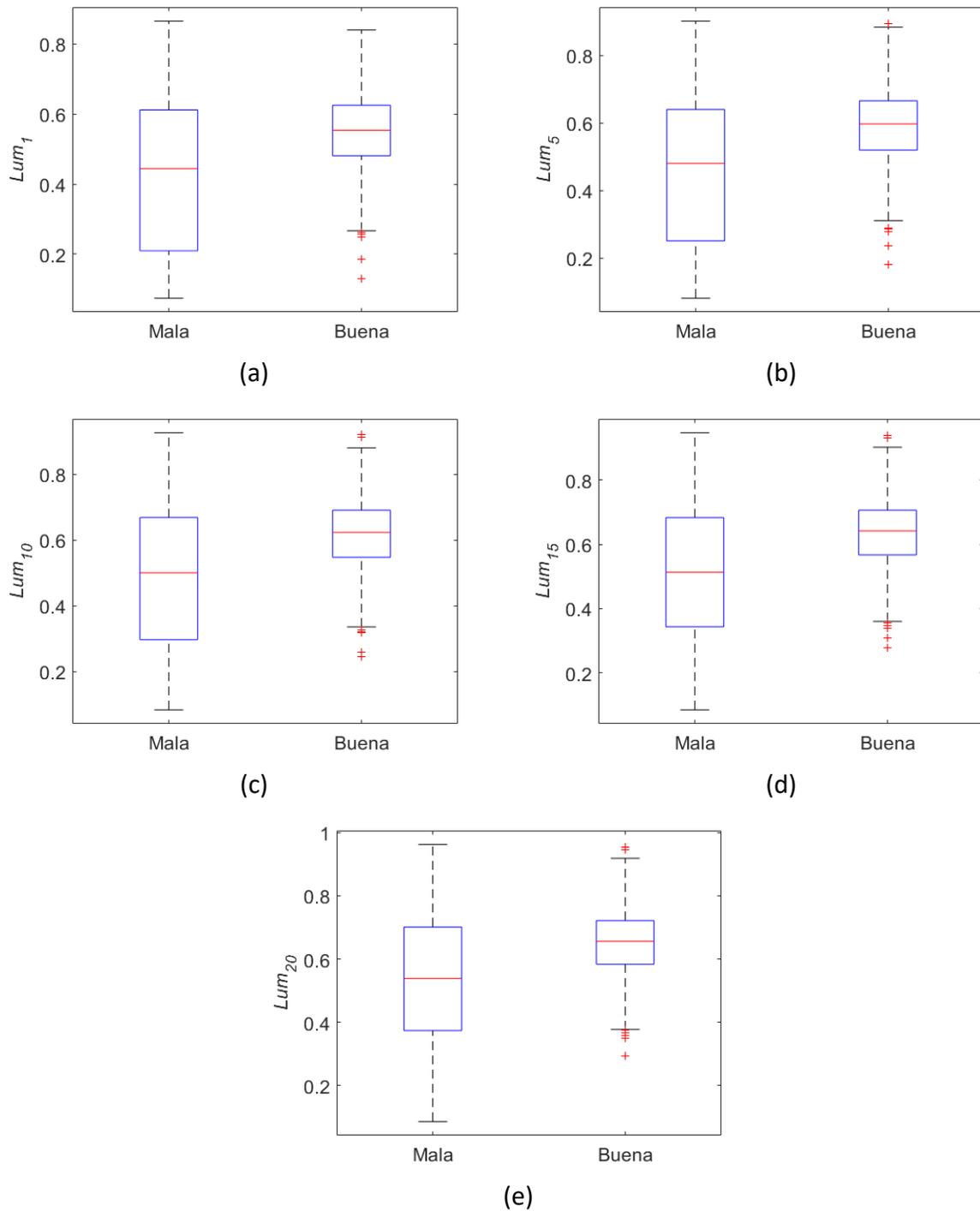
**Figura 5.6:** Representación mediante *boxplots* de  $CWT\_Med$  para diferentes escalas. (a)  $s=2$ ; (b)  $s=4$ ; (c)  $s=8$ ; (d)  $s=16$ ; (e)  $s=32$ ; (f)  $s=64$ .

Finalmente, los resultados obtenidos al analizar las características relativas a la luminosidad del fondo de la imagen en el espacio HSV fueron más discretos que los anteriores.



**Figura 5.7:** Representación mediante *boxplots* de  $CWT\_Desv$  para diferentes escalas. (a)  $s=2$ ; (b)  $s=4$ ; (c)  $s=8$ ; (d)  $s=16$ ; (e)  $s=32$ ; (f)  $s=64$ .

Los *boxplots* correspondientes se muestran en la Figura 5.8. Se puede apreciar que  $Lum_1$ ,  $Lum_5$ ,  $Lum_{10}$ ,  $Lum_{15}$  y  $Lum_{20}$  abarcan un gran rango de valores, aunque en las imágenes de buena calidad éstos se concentran en valores más elevados. Esto indica por una parte que la baja luminosidad del fondo o la presencia de zonas oscurecidas es solo una de las posibles

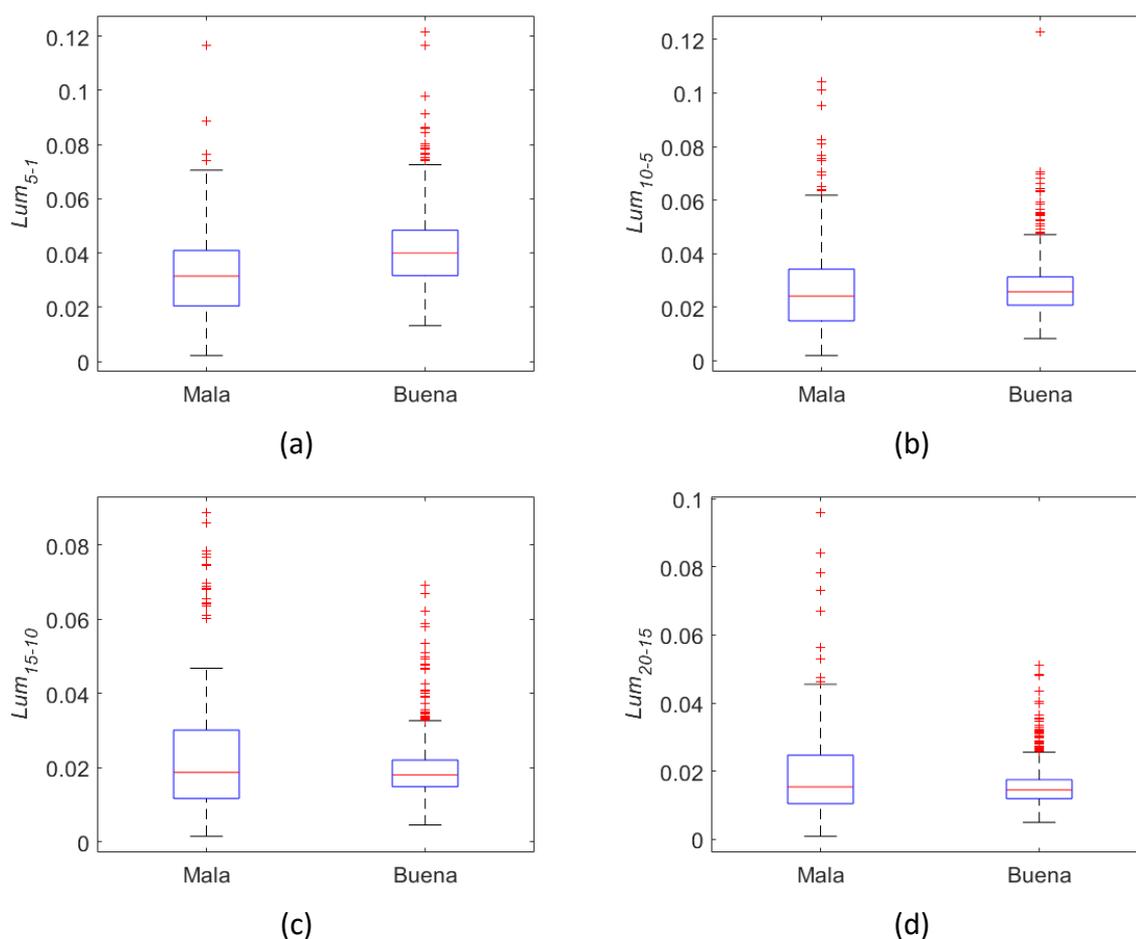


**Figura 5.8:** Representación mediante *boxplots* de las características extraídas a partir de la luminosidad del fondo. (a)  $Lum_1$ ; (b)  $Lum_5$ ; (c)  $Lum_{10}$ ; (d)  $Lum_{15}$ ; (e)  $Lum_{20}$ .

descripciones de una imagen de mala calidad. Por otra parte, las imágenes de buena calidad tienen en general una luminosidad elevada. Ciertamente, nuestra BD incluía imágenes de calidad insuficiente con iluminación aceptable y uniforme, por lo que la calidad no puede evaluarse únicamente con características relacionadas con la luminosidad. Estas

características deben entenderse como un complemento de las características obtenidas previamente. Por otra parte, la información extraída de estas características es redundante (como se puede comprobar en las filas/columnas 32-36 de la Figura 3.9). El algoritmo de selección de características identificó  $Lum_5$  como la característica más relevante de este grupo, y descartó las demás al ser redundantes con ésta.

En cuanto a las características  $Lum_{5-1}$ ,  $Lum_{10-5}$ ,  $Lum_{15-10}$ , y  $Lum_{20-15}$ , se puede observar en la Figura 5.9 que no proporcionan una buena separación de clases. Salvo en el caso de  $Lum_{5-1}$ , los *boxplots* están bastante solapados entre clases. El rango de variación de la luminosidad en el caso de  $Lum_{5-1}$  es más reducido en las imágenes de mala calidad, lo que puede indicar que en este grupo de imágenes las zonas oscuras son más extensas y/o uniformes. No obstante, 2 de estas 4 características fueron seleccionadas por FCBF:  $Lum_{5-1}$  y  $Lum_{15-10}$ . A partir del análisis del *boxplot* de la Figura 5.9(c), parece que el poder discriminatorio de esta característica es reducido. No obstante, es posible que el algoritmo



**Figura 5.9:** Representación mediante *boxplots* de las características extraídas a partir de la luminosidad del fondo. (a)  $Lum_{5-1}$ ; (b)  $Lum_{10-5}$ ; (c)  $Lum_{15-10}$ ; (d)  $Lum_{20-15}$ .

FCBF la haya seleccionado porque aporte información reducida pero suficientemente relevante.

#### 5.4 SELECCIÓN DE CARACTERÍSTICAS

En la fase de selección de características se redujo el número de características, obteniendo una representación compacta de los datos extraídos evitando características redundantes. Como se indicó en el apartado 3.5, se aplicó el algoritmo FCBF con el objetivo principal de reducir la redundancia de las características. FCBF utilizó  $SU$  como una medida de correlación entre las características y las anotaciones realizadas en la BD que permitiera determinar la relevancia de cada una de ellas. También se utilizó  $SU$  como medida de correlación entre características para identificar las características redundantes. Como se indicó en la Tabla 4.1, las características con mayor  $SU$  fueron las basadas en la CWT y el evaluador NIQE. Estos resultados son consistentes con los resultados sobre el poder discriminatorio de cada característica individual (*boxplots* mostrados en las Figuras 5.1, 5.5 (b-d) y 5.7(e)). Empleando el algoritmo FCBF, se seleccionaron únicamente 3 de las características extraídas de la CWT 2D:  $CWT\_Ent$  ( $s=4$ ),  $CWT\_Ent$  ( $s=16$ ) y  $CWT\_Desv$  ( $s=32$ ). Es probable que el resto de las características fueran eliminadas por ser redundantes con estas. Además, se ha comprobado que la característica  $D\_NIQE$  resulta útil y aporta información complementaria respecto a las demás características. Es necesario mencionar que el algoritmo FCBF también seleccionó algunas de las características relacionadas con las entropías calculadas a partir del algoritmo SSEQ, especialmente en la escala 1. En concreto, el algoritmo incluyó  $SpacEn_{ASIM}$ ,  $SpecEn_{MED}$  y  $SpecEn_{ASIM}$  (escalas 1 y 3) como características relevantes. Sin embargo,  $SpacEn_{MED}$  (escalas 1, 2, y 3) se consideró redundante. Las 4 características obtenidas con el algoritmo SSEQ en la escala 2 fueron descartadas por FCBF. Como ya se indicó, estas características no proporcionaron buena separación entre clases en general (Figura 5.3). Finalmente, cabe señalar que también se seleccionaron algunas características relacionadas con la luminosidad de la imagen. En concreto, se incluyeron las características  $Lum_5$ ,  $Lum_{5-1}$  y  $Lum_{15-10}$ . Como se puede ver, el algoritmo FCBF seleccionó características procedentes de todos los métodos propuestos en este TFM. Por lo tanto, las características extraídas se complementan bien entre ellas y aportan información para poder evaluar la calidad de las imágenes.

Es necesario mencionar que algunas de las características que menor  $SU$  obtuvieron también se incluyeron en el conjunto de características seleccionadas por FCBF. Como se puede observar a partir de la matriz de coeficientes de correlación entre las características extraídas (Figura 3.9), estas características no son redundantes entre sí al no estar correladas. Su relevancia es baja, pero superó el umbral fijado en  $SU=0.05$ , por lo que finalmente se seleccionaron. El umbral escogido para  $SU$  es bajo, pero gracias a ello se pudieron incluir en el modelo características que se complementaron con las más relevantes. En este caso, se priorizó la inclusión de más características que aporten algo de información relevante para

mejorar la precisión final del método a minimizar el número de características para reducir la carga computacional.

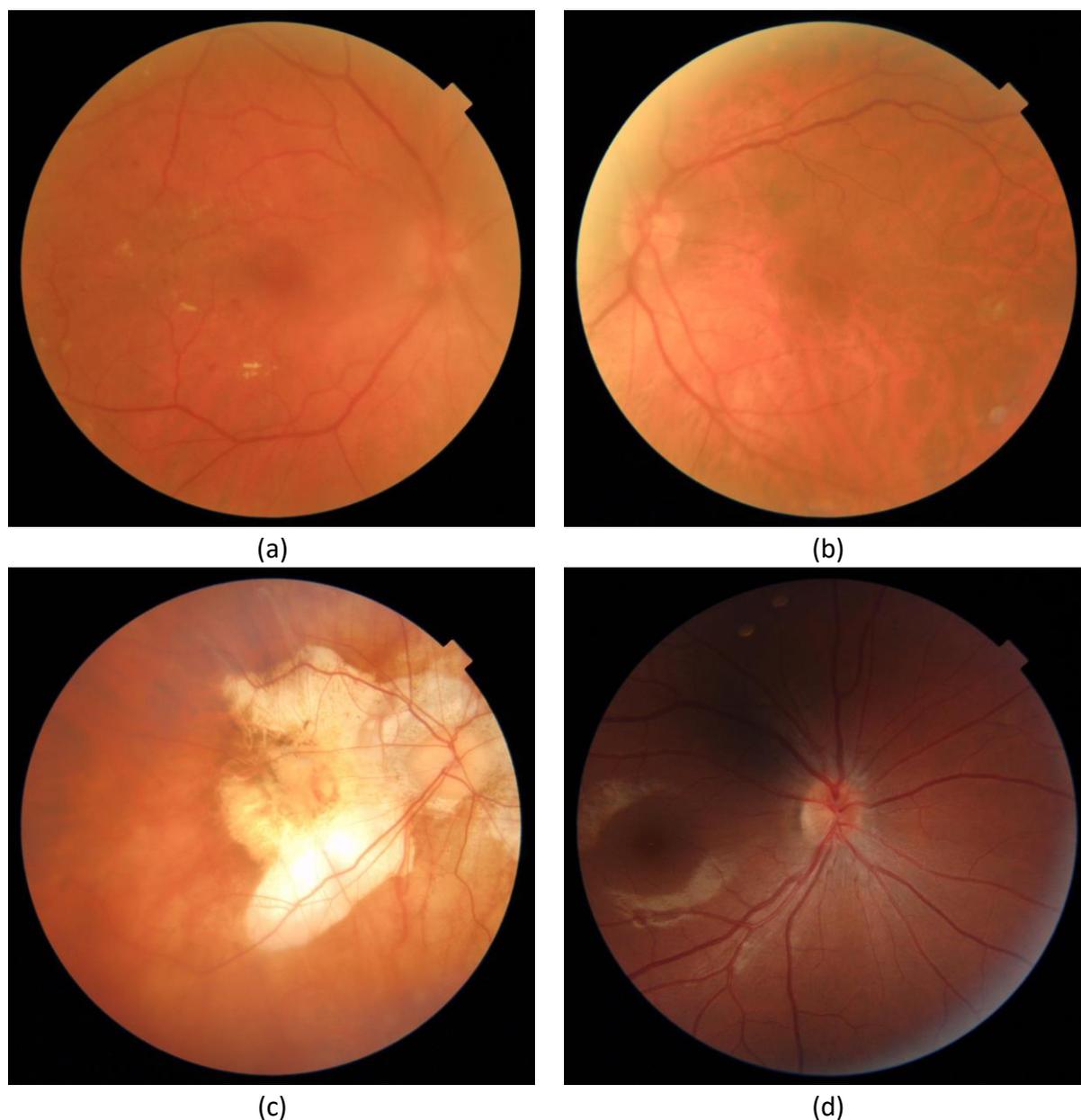
## 5.5 CLASIFICACIÓN MEDIANTE MLP

Las características seleccionadas se emplearon como entradas a un clasificador MLP para obtener la calidad final de las imágenes. Tras las pruebas realizadas durante el entrenamiento empleando una red MLP de tres capas (una capa de entrada, una oculta y una de salida), se determinó que los mejores resultados se consiguieron con 30 neuronas en la capa oculta. Se entrenó la red con un conjunto de entrenamiento preprocesado mediante la técnica de sobremuestreo SMOTE, que incrementa de manera artificial la cantidad de instancias de la clase minoritaria (en este caso, las instancias relativas a imágenes de mala calidad). Para evitar el riesgo de *overfitting* se empleó un criterio de parada temprana junto con la técnica de validación cruzada. Sin embargo, se observó que este criterio es insuficiente para evitar el sobreajuste del clasificador. Por ello, se aplicó adicionalmente la técnica de regularización en la fase de entrenamiento. Con la técnica de validación cruzada se determinó durante el entrenamiento que el valor más adecuado del parámetro de regularización para este caso es  $\nu = 0.2$ . Las razones por las que el criterio de parada temprana no fue suficiente para evitar el sobreajuste podrían estar relacionadas con el empleo de SMOTE. Al aplicar esta técnica se crearon 740 muestras sintéticas a partir de los datos reales, que se dividieron de forma aleatoria entre los conjuntos de entrenamiento y validación al aplicar validación cruzada. Es posible que la técnica de parada temprana no resulte del todo efectiva si las muestras de la clase minoritaria están interrelacionadas entre sí los conjuntos de entrenamiento y validación.

Los resultados de precisión obtenidos sobre el conjunto de test alcanzaron  $Acc=92.13\%$ , y  $F1=0.9531$ . Este último resultado es especialmente relevante, ya que  $F1$  representa el compromiso entre una elevada tasa de detección ( $Se$ ) frente a la probabilidad de acierto en caso de producirse un positivo ( $VPP$ ) (Saito and Rehmsmeier, 2015). Ciertamente, es importante que el valor de  $F1$  sea elevado para que el método desarrollado se pueda trasladar a la práctica clínica. El método de evaluación de la calidad debe ser capaz de detectar correctamente el mayor número posible de retinografías aptas para ser analizadas ( $Se$  elevada, FN bajo) y, a su vez, cometer pocos fallos al considerar de buena calidad imágenes no graduables ( $VPP$  elevado, FP bajo). De todas formas, hay que tener en cuenta que un FP y un FN no tiene las mismas consecuencias en un sistema como el propuesto en este TFM. Un FN implicaría que el oftalmólogo tuviese que repetir una imagen que ya era de buena calidad. Sin embargo, un FP implicaría que una imagen de mala calidad se pasase a las siguientes etapas de un sistema automático más complejo de ayuda al diagnóstico, pudiéndose producir diagnósticos erróneos. Es importante señalar que  $Sp$  fue mucho menor en el conjunto de test que en el conjunto de entrenamiento. Creemos que esto puede ser debido a que en el conjunto de test el número de imágenes de calidad insuficiente es mucho menor que el número de imágenes de buena calidad. La red MLP se entrenó empleando SMOTE, y se evaluó

en un conjunto de test desbalanceado. Entre los resultados obtenidos, se observó que FN=61 y FP=22 en el conjunto de test. Se cometieron pocos FP, pero al contar solo con 149 imágenes de mala calidad  $Sp$  se redujo considerablemente. Esto es relevante a la hora de trasladar este método al entorno clínico. Las consecuencias de reducir el número de FP son más importantes (puesto que pueden acarrear fallos en el diagnóstico) que incrementar el número de VN o reducir el número de FN (puesto que implica únicamente capturar de nuevo una imagen). De cualquier modo, cabe esperar que si el oftalmólogo está bien entrenado, cometa pocos fallos al capturar las imágenes. De hecho, en la BD utilizada, el número de imágenes de buena calidad (1810) era muy superior al número de imágenes de mala calidad (297). Si un oftalmólogo con experiencia captura las imágenes para su posterior análisis, es probable que el método las acepte cuando se hayan obtenido correctamente. Además, tanto el oftalmólogo como el método propuesto coincidirían en su valoración en la mayoría de los casos. Por otra parte, si el oftalmólogo cometiera ocasionalmente un fallo al capturar la imagen, el método propuesto detectaría que la imagen no tiene calidad suficiente en muchas ocasiones. Sin embargo, el valor de  $Sp$  obtenido revela que en algunos casos el sistema permitiría analizar imágenes de mala calidad, con los inconvenientes que conlleva.

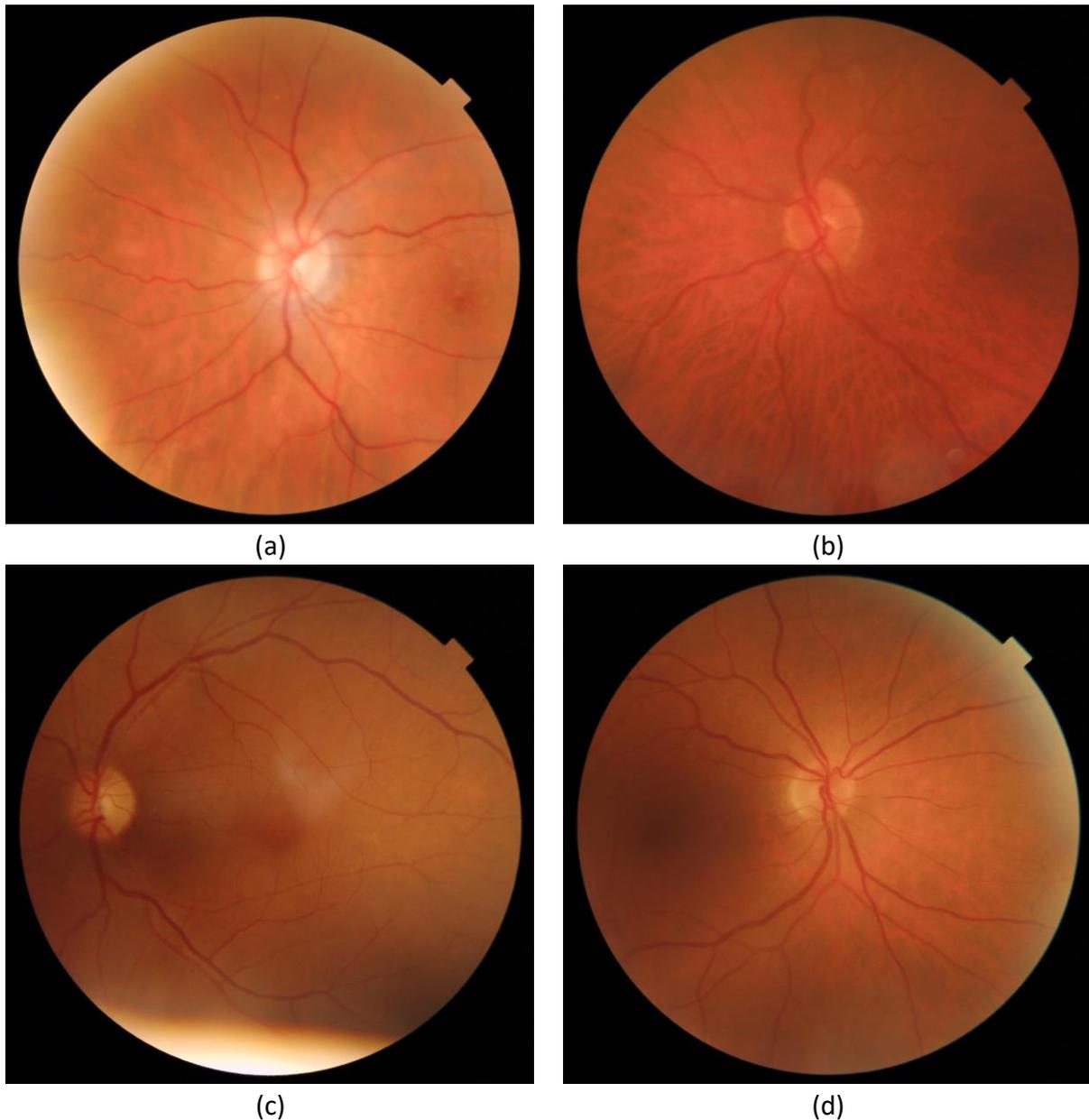
En la Figura 5.10 se muestran ejemplos de imágenes consideradas erróneamente como aptas para su evaluación (casos de FP). Estas imágenes fueron anotadas por los expertos como imágenes de mala calidad y el método de evaluación automática consideró que eran de buena calidad. En la Figura 5.10 (a) se aprecia claramente que los vasos sanguíneos no son visibles en todas las zonas de la imagen, y que la papila no se muestra bien definida. No obstante, se aprecian zonas bien contrastadas en la zona inferior de la imagen y la iluminación es aceptable. Quizás por ello el método de evaluación de la calidad determinó que la imagen tenía calidad suficiente. Un caso similar se muestra en la Figura 5.10 (b), donde la imagen es borrosa en algunas zonas, pero los vasos se aprecian correctamente en la mayor parte de la imagen. También esta imagen se clasificó incorrectamente como imagen de buena calidad empleando el método automático. Los especialistas también marcaron la imagen de la Figura 5.10 (c) como de mala calidad, a pesar de mostrar zonas con buen contraste. No obstante, se puede observar que el contraste no es uniforme a lo largo de la imagen y que en las zonas izquierda e inferior no se aprecian bien los vasos sanguíneos. De nuevo, el método automático clasificó esta imagen como de buena calidad de forma incorrecta. Finalmente, en la Figura 5.10 (d) se muestra una retinografía con buena nitidez en la parte inferior, pero que los especialistas marcaron como de mala calidad debido a las zonas oscurecidas de la parte superior.. Como se puede comprobar, la evaluación de la calidad de las retinografías no es sencilla, debido a que en algunas zonas hay una pérdida de claridad notable frente a las demás. Esto hace que, a pesar de disponer de unos criterios claros para definir la calidad, en algunos casos los expertos tengan dudas al clasificar las imágenes. Sería interesante disponer de las anotaciones de un segundo experto para poder comparar los resultados obtenidos. De esta manera, sería posible comparar los estos resultados con la variabilidad de las anotaciones realizadas por los expertos.



**Figura 5.10:** Imágenes de mala calidad clasificadas erróneamente por el método de evaluación automática (a) Papila y parte de los vasos borrosos; (b) Vasos borrosos en algunas zonas de la imagen; (c) Gran diferencia de contraste entre zonas central y periférica; (d) Imagen parcialmente oscurecida.

De forma complementaria, en la Figura 5.11 se muestran algunos casos de FN. Estas imágenes fueron anotadas por los expertos como imágenes aptas para su análisis, pero el clasificador las rechazó como imágenes con calidad insuficiente. En la Figura 5.11(a) se aprecia que el contraste es bueno y que las estructuras oculares se aprecian correctamente en la mayor parte de la imagen. Sin embargo, la imagen aparece ligeramente borrosa debido al movimiento en el momento de la captura y se aprecia un brillo en la parte inferior izquierda que impide la visibilidad en esa zona. Se puede apreciar también que algunas lesiones aparecen borrosas, lo que dificulta su detección por parte de sistemas automáticos. Por ello, como se comentaba anteriormente, sería deseable poder determinar si habría discrepancias

entre expertos a la hora de determinar la calidad en este tipo de imágenes. En la retinografía de la Figura 5.11 (b) la nitidez de los vasos es baja, y no se aprecian correctamente en algunas zonas. Por lo tanto, y teniendo en cuenta las características que se han considerado en este TFM para entrenar el clasificador, resulta comprensible que esta imagen haya sido considerada de mala calidad por el método automático. Los casos mostrados en las Figuras 5.11 (c-d) son algo distintos, puesto que la papila y los vasos sanguíneos se aprecian claramente, y solo la zona de la mácula aparece un poco menos definida. Por ello, y a pesar de que las imágenes son algo oscuras en algunas zonas y que hay una zona brillante en la Figura 5.11 (c), las imágenes cumplen claramente los criterios para que sean consideradas



**Figura 5.11:** Imágenes de buena calidad clasificadas erróneamente por el método de evaluación automática (a) Imagen borrosa debido al movimiento en la captura; (b) Contraste de los vasos escaso, posible subjetividad en la anotación; (c), (d) Imágenes con calidad aceptable en general.

como de buena calidad. Sin embargo, en estos casos la clasificación automática ha sido claramente incorrecta.

Es necesario mencionar que el número de fallos cometidos en la clasificación automática es bajo con respecto al total de imágenes ( $Acc=92.13\%$ ). En general, los fallos se pueden atribuir a que las características que se han considerado para determinar la calidad de las imágenes son medidas globales sobre la imagen. Por lo tanto, no reflejan pérdidas de calidad en algunas zonas concretas de las imágenes, como sucede en algunas de las imágenes de la BD (por ejemplo, en la Figura 5.10(c y d) o 5.11(a)). Este problema es habitual cuando se emplean métodos genéricos como el propuesto en este TFM (Fasih *et al.*, 2014; Pires Dias *et al.*, 2014). Para mejorar los resultados, sería deseable emplear algún método híbrido que considerase las diferencias de calidad entre zonas de la imagen (Sevik *et al.*, 2014). A pesar de estos errores, en líneas generales, el método desarrollado en este TFM obtuvo unos resultados bastante buenos. Salvo algunos fallos que se han indicado, el método fue capaz de clasificar correctamente la mayoría de las imágenes del conjunto de test.

## 5.6 COMPARATIVA CON ESTUDIOS PREVIOS

Se ha realizado una comparativa de los resultados obtenidos en este TFM con los obtenidos en estudios previos (Tabla 5.1). La mayoría de los estudios previos miden sus resultados en base a  $Se$  y  $Sp$ , aunque en algunos casos se incluyó también  $AUC$  o  $F1$ . En este TFM, se decidió incluir, adicionalmente, medidas de  $Acc$ ,  $VPP$  y  $F1$  para poder analizar los resultados con mayor detalle. Entre los estudios previamente desarrollados, destacan aquellos basados en *Deep Learning*, que consiguieron clasificar correctamente todas las imágenes en los experimentos realizados (Costa *et al.*, 2017; Saha *et al.*, 2018). El resto de los métodos logran  $Se$  y  $Sp$  superiores al 90% en la mayoría de los casos. En el caso de los métodos estructurales, Fleming *et al.* (2012) y Welikala *et al.* (2016) obtuvieron buenos resultados expresados mediante  $Se$  y  $Sp$  utilizando el clasificador SVM. Las características genéricas de nitidez, iluminación y textura empleadas por Davis *et al.* (2009) y Fasih *et al.* (2014) fueron diferentes a las utilizadas en este estudio, obteniéndose resultados similares. En este TFM,  $Sp$  fue menor mientras que  $Se$  fue aceptable, aunque es complicado establecer comparaciones. En cuanto a  $Acc$ , los métodos estructurales de Niemeijer *et al.* (2006) y Giancardo *et al.* (2008) alcanzaron en ambos casos el 97%, superior a la lograda en este estudio. El método híbrido propuesto por Paulus *et al.* (2010) obtuvo  $Acc=91.70$ , inferior a la lograda en estudios similares. Finalmente, Shao *et al.* (2018) empleó NIQE como medida de calidad, obteniendo  $Acc=93.60\%$  similar a la obtenida en este TFM.

En algunos estudios previos se emplearon también curvas ROC para evaluar los resultados. El  $AUC$  obtenido en estos estudios es elevado en todos los casos. Cabe destacar que método propuesto en este TFM empleó una red MLP, al igual que el estudio de Pires Dias *et al.* (2014). El  $AUC$  es muy similar en ambos métodos, y es superior al conseguido con métodos similares que utilizaron SVM como clasificador (Wang *et al.*, 2016). Algunos métodos

	<b>Acc</b>	<b>Se</b>	<b>Sp</b>	<b>VPP</b>	<b>F1</b>	<b>AUC</b>
Niemeijer <i>et al.</i> (2006)	97.40%	-	-	-	-	0.9968
Giancardo <i>et al.</i> (2008)	97.00%	-	-	-	-	-
Fleming <i>et al.</i> (2012)	-	92.60%	90.00%	-	-	0.9690
Yu <i>et al.</i> (2012)	-	95.30%	80.00%	-	-	0.9580
Kohler <i>et al.</i> (2013)	-	89.00%	90.00%	-	-	0.8830
Welikala <i>et al.</i> (2016)	-	95.33%	91.13%	-	-	0.9828
Davis <i>et al.</i> (2009)	-	96.00%	100.00%			
Fasih <i>et al.</i> (2014)	-	92.00%	94.00%			
Pires Dias <i>et al.</i> (2014)	-	99.49%	99.76%	-	-	0.9987
Wang <i>et al.</i> (2016)	-	87.45%	91.66%	-	-	0.9452
Abdel-Hamid <i>et al.</i> (2017)	-	-	-	-	0.8780	-
Paulus <i>et al.</i> (2010)	91.70%	96.90%	80.00%	-	-	0.9530
Sevik <i>et al.</i> (2014)	-	-	-	-	0.9960	-
Shao <i>et al.</i> (2018)	93.60%	94.69%	92.29%	-	-	-
Costa <i>et al.</i> (2017)	100.00%	100.00%	100.00%	-	-	-
Saha <i>et al.</i> (2018)	100.00%	100.00%	100.00%	-	-	-
Método propuesto en este TFM	92.13%	93.26%	85.23%	97.46%	0.9531	0.9992

Tabla 5.1: Comparación de los resultados logrados en estudios previos.

recientes incluyen resultados mediante el valor *F1*. Los resultados obtenidos por Abdel-Hamid *et al.* (2017) fueron inferiores a los alcanzados en este TFM, mientras que el valor *F1* obtenido por otros autores fue superior utilizando más características junto con la segmentación de las estructuras oculares (Sevik *et al.*, 2014).

En resumen, los resultados obtenidos en este TFM son buenos y similares a los obtenidos en estudios previos como se puede observar en la Tabla 5.1. Sin embargo, es necesario mencionar que es difícil establecer comparaciones entre estudios. En la mayoría de los casos, los resultados se obtuvieron en BD propietarias, por lo que son diferentes para cada estudio. La única excepción es el trabajo de Sevik *et al.* (2014), en el que se propuso una nueva BD pública para evaluar la calidad en retinografías. En el futuro, sería deseable obtener resultados del método propuesto sobre esta BD pública, de modo que se puedan comparar los resultados con mayor exactitud.



## Capítulo 6. Conclusiones y líneas futuras

---

### 6.1 INTRODUCCIÓN

En este capítulo se recogen las principales contribuciones y conclusiones del estudio, así como las líneas futuras de investigación.

### 6.2 GRADO DE CUMPLIMIENTO DE LOS OBJETIVOS

El principal objetivo de este TFM fue el desarrollo de un método automático para evaluar la calidad en retinografías. Para ello, se extrajeron y seleccionaron características globales de las imágenes para su posterior clasificación. El método desarrollado se probó en una BD de 2107 retinografías con anotaciones sobre su calidad. De esta manera, se puede afirmar que se han cumplido los objetivos de este TFM. Concretamente, se han cumplido los siguientes objetivos:

- Se ha realizado una revisión bibliográfica con el objetivo de obtener toda la información y comprender todos los conceptos necesarios para la realización de este TFM. Además, se identificaron y estudiaron numerosos métodos de análisis de retinografías que permitieron desarrollar el método propuesto en este TFM. Esta búsqueda exhaustiva de información posibilitó la consecución de objetivos posteriores.
- Se ha colaborado en la creación de una BD de retinografías procedentes de programas de cribado de la RD. Esta BD es útil para la implementación y validación de métodos de análisis de retinografías que sirvan para el desarrollo de un sistema automático de ayuda al diagnóstico de la RD.
- Se han implementado los métodos descritos utilizando el *software* MATLAB®. Se comprobaron las capacidades de su *toolbox* de procesamiento de imágenes y de otros *toolbox* complementarios, como el *Statistics and Machine Learning Toolbox* y el *Neural Network Toolbox*. Por lo tanto, MATLAB® resultó útil para alcanzar los objetivos del TFM.
- Se evaluó la utilidad de los métodos propuestos empleando la BD disponible. Para ello se emplearon diversos tipos de medidas: *Se*, *Sp*, *Acc*, *VPP*, *F1* y *AUC*.
- Los resultados obtenidos se pueden considerar satisfactorios y fueron comparables a los obtenidos en estudios previos. Cada una de las etapas del método propuesto contribuyeron a estos resultados finales:
  - En la fase de preprocesado se consiguió eliminar los efectos de borde que provocan las etapas de procesamiento posteriores. Eliminando el borde negro que

- delimita la apertura de la cámara se obtuvieron imágenes más homogéneas y se mejoró la extracción de características.
- Se extrajeron un total de 40 características para la descripción de las imágenes de buena y mala calidad. Para ello, se emplearon diversos tipos de métodos que permitieron obtener características relevantes y complementarias. El subconjunto óptimo de estas características se determinó utilizando FCBF, obteniéndose las 13 características más relevantes para la clasificación.
  - Se empleó una red MLP para evaluar la calidad de las imágenes. De este modo, se consiguió clasificar las retinografías de buena y mala calidad empleando una arquitectura de RNA sencilla.
  - Se logró desarrollar un método de evaluación automática de la calidad de las retinografías. Este método podría emplearse en un sistema más general de análisis de retinografías para la ayuda al diagnóstico de diversas patologías oculares, entre las que se encontraría la RD.
  - Se ha tratado de difundir los resultados obtenidos en este TFM en foros científicos. En este sentido, se ha enviado una comunicación al XXXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2018).

### 6.3 CONTRIBUCIONES ORIGINALES

En este TFM se obtuvieron medidas globales de la calidad de las retinografías. Para ello, se aplicaron, entre otros, los métodos NIQE y SSEQ. Estos métodos han sido desarrollados fuera del ámbito médico, con un propósito más general y en el contexto de la evaluación de la calidad de diversos tipos de imágenes. En este trabajo se ha podido comprobar su aplicabilidad sobre imágenes de retina. De manera complementaria, se analizaron las imágenes mediante la CWT y se obtuvieron características relativas a la luminosidad del fondo de las imágenes empleando el espacio de color HSV. Los resultados obtenidos en este TFM son comparables a los obtenidos en estudios previos para detectar de forma automática la calidad de retinografías, lo que pone de manifiesto la utilidad del método propuesto. Las principales contribuciones del trabajo realizado son:

- Se ha colaborado en la obtención de una BD de retinografías para la prueba y evaluación de algoritmos de análisis automático de imágenes de fondo de ojo. La revisión de las anotaciones siguiendo las indicaciones establecidas sobre la calidad de las imágenes permitió reducir la subjetividad de los observadores humanos. Mediante la obtención y anotación de esta BD, es posible desarrollar, probar y validar métodos de procesamiento de retinografías para la ayuda al diagnóstico de la RD.
- Se han implementado las técnicas NIQE y SSEQ, utilizadas previamente en tareas de análisis de calidad de propósito general fuera del ámbito médico. Se ha comprobado que estas sencillas técnicas son aplicables al análisis de la calidad de las retinografías. Los resultados obtenidos fueron satisfactorios y permitieron obtener características

útiles para la evaluación de la calidad de las retinografías. Hasta donde tenemos conocimiento, no se ha aplicado la técnica SSEQ en el análisis de retinografías.

- Se ha estudiado la utilidad de la CWT 2D para la evaluación de la nitidez de las retinografías. Se ha comprobado que esta herramienta es útil para evaluar la claridad con la que se muestran las estructuras oculares, especialmente los vasos sanguíneos y la papila. Los resultados obtenidos fueron buenos y también se obtuvieron características útiles para la evaluación de la calidad. La CWT 2D se ha aplicado en numerosos estudios relacionados con el análisis de retinografías, pero hasta donde tenemos conocimiento, no se ha empleado en la tarea de evaluación de la calidad.
- Se ha estudiado la utilidad del espacio de color HSV para la implementación de un método de análisis de la luminosidad basado en la extracción del fondo de las retinografías. En este caso, los resultados fueron parcialmente satisfactorios. Algunas características fueron útiles para la evaluación de la calidad, pero otras resultaron poco relevantes.
- Se ha utilizado el algoritmo FCBF de selección de características, con el que se obtuvo un conjunto de características más compacto. Se incluyeron las características relevantes y se eliminó la redundancia entre las características extraídas. Este algoritmo se ha utilizado en muchas aplicaciones (Saeys *et al.*, 2007), pero hasta donde tenemos conocimiento no se ha aplicado en la evaluación de la calidad en retinografías. Los resultados obtenidos con esta técnica permitieron una implementación más eficiente del algoritmo de clasificación, a la vez que se obtuvieron mejores resultados. Además, proporcionó información valiosa sobre las características que menos contribuían a la diferenciación de imágenes de buena y mala calidad.
- El desarrollo de este método ha hecho posible la diferenciación automática de las imágenes de la BD proporcionada. Los resultados obtenidos indican un alto nivel de concordancia entre los expertos y el método de evaluación automática de la calidad.

## 6.4 CONCLUSIONES

En este trabajo se han estudiado métodos de análisis de la calidad en retinografías. Para ello, se han analizado técnicas genéricas, técnicas estructurales y técnicas híbridas que combinan ambos enfoques. Estas técnicas de análisis de la calidad en retinografías no han recibido tanta atención como los métodos para la detección de lesiones o estructuras oculares, pero los logros obtenidos son destacables. Los métodos propuestos en la literatura para determinar la calidad en retinografías incluyen un gran número de técnicas de análisis de imagen, incluyendo técnicas avanzadas y relativamente recientes como *Deep Learning*. La principal conclusión de este TFM es que el método propuesto permite diferenciar de forma automática entre imágenes de buena y mala calidad. Los resultados son adecuados y sería posible incluir el método propuesto como primer paso en un sistema de análisis de retinografías como ayuda al diagnóstico de la RD. De manera más específica, se puede

concluir que las características extraídas mediante los métodos de propósito general NIQE y SSEQ han resultado útiles para la evaluación de la calidad de las imágenes. También se demostró que la CWT 2D es útil para extraer características relacionadas con la claridad de las imágenes, si bien algunas de ellas fueron redundantes. Gracias al algoritmo FCBF fue posible obtener un subconjunto óptimo para la clasificación de las características. La clasificación mediante la red MLP produjo buenos resultados, si bien fue necesario controlar el sobreajuste con técnicas de regularización y *early stopping* de manera conjunta para lograr un clasificador generalizable.

## 6.5 LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

A pesar de que los resultados obtenidos son satisfactorios y son similares a los de estudios previos, este estudio no está exento de limitaciones. En primer lugar, es necesario mencionar que se ha realizado un gran esfuerzo por parte de oftalmólogos del IOBA y del HCUV, así como por parte de los ingenieros del Grupo de Ingeniería Biomédica de la Universidad de Valladolid para obtener la BD que se ha utilizado en este TFM. No obstante, el reducido número de imágenes de calidad insuficiente en la BD es una limitación reseñable. Hubiese sido deseable disponer de más imágenes de calidad insuficiente para poder proporcionar a la red MLP muchos más ejemplos reales de imágenes de mala calidad. De esta manera, se habría logrado una implementación de la red MLP más robusta. Para tratar de solventar esta limitación, se implementó la técnica SMOTE. Con esta técnica se han generado muestras sintéticas para el entrenamiento y configuración de la red MLP. Debido a ello, ha sido necesario emplear varias técnicas para evitar el sobreajuste. Sería deseable estudiar si incluir más imágenes de mala calidad en la BD permitiría reducir la complejidad del método desarrollado.

Por otra parte, cabe destacar que la anotación manual de calidad de las imágenes no está exenta de cierto grado de subjetividad. Los expertos valoraron la calidad siguiendo unos criterios claros y pactados. No obstante, en algunas imágenes las características varían entre las diversas zonas, lo que hace difícil determinar la calidad en algunos casos. Este problema es habitual cuando se trata de evaluar la calidad de las imágenes (Niemeijer *et al.*, 2006; Paulus *et al.*, 2010). El haber incluido una fase de revisión de la base de datos por parte de un mayor número de expertos disminuyó la subjetividad y mejoró las anotaciones. No obstante, sería deseable contar con las anotaciones de diversos expertos para poder evaluar los resultados obtenidos empleando el método automático con la variabilidad entre expertos.

Existen algunas limitaciones que afectan al método de evaluación desarrollado. En la etapa de preprocesado, el método de extensión del FOV necesita aproximadamente un minuto por cada imagen para eliminar por completo el borde del FOV. En el futuro sería deseable estudiar alternativas para mejorar la eficiencia del algoritmo, así como emplear una resolución más baja en esta parte del procesado. En la fase de extracción de características únicamente se han empleado métodos genéricos de evaluación de la calidad. Los métodos

estructurales e híbridos han demostrado su utilidad en la evaluación de la calidad de las imágenes (Sevik *et al.*, 2014; Abdel-Hamid *et al.*, 2016). Por lo tanto, se plantea como línea futura combinar los métodos desarrollados con métodos estructurales para incrementar la eficacia del método. Los resultados obtenidos empleando la red MLP como clasificador han sido satisfactorios. Sin embargo, sería deseable estudiar y comparar varios esquemas de clasificación en el futuro, con el objetivo de comparar los resultados obtenidos con cada uno de ellos. Por último, en este trabajo no se ha planteado la posibilidad de que el método pueda utilizarse para evaluar la calidad de retinografías de diferentes resoluciones y grados de apertura del FOV. La adaptabilidad a imágenes capturadas con cámaras diferentes (lo que implica, generalmente, diferente resolución y FOV) es importante para la implementación exitosa de métodos de evaluación de la calidad (Abdel-Hamid *et al.*, 2018). Como línea futura se plantea explorar las posibilidades de adaptar el método desarrollado a una mayor variabilidad de imágenes.

Otro de los aspectos relevantes a desarrollar en futuros estudios sería integrar el método desarrollado en un sistema completo de análisis de imágenes de fondo de ojo como herramienta de diagnóstico asistido por ordenador para el diagnóstico y cribado de la RD. Finalmente, sería deseable difundir los resultados del estudio en revistas y/o congresos relacionados con el análisis de imágenes médicas o la ingeniería biomédica.



## Bibliografía

---

Abdel-Hamid L, El-Rafei A, El-Ramly S, Michelson G. Performance dependency of retinal image quality assessment algorithms on image resolution: analyses and solutions. *Signal, Image Video Process* 2018; 12: 9–16.

Abdel-Hamid L, El-Rafei A, El-Ramly S, Michelson G, Hornegger J. Retinal image quality assessment based on image clarity and content. *J Biomed Opt* 2016; 21: 096007.

Abdel-Hamid L, El-Rafei A, Michelson G. No-reference quality index for color retinal images. *Comput Biol Med* 2017; 90: 68–75.

Abràmoff MD, Garvin MK, Sonka M. Retinal Imaging and Image Analysis. *Biomed Eng IEEE Rev* 2010; 3: 169–208.

Aguirre-Díaz O. Diseño y desarrollo de métodos automáticos para la evaluación de la calidad en imágenes de fondo de ojo. *Proy Fin Carrera* 2014

Aja-Fernandez S, San José Estepar R, Alberola-Lopez C, Westin CF. Image quality assessment based on local variance. *IEEE Int Conf Eng Med Biol Soc* 2006; 1: 4815–8.

Alghadyan AA. Diabetic retinopathy - An update. *Saudi J Ophthalmol* 2011; 25: 99–111.

Alio J, Fernández J, García J, Miralles J, Pastor JC, Zato MA. Guiones de oftalmología. 9ª Edición. Valladolid: Secretariado de publicaciones de la Universidad de Valladolid; 1993.

American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2009; 32

Antoine J-P, Carrette P, Murenzi R, Piette B. Image analysis with two-dimensional continuous wavelet transform. *Signal Processing* 1993; 31: 241–72.

Antoine J-P, Murenzi R. Two-dimensional directional wavelets and the scale-angle representation. *Signal Processing* 1996; 52: 259–81.

Bartling H, Wanger P, Martin L. Automated quality evaluation of digital fundus photographs. *Acta Ophthalmol* 2009; 87: 643–7.

Bernardes R, Serranho P, Lobo C. Digital ocular fundus imaging: A review. *Ophthalmologica* 2011; 226: 161–81.

Besenczi R, Tóth J, Hajdu A. A review on automatic analysis techniques for color fundus photographs. *Comput Struct Biotechnol J* 2016; 14: 371–84.

Bishop CM. *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc.; 1995.

Boucher MC, Gresset JA, Angioi K, Olivier S. Effectiveness and safety of screening for diabetic retinopathy with two nonmydriatic digital images compared with the seven standard

- stereoscopic photographic fields. *Can J Ophthalmol* 2003; 38: 557–68.
- Bronzino JD. *The biomedical engineering handbook*. CRC Press - Taylor & Francis Group; 2006.
- Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–57.
- Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet* 2010; 376: 124–36.
- Corcostegui B. *El fondo de ojo en la medicina práctica*. Barcelona: ESPAXS; 1983.
- Costa P, Campilho A, Hooi B, Smailagic A, Kitani K, Liu S, *et al.* EyeQual: Accurate, Explainable, Retinal Image Quality Assessment. 2017 16th IEEE Int Conf Mach Learn Appl 2017: 323–30.
- Cuevas E, Zaldívar D, Pérez-Cisneros M. *Procesamiento digital de imágenes usando MatLAB & Simulink*. Ra-Ma; 2010.
- Davis H, Russell S, Barriga E, Abramoff M, Soliz P. Vision-based, real-time retinal image quality assessment. In: 2009 22nd IEEE International Symposium on Computer-Based Medical Systems. IEEE; 2009. p. 1–6.
- Department of Health. *National Service Framework for Diabetes: Delivery Strategy*. Londres: 2002.
- Dhawan AP. *Medical image analysis*. John Wiley & Sons; 2011.
- Fasih M, Langlois JMP, Ben Tahar H, Cheriet F. Retinal image quality assessment using generic features. In: Aylward S, Hadjiiski LM, editor(s). *Proc.SPIE Medical Imaging 2014: Computer-Aided Diagnosis*. 2014. p. 90352Z.
- Fathi A, Naghsh-Nilchi AR. Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. *Biomed Signal Process Control* 2013; 8: 71–80.
- Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006; 27: 861–74.
- Federación Internacional de Diabetes. *Diabetes Atlas de la FID*. Octava Edi. 2017.
- Fleming AD, Philip S, Goatman KA, Olson JA, Sharp PF. Automated assessment of diabetic retinal image quality based on clarity and field definition. *Investig Ophthalmol Vis Sci* 2006; 47: 1120–5.
- Fleming AD, Philip S, Goatman KA, Sharp PF, Olson JA. Automated clarity assessment of retinal images using regionally based structural and statistical measures. *Med Eng Phys* 2012; 34: 849–59.
- Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, *et al.* Diabetic Retinopathy. *Diabetes Care* 2003; 26: 226–9.
- Fong DS, Aiello LP, Ferris FL, Klein R. Diabetic Retinopathy. *Diabetes Care* 2004; 27: 2540–53.
- Fraz MM, Remagnino P, Hoppe A, Uyyanonvara B, Rudnicka AR, Owen CG, *et al.* Blood vessel segmentation methodologies in retinal images - A survey. *Comput Methods Programs Biomed* 2012; 108: 407–33.

- 
- García M. Procesado de retinografías basado en redes neuronales para la detección automática de lesiones asociadas a la retinopatía diabética. PhD Diss 2008
- García M, Sánchez CI, López MI, Abásolo D, Hornero R. Neural network based detection of hard exudates in retinal images. *Comput Methods Programs Biomed* 2009; 93: 9–19.
- Giancardo L, Abramoff MD, Chaum E, Karnowski TP, Meriaudeau F, Tobin KW. Elliptical local vessel density: A fast and robust quality metric for retinal images. 2008 30th Annu Int Conf IEEE Eng Med Biol Soc 2008: 3534–7.
- Gilmartin B, Amer AC, Ingleby S. Reversal of tropicamide mydriasis with single instillations of pilocarpine can induce substantial pseudo-myopia in young adults. *Ophthalmic Physiol Opt* 1995; 15: 475–9.
- Gómez-Ulla F. Nuevos métodos de exploración retiniana. Madrid: Sociedad Española de Oftalmología; 1994.
- Gonzalez RC, Woods RE. *Digital image processing*. 3rd Ed. Prentice Hall; 2008.
- Graue E. Estructura y función. In: Graue E, editor(s). *Oftalmología*. Méjico D.F.: Interamericana McGraw-Hill; 1995. p. 1–17.
- Grisan E. *Automatic Analysis of Retinal Images: Retinopathy Detection and Grading*. PhD Diss 2005
- Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003; 3: 1157–82.
- Hackel RE, Saine PJ. Creating Retinal Fundus Maps. *J Ophthalmic Photogr* 2005; 27: 10–8.
- Hunter A, Lowell JA, Habib M, Ryder B, Basu A, Steel D. An automated retinal image quality grading algorithm. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. 2011. p. 5955–8.
- International Diabetes Federation, The Fred Hollows Foundation. *Diabetes y salud ocular. Una guía para los profesionales de la salud*. Bruselas, Bélgica: 2015.
- Jain AK, Duin RPW, Mao J. Statistical pattern recognition: a review. *IEEE Trans Pattern Anal Mach Intell* 2000; 22: 4–37.
- Kohler T, Budai A, Kraus MF, Odstrcilik J, Michelson G, Hornegger J. Automatic no-reference quality assessment for retinal fundus images using vessel segmentation. *Proc 26th IEEE Int Symp Comput Med Syst* 2013: 95–100.
- Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 2016; 5: 221–32.
- Lalonde M, Gagnon L, Boucher M-C, others. Automatic visual quality assessment in optical fundus images. In: *Proceedings of vision interface*. 2001. p. 259–64.
- Lee SC, Wang Y. Automatic retinal image quality assessment and enhancement. In: Hanson KM, editor(s). *Medical Imaging 1999: Image Processing*. 1999. p. 1581–90.

- Lim JS. Two-dimensional Signal and Image Processing. Prentice Hall; 1990.
- Lin J-W, Weng Q, Xue L-Y, Cao X-R, Yu L. A retinal image sharpness metric based on histogram of edge width. *J Algorithm Comput Technol* 2017; 11: 292–300.
- Liu L, Liu B, Huang H, Bovik AC. No-reference image quality assessment based on spatial and spectral entropies. *Signal Process Image Commun* 2014; 29: 856–63.
- Mansour RF. Evolutionary Computing Enriched Computer-Aided Diagnosis System for Diabetic Retinopathy: A Survey. *IEEE Rev Biomed Eng* 2017; 10: 334–49.
- Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans Med Imaging* 2011; 30: 146–58.
- Martínez Rubio M, Moya Moya M, Bellot Bernabé A, Belmonte Martínez J. Cribado de retinopatía diabética y teleoftalmología. *Arch Soc Esp Oftalmol* 2012; 87: 392–5.
- Mitra A, Roy S, Roy S, Setua SK. Enhancement and restoration of non-uniform illuminated Fundus Image of Retina obtained through thin layer of cataract. *Comput Methods Programs Biomed* 2018; 156: 169–78.
- Mittal A, Moorthy AK, Bovik AC. No-reference image quality assessment in the spatial domain. *IEEE Trans Image Process* 2012; 21: 4695–708.
- Mittal A, Soundararajan R, Bovik AC. Making a ‘Completely Blind’ Image Quality Analyzer. *IEEE Signal Process Lett* 2013; 20: 209–12.
- Mompín Poblet J. Introducción a la bioingeniería. Marcombo; 1988.
- Mookiah MRK, Acharya UR, Fujita H, Tan JH, Chua CK, Bhandary S V., *et al.* Application of different imaging modalities for diagnosis of Diabetic Macular Edema: A review. *Comput Biol Med* 2015; 66: 295–315.
- Moorthy AK, Bovik AC. A Two-stage Framework for Blind Image Quality Assessment. *IEEE Int Conf Image Process* 2010; 17: 2481–4.
- Mustafi D, Engel AH, Palczewski K. Structure of cone photoreceptors. *Prog Retin Eye Res* 2009; 28: 289–302.
- Nentwich MM, Ulbig MW. Diabetic retinopathy - ocular complications of diabetes mellitus. *World J Diabetes* 2015; 6: 489–99.
- Ng EYK, Acharya UR, Suri JS, Campilho A. *Image Analysis and Modeling in Ophthalmology*. Taylor & Francis; 2014.
- Nielsen MA. *Neural Networks and Deep Learning*. Determination Press; 2015.
- Niemeijer M, Abramoff MD, van Ginneken B. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal* 2006; 10: 888–98.

- 
- Park SH, Goo JM, Jo C-H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J Radiol* 2004; 5: 11.
- Paulus J, Meier J, Bock R, Hornegger J, Michelson G. Automated quality assessment of retinal fundus photos. *Int J Comput Assist Radiol Surg* 2010; 5: 557–64.
- Pinies JA. Retinografía con cámara no midriática. *Av en Diabetol* 2005; 21: 217–21.
- Pires Dias JM, Oliveira CM, Da Silva Cruz LA. Retinal image quality assessment using generic image quality indicators. *Inf Fusion* 2014; 19: 73–90.
- Rangayyan RM. *Biomedical image analysis*. 2005.
- Romero-Oraá R. Análisis automático de retinografías en Retinopatía Diabética: evaluación de la calidad y localización de las estructuras anatómicas del ojo mediante técnicas de procesamiento digital de imágenes. *Trab Fin Master* 2016
- Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007; 23: 2507–17.
- Saha SK, Fernando B, Cuadros J, Xiao D, Kanagasingam Y. Automated Quality Assessment of Colour Fundus Images for Diabetic Retinopathy Screening in Telemedicine. *J Digit Imaging* 2018: 1–23.
- Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015; 10: 1–21.
- Scanlon PH, Malhotra R, Greenwood RH, Aldington SJ, Foy C, Flatman M, *et al*. Comparison of two reference standards in validating two field mydriatic digital photography as a method of screening for diabetic retinopathy. *Br J Ophthalmol* 2003; 87: 1258–63.
- Sevik U, Köse C, Berber T, Erdöl H. Identification of suitable fundus images using automated quality assessment methods. *J Biomed Opt* 2014; 19: 046006.
- Shao F, Yang Y, Jiang Q, Jiang G, Ho YS. Automated quality assessment of fundus images via analysis of illumination, naturalness and structure. *IEEE Access* 2018; 6
- Soares JVB, Leandro JG, Cesar Jr RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *Med Imaging, IEEE Trans* 2006; 25: 1214–22.
- Tebar FJ, Escobar F. *La Diabetes Mellitus en la Práctica Clínica*. Madrid: Editorial Médica Panamericana S.A.; 2009.
- Thibodeau GA, Patton KT. *Enciclopedia de enfermería Mosby*. Tomo 1: Anatomía y fisiología. 4ª Edición. Madrid: Elsevier Iberoamericana; 2000.
- Toennies KD. *Guide to Medical Image Analysis: Methods and Algorithms*. 2012.
- Topcon. TRC-NW400 [Internet].
- Usher DB, Himaga M, Dumskyj MJ, Boyce JF. Automated assessment of digital fundus image

quality using detected vessel area. In: Proceedings of Medical Image Understanding and Analysis. 2003. p. 81–4.

Wang S, Jin K, Lu H, Cheng C, Ye J, Qian D. Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs. *IEEE Trans Med Imaging* 2016; 35: 1046–55.

Webb RH, Hughes GW, Delori FC. Confocal scanning laser ophthalmoscope. *Appl Opt* 1987; 26: 1492.

Welikala RA, Fraz MM, Foster PJ, Whincup PH, Rudnicka AR, Owen CG, *et al.* Automated retinal image quality assessment on the UK Biobank dataset for epidemiological studies. *Comput Biol Med* 2016; 71: 67–76.

Winder RJ, Morrow PJ, McRitchie IN, Bailie JR, Hart PM. Algorithms for digital image processing in diabetic retinopathy. *Comput Med Imaging Graph* 2009; 33: 608–22.

Wong T, Mitchell P. The eye in hypertension. *Lancet* 2007; 369: 425–35.

World Health Organization. WHO | Medical imaging [Internet]. 2017

Yannuzzi LA, Ober MD, Slakter JS, Spaide RF, Fisher YL, Flower RW, *et al.* Ophthalmic fundus imaging: Today and beyond. *Am J Ophthalmol* 2004; 137: 511–24.

Yau JWY, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, *et al.* Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012; 35: 556–64.

Yolcu U, Sahin OF, Gundogan FC. Imaging in Ophthalmology. In: *Ophthalmology - Current Clinical and Research Updates*. InTech; 2014.

Yu H, Agurto C, Barriga S, Nemeth SC, Soliz P, Zamora G. Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening. *2012 IEEE Southwest Symp Image Anal Interpret* 2012: 125–8.

Yu L, Liu H. Efficient Feature Selection via Analysis of Relevance and Redundancy. *J Mach Learn Res* 2004; 5: 1205–1224.

Zhou M, Jin K, Wang S, Ye J, Qian D. Color Retinal Image Enhancement Based on Luminosity and Contrast Adjustment. *IEEE Trans Biomed Eng* 2017; 65: 1–1.

Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin Chem* 1993; 39: 561–77.

## Listado de acrónimos

---

- **AZIFO:** Análisis Automático de Imágenes de Fondo de Ojo
- **AGGD:** *Asymmetric Generalized Gaussian Distribution* (Distribución Gaussiana Generalizada Asimétrica)
- **AUC:** *Area Under the ROC Curve* (Área bajo la curva ROC)
- **BD:** Base de Datos
- **CNN:** *Convolutional Neural Network* (Red Neuronal Convolutacional)
- **CPBD:** *Cumulative Probability of Blur Detection* (Probabilidad Acumulativa de Detección de Borrosidad)
- **CW:** *Cotton Wool spots* (Exudados Algodonosos)
- **CWT:** *Continuous Wavelet Transform* (Transformada *Wavelet* Continua)
- **DCT:** *Discrete Cosine Transform* (Transformada del Coseno Discreta)
- **DM:** Diabetes Mellitus
- **DMAE:** Degeneración Macular Asociada a la Edad
- **EX:** Exudados Duros
- **F1:** *F-Score* (Medida F)
- **FCBF:** *Fast Correlation-Based Filter* (Filtro Rápido Basado en Correlación)
- **FEDER:** Fondo Europeo de Desarrollo Regional
- **FID:** Federación Internacional de Diabetes
- **FN:** Falso Negativo
- **FOV:** *Field Of View* (Campo de Visión)
- **FP:** Falso Positivo
- **GGD:** *Generalized Gaussian Distribution* (Distribución Gaussiana Generalizada)
- **HCUV:** Hospital Clínico Universitario de Valladolid
- **HE:** Hemorragias intrarretinianas
- **HSV:** *Hue Saturation Value* (Matiz Saturación Valor)
- **HVS:** *Human Vision System* (Sistema Visual Humano)
- **IG:** *Information Gain* (Ganancia de Información)
- **IOBA:** Instituto Universitario de Oftalmobiología Aplicada
- **IPT:** *Image Processing Toolbox* (Utilidades de Procesado de Imagen)
- **ISC:** *Image Structure Clustering* (Agrupamiento de Estructuras de Imagen)
- **JNB:** *Just Noticeable Blur* (Borrosidad Detectable)
- **JPEG:** *Joint Photographic Experts Group*
- **LoG:** *Laplacian of Gaussian* (Laplaciano de Gaussiana)
- **MA:** Microaneurismas
- **MC:** Matriz de Confusión

- **MLP:** *Multilayer Perceptron* (Perceptrón Multicapa)
- **MVG:** *Multivariate Gaussian Model* (Modelo Gaussiano Multivariante)
- **NICE:** *National Institute for Health and Care Excellence*
- **NIQE:** *Natural Image Quality Evaluator* (Evaluador de calidad de imágenes naturales)
- **NR-IQA:** *No-Reference Image Quality Assessment* (Evaluación de Calidad de Imagen Sin Referencia)
- **NSS:** *Natural Scenes Statistics* (Estadísticas de Escenas Naturales)
- **OCT:** *Optical Coherence Tomography* (Tomografía de Coherencia Óptica)
- **RBF:** *Radial Basis Function* (Función de Base Radial)
- **RD:** Retinopatía Diabética
- **RDNP:** Retinopatía Diabética No Proliferativa
- **RDP:** Retinopatía Diabética Proliferativa
- **RGB:** *Red Green Blue* (Rojo Verde Azul)
- **RNA:** Red Neuronal Artificial
- **ROC:** *Receiver Operating Characteristic* (Característica Operativa del Receptor)
- **SLO:** *Scanning Laser Ophthalmoscope* (Oftalmoscopia de Barrido Láser)
- **SMOTE:** *Synthetic Minority Oversampling Technique* (Técnica de Sobremuestreo Sintético de Minoría)
- **SSEQ:** *Spatial and Spectral Entropy-based Quality* (Calidad basada en Entropía Espacial y Espectral)
- **SU:** *Symmetrical Uncertainty* (Incertidumbre Simétrica)
- **SVM:** *Support Vector Machine* (Máquina de Vectores Soporte)
- **TFM:** Trabajo Fin de Máster
- **UWT:** *Undecimated Wavelet Transform* (Transformada *Wavelet* sin Diezmado)
- **VN:** Verdadero Negativo
- **VP:** Verdadero Positivo
- **VPP:** Valor Predictivo Positivo