

UNIVERSIDAD DE VALLADOLID

FACULTAD DE MEDICINA

ESCUELA DE INGENIERÍAS INDUSTRIALES

TRABAJO DE FIN DE GRADO

GRADO EN INGENIERÍA BIOMÉDICA

Análisis de retinografías para la ayuda al diagnóstico de la Degeneración Macular Asociada a la Edad (DMAE).

Autor:

D. Guillermo Ordax Galindo

Tutores:

Dra. D^a. María García Gadañón

Valladolid, 29 de septiembre de 2023

TÍTULO:

Análisis de retinografías para la ayuda al diagnóstico de la Degeneración Macular Asociada a la Edad (DMAE)

AUTOR:

D. Guillermo Ordax Galindo

TUTORA:

Dra. D^a. 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. Carlos Gómez Peña

VOCAL:

Dra. D^a. María García Gadañón

SUPLENTE:

Dr. D. Jesús Poza Crespo

SUPLENTE:

Dr. D. Daniel Álvarez González

FECHA:

CALIFICACIÓN:

A mis padres

AGRADECIMIENTOS

Con estas líneas quiero expresar mi profundo agradecimiento a todos aquellos que con su ayuda han colaborado en la realización de este trabajo. En primera instancia, he de agradecer enormemente a todos los miembros del Grupo de Ingeniería Biomédica de la Universidad de Valladolid. En especial, quiero agradecer a Roberto Romero su apoyo, su dedicación, sus consejos y su amabilidad frente algunas dudas que me han ido surgiendo a la hora de realizar mi TFG. También, quiero agradecer a mi tutora María García, por siempre estar disponible para ayudarme en la realización de este trabajo.

A nivel personal, tampoco puedo olvidar la inestimable ayuda que me han prestado muchas personas. A mis amigos les agradezco su apoyo en los momentos de mayor presión y estrés. A mi familia, y en especial a mis padres, les agradezco de todo corazón su constante aliento, comprensión y amor incondicional. Gracias por creer en mí y estar a mi lado en cada paso que he dado.

Muchas gracias a todos.

RESUMEN

La Degeneración Macular Asociada a la edad (DMAE) representa el 8.7% de ceguera en todo el mundo, siendo la causa más común en los países desarrollados en individuos mayores de 60 años. De acuerdo con el informe global sobre la salud ocular, publicado por la Organización Mundial de la Salud en el año 2019, se estima que la prevalencia de la DMAE experimentará un aumento del 20% entre 2020 y 2030. El envejecimiento sistemático y acelerado de la población, junto con la estrecha relación existente entre la edad y la DMAE, augura que estas cifras seguirán creciendo con el paso del tiempo. Con el objetivo de evitar las complicaciones más graves de la DMAE, es sumamente importante abordarla en sus primeras etapas. Para este propósito, los exámenes oftalmológicos de *screening* constituyen una herramienta esencial. En estos exámenes se utilizan diversas técnicas para capturar imágenes de la retina de los pacientes, siendo la retinografía una de las más baratas y accesibles. Posteriormente, dichas imágenes son revisadas por parte de los especialistas en busca de signos de la enfermedad. Debido a su creciente incidencia, el número de imágenes que deberá revisarse será cada vez más elevado. Esto, unido a la escasez de oftalmólogos, hace que sea cada vez más complicado obtener un diagnóstico rápido y preciso de la enfermedad.

En este trabajo se propone un método automático basado en *Machine Learning* para la detección de la DMAE mediante el análisis de retinografías. Para ello se empleó la base de datos pública *Automatic Detection Challenge on Age-related Macular degeneration* (ADAM), que contiene 400 retinografías de pacientes sanos y con DMAE. Dicha base de datos se dividió en un conjunto de entrenamiento con 300 imágenes y un conjunto de test con 100 imágenes. El método desarrollado se dividió en diferentes etapas. En primer lugar, se llevó a cabo una etapa de preprocesado para mejorar el contraste de las imágenes mediante la aplicación del algoritmo *Contrast-Limited Histogram Equalization* (CLAHE). A continuación, se seleccionó una región de interés para enfocar las imágenes y eliminar elementos innecesarios como el fondo. Posteriormente, se llevó a cabo una fase de extracción de características de las imágenes. En ella, se emplearon características basadas en color, estadísticos texturales e histogramas de gradiente orientados. De entre todas las características extraídas, se seleccionaron aquellas más relevantes para la clasificación de la presencia o ausencia de DMAE empleando el algoritmo *Fast Correlation Based Filter* (FCBF). Finalmente, se llevó a cabo una etapa de clasificación en la que se entrenó un clasificador basado en *Ensemble Methods* formado por árboles de decisión.

En cuanto a los resultados sobre el conjunto de test, se alcanzó un 55.1% de sensibilidad, 91.5% de especificidad, 91.2% de precisión, 72.7% de valor predictivo positivo (VPP), 83.3% de valor predictivo negativo (VPN) y 62.5% en la métrica F1. Además, se obtuvo la curva ROC y se calculó el área bajo la curva (*Area Under ROC Curve*, AUC), que alcanzó un valor de 0.86. Finalmente se calculó la matriz de confusión, con el fin de dar una visión más detallada de los resultados del método propuesto.

Los resultados obtenidos indican que es posible llevar a cabo la detección de la DMAE de forma precisa mediante el análisis automático de retinografías. El método propuesto, por tanto, permitiría acortar el tiempo de diagnóstico, reducir la carga de trabajo de los expertos y, como consecuencia, disminuir los costes económicos asociados al tratamiento de la enfermedad.

PALABRAS CLAVE

Análisis de retinografías; Detección automática; DMAE; Extracción de características; *Machine Learning*; Procesado digital de imágenes; Retina.

ABSTRACT

Age-related macular degeneration (AMD) accounts for 8.7% of blindness worldwide, being the most common cause in developed countries in individuals over the age of 60. According to the global report on eye health, published by the World Health Organization in 2019, it is estimated that the prevalence of AMD will experience a 20% increase between 2020 and 2030. The systematic and accelerated aging of the population, coupled with the close relationship between age and AMD, predicts that these numbers will continue to grow over time. In order to avoid the most serious complications of AMD, it is extremely important to address it in its early stages. For this purpose, ophthalmologic screening tests are an essential tool. In these examinations, various techniques are used to capture images of the retina of patients, being retinography one of the cheapest and most accessible. These images are then reviewed by specialists for signs of the disease. Due to its increasing incidence, the number of images to be reviewed will continue to increase. This, linked with the shortage of ophthalmologists, makes it increasingly difficult to obtain a quick and accurate diagnosis of the disease.

In this work we propose an automatic method based on Machine Learning for the detection of AMD through the analysis of retinographies. For this purpose, the database Automatic Detection Challenge on Age-related Macular degeneration (ADAM), which contains 400 retinographies of healthy and AMD patients, was used. This database was divided into a training set with 300 images and a test set with 100 images. The developed method was separated into different stages. First, a preprocessing stage was carried out to improve the contrast of the images by applying the Contrast-Limited Histogram Equalization (CLAHE) algorithm. Next, a region of interest was selected to focus the images and eliminate unnecessary elements such as the background. Subsequently, a feature extraction phase was performed on the images. Here, color-based features, textural statistics and oriented gradient histograms were used. Among all the extracted features, the most relevant ones were selected for the classification of the presence or absence of AMD using the Fast Correlation Based Filter (FCBF) algorithm. Finally, a classification stage was carried out in which an Ensemble Methods-based classifier was trained based on a combination of decision tree classifiers.

Regarding the results on the test set, 55.1% sensitivity, 91.5% specificity, 91.2% accuracy, 72.7% positive predictive value (PPV), 83.3% negative predictive value (NPV) and 62.5% in the F1 metric were achieved. For this classifier, the ROC curve and the area under the curve (Area Under ROC Curve, AUC) were calculated, which reached a value of 0.86. Finally, to provide a more detailed view of the results the confusion matrix was determined.

The results obtained indicate that it is possible to accurately detect AMD by means of automatic analysis of retinographies. The proposed method, therefore, would shorten the diagnostic time, reduce the workload of experts and, consequently, reduce the economic costs associated with the treatment of the disease.

KEYWORDS

AMD; Blindness; Detection; Image processing; Machine Learning; Retina; Retinographies.

GLOSARIO DE TÉRMINOS Y ACRÓNIMOS

ACO-GA. Algoritmo Genético de optimización de colonias de hormigas (*Ant Colony Optimization*)

ADAM. *Automatic Detection Challenge on Age-related Macular degeneration*

ANOVA. Análisis de la varianza (*Analysis of Variance*)

AREDS. *Age-Related Eye Disease Study*

AUC. Área bajo la curva (*Area Under Roc Curve*)

BD. Base de datos.

CLAHE. *Contrast Limited Adaptive Histogram Equalization*

CNN. Red neuronal convolucional (*Convolutional Neural Network*)

DT. Árbol de decisión (*Decision Tree*)

DMAE. Degeneración Macular Asociada a la Edad.

DWT. transformada wavelet discreta (*Discrete Wavelet Transform*)

GLCM. Matriz de coocurrencia de niveles de gris (*Gray Level Co-occurrence Matrix*)

HOG. Histogramas de gradiente orientado (*Histograms of Oriented Gradients*)

ISBI. *Symposium on Biomedical Imaging*

KNN. Algoritmo de los k vecinos más cercanos (*K-Nearest Neighbors*)

LBP. Patrones binarios locales (*Local Binnary Patterns*)

LSDA. Análisis discriminante sensible a la localidad

OCT. Tomografía de coherencia óptica (*Optical Coherence Tomography*)

OMS. Organización mundial de la salud.

PNN. Red neuronal de probabilidad (*Probabilistic Neural Network*)

PSO. Optimización del enjambre de partículas (*Particle Swarm Optimization.*)

ReLU. Unidad Lineal Rectificada (*Rectified Linear Unit*)

REFUGE. *Retinal Fundus Glaucoma Challenge.*

RGB. Espacio de color Rojo-Verde-Azul (*Red-Green-Blue*)

ROC. Característica de operación del receptor (*Receiver Operating Characteristic*)

ROI. Región de interés (*Region of Interest*)

SVM. Máquina de vector de soporte (*Support Vector Machine*)

SMOTE. Técnica de Sobremuestreo de Minorías Sintéticas (*Synthetic Minority Oversampling Technique*)

SU. Incertidumbre simétrica (*Symetric Uncertainty*)

VEGF. Factor de crecimiento endotelial vascular (*Vascular Endothelial Growth Factor*)

ÍNDICE GENERAL

Agradecimientos.....	vii
Resumen	ix
Palabras clave	x
Abstract.....	xi
Keywords	xii
Glosario de términos y acrónimos.....	xiii
ÍNDICE DE FIGURAS.....	xviii
ÍNDICE DE TABLAS.....	xix
CAPÍTULO 1. INTRODUCCIÓN	20
1.1 INGENIERÍA BIOMÉDICA	20
1.2 IMAGEN MÉDICA.....	20
1.3 IMÁGENES DE RETINA. RETINOGRAFÍAS.....	21
1.4 DEGENERACIÓN MACULAR ASOCIADA A LA EDAD.....	26
1.4.1 PATOGÉNESIS.....	26
1.4.2 DIAGNÓSTICO	28
1.4.3 PREVALENCIA.....	29
1.5 HIPÓTESIS DE TRABAJO.....	30
1.6 OBJETIVOS DEL TFG.....	31
1.7 METODOLOGÍA EMPLEADA	31
1.8 ESTRUCTURA DEL TFG.....	32
CAPÍTULO 2. REVISIÓN DEL ESTADO DE LA TÉCNICA.....	34
2.1 INTRODUCCIÓN.....	34
2.2 MÉTODOS BASADOS EN <i>MACHINE LEARNING</i>	34
2.3 MÉTODOS BASADOS EN <i>DEEP LEARNING</i>	37
CAPÍTULO 3. MATERIALES Y MÉTODOS.....	39
3.1 INTRODUCCIÓN.....	39
3.2 BASE DE DATOS DE RETINOGRAFÍAS	39

3.3 PREPROCESADO DE RETINOGRAFÍAS.....	41
3.3.1 AUMENTO DEL CONTRASTE MEDIANTE CLAHE.....	41
3.2.1 EXTRACCIÓN DE ROI	42
3.4 EXTRACCIÓN DE CARACTERÍSTICAS	43
3.4.1 HISTOGRAMAS DE COLOR.....	43
3.4.2 ESTADÍSTICOS TEXTURALES	44
3.4.3 HISTOGRAMAS DE GRADIENTE ORIENTADOS	47
3.5 SELECCIÓN DE CARACTERÍSTICAS	49
3.6 <i>SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE</i>	49
3.7 CLASIFICACIÓN MEDIANTE <i>MACHINE LEARNING</i>	50
3.7.1 ENSEMBLE METHODS	51
CAPÍTULO 4. RESULTADOS.....	53
4.1 INTRODUCCIÓN.....	53
4.2 MODO DE EVALUACIÓN.....	53
4.2.1 MATRIZ DE CONFUSIÓN.....	53
4.2.2 PRECISIÓN	54
4.2.3 VALOR PREDICTIVO POSITIVO Y VALOR PREDICTIVO NEGATIVO	55
4.2.4 CURVA <i>RECEIVER OPERATING CHARACTERISTIC</i>	55
4.2.5 F1	56
4.3 RESULTADOS DE LA ETAPA DE ENTRENAMIENTO	56
4.3.1 CARACTERÍSTICAS SELECCIONADAS.....	57
4.3.2 ETAPA DE ENTRENAMIENTO.....	57
4.4 RESULTADOS SOBRE EL CONJUNTO DE TEST	59
CAPÍTULO 5. DISCUSIÓN	61
5.1 INTRODUCCIÓN.....	61
5.2 PREPROCESADO	61
5.3 EXTRACCIÓN DE CARACTERÍSTICAS	62
5.4 RESULTADOS DE LA FASE DE ENTRENAMIENTO	63
5.5 RESULTADOS SOBRE EL CONJUNTO DE TEST	64
5.5 COMPARACIÓN CON ESTUDIOS PREVIOS.....	65

CAPÍTULO 6. CONCLUSIÓN	67
6.1 INTRODUCCIÓN.....	67
6.2 CONTRIBUCIONES ORIGINALES.....	67
6.3 CONCLUSIONES.....	67
6.4 LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN	68
REFERENCIAS	70

ÍNDICE DE FIGURAS

Figura 1.1: Corte del globo ocular en el que se señalan las principales estructuras del ojo humano (Abramoff, Garvin, and Sonka 2010).	23
Figura 1.2: Retinografías; (a) DMAE con presencia de drusas duras; (b) DMAE con presencia de drusas blandas (Thomas, Mirza, and Gill 2021).	27
Figura 1.3: Evolución de la prevalencia de la DMAE en los diferentes rangos de edad (adaptado de (Organisation 2019))	30
Figura 1.4: Esquema de la metodología seguida en este trabajo	32
Figura 3.1: Ejemplo de obtención de ROI; (a) Imagen original; (b) ROI obtenido.....	43
Figura 3.2: Explicación funcionamiento HOG (Khalid et al. 2021).....	48
Figura 4.2: Curva ROC promedio obtenida sobre el conjunto de validación	59
Figura 4.3: Matriz de confusión promedio obtenida sobre el validación	59
Figura 4.4: Curva ROC obtenida sobre al conjunto de test	60
Figura 4.5: Matriz de confusión obtenida sobre el conjunto de test	60
Figura 5.1: Ejemplo ROIS; (a) Canal verde; (b) RGB.....	62

ÍNDICE DE TABLAS

Tabla 2.1: Descripción estudios Machine Learning presentes en la literatura.	36
Tabla 3.1: Descripción base de datos ADAM	40
Tabla 4.1: Distribución de las imágenes en los diferentes subconjuntos	41
Tabla 4.2: Características seleccionadas por el algoritmo FCBF	58
Tabla 4.3: Resultados conjuntos de validación	58
Tabla 5.1: Resultados conjuntos de otros autores y nuestro sistema	66

CAPÍTULO 1. INTRODUCCIÓN

En este capítulo se expondrán las principales ideas a desarrollar en el Trabajo de Fin de Grado. Se definirán los pilares básicos sobre los que se sustenta, tales como la Ingeniería Biomédica, la imagen médica, las retinografías y la Degeneración Macular Asociada a la Edad (DMAE). Asimismo, se expondrán las hipótesis del trabajo y los objetivos del TFG. Por último, se presentará la metodología empleada y la estructura del documento.

1.1 INGENIERÍA BIOMÉDICA

Se denomina Ingeniería Biomédica al campo científico encargado de integrar las cualidades propias de disciplinas como ingeniería, matemáticas, física e informática con el objetivo de, a través del avance y desarrollo en el estudio de la medicina y la biología, mejorar la calidad de vida y la salud de las personas (Houssein et al. 2019). Este ámbito de conocimiento busca construir un puente entre la ingeniería y la medicina mediante la combinación natural de características de ambas disciplinas. De este modo se consigue mejorar el tratamiento, y monitorización de enfermedades, reducir la brecha tecnológica en el sistema de salud y mejorar su rendimiento gracias al trabajo conjunto entre el profesionales de diferentes especialidades (M. Y. Khan, Gupta, and Verma 2018).

Hoy en día, los retos que se presentan en la medicina global son cada vez más complejos, los recursos son más limitados y la población se encuentra cada vez más envejecida (M. Y. Khan, Gupta, and Verma 2018). Por ello, es necesario llevar a cabo una modernización de los sistemas de salud actuales (Houssein et al. 2019). En esta tarea, el papel desempeñado por la figura del ingeniero biomédico es fundamental. Los profesionales con este perfil pueden aportar sus conocimientos para el desarrollo de algoritmos, sistemas y técnicas innovadoras que ayuden a mejorar la eficiencia de todos los procesos relacionados con el cuidado del paciente (Houssein et al. 2019).

1.2 IMAGEN MÉDICA

Se define el término imagen como una función bidimensional, donde, las variables x, y representan las coordenadas en el plano, mientras que $f(x, y)$ representa la intensidad o nivel de gris (Ponce-Rivera et al. 2017). Cada uno de los elementos que componen una imagen se denominan píxeles (Ponce-Rivera et al. 2017). Es importante señalar que, en las imágenes digitales, tanto las coordenadas como las intensidades de los píxeles van a tomar valores discretos y finitos (Ponce-Rivera et al. 2017).

La imagen médica es el conjunto de técnicas y procesos empleados con el objetivo de visualizar el interior del cuerpo humano. Las diferentes modalidades de imagen médica

permiten revelar, diagnosticar y examinar enfermedades. Asimismo, permiten estudiar la fisiología y anatomía del cuerpo humano (Ponce-Rivera et al. 2017). En la actualidad, se requiere una revisión manual de las imágenes en la mayoría de las aplicaciones médicas. Esta práctica requiere una importante inversión de tiempo y esfuerzo por parte de los profesionales (Ponce-Rivera et al. 2017). Dada la escasez de personal capacitado para analizar algunas modalidades de imágenes médicas, existe la necesidad de desarrollar e implementar métodos automatizados de ayuda al diagnóstico que sean capaces de aliviar la presión sobre los sistemas de salud (Dey, Ashour, and Borra 2018).

El enfoque principal del análisis y procesamiento de imágenes médicas es el estudio de propiedades como el tamaño, la forma, la textura o el color, con el fin de detectar lesiones visibles o regiones clínicamente relevantes. Entre los procedimientos más importantes en el procesado de imagen médica, se pueden destacar (Herrera-Romero et al. 2021):

- **Preprocesado.** La etapa de preprocesado se realiza con el fin de mejorar la calidad de la imagen, mejorando aspectos como el contraste o el equilibrio del espacio de color, entre otros.
- **Segmentación.** Definida como la división de una imagen en regiones u objetos completos.
- **Extracción de características.** Esta etapa tiene el propósito de extraer y describir las propiedades más importantes de los objetos o de la imagen.

En los últimos años, estos procedimientos se han convertido en una herramienta indispensable en el ámbito de la imagen médica. Su creciente importancia se debe a que permiten mejorar el diagnóstico precoz de enfermedades, monitorizar su progresión y evaluar la eficacia de los tratamientos (Herrera-Romero et al. 2021).

Al igual que en otras ramas de la medicina, los exámenes de diagnóstico por imágenes desempeñan un papel fundamental en el estudio de pacientes con enfermedades oftalmológicas (Behar-Cohen et al. 2020). En primer lugar, las técnicas de diagnóstico por imagen son un componente esencial en la identificación temprana de enfermedades oculares. Asimismo, ofrecen a los especialistas la posibilidad de obtener información detallada sobre la ubicación y extensión de las diferentes enfermedades. Para, de esta manera, servir como guía tanto en la aplicación de los diferentes tratamientos e intervenciones como en el seguimiento de la progresión de la enfermedad (Behar-Cohen et al. 2020). Dentro de la amplia variedad de modalidades de imágenes médicas enfocadas a la oftalmología, este TFG se centrará en la retinografía (Behar-Cohen et al. 2020).

1.3 IMÁGENES DE RETINA. RETINOGRAFÍAS

La retina es una capa delgada y transparente que recubre la parte posterior del ojo (Behar-Cohen et al. 2020). Se encuentra ubicada entre la coroides y el cuerpo vítreo y se extiende desde la ora serrata, por delante, hasta el nervio óptico. La retina es la encargada de transformar los estímulos visuales del medio externo en impulsos nerviosos que se transmitirán hacia la corteza cerebral a través del nervio óptico (Sakata et al. 2009). Por ello, la retina es considerada la capa neuroceptiva del ojo. De igual forma que otras partes

como la coroides o la esclera, esta tiene forma de una esfera abierta por delante, lo que permite distinguir dos diferentes caras, una externa y otra interna (Sakata et al. 2009). Desde un punto de vista histológico y microanatómico, la retina presenta diez diferentes capas diferenciadas, que a su vez se dividen en dos partes principales (Varón, Jaramillo, and Tello 2010). La retina neurosensorial es la capa más interna y está formada por nueve capas. Por otra parte, el epitelio pigmentario de la retina (EPR) es la capa más externa. Se encuentra ubicado entre la coroides y la retina neurosensorial y contribuye al correcto mantenimiento de las propiedades funcionales y metabólicas de la retina (Varón, Jaramillo, and Tello 2010). Tiene funciones como la fagocitosis de las membranas de los fotorreceptores o la absorción de luz que entra al ojo, reduciendo así la dispersión de esta (Varón, Jaramillo, and Tello 2010). A su vez, la membrana de Bruch, el coriocapilaris y la coroides subyacente también presentan funciones destacables. Por su parte, el coriocapilaris y la coroides contienen la red vascular que nutre las capas externas de la retina. La membrana de Bruch es responsable de mediar las interacciones entre el EPR y el coriocapilaris y juega un importante papel en el desarrollo de las lesiones de tipo neovascular en la DMAE (Mitchell et al. 2018).

Dentro de la retina neurosensorial se encuentra la capa de fotorreceptores. Esta capa contiene dos tipos de células altamente especializadas denominados conos y bastones (Varón, Jaramillo, and Tello 2010). Al entrar los rayos de luz al ojo, estos atraviesan las diversas capas de la retina sensorial hasta alcanzar los fotorreceptores, estimulándolos (Varón, Jaramillo, and Tello 2010). Los conos y bastones son los encargados en convertir la energía luminosa en potenciales de acción que serán posteriormente transmitidos a las neuronas integradoras y ganglionares, sucesivamente. Estas son las responsables de formar el nervio óptico, encargado de transmitir el estímulo al sistema nervioso central (Varón, Jaramillo, and Tello 2010). Los bastones son los encargados de la visión en la oscuridad y los tonos de grises, es decir, de la visión escotópica. Los conos, por el contrario, son los responsables de proporcionar la visión en alta definición y en colores (visión fototópica), cabe destacar que este tipo de fotorreceptores se encuentran mayoritariamente en la fóvea (Varón, Jaramillo, and Tello 2010).

En la Figura 1.1 se pueden observar las diferentes estructuras que forman el ojo humano. En el caso de la retina, si se realizase una división topográfica, esta quedaría dividida en diferentes regiones anatómicas (Behar-Cohen et al. 2020):

Mácula. Pequeña depresión en el centro de la retina, de unos 6 mm de diámetro, que constituye la región ubicada en el centro del eje visual. Al poseer una gran concentración de conos desempeña la función de proporcionar la mayor parte de la visión fototópica y la agudeza visual.

Fóvea. Situada en el centro de la mácula, se encuentra compuesta exclusivamente por conos en una densidad muy alta, por lo que es la zona de mayor agudeza visual de todo el globo ocular. Presenta una zona totalmente avascular denominada zona avascular foveal.

Papila óptica. Zona circular situada en el centro de la retina a través de la cual emergen los axones de las células ganglionares retinianas que conformarán el nervio óptico. Es una zona huérfana de fotorreceptores por lo que se le considera el “punto ciego” del ojo.

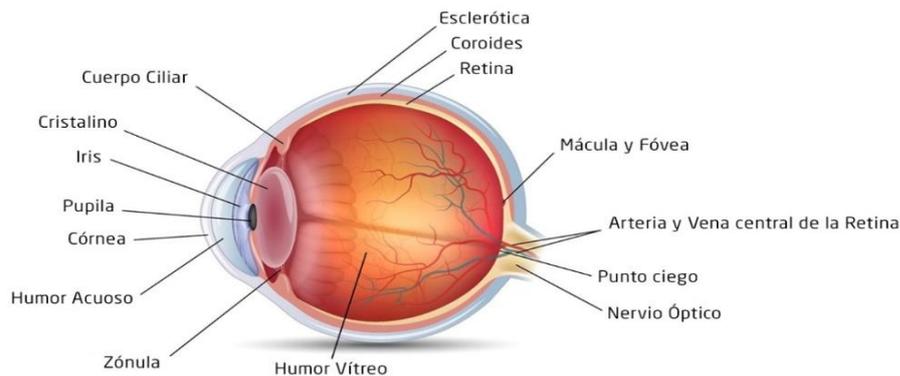


Figura 1. 1: Corte del globo ocular en el que se señalan las principales estructuras del ojo humano (Abramoff, Garvin, and Sonka 2010).

Debido a su destacada visibilidad en el examen de fondo de ojo, el estudio y análisis de dicha zona tiene una importancia clínica muy significativa.

Retina periférica. Contiene una mayor proporción de conos que de bastones, siendo la encargada de la visión periférica.

La retina desempeña un papel fundamental en la visión del mundo exterior. Por ello, es de suma importancia que sus estructuras sean ópticamente transparentes para posibilitar la formación de imágenes. Esta transparencia es fundamental ya que permite que, el tejido retiniano y, por tanto, el tejido cerebral sea visible desde el exterior (Abramoff, Garvin, and Sonka 2010). De esta forma, los oftalmólogos, a través de diferentes técnicas de imagen, pueden realizar una observación directa y determinar las diferentes lesiones producidas por las enfermedades infecciosas o autoinmunes en la retina. Entre las diferentes técnicas más empleadas para la obtención de imágenes retinianas se destacan las siguientes (Teus et al. 2007):

Oftalmoscopia. Esta es la técnica más sencilla para visualizar el fondo de ojo sin necesidad de obtener una imagen de él. Existen dos tipos de oftalmoscopios:

- El oftalmoscopio directo es un instrumento óptico que emite una luz directamente al fondo de ojo del paciente a través de un prisma. La porción de luz reflejada permitirá al profesional la inspección, desde un punto de vista amplificado, de las diferentes estructuras oculares del fondo de ojo. A través de este procedimiento, el profesional es capaz de examinar en profundidad el estado de la retina y detectar posibles enfermedades oculares. Este tipo de técnica es de gran utilidad debido a su bajo coste, facilidad de uso y capacidad de amplificación (Teus et al. 2007).
- El oftalmoscopio indirecto, por su parte, es un artefacto con binóculos incorporados que se coloca mediante un casco en la cabeza del

oftalmólogo. Dicho casco emite una fuente de luz que se dirige hacia una lente sujeta por el propio profesional proporcionando una visualización indirecta e invertida del ojo del paciente. Esta técnica es menos utilizada que el oftalmoscopio directo debido a la dificultad de aprendizaje por parte del especialista y su alto costo. Sin embargo, ofrece ventajas como la posibilidad de visualizar un mayor perímetro, permitiendo explorar zonas más alejadas del centro de la retina (Teus et al. 2007).

Angiografía fluoresceínica. Este método permite la evaluación de la salud vascular de la retina. Consiste en la aplicación de un tinte fluorescente, conocido como fluoresceína, el cual se inyecta por vía intravenosa y es distribuido por todo el cuerpo. En la región ocular, dicho tinte reacciona a luz azul de una longitud de onda específica (Nentwich 2015). De esta manera, los vasos sanguíneos se observarán brillantes y contrastados. Este procedimiento permite la evaluación de diferentes anomalías como la presencia de fugas vasculares, obstrucción de capilares, creación de áreas isquémicas y neovascularización. Además, proporciona información sobre el área en la que se ha producido fuga y ayuda a establecer la ubicación de zonas no perfundidas de la retina (Nentwich 2015).

Oftalmoscopia de laser de barrido. Se trata de una técnica desarrollada en la década de los 80 que permite la obtención de imágenes de alta resolución de retina sin dilatar la pupila. El funcionamiento de esta técnica se basa en proyectar puntos iluminados de forma secuencial sobre una pequeña área de la retina, de forma que simule una iluminación homogénea y continua (Webb, Hughes, and Delori 1987). Posteriormente, se capturan los reflejos generados por el láser en cada punto con el objetivo de crear una imagen detallada de la retina. Sus principales aplicaciones clínicas incluyen: estudio hemodinámicos retinianos, medición de la densidad en la retina y examinación de la topografía papilar (Webb, Hughes, and Delori 1987).

Tomografía de coherencia óptica (OCT). Este procedimiento no invasivo utiliza interferometría de baja coherencia para producir una imagen bidimensional de la dispersión óptica de las microestructuras internas del tejido retiniano (Huang et al. 1991). Esta técnica de imagen nos brinda información acerca del contorno y el grosor de las diferentes estructuras retinianas, las cuales pueden ser fundamentales para la creación de nuevos tratamientos. Además, los sistemas de OCT también pueden operar en múltiples longitudes de onda para medir otras propiedades espectrales (Huang et al. 1991).

Autofluorescencia. Técnica no invasiva que se aprovecha de la propiedad de autofluorescencia propia de la retina con el fin de detectar daños de tipo oxidativo. Dicha propiedad, se debe principalmente a la presencia de una molécula denominada lipofuscina. Estas moléculas se acumulan en el EPR como resultado de una degradación incompleta por parte de los segmentos exteriores de los fotorreceptores del ojo. El procedimiento que se lleva a cabo para conseguir imágenes se basa en la emisión de luz a una longitud de onda de entre 500 y 700 nm con el objetivo de obtener la, ya mencionada, propiedad de autofluorescencia. De este modo, se consigue revelar signos de daño

oxidativo permitiendo detectar cambios metabólicos a nivel del EPR (Schmitz-Valckenberget al. 2008). La autofluorescencia se presenta como una buena herramienta en la evaluación de daños causados por la DMAE, pudiendo llegar a convertirse en una técnica muy útil y complementaria debido al carácter poco invasivo que presenta (Schmitz-Valckenberg et al. 2008).

Retinografía. La retinografía, también denominada “imagen de fondo de ojo”, es un procedimiento seguro e indoloro basado en la captura de imágenes en color de la retina. Esta técnica requiere el empleo de cámaras especializadas que registren imágenes digitales detalladas del fondo de ojo. Esto es posible debido a las características del ojo humano, el cual funciona de manera similar a una cámara fotográfica. La transparencia de elementos como la córnea, el cristalino y el humor vítreo permite la visión directa de diferentes estructuras oculares del fondo de ojo como la retina central, la papila óptica, la mácula y los vasos sanguíneos retinianos (Bennett and Barry 2009). La captura de imágenes de fondo de ojo se lleva a cabo utilizando un retinógrafo o cámara especializada conectado a un ordenador que se encuentra equipado con un *software* específico para la visualización de imágenes (Bennett and Barry 2009). La fotografía especializada de fondo de ojo consiste en un microscopio modificado y conectado a una cámara *flash*, la cual proporciona una vista vertical y magnificada. Las cámaras clásicas, normalmente, permiten la visión en un ángulo de 30 grados, con un aumento de 2.5 veces el tamaño real. Por su parte, las cámaras de ángulo variable proporcionan capacidades de gran angular entre 45 y 60 grados. No obstante, en estas técnicas sería necesario iluminar un área más amplia de la retina, por lo que haría falta una mayor dilatación pupilar (Bennett and Barry 2009). También existe otra modalidad de retinografía denominada de campo amplio. Esta técnica revolucionaria consigue captar más de 200 grados de visión, lo que supone un 80% de la retina en una imagen de tipo panorámico, que, a mayores de las estructuras mencionadas anteriormente, consigue mostrar una imagen completa de la retina periférica (Kumar et al. 2021). Los retinógrafos se pueden dividir en dos categorías: midriáticos y no midriáticos. La midriasis es la respuesta natural del ojo que provoca el aumento en el diámetro de la pupila en condiciones de poca iluminación. El oftalmólogo induce la midriasis utilizando un colirio conocido como tropicamida (Tsikata et al. 2017). Dicho colirio presenta un retardo en su efecto de 15 minutos y provoca síntomas en el paciente como visión borrosa horas después de su aplicación. La retinografía es una prueba muy rápida de realizar en ambos procedimientos. La calidad de la imagen de fondo de ojo depende de varios factores (Tsikata et al. 2017). Por un lado, se encuentran las características propias del paciente, como el tamaño de la pupila y las opacidades en los medios oculares (Tsikata et al. 2017). Además, propiedades típicas de la cámara, como la resolución del sensor o la respuesta cromática también juegan un papel importante en la calidad de la imagen. Asimismo, el *software* de procesamiento y las técnicas a utilizar por el fotógrafo serán de gran importancia en el resultado final. Por lo tanto, todos estos factores han de tenerse en cuenta para obtener imágenes de alta calidad y precisión (Tsikata et al. 2017).

En los últimos años, métodos como la OCT han ganado una importancia muy significativa en la detección y clasificación de la DMAE. A pesar de ello, el examen mediante retinografías sigue representando una parte fundamental en este ámbito. El procedimiento de examen de retinografías correspondiente a la DMAE se basa en la

determinación manual del tamaño y la extensión de las drusas (van Grinsven et al. 2013). Esto permite a los oftalmólogos monitorizar la progresión de la enfermedad y, de este modo, ajustar el plan de tratamiento a realizar. Sin embargo, la clasificación realizada por parte de observadores humanos requiere mucho tiempo y está sujeta a variaciones entre profesionales. Además de la velocidad, objetividad y reproducibilidad, la implementación de un sistema automatizado de detección y cuantificación de drusas puede ser beneficioso de varias formas (van Grinsven et al. 2013). Por un lado, puede permitir la creación de programas de detección rentables para pacientes en riesgo, así como identificar y clasificar a los pacientes con DMAE en grandes estudios de cohortes. Además, las mediciones cuantitativas precisas pueden ser útiles en estudios clínicos para evaluar y monitorear la progresión del área de drusas (van Grinsven et al. 2013).

1.4 DEGENERACIÓN MACULAR ASOCIADA A LA EDAD

La DMAE es una enfermedad multifactorial desarrollada especialmente en etapas avanzadas de la vida. Se caracteriza por la formación de depósitos extracelulares, inflamación localizada y neurodegeneración de la mácula, causando la pérdida progresiva de la visión central (Mitchell et al. 2018). La DMAE se define por la presencia de cambios específicos en la macula, particularmente por la deposición de depósitos extracelulares focales conocidos como drusas (Thomas, Mirza, and Gill 2021). La acumulación de drusas es el hallazgo clínico más distintivo en la DMAE. En las primeras etapas, la enfermedad presenta diferentes signos clínicos como drusas y anomalías en el epitelio pigmentario de la retina (Mitchell et al. 2018). En sus fases más avanzadas, esta puede adoptar diferentes formas. Puede manifestarse en forma neovascular, que representa el 15 % de los casos y conlleva un pronóstico más desfavorable, o no neovascular, conocida como atrófica seca o no exudativa. La DMAE en estadios tardíos causa una pérdida de la agudeza visual central, lo que conduce a una discapacidad visual significativa y permanente (Mitchell et al. 2018).

1.4.1 PATOGÉNESIS

La pérdida de visión ocasionada por la DMAE se debe fundamentalmente al daño progresivo ocasionado a los fotorreceptores, que puede deberse a diferentes factores (Mitchell et al. 2018). Estos incluyen la disfunción y atrofia del EPR y la alteración en el transporte de oxígeno, nutrientes y metabolitos entre los vasos sanguíneos y las células externas de la retina (Fritsche et al. 2014).

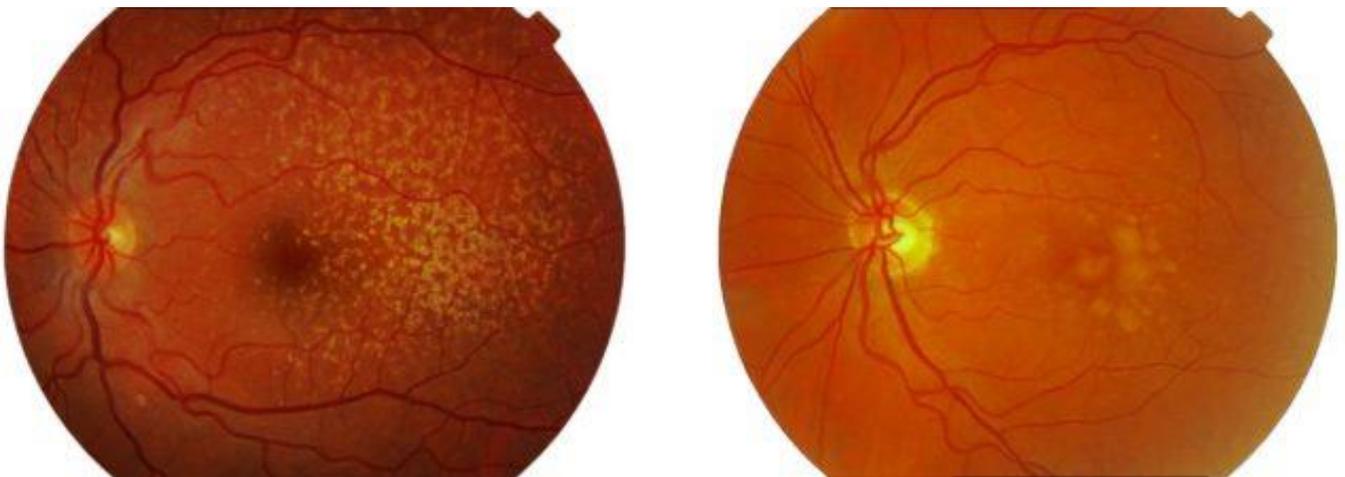
La DMAE se caracteriza por la presencia de drusas. Como ya se ha comentado, la aparición de estos depósitos puede considerarse el primer signo de la forma seca de la enfermedad (Mitchell et al. 2018). Las drusas son las lesiones distintivas de la DMAE pudiendo observarse tanto en la mácula como en la periferia retiniana. Este tipo de lesiones puede diferenciarse en duras y blandas. Las drusas duras presentan bordes discretos y generalmente son pequeñas, habitualmente son consideradas un signo de envejecimiento normal y no conllevan riesgos. A su vez, las drusas blandas son más grandes y presentan bordes menos definidos. Asimismo, pueden aparecer agrupadas en lesiones

de mayor tamaño (Thomas, Mirza, and Gill 2021). En la Figura 1.2 se puede observar un ejemplo de retinografía en la que aparecen drusas duras y blandas. En función de su tamaño estas lesiones pueden clasificarse de la siguiente manera (Thomas, Mirza, and Gill 2021):

- Drusas pequeñas: diámetro inferior a $63 \mu m$
- Drusas medianas: entre 63 y $124 \mu m$
- Drusas grandes: diámetro superior a $125 \mu m$

Si se definen desde un punto de vista histológico y microscopio, las drusas, especialmente las de gran tamaño, corresponden a depósitos lineales basales compuestas por material membranoso. Estas drusas se ubican entre el EPR y la capa interna de colágeno de la membrana de Bruch (Mitchell et al. 2018). Por su parte, los depósitos laminares basales, otro tipo de depósito relacionado con la DMAE, se localizan entre la membrana basal del EPR. En cuanto a la composición de las drusas, se destaca la presencia de lípidos neutros con colesterol esterificado y no esterificado, multitud de proteínas e iones de zinc y hierro (K. N. Khan et al. 2016; Mitchell et al. 2018). Existen diferentes niveles de severidad de la DMAE: DMAE no neovascular temprana, intermedia o avanzada y DMAE neovascular avanzada. La DMAE generalmente no presenta síntomas en etapas tempranas (Stahl 2020). En etapas intermedias, algunos pacientes pueden percibir una leve distorsión de la visión central, especialmente al leer o en condiciones de baja luminosidad. Por el contrario, en las etapas avanzadas de la enfermedad, la afectación de la visión central puede evolucionar más rápidamente, como en el caso de la DMAE neovascular, o más lentamente, como sucede en la DMAE no neovascular avanzada también denominada, forma atrófica (Mitchell et al. 2018). Se han identificado ciertas características morfológicas que permiten clasificar con mayor precisión la DMAE en función de la presencia o ausencia de proliferación anormal de vasos coroideos neovasculares (Mitchell et al. 2018).

- **DMAE no neovascular.** La DMAE “seca” se refiere tanto a la presencia de drusas en la enfermedad temprana o intermedia como a la atrofia geográfica. En el momento en el que se produce afectación de la fovea, la



(a)

Figura 1. 3: Retinografías; (a) DMAE con presencia de drusas duras; (b) DMAE con presencia de drusas blandas (Thomas, Mirza, and Gill 2021).

atrofia geográfica se identifica como DMAE seca avanzada pudiendo conducir a una pérdida progresiva de la visión. Los cambios atróficos se producen como resultado de la disminución de tejido retiniano externo y la red vascular circundante, incluyendo el EPR, la membrana de Bruch y el coriocapilaris (Thomas, Mirza, and Gill 2021).

- **DMAE neovascular.** La neovascularización que se desarrolla en esta forma de la enfermedad provoca la aparición de hemorragia retiniana, desprendimientos epiteliales de pigmento retiniano, y tejido cicatricial fibroso en las capas internas de la retina y el espacio subretiniano. Estos procesos conducen a la formación de fibrosis y, en ausencia de un tratamiento adecuado, a la pérdida permanente de la visión (Mitchell et al. 2018). La formación anormal de nuevos vasos sanguíneos, conocida como angiogénesis, desempeña un papel crucial en el desarrollo de la neovascularización coroidea. Dicha angiogénesis se ve mediada por el factor de crecimiento endotelial vascular (VEGF). El VEGF es un componente esencial para el correcto funcionamiento circulatorio de la retina. Sin embargo, la expresión anormal de ciertos subtipos específicos, inducida por condiciones de hipoxia, conducirá a la proliferación de nuevos vasos sanguíneos que son susceptibles a fugas y sangrado característicos (Mitchell et al. 2018).

1.4.2 DIAGNÓSTICO

De manera general, la DMAE en etapas tempranas es asintomática. Hasta las etapas intermedias no se empiezan a percibir síntomas como visión distorsionada o dificultad para leer (Stahl 2020). La tasa de progresión de DMAE intermedia a tardía en el curso natural de la enfermedad es del 28% en cinco años (Stahl 2020). Por ello, es de suma importancia tratar la enfermedad cuando aún se encuentra en las etapas prematuras. Para este fin, se recomienda que todos los pacientes que presenten sospecha se sometan a un examen oftalmológico para detectar si existe presencia de DMAE. En este examen, con el objetivo de realizar una exploración completa, el especialista evaluará la mácula en busca de signos como depósitos de drusas, cambios en la pigmentación, atrofia geográfica, hemorragia, formación de cicatrices y fibrosis (Thomas, Mirza, and Gill 2021). En este tipo de examen se presta especial atención al número y distribución de las drusas. Además, se realiza un examen completo para descartar cualquier tipo de enfermedad ocular coexistente. El estadiaje de la enfermedad se realiza, mayoritariamente, en base al examen realizado (Thomas, Mirza, and Gill 2021). En la actualidad existen diversos sistemas para clasificar la DMAE. Tradicionalmente, los estudios de población han dividido la DMAE en etapas tempranas y tardías. Por su parte, los estudios y ensayos clínicos, habitualmente, utilizan la escala de gravedad del Estudio de Enfermedades Oculares Relacionadas con la Edad (AREDS) (Ferris et al. 2005). Para emplear dicha escala, se asigna un factor de riesgo a cada ojo que presente drusas grandes, se asigna otro factor de riesgo a cada ojo con anomalías pigmentarias y se asigna otro factor de riesgo si ninguno de los ojos tiene drusas grandes, pero ambos ojos presentan drusas medianas (Ferris et al. 2005).

1.4.3 PREVALENCIA

La DMAE representa el 8.7 % de ceguera en todo el mundo, siendo la causa más común en los países desarrollados en individuos mayores de 60 años (Wong et al. 2014). Existe una gran probabilidad de que, a causa del envejecimiento acelerado de la población, su prevalencia aumente considerablemente en los próximos años (Wong et al. 2014). En los últimos tiempos, se han producido avances significativos en el tratamiento de la DMAE húmeda con la introducción de novedosos procedimientos como la terapia antiangiogénica. De este modo, se presenta al paciente una opción de cuidado efectiva que puede contribuir a prevenir la ceguera y, en algunos casos, incluso lograr la recuperación de la visión (Wong et al. 2014). No obstante, este tipo de tratamientos tiene un coste muy elevado y no se encuentra disponible en la gran mayoría de países. Como se puede observar, comprender la prevalencia y el impacto de la población resulta fundamental poder planificar y ofrecer una atención sanitaria adecuada (Wong et al. 2014). Para ello, se requiere contar con estimaciones precisas y actualizadas de la prevalencia de la enfermedad (Wong et al. 2014).

A pesar de que se han realizado multitud de estudios poblacionales sobre la DMAE en todo el mundo, todavía no existen datos concluyentes que permitan desarrollar estrategias globales (Wong et al. 2014). Los resultados alcanzados en varios estudios han sugerido que las diferencias raciales o étnicas tendrían influencia en cuanto a la prevalencia de la enfermedad (Wong et al. 2014). Según el *Baltimore Eye Study* las personas de ascendencia europea presentaban más probabilidades de desarrollar la enfermedad tanto en etapa temprana como tardía que las personas de descendencia africana (Klein et al. 2006; Wong et al. 2014). A su vez, se sugirió que la prevalencia específica de presentar DMAE en edad tardía en las personas de ascendencia asiáticas era similar a las europeas (Mitchell et al. 2018).

Aunque la ceguera es la consecuencia más grave de la DMAE, no es la única. Las personas mayores con una visión disminuida se enfrentan un mayor riesgo de sufrir caídas, depresión y dependencia a largo plazo (Mitchell et al. 2018). Las afecciones visuales asociadas a la DMAE pueden agravar las preocupaciones de salud en personas mayores, afectando significativamente a su calidad de vida. Por ello, es importante conocer los signos y síntomas de la DMAE para guiar a los pacientes en su búsqueda de atención oftalmológica y prevenir, de esta forma, una mayor pérdida de visión (Mitchell et al. 2018). Un aspecto vital en el manejo esta enfermedad es trabajar en reducir la exposición a factores de riesgo. La edad es el factor de riesgo más relevante en el desarrollo de la DMAE (Mitchell et al. 2018). Sin embargo, existen otros factores como el tabaquismo y la alimentación, que desempeñan un papel importante. Se ha demostrado que fumar es el factor riesgo modificable más fuerte, conllevando un riesgo asociado hasta dos veces mayor de desarrollar DMAE tardía, y adelantando el inicio de la enfermedad en aproximadamente 10 años (Mitchell et al. 2018). Otros factores, como la exposición a la luz solar, el consumo de alcohol y el color del iris tienen una evidencia menos sólida sobre su influencia en el desarrollo de DMAE (Thomas, Mirza, and Gill 2021). En la Figura 1.3 se puede observar como la prevalencia de la DMAE aumenta exponencialmente con la edad.

De acuerdo con el informe global sobre la salud ocular publicado por la Organización Mundial de la Salud (OMS), aproximadamente 2200 millones de personas en el mundo padecen alguna forma de discapacidad visual. De este grupo, alrededor de 1000 millones presentan problemas de visión que podrían haberse evitado o no se han tratado (Organisation 2019). Dicho informe enumera 7 causas principales de discapacidad visual, una de las cuales es la DMAE. Según dicho informe, se estima que la prevalencia de la DMAE experimentará un aumento del 20% entre 2020 y 2030. Esto significa que el número de personas afectadas pasará de 196 millones aproximadamente a 243.3 millones para el año 2030 (Organisation 2019). El envejecimiento sistemático de la población y la estrecha relación entre la edad y la enfermedad en sí misma, augura que estos datos no harían sino empeorar con el paso del tiempo. Varios estudios indican que, para la fecha de 2040, se espera que la cifra de personas afectadas por la DMAE alcance los 288 millones (Wong et al. 2014).

1.5 HIPÓTESIS DE TRABAJO

A causa de la creciente incidencia de la DMAE, el número de imágenes que deberán ser revisadas por los especialistas aumentará exponencialmente. Por ello, será cada vez más complicado obtener un diagnóstico rápido y eficaz sobre la enfermedad. En los últimos años se han ido adoptando medidas para paliar estas dificultades. Entre ellas destacan el empleo de telesalud y el uso de aplicaciones *software* móviles para evaluar la visión (Organisation 2019). Como se ha observado, la presencia de drusas en la retina es considerado un indicador temprano fundamental. El tamaño, la cantidad y la forma de estas se correlaciona de manera significativa con el riesgo de desarrollar DMAE. El etiquetado manual de estas drusas requiere de una gran cantidad de trabajo manual por

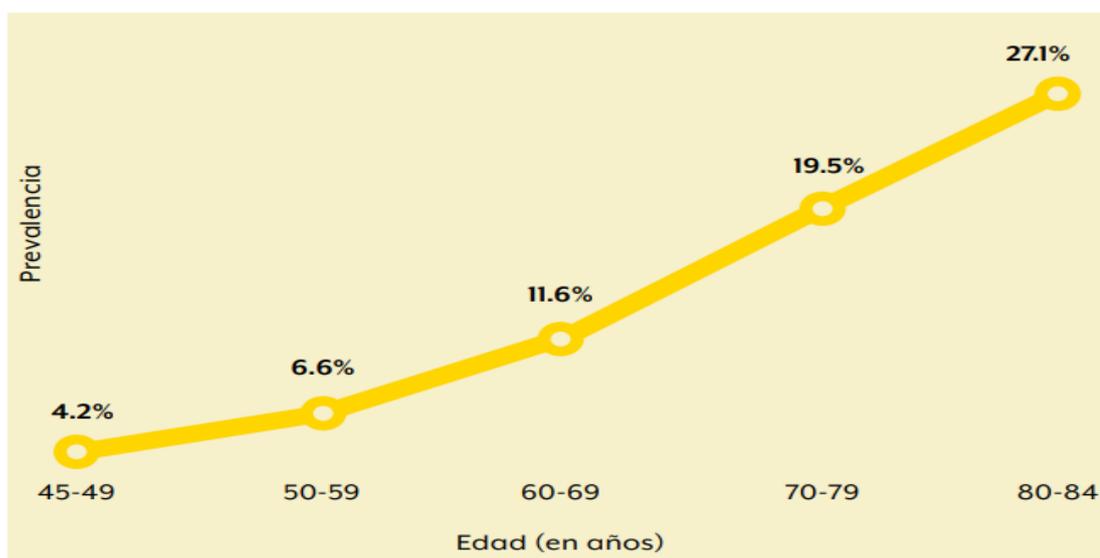


Figura 1. 4: Evolución de la prevalencia de la DMAE en los diferentes rangos de edad (adaptado de (Organisation 2019))

parte de los profesionales. En este sentido, la detección automática mediante métodos computarizados sería vital para mejorar la atención al paciente, ya que permitiría agilizar el proceso y reducir la carga de trabajo. De esta manera, se plantea la hipótesis de que una herramienta de detección automática de DMAE podría facilitar su proceso de diagnóstico, siendo de gran utilidad para su combinación con las técnicas tradicionales (Pead et al. 2019a).

Mediante la realización de este trabajo se busca desarrollar un sistema de detección automatizada de la DMAE. Para ello, se han estudiado diferentes modelos de *Machine Learning* con la finalidad de obtener un método preciso y eficaz que sea capaz de funcionar correctamente para una amplia variedad de imágenes.

1.6 OBJETIVOS DEL TFG

El objetivo principal de este trabajo ha sido el diseño, desarrollo y validación de un sistema de detección automática de la presencia de DMAE basado en *Machine Learning*. Para conseguir dicha meta, se plantearon los siguientes objetivos específicos.

- 1) Estudio de la literatura existente acerca del análisis de retinografías con el objetivo de comprender los principales conceptos necesarios para realizar este trabajo y obtener documentación sobre la variedad de técnicas empleadas. Dichos conceptos engloban desde la descripción patológica de la DMAE hasta los criterios de selección del modelo de *Machine Learning* empleado para llevar a cabo la detección de la DMAE. A partir de esta revisión, se han establecido las bases sobre las cuales se fundamenta el algoritmo desarrollado.
- 2) Análisis de la base de datos *Automatic Detection Challenge on Age-related Macular degeneration* (ADAM), con el propósito de inspeccionar los diferentes aspectos que ofrece y llevar a cabo una primera toma de contacto con las imágenes a examinar (Fang et al. 2022a).
- 3) Familiarización con el *software* MATLAB®.
- 4) Desarrollo de los métodos de procesado basados en la utilización del *toolbox* de MATLAB®: “*Image proccessing*”
- 5) Desarrollo del modelo de *Machine Learning* mediante la utilización de la aplicación de MATLAB®: “*Classification Learner*”
- 6) Procesado de las retinografías de la base de datos ADAM para la obtención de los resultados del estudio.
- 7) Análisis de los resultados obtenidos.
- 8) Extracción de conclusiones acerca del trabajo realizado.

1.7 METODOLOGÍA EMPLEADA

Para lograr los objetivos establecidos en este TFG, se ha procedido en base a la metodología de investigación presentada en la Figura (1.4). En este esquema de bloques se muestran las diferentes etapas del trabajo, las cuales se describen a continuación.

- 1) Búsqueda de la información para comprender el problema a tratar. Fundamentalmente se revisaron libros, artículos científicos y revistas médicas. Los buscadores principales en los que se basó la tarea de búsqueda fueron tanto PubMed como Google Scholar.
- 2) Familiarización con el *software* MATLAB® y sus diferentes funciones y aplicaciones.
- 3) Búsqueda de métodos de preprocesado y procesado de retinografías orientados al desarrollo del sistema propuesto.
- 4) Implementación del método seleccionado. Para llevar a cabo esta tarea, se utilizaron funcionalidades de MATLAB®.
- 5) Aplicación del método desarrollado sobre las imágenes de la base de datos ADAM.
- 6) Análisis de resultados y extracción de conclusiones.

1.8 ESTRUCTURA DEL TFG

En el siguiente apartado se describe la organización de esta memoria. En este primer capítulo introductorio se ha abordado el problema a resolver en este TFG, junto con sus respectivos objetivos, la hipótesis de trabajo y la metodología aplicada. Los capítulos restantes se organizaron de la siguiente manera:

- En el capítulo 2, “Revisión del estado de la técnica”, se presenta descripción de diversos estudios previos dedicados a la detección de la DMAE en retinografías. Se hace hincapié en los procedimientos que emplean *Machine Learning*.
- En el capítulo 3, “Materiales y métodos”, se presenta la base de datos utilizada en este TFG, así como los métodos empleados.

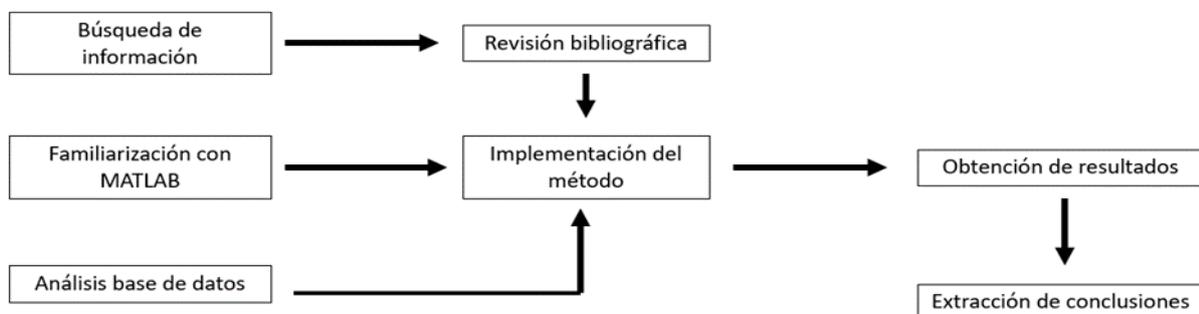


Figura 1.5: Esquema de la metodología seguida en este trabajo

Capítulo 1. introducción

- En el capítulo 4, “Resultados”, se muestran los resultados del modelo desarrollado aplicado a la base de datos ADAM.
- En el capítulo 5, “Discusión”, se presenta una interpretación de los resultados obtenidos en el TFG. Además, se comparan estos con los obtenidos en estudios previos.
- En el capítulo 6, “Conclusiones y líneas futuras”, se recopilan las conclusiones más destacadas derivadas de este estudio. De la misma manera, se describen en detalle las contribuciones del trabajo realizado y se sugieren posibles direcciones para futuras investigaciones.

CAPÍTULO 2. REVISIÓN DEL ESTADO DE LA TÉCNICA

2.1 INTRODUCCIÓN

El creciente aumento en la prevalencia de enfermedades oculares relacionadas con el envejecimiento plantea desafíos cada vez mayores para los profesionales de la salud (Pead et al. 2019b). En los últimos años, se han dedicado numerosos esfuerzos al desarrollo de métodos de análisis automatizado de imágenes retinianas para la ayuda al diagnóstico (García-Florian et al. 2019). En el caso específico de la DMAE, la presencia de drusas en la retina se considera un indicador cuantitativo que informa acerca del riesgo de progresión de la enfermedad. El proceso de etiquetado manual de drusas en retinografías requiere mucho tiempo y esfuerzo por parte de los especialistas. Es en este punto donde los métodos de detección automática pueden ofrecer una valiosa ayuda en la atención al paciente, reduciendo la carga de trabajo y el tiempo de diagnóstico (Pead et al. 2019b).

En este capítulo se presenta un análisis de las investigaciones más significativas llevadas a cabo en el campo de la detección y clasificación de DMAE a partir de retinografías. Para ello se revisarán tanto modelos basados en *Machine Learning* como aquellos fundamentados en el *Deep Learning*.

2.2 MÉTODOS BASADOS EN *MACHINE LEARNING*

El *Machine Learning* es un enfoque de inteligencia artificial parcialmente inspirado en como aprenden los humanos. El aprendizaje sucede con el ejemplo (Pead et al. 2019b). De esta manera, los algoritmos de *Machine Learning* emplean características extraídas de ejemplos de entrenamiento para descubrir o validar patrones que permitan clasificar diferentes subconjuntos (Pead et al. 2019b).

En el contexto diagnóstico de la DMAE, estos métodos se basan en detectar diferentes lesiones correspondientes a la enfermedad (D^a. María Herrero Tudela. 2021.). Como se mencionó previamente, la DMAE está estrechamente vinculada con la acumulación de depósitos focales de material extracelular, conocidos como drusas. A través de la obtención de diferentes características mediante técnicas de procesamiento digital de imagen y el uso de clasificadores basados en *Machine Learning* es posible la detección de la enfermedad, e incluso, la estimación de su gravedad (Acharya et al. 2017). En la literatura se han propuesto diversos métodos para la detección automatizada de la DMAE (Tabla 2.1). Estos estudios se agruparán en función de las características escogidas para llevar a

cabo su sistema de detección. Por otro lado, es preciso destacar que, en este trabajo, se prestara especial atención a los estudios que busquen detectar la enfermedad sin realizar ningún tipo de estadificación de esta. En la Tabla 2.1 se muestran, detalladamente, los diferentes procedimientos empleados en la literatura para llevar a cabo la detección y/o clasificación de la DMAE en sus diferentes estadios.

Diversos estudios centraron sus esfuerzos en la extracción de características no lineales, entropías y espectros de orden superior. Mookiah et al, en su investigación, extrajeron diferentes características como la dimensión fractal o la entropía. Dichas características se evaluaron mediante un proceso de ranking utilizando diferentes métodos. Posteriormente, aquellas que fueron seleccionadas, se sometieron a una fase de clasificación empleando clasificadores como *k*-Nearest Neighbors (*k*-NN) o *Support Vector Machine* (SVM) (Mookiah, Acharya, Koh, Chandran, et al. 2014). Por otro lado, otros autores propusieron un algoritmo para la clasificación de DMAE en sus variantes seca, húmeda o sana (Acharya et al. 2017). De manera inicial, estos aplicaron la pirámide de histogramas de gradiente orientados sobre las imágenes. Esta técnica permitió capturar la forma y los patrones espaciales presentes en las retinografías. A partir de esta representación, se extrajeron diferentes características lineales, energéticas y no lineales. La etapa de selección de características se llevó a cabo a través de dos diferentes técnicas basadas en inteligencia de enjambre: algoritmo genético de optimización de colonias de hormigas (ACO-GA) y *Particle Swarm Optimization* (PSO). Finalmente, las características seleccionadas se clasificaron mediante SVM uno contra todos.

Algunos autores presentaron un enfoque interesante para la detección de drusas. Dicho enfoque buscó combinar técnicas básicas de procesamiento digital y morfológico de imágenes para distinguir entre sujetos enfermos y sanos (García-Florianio et al. 2019). La metodología propuesta consta de dos fases principales: procesamiento de imágenes y clasificación de patrones (García-Florianio et al. 2019). En la etapa de procesamiento se aplicaron un conjunto de operaciones estándar y técnicas de morfología matemática con el objetivo de resaltar las drusas y la región macular. Posteriormente, se generó un patrón derivado de los momentos de Hu (García-Florianio et al. 2019). En la segunda etapa, se presentó el patrón generado a una SVM. La SVM clasificó el patrón como perteneciente a una clase u otra: positivo para drusas y, por lo tanto, DMAE; o negativo (García-Florianio et al. 2019).

Por su parte, estudios como el realizado por Phan et al optaron por la caracterización de la DMAE en función de atributos clásicos como el color, la textura o el contexto visual en las imágenes de fondo de ojo (Phan et al. 2016). Con respecto al color, se propuso la extracción de los histogramas propios del espacio RGB y $L^*a^*b^*$. Para las características texturales, se utilizaron *Local Binary Patterns* (LBP). Por último, para obtener información acerca de las relaciones espaciales entre los píxeles se introdujeron los *Histogram of oriented gradients* (HOG). Todas estas características se sometieron a un proceso de selección y, finalmente, a una etapa de clasificación a través de SVM. Otros estudios, como el presentado por Khalid et al, buscaron replicar esta fórmula para categorizar la mácula como enferma o sana (Khalid, Akram, and Khalil 2017). En este caso, se extrajeron características híbridas que, al igual que en el estudio previo, se basan en los patrones de textura color y forma. A diferencia del estudio anterior (Khalid, Akram, and Khalil 2017) utilizaron clasificadores SVM, *k*-NN y redes neuronales.

Autor/res, año	Preprocesamiento	¿Segmentación de drusas?	Extracción de características	Método de clasificación/detección
(García-Florianio et al. 2019)	Se llevo a cabo un escalado de imágenes, seguido de la aplicación de CLAHE y la selección del canal verde	No existe segmentación de drusas.	Método de reconocimiento de patrones a partir de los 7 momentos de Hu.	Detección de DMAE mediante SVM
(Acharya et al. 2017)	Se llevo a cabo la aplicación de CLAHE	No existe segmentación de drusas.	Cálculo de características lineales, energéticas y de entropía a partir de las pirámides de histograma de gradiente orientados.	SVM uno contra todos con kernel RBF para clasificación automatizada de tres clases de DMAE.
(Acharya et al. 2016)	Se llevo a cabo la aplicación de CLAHE	No existe segmentación de drusas.	Cálculo de características mediante la descomposición en tres niveles de transformación mediante la wavelet discreta.	Diferentes métodos de detección: k-NN, PNN, SVM y arboles de decisión
(Phan et al. 2016)	Se llevo a cabo la normalización de la imagen, seguido de la selección de ROI y la aplicación de un filtro de mediana	No existe segmentación de drusas.	Extracción de diferentes características como histogramas de color para cada canal, LBP y HOG	Clasificación mediante SVM uno contra todos
(Mookiah, Acharya, Koh, Chua, et al. 2014)	Se llevo a cabo la aplicación de CLAHE	No existe segmentación de drusas.	Descomposición wavelet de tres niveles.	Detección mediante SVM con diferentes núcleos
(Mookiah, Acharya, Koh, Chandran, et al. 2014)	Se llevo a cabo la aplicación de CLAHE sobre el canal verde	No existe segmentación de drusas.	Extracción de características como 1dimensión fractal, wavelet de gabor, y espectros de orden superior.	Clasificadores para la detección entre los que destacamos: SVM, DT, PNN, k-NN
(Khalid et al. 2021)	Eliminación de parte ruidosa de la imagen	No existe segmentación de drusas.	Extracción de diferentes características como LBP, HOG e histogramas de intensidad	Clasificadores para la detección: SVM, DT, KNN

Tabla 2. 1: Descripción estudios Machine Learning presentes en la literatura.

transformada *wavelet* discreta (DWT) con el objetivo de extraer características de las retinografías. En el artículo (Acharya et al. 2016) se presentó una propuesta, que combina la transformada de radón y la DWT. En primera instancia, las imágenes se sometieron a la transformada de radón seguida de la DWT. Esta se encargó de generar coeficientes aproximados a partir de la descomposición de las proyecciones de radón. Dichos coeficientes, se concatenaron y utilizaron como características. Finalmente, se evaluó el rendimiento de diversos clasificadores supervisados como SVM, *Probabilistic Neural Network* (PNN), *k*-NN y árboles de decisión (Acharya et al. 2016). Por otra parte, en el estudio (Mookiah, Acharya, Koh, Chua, et al. 2014) se planteó un sistema automatizado para detectar DMAE mediante el uso de DWT. La metodología empleada consistía en aplicar la DWT a las imágenes para, seguidamente, extraer características como varianza, media, asimetría, curtosis, energía y entropía. A continuación, se evaluaba el rendimiento del conjunto de características designado por medio de clasificadores supervisados como SVM o *k*-NN (Mookiah, Acharya, Koh, Chua, et al. 2014).

2.3 MÉTODOS BASADOS EN *DEEP LEARNING*

El *Deep Learning* es un subcampo de la inteligencia artificial basado en el uso de redes neuronales artificiales profundas. Estas redes están compuestas por múltiples capas ocultas que adoptan una estructura jerárquica. Dentro de la amplia variedad de arquitecturas de *Deep Learning*, las redes neuronales convolucionales (CNN) destacan por su alto rendimiento en el análisis de imágenes (Dong et al. 2021). Este tipo de redes cuenta con capas especiales, basadas en la aplicación de un procedimiento matemático de filtrado conocido como convolución (Dong et al. 2021). Estos filtros funcionan como detectores de características especiales. A medida que la imagen de entrada se procesa en sucesivas capas convolucionales de la red, los filtros se irán acumulando y generando progresivamente detectores de características más descriptivos y sofisticados (Schmidt-Erfurth et al. 2018). La arquitectura CNN se compone de convolución, *pooling* y capa totalmente conectada (Tan et al. 2018). El propósito principal de la operación de convolución es, como ya se ha comentado, recoger características de las imágenes de entrada. La operación de agrupación tiene como objetivo disminuir la dimensionalidad de la salida y asegurar un tamaño constante. Por último, la capa totalmente conectada se encarga de predecir la imagen de entrada en clases normales o enfermas (Tan et al. 2018). Una de las grandes ventajas que presentan estos sistemas es que no es necesaria la extracción manual de características por parte de expertos. Sino que, la propia red realiza el análisis y extracción de características que permite distinguir las clases del problema (D^a. María Herrero Tudela. 2021.).

El enfoque más empleado en la literatura es la utilización de CNN para la detección y/o clasificación de DMAE. En algunos estudios se empleó un modelo basado en CNN con el objetivo de clasificar las imágenes de fondo de ojo en dos clases: normal y DMAE (Tan et al. 2018). Dicho modelo, cuenta con siete capas de convolución, cuatro capas de agrupación y tres capas totalmente conectadas. El entrenamiento se realizó mediante *backpropagation* con un tamaño de lote de 50 (Tan et al. 2018). Asimismo, el modelo desarrollado por Zapata planteó una clasificación binaria de la DMAE a partir de una nueva arquitectura denominada AMDNET (Zapata et al. 2020). Esta arquitectura permite extraer

detalles y características precisas, al mismo tiempo que se realizan predicciones con una cantidad pequeña de parámetros entrenables en comparación con otras arquitecturas de *Deep Learning* como Inception-V3 (Zapata et al. 2020).

Del mismo modo, el estudio de Keel et al incorporó tres modelos de *Deep Learning*, todos ellos derivados de la arquitectura Inception-V3. Estos abordaron tareas para la clasificación de DMAE en sus etapas avanzadas, evaluación de la calidad de imagen y evaluación de la disponibilidad de campo (Keel et al. 2019). Otros estudios buscaron no sólo detectar la enfermedad, sino también clasificar la DMAE en diferentes estadios mediante el uso de técnicas de *Deep Learning* junto con técnicas asociadas, como *transfer learning* y aprendizaje de conjuntos (Govindaiah, Smith, and Bhuiyan 2018). Para ello, se implementó la técnica de aprendizaje por conjuntos a partir de dos redes neuronales profundas Inception-Resnet-V2 y Xception (Govindaiah, Smith, and Bhuiyan 2018).

Xie et al, en su estudio, aplicaron un método de estadiaje de la DMAE basado en la identificación de los tres componentes principales que determinan la gravedad de la enfermedad: DMAE avanzada, drusas, y anomalías pigmentarias. El objetivo principal era automatizar la predicción del riesgo de que la enfermedad se desarrollase (Xie et al. 2023). Para llevar a cabo este cometido, se entrenaron tres CNN individuales. Cada una de ellas, se creó mediante la estructura EfficientNetB0 y se organizaron en una arquitectura “en cascada” (Xie et al. 2023).

A pesar de que las arquitecturas basadas en *Deep Learning* son muy precisas, es necesario destacar sus múltiples inconveniente y limitaciones. El más importante es su falta de transparencia e interpretabilidad (Romero-Oraá et al. 2020). Asimismo, el entrenamiento de este tipo de redes implica la optimización y ajuste de una gran cantidad de hiperparámetros, como la regularización o la tasa de aprendizaje inicial. Encontrar los valores ideales de estos hiperparámetros es una tarea desafiante y crítica. Además, la configuración de la cantidad de capas de convolución y *pooling*, y el tamaño y número de filtros en cada capa, suele determinarse empíricamente (Al-Bander et al. 2018). En consecuencia, tanto la arquitectura óptima de la red como la configuración adecuada de los hiperparámetros son decisiones basadas en la experiencia, lo que dificulta su identificación por parte de personas no expertas en el campo. Finalmente, es preciso destacar que los métodos basados en *Deep Learning* requieren una cantidad de datos de entrenamiento, que no siempre se encuentran disponibles (Al-Bander et al. 2018) (Romero-Oraá et al. 2020). Debido a todo ello, nuestro estudio se basará en el desarrollo de un algoritmo para la detección automática de DMAE mediante *Machine Learning*.

CAPÍTULO 3. MATERIALES Y MÉTODOS

3.1 INTRODUCCIÓN

El análisis automático de retinografías para la detección de DMAE puede ser de gran utilidad en el diagnóstico temprano de esta enfermedad. Por ello, en los últimos años, se han desarrollado numerosos estudios con el fin de detectar la DMAE en imágenes de fondo de ojo. Entre ellos, destaca la gran variedad de técnicas empleadas para llevar a cabo este cometido

En este capítulo se expondrán tanto las imágenes empleadas en esta investigación como el algoritmo propuesto para la detección automática de DMAE en retinografías. En primer lugar, se describirá detalladamente la base de datos utilizada en este estudio. Posteriormente, tanto la fase de preprocesado como la de extracción y selección de características serán explicadas. Finalmente, se abordará en profundidad el clasificador empleado en la investigación.

3.2 BASE DE DATOS DE RETINOGRAFÍAS

La base de datos utilizada en este trabajo corresponde al desafío ADAM celebrado en el año 2020. Este desafío tuvo lugar en el seno del *Symposium on Biomedical Imaging (ISBI)* celebrado en Iowa (USA), vía conferencia (Fang et al. 2022b). Según el estudio AREDS, el diagnóstico de DMAE implica no solo detectar lesiones relacionadas con la enfermedad, sino analizar su tamaño y ubicación para determinar su gravedad (Fang et al. 2022b). Por ello, el objetivo clínico de este reto no fue únicamente la detección de la presencia o ausencia de DMAE, sino que también se propuso el reconocimiento de estructuras como la papila y la fovea y la segmentación de lesiones (Fang et al. 2022b).

El conjunto de datos ADAM contiene 1200 imágenes etiquetadas de fondo de ojo en color, almacenadas en formato JPEG con una profundidad de 8 bits. Estas imágenes se capturaron en el centro oftalmológico Zhongshan de la Universidad Sun Yat-sen, China (Fang et al. 2022b). La adquisición se realizó mediante el uso de dos retinógrafos distintos. Por un lado, se utilizó el retinógrafo Zeiss Visucam 500 para capturar 824 imágenes con una resolución de 2124×2056 píxeles. Por su parte, con el retinógrafo Canon CR-2 se realizó una captura de 376 imágenes con una resolución de 1444×1444 píxeles (Fang et al. 2022b). Cabe mencionar que se capturó una imagen por cada uno de los ojos de los 1200 pacientes que participaron en el estudio. Todos los pacientes eran de nacionalidad china y presentaban edades comprendidas entre los 53 y los 67 años. La colección de datos final se construyó mediante la selección manual de imágenes de alta calidad obtenidas de

cualquiera de los dos ojos de los pacientes. Estas imágenes procedían tanto de pacientes sin DMAE como de pacientes afectados por la enfermedad (Fang et al. 2022b). En cuanto al porcentaje de muestras de sujetos enfermos y sanos, aquellos sujetos con DMAE representan el 22.5% del total (Fang et al. 2022b). A su vez, Este conjunto de datos, se subdividió aleatoriamente en tres subconjuntos (Fang et al. 2022b):

- Conjunto de entrenamiento (400 imágenes).
- Conjunto en línea empleado para test (400 imágenes).
- Conjunto in situ empleado para test (400 imágenes).

En relación con las etiquetas presentadas por el desafío ADAM, además de las clasificaciones de categoría binaria DMAE/no DMAE, se incluyeron máscaras de segmentación en píxeles del disco óptico completo, máscaras de segmentación de lesiones, así como las coordenadas de la ubicación de la fovea macular. Hasta donde se tiene conocimiento, ADAM es el primer conjunto de datos en incluir un conjunto de etiquetas tan completo (Fang et al. 2022b).

Para este TFG se ha empleado únicamente el conjunto de entrenamiento formado por 400 imágenes, dado que las anotaciones sólo se encontraban disponibles para dicho subconjunto. Dentro de este contexto, se disponía de las siguientes anotaciones (Fang et al. 2022b):

- Etiqueta de clasificación binaria (enfermo/sano) de la DMAE.
- Máscara de la segmentación de la papila.
- Coordenadas del centro de la fovea.
- Cinco máscaras con la segmentación de cuatro tipos de lesiones diferentes que pueden asociarse a la DMAE (drusas, exudados, hemorragias y cicatrices).

En la Tabla 3.1 se puede visualizar una descripción del conjunto de imágenes empleado en este estudio, dividiéndose tanto por la presencia de DMAE como por la resolución presente en cada una de las imágenes.

A su vez, el conjunto de imágenes empleado en este estudio se ha dividido en 2 subconjuntos: entrenamiento y test. El grupo de entrenamiento se utilizó para ajustar los parámetros de los diferentes modelos. Por su parte, el conjunto de test, integrado por un conjunto de imágenes independientes, sirvió para comprobar el funcionamiento del modelo implementado.

En este TFG, las imágenes se dividieron entre los 2 subconjuntos de manera aleatoria, mediante una proporción aproximada de 75:25. De esta forma, se seleccionaron 100 imágenes para el conjunto de test y 294 imágenes para el entrenamiento. parámetros del

Estado	Nº imágenes	Resolución 2124x2056 píxeles	Resolución 1444x1444 píxeles
Enfermo	89	89	0
Sano	311	186	125
Total	400	275	125

Tabla 3. 1: Descripción de la base de datos ADAM.

modelo . En la Tabla 3.2 se puede observar la distribución realizada en este estudio de las imágenes de la base de datos en los diferentes conjuntos de entrenamiento y test.

CONJUNTO DE IMÁGENES	NÚMERO DE IMAGENES	Tipo de imágenes	
		DMAE	SANO
Conjunto de entrenamiento	294	60	234
Conjunto de test	100	29	71
Total	400	263	305

Tabla 3. 2: Distribución de las imágenes en los diferentes subconjuntos

3.3 PREPROCESADO DE RETINOGRAFÍAS

En el análisis de retinografías es común la utilización de algoritmos de preprocesado previo a la extracción de características (Winder et al. 2009). En este estudio, con el objetivo de aumentar el contraste y la precisión en el análisis de las imágenes, así como de reducir el tiempo de computación, la etapa de preprocesado se dividió en dos fases principales: aumento del contraste mediante el algoritmo CLAHE y extracción de la ROI.

3.3.1 AUMENTO DEL CONTRASTE MEDIANTE CLAHE

En general, las retinografías presentan un contraste bajo y variable en función de las zonas de la imagen (Winder et al. 2009). Por ello, es necesario incluir una fase inicial de preprocesado en este tipo de imágenes, que tenga como objetivo resaltar las estructuras importantes de la imagen con respecto al fondo y mejorar su calidad. En este TFG se ha aplicado el algoritmo de realce de contraste CLAHE en cada uno de los tres canales de color de la imagen. CLAHE es una versión modificada de los algoritmos tradicionales de ecualización de histograma. A diferencia de la ecualización de histograma convencional, este método adaptativo calcula múltiples histogramas, cada uno correspondiente a una sección diferente de la imagen (Sonali et al. 2019). Luego, utiliza estos histogramas para redistribuir los valores de luminosidad en la imagen. Como resultado, es especialmente eficaz para mejorar el contraste local y aumentar la definición de los bordes en cada región de la imagen (Musa, Rafi, and Lamsani 2018). Mediante la utilización de este algoritmo se consiguió resaltar las estructuras visibles de la imagen consiguiendo, de esta manera, un mayor diferenciación con el fondo de las imagen.

En el algoritmo CLAHE se definieron una serie de parámetros:

- Número de mosaicos (*'NumTiles'*): este parámetro se refiere al número de mosaicos (*tiles*) en los que se va a dividir la imagen. Es necesario indicar el número de filas y columnas para estos mosaicos. A medida que se aumenta la cantidad de mosaicos establecidos, se logra una mayor precisión en las áreas

más pequeñas, lo que permite obtener detalles más finos (Musa, Rafi, and Lamsani 2018). Con el objetivo de obtener un alto grado de detalle, se ha fijado el tamaño de los bloques a 16x16 píxeles (GeethaRamani and Balasubramanian 2016).

- Límite de contraste (*'ClipLimit'*): parámetro que limita la mejora del contraste, evitando, de esta manera, la sobresaturación de la imagen en zonas homogéneas. Se consigue recortando el histograma por encima de un valor concreto y ecualizándolo. A medida que aumenta el valor de este límite, se logra una mayor mejora en el contraste de la imagen. Sin embargo, un valor más alto puede conllevar una amplificación en el ruido (Musa, Rafi, and Lamsani 2018). Es preciso destacar que el parámetro debe estar en el rango de 0 a 1 para garantizar un ajuste adecuado sin afectar negativamente a la calidad de la imagen. El valor fijado en este TFG fue de 0.005 (GeethaRamani and Balasubramanian 2016).

3.2.1 EXTRACCIÓN DE ROI

La ubicación de las drusas en la DMAE es tan importante como su presencia. Aquellas drusas situadas a una distancia considerable de la mácula tendrán un menor impacto en el desarrollo de la DMAE en comparación con aquellas situadas cerca de esta estructura ocular (Ziyang Liang et al. 2010). Algunos autores consideran que las drusas más relevantes en el caso de la DMAE son aquellas que se encuentran a menos de 3 diámetros de la papila de distancia con respecto al centro de la mácula (Shen et al. 2008; Ziyang Liang et al. 2010). Por consiguiente, en este TFG se ha adoptado esta definición para delimitar una ROI dentro de la imagen en el contexto de la DMAE:

En la base de datos empleada en este estudio, se proporcionaba una máscara de la papila para la mayoría de las imágenes. No obstante, en 125 de las 400 imágenes del conjunto de entrenamiento, la papila aparecía incompleta y para ellas no se proporcionó la máscara correspondiente. Por ello, en este trabajo se optó por calcular el diámetro promedio de la papila para las imágenes de las que se disponía de la máscara correspondiente. Este promedio es de 316 píxeles en las imágenes de la base de datos ADAM. De esta forma, se recortaron las imágenes a un tamaño de 950x950 píxeles (3 veces el diámetro promedio de la papila) tomando como centro el punto central de la fovea (también indicado en las anotaciones de la base de datos).

Tras llevar a cabo una examinación visual se determinó que el algoritmo de extracción de ROI falló en 6 de las 400 imágenes debido a fallos en las coordenadas proporcionadas por la base de datos. Estas fueron automáticamente eliminadas para no comprometer el entrenamiento del modelo.

En la Figura 3.1 se muestra un ejemplo extraído de la base de datos empleada en este TFG y la ROI correspondiente tras la aplicación de las dos etapas de preprocesado.



Figura 3. 1: Ejemplo de obtención de ROI; (a) Imagen original; (b) ROI obtenido

3.4 EXTRACCIÓN DE CARACTERÍSTICAS

Debido a su comprobada eficacia en el análisis de retinografías, en este estudio se seleccionaron 3 tipos de características basadas en el color, la textura y el contexto visual (Phan et al. 2016):

- Histogramas de color
- *Gray Level Cooccurrence Matrix* (GLCM)
- HOG

Las drusas son la lesión más característica de la DMAE. Por ello, este estudio centrará sus esfuerzos en su detección con el objetivo de determinar si un sujeto presenta o no la enfermedad (Phan et al. 2016). Dado que una de las características más distintivas de las drusas asociadas a la DMAE es su color blanquecino-amarillento, la información aportada por los diferentes canales de color representa una importancia muy significativa a la hora de analizar las imágenes afectadas (Phan et al. 2016). Asimismo, la alteración de las estructuras retinianas también es un indicativo de presencia de DMAE. En este caso, tanto los patrones texturales como el estudio de los histogramas de gradiente orientados fueron empleados para obtener información relevante acerca del estado de estas estructuras.

3.4.1 HISTOGRAMAS DE COLOR

El ojo humano tiene la capacidad de detectar una fracción limitada del espectro electromagnético correspondiente a la luz visible. La luz visible está compuesta por ondas con diversas longitudes, que son percibidas en forma de colores variados. Esencialmente,

el color que experimentamos no es más que el resultado de dichas longitudes de onda reflejándose en nuestros ojos. Este espectro visible se extiende desde aproximadamente 700 nanómetros, representando el extremo rojo, hasta alrededor de 400 nanómetros en la región violeta (Pietikainen et al. 2002).

En relación con esto, Los espacios de color se refieren a la estructura mediante la cual se organizan los colores en una imagen y la forma en la que percibimos su representación visual. Un espacio de color RGB se define por las tres cromaticidades de los colores primarios aditivos: rojo, verde y azul (Pietikainen et al. 2002). La luz verde es menos absorbida por las capas de fondo de ojo en comparación con la porción azul del espectro (Walter et al. 2007). Por su parte, la luz roja penetra más profundamente en las capas del ojo interno y se refleja principalmente en la coroides. Esto explica porque el canal verde muestra un contraste más nítido, mientras que el canal rojo a menudo se satura y presenta bajo contraste, y el canal azul tiende a mostrar un alto nivel de ruido (Walter et al. 2007). Sin embargo, a pesar de las consideraciones expuestas sobre los canales rojo y azul, en el contexto de la detección de la DMAE, es fundamental considerar los tres canales de color. Esto resulta necesario para distinguir de manera efectiva entre drusas y exudados, ya que estas lesiones presentan considerables similitudes (Phan et al. 2016).

Por su parte, las características del espacio de color $L^*a^*b^*$, juegan un papel especialmente relevante en la detección de la DMAE. Las tonalidades amarillentas que presentan las drusas aportan una gran cantidad de información sobre su presencia. Esta información se ve reflejada especialmente bien en la coordenada b^* del espacio $L^*a^*b^*$ donde: el valor L^* representa la luminosidad, mientras que a^* y b^* representan las coordenadas cromáticas (Pietikainen et al. 2002).

- L^* : representa la luminosidad.
- a^* : coordenadas rojo/verde. A mayor valor mayor intensidad de rojo y menor de verde.
- b^* coordenadas amarillo/azul. A mayor valor mayor intensidad de azul y menor de amarillo.

En el presente estudio se calcularon los histogramas de 8 *bins* para cada uno de los 6 canales dando como resultado un vector formado por 48 características. Basándonos en las afirmaciones establecidas por Phan et al, el número de *bins* se estableció en 8 debido a que, a partir de este número no se presentaron mejoras reseñables en el modelo (Phan et al. 2016). Además, de esta manera, se consiguió mejorar la eficiencia computacional, sin comprometer la precisión del modelo (Phan et al. 2016).

3.4.2 ESTADÍSTICOS TEXTURALES

La textura puede describirse como un atributo que representa la disposición espacial de los niveles de gris de los píxeles en una región de una imagen digital (Bharati, Liu, and MacGregor 2004). Con el paso de los años, el análisis de textura ha desempeñado un papel importante en diversos áreas del análisis de imágenes (Bharati, Liu, and MacGregor 2004).

En el ámbito de la detección de la DMAE, el análisis de textura presenta un gran potencial para diferenciar entre patrones enfermos y sanos (Khalid et al. 2021). Con el objetivo de lograr una detección más precisa de los patrones en la región conteniente de

drusas, hemos integrado un conjunto de características de textura en el sistema propuesto. El análisis del patrón de textura nos permite identificar de manera efectiva la presencia y propagación de las drusas, así como contemplar las alteraciones de las diferentes estructuras retinianas, lo que contribuye a una clasificación más precisa de los pacientes enfermos y sanos de la DMAE (Khalid et al. 2021).

En este trabajo se ha decidió emplear técnicas estadísticas de análisis de textura (Bharati, Liu, and MacGregor 2004). Estas técnicas se centran en describir la textura de las regiones de una imagen a través de momentos de orden superior de sus histogramas en escala de grises (Bharati, Liu, and MacGregor 2004). El método más comúnmente utilizado para ello es la GLCM. El método GLCM se utiliza para calcular la frecuencia con la que un nivel de gris aparece en una relación con otro valor de gris dentro del área de una ventana determinada (Presutti 2004). Dicha frecuencia se va a representar mediante una matriz cuadrada, cuyo tamaño coincide con el número de niveles de gris que se van a analizar. En este caso tendrá un tamaño de 8×8 , ya que se han empleado histogramas de 8 *bins* (Bharati, Liu, and MacGregor 2004).

La GLCM examina la relación espacial entre dos píxeles en una imagen. Estos se conocen como píxel de referencia y píxel vecino. Por ejemplo, si seleccionamos el píxel vecino que se encuentra un píxel a la derecha de cada píxel de referencia, esta relación, también llamado *offset*, se expresa como (1,0): 1 píxel en la dirección horizontal x y 0 píxeles en la dirección vertical y (Presutti 2004). En el proceso de cálculo, cada píxel dentro de la ventana se convierte progresivamente en el píxel de referencia, comenzando desde la esquina superior izquierda y avanzando hacia abajo y a la derecha (Presutti 2004). Este proceso continuaría hasta completar todo el área de la imagen. Una vez se obtiene la matriz con las diferentes combinaciones, el siguiente paso es calcular la probabilidad de co-ocurrencia, para cada uno de los casos, es decir el número de veces que una relación entre píxeles ocurre, dividido por el número total de posibles relaciones (Presutti 2004). Después de que se haya construido esta matriz, se derivarán de ella diferentes estadísticos texturales (Presutti 2004).

En este trabajo el método GLCM se implementó a cada uno de los tres canales de color RGB a través del *software* de MATLAB®. Posteriormente, basándonos en los estudios revisados en la literatura se concluyó que los estadísticos texturales más utilizados en el eran los siguiente (Presutti 2004):

- Homogeneidad
- Contraste
- Correlación
- Energía

Por ello, tras implementar el método GLCM, se decidió a extraer dichos estadísticos texturales para caracterizar la enfermedad. Tras la obtención de estos estadísticos se obtuvieron 12 características para cada imagen, cuatro para cada uno de los 3 canales RGB.

En primer lugar, la homogeneidad mide la cercanía de la distribución de los elementos de la GLCM a su diagonal. Sus valores se mueven en el rango $[0,1]$, siendo 1 para toda la diagonal (Chaki and Dey 2020). Esta se calcula como:

$$Homogeneidad = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (3.1)$$

Siendo $P_{i,j}$ la probabilidad de co-ocurrencia de los niveles de gris i, j para una distancia dada y N el número de niveles de gris empleados (Presutti 2004).

Por su parte, el contraste es la medida opuesta a la homogeneidad. Se encarga de reflejar la diferencia entre el contraste de intensidad de un píxel y su vecino en toda la imagen (Presutti 2004). Esta se calcula como:

$$Contraste = \sum_{i,j=0}^{N-1} P_{i,j}(i - j)^2 \quad (3.2)$$

Siendo $P_{i,j}$ la probabilidad de co-ocurrencia de los niveles de gris i, j para una distancia dada y N el número de niveles de gris empleados (Presutti 2004).

La correlación es una medida utilizada para cuantificar la relación entre la intensidad de los píxeles en una imagen y su disposición espacial. El valor de la correlación está en el rango $[-1,1]$, donde valores cercanos -1 indican relación lineal negativa entre la intensidad de los píxeles en esa dirección. Por su parte, el valor 0 indica una correlación nula y valores cercanos a 1 sugieren una relación lineal positiva entre los píxeles en esa dirección (Ref). Se calcula como (Chaki and Dey 2020):

$$Correlación = \sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right] \quad (3.3)$$

Siendo $P_{i,j}$ la probabilidad de co-ocurrencia de los niveles de gris i, j para una distancia dada y N el número de niveles de gris empleados (Presutti 2004). A su vez, σ representa la desviación estándar de $P_{i,j}$ y μ la media de $P_{i,j}$ (Presutti 2004).

Por último, la energía se encarga de medir la homogeneidad mediante la suma del cuadrado de las entradas. Se enmarca en el rango $[0,1]$ (Presutti 2004).

$$Energía = \sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (3.4)$$

Siendo $P_{i,j}$ la probabilidad de co-ocurrencia de los niveles de gris i, j para una distancia dada y N el número de niveles de gris empleados (Presutti 2004).

3.4.3 HISTOGRAMAS DE GRADIENTE ORIENTADOS

HOG es un método utilizado en el procesamiento de imágenes con el fin de detectar objetos. Su idea central es que la forma y apariencia de un objeto local dentro de una imagen pueden describirse mediante la distribución de sus gradientes de intensidad (H. Zhou and Yu 2021). En este método, la imagen se divide en pequeñas regiones conectadas denominadas celdas. Posteriormente, para los píxeles dentro de cada celda se compila un histograma de gradientes. (H. Zhou and Yu 2021). Los histogramas de estas celdas se irán agrupando para formar el vector final de características (W. Zhou et al. 2020).

Para llevar a cabo la implantación de HOG, es necesario llevar a cabo una serie de pasos:

1. **Cálculo de gradientes.** El primer paso se basa en el cálculo de los valores de gradiente en las direcciones vertical y horizontal. Estos dos gradientes se utilizarán más adelante para determinar las magnitudes y los ángulos de gradiente (W. Zhou et al. 2020).
2. **Emisión de voto.** Este procedimiento consiste en generar los histogramas de gradiente de cada una de las celdas. En estas, cada uno de los píxeles emite un voto ponderado para conformar un histograma basado tanto en la orientación como en la magnitud de gradiente. Los *bins* del histograma se distribuyen uniformemente entre 0° y 180° . En la Figura 3.3 se puede observar una explicación de esta idea (W. Zhou et al. 2020). Pongamos el ejemplo de que la imagen se divide en celdas de 8×8 píxeles. Estos píxeles nos aportan información sobre la dirección y la magnitud del gradiente de intensidad. Cada uno de los *bins* del histograma se seleccionan en función del ángulo de gradiente (dirección) y su valor agregado corresponde a la magnitud de gradiente (W. Zhou et al. 2020). Si se presta atención al círculo azul, se puede observar cómo su ángulo de gradiente es de 80° y su magnitud es de 2, con lo cual se agregaría la magnitud 2 al quinto *bin* correspondiente al ángulo 80° (W. Zhou et al. 2020).
3. **Creación de bloques.** Con el objetivo de tener en cuenta las variaciones en la iluminación y el contraste, las intensidades del gradiente se normalizan localmente. Para lograr esto, se agrupan las celdas adyacentes en áreas más grandes, denominadas bloques. De este modo, el descriptor HOG se forma mediante la concatenación de los histogramas de celdas normalizadas de cada una de las regiones de los bloques (W. Zhou et al. 2020). Hay que tener en cuenta que estos bloques, en la mayoría de los casos, se superponen, por lo que algunas celdas contribuyen más de una vez al descriptor final (W. Zhou et al. 2020).

En este trabajo la aplicación de HOG se realizó mediante la utilización del *software* MATLAB®. En base a las investigaciones realizadas por Khalid et al, se seleccionaron una serie de parámetros fundamentales para el cálculo de HOG (Khalid et al. 2021). Entre ellos se encuentran:

- Tamaño de celda (*'CellSize'*): la elección del tamaño de celda es imprescindible para obtener un buen descriptor. En función del nivel de detalle que se requiera se incrementará o disminuirá la dimensión de las celdas. Es importante tener en cuenta que un tamaño pequeño de celda conseguirá un nivel de detalle muy destacable, pero un vector de características demasiado

grande, con el coste computacional que ello conlleva. En base a estudios anteriores se seleccionó un tamaño de 16x16 píxeles por celda (Khalid et al. 2021).

- Tamaño de bloque (*'BlockSize'*): indica el número de celdas por bloque. Un tamaño grande de bloque reduce la habilidad de HOG para suprimir cambios locales en la iluminación. Por su parte, reducir el tamaño de bloque ayuda a capturar la importancia de los píxeles locales. El tamaño de bloque elegido fue de 2x2 (Dalal and Triggs 2005).
- Solapamiento de bloques (*'BlockOverlap'*): señala el número de celdas superpuestas entre bloques adyacentes. Para garantizar una normalización adecuada del contraste se determinó una superposición de al menos la mitad del tamaño del bloque. Los valores de solapamiento grande capturan más información, pero producen vectores de mayor tamaño.
- Número de *bins* (*'NumBins'*): muestra el número de *bins* presentes en un histograma. A mayor número de *bins*, mayor nivel de detalle, pero también mayor tamaño del vector resultante. En este estudio se seleccionaron 5 *bins* (Dalal and Triggs 2005).

Con todos estos parámetros expuestos, se extrajeron los HOG de las imágenes de ROI de tamaño 950x950 píxeles. Cada una de ellas se dividió en 58x58 bloques, creando un vector final de 67820 características de longitud. Este fue calculado mediante el producto entre el número de bloques, el solapamiento entre bloques y el número de *bins* de los histogramas. Basándonos en estudios previos, estas se empaquetaron de dos formas distintas. En primer lugar, se dividieron en histogramas de 32 *bins*. Posteriormente, se obtuvieron estadísticos texturales de los HOG, obteniendo, respectivamente, 32 y 4 nuevas características por cada imagen (Khalid et al. 2021).

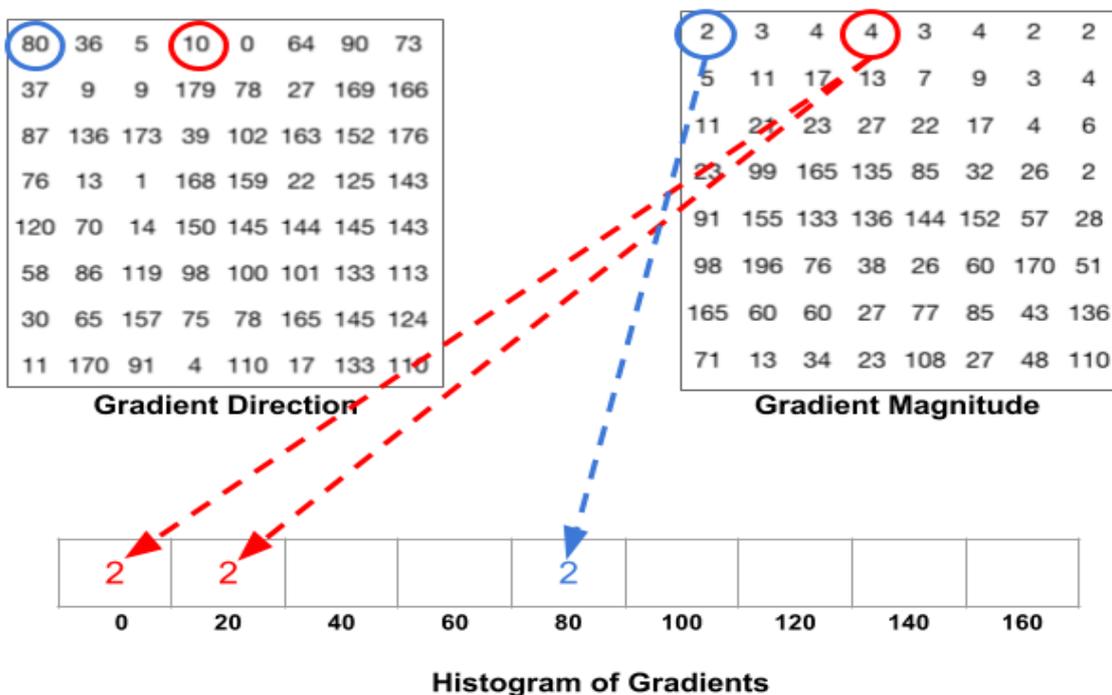


Figura 3.2: Explicación del funcionamiento de HOG (Khalid et al. 2021).

3.5 SELECCIÓN DE CARACTERÍSTICAS

La selección de características es un enfoque empleado para reducir la dimensionalidad de los datos (Li et al. 2017). Esta técnica se fundamenta en elegir las características más relevantes descartando aquellas redundantes y carentes de importancia. Dicho proceso se realiza manteniendo el significado físico de las características, lo que otorga al modelo una mayor capacidad de interpretación y análisis (Li et al. 2017). Al reducir la dimensionalidad de los datos, se obtienen una serie de mejoras en el rendimiento del modelo. Entre ellas destacan el aumento de la capacidad de generalización, la reducción de la complejidad y el incremento en la eficiencia computacional (Remeseiro and Bolon-Canedo 2019).

En este estudio se seleccionó el método FCBF. FCBF es un procedimiento de selección de características multivariante, basado en considerar tanto la correlación entre clase y característica como la correlación de las características entre sí (Yu and Liu 2003). Su objetivo es identificar las características que presentan una alta correlación con la variable objetivo y, al mismo tiempo, una baja correlación entre ellas. En primer lugar, se calcula la correlación entre las características de nuestro conjunto de datos y la variable objetivo que se busca predecir (Yu and Liu 2003). A medida que se seleccionan las diferentes características el algoritmo verifica la correlación entre ellas, si se da el caso de que dos características se encuentran altamente correlacionadas, se eliminará una de ellas para evitar información redundante. Para llevar a cabo este proceso se utiliza la incertidumbre simétrica (SU) (Yu and Liu 2003).

$$SU = 2 \left[\frac{H(X) - H(X|Y)}{H(X) + H(Y)} \right] \quad (3.5)$$

Siendo $H(X)$ la entropía de la variable X , $H(Y)$ la entropía de la variable Y , $H(X|Y)$ la entropía de la variable X después de observar los valores de otra variable Y

Para este algoritmo es necesario elegir el valor del umbral δ . En este estudio, el parámetro δ se estableció en 5 en base a trabajos previos (Yu and Liu 2003). Aplicando el algoritmo FCBF con este valor del umbral, se consiguió reducir el número de características de 94 hasta 19, que serán las más relevantes para la tarea de determinar la presencia o ausencia de DMAE en una imagen.

3.6 SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE

El desbalanceo de datos es un problema muy común en los modelos de *Machine Learning*. Existen multitud de razones por las que un conjunto de datos puede presentar un desequilibrio. Por ejemplo, la categoría desequilibrada podría ser poco frecuente en la población o los datos podrían ser difíciles de recopilar (Torgo et al. 2013). Una estrategia utilizada para tratar conjuntos de datos desequilibrados es el sobre muestreo de la clase minoritaria. El método *Synthetic Minority Over-sampling Technique* (SMOTE) es una técnica estadística utilizada para aumentar equilibradamente el número de casos de la

clase minoritaria en un conjunto de datos. Este enfoque genera nuevas instancias tomando como entrada los casos minoritarios existentes (Torgo et al. 2013). La implementación de SMOTE no va a modificar el número de casos de la clase mayoritaria, centrándose únicamente en generar instancias sintéticas para la clase minoritaria. Estas nuevas instancias no son simples copias de los casos minoritarios existentes. En cambio, el algoritmo toma muestras del espacio de características de cada clase objetivo y sus vecinos más cercanos (Torgo et al. 2013). Posteriormente, combina características del caso actual con características de sus vecinos para generar ejemplos nuevos. Este enfoque aumenta las características disponibles para cada clase y, en consecuencia, mejora la generalización de las muestras generadas (Torgo et al. 2013).

Uno de los problemas de la base de datos ADAM era su notable desbalanceo, ya que 305 de las 394 imágenes pertenecían a sujetos sanos. Concretamente, dentro del conjunto de entrenamiento, únicamente, 60 de las 234 imágenes pertenecían a sujetos enfermos. Para solventar este problema, en este TFG se implementó el método SMOTE (Chawla et al. 2002). La idea principal detrás de este procedimiento era conseguir tener el mismo número de imágenes de sujetos enfermos y sanos. Para lograrlo, se aumentó el número de sujetos enfermos, en una proporción igual al desbalanceo. Con el objetivo de calcular dicha proporción, se restaron el número de sujetos sanos y enfermos y posteriormente se dividieron entre el número de sujetos enfermos, obteniendo una proporción cercana a 3 y generando 174 muestras sintéticas de sujetos enfermos. De esta manera, se consiguió igualar a 234 el número de sujetos enfermos y sanos. Además, el parámetro correspondiente al número de vecinos se determinó en 5 teniendo en cuenta el trabajo de Chawla et al (Chawla et al. 2002).

3.7 CLASIFICACIÓN MEDIANTE *MACHINE LEARNING*

Para realizar la clasificación del modelo en este estudio se utilizó la aplicación “*Classification Learner*” del *software* de MATLAB®. En ella, se seleccionó la opción de entrenar todos los modelos de *Machine Learning* disponibles para escoger aquel que ofreciese un mayor rendimiento. Para obtener estos resultados, y debido al tamaño reducido del conjunto de entrenamiento, se empleó la técnica de validación *K-fold Cross-Validation*. La validación cruzada es un procedimiento de remuestreo utilizado para evaluar modelos de aprendizaje automático en una muestra de datos limitada (Fushiki 2011). El enfoque de validación cruzada implica dividir aleatoriamente el conjunto de entrenamiento en K grupos de aproximadamente el mismo tamaño. El primero de los grupos se utilizará como conjunto de validación, ajustándose el método para los $K-1$ grupos restantes (Fushiki 2011). En este estudio, el valor de K se estableció en 10 en base a trabajos previos (Phan et al. 2016). De esta manera aquel que ofreció un mejor rendimiento fueron los *Ensemble Methods*.

Con el objetivo de obtener el mejor resultado posible, en este TFG se decidió emplear una optimización automática de los hiperparámetros del modelo. Para este propósito se implementó la técnica de optimización bayesiana para ajustar los hiperparámetros del *Ensemble* (Wu et al. 2019). Este enfoque se desarrolla mediante una serie de pasos destinados a descubrir la combinación óptima de hiperparámetros (Wu et al. 2019):

1. Se define un espacio de búsqueda que contiene los posibles valores para optimizar el método Ensemble.
2. Se construye un modelo probabilístico basado en el teorema de Bayes para representar la relación entre los hiperparámetros y una función objetivo.
3. Se elige una función objetivo para evaluar el rendimiento del Ensemble con una configuración específica de hiperparámetros, como la precisión o el área bajo la curva ROC, entre otras opciones.
4. Se realiza una evaluación inicial utilizando un conjunto de hiperparámetros seleccionados de forma aleatoria.
5. Se busca la combinación óptima de hiperparámetros mediante la exploración en el modelo probabilístico, seleccionando nuevos conjuntos de hiperparámetros con el objetivo de mejorar la función objetivo.
6. Se actualiza el modelo probabilístico después de cada evaluación y se repiten los pasos anteriores hasta alcanzar una solución convergente.

3.7.1 ENSEMBLE METHODS

Ensemble Learning es un enfoque de aprendizaje automático basado en entrenar varios clasificadores individuales, usualmente denominados “*weak learners*”, para abordar la misma tarea y luego combinarlos para obtener resultados mejorados (X. Dong et al. 2020). La premisa fundamental es que, al combinar adecuadamente estos clasificadores individuales, se consiguen modelos más precisos y robustos (X. Dong et al. 2020). Se denominan clasificadores individuales a aquellos modelos que pueden utilizarse como bloques de construcción para diseñar modelos más complejos combinando varios de ellos. En la mayoría de los casos, estos clasificadores individuales presentan un sesgo elevado o una alta variabilidad. Por lo tanto, la idea detrás de los *Ensemble Methods* es intentar reducir el sesgo y/o la varianza de estos clasificadores combinando varios de ellos para crear un modelo que logre un mejor rendimiento (X. Dong et al. 2020).

Para establecer un método de *Ensemble*, en primer lugar, debemos seleccionar los clasificadores individuales que se van a incorporar. Un punto a tener en cuenta es que la elección de los clasificadores debe ser coherente con la forma en la que se agregan los modelos. Si optamos por clasificadores individuales con bajo sesgo, pero alta varianza, se debería aplicar un método de agregación que tienda a reducir la varianza. Existen diferentes planteamientos para llevar a cabo estas agregaciones (X. Dong et al. 2020):

- **Bagging.** El objetivo de *bagging* es entrenar múltiples clasificadores independientes y combinar sus predicciones para obtener un modelo con menor varianza. Los árboles de decisión son muy empleados en la literatura como clasificadores individuales.
- **Boosting.** Los métodos de *boosting* se basan en construir un modelo predictivo sólido mediante el ajuste sucesivo de múltiples clasificadores individuales. Los clasificadores son agregados de manera secuencial, de forma que cada uno actúa corrigiendo a su predecesor. Esto se repite hasta que los datos se predigan de manera perfecta o se alcance el número máximo de modelos permitidos. El objetivo principal del *boosting* es reducir el sesgo de error que se produce cuando los modelos no logran identificar patrones relevantes en

los datos. Esto se logra mediante la evaluación de la discrepancia entre el valor predicho y el valor real. *LogitBoost* o *AdaBoost* son algunos de los tipos de los modelos base basados en *Boosting*.

- **Stacking.** Difiere principalmente de *Bagging* y *Boosting* en dos aspectos. En primer lugar, el método de *Stacking* emplea diferentes algoritmos de aprendizaje, mientras que *Bagging* y *Boosting* emplean el mismo. Por otra parte, *stacking* combina los clasificadores individuales utilizando un metamodelo, mientras que los demás emplean algoritmos deterministas. Por ejemplo, en un problema de clasificación se podrían elegir clasificadores individuales un clasificador *k*-NN y un SVM y decidir aprender una red neuronal como metamodelo. De esta manera, la red neuronal tomará como entradas la salida de los clasificadores individuales y aprenderá a devolver predicciones basadas en ellos.

Como se mencionó previamente, en este estudio se optó por optimizar automáticamente los hiperparámetros de *Ensemble*. En este contexto, se pueden ajustar una serie hiperparámetros para optimizar el rendimiento del clasificador. Entre ellos destacan: el número de clasificadores individuales, el tipo de clasificadores, el modelo base y los hiperparámetros dependientes del modelo base.

Entre las posibles opciones disponibles para el modelo base se incluían *Adaboost*, *RUSboost*, *LogitBoost*, *Gentleboost* y *Bag*. Después de aplicar el optimizador bayesiano, se seleccionó *LogitBoost* como el modelo base óptimo. El funcionamiento de este método se basa en los fundamentos de *Boosting*. Entre los tipos de clasificadores individuales aptos para la selección se encontraban: árboles de decisión, regresión logística, y SVM. El tipo de clasificador que seleccionó la optimización bayesiana fueron los árboles de decisión. El número de árboles de decisión podía variar en un rango entre 10 y 500. Finalmente, el optimizador bayesiano eligió un total de 118 árboles, cada uno dividido en un máximo de 16 nodos. Además, al seleccionar *LogitBoost*, se introdujo un nuevo hiperparámetro ajustable: la tasa de aprendizaje, que variaba un rango de [0.001, 1]. El optimizador bayesiano determinó que la tasa de aprendizaje óptima era de 0.8940.

CAPÍTULO 4. RESULTADOS

4.1 INTRODUCCIÓN

En este trabajo, en primer lugar, se ha llevado a cabo una fase de preprocesado de las imágenes. Está se fundamentó en el realce del contraste y la extracción de los ROIS. Posteriormente, sobre dichas imágenes se aplicaron diferentes técnicas de extracción de características: histogramas de color, GLCM y HOG. A continuación, con el objetivo de reducir la dimensionalidad de los datos se llevó a cabo una fase de selección de características mediante FCBF. Seguidamente, para solucionar el desbalanceo presente en la base de datos se aplicó el método SMOTE. Para finalizar, partir de las características extraídas, se entrenó un modelo de *Machine Learning* y se recogieron sus resultados.

Una vez se ha detallado la metodología empleada, en este capítulo se presentarán los principales resultados obtenidos en la detección automática de DMAE. En primer lugar, se describirán las métricas empleadas para la evaluación del modelo. Seguidamente, se presentarán los resultados obtenidos tanto para la fase de entrenamiento como para la fase de test.

4.2 MODO DE EVALUACIÓN

Un aspecto muy importante en la elaboración de un sistema de clasificación es la evaluación de los resultados. Para ello, es crucial establecer medidas que brinden información sobre el desempeño del modelo desarrollado.

El problema abordado en este trabajo es una clasificación, lo que significa que la variable objetivo solo puede adquirir un conjunto específico de valores discretos. En este caso estamos ante un problema con dos posibles clases: enfermo y sano. Al tratarse de una clasificación binaria, el modelo de *Machine Learning* desarrollado se han evaluado en términos de: precisión, sensibilidad, especificidad, valor predictivo positivo (VPP), valor predictivo negativo (VPN), F1, curva *Receiver Operating Characteristic (ROC)* y matrices de confusión tanto en la fase de entrenamiento como en la fase de test (Vizcaíno-Salazar 2017).

4.2.1 MATRIZ DE CONFUSIÓN

La matriz de confusión es una tabla cruzada que permite comparar la clasificación realizada por dos evaluadores (Vizcaíno-Salazar 2017). En este contexto, se utilizará para comparar la clasificación real proporcionada por el especialista con la clasificación pronosticada por el método propuesto. Las columnas de la matriz representan las

predicciones del modelo, mientras que las filas muestran la clasificación real. Los parámetros que se presentan en una matriz de confusión son (Vizcaíno-Salazar 2017):

- Verdadero positivo (VP): número de instancias positivas que el modelo predice como positivas.
- Falso positivo (FP): número de instancias negativas que el modelo predice como positivas
- Verdadero negativo (VN): número de instancias negativas que el modelo predice como negativas.
- Falso negativo (FN): número de instancias negativas que el modelo predice como positivas (Vizcaíno-Salazar 2017).

4.2.2 PRECISIÓN

Con el objetivo de entender la métrica de la precisión, es preciso establecer, previamente, definiciones de algunos conceptos clave. Por un lado, la sensibilidad indica la proporción de instancias positivas que han sido identificadas correctamente por el modelo entre todas las instancias positivas. La especificidad, por su parte, señala la proporción de instancias negativas que han sido identificadas correctamente por el modelo entre todas las instancias negativas (Martínez Pérez y Pérez Martín 2023).

Tanto la sensibilidad como la especificidad pueden expresarse como (Martínez Pérez y Pérez Martín 2023):

$$Sensibilidad = \frac{VP}{VP + FN} \quad (4.4)$$

$$Especificidad = \frac{VN}{VN + FP} \quad (4.5)$$

Con respecto a la precisión, la métrica utilizada ha sido la precisión categórica. La precisión categórica indica la fracción de predicciones que el modelo realizó correctamente. Se calcula como (Vizcaíno-Salazar 2017):

$$Precisión = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.1)$$

4.2.3 VALOR PREDICTIVO POSITIVO Y VALOR PREDICTIVO NEGATIVO

Los valores predictivos indican la capacidad de un modelo para predecir correctamente la presencia o ausencia de un evento o condición (Vizcaíno-Salazar 2017). El VPP indica la relación entre los verdaderos positivos y el total de positivos. Se define como (Vizcaíno-Salazar 2017):

$$VPP = \frac{VP}{(VP + FP)} \quad (4.2)$$

El VPN indica la relación entre los verdaderos negativos y el total de negativos. Se define como (Vizcaíno-Salazar 2017):

$$VPN = \frac{VN}{(VN + FN)} \quad (4.3)$$

4.2.4 CURVA RECEIVER OPERATING CHARACTERISTIC

La curva ROC es una herramienta estadística empleada para evaluar la capacidad de discriminación de una prueba diagnóstica dicotómica (Martínez Pérez y Pérez Martín 2023). Estas curvas representan gráficamente la sensibilidad frente al complemento de la especificidad (1-especificidad) para diferentes puntos de corte. La sensibilidad se situará en el eje de ordenadas y el complemento de la especificidad, en el de abscisas. Ambos ejes englobarán valores situados entre 0 y 1. La curva ROC se formará al unir los pares de valores correspondientes a cada uno de los puntos de corte. Estos puntos de corte representan umbrales específicos que permiten tomar decisiones sobre como clasificar las instancias de un modelo. En la figura 4.1 se puede visualizar un ejemplo de curva ROC.

La utilidad de la curva ROC se extiende a varios aspectos clave: seleccionar el punto de corte óptimo de una prueba, comprender el rendimiento global de la misma y comparar la capacidad discriminativa de dos o más pruebas diagnósticas (Martínez Pérez y Pérez Martín 2023).

Es preciso destacar que la información de la curva ROC puede resumirse en un solo parámetro, conocido como *Area Under the ROC Curve* (AUC). El parámetro AUC puede definirse como la probabilidad de clasificar correctamente a un par de individuos seleccionados al azar, uno sano y otro enfermo, al aplicar la prueba. Esta medida proporciona una medida global de la capacidad discriminativa de la prueba a lo largo de todos los posibles puntos de corte (Martínez Pérez y Pérez Martín 2023). Una prueba con

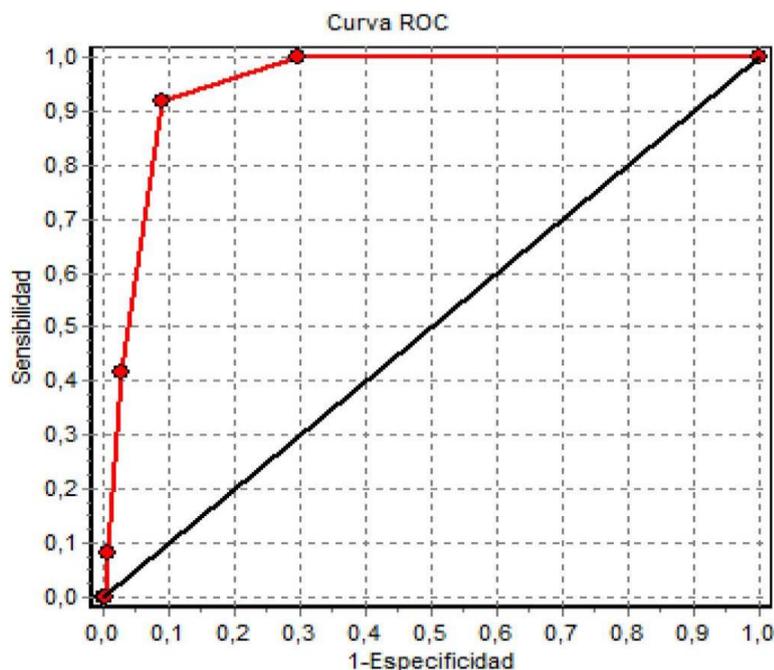


Figura 4. 1: Ejemplo de Curva ROC (Martínez Pérez y Pérez Martin 2023).

una discriminación perfecta se representaría por una curva ROC localizada en la parte superior izquierda del gráfico, con sensibilidad igual a 1 y especificidad igual a 1. En contraste, una curva sin capacidad discriminativa se representaría por la diagonal del gráfico, con sensibilidad y especificidad iguales a 0,5 (Martínez Pérez y Pérez Martin 2023). Por lo tanto, un AUC=1 indicaría una clasificación perfecta mientras que AUC=0.5 sería equivalente a un clasificador aleatorio (Martínez Pérez y Pérez Martin 2023).

4.2.5 F1

La métrica F1 se define como la media armónica de VPP y sensibilidad (Sokolova, Japkowicz, y Szpakowicz 2006). El objetivo de la puntuación F1 es combinar estas métricas en un solo parámetro. F1 es utilizada normalmente cuando se desea encontrar un equilibrio entre la sensibilidad y la VPP en los modelos, debido a que penaliza a aquellos que presentan un desequilibrio entre estas dos métricas (Sokolova, Japkowicz, y Szpakowicz 2006). Se puede definir como (Sokolova, Japkowicz, y Szpakowicz 2006):

$$F1 = 2 \times \frac{VPP \times Sensibilidad}{VPP + Sensibilidad} \tag{4.6}$$

4.3 RESULTADOS DE LA ETAPA DE ENTRENAMIENTO

4.3.1 CARACTERÍSTICAS SELECCIONADAS

En este estudio, como se detalló en el capítulo 3, con el objetivo de seleccionar las características más relevantes, se llevó a cabo la implementación del algoritmo FCBF. La aplicación de este algoritmo resultó en una reducción significativa del número de características, pasando de un total inicial de 94 a 19. En la Tabla 4.2 muestran las características que se seleccionaron en el proceso.

4.3.2 ETAPA DE ENTRENAMIENTO

Durante la etapa de entrenamiento se ajustaron los parámetros del método desarrollado y se seleccionó el clasificador que proporcionó mejor rendimiento en la clasificación de imágenes correspondientes a sujetos sanos y a sujetos con DMAE. Para ello, se entrenó el modelo de *Machine Learning* empleando *10-fold cross validation* y se promediaron los resultados correspondientes. Posteriormente, se calcularon los valores medios de precisión, sensibilidad, especificidad, y valores predictivos sobre el conjunto de validación. Además, para ofrecer una evaluación más completa de los resultados, se calcularon la curva ROC y la matriz de confusión.

En la Tabla 4.3 se muestran los resultados promedio del clasificador sobre el conjunto de validación. Cabe destacar que los valores de precisión están por encima del 87% en todos los casos. Además, los valores de sensibilidad y especificidad están bastante balanceados en todos los casos. No obstante, para el método *Ensemble* se obtuvieron mejores resultados en todas las métricas. Por lo tanto, este ha sido el método seleccionado para este TFG. En las Figuras 4.2 y 4.3 respectivamente se pueden observar tanto las curvas ROC como las matrices de confusión para cada uno de los métodos empleados.

ORDEN DE SELECCIÓN	CARACTERÍSTICA SELECCIONADA
1.	<i>Histograma canal B</i>
2.	<i>Histograma canal a</i>
3.	<i>Histograma canal a</i>
4.	<i>Histograma canal b</i>
5.	<i>Homogeneidad estadísticos HOG</i>
6.	<i>Histogramas HOG</i>
7.	<i>Histograma canal R</i>
8.	<i>Correlación canal B</i>
9.	<i>Histograma canal b</i>
10.	<i>Histogramas HOG</i>
11.	<i>Histograma canal b</i>
12.	<i>Homogeneidad canal G</i>
13.	<i>Histogramas HOG</i>
14.	<i>Histograma canal B</i>

Capítulo 4. Resultados

15.	<i>Histograma HOG</i>
16.	<i>Histograma HOG</i>
17.	<i>Histograma canal b</i>
18.	<i>Histograma canal a</i>
19.	<i>Histograma canal b</i>

Tabla 4. 1: Características seleccionadas por el algoritmo FCBF.

MÉTODOS	PRECISIÓN	SENSIBILIDAD	ESPECIFICIDAD	VPP	F1	VPN	AUC
<i>Ensemble</i>	91.2%	92.7%	89.7%	90%	91.33%	92.5%	0.96

Tabla 4.2: Resultados promedio sobre el conjunto de validación para el clasificador empleado en el estudio.

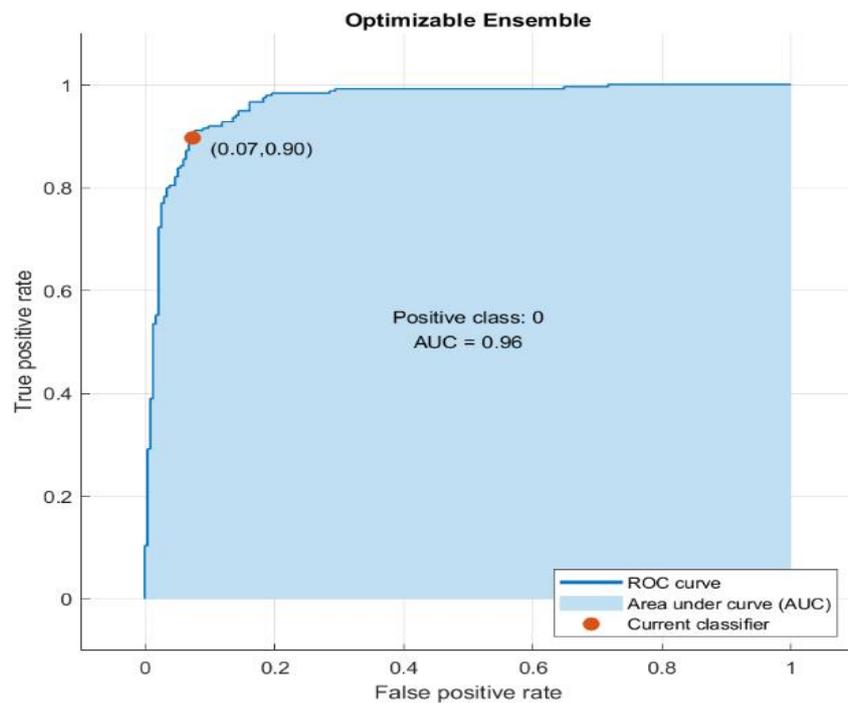


Figura 4. 2: Curva Roc promedio obtenida sobre el conjunto de validación

Figura 4. 2: Curva Roc promedio obtenida sobre el conjunto de validación

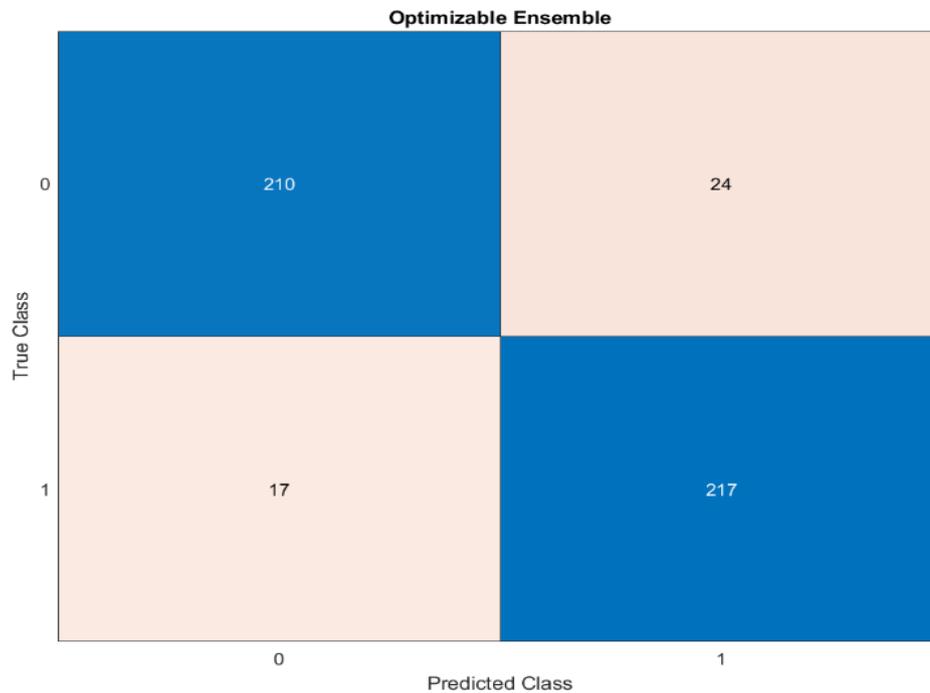


Figura 4. 3: Matriz de confusión promedio obtenidas sobre el conjunto de validación

Figura 4. 4: Curva ROC perteneciente al conjunto de test
 Figura 4. 5: Matriz de confusión promedio obtenidas sobre el conjunto de validación

4.4 RESULTADOS SOBRE EL CONJUNTO DE TEST

En la etapa de test, se evaluó el clasificador seleccionado sobre el conjunto de imágenes de test. De esta manera, se puede verificar la efectividad del algoritmo desarrollado en un conjunto de imágenes distinto al utilizado durante el entrenamiento. En la Tabla 4.4, se muestran los resultados obtenidos sobre el conjunto de prueba.

Asimismo, para poder realizar un análisis más detallado de los resultados de test, se calculó tanto la curva ROC como las matriz de confusión para cada uno de los clasificadores. Se pueden observar en las Figuras 4.4 y 4.5, respectivamente.

MÉTODOS	PRECISIÓN	SENSIBILIDAD	ESPECIFICIDAD	VPP	F1	VPN	AUC
---------	-----------	--------------	---------------	-----	----	-----	-----

Capítulo 4. Resultados

Ensemble	81%	55.1%	91.5%	72.7%	62.6%	83.3%	0.86
----------	-----	-------	-------	-------	-------	-------	------

Tabla 4. 3: Resultados sobre el conjunto de test

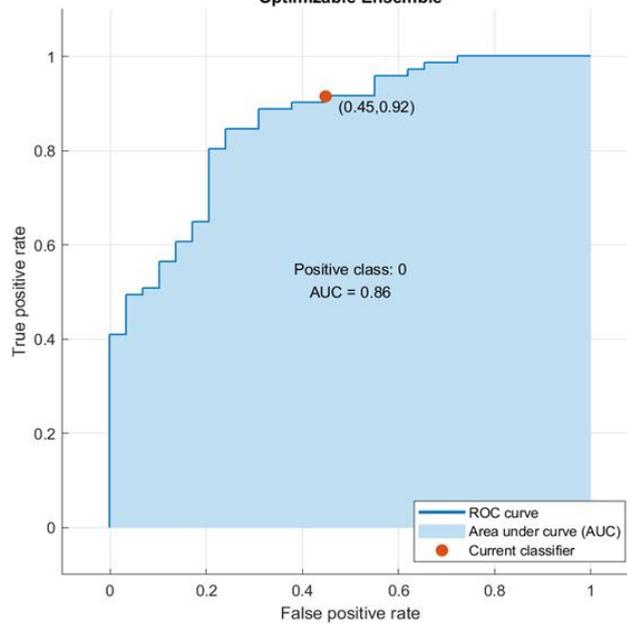


Figura 4. 9: Curva ROC perteneciente al conjunto de test

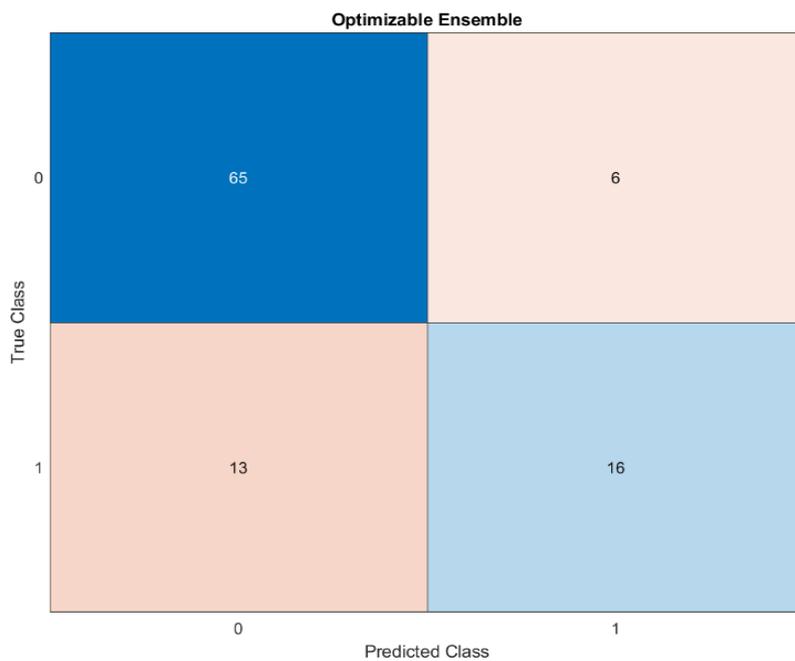


Figura 4. 12: Matriz de confusión pertenecientes al conjunto de test

CAPÍTULO 5. DISCUSIÓN

5.1 INTRODUCCIÓN

En este TFG se ha desarrollado y evaluado un sistema de detección de DMAE en retinografías. Este sistema se ha entrenado utilizando un conjunto de retinografías pertenecientes a la base de datos ADAM, verificando su rendimiento mediante un subconjunto de imágenes de la propia base de datos.

En este capítulo, se presentan una discusión de los resultados obtenidos durante el estudio. Además, se presentan sus principales ventajas e inconvenientes y se realiza una comparación con los resultados de otros estudios publicados previamente.

5.2 PREPROCESADO

En este estudio se ha incorporado una fase de preprocesado, previa a la extracción de características. Esta tiene como objetivo tanto mejorar la calidad de las imágenes como resaltar las estructuras más importantes con respecto al fondo. En primera instancia, se realizó un realce del contraste de cada uno de los canales RGB mediante el algoritmo CLAHE. De esta manera, se realzaron las estructuras principales de la imagen. Uno de los inconvenientes que presenta este algoritmo es que se debe implementar en cada uno de los canales de color por separado, aumentando el coste computacional del sistema. A pesar de que la componente más utilizada en la literatura para la detección de DMAE es el canal verde, en este estudio se optó por utilizar los tres canales de color porque se quería evitar la pérdida de la información contenida en dichos canales (García-Floriano et al., 2019).

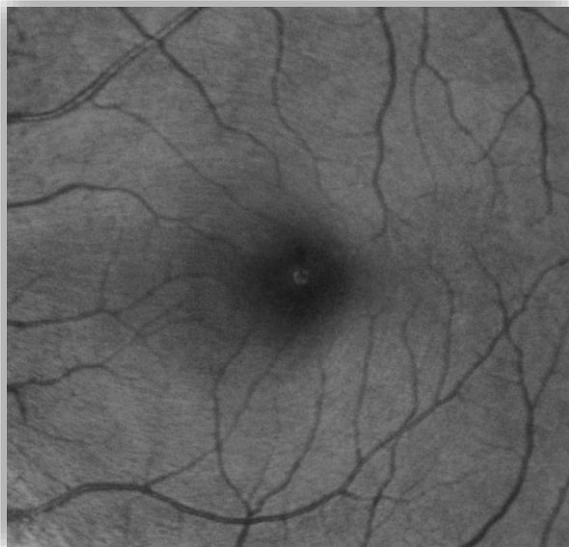
Posteriormente, se llevó a cabo la fase de extracción de ROI. En esta fase se extrae una región de 950x950 píxeles alrededor del centro de la mácula. Esta es la zona más relevante para la detección de la DMAE, por lo que la mayoría de los estudios se centran en analizar esta área [Referencias]. Este proceso, además, permite reducir el coste computacional del algoritmo. El método propuesto para la extracción de ROI se basaba, fundamentalmente, en las coordenadas de la fóvea y el diámetro del disco óptico. Puesto que las anotaciones de la base de datos no contenían la máscara de la papila para todas las imágenes (Shen et al., 2008), se decidió partir del diámetro promedio de las máscaras disponibles. No obstante, se ha observado que esta decisión de diseño no ha afectado a los resultados, puesto que todas las ROI extraídas para esta base de datos se corresponden con la zona correcta alrededor de la mácula

Aunque no haya constituido la fase central de este trabajo, el preprocesado desempeñó un papel fundamental al eliminar información redundante y preparar las imágenes para la extracción de características. En la Figura 5.1 se puede observar los diferentes ROIS correspondiente a una imagen de canal verde e imagen RGB.

5.3 EXTRACCIÓN DE CARACTERÍSTICAS

Durante la fase de extracción de características se han estudiado diferentes tipos de características extraídas de las imágenes. Los histogramas de color y los estadísticos texturales son ampliamente utilizados en la literatura, tanto para la detección de DMAE como de otras enfermedades (Khalid et al. 2021; Phan et al. 2016). Por su parte, el método HOG ha sido predominantemente empleado en otros contextos como la detección de objetos y personas (Khalid et al. 2021). Por tanto, hasta donde se tiene conocimiento, su aplicación en la detección de la DMAE resulta innovador en este campo.

Los histogramas de color en el espacio RGB son utilizados habitualmente como descriptores (Phan et al. 2016). Además, en este estudio se han empleado también los histogramas en el espacio de color $L^*a^*b^*$. Las características pertenecientes a estos histogramas de color han sido muy relevantes para la tarea de clasificación puesto que hasta un 58 % de las características seleccionadas por el algoritmo FCBF pertenecen a los histogramas de color en ambos espacios (RGB y $L^*a^*b^*$). Es interesante destacar que 8 de las 11 características seleccionadas se basan en el espacio $L^*a^*b^*$, Esto puede ser debido a que el color amarillento de las drusas resulta muy relevante para detectar la presencia o ausencia de DMAE.



(a)



(b)

Figura 5. 1: Ejemplo ROIS; (a) Canal verde; (b) RGB

Por su parte, los estadísticos basados en textura son frecuentemente utilizados el ámbito de la DMAE (Khalid et al. 2021; Phan et al. 2016). En este estudio se empleó como descriptor principal de textura el método GLCM. A partir de este, se calcularon cuatro medidas de textura para cada uno de los canales: energía, homogeneidad, contraste y correlación. A pesar de que estas medidas han sido empleadas en estudios previos en el contexto de la DAME, en este TFG únicamente se seleccionaron 2 de las 19 características extraídas. Esto indica que, al emplearse en combinación con otro tipo de características, el poder descriptivo de las medidas texturales ha resultado ser menor en comparación con el resto.

Finalmente, el método HOG se basa en la descripción de un objeto dentro de una imagen mediante la distribución de los gradientes de intensidad o las direcciones de sus bordes (Zhou & Yu, 2021). Uno de los principales problemas relacionados con este método es su alto coste computacional. Aunque el tamaño de las imágenes se redujo mediante la extracción de los ROIS, las dimensiones seguían siendo relativamente grandes (950 x 950 píxels). Para ponerlo en perspectiva, en la literatura (Khalid et al., 2021) empleaba imágenes de 224 x 224 píxeles, un tamaño considerablemente menor al propuesto en nuestro estudio. Otro de los problemas asociados a este método es la gran cantidad de características extraídas para cada imagen. En total, se obtuvieron 67820 características. Para integrar estas características en los modelos de *Machine Learning*, fue necesario realizar un proceso de empaquetado basándonos en el estudio de Khalid et al (Khalid et al., 2021). Como se mencionó en el capítulo 3, se llevaron a cabo dos diferentes procesos. Por un lado, se llevó a cabo la división en histogramas de 32 *bins* de las características para capturar, de esta manera, la frecuencia de aparición de las intensidades de gradiente (Khalid et al., 2021). Por otra parte, se realizó la extracción de estadísticas texturales de los diferentes HOG, hasta donde se tiene constancia esto no se empleó en ningún estudio hasta la fecha. A pesar de las dificultades encontradas en la implementación de esta técnica, el algoritmo de selección de características eligió 5 de las 19 características finales pertenecientes a los HOG. Esto representa aproximadamente el 26% del total de las características, otorgándole un gran poder descriptivo a este estadístico.

5.4 RESULTADOS DE LA FASE DE ENTRENAMIENTO

Durante la fase de entrenamiento se implementó *10-fold cross validation* con el propósito de medir la robustez y reproducibilidad de los resultados obtenidos (Vizcaíno-Salazar, 2017). En relación con la precisión del sistema en el conjunto de validación, se puede observar que el modelo *Ensemble* logró alcanzar un 91.2%. Los datos de precisión indican que se consiguió clasificar de manera correcta la mayoría de las imágenes durante esta fase. Al analizar la Tabla 4.3, se puede observar como el modelo *Ensemble* presenta una sensibilidad del 92.7% y una especificidad del 89.7%. Como indican los resultados, estas dos métricas se encuentran notablemente balanceadas. Esto significa que el modelo tiene una buena capacidad para identificar tanto los casos negativos como positivos. En cuanto al VPP y VPN se obtuvieron resultados del 90 y 92,5 %, respectivamente. Esto indica que el modelo tiende a acertar tanto cuando indica que una imagen pertenece a un paciente sano como cuando indica que una imagen pertenece a un paciente enfermo. Por su parte, el AUC obtenido en el caso del modelo *Ensemble* fue de 0.96, lo que indica que el

modelo presenta una capacidad discriminativa prometedora. Por último, se alcanzó un F1 de 91.33%, lo que significa que el rendimiento combinado de la sensibilidad y el VPP es satisfactorio.

5.5 RESULTADOS SOBRE EL CONJUNTO DE TEST

En cuanto a los resultados en la fase de test. La precisión del sistema alcanzó un 81%. Esto indica que el sistema conseguía clasificar correctamente la mayoría de las imágenes empleadas para esta fase de test. Al examinar la tabla 4.4 se puede contemplar como el sistema presenta una sensibilidad del 55.1 % y una especificidad del 91.5 %. En este caso dichas métricas no se encuentran en absoluto balanceadas. Esto puede deberse a la notable disminución sufrida por parte de la sensibilidad. En cuanto al VPP y VPN se obtuvieron resultados de 73.7 % y 83,3 %, respectivamente, dando lugar a una notable diferencia entre las dos métricas. Por su parte, el AUC obtenido fue de 0.86, lo que muestra que el modelo continúa teniendo una destacada capacidad discriminativa. En último término, cabe destacar que se obtuvo un F1 de 62.6 %.

En los resultados de test puede observarse una disminución general en el rendimiento de nuestro sistema de detección, en comparación con los resultados de entrenamiento. Nuestro modelo experimentó una reducción de más de un 9% en su precisión. Estos indicadores no son favorables, ya que sugieren que es posible que el modelo haya sufrido, de algún modo, un sobreajuste. Dicho sobreajuste puede deberse a diferentes motivos. En primer lugar, es preciso destacar que, la cantidad de datos con la que se trabajó fue menor de la esperada. Únicamente se dispuso de 394 imágenes para la consecución del sistema, de las cuales 100 se seleccionaron para el testeo y el resto para el entrenamiento del modelo. Por otro lado, como se comentó en el capítulo 3, el ajuste de hiperparámetros se llevó a cabo mediante 100 iteraciones del sistema optimizador de la aplicación "Classification Designer" de MATLAB®. Este entrenamiento pudo haber sobreajustado los hiperparámetros para el conjunto de entrenamiento disminuyendo el sesgo, pero aumentando la varianza del modelo. Es importante destacar la disminución en la sensibilidad del modelo sobre el conjunto de test. Esto indica que nuestro sistema presenta dificultades para identificar correctamente a los nuevos sujetos enfermos. Una posible explicación de esta caída tan abrupta podría recaer en la distribución irregular de los sujetos en el conjunto de test. De hecho, únicamente 29 de los 100 sujetos en este conjunto presentaban DMAE. Por lo tanto, el margen de error se reduce considerablemente. A pesar de esta notable reducción en la sensibilidad, el modelo presenta una especificidad más que notable. Del mismo modo ocurre en el caso de los valores predictivos, el modelo presenta muchos fallos a la hora de predecir aquellos sujetos positivos mientras que mantiene su fiabilidad en el caso de los sujetos negativos.

Debido a que los resultados obtenidos en el conjunto de test fueron inferiores a los de entrenamiento, se analizó cada imagen perteneciente a este conjunto de datos. De esta manera, se descubrió que algunas imágenes eran inadecuadas para el análisis automático debido a su baja calidad y falta de contraste.

5.5 COMPARACIÓN CON ESTUDIOS PREVIOS

Se ha realizado una evaluación comparativa de los resultados obtenidos en este TFG con los obtenidos en estudios previos. En la Tabla 5.1 se muestran los resultados obtenidos en estudios previos junto a los resultados obtenidos en este estudio. Cabe destacar que ninguno de los estudios analizados empleó la misma base de datos que este TFG. Por ello, es necesario ser cautos a la hora de comparar los diferentes resultados. Además, tampoco se emplearon las mismas métricas en la medición de resultados (Acharya et al. 2017 ;García-Floriano et al., 2019). Se puede observar que la mayoría de los estudios utilizan la precisión como medida principal de los resultados. Asimismo, muchos estudios emplean las métricas de sensibilidad, especificidad y VPP (Acharya et al. 2017 ;García-Floriano et al., 2019). En este TFG se ha optado por complementar estas medidas con el empleo de curvas ROC y métricas como el AUC, la matriz de confusión y F1. De esta forma se puede ofrecer una visión más detallada del rendimiento del algoritmo propuesto.

En la Tabla 5.1 se muestra que la precisión alcanzada por el método propuesto en este TFG es similar a la obtenida en estudios previos. Si se realiza una comparación más al detalle de los diferentes trabajos presentes en la literatura, se puede observar que el estudio realizado por García-Floriano et al utilizando su base de datos original y sin eliminar las imágenes de baja calidad logró un 83.4% de precisión (García-Floriano et al., 2019). Sin embargo, después de aplicar una fase de cribado y eliminar las imágenes de baja calidad, el rendimiento ascendió hasta el 93.2% de precisión. Es importante tener en cuenta que, en primera instancia, nuestro modelo consiguió mejorar los resultados de García-Floriano et al (García-Floriano et al., 2019). Por ello, la implementación de una fase de cribado de las imágenes de baja calidad puede ser una interesante línea de investigación futura para mejorar los resultados obtenidos. Por otro lado, el estudio realizado por Acharya et al se enfocó especialmente en la fase de selección de características (Acharya et al., 2017). En ella se implementaron dos tipos de algoritmos de selección altamente sofisticados basados en la inteligencia de enjambre: ACO-GA y PSO. A pesar de que los resultados obtenidos en el estudio de Acharya et al fueron ligeramente inferiores y que se considera que el algoritmo de selección utilizado en este TFG (FCBF) es el más adecuado en este caso, es probable que una mayor investigación en la fase de selección de características pueda mejorar el rendimiento global del sistema (Acharya et al., 2017). Los estudios (Acharya et al., 2016), (Mookiah, Acharya, Koh, Chua, et al., 2014) y (Khalid et al., 2021) presentaron resultados superiores en términos de precisión en comparación con el estudio propuesto. Por su parte, los estudios de Phan et al y Mookiah et al mostraron resultados ligeramente inferiores en cuanto a precisión a los que se detallan en este TFG (Phan et al. 2016; Mookiah, Acharya, Koh, Chandran, et al. 2014). En conclusión, basándonos en los resultados obtenidos, se puede afirmar que el método propuesto tiene potenciales aplicaciones prácticas en entornos clínicos al facilitar la detección automática de la DMAE.

Otro aspecto importante a subrayar es el hecho de que la diferencia entre sensibilidad y especificidad de este estudio es menor a la presente en los mencionados en la literatura. Esto significa que el modelo propuesto presenta una eficacia más que notable tanto para la detección de pacientes enfermos como sanos. Por ello, se puede afirmar que nuestro modelo no presenta ningún tipo de sesgo hacia una de las clases y, por tanto, presenta una capacidad efectiva para llevar a cabo ambas tareas. Por otro lado, es preciso resaltar que

Capítulo 5. Discusión

aquellos métodos que presentaban un mejor rendimiento que el propuesto trabajaron con bases de datos de mayor tamaño, lo que supone una desventaja para el entrenamiento.

Autor/res, año	Descripción breve del método	Resultados sobre el conjunto de validación			
		Precisión	VPP	Sensibilidad	Especificidad
(García-Florian et al., 2019)	Reconocimiento de patrones a partir de los momentos de Hu.	92.15%	93.21%	-	-
(Acharya et al., 2017)	Cálculo de características a partir de las pirámides de histograma de gradiente orientados.	85.12%	91.24%	87.24%	80.09%
(Acharya et al., 2016)	Transformada Wavelet discreta.	96.89%	95.28%	100%	91.67%
(Phan et al., 2016)	Estadísticos texturales, de color y contexto visual.	75.6%	-	-	-
(Mookiah, Acharya, Koh, Chua, et al., 2014)	Transformada Wavelet discreta	93.7%	-	-	-
(Mookiah, Acharya, Koh, Chandran, et al., 2014)	Dimensión fractal y Wavelet de Gabor	90.19%	91.28%	88.89%	91.48%
(Khalid et al., 2021)	Estadísticos como LBP, HOG e histogramas de intensidad.	92.34%	-	93.3%	92%
Método propuesto en este TFG	Características basadas en histogramas de color, estadísticos texturales y HOG.	91.2%	90%	92.7%	89.7%

Tabla 5. 1: Resultados conjuntos de otros autores y nuestro sistema sobre el conjunto de validación

CAPÍTULO 6. CONCLUSIÓN

6.1 INTRODUCCIÓN

En este capítulo se exponen las conclusiones referentes al TFG. Tras la exposición en los capítulos anteriores del método propuesto, los resultados alcanzados y su correspondiente discusión, se detallan a continuación las principales contribuciones y conclusiones extraídas del estudio, así como las líneas futuras de investigación.

6.2 CONTRIBUCIONES ORIGINALES

En este TFG se ha introducido un sistema de análisis automático de retinografías para la ayuda al diagnóstico de la DMAE. Por ello, la utilización de retinografías para su detección resulta valiosa teniendo en cuenta el bajo costo y la naturaleza no invasiva de esta modalidad de imagen.

En este trabajo se ha desarrollado y evaluado un modelo de clasificación basado en *Machine Learning*. Entre las principales contribuciones de este estudio, se pueden resaltar las siguientes:

1. Implementación de técnicas de extracción de características basadas en histogramas de color, estadísticos texturales y HOG. En este sentido, cabe destacar que, hasta donde tenemos conocimiento, estas características no se han empleado de forma conjunta hasta ahora para la caracterización de la DMAE.
2. Desarrollo de una etapa de selección de características basada en el algoritmo FCBF. Hasta donde se tiene constancia, este es el primer estudio relacionado con la detección automática de la DMAE en el que se emplea este método de selección de características todas estas métricas para la evaluación de los resultados.
3. Evaluación de los resultados obtenidos mediante la utilización de diversas métricas: precisión, sensibilidad, especificidad, VPP, VPN, curvas ROC, F1 y matrices de confusión. Hasta donde tenemos conocimiento, este es el primer estudio relacionado con la detección automática de la DMAE en el que se emplea tal cantidad de métricas para la evaluación de los resultados.

6.3 CONCLUSIONES

En este TFG se ha afrontado la tarea de llevar a cabo un sistema automatizado de detección de DMAE en retinografías. Los resultados obtenidos indican que el método propuesto es apropiado para la detección de la enfermedad y que puede suponer un apoyo

significativo para la práctica clínica de los oftalmólogos, ayudando en la reducción de costes y reduciendo el tiempo necesario para obtener un diagnóstico.

Como conclusiones del estudio se pueden indicar las siguientes:

1. La implementación de una fase de preprocesado consigue una mejora sustancial en los resultados obtenidos por el modelo.
2. La combinación propuesta de las características basadas en color, textura y HOG es adecuada para la caracterización de las imágenes en el contexto de la DMAE.
3. Debido al gran número de características extraídas, La fase de selección de características es imprescindible para poder llevar a cabo la realización del entrenamiento del modelo.
4. La metodología de evaluación del modelo empleada en este estudio brinda una visión muy completa de su rendimiento, ya que se consideran una mayor cantidad de parámetros de evaluación.
5. Los resultados obtenidos son satisfactorios y acorde con los expresados en la literatura. A pesar de ello, es necesario seguir investigando para tratar de implementar mejoras en el rendimiento del modelo, especialmente en la fase de preprocesado, selección de características y en la detección sobre el conjunto de test.

6.4 LIMITACIONES Y LÍNEAS FUTURAS DE INVESTIGACIÓN

Aunque los resultados alcanzados son satisfactorios y equiparables a los de investigaciones anteriores, este estudio no está exento de limitaciones que es necesario mencionar:

1. La principal limitación del presente estudio es el reducido tamaño de la BD con la que se ha trabajado. Con el objetivo de generalizar las conclusiones del TFG realizado, sería conveniente completar la base de datos con un mayor número de imágenes. Asimismo, sería conveniente que estas nuevas imágenes reflejasen, en la medida de lo posible, la variabilidad existente en la práctica clínica. Es decir, sería deseable emplear imágenes de distintos retinógrafos, anotadas por distintos especialistas y capturadas en diferentes lugares.
2. Otra de las limitaciones relacionadas con la base de datos es el desbalanceo existente entre las dos clases. La clase predominante (imágenes sin DMAE) representaba aproximadamente el 78% de las imágenes totales. Por su parte, únicamente 89 de las 400 imágenes empleadas pertenecían a la clase minoritaria (imágenes con DMAE).
3. Los resultados sobre el conjunto de test son menores que durante el entrenamiento, especialmente en lo que se refiere a la sensibilidad. Esto puede deberse a que exista un sobreajuste del modelo por el reducido tamaño de la base de datos. Asimismo, el desbalanceo presente en el conjunto de test también ha podido influir en los resultados finales.

Considerando estas limitaciones, se proponen algunas líneas futuras de investigación enfocadas en efectuar ciertas modificaciones sobre el método propuesto. En este contexto, se plantea:

1. Aumentar el número de imágenes de la BD con el fin de evaluar la utilidad diagnóstica del método desarrollado en imágenes que presenten un mayor variabilidad en términos de color, contraste y calidad. Además, sería sumamente beneficioso que la base de datos se encuentre más balanceada.
2. Realizar ciertas modificaciones en el método propuesto con el objetivo de intentar mejorar los resultados obtenidos hasta el momento. Una manera de lograr esto es emplear métodos más complejos en la etapa de preprocesado, añadir una etapa de cribado, mejorar la etapa de selección de características, evaluar diferentes métodos de selección de características y otros modelos de clasificación. En este sentido se podrían explorar algoritmos basados en la inteligencia de enjambre, en el caso de la selección de características. Finalmente, se plantea el empleo de técnicas de *Deep Learning* para la detección de DMAE en retinografías.
3. Integración del modelo desarrollado en un sistema completo de análisis de imágenes de fondo de ojo como herramienta de diagnóstico asistido por ordenador (CAD, *Computer Aided Diagnosis*) aplicado al cribado y diagnóstico de la DMAE.

REFERENCIAS

- Abramoff, Michael D., Mona K. Garvin, and Milan Sonka. 2010. "Retinal Imaging and Image Analysis." *IEEE Reviews in Biomedical Engineering* 3: 169–208.
- Acharya, U. R., Mookiah, M. R. K., Koh, J. E., Tan, J. H., Noronha, K., Bhandary, S. V., ... & Laude, A. (2016). Novel risk index for the identification of age-related macular degeneration using radon transform and DWT features. *Computers in biology and medicine*, 73, 131-140
- Acharya, U. R., Hagiwara, Y., Koh, J. E., Tan, J. H., Bhandary, S. V., Rao, A. K., & Raghavendra, U. (2017). Automated screening tool for dry and wet age-related macular degeneration (ARMD) using pyramid of histogram of oriented gradients (PHOG) and nonlinear features. *Journal of Computational Science*, 20, 41-51.
- Al-Bander, B., Al-Nuaimy, W., Williams, B. M., & Zheng, Y. (2018). Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. *Biomedical Signal Processing and Control*, 40, 91-101.
- Arriagada, Miguel. 2015. "Comparación de Métricas de Distancia En El Algoritmo K-Vecinos Más Cercanos Para El Problema de Reconocimiento Automático de Dígitos Manuscritos." : 1–40. http://opac.pucv.cl/pucv_txt/txt-3000/UCD3128_01.pdf.
- Behar-Cohen, Francine, Emmanuelle Gelizé, Laurent Jonet, and Patricia Lassiaz. 2020. "Anatomie de La Rétine." *médecine/sciences* 36(6–7): 594–99.
- Bennett, T. J., & Barry, C. J. (2009). Ophthalmic imaging today: an ophthalmic photographer's viewpoint—a review. *Clinical & experimental ophthalmology*, 37(1), 2-13.
- Bharati, M. H., Liu, J. J., & MacGregor, J. F. (2004). Image texture analysis: methods and comparisons. *Chemometrics and intelligent laboratory systems*, 72(1), 57-71.
- Bhuiyan, A., Wong, T. Y., Ting, D. S. W., Govindaiah, A., Souied, E. H., & Smith, R. T. (2020). Artificial intelligence to stratify severity of age-related macular degeneration (AMD) and predict risk of progression to late AMD. *Translational vision science & technology*, 9(2), 25-25.
- Chaki, J., & Dey, N. (2020). *Texture feature extraction techniques for image recognition*. Springer Singapore.
- Chandra, Mayank Arya, and S. S. Bedi. 2021. "Survey on SVM and Their Application in Image Classification." *International Journal of Information Technology* 13(5): 1–11.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16: 321–57.

Dalal, Navneet, and Bill Triggs. 2005. "Histograms of Oriented Gradients for Human Detection." *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* 1: 886–93.

Dey, Nilanjan, Amira S. Ashour, and Surekha Borra. 2018. 26 Lecture Notes in Computational Vision and Biomechanics *Classification in BioApps*.

Dong, Li, Qiong Yang, Rui Heng Zhang, and Wen Bin Wei. 2021. "Artificial Intelligence for the Detection of Age-Related Macular Degeneration in Color Fundus Photographs: A Systematic Review and Meta-Analysis." *EClinicalMedicine* 35: 100875.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14, 241-258.

Fang, H., Li, F., Fu, H., Sun, X., Cao, X., Lin, F., ... & Xu, Y. (2022). Adam challenge: Detecting age-related macular degeneration from fundus images. *IEEE Transactions on Medical Imaging*, 41(10), 2828-2847.

Ferris, Frederick L. et al. 2005. "A Simplified Severity Scale for Age-Related Macular Degeneration: AREDS Report No. 18." *Archives of Ophthalmology* 123(11): 1570–74.

Fritsche, Lars G. et al. 2014. "Age-Related Macular Degeneration: Genetics and Biology Coming Together." *Annual Review of Genomics and Human Genetics* 15: 151–71.

Fu, Huazhu et al. 2020. "AGE Challenge: Angle Closure Glaucoma Evaluation in Anterior Segment Optical Coherence Tomography." *Medical Image Analysis* 66: 101798.

Fushiki, Tadayoshi. 2011. "Estimation of Prediction Error by Using K-Fold Cross-Validation." *Statistics and Computing* 21(2): 137–46. <http://link.springer.com/10.1007/s11222-009-9153-8>.

García-Floriano, A., Ferreira-Santiago, Á., Camacho-Nieto, O., & Yáñez-Márquez, C. (2019). A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images. *Computers & Electrical Engineering*, 75, 218-229.

GeethaRamani, R., and Lakshmi Balasubramanian. 2016. "Retinal Blood Vessel Segmentation Employing Image Processing and Data Mining Techniques for Computerized Retinal Image Analysis." *Biocybernetics and Biomedical Engineering* 36(1): 102–18.

Govindaiah, Arun, Roland Theodore Smith, and Alauddin Bhuiyan. 2018. "A New and Improved Method for Automated Screening of Age-Related Macular Degeneration Using Ensemble Deep Neural Networks." In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 702–5.

van Grinsven, M. J., Lechanteur, Y. T., van de Ven, J. P., van Ginneken, B., Hoyng, C. B., Theelen, T., & Sánchez, C. I. (2013). Automatic drusen quantification and risk assessment of age-related macular degeneration on color fundus images. *Investigative ophthalmology & visual science*, 54(4), 3019-3027

Herrera-Romero, Bryan et al. 2021. "Procesamiento y Análisis de Imágenes Digitales Para El Diagnóstico de Enfermedades Oftalmológicas: Análisis Descriptivo Actual de La Investigación Científica." : 658–75.

Herrero Tudela, M. (2021). Clasificación automática de la severidad de la retinopatía diabética mediante técnicas de Deep Learning.

Houssein, Alexandros et al. 2019. "BMC Biomedical Engineering: A Home for All Biomedical Engineering Research." *BMC Biomedical Engineering* 1(1): 1–4.

Huang, David et al. 1991. "Optical Coherence Tomography." *Science* 254(5035): 1178–81.

Keel, Stuart et al. 2019. "Development and Validation of a Deep-learning Algorithm for the Detection of Neovascular Age-related Macular Degeneration from Colour Fundus Photographs." *Clinical & Experimental Ophthalmology* 47(8): 1009–18.

Khalid, Samina et al. 2021. "Automated Diagnosis System for Age-Related Macular Degeneration Using Hybrid Features Set from Fundus Images." *International Journal of Imaging Systems and Technology* 31(1): 236–52.

Khalid, Samina, M. Usman Akram, and Tehmina Khalil. 2017. "Hybrid Textural Feature Set Based Automated Diagnosis System for Age Related Macular Degeneration Using Fundus Images." *Proceedings of 2017 International Conference on Communication, Computing and Digital Systems, C-CODE 2017*: 390–95.

Khan, Kamron N. et al. 2016. "Differentiating Drusen: Drusen and Drusen-like Appearances Associated with Ageing, Age-Related Macular Degeneration, Inherited Eye Disease and Other Pathological Processes." *Progress in Retinal and Eye Research* 53: 70–106.

Khan, Mohd Yaqub, Poonam Gupta, and Vikas Kumar Verma. 2018. "A Review-Biomedical Engineering-Present and Future Prospective A Review- Biomedical Engineering-Present and Future Prospective." *Asian journal Pharma* 3(4): 202–6.

Klein, Ronald et al. 2006. "Prevalence of Age-Related Macular Degeneration in 4 Racial/Ethnic Groups in the Multi-Ethnic Study of Atherosclerosis." *Ophthalmology* 113(3): 373–80.

Kumar, V., Surve, A., Kumawat, D., Takkar, B., Azad, S., Chawla, R., ... & Venkatesh, P. (2021). Ultra-wide field retinal imaging: A wider clinical perspective. *Indian Journal of Ophthalmology*, 69(4), 824. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM computing surveys (CSUR)*, 50(6), 1-45.

Martínez Pérez, J.A., y P.S. Pérez Martín. 2023. «La curva ROC». *Medicina de Familia. SEMERGEN* 49(1): 101821. <https://linkinghub.elsevier.com/retrieve/pii/S1138359322001952>.

Mitchell, Paul, Gerald Liew, Bamini Gopinath, and Tien Y. Wong. 2018. "Age-Related Macular Degeneration." *The Lancet* 392(10153): 1147–59.

Mookiah, Muthu Rama Krishnan, U. Rajendra Acharya, Joel E.W. Koh, Vinod Chandran, et al. 2014. "Automated Diagnosis of Age-Related Macular Degeneration Using Greyscale Features from Digital Fundus Images." *Computers in Biology and Medicine* 53: 55–64. <https://linkinghub.elsevier.com/retrieve/pii/S0010482514001802>.

Mookiah, Muthu Rama Krishnan, U. Rajendra Acharya, Joel E.W. Koh, Chua Kuang Chua, et al. 2014. "Decision Support System for Age-Related Macular Degeneration Using Discrete Wavelet Transform." *Medical and Biological Engineering and Computing* 52(9): 781–96.

Musa, Purnawarman, Farid Al Rafi, and Missa Lamsani. 2018. "A Review: Contrast-Limited Adaptive Histogram Equalization (CLAHE) Methods to Help the Application of Face Recognition." In *2018 Third International Conference on Informatics and Computing (ICIC)*, IEEE, 1–6.

Nalepa, Jakub, and Michal Kawulok. 2019. "Selecting Training Sets for Support Vector Machines: A Review." *Artificial Intelligence Review* 52(2): 857–900.

El Naqa, Issam et al. 2018. "Machine Learning and Modeling: Data, Validation, Communication Challenges." *Medical Physics* 45(10): e834–40. <https://onlinelibrary.wiley.com/doi/10.1002/mp.12811>.

Nentwich, Martin M. 2015. "Diabetic Retinopathy - Ocular Complications of Diabetes Mellitus." *World Journal of Diabetes* 6(3): 489.

Organisation, World health. 2019. 214 World health Organisation *Informe Mundial Sobre La Visión*.

Orlando, José Ignacio et al. 2020. "REFUGE Challenge: A Unified Framework for Evaluating Automated Methods for Glaucoma Assessment from Fundus Photographs." *Medical Image Analysis* 59: 101570.

Pead, E., Megaw, R., Cameron, J., Fleming, A., Dhillon, B., Trucco, E., & MacGillivray, T. (2019). Automated detection of age-related macular degeneration in color fundus photography: a systematic review. *survey of ophthalmology*, 64(4), 498-511.

Phan, Thanh Vân, Lama Seoud, Hadi Chakor, and Farida Cheriet. 2016. "Automatic Screening and Grading of Age-Related Macular Degeneration from Texture Analysis of Fundus Images." *Journal of Ophthalmology* 2016.

Ponce-Rivera, Enrique et al. 2017. "Procesamiento y Análisis de Imágenes Digitales de Fondo de Ojo Para Facilitar El Tele-Diagnóstico Oportuno." *Artículo Revista de Cómputo Aplicado Septiembre* 1(3): 22–37.

Presutti, M. (2004). La matriz de co-ocurrencia en la clasificación multispectral: tutorial para la enseñanza de medidas texturales en cursos de grado universitario. 4ª Jornada de Educação em Sensoriamento Remoto no Âmbito do Mercosul, 1-9.

Remeseiro, Beatriz, and Veronica Bolon-Canedo. 2019. "A Review of Feature Selection Methods in Medical Applications." *Computers in Biology and Medicine* 112: 103375.

Romero-Oraá, R., García, M., Oraá-Pérez, J., López, M. I., & Hornero, R. (2020). A robust method for the automatic location of the optic disc and the fovea in fundus images. *Computer Methods and Programs in Biomedicine*, 196, 105599.

Rocca, J. (2019). Ensemble methods: bagging, boosting and stacking. *Towards Data Science*, 5, 410.

Russ, John C., James R. Matey, A. John Mallinckrodt, and Susan McKay. 1994. "The Image Processing Handbook." *Computers in Physics* 8(2): 177.

Sakata, Lisandro M, Julio DeLeon-Ortega, Viviane Sakata, and Christopher A Girkin. 2009. "Optical Coherence Tomography of the Retina and Optic Nerve - a Review." *Clinical &*

Experimental Ophthalmology 37(1): 90–99.
<https://onlinelibrary.wiley.com/doi/10.1111/j.1442-9071.2009.02015.x>

Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M., & Bogunović, H. (2018). Artificial intelligence in retina. *Progress in retinal and eye research*, 67, 1-29.

Schmitz-Valckenberg, S., Holz, F. G., Bird, A. C., & Spaide, R. F. (2008). Fundus autofluorescence imaging: review and perspectives. *Retina*, 28(3), 385-409.

Shen, S. Y., Wong, T. Y., Foster, P. J., Loo, J. L., Rosman, M., Loon, S. C., ... & Aung, T. (2008). The prevalence and types of glaucoma in Malay people: the Singapore Malay eye study. *Investigative ophthalmology & visual science*, 49(9), 3846-3851.

Sokolova, Marina, Nathalie Japkowicz, y Stan Szpakowicz. 2006. «Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation». *AAAI Workshop - Technical Report WS-06-06*(June 2014): 24-29.

Sonali et al. 2019. "An Approach for De-Noising and Contrast Enhancement of Retinal Fundus Image Using CLAHE." *Optics & Laser Technology* 110: 87–98.

Stahl, A. (2020). The diagnosis and treatment of age-related macular degeneration. *Deutsches Ärzteblatt International*, 117(29-30), 513.

Tan, Jen Hong et al. 2018. "Age-Related Macular Degeneration Detection Using Deep Convolutional Neural Network." *Future Generation Computer Systems* 87: 127–35.

Teus, Miguel Ángel, Esther Arranz-Márquez, Lorenzo López-Guajardo, and Rafael Jiménez-Parras. 2007. "Fondo de Ojo." *Anales de Pediatría Continuada* 5(3): 163–66.

Thomas, Catherine J., Rukhsana G. Mirza, and Manjot K. Gill. 2021. "Age-Related Macular Degeneration." *Medical Clinics of North America* 105(3): 473–91.

Torgo, Luís, Rita P. Ribeiro, Bernhard Pfahringer, and Paula Branco. 2013. "SMOTE for Regression." In , 378–89.

Tsikata, Edem et al. 2017. "Automated Brightness and Contrast Adjustment of Color Fundus Photographs for the Grading of Age-Related Macular Degeneration." *Translational Vision Science & Technology* 6(2): 3.

Varón, Clara, Sergio Jaramillo, and Alejandro Tello. 2010. "La Retina Para El Médico No Oftalmólogo." *MedUNAB* 13(1): 31–37.

Vizcaíno-Salazar, Gilberto J. 2017. «Importancia del cálculo de la sensibilidad, la especificidad y otros parámetros estadísticos en el uso de las pruebas de diagnóstico clínico y de laboratorio». *Medicina y Laboratorio* 23(7-8): 365-86.

Walter, Thomas et al. 2007. "Automatic Detection of Microaneurysms in Color Fundus Images." *Medical Image Analysis* 11(6): 555–66.

Wang, L. (2019, December). Research and implementation of machine learning classifier based on KNN. In *IOP Conference Series: Materials Science and Engineering* (Vol. 677, No. 5, p. 052038). IOP publishing.

Webb, Robert H., George W. Hughes, and Francois C. Delori. 1987. "Confocal Scanning Laser Ophthalmoscope." *Applied Optics* 26(8): 1492.

Winder, R.J. et al. 2009. "Algorithms for Digital Image Processing in Diabetic Retinopathy." *Computerized Medical Imaging and Graphics* 33(8): 608–22.

Wong, Wan Ling et al. 2014. "Global Prevalence of Age-Related Macular Degeneration and Disease Burden Projection for 2020 and 2040: A Systematic Review and Meta-Analysis." *The Lancet Global Health* 2(2): e106–16.

Wu, Jia et al. 2019. "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization." *Journal of Electronic Science and Technology* 17(1): 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>.

Xie, Li et al. 2023. "Automation of Macular Degeneration Classification in the AREDS Dataset, Using a Novel Neural Network Design." *Clinical Ophthalmology* Volume 17: 455–69.

Yu, L., & Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *The Journal of Machine Learning Research*, 5, 1205-1224. Zapata, Miguel Angel et al. 2020. "Artificial Intelligence to Identify Retinal Fundus Images, Quality Validation, Laterality Evaluation, Macular Degeneration, and Suspected Glaucoma." *Clinical Ophthalmology* 14: 419–29.

Zhang, Zhihua. 2018. "Artificial Neural Network." In *Multivariate Time Series Analysis in Climate and Environmental Research*, Cham: Springer International Publishing, 1–35.

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11).

Zhou, Hongzhi, and Gan Yu. 2021. "Research on Pedestrian Detection Technology Based on the SVM Classifier Trained by HOG and LTP Features." *Future Generation Computer Systems* 125: 604–15. <https://linkinghub.elsevier.com/retrieve/pii/S0167739X21002119>.

Zhou, Wei, Shengyu Gao, Ling Zhang, and Xin Lou. 2020. "Histogram of Oriented Gradients Feature Extraction From Raw Bayer Pattern Images." *IEEE Transactions on Circuits and Systems II: Express Briefs* 67(5): 946–50.

Ziyang Liang et al. 2010. "Towards Automatic Detection of Age-Related Macular Degeneration in Retinal Fundus Images." In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, IEEE, 4100–4103.