

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



DETECCIÓN AUTOMATIZADA DE ESTRUCTURAS ANATÓMICAS RETINALES EN RETINOGRAFÍAS DIGITALES

Mitcheel Ariel Lanas Soza

TESIS PARA OPTAR AL GRADO DE MAGISTER EN INFORMÁTICA
MÉDICA.

Director de Tesis: Prof. Dr. Mauricio Cerda Villablanca

Co-Director de Tesis: Prof. Tomás Arenas Cavalli

2022

*A mi familia, mi amada Piera y mis queridos hijos Catalina y Bruno
ellos son mi alegría, mi inspiración y la razón por la cual continúo aprendiendo.*

AGRADECIMIENTOS

Quisiera agradecer a mi Director de Tesis, el profesor Mauricio Cerda, por su apoyo y visión para guiar la investigación en sus etapas iniciales y la claridad para resolver las dudas que surgieron durante las etapas finales del presente trabajo. Asimismo, también agradecer a mi Co-Director, Tomás Arenas Cavalli y al equipo de DART por toda la ayuda proporcionada y su buena disposición procesando la información necesaria para el desarrollo del presente proyecto de tesis.

Agregar un especial agradecimiento a la profesora Dra Eugenia Díaz por su preocupación, guía y orientación que siempre fue más allá del ámbito académico. También quisiera agradecer a la Dra Daniela Faivovich, gracias a su desinteresada colaboración contribuyó con información fundamental para las primeras etapas del proyecto, sin el aporte de ambas mi trabajo no hubiese podido desarrollarse de la misma manera. Quiero agradecer a mis compañeros de cohorte del Magíster en Informática Médica, especialmente a Fabian Villena, Paulina Cubillos, Patricia Menares y José Carrasco por su generosidad, apoyo y dedicación. Su buena disposición y compañía hicieron mucho más agradable este programa académico.

Agradecer profundamente a mi familia, mis abuelos, hermanos y amigos por todo el apoyo brindado, su cariño incondicional me permitió concluir exitosamente esta nueva etapa de formación académica, profesional y personal. Muchas gracias a todos.

Índice

1.- Resumen	4
2.- Abstract	6
3.- Antecedentes	9
4.- Hipótesis y objetivos	33
4.1.- Objetivo General	33
4.2.- Objetivos Específicos	33
5.- Material y Métodos	33
5.1.- Material	34
5.2.- Métodos	41
6.- Resultados	53
7.- Discusión	66
8.- Conclusiones	72
9.- Bibliografía	73
10.- Anexos	77

1.- Resumen

La Retinopatía Diabética (RD) es una microangiopatía progresiva caracterizada por la presencia de lesiones y oclusión de vasos retinales. Se estima que entre un 5% a un 10% de los pacientes que presentan RD pueden desarrollar Maculopatía Diabética (MD), patología que se caracteriza por presentar alteraciones en la agudeza visual causadas por lesiones localizadas en el área macular de la retina. Cabe señalar que la mácula es una región ubicada en el centro de la retina, que corresponde al punto de máxima discriminación visual, por lo tanto las lesiones retinales en esta zona son un factor muy relevante en la clasificación del estado macular.

En nuestro país, el control de pacientes con RD se realiza anualmente mediante la adquisición de imágenes retinales o, en su defecto, mediante un examen de fondo de ojo. Sin embargo, en la última década se ha observado una disminución importante en las atenciones oftalmológicas realizadas de los pacientes con RD. A partir del año 2018, el sistema de Diagnóstico Automatizado de Retinografías Telemáticas (DART) ha contribuido a aumentar el control y acceso oportuno de los pacientes a exámenes oftalmológicos, utilizando métodos de aprendizaje de máquinas, identificando a los pacientes que presentan signos patológicos con sospecha de RD y que deben ser derivados a controles presenciales con especialistas. No obstante, el módulo de inteligencia artificial del sistema DART no contempla dentro de su algoritmo de clasificación la localización y la extensión de las lesiones presentes en la zona macular. Por lo anteriormente descrito, en este trabajo de tesis se propone relacionar la ubicación del área macular con las lesiones retinales, incorporando un clasificador binario del estado macular que contribuya a mejorar la precisión de la clasificación automática de RD realizada actualmente, información que permitirá la priorización de los casos con MD.

Los resultados obtenidos por los algoritmos *Logistic Regression* y *Decision Tree*, utilizando información de la posición macular y lesiones *ground truth* (MGLG), permitieron establecer que la relación de las lesiones retinales con respecto al área macular constituyen el factor más relevante para la determinación del estado macular,

alcanzando un rendimiento de un 100% en *Precision*, *Recall* y *F1-Score* en la clasificación de cada clase de interés (presencia o ausencia de maculopatía).

Con respecto a la tarea de estimar automáticamente la posición macular, el algoritmo que logró mejores resultados entre los propuestos fue el método de aprendizaje de máquinas *U-net*, el que en una muestra de 318 retinografías alcanzó un error acumulado medio de 0.9 μm de distancia del centro real macular o el equivalente a $\frac{1}{5}$ de diámetro de disco (DD) aproximadamente.

Al combinar los resultados anteriores, utilizando la posición macular y las lesiones estimadas por DART, se logró corroborar que los algoritmos *Logistic Regression* y *Support Vector Machine* fueron los clasificadores con mejor desempeño, alcanzando un ROC AUC de 0.9761 y 0.9680 respectivamente, lo que demuestra la factibilidad de identificar MD automáticamente.

2.- Abstract

Diabetic Retinopathy (DR) is a progressive microangiopathy characterized by lesions and occlusion of retinal vessels. It is estimated that between 5% to 10% of patients with DR may develop Diabetic Maculopathy (DM), a pathology characterized by alterations in the visual acuity caused by lesions located in the retinal macular area. It should be noted that the macula is located in the center of the retina and correspond to the point of maximum visual discrimination; therefore, retinal lesions in this area are a very relevant factor in the classification of the macular status.

In our country, the control of patients with DR is performed yearly by acquiring retinal images or, in its absence, a fundus examination. However, in the last decade a significant gap has been observed in the performance of ophthalmologic care of patients with DR. From 2018 onwards, the Diagnostic Automated Retinography Telematics Diagnostics (DART) system has contributed to increasing the control and timely access of patients to ophthalmological examinations using machine learning methods, identifying patients who present pathological signs with suspected DR and who should be referred to on-site controls with specialists. It must be pointed out, however, that the classification algorithm of the artificial intelligence module of DART system does not include the location and extent of the lesions present in the macular area.

Therefore, this thesis proposes to correlate the location of the macular area with the retinal lesions, incorporating a binary classifier of the macular state that contributes to the accuracy of the automatic classification of DR currently performed. Furthermore, obtaining this information will allow the prioritization of cases with MD.

The results obtained by the Logistic Regression and Decision Tree algorithms, using the information on the macular position and ground truth lesions (MGLG), allowed us to establish that the macular area and its relationship with retinal lesions are the most relevant factor for the determination of macular status, reaching the performance of 100% in Precision, Recall and F1-Score in the classification of each class of interest (presence or absence of maculopathy).

For automatically estimating macular position, the algorithm that achieved the best results among the proposed methods was the U-net machine learning method. A sample of 318 retinographies, reached a mean cumulative error of 0.9 μm distance from the actual macular center or the equivalent of approximately of $\frac{1}{5}$ of a disc diameter (DD).

By combining the previous results, using macular position and lesions estimated by DART, this thesis corroborated the Logistic Regression and Support Vector Machine algorithms as the best performing classifiers, reaching a ROC AUC of 0.9761 and 0.9680 respectively, showing the high feasibility of identifying MD automatically.

3.- Antecedentes

En esta sección se describen algunos conceptos relacionados con la epidemiología y fisiopatología de la retinopatía diabética y el impacto visual en pacientes que presentan maculopatía diabética. Se reseñan brevemente las características de los métodos automatizados en el procesamiento de imágenes y aprendizaje de máquinas utilizados en la detección automatizada de la posición macular y se resumen los rendimientos alcanzados por diversos autores en la tarea de detección macular y clasificación de MD automáticamente. En la última subsección se presenta el flujo de atención primaria actualmente utilizado en los pacientes con RD y la implementación de la estrategia DART.

3.1.- Epidemiología e impacto visual causado por la Retinopatía Diabética

La Retinopatía Diabética (RD) es una microangiopatía progresiva que se caracteriza por la presencia de lesiones y oclusión de vasos retinales en pacientes con *Diabetes Mellitus*. A nivel mundial la incidencia de RD se ha cuadruplicado en las últimas décadas. Actualmente se estima que aproximadamente un 40% de los pacientes diabéticos desarrollará en forma concomitante algún grado de alteración retinal [1].

Esta patología corresponde a la principal causa de ceguera adquirida, tanto en Chile como en el mundo. En nuestro país, se garantiza su control anual mediante el programa de Garantías Explícitas en Salud (GES) por medio del examen de fondo de ojo, que consiste en una fotografía de fondo de ojo estandarizada denominada retinografía. No obstante, durante la última década se ha observado una brecha importante en el acceso oportuno a atenciones oftalmológicas. Datos obtenidos del Departamento de Estadísticas e Información de la Salud (DEIS), muestran que de un total de 890.000 prestaciones referidas desde el Programa de Salud Cardiovascular (PSCV) durante el año 2018, tan solo un 37,5% de los pacientes accedieron efectivamente a su evaluación anual de fondo de ojo.

Las lesiones presentes en la RD corresponden a microaneurismas, hemorragias superficiales y exudados duros. Estos tres tipos de lesiones son identificables en las fotografías de fondo de ojo debido a la diferencia de color que presentan: rojo para las hemorragias y microaneurismas, y blanco en los exudados duros o blandos. Dichas

características también permiten diferenciarlas utilizando técnicas de procesamiento de imágenes.

De acuerdo a protocolos de clasificación internacionales tales como “*Early Treatment Diabetic Retinopathy Study*” [2] cuyas recomendaciones son adoptadas en la “Guía Clínica de Retinopatía Diabética” del Ministerio de Salud de nuestro país, la RD puede ser clasificada en 5 estadios: i) Sin presencia de retinopatía diabética, ii) Retinopatía diabética no proliferativa, iii) Retinopatía proliferativa leve a moderada, iv) Retinopatía diabética de alto riesgo y, en paralelo, presencia o ausencia de v) Edema macular clínicamente significativo (EMCS) [3]. El EMCS es una subclasificación de la maculopatía diabética (MD) y de ahora en adelante nos referiremos a ella como términos equivalentes. Ambas patologías pueden progresar independientemente y de manera desigual pero la MD está siempre asociada a RD.

La manifestación del daño retinal depende de la localización y extensión de las lesiones observadas. En la RD es posible encontrar lesiones en toda la extensión del fondo de ojo, existiendo casos en donde las lesiones periféricas no afectan la agudeza visual del paciente. En la MD, por el contrario, las lesiones se encuentran localizadas en el área central de la retina y pueden, por consiguiente, alterar la agudeza visual. Por lo tanto la localización de las lesiones retinianas constituye un factor muy relevante en la clasificación del estado macular [4]–[6].

La importancia de discriminar entre RD y MD reside en que si hay presencia de MD el tratamiento clínico debe ser iniciado de inmediato. Es necesario mencionar que el hallazgo de MD no mejoraría la detección de RD, pero permitiría seleccionar dentro de los resultados positivos de RD los casos más agudos de esta patología, que, como se mencionó, necesitan un procedimiento clínico perentorio, por lo que la importancia de discriminarlos radica en que permite otorgarles un carácter de urgencia para su tratamiento.

Es importante destacar que la progresión de la RD se puede prevenir en buena manera si esta es detectada y tratada oportunamente, dado que su prevalencia está relacionada con el tiempo transcurrido entre el diagnóstico y el tratamiento de la *Diabetes Mellitus*. De lo anterior se desprende que es por lo tanto fundamental la implementación de

estrategias de detección, prevención y tratamiento temprano de esta enfermedad [1], [7].

3.2.- Anatomía retinal

La retina es una estructura del sistema nervioso que forma parte del globo ocular, cuya porción neurosensorial es la encargada de captar y transmitir la información visual a la corteza cerebral visual, localizada en la región occipital del encéfalo. En la retina es posible identificar diversos hitos anatómicos, los que se describen brevemente a continuación:

3.2.1.- Mácula

Es un área de forma elíptica ubicada en el centro del polo posterior de la retina, en cuyo centro se forma una depresión denominada fovea. En esta zona los capilares retinales no están presentes y abarca un área de 0.5 mm de diámetro (3 mm temporal al disco y 8 mm por debajo del meridiano horizontal). Además de esta zona avascular denominada fovea es posible identificar una zona denominada foveola, cuyo tamaño corresponde a 0.2 mm de diámetro y es el punto de máxima discriminación visual en la retina [8]. Es posible distinguirla con respecto a la retina periférica debido a que presenta un color más oscuro que el resto de las estructuras retinales (Ver Figura 1 a).

3.2.2.- Nervio Óptico

Compuesto por alrededor de un millón de axones de células ganglionares retinales. Mide aproximadamente 1.5 mm de diámetro y está conformado por un borde neurorretiniano que corresponde a la proyección de las fibras nerviosas en el nervio óptico y un espacio central que no contiene axones, denominado copa óptica [9]. La relación de tamaño copa-disco es variable y permite dar cuenta de ciertas patologías según se manifieste la pérdida de fibras nerviosas. Su aspecto oval, de coloración más clara que la retina, hace que sea fácilmente distinguible en la imagen de fondo de ojo normal (Ver Figura 1 b).

3.2.3.- Componente Vascular

Constituido por dos ramas arteriales principales, la papilar superior y la papilar inferior, que a su vez se dividen en Temporal (arqueada) y Nasal (recta). El componente vascular está encargado de la nutrición de las capas más internas de la retina y proviene

de la arteria central de la retina, rama de la arteria oftálmica que emerge a través del nervio óptico, figura 1(c). Las anomalías arteriales, como por ejemplo la hipertensión arterial, tienden a afectar a los capilares del plexo de las fibras nerviosas, mientras que la *Diabetes Mellitus* que es una anomalía de tipo venosa, afecta principalmente a los capilares de la capa nuclear interna, tal como se puede observar en la figura 2.

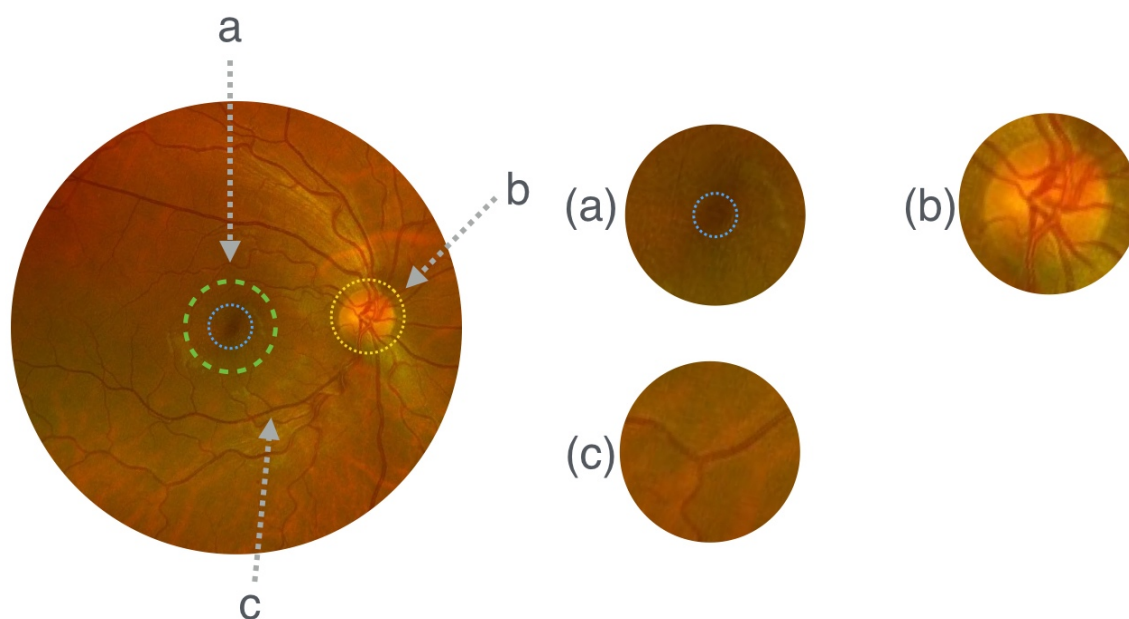


Figura 1. Anatomía retinal normal. Se muestra una retina de aspecto normal, donde es posible identificar la mácula, destacada en el círculo verde y la fovea central, destacada en el círculo azul (a), el nervio óptico (b) y los vasos retinales (c).

3.3.- Relaciones anatómicas entre las estructuras retinales

La localización y tamaño de las estructuras retinales, pueden ser utilizadas para definir ciertas relaciones anatómicas que establecen patrones de distancia y posición entre ellas. Esta estrategia permite localizar de manera más precisa la posición de cada uno de estos hitos morfológicos al verificar complementariamente su posición en el contexto anatómico normal.

3.3.1.- Nervio óptico y posición macular

Es posible identificar la localización relativa de la mácula en una retinografía, tomando la posición del nervio óptico como referencia, debido a que la mácula está localizada en una zona ubicada 5 grados hacia inferior y temporal de esta estructura. El diámetro

del nervio óptico o diámetro de disco óptico (DD) es otra constante que es utilizada en la localización de la zona macular, su tamaño depende del criterio utilizado para cuantificar su diámetro y puede ser expresado en milímetros o en retinografías digitales en píxeles [4].

3.3.2.-Vasos retinales y posición macular

Las arcadas superior e inferior de los vasos retinales se originan en el centro del nervio óptico y dirigen sus ramas casi hasta llegar a la región macular, esta define a la zona avascular foveal (ZAF) que corresponde al centro macular. Es posible identificar la región central utilizando la información de los terminales vasculares que apuntan en dirección a la mácula. Estos puntos pueden ser individualizados y agrupados, generando un centroide que coincide con el área macular.

3.4.- Tipos de lesiones retinales presentes en la Retinopatía Diabética

3.4.1.- Microaneurismas

Los primeros signos oftalmoscópicos de la RD se localizan en la capa nuclear interna de la retina. Estos se observan como pequeños puntos rojos cuyos diámetro oscila entre 15 a 125 micrones. Son de forma redondeada, con bordes lisos bien definidos y se localizan con mayor frecuencia en el lado temporal de la mácula. Se forman mayoritariamente a partir de capilares venosos y en un menor grado de los capilares arteriales (Ver Figura 2 a).

3.4.2.- Hemorragias Intrarretinales

Se producen por la filtración del flujo sanguíneo desde los microaneurismas, capilares o vénulas. Producen un depósito cuya forma depende de su localización en las distintas capas de la retina, pudiendo observarse hemorragias profundas o superficiales (Ver Figura 2 b).

3.4.3.- Exudados Duros

Depósitos blancos o blanco amarillentos (céreos) de tamaño variable y límites irregulares pero precisos. Se localizan entre la capa plexiforme externa y la nuclear interna de la retina y son producto de la filtración activa de los microaneurismas, conformando depósitos extracelulares de lípidos y lipoproteínas. Se observan aislados

o agrupados, con forma de estrella, anillo (parcial o completo) o placas compactas (Ver Figura 2 d) [7].

3.4.4.- Exudados Blandos

Son estructuras que evidencian la ocurrencia de microinfartos retinales. Presentan forma redondeada u ovalada, de color blanco amarillento, de bordes imprecisos, localizados superficialmente en capa de fibras nerviosas, Son causados por una oclusión capilar que determina la interrupción del flujo axoplásmico originado por la isquemia y la acumulación subsiguiente del material transportado. Son visibles por solo 4 a 6 semanas post microinfarto y es el signo patognomónico de isquemia (Ver Figura 2 c) [7].

3.4.5.- Edema Macular o Maculopatía Diabética

La evolución de una lesión retinal depende de su ubicación y extensión, principalmente en la zona macular [10]. En esta región se modifica la estructura compacta de la retina, debido al desplazamiento lateral de sus capas, lo que permite una mejor captación de la luz pero incrementando la susceptibilidad a la acumulación de líquido extracelular.

El edema macular, definido como la presencia de cualquier engrosamiento de la retina o de exudados duros a una distancia menor a un diámetro papilar del centro de la fóvea (1500 micras), es el causante más común de pérdida visual en pacientes con RD (Ver figura 3) [11], [12].

El edema macular clínicamente significativo (EMCS) se define mediante la presencia de :

1. Engrosamiento de retina a menos de 500 micras del centro de la fóvea.
2. Exudados duros a menos de 500 micras de la fóvea asociados a un engrosamiento retiniano adyacente.
3. Engrosamiento retiniano de tamaño igual o superior a 1 diámetro papilar, una de las partes del mismo a menos de 1 diámetro papilar del centro de la fóvea.

3.4.6.- Neovascularización papilar

Por otra parte, el estado del nervio óptico se puede ver afectado debido al desarrollo de neovasos (vascularización papilar). Si esta neovascularización se da en el polo anterior ocular puede incluso afectar el flujo normal del humor acuoso, aumentando el riesgo de desarrollar glaucoma, entre otras afecciones (Ver Figura 2 e).

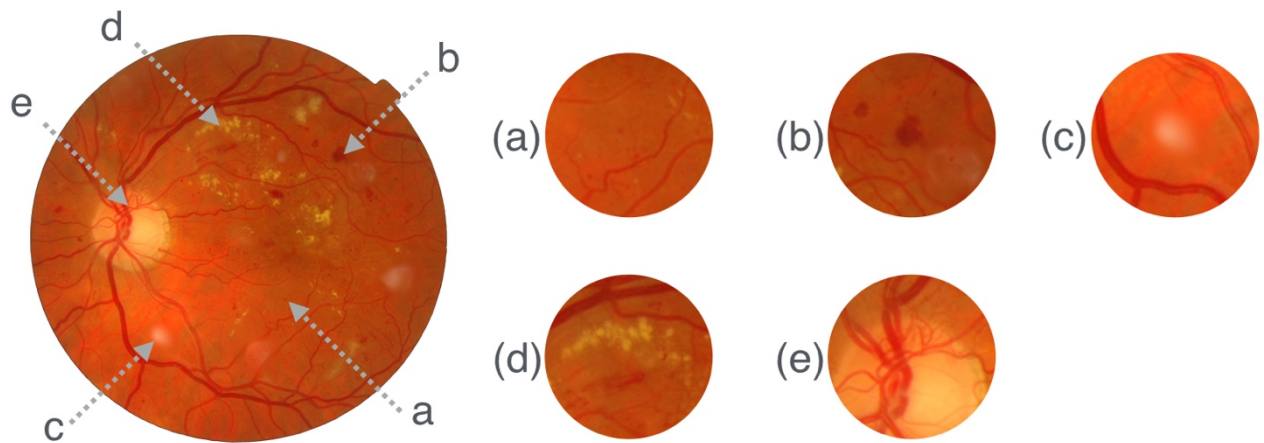


Figura 2. Anatomía retinal mostrando alteraciones propias de la RD. En una retina afectada por RD es posible identificar la presencia de microaneurismas (a), hemorragias (b), exudados blandos (c) y exudados duros (d). Además, en el tipo de RD proliferante es posible identificar desarrollo de neovasos (e).

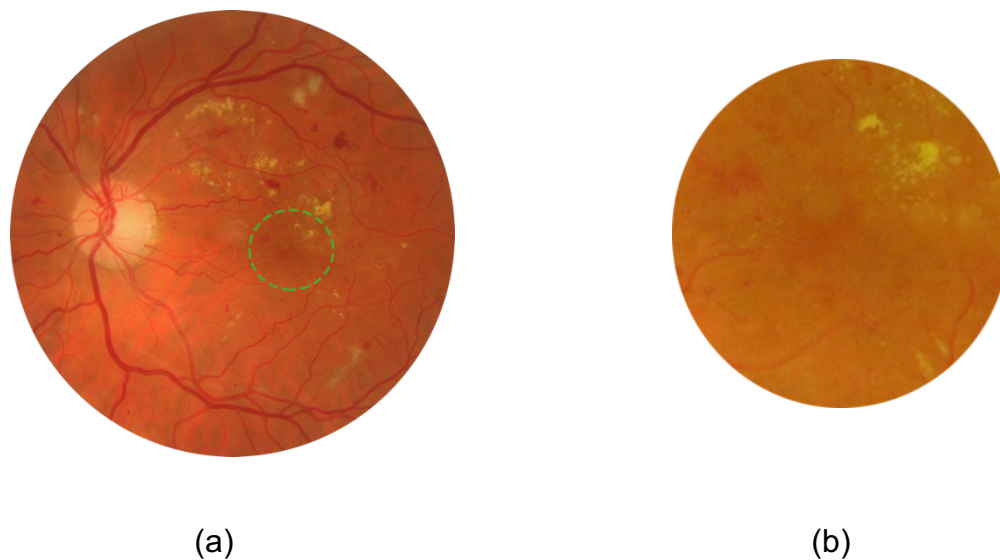


Figura 3. Anatomía retinal mostrando alteraciones propias de la MD. (a) la localización de la zona macular se destaca mediante un círculo verde, (b) Acercamiento de la zona macular afectada por MD se pueden identificar microaneurismas, hemorragias, exudados blandos y exudados duros que alteran la coloración uniforme de la zona macular.

3.5.- Factores que alteran la calidad de la retinografía digital

La calidad de la imagen retinográfica puede verse afectada por dos factores principales:

i) factores anatomopatológicos que tienen relación con las características morfológicas

de la retina en estudio y ii) factores instrumentales o de adquisición, relacionados con los aspectos técnicos de la adquisición de la imagen [13].

3.5.1.- Factores anatomopatológicos

La gran diversidad y tamaños de las lesiones presentes en la RD, pueden alterar de muchas maneras la calidad de la captura de la imagen retinal. Algunas patologías como las cataratas modifican de manera variable la transparencia de los medios refractivos, disminuyendo el contraste de la imagen. En otros casos, patologías tales como atrofas pigmentarias retinales, atrofas del nervio óptico (asociado a altas miopías), casos avanzados de edema macular y hemorragias vítreas, generan contornos difusos en las estructuras retinales afectadas (ver figura 2). El diámetro pupilar también es un factor importante a considerar, dado que su tamaño incide directamente en la obtención de una adecuada iluminación y por consiguiente en la nitidez y amplitud de campo obtenida en la imagen retinal (ver figura 4 a y b).

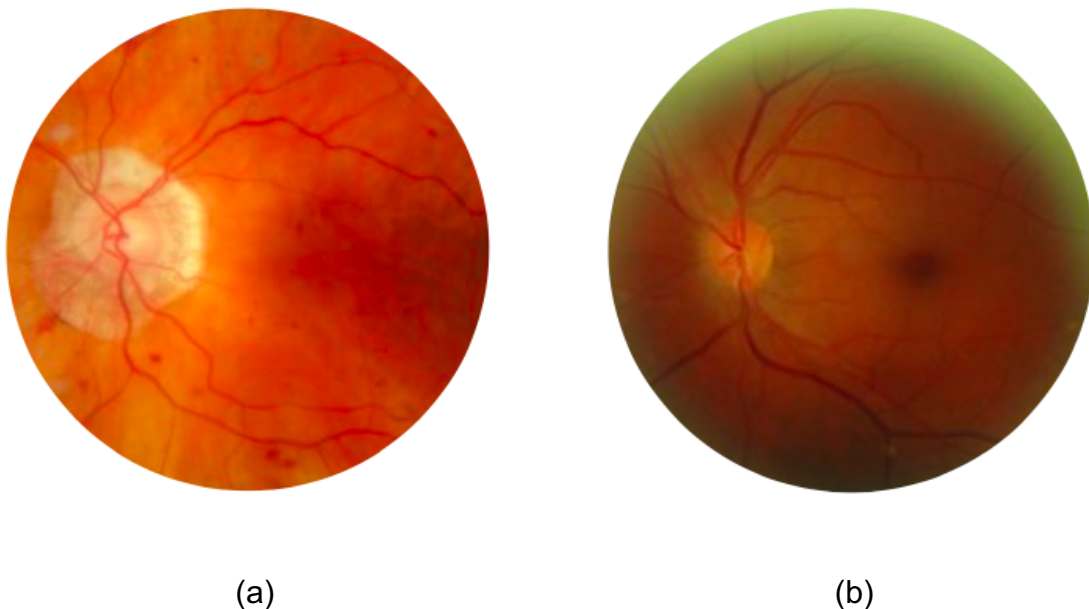


Figura 4. Factores patológicos y fisiológicos que alteran la calidad de las retinografías. Se observan variaciones en la definición de estructuras anatómicas, el contraste o disminución en la amplitud de campo de la retinografía: (a) variación del tamaño del nervio óptico debido a una alta miopía, (b) inadecuada iluminación debido a un tamaño pupilar pequeño.

3.5.2.- Factores instrumentales y de adquisición

El desenfoque y el descentramiento son factores que se asocian a una baja calidad y nitidez de las imágenes retinográficas. Ambos pueden generar diferencias en el tono, iluminación y contraste, incorporando un elemento de variabilidad que dificulta el posterior análisis de las retinografías digitales. Tal como se muestra en la figura 5 (a) y (b), la presencia de artefactos ópticos (como por ejemplo halos) disminuye la porción de la retina que puede ser analizada e introducen ruido que puede incluso enmascarar las lesiones presentes en la imagen del fondo de ojo.

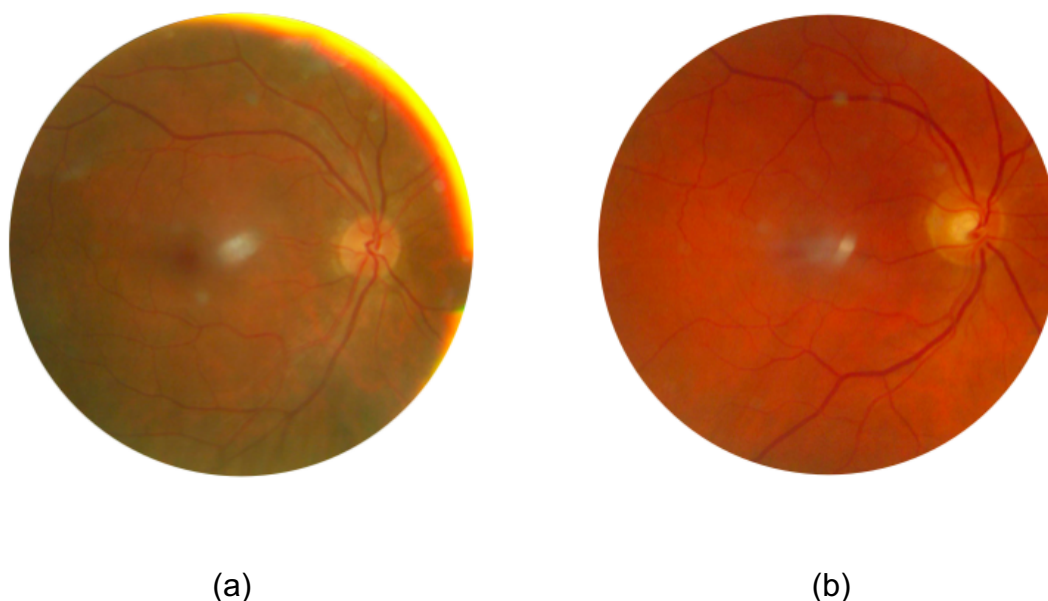


Figura 5. Artefactos ópticos que alteran la calidad de las retinografías. Ambas retinografías presentan artefactos ópticos descritos a continuación: la retinografía puede verse afectada por descentramientos del eje óptico durante su captura generando un halo periférico, sumado a esto se observa un brillo central al lado de la mácula (a), presencia de brillo central blanco sobrepuesto a la posición macular (localizada en la zona central de la retinografía y de coloración más oscura) que obstaculiza su correcta visualización(a) y (b).

3.6.- Características de los métodos de procesamiento de imágenes

Los métodos basados en procesamiento de imágenes comprenden una amplia gama de técnicas orientadas al análisis descriptivo. Se componen de operadores matemáticos que seleccionan, extraen y clasifican la imagen en alguna clase determinada o bien detectan objetos. En esta clase podemos encontrar filtros de umbral, descriptores de contorno y detectores de esquinas cuya implementación no requiere de un gran poder de cómputo. Una importante limitación de estos métodos es la dependencia de supervisión experta en la selección de las características de interés.

El equilibrio del número de variables seleccionado también es un factor a considerar, razón por la cual este proceso se basa frecuentemente en heurísticas que ajustan los parámetros de manera iterativa.

3.7.- Características de los métodos de aprendizaje de máquinas

Los métodos basados en aprendizaje de máquinas, en el contexto de las retinografías digitales, se enfocan en análisis de tipo predictivo, simplificando el proceso de extracción de características de interés, clasificando o segmentando imágenes a partir de un conjunto de datos de entrenamiento. Las imágenes son procesadas estableciendo las reglas de clasificación de manera automática, a partir de datos etiquetados por especialistas o expertos. Estos métodos aprovechan la gran cantidad de datos disponibles, alcanzando resultados con una alta precisión y eficiencia, no necesitan la supervisión de expertos en cada etapa del proceso y requieren de un menor ajuste fino de sus parámetros, debido a que el sistema corrige las diferencias de clasificación de manera automática. Entre sus inconvenientes se encuentran i) el etiquetado o clasificación de los datos, el cual requiere no solo de conocimientos específicos del problema sino que de muchas horas de etiquetado manual, generando el mayor cuello de botella en el flujo de trabajo y ii) la alta variabilidad en el etiquetado de la muestra, causada por las diferencias de criterio que existe entre los expertos.

3.7.1.- Aprendizaje de máquinas

En esta sección se reseña el desarrollo del campo del aprendizaje de máquinas, describiendo sus inicios con el Perceptrón hasta llegar al análisis de imágenes mediante segmentación automatizada [14], [15].

Perceptrón

El perceptrón fue un algoritmo desarrollado entre las décadas de 1950 y 1960 por el científico Frank Rosenblatt, quien definió la unidad básica de procesamiento presente en las primeras redes neuronales artificiales, basando su modelo de activación en la respuesta sináptica de una neurona. El perceptrón es capaz de recibir diversos estímulos mediante múltiples conexiones de entrada, estableciendo el valor (peso o w) que define la intensidad con que cada variable afecta al vector de salida o resultado final. Este valor podrá tomar el estado activado (1) o desactivado (0) dependiendo si la

suma de la multiplicación de sus pesos es mayor o menor al umbral predefinido (sesgo o *bias*). Mediante este método, un perceptrón es capaz de separar dos conjuntos de datos en el espacio por medio de una recta o un hiperplano. Sin embargo con la finalidad de resolver problemas de mayor complejidad, fue necesaria la implementación de nuevas arquitecturas compuestas por un mayor número de perceptrones, lo que llevó al desarrollo del perceptrón multicapa el cual se describe a continuación.

Perceptrón Multicapa

Esta red neuronal artificial está compuesta por varias capas de perceptrones, los que interconectados de manera secuencial procesan la información de manera jerarquizada, ponderando los valores de entrada mediante cierto umbral o función de activación, siendo una de las más utilizadas la unidad rectificadora lineal (RELU). La agrupación en capas permite generar hiperplanos de mayor complejidad y según su ubicación en la red se pueden clasificar como: capas de entrada, intermedias (ocultas) y de salida. Su principal ventaja es posibilitar la separación de conjuntos de datos de manera no lineal.

Back propagation y descenso de la gradiente

En el año 1986 los autores David Rumelhart, Geoffrey Hinton y Ronald Williams publican "*Learning representations by back-propagating errors*" [16], trabajo en donde se plantea la disminución del error de predicción del modelo sobre el conjunto de datos, por medio del ajuste iterativo de los parámetros de la red neuronal. La retropropagación se consigue imputando el error a cada neurona de las capas previas de la red según fuese su porcentaje de responsabilidad en la obtención del resultado del vector de salida. A esta función de error se le suele denominar función de coste. Para identificar el mínimo de la función se determina la derivada parcial del punto de interés con respecto a cada parámetro, este proceso es conocido como descenso del gradiente y es utilizado para analizar el comportamiento de la función de coste o de error total, según se modifique algún parámetro de la red. Es mediante este proceso que la red ejecuta el autoajuste de sus parámetros.

Convolutional neural network (CNN)

Las redes neuronales convolucionales (CNN) hicieron su aparición en la publicación de Yann LeCun, "*Object Recognition with Gradient-Based Learning*" [17]. Este autor

implementó el procesamiento directo de la información de los datos de entrada, permitiendo que la red aprendiera a extraer y seleccionar dichos atributos de forma iterativa, logrando que tanto el entrenamiento como el ajuste de los parámetros necesarios para la clasificación se produjeran de manera automática.

Para realizar esta tarea, las CNN reciben como entrada un mapa de atributos que son representados por medio de una matriz compuesta por tres dimensiones: largo y ancho de la imagen medidos en píxeles y una última dimensión que representa los 3 canales de la imagen de color: rojo, verde y azul; en donde se aplica una operación matemática llamada “convolución”, utilizando un conjunto de *kernels* matriciales que ajustan sus parámetros a través del entrenamiento, entregando como producto un conjunto de mapas con características que permiten la posterior clasificación del objeto de interés. Una CNN está compuesta por una capa convolucional, una capa de agrupación y una capa completamente conectada. A continuación se describen brevemente los componentes o capas más comunes de una CNN.

Capa de convolución

Es el proceso de transformación de una imagen aplicando un operador matemático conocido como *kernel* o filtro sobre un píxel de interés y el conjunto de píxeles vecinos de una imagen. Es determinado desplazando el filtro convolucional en forma de ventana deslizante, determinando el producto punto entre ambas matrices en cada píxel perteneciente a la imagen. Como resultado se logra extraer diversos atributos, preservando la relación espacial entre los píxeles que la componen. Durante el entrenamiento, la CNN selecciona los valores óptimos para las matrices de filtro, lo que le permite extraer los atributos de interés principales (texturas, bordes, formas) del mapa de atributos [18].

A medida que la imagen es procesada, la operación de convolución incrementa la profundidad del mapa de atributos de salida, utilizando múltiples *kernels*, conformando un conjunto más amplio y complejo de características pertenecientes al objeto, convirtiéndose en el vector de entrada de las siguientes capas de la red.

Capa de sub muestreo (*Pooling*)

Una vez realizada la operación de convolución es necesario reducir las dimensiones del mapa de atributos, preservando la información más importante. En este proceso

normalmente se utiliza un algoritmo denominado operación de agrupación o submuestreo.

Este algoritmo funciona de manera similar a la convolución. La operación de agrupación se realiza sobre el mapa de atributos, extrayendo matrices de un tamaño predeterminado para luego escoger el valor máximo de cada matriz, descartando el resto de los valores. El resultado será la salida de un nuevo mapa de atributos con un menor tamaño y un menor costo computacional asociado.

Una vez procesada la imagen mediante la combinación de varias capas convolucionales seguidas de funciones de *max-pooling*, se obtiene una imagen pequeña con muchos canales (*feature maps*) que es transformada en un vector unidimensional, el que posteriormente es filtrado mediante un regresor logístico. Este operador permite identificar qué características son las más importantes para resolver el problema de clasificación.

Entrenamiento de una red CNN

En esta etapa se realiza el ajuste de los parámetros de la red con el fin de clasificar al objeto de interés en la imagen. En este punto son seleccionados los filtros más adecuados y se asignan los pesos correspondientes a cada parámetro en la red. Esta tarea es llevada a cabo utilizando una función de coste o error total, que utiliza los datos de salida de la red como un producto parcial que es comparado de manera iterativa con el resultado deseado. Su diferencia o error será ajustado utilizando el algoritmo de *backpropagation* buscando disminuir la diferencia entre el resultado obtenido por la red y el resultado deseado.

3.8.- Arquitecturas de la red neuronal utilizada en el procesamiento automático de imágenes

3.8.1.- Generative Adversarial Networks

La arquitectura de redes neuronales de tipo *Generative Adversarial Networks* (GAN) [19] se basa en un escenario de teoría de juegos, en donde una red generadora debe competir contra un adversario. La red generadora produce imágenes a partir de ruido aleatorio, mientras que una red discriminadora trata diferenciar si la imagen analizada es real entre las muestras obtenidas de los datos de entrenamiento y las aportadas por

la red generadora. El discriminador retorna entonces la probabilidad de que la muestra en cuestión sea un dato real y no uno simulado.

Esto motiva a que el discriminador aprenda a clasificar correctamente entre muestras reales y falsas, mientras que simultáneamente la red generadora intenta producir imágenes que el clasificador etiquete como reales.

En el campo del aprendizaje de máquinas, los modelos generativos han tomado recientemente una mayor relevancia debido a su adaptabilidad en variados campos. Su capacidad de representar datos complejos y de alta dimensionalidad puede ser utilizado en el procesamiento de imágenes, vídeos, generación de música, procesamiento de lenguaje natural y otras áreas, entre las cuales destaca el procesamiento de imágenes médicas. Estos modelos son de gran utilidad en la transferencia de estilo desde un dominio a otro, por ejemplo, con problemas como la generación de super resolución, color, procesamiento, corrección de imágenes, plegamiento de proteínas o problemas de segmentación.

3.8.2.- U-Net

La primera red tipo *U-Net* apareció en la publicación "*U-Net: Convolutional Networks for Biomedical Image Segmentation*" escrita por Ronneberger *et al* [20]. La arquitectura tipo *U-Net*, en el que se implementan las cGAN, está compuesta de una red codificadora que extrae las características espaciales de las imágenes de entrada y una red decodificadora, que genera mapas de segmentación a partir de las características extraídas previamente. Ambas redes se entrelazan desde el segmento codificador al decodificador concatenando dos operaciones de convolución. Este tipo de redes cuenta además con conexiones intermedias o de salto (*skip connections*) que permiten recuperar información espacial perdida durante el submuestreo de la imagen (*downsampling*)[21].

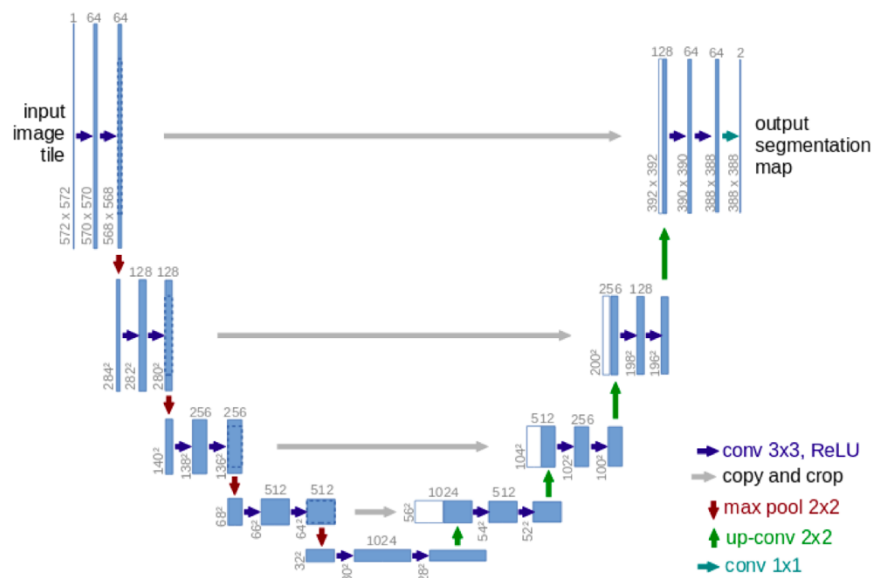


Figura 6. Arquitectura de una red tipo U-net. Se muestra la conformación de la red convolucional-convolucional inversa. Este modelo comprende cuatro niveles de bloques que contienen dos capas convolucionales con normalización por lotes y funciones de activación ReLU, una capa de *max pooling* en la parte de codificación y capas convolucionales en la parte de decodificación. El número de convoluciones en cada bloque es de 32, 64, 128 y 256. La capa de cuello de botella tiene 512 filtros. Desde las capas de codificación, se utilizan conexiones de salto a las capas correspondientes en la parte de decodificación. Fuente: "Image-to-Image Translation with Conditional Adversarial Networks" [21]

3.9.- Otros métodos de clasificación utilizados en este trabajo.

La implementación y selección de clasificadores basados en métodos de aprendizaje supervisado se basó en la evidencia entregada por diversos estudios que apuntan a un mejor desempeño en los métodos que implementan aprendizaje de tipo supervisado [22] y al hecho de contar con la información correspondiente al *ground truth* de clasificación, posición macular y localización de lesiones de las bases de datos IDRiD. Para llevar a cabo el entrenamiento supervisado es necesario disponer de un conjunto de instancias como dato de entrada (llamadas instancias de entrenamiento), pertenecientes a una clase (etiqueta) particularmente conocida y poseer un conjunto de atributos asociados con valores que permiten un aprendizaje que discrimine a cada clase de interés en base a la información entregada. A partir de este proceso, el modelo es capaz de predecir la clase a la cual pertenece un nuevo dato utilizando los atributos propios del objeto [23].

En base a los antecedentes expuestos anteriormente, es que se escogieron 4 algoritmos cuyos resultados serán contrastados analizando cada base de datos, método de balance y vector de características utilizado. A continuación se describen brevemente los algoritmos propuestos en la clasificación de maculopatía.

3.9.1.- Logistic Regression (LR)

Es un algoritmo de tipo discriminativo, aplicado en problemas linealmente separables que clasifica de manera binaria la probabilidad de que cierto dato pertenezca a una clase en particular. Para llevar a cabo esta clasificación se separan las regiones pertenecientes a cada clase mediante una curva o función logística cuyo ajuste es realizado implementando el método de estimación por máxima verosimilitud (*maximum likelihood estimation*), un método de estimación de los parámetros de una supuesta distribución de probabilidad. El modelo de LR puede utilizar entradas de tipo categóricas (codificados como valores 0 y 1) o continuas (toma cualquier valor en un intervalo real) [24]. Entre sus ventajas podemos mencionar su fácil implementación, interpretación y eficiencia de recursos computacionales. Por otro lado, entre sus desventajas se mencionan el sobreajuste en bases de datos de alta dimensionalidad y su bajo rendimiento en datos no lineales.

3.9.2.-Support Vector Machine (SVM)

SVM es un conjunto de modelos de aprendizaje supervisado utilizado en clasificación, regresión y detección de *outliers*. Estos modelos son capaces de separar regiones pertenecientes a diferentes clases, utilizando hiperplanos que buscan maximizar la separación entre los datos y las regiones correspondientes a cada clase. SVM utiliza un subconjunto de datos de entrenamiento que definen el margen de separación para cada categoría, denominados vectores de soporte, cuya función es restringir la rotación y expansión del margen de decisión. En el caso de problemas no separables linealmente, es posible aplicar una función *kernel* que permite el ajuste y la subsecuente separación de clases, aumentando con ello las dimensiones del conjunto de datos a clasificar previo al ajuste del hiperplano [24]. Entre sus principales ventajas se menciona su buen desempeño separando bases de datos de alta dimensionalidad y que los datos marginales “*outliers*” tienen un menor impacto en sus resultados. Por otra parte, entre sus desventajas se encuentra un bajo rendimiento en bases de datos con

categorías superpuestas y que la selección de hiper-parámetros es compleja para lograr la generalización del modelo.

3.9.3.-Decision Tree (DT)

Un DT es un modelo de predicción expresado como una partición recursiva del espacio de características en subespacios que constituyen una base para la siguiente predicción. Este método de clasificación utiliza un conjunto de decisiones jerárquicas sobre características denominadas nodos. Las decisiones tomadas en base a las características evaluadas en los nodos internos constituyen el criterio de división, el cual separa y entrega al siguiente nodo la información. Si la ruta de decisión está completa, el resultado se entrega a un nodo terminal u hoja, la cual finalmente asigna una clase o probabilidad a la característica de interés [25]. Entre sus ventajas se encuentran su fácil interpretación, selección automática de atributos de importancia y ser un modelo no paramétrico. Un problema que presenta este método consiste en que pequeñas variaciones en el conjunto de entrenamiento generan diferentes divisiones, conducentes a DT diferentes.

3.9.4.-Random Forest (RF)

RF determina la clase a la cual pertenece la variable de interés utilizando una combinación de múltiples árboles de decisión, de forma que a cada árbol le es asignado una parte aleatoria de datos (*tree bagging*) y una parte aleatoria de características (*feature sampling*). La clase a la cual pertenece la variable de interés es determinada a partir del promedio del conjunto de salidas de cada árbol individual. De esta manera, el sistema corrige los errores individuales de clasificación, mejorando el resultado general de clasificación del modelo [26]. Esta estructura permite que el conjunto de árboles incremente las propiedades de generalización del modelo, conformando un clasificador más robusto debido a la estrategia de clasificación basada en la ponderación de los resultados de diversos árboles de decisión.

3.10.- Estrategias de evaluación de los clasificadores

3.10.1.- Coeficiente de correlación de Pearson

El coeficiente de correlación de Pearson es una medida de asociación lineal entre dos variables X e Y cuyo resultado puede variar entre los valores +1 y -1, es calculado mediante el valor de la covarianza de ambas variables dividido por el producto de sus desviaciones estándar. Los resultados de la correlación se clasifican de acuerdo a si los valores obtenidos son positivos (ambas variables tienden a aumentar) cercanos a cero (no hay variación entre las variables en análisis) o negativos (mientras una variable aumenta la otra tiende a disminuir).

3.10.2.- Matriz de confusión

Consiste en una tabla que muestra el rendimiento de un modelo predictivo de clasificación, comparando el valor previsto de la variable de destino con su valor real. Se utiliza en problemas de clasificación en donde la salida puede tener dos o más tipos de clase.

Tabla 1: Matriz de confusión

Resultado del Clasificador (predicción)	Característica Observada (real)		Total
	Presente	Ausente	
Positivo	Verdadero Positivo (VP)	Falsos Positivo (FP)	VP+FP
Negativo	Falso Negativo (FN)	Verdadero Negativo(VN)	FN+VN
Total	VP+FN	FP+VN	n

A partir de la matriz de confusión es posible evaluar el rendimiento de un modelo de clasificación mediante el cálculo de métricas de rendimiento, tales como: i) exactitud (accuracy), ii) precisión (*precision*), iii) exhaustividad (*recall*) y iv) puntuación F1(*F1 score*) descritas en la tabla 2 [27].

Tabla 2: Métricas de desempeño y su respectiva fórmula.

Parámetro	Fórmula	Interpretación
Exactitud (<i>Accuracy</i>)	$\frac{VN + VP}{VN + FP + TP + FN}$	$\frac{\text{Total de predicciones correctas}}{\text{Total de datos observados}}$
Precisión (<i>Precision</i>)	$\frac{VP}{VP + FP}$	$\frac{\text{Datos correctamente clasificados como verdaderos}}{\text{Total de datos clasificados como verdaderos}}$
Especificidad o Exhaustividad (<i>Recall</i>)	$\frac{VN}{VN + FP}$	$\frac{\text{Datos correctamente clasificados como falsos}}{\text{Total de observaciones falsas}}$
Puntuación F1 (<i>F1-Score</i>)	$2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$	Media armónica entre <i>Precision</i> y <i>Recall</i>

3.10.3.- Receiving operator curve (ROC)

Es una representación gráfica, que resume el desempeño de un modelo de clasificación en todos los umbrales de clasificación. Esta curva utiliza dos parámetros, el valor predictivo negativo (1- especificidad) en el eje de las abscisas y el valor predictivo positivo (sensibilidad) en el eje de las ordenadas.

3.10.4.- Area under the ROC curve (AUC)

Entrega una medición agregada del rendimiento de todos los umbrales de clasificación, valor que se interpreta como la probabilidad de que el modelo clasifique con un mayor porcentaje de acierto a un ejemplo positivo aleatorio, que a un ejemplo negativo aleatorio. A mayor valor de AUC mejor es la predicción del modelo diferenciando clases [28].

3.10.5.- Validación cruzada de K interacciones

La validación cruzada nos permite comparar diversos métodos de clasificación y nos da una idea de qué tan bueno es su desempeño. En este método, los datos son divididos en subconjuntos de entrenamiento y prueba, asignando aleatoriamente el grupo correspondiente al de prueba, se calculan los resultados de este primer ciclo y se escoge un nuevo subconjunto de prueba; el proceso continúa escogiendo diferentes combinaciones hasta utilizar todas las alternativas posibles, promediando y resumiendo los resultados de cada ciclo. El número de divisiones utilizadas se denomina K. En este

caso corresponde a 10 subconjuntos que corresponde al número de divisiones más utilizado en los estudios que lo implementan [29].

3.11.- Métodos automatizados de detección macular

Existe una gran diversidad de criterios utilizados con la finalidad de reportar los resultados alcanzados en la tarea de detectar la posición macular de manera automática. En la mayoría de los trabajos publicados evaluaron la posición macular considerando la distancia euclidiana entre la posición del centro de la mácula obtenida automáticamente y la posición real de esta estructura [5], [6]. Esta distancia suele expresarse en función del radio del nervio óptico (DD), que se toma como medida de referencia. Esta forma de evaluación se justifica por la falta de definición de los bordes en la región macular [4]. La información de los rendimientos alcanzados por distintos autores resumidos en la tabla 3.

Tabla 3: Comparación de los desempeños alcanzados en la determinación automática de la posición macular reportadas por otras metodologías encontradas en la literatura.

Method	N° Images	Quality scale				
		1/8 R	1/4 R	1/2 R	1 R	2 R
Fleming <i>et al</i>	1056	-	-	-	96.50	-
Niemejer <i>et al</i>	500	-	-	-	94.40	-
Niemejer <i>et al</i>	500	-	-	-	96.80	-
Tobin <i>et al</i>	345	-	-	-	-	92.50
Yu <i>et al</i>	1200	-	-	95.00	-	-
Gegundez-Arias <i>et al</i>	1200	80.2	93.92	96.08	96.92	97.83

Los resultados se expresan en porcentaje y no necesariamente utilizan la misma base de datos. Tabla adaptada de Gegundez-Arias *et al* [4]

3.12.- Clasificación automatizada de maculopatía diabética

Los prometedores avances en el diagnóstico automatizado de imágenes retinales han permitido su implementación como herramientas de apoyo clínico en el tamizaje de pacientes con RD. Entre sus ventajas podemos destacar el corto tiempo de procesamiento, su adquisición no invasiva y la obtención de resultados con muy baja variabilidad, convirtiéndose en una herramienta confiable y eficiente que contribuye a alcanzar las metas sanitarias propuestas a nivel ministerial.

Es posible encontrar diversas implementaciones utilizadas en la evaluación del riesgo de presentar MD, algunas de ellas se basan en algoritmos especializados en la detección de características, otras en la clasificación automatizada de MD a partir de imágenes etiquetadas. Un problema de importancia mencionado frecuentemente en la literatura es la ausencia de KPI (*Key Performance Indicators*) estandarizados para medir el rendimiento de un modelo particular [30], las métricas utilizadas son variadas: sensibilidad, especificidad y AUC por nombrar algunas de ellas. Cada publicación utiliza métricas distintas para medir su trabajo y debido a que se utilizaron diferentes condiciones objetivo y escalas de clasificación de MD, es difícil comparar directamente los estudios incluidos. Se resumen algunos de los trabajos publicados que clasifican la maculopatía diabética en la tabla 4 (Ver información completa en la tabla anexo).

Tabla 4: Extracto de los desempeños alcanzados en la detección de maculopatía diabética (MD).

Author	Database	Method and Classifiers	Outcomes and Comments
Fleming et al (2007)	Private (14406)	Morphological image processing, exudate location.	Accuracy – 99.2% (NCSME) Accuracy – 97.3% (CSME)
Lim et al (2011)	MESSIDOR (88)	CLAHE and Watershed transform, exudate location	Sensitivity –80.90% Specificity –90.20% Accuracy – 85.20%
Giancardo et al (2012)	HEI-MED and MESSIDOR (1200)	Wavelet transform, Kirsch edge detection, colour, and support vector machine	AUC – 0.94
Lim et al (2017)	MESSIDOR	Local Binary pattern -extract gray level features and SVM classifier	Sensitivity –80% Specificity -70%
Sulaiman (2020)	IDRiD	Deep learning- based DME-grading model is proposed for automatic DME grading of retinal fundus images	Accuracy 68% Precision 66% Recall 68% F1-score 65%
Singh (2020)	IDRiD MESSIDOR	DME Net Algorithm which is built using Hierarchical Ensemble of CNNs (HE-CNN) for DME diagnosis	Accuracy of 96.12% Sensitivity of 96.32% Specificity of 95.84% F – 1 score of 0.9609 AUC 0.9654 MESSIDOR and IDRiD dataset

Se resumen algunos de los resultados obtenidos utilizando la base de datos MESSIDOR, también se incluye un estudio que utilizó la base de datos IDRiD. Fuente “*Diabetic retinopathy detection through artificial intelligent techniques: A review and open issues*” [22]

3.12.1.- Interpretación de los modelos de decisión

Un aspecto ampliamente discutido en la literatura ha sido la dificultad para interpretar los mecanismos internos por los cuales un modelo con determinada arquitectura escoge ciertos procesos o decisiones por sobre otros. En la actualidad se está abogando cada vez más por el uso de modelos interpretables por sobre modelos poco transparentes (*black box models*) [31]. Estos aspectos se hacen aún más relevantes en problemas que implican decisiones de gran impacto, como por ejemplo en el ámbito de la salud, donde la determinación de cualquier diagnóstico está fuertemente controlado e implementado mediante guías y protocolos clínicos [12]. Es en estas circunstancias que se hace cada vez más relevante el concepto de interpretabilidad del modelo (*interpretability*), entendiéndose como la capacidad de explicar o presentar los procesos

de toma de decisiones del modelo en términos comprensibles para el ser humano. Esta carencia de interpretación de los *black box models* conlleva a una formalización incompleta del problema, creando una importante barrera para la optimización y evaluación del proceso, aspecto fundamental para lograr la consolidación de la confianza por parte de los usuarios finales de estas herramientas tecnológicas [32].

3.13.- Flujo de la atención primaria oftalmológica en pacientes con RD

3.13.1.- Unidades de Atención Primaria Oftalmológicas UAPO

En Chile, el examen de fondo de ojo es realizado en las Unidades de Atención Primaria Oftalmológicas (UAPO). Consiste principalmente en la adquisición de imágenes de la retina o en su defecto, cuando no es posible obtener la retinografía, mediante la observación directa de la zona posterior del ojo a través de un oftalmoscopio, procedimiento realizado por el médico especialista. La utilización de retinografías tiene la ventaja de permitir su revisión por profesionales especializados de manera asincrónica. Otra gran ventaja es que pueden ser analizadas utilizando técnicas de procesamiento de imágenes, metodologías actualmente validadas como técnicas de tamizaje. En nuestro país se han implementado diversas metodologías de análisis, siendo la más ampliamente difundida el sistema de Diagnóstico Automatizado de Retinografías Telemáticas (DART) el cual, desde su implementación en el año 2018, ha contribuido a aumentar el control y acceso oportuno de los pacientes a exámenes oftalmológicos.

3.13.2.- Diagnóstico Automatizado de Retinografías Telemáticas DART

El sistema DART utiliza métodos de aprendizaje de máquinas que detectan las lesiones presentes en el área retinal para luego clasificar la probabilidad de que el paciente presente RD. Este sistema identifica a los pacientes que presentan signos patológicos con sospecha de RD, los que serán derivados a controles presenciales con especialistas. Para llevar a cabo esta tarea DART utiliza dos componentes: (i) un módulo de inteligencia artificial que permite hacer un primer tamizaje discriminando entre RD positiva o negativa, y (ii) una interfaz de teleinforme utilizada por médicos oftalmólogos, quienes emiten un informe y deciden si es necesaria la derivación a controles presenciales. Cabe señalar que las redes neuronales del algoritmo de

inteligencia artificial de DART encargadas de preprocesar la información no están diseñadas para entregar un resultado exacto de la localización de lesiones retinales. Tales funciones más bien corresponden a pasos intermedios usados con fines ilustrativos, con el propósito de mejorar la interpretabilidad de los datos procesados [27], [33].

3.14.- Problema y solución propuesta

El módulo de inteligencia artificial del sistema DART no contempla dentro de su algoritmo de clasificación la localización y la extensión de las lesiones presentes en la zona macular dejando fuera un importante factor de categorización de las lesiones encontradas.

En esta tesis se propone relacionar la ubicación del área macular con las lesiones retinales correspondientes a la zona central, incorporando un clasificador del estado macular que contribuya a mejorar la precisión de la clasificación automática de RD realizada actualmente. Es decir, se busca contribuir mediante la construcción de un clasificador binario para maculopatía que permita la priorización en la evaluación de los casos con MD en el módulo de teleinforme.

4.- Hipótesis y objetivos

La determinación de la ubicación del área macular, y su relación con las lesiones retinales detectadas previamente, incrementará la precisión diagnóstica del sistema de clasificación automatizada.

4.1.- Objetivo General

Evaluar la contribución de la integración de la posición macular y la presencia de lesiones retinales para la detección de maculopatía de manera automatizada.

4.2.- Objetivos Específicos

Objetivo específico 1: Evaluar el riesgo de presentar maculopatía, integrando en un clasificador la información manual de la posición macular y la presencia de lesiones retinales.

Objetivo específico 2: Determinar el método que mejor estime la posición macular, comparando algoritmos de procesamiento de imágenes y aprendizaje de máquinas.

Objetivo específico 3: Comparar el riesgo de presentar maculopatía utilizando la posición manual de las lesiones retinales frente a la posición de las lesiones retinales obtenidas a partir del sistema automatizado DART.

Objetivo específico 4: Evaluar el riesgo de presentar maculopatía integrando un estimador de posición macular, la localización de lesiones DART y un clasificador automático.

5.- Material y Métodos

A continuación se describe el material y método utilizados en la implementación del presente trabajo de tesis. Se reseñan las características principales de las bases de datos seleccionadas y el proceso de construcción de las entradas de datos necesarias. La sección de métodos está subdividida en base a los objetivos específicos planteados y describe las etapas necesarias para evaluar el riesgo de presentar maculopatía integrando información manual y automática de la localización de lesiones y posición macular.

5.1.- Material

5.1.1.- Bases de datos

Con respecto a las bases de datos disponibles de manera pública, en ellas se observan dos tipos de captura: i) centrada en la zona macular y nervio óptico periférico y ii) con el nervio óptico al centro y retina nasal periférica, abarcando ambas un campo visual de 45° no estereoscópico. En base a estas imágenes son localizados de manera manual, el centro foveal, el centro del nervio óptico y los vasos retinales. Algunas bases de datos entregan coordenadas correspondientes al centro de dichas estructuras, otras duplican la imagen y las delimitan directamente sobreponiendo marcas sobre lesiones detectadas. Esta información es la que permite definir qué base de datos y en qué etapa puede ser utilizada.

Tabla 5: Bases de datos disponibles utilizadas para el entrenamiento y evaluación de diversas tareas de clasificación y localización de estructuras y patologías retinales.

Dataset	Tamaño del Dataset	Peso por Imágen	Dimension Pixels	Patología/ Diagnóstico	Cantidad de Imágenes	FOV	Clasificación/ Identificación	Ground Truth Mácula Fóvea
APTOS 2019	Test 1,61 GB/ Train 8,6 GB	Variable	Train variable Test 640 X 480	Retinopatía Diabética	Test 1929 / Train 3663	-	Diagnóstico de RD	NO DISPONIBLE
DIARETDB0	216 MB	1,6 MB	1500 X 1152	Retinopatía Diabética	130 (20 normales, 110 alteradas)	50°	Microaneurismas , hemorragias, exudados blandos y duros	NO DISPONIBLE
DIARETDB1	143 MB	1,6 MB	1500 X 1153	Retinopatía Diabética	89 (5 normales, 84 alteradas)	50°	Microaneurismas , hemorragias, exudados blandos y duros	NO DISPONIBLE
DRIVE	Test 14.8 MB/Train 14.8 MB	750-780 KB	565 X 584		Test 20 / Train 20	45°	Vasos Retinales	Nervio Óptico/ Vasos Retinales Segmentados
MESSIDOR	8.7 GB	4,2 / 10 MB	1440 X 960 2240 X 1488 2304 X 1536	Retinopatía Diabética	1200 imágenes	45°	Diagnóstico de RD	Retinopatía y graduación de riesgo de edema macular
RIGA	Magrabi 2,8 GB BinRushed 395 MB MESSIDOR 9,42 GB	-	Magrabi 2743 X 1936 BinRushed 2376 X 1584 MESSIDOR 1440 X 960 2240 X 1488 2304 X 1536	Glaucoma	-	-	6 expertos delimitan reborde neural	Excavación del nervio óptico

STARE	504 MB	1,3 MB	700 X 605	diagnósticos Variados	402 imágenes	-	2 expertos clasifican, diagnostican, segmentan vasos y determinan ubicación del Nervio óptico	Nervio Óptico
IDRiD	1.01 GB	Variable	4288 x 2848	Retinopatía Diabética	516 imágenes Clasificación RD 516 Localización estructuras 80 segmentación lesiones		Diagnóstico de RD Diagnóstico de MD Localización Lesiones	No disponible en subconjunto de segmentación de lesiones
UoA-DR	500 MB	1,8/2 MB	2124 x 2056	Retinopatía Diabética	200 sets (4 por cada set, 200 retinografías)	45°	Centro del Nervio Óptico, localización Foveal , clasificación de la Retinopatía Diabética (proliferativa, No proliferativa, sana)	Vasos retinales, reborde y centro del nervio óptico, localización del centro foveal

En la tabla podemos observar la información disponible y caracterización de cada base de datos, tales como etiqueta por diagnóstico, segmentación de estructuras. Sus datos públicos permiten evaluar distintos algoritmos y comparar sus desempeños. Se destaca en negrita la información relevante para seleccionar la base de datos MESSIDOR e IDRiD.

Para el desarrollo del presente trabajo de tesis, se utilizaron dos conjuntos de bases de datos que cuentan con una gran parte de la información necesaria para procesar y clasificar las variables de interés: 1) “*Méthodes d'Évaluation de Systèmes de Segmentation et d'Indexation Dédiées à l'Ophtalmologie Rétinienne*” (MESSIDOR) , que consta de 1744 retinografías (151 con maculopatía y 1593 sin maculopatía) [34] y 2) “*Indian Diabetic Retinopathy Image Dataset*” (IDRiD) compuesta por 80 retinografías con lesiones segmentadas (69 retinografías con maculopatía y 11 sin maculopatía) [35].

5.1.2.-Base de datos MESSIDOR 1/MESSIDOR 2

MESSIDOR es un proyecto de investigación fundado por el ministerio francés de investigación y defensa; tres departamentos de oftalmología estuvieron a cargo de la captura y posterior clasificación de las 1200 imágenes con que cuenta. Equipos de oftalmólogos de cada servicio clasificaron el tipo de retinopatía diabética y el riesgo del paciente de desarrollar edema macular. Esta clasificación es de gran utilidad para entrenar y probar distintos modelos basados en aprendizaje de máquinas.

Tabla 6: Clasificación del grado de progresión de RD utilizada en la base de datos MESSIDOR.

Clasificación de Retinopatía		
0 (Normal)	$(\mu A = 0) \text{ y } (H = 0)$	μA : número de microaneurismas H: número de hemorragias NV = 1: neovascularización NV = 0: sin neovascularización
1	$(0 < \mu A \leq 5) \text{ y } (H = 0)$	
2	$((5 < \mu A < 15) \text{ o } (0 < H < 5)) \text{ y } (NV = 0)$	
3	$(\mu A \geq 15) \text{ o } (H \geq 5) \text{ o } (NV = 1)$	

En la primera columna se presenta la clasificación en grados de 0 a 3 (normal y con RD respectivamente) y el criterio utilizado para clasificar estas distintas etapas de progresión de RD (columnas 2 y 3).

Tabla 7: Clasificación del riesgo de presentar edema macular utilizada en la base de datos MESSIDOR.

Riesgo de edema macular: Se utilizó la ubicación de los exudados duros al respecto del centro macular para graduar el riesgo de edema macular.	
0	No hay exudados duros visibles (Sin riesgo)
1	Distancia más corta entre la mácula y un exudado duro > un diámetro papilar
2	Distancia más corta entre la mácula y un exudado duro \leq un diámetro papilar

En la primera columna se presenta la clasificación en grados de 0 a 2 (normal y con edema macular respectivamente) y a la derecha de la tabla, el criterio utilizado para clasificar las distintas etapas de progresión.

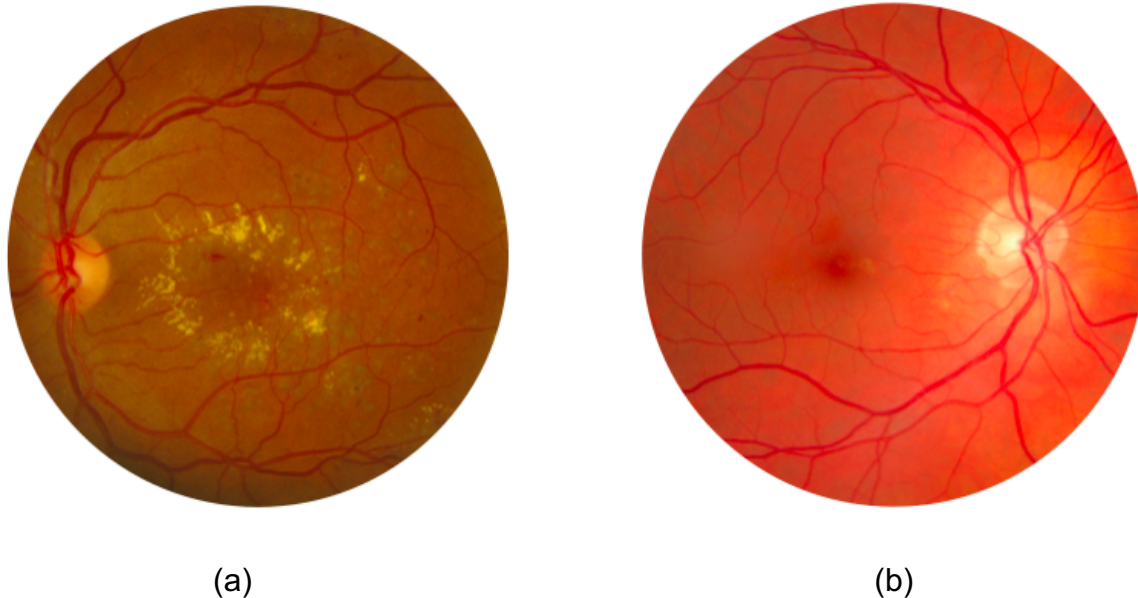


Figura 7. Ejemplo de las diferencias en las retinografías presentes en la base de datos MESSIDOR. Se muestra una retinografía centrada en la mácula y a un extremo el nervio óptico en conjunto con las arcadas superior e inferior de los vasos retinales. La retinografía (a) corresponde al ojo izquierdo con presencia de lesiones brillantes y algunas lesiones rojas, retinografía (b) corresponde a un ojo derecho con la retina sin lesiones.

La base de datos MESSIDOR 2 es una actualización de la base de datos original. En MESSIDOR 1 algunas imágenes solo incluían una retinografía por paciente, en cambio otras incluían tanto el ojo derecho como el izquierdo. En MESSIDOR 2 se incluyeron todas las parejas de las retinografías individuales faltantes, es decir, 529 (1058 imágenes en formato PNG). Además se reclutaron nuevos pacientes para ampliar la base de datos ya existente, incluyendo retinografías centradas solamente en mácula. MESSIDOR ampliada incluye 345 exámenes (690 imágenes en formato JPG).

Un aspecto no mencionado en los estudios que utilizan esta actualización de la base de datos MESSIDOR, es la presencia de un artefacto óptico en el centro de la retinografía, el que es coincidente con la posición macular. Su presencia implica limitantes y dificultades que pueden generar sesgos importantes en el presente trabajo, sobre todo en la etapa de entrenamiento en la detección de la posición macular. Por tal razón, estas imágenes fueron excluidas del flujo de procesamiento en el presente trabajo de tesis.

5.1.3.-Base de datos IDRiD

Esta base de datos fue recopilada para ser utilizada en el “*Diabetic Retinopathy: Segmentation and Grading Challenge*” realizado en conjunto con el IEEE *International Symposium on Biomedical Imaging*. Está compuesta por 516 retinografías que fueron subdivididas para ser utilizados en 3 fases del certamen: segmentación, diagnóstico automatizado y localización de estructuras anatómicas retinales. Cabe hacer notar que las etiquetas del conjunto de datos total fue recopilado como un dato independiente para cada fase del certamen y no cuenta con la totalidad de la información necesaria para segmentar, localizar o clasificar a cada uno de los subconjuntos de imágenes. La composición de cada base de datos se describe a continuación:

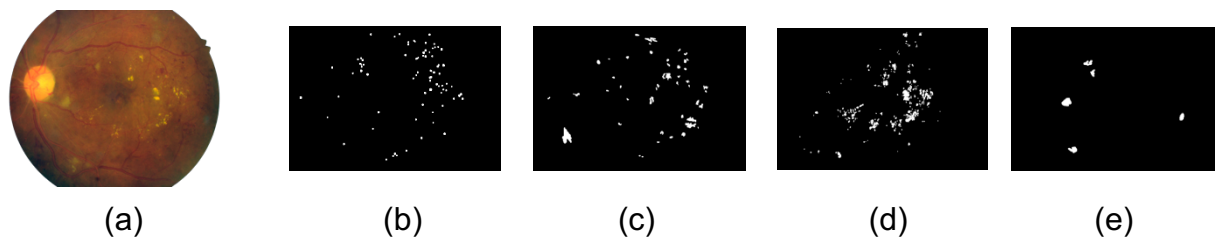


Figura 8. Ejemplo de retinografía y sus correspondientes imágenes segmentadas con lesiones presentes en la base de datos IDRiD. La retinografía representa (a) retinografía original, (b) segmentación de microaneurismas, (c) segmentación de hemorragias, (d) segmentación de exudados duros y (e) segmentación de exudados blandos.

Tabla 8: Composición de los subconjuntos de datos correspondientes a la base de datos IDRiD.

Database	N de Retinografías	Máscaras de Segmentación Disponibles	Tipo de Clasificación Disponibles	Subdivisiones Disponible
Segmentation	81	Microaneurismas, Hemorragias, Exudados Duros, Blandos y Nervio Óptico		Entrenamiento Prueba
Clasificación de la RD	516	No disponible	Nivel de riesgo de RD, clasificación desde 0 (sin riesgo aparente de presentar RD) a 4(RD grave). Riesgo de Edema Macular clasificación desde 0 (sin riesgo aparente de presentar EM) a 2 (EM grave).	Entrenamiento Prueba
Localización	516	No disponible	Localización del Nervio Óptico en el eje X e Y expresados en píxeles. Localización Macular en el eje X e Y expresados en píxeles.	Entrenamiento Prueba

Se puede observar tres subconjuntos de datos con información correspondiente a la segmentación, localización de estructuras y clasificación de retinopatía diabética. Del total de imágenes, solamente el subconjunto de segmentación de lesiones fue utilizado en el presente estudio, la información correspondiente a la posición macular y clasificación del riesgo de presentar maculopatía fue determinada por un especialista en oftalmología.

5.1.4.-Preparación de las bases de datos

En este trabajo de tesis fue necesario incorporar la información correspondiente a la posición macular en ambas bases de datos. En el caso de MESSIDOR esto se realizó a partir del trabajo de Gegundez-Arias *et al* [4] que determinó la posición macular de manera manual . En la base IDRiD la posición fue determinada por un especialista en oftalmología que colaboró en este trabajo, etiquetando adicionalmente el riesgo de desarrollar maculopatía de manera binaria, utilizando un formulario desarrollado especialmente para la clasificación de las retinografías utilizadas en la plataforma *Google Colabs*, etiquetando ausencia o presencia de maculopatía para cada imagen perteneciente al subconjunto de datos utilizados en el entrenamiento de algoritmos para segmentación de imágenes. Con la finalidad de obtener la posición de las lesiones presentes en las retinografías, se llevaron a cabo dos procesos dependiendo de la base

de datos utilizada: en el caso de IDRiD las imágenes segmentadas manualmente fueron binarizadas, determinando el centroide de cada región de interés (ROI) obteniendo las coordenadas correspondientes a cada lesión y separando las coordenadas obtenidas según si su clasificación era de lesión de tipo roja (hemorragias y microaneurismas) o de tipo brillante (exudados duros o blandos). Ambas bases (IDRID y MESSIDOR) fueron procesadas por redes neuronales intermedias del algoritmo de clasificación DART, obteniendo mapas de calor (*heat maps*) que determinan la probabilidad de que un píxel pertenezca a la clase lesión roja o brillante. Para finalizar, se determinaron las coordenadas de las lesiones obtenidas mediante el procesamiento DART y las correspondientes a máscaras de segmentación manual disponibles en la base de datos IDRiD. El resumen de las combinaciones posibles con la información disponible y bases de datos resultantes se muestra a continuación en la Tabla 9.

Tabla 9: Codificación de cada base de datos generada a partir de IDRiD y sus diversas entradas de información que la componen.

Base de Datos	Origen de la información que compone cada base de datos
MGLG	posición macular <i>ground truth</i> y lesiones <i>ground truth</i>
MGLD	posición macular <i>ground truth</i> y lesiones provenientes del sistema DART
UMLG	posición macular <i>U-net</i> y lesiones <i>ground truth</i>
UMLD	posición macular <i>U-net</i> y lesiones provenientes del sistema DART

Se listan de manera descriptiva los componentes de cada base de datos utilizada en la evaluación del rendimiento de cada clasificador de maculopatía.

5.1.5.- Caracterización del balance de clases

La base de datos MESSIDOR presenta un 54.5 % de la muestra (654 retinografías en total) clasificada con algún grado de RD; de éstas, un 34 % (226 retinografías) fueron clasificadas con riesgo tipo 1 o 2 de maculopatía. Un total de 75 pacientes presentaban riesgo tipo 1, es decir un 11.4 % y en 151 pacientes fue determinado que correspondían a la clasificación de riesgo tipo 2, representando un 23 % de la muestra de pacientes clasificados con RD.

La base de datos total de IDRiD está compuesta por 516 imágenes de las cuales un 33 % (168 imágenes) está clasificada con algún grado de RD. Del subconjunto utilizado en el presente trabajo de tesis (80 retinografías correspondientes a la tarea de segmentación) un 86 % (69 retinografías) fueron clasificadas con algún grado de maculopatía. Además, la totalidad de las imágenes del subconjunto de datos presentaron lesiones centrales o periféricas incorporando su correspondiente segmentación.

5.2. - Métodos

5.2.1.- Evaluar el riesgo de presentar maculopatía, integrando en un clasificador la información manual de la posición macular y la presencia de lesiones retinales

Preparación (preprocesamiento) de las bases de datos disponibles

Con la finalidad de estimar la contribución de la posición macular en la determinación de maculopatía en pacientes con RD, se utilizó la base de datos IDRiD debido a que cuenta con el etiquetado manual necesario para establecer el *ground truth* en las siguientes características: i) coordenadas correspondientes a la zona macular y del nervio óptico, ii) clasificación del riesgo de presentar RD, iii) presencia de edema macular o maculopatía y iv) determinación de lesiones presentes en la retina. Con esta información fue conformada la base de datos “posición macular *ground truth* lesiones *ground truth*” (MGLG) necesaria para establecer el rendimiento referencial que fue utilizado para comparar el rendimiento en los objetivos 3 y 4 de la tesis.

Balance de clases

Debido al desbalance en el número de casos que componen la base de datos IDRiD, es que fue necesario implementar métodos de balance que permitieran incrementar de manera sintética los ejemplos disponibles de las clases minoritarias sin alterar su distribución. Las bases de datos resultantes del proceso de balance de clases se describen a continuación:

Tabla 10: Abreviaturas de nombre y composición de cada base de datos generada según el método de balance de clases implementado.

Base de Datos	Método de balance implementado
80	80 retinografías originales sin balancear sus clases
80_80sm	80 retinografías originales adicionando 80 casos sanos sin lesiones
80_BC	80 retinografías originales aplicando balance de clases <i>Synthetic Minority Oversampling Technique</i> y <i>Undersampling Technique</i>
80_9_BC	80 retinografías originales adicionando 9 casos sanos sin lesiones y aplicando balance de clases <i>Synthetic Minority Oversampling Technique</i> y <i>Undersampling Technique</i>

Variables necesarias para evaluar la contribución de la posición macular en la clasificación de maculopatía

Se estableció un conjunto de datos que poseen la ubicación macular y del nervio óptico, determinación de lesiones retinales rojas y brillantes, clasificación general en estadios de RD y presencia de edema macular o riesgo de presentar maculopatía. A partir de la consolidación de estos datos, se determinaron una serie de características de interés con las que se conformaron tres vectores (S, E y F) y que se originaron en el conjunto de datos basados en la posición macular y lesiones *ground truth*. El vector “S” utiliza la información general de las lesiones retinales, sin considerar la posición macular (ver figura 9 a); el vector “E” utiliza exclusivamente la información de la posición macular y la distancia a las lesiones retinales (ver figura 9 b y figura 10); y el vector “F”, que es la concatenación de ambos vectores, incorpora toda la información disponible (ver figura 9 c). Una descripción detallada de estos vectores se presenta en la Tabla 11.

Tabla 11: Variables que componen cada vector de características

Vector F (Vector E + Vector S)	
Vector E	Vector S
Presencia de lesión macular dentro de un radio de 200 píxeles (0=ausencia,1=presencia)	Número total de lesiones retinales.
Número de lesiones rojas localizadas a una distancia menor a un radio de 100 píxeles desde la mácula.	Número total de lesiones retinales rojas.
Número de lesiones brillantes localizadas a una distancia menor a un radio de 100 píxeles desde la mácula.	Número total de lesiones retinales brillantes.
Lesión roja más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles).	Porcentaje: Lesiones rojas vs total de lesiones.
Lesión brillante más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles).	Porcentaje: Lesiones brillantes vs total de lesiones.
Lesión más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles).	Centroide del total de lesiones rojas.
Número de lesiones rojas localizadas a una distancia menor a un radio de 200 píxeles desde la mácula.	Centroide del total de lesiones brillantes.
Número de lesiones brillantes localizadas a una distancia menor a un radio de 200 píxeles desde la mácula.	
Número total de lesiones localizadas a una distancia menor a un radio de 200 píxeles desde la mácula.	

Se listan de manera descriptiva cada característica elaborada a partir de la posición macular y la distancia a las lesiones *ground truth*, También se muestran las características incorporadas a partir del tipo de lesión y su cantidad presente en la retina.

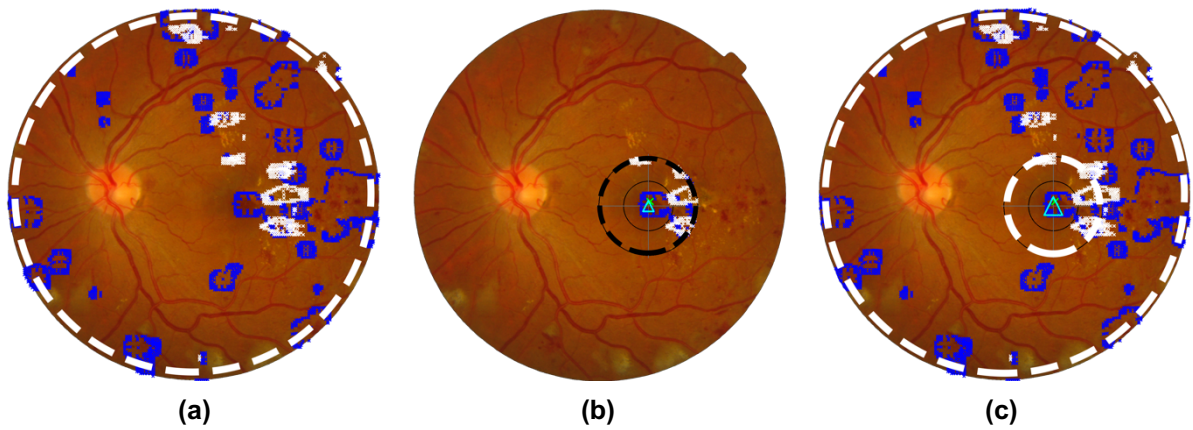


Figura 9. Ejemplo de lesiones utilizadas en cada vector de características. Representación de las zonas con lesiones utilizadas para generar la información correspondiente a los vectores S, E y F: (a) vector S, información de las lesiones periféricas (b) vector E, información correspondiente solo a la región macular (c) vector F, concatenación de ambos vectores. La posición de las lesiones brillantes se representan mediante zonas blancas, las lesiones rojas mediante zonas azules, la posición macular real se representa por una cruz verde y la posición macular U-net mediante un triángulo, cada área de interés está demarcado en un círculo de línea segmentada.

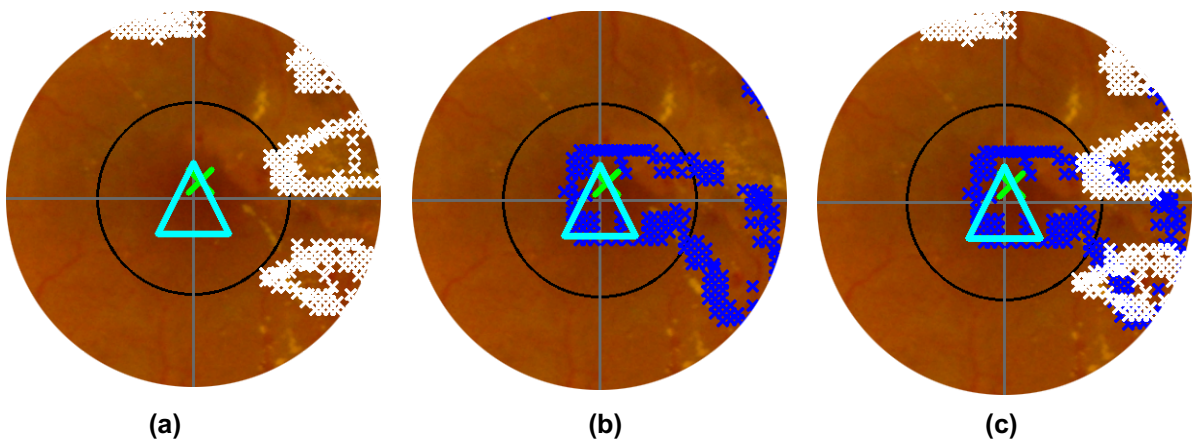


Figura 10. Detalle de la región macular, mostrando los elementos que conforman al vector E. (a) total de lesiones brillantes (exudados duros y blandos), (b) total de lesiones rojas (microaneurismas y hemorragias) y (c) total de lesiones presentes en la región macular en un radio de 100 y 200 píxeles respectivamente. La posición de las lesiones brillantes se representan mediante cruces blancas, las lesiones rojas mediante cruces azules, la posición macular real se representa por una cruz verde y la posición macular U-net mediante un triángulo.

Convencionalmente, para llevar a cabo la determinación automatizada de MD se ha utilizado un único criterio: la detección de exudados duros dentro de un DD a partir del centro de la mácula. Autores como Li *et al.* [36] emplearon este criterio o la presencia de exudados duros en la región macular abarcando al menos un 50% del área de DD, en tanto que Ting *et al* [37] usaron un criterio menos restrictivo detectando cualquier exudado duro presente en la retina. Además de los exudados, Abramoff *et al* [11] utilizaron el engrosamiento de la retina o los microaneurismas dentro de un DD de la fovea como indicadores de MD.

Selección de clasificadores y regresores

Se escogieron 4 tipos de algoritmos clasificadores pertenecientes al tipo de algoritmos de aprendizaje automático supervisado, para comparar el rendimiento alcanzado por cada vector de características (S,E,F): i) *Logistic Regression*, ii) *Support Vector Machine*, iii) *Decision Tree*, iv) *Random Forest*.

Los resultados obtenidos permitieron identificar qué entradas de datos y vectores tuvieron mejores desempeños y su contribución en la determinación de maculopatía.

Se utilizó el Coeficiente de Correlación de Pearson con la finalidad de identificar qué componente del vector de características se comportó como el factor predominante en la determinación del riesgo de maculopatía utilizando la posición del área macular como dato de entrada.

5.2.2.- Selección del método que determine con mayor precisión la posición macular

En la etapa de determinación de la posición macular se escogió la base de datos MESSIDOR debido a que posee la mayor cantidad de imágenes disponibles entre las bases seleccionadas, además de contar con la información de la localización de estructuras retinales, características de gran utilidad en el entrenamiento de algoritmos de aprendizaje de máquinas.

La información de las coordenadas que determinan la posición macular fue utilizada para generar las máscaras de segmentación correspondientes a cada retinografía, conformando el *ground truth* utilizado para entrenar al modelo escogido de aprendizaje de máquinas, en este caso una red neuronal tipo *U-net*. El total de datos fue dividido en

un subconjunto de entrenamiento (1408 imágenes incluyendo aumento de datos), prueba (106 imágenes) y validación (318 imágenes), siendo este último utilizado para la comparación de los resultados obtenidos de la estimación de la posición macular entre los métodos de procesamiento de imágenes y de aprendizaje de máquinas.

Las retinografías utilizadas en el presente trabajo fueron reescaladas a un tamaño de 1514 x 1514 píxeles, producto del preprocesamiento del sistema de clasificación DART, definiendo como consecuencia la escala utilizada para realizar la estimación de la posición macular. Adicionalmente se definió el tamaño del nervio óptico (DD) en 100 píxeles de diámetro en base al análisis del tamaño promedio encontrado en el conjunto de datos utilizado.

Métodos automatizados de detección de la localización macular

Implementación de los métodos de procesamiento de imágenes

Con la finalidad de estimar la posición macular se implementó un conjunto de técnicas basadas en características tales como el contraste de la mácula y el nervio óptico con respecto al resto de la retina, o en las relaciones espaciales que comparte la mácula junto con los hitos anatómicos retinales (nervio óptico y vasos retinales). La estimación de la posición macular fue evaluada mediante tres filtros: el umbral de media localizando directamente la mácula (ver figura 7), el umbral *Yen* determinando el nervio óptico y localizando la mácula mediante la distancia angular a este punto (ver figura 8) y finalmente aplicando el filtro *Meijering* para la localización los vasos retinales, obteniendo un *skeleton* cuyo centroide ajustado corresponde a la región predicha de la posición macular (ver figura 9)[5].

A continuación se describe la implementación realizada para determinar la posición macular:

i) Filtro de intensidad

La posición macular queda definida como la zona central de coloración más oscura con respecto al resto de la retina. El objetivo de este método es obtener una imagen binarizada a partir del umbral calculado con la media y la desviación estándar de la imagen total [38]. El algoritmo de detección macular basado en filtro de media comprende los siguientes pasos:

1. Separación de canales
2. Filtrado gaussiano
3. Binarización aplicando umbral de media
4. Utilización de operadores matemáticos
5. Identificación de regiones de interés
6. Determinación del centroide de la región de interés

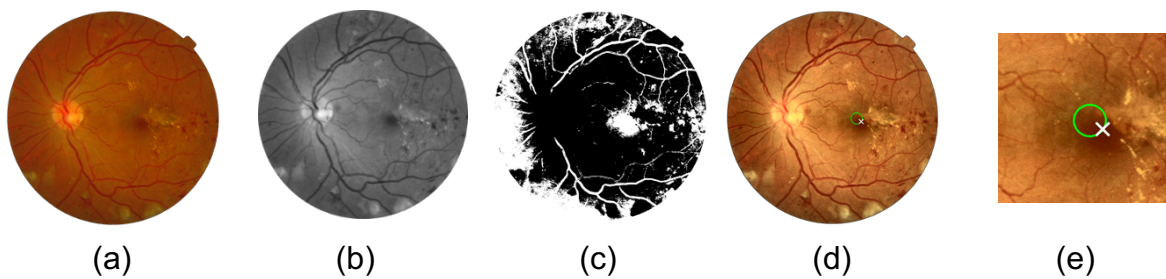


Figura 7. Método de detección macular basado en umbral macular. (a) retinografía original, (b) extracción del canal verde, (c) imagen binarizada aplicando filtro de media, (d) determinación del centroide macular, (e) representación de la estimación de la posición macular mediante un círculo verde, posición del nervio óptico en un círculo azul y posición macular real localizada mediante una cruz blanca.

ii) Relación espacial con el nervio óptico

Primero es necesario determinar la posición del nervio óptico, basándose en que es la estructura que posee mayor contraste con el resto de la retina. Una vez determinada la posición del nervio óptico, la localización macular queda definida por el área subtendida entre el centro de esta estructura y una distancia de dos diámetros de disco óptico (DD) desplazándose hacia temporal inferior en un ángulo de 5 grados aproximadamente (ver figura 8) [39].

El algoritmo de detección macular basado en detección del nervio óptico comprende los siguientes pasos:

1. Separación de canales
2. Binarización aplicando umbral Yen
3. Utilización de operadores matemáticos
4. Identificación de regiones de interés
5. Determinación del centroide de la región de interés (nervio óptico)
6. Determinación de la lateralización de la retinografía (ojo derecho o izquierdo)
7. Desplazamiento desde el centro del nervio óptico 500 píxeles hacia temporal inferior en un ángulo de 5 grados

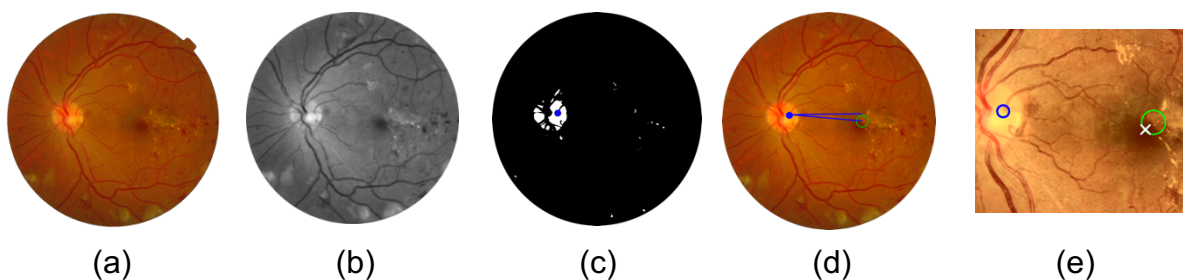


Figura 8. Método de detección macular basado en detección del nervio óptico y la distancia angular a la mácula. (a) retinografía original, (b) extracción del canal verde, (c) imagen binarizada aplicando filtro Yen y determinación del centroide, (d) determinación de la distancia angular al centro macular, (e) representación de la estimación de la posición macular mediante un círculo verde, posición del nervio óptico en un círculo azul y posición macular real localizada mediante una cruz blanca.

iii) Basado en la información entregada por la arcada temporal vascular

Este método se basa en la información entregada por la arcada temporal vascular y sus vasos terminales que apuntan al área macular, delimitando al área macular sin presencia de vasos retinales (ver figura 9)[40].

El algoritmo de detección macular basado en los vasos retinales comprende los siguientes pasos:

1. Separación de canales
2. Filtrado gaussiano
3. Binarización aplicando Meijering
4. Utilización de operadores matemáticos
5. Determinación de un skeleton
6. Determinación del centroide mediante los puntos terminales cercanos a la zona macular

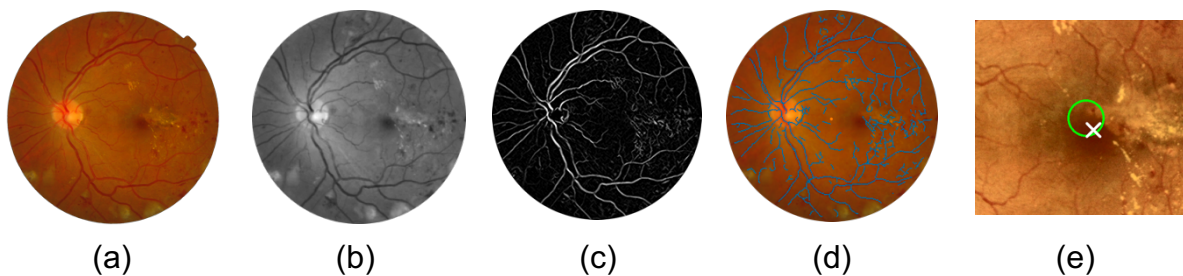


Figura 9. Método de detección macular basado en la localización de los vasos retinales. (a) retinografía original, (b) extracción del canal verde, (c) imagen binarizada aplicando filtro Meijering, (d) *skeleton* de los vasos retinales determinación del centroide, (e) representación de la estimación de la posición macular mediante un círculo verde y posición macular real localizada mediante una cruz blanca.

Implementación del método de aprendizaje de máquinas *U-net*

La estimación de la posición macular fue determinada mediante el entrenamiento de una red generativa adversaria tipo *U-net*, entrenando con 1480 retinografías y utilizando alrededor de 1,9 millones de parámetros. Se generaron máscaras binarias construidas a partir de la información de la posición macular de la base de datos MESSIDOR, implementando técnicas de aumento de datos específicos para retinografías, buscando evitar el sobreajuste de la red. Una vez entrenado el modelo, se procesaron las imágenes pertenecientes al conjunto de validación, obteniendo imágenes segmentadas compuestas por un ROI (*region of interest*) cuyo centroide correspondió a las coordenadas estimadas de la posición macular (ver figura 10).

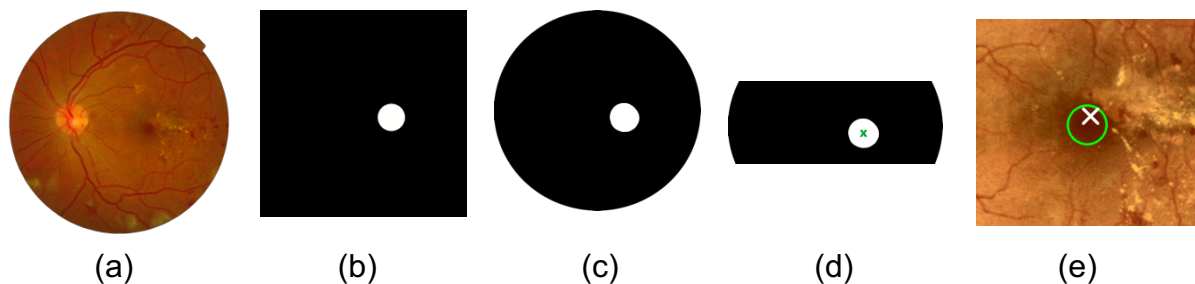


Figura 10. Método de detección macular basado en aprendizaje de máquinas. Se representa el procesamiento realizado a partir de *U-net* para localizar la posición macular: (a) retinografía original, (b) máscara binaria (*ground truth*) utilizada en el entrenamiento de la red, (c) imagen segmentada entregada por la *U-net*, (d) determinación del centroide del ROI de la imagen segmentada, (e) representación de la posición macular estimada por la *U-net* en un círculo verde y posición macular real localizada mediante una cruz blanca.

5.2.3. - Comparación el riesgo de presentar maculopatía utilizando la posición de las lesiones determinadas por el algoritmo DART

La posición de las lesiones obtenidas mediante el procesamiento de retinografías DART se incorporó a los datos obtenidos de la posición macular y lesiones *ground truth*. Para llevar a cabo el análisis, se determinó la posición de las lesiones localizadas por el sistema DART (ver figura 11) mediante el siguiente procedimiento:

1. Recorte de la retinografía (crop) disminuyendo área sin información en la retinografía (marco negro)
2. Separación de canales rojo (información de las lesiones rojas) y verde (información de las lesiones brillantes)
3. Binarización de la imagen utilizando umbral de 140 (medido en una escala de grises de 0 a 255). Dato recomendado por el equipo DART
4. Obtención de las coordenadas mediante algoritmo de *blob detection*
5. Recopilación en una base de datos consolidando información del centro macular, diagnóstico de MD y posición de las lesiones brillantes y rojas (base de datos MGLD, compuesta por la posición macular *ground truth* y lesiones obtenidas mediante el sistema DART)
6. Generación de vectores de características E,S y F
7. Implementación de balance de clases, obteniéndose las bases de datos 80, 80_80, 80_BC y 80_9 a partir de base de datos MGLD

Se evaluaron en paralelo los rendimientos de los clasificadores *Logistic Regression*, *Support Vector Machine*, *Decision Tree*, y *Random Forest* mediante validación simple para posteriormente consolidar los resultados obtenidos en el objetivo 4.

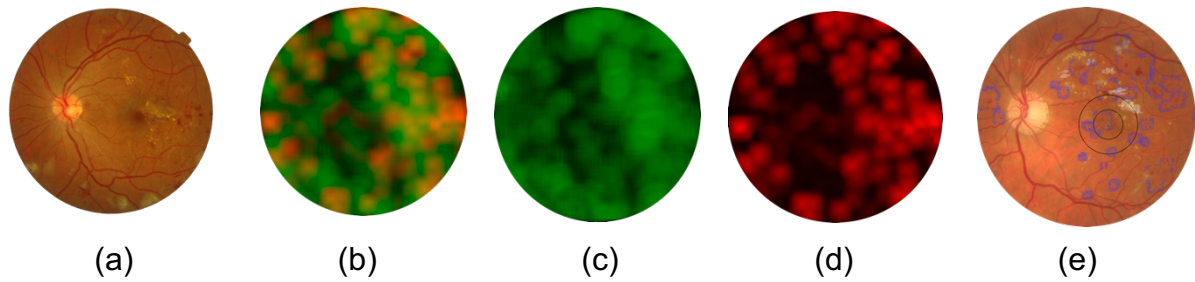


Figura 11. Procesamiento de la información entregada por el algoritmo DART. El mapa de calor entregado por DART es procesado para determinar las lesiones presentes en cada retinografía: (a) retinografía sin procesar, (b) mapa de calor (*Heatmap*), (c) canal verde, corresponde a la información de lesiones brillantes (exudados duros y blandos), (d) canal rojo, corresponde a la información de lesiones rojas (microaneurismas y hemorragias), (e) representación gráfica de las lesiones sobre la retinografía original, lesiones brillantes corresponden a zonas blancas y lesiones rojas a zonas azules.

5.2.4. - Evaluación del riesgo de presentar maculopatía integrando un estimador de posición macular, la localización de lesiones DART y un clasificador automático

Este objetivo busca analizar el comportamiento de los clasificadores y el peso atribuible a cada entrada separada según el tipo de procesamiento (manual o automatizado para las coordenadas) y entrada de datos correspondiente (tipo de balance de clase implementado o vector de características utilizado). Para ello fue seleccionada la base de datos IDRiD la cual fue pre-procesada con el objetivo de determinar tanto la posición de las lesiones presentes en cada retinografía como la posición macular. La segmentación manual de las lesiones ya estaba presente como parte de la base de datos y conformó el *ground truth* de posición de las lesiones retinales. La segmentación automática fue incorporada utilizando las redes intermedias del módulo de inteligencia artificial del sistema DART, entregando mapas de calor (*heat maps*) que definieron la localización de las lesiones a un determinado umbral (140 en una escala de grises de 0 a 255) determinando la probabilidad de que un píxel pertenezca a la clase lesión roja o brillante. La información de la posición macular se incorporó mediante su etiquetado manual, conformando el *ground truth* de la posición macular. El etiquetado automático de la localización macular se logró mediante el algoritmo de aprendizaje de máquinas *U-net*, método escogido en el objetivo 2.

Esta información fue consolidada determinando las distancias entre las lesiones *ground truth* a la posición macular *ground truth* y la posición macular *U-net*. Asimismo, se determinó las distancias de las lesiones obtenidas a partir del algoritmo DART y la posición macular *ground truth* y la posición macular *U-net* respectivamente, conformando las entradas MGLG, MGLD, UMLG y UMLD (Ver Tabla 1 Abreviaturas de cada base de datos generada a partir de IDRiD). Cada entrada de datos fue procesada balanceando sus clases y generando los vectores de características E, S y F.

Se seleccionó el algoritmo de validación cruzada, para determinar el rendimiento de cada clasificador en una sola métrica: *Compute Area Under the Receiver Operating Characteristic Curve* (ROC AUC), permitiendo visualizar de manera global el rendimiento alcanzado por cada clasificador.

La validación cruzada permite comparar los diversos métodos de clasificación y entrega una idea de qué tan bueno es su desempeño. Una vez escogido el clasificador y el vector de origen, fue posible graficar sus desempeños y analizar su comportamiento.

6.- Resultados

6.1.- Evaluación del desempeño alcanzado en la determinación del riesgo de presentar maculopatía

6.1.1.- Validación simple: *Logistic Regression*, *Decision Tree*

Los rendimientos obtenidos al procesar el conjunto de datos MGLG utilizando los clasificadores *Logistic Regression*, *Support Vector Machine*, *Decision Tree* y *Random Forest* permitieron identificar a los dos clasificadores que alcanzaron el mejor desempeño y diferenciar si utilizaron la posición macular (vector “E”) o la información general de las lesiones retinales sin considerar la posición macular (vector “S”). Los rendimientos alcanzados fueron obtenidos utilizando la siguiente información:

i) El clasificador *Logistic Regression*, incorporó exclusivamente la información macular (vector “E”) y utilizó balance de clases (submuestreo de la clase mayoritaria y sobremuestreo de la clase minoritaria equiparando la cantidad de ejemplos para su entrenamiento, base de datos 80_BC).

ii) El clasificador *Decision Tree*, incorporó exclusivamente la información macular (vector “E”), sin utilizar balance de clases.

El rendimiento alcanzado en las métricas de *Precision*, *Recall*, *F1 Score* y *Accuracy* fue de 100% para ambas clases de interés (sin maculopatía o con maculopatía), tal como se muestra en la tabla 12.

Tabla 12: Determinación del riesgo de presentar maculopatía mediante validación simple, utilizando el algoritmo de *Logistic Regression* y *Decision Tree* procesando los datos provenientes de la posición macular *ground truth* y lesiones *ground truth*.

	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>	<i>Support</i>
Sin Maculopatía	1.0	1.0	1.0	3
Con Maculopatía	1.0	1.0	1.0	21
<i>accuracy</i>			1.0	24
<i>macro avg</i>	1.0	1.0	1.0	24
<i>weighted avg</i>	1.0	1.0	1.0	24

Se observa que ambos clasificadores logran un alto desempeño identificando ambas clases de interés. Su diferencia solo radica en el manejo del balance de clases.

El mayor incremento lo obtuvo el clasificador *Logistic Regression*, al procesar los datos generados a partir del vector E, clasificando a ambas clases de interés de manera exitosa (presencia o ausencia de MD) logrando rendimientos de un 100% en las métricas de *Precision* y *Recall* mediante validación simple y un 99% en ROC AUC utilizando validación cruzada. También fue posible observar comportamientos similares en los resultados de los clasificadores *Support Vector Machine*, *Random Forest* y *Decision Tree*: al utilizar la información correspondiente a los vectores F y E sus

desempeños aumentaron pero al utilizar el vector S (procesando información general de las lesiones retinales) tendieron a disminuir.

Es importante considerar que los altos rendimientos alcanzados en esta primera etapa fueron debidos en parte al método de balance implementado: 80_80 y las entradas construidas a partir de los valores de referencia etiquetados por expertos, por lo que su estructura proporciona un porcentaje mínimo de error atribuible a la detección de lesiones y posición macular, en combinación con una distribución de datos y balance uniforme.

6.1.2.- Coeficiente de correlación de Pearson

Los resultados del Coeficiente de Correlación de Pearson permiten identificar entre la totalidad de factores que componen los vectores de interés a aquellos que logran una mejor correlación con el diagnóstico “presencia de maculopatía”. Al analizar los resultados alcanzados por el vector E sin realizar balance (base de datos 80) la característica “Lesión roja más cercana a la mácula” alcanzó la mejor correlación, con un coeficiente de -0.4426, “Lesión brillante más cercana a la mácula” con un coeficiente de -0.3966 seguida de “Presencia de lesión macular” con un coeficiente de -0.3901; cabe destacar que ambas características utilizaron la posición macular como dato de entrada. Si se analizan los resultados obtenidos utilizando balance de clases y al vector F, se identificaron otros factores no menos importantes en el peso de la clasificación, tales como “número de lesiones retinales totales” o “centroide de lesiones rojas” y “centroide de lesiones brillantes” que no consideran directamente la posición macular, se observa además que al implementar balanceo de clases el porcentaje de correlación mejora. Los resultados obtenidos para un conjunto de clases con y sin balance utilizando el vector E son presentados en la Tabla 13.

Tabla 13: Determinación del Coeficiente de Correlación de Pearson en base al vector E utilizando cada característica y mejor resultado obtenido utilizando balance de clases.

Característica	clasificación macular 80_80	clasificación macular 80
Lesión roja más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.881594	-0.442684
Lesión brillante más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.879311	-0.396637
Lesión más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.876452	-0.445612
Número de lesiones rojas localizadas a una distancia menor a un radio de 100 píxeles desde la mácula. (E)	0.214629	0.107423
Número de lesiones brillantes localizadas a una distancia menor a un radio de 100 píxeles desde la mácula. (E)	0.365432	0.174165
Número de lesiones rojas localizadas a una distancia menor a un radio de 200 píxeles desde la mácula. (E)	0.497790	0.195155
Número de lesiones brillantes localizadas a una distancia menor a un radio de 200 píxeles desde la mácula (E)	0.527206	0.271221
Número total de lesiones localizadas a una distancia menor a un radio de 200 píxeles desde la mácula. (E)	0.573545	0.294679
Presencia de lesión macular dentro de un radio de 200 píxeles (0=ausencia, 1=presencia) (E)	0.874599	0.390189

La tabla presenta los resultados obtenidos por el vector E, ordenados de manera decreciente según su correlación sea positiva, cercana a cero y negativo.

Los análisis realizados a partir de los resultados de las métricas de *Precision*, *Recall*, *F1 Score* y *Accuracy*, determinaron que el vector que alcanzó mejor rendimiento utilizó exclusivamente información proveniente de la posición macular, logrando un rendimiento de 1.0 en la clasificación de cada clase de interés.

Estos resultados permiten establecer que el área macular y su relación con las lesiones retinales se comportan como el factor más relevante para la determinación del riesgo de maculopatía.

6.2.- Comparación de algoritmos de procesamiento de imágenes y aprendizaje de máquinas en la estimación de la posición macular

El algoritmo que alcanzó mejor desempeño fue el método de aprendizaje de máquinas (U-net), alcanzando un error acumulado medio (17.6172 ± 58.73 píxeles) equivalentes a $0.9 \mu\text{m}$ de distancia al centro real macular, significativamente menor a las tres técnicas de procesamiento de imágenes: método de vasos retinales (46.9965 ± 65.1321 píxeles), umbral de media (52.7355 ± 67.8410 píxeles) y nervio óptico (77.8048 ± 99.8321 píxeles). Este resultado se obtuvo calculando la distancia euclidiana entre la posición predicha y la posición real macular en un total de 318 imágenes, para luego sumar las diferencias mediante el error medio acumulado de cada método, tal como se puede observar en la Figura 12.

Estos resultados permiten establecer al método de aprendizaje de máquinas (*U-net*) como el mejor algoritmo en la tarea de detección macular entre los algoritmos implementados en este trabajo, alcanzando un desempeño que sobrepasó a cada uno de los métodos de procesamiento de imágenes. Este método fue utilizado en la detección de la posición macular en la base de datos IDRiD para su posterior procesamiento en conjunto con la posición de lesiones *ground truth* y de lesiones del sistema DART.

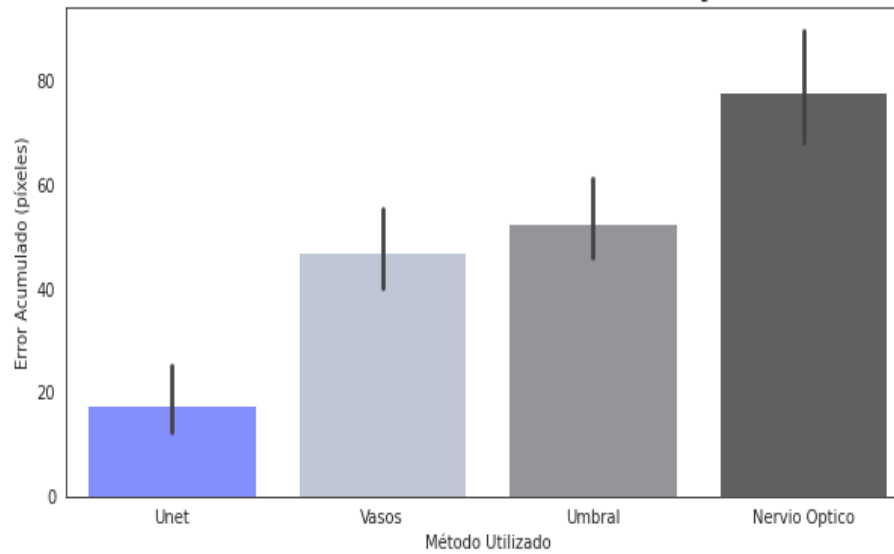


Figura 12. Resultados obtenidos en la determinación de la posición macular utilizando métodos de aprendizaje de máquinas (AM) y procesamiento de imágenes (PI). Se observa que el mejor rendimiento obtenido corresponde al método de aprendizaje de máquinas (*U-net*) con un error medio acumulado de 17.61 píxeles de diferencia respecto a la posición macular real. Estos resultados fueron obtenidos procesando la base de datos MESSIDOR en 318 retinografías correspondientes al subconjunto de pruebas.

Tabla 14: Datos de origen en la comparación entre los métodos de aprendizaje de máquinas (AM) y procesamiento de imágenes (PI).

Método utilizado en detección Macular	Error Acumulado (píxeles)	Desviación estándar (píxeles)	Mínimo (píxeles)	Máximo (píxeles)
Método <i>U-net</i> (AM)	17.617279	58.731213	0.098235	971.258178
Método vasos retinales (PI)	46.996543	65.132183	1.618997	953.718730
Método umbral macular(PI)	52.735591	67.841020	1.240499	970.382439
Método distancia angular al nervio óptico (PI)	77.804887	99.832163	1.702432	764.372502

6.3.- Determinación del riesgo de presentar maculopatía utilizando la posición de las lesiones retinales obtenidas a partir del sistema automatizado DART

6.3.1.- Validación simple: *Logistic Regression*, *Support Vector Machine*

Los rendimientos obtenidos en *Precision*, *Recall* y *F1-score* sobre el conjunto de datos que utilizaron la posición macular *ground truth* y lesiones provenientes del sistema DART (MGLD), permitieron comparar la diferencia entre el desempeño alcanzado utilizando la entrada posición macular *ground truth* y lesiones *ground truth* (MGLG). En la Tabla 15 se puede observar la diferencia en los resultados alcanzados por el algoritmo *Logistic Regression*, clasificador seleccionado en el objetivo 1 como el que alcanzó el mejor rendimiento utilizando la base de datos MGLG.

Tabla 15: Comparación de los resultados obtenidos mediante validación simple, por el algoritmo de *Logistic Regression* procesando la entrada de datos *ground truth* y la información obtenida a partir del sistema DART

	<i>Precision</i>		<i>Recall</i>		<i>F1 Score</i>		<i>Support</i>
	Mácula GT Lesión GT	Mácula GT Lesión DART	Mácula GT Lesión GT	Mácula GT Lesión DART	Mácula GT Lesión GT	Mácula GT Lesión DART	
Sin Maculopatía	1.0	0.50	1.0	0.33	1.0	0.4	3
Con Maculopatía	1.0	0.90	1.0	0.95	1.0	0.93	21
<i>accuracy</i>						0.87	24
<i>macro avg</i>	1.0	0.70	1.0	0.64	1.0	0.66	24
<i>weighted avg</i>	1.0	0.85	1.0	0.85	1.0	0.86	24

Se observa un descenso al comparar los rendimientos alcanzados al utilizar la información de la posición de las lesiones determinada por el sistema DART en contraste con los alcanzados por la posición *ground truth* de las lesiones, esta diferencia se hace más notoria en la clase “sin maculopatía” en donde disminuye a la mitad.

6.3.2.- Validación cruzada

Los resultados obtenidos utilizando el algoritmo de validación cruzada (ROC AUC) permitieron identificar al clasificador *Support Vector Machine* como el algoritmo que obtuvo el mejor desempeño utilizando como dato de entrada MGLD. Adicionalmente es posible analizar los rendimientos alcanzados mediante la validación simple, permitiendo visualizar las métricas de *Precision*, *Recall*, *F1 Score* y *Accuracy* que se resumen en la Tabla 16.

Tabla 16: Resultados obtenidos por el algoritmo de *Support Vector Machine* procesando la entrada de datos *ground truth* y la información obtenida a partir del sistema DART.

	<i>Precision</i>		<i>Recall</i>		<i>F1 Score</i>		<i>Support</i>
	Mácula GT Lesión GT	Mácula GT Lesión DART	Mácula GT Lesión GT	Mácula GT Lesión DART	Mácula GT Lesión GT	Mácula GT Lesión DART	
Sin Maculopatía	0.92	0.96	0.85	0.88	0.88	0.92	27
Con Maculopatía	0.82	0.86	0.90	0.95	0.86	0.90	21
<i>accuracy</i>					0.87	0.91	48
<i>macro avg</i>	0.87	0.91	0.87	0.92	0.87	0.91	48
<i>weighted avg</i>	0.87	0.92	0.87	0.91	0.87	0.91	48

Se puede observar un mejor rendimiento al alcanzado por el clasificador Logistic Regression al utilizar la información proveniente del sistema DART. Sus resultados igualmente muestran una disminución en el rendimiento de la clase de interés "Con Maculopatía".

Los resultados obtenidos permiten establecer valores de referencia que caracterizan el comportamiento de los distintos clasificadores seleccionados al utilizar los valores de entrada de las lesiones detectadas por el sistema DART. Es posible observar un descenso en el rendimiento del clasificador *Logistic Regression* (clasificador identificado en el objetivo 1 como el de mejor rendimiento al utilizar la posición macular y lesiones *ground truth*) en contraste con un menor descenso en los resultados alcanzados con el algoritmo *Support Vector Machine*.

6.4.- Evaluación del riesgo de presentar maculopatía integrando un estimador de posición macular, la localización de lesiones DART y un clasificador automático

Al analizar los resultados obtenidos a partir del procesamiento de cada base de datos, tipo de vector y método de balance, se pudo establecer que el clasificador *Support Vector Machine* fue el que logró el mejor desempeño, seguido de los clasificadores *Logistic Regression* y *Random Forest*. Estos resultados fueron obtenidos incorporando toda la información disponible sobre el conjunto de características de interés (vector F) tal como se puede observar en la Tabla 17.

Tabla 17: Extracto de los rendimientos utilizando validación cruzada (ROC AUC)

Clasificador	Vector	Balance	Score MGLG	std MGLG	Score MGLD	std MGLD	Score UMLG	std UMLG	Score UMLD	std UMLD
Support Vector Machine	F	80_80	0,993	0,013	0,965	0,048	0,987	0,017	0,961	0,047
Logistic Regression	F	80_80	0,992	0,013	0,976	0,029	0,981	0,023	0,976	0,022
Random Forest	F	80_80	0,980	0,037	0,965	0,057	0,975	0,043	0,962	0,052
Support Vector Machine	F	80_9	0,955	0,062	0,869	0,146	0,926	0,092	0,851	0,164
Logistic Regression	F	80_9	0,946	0,074	0,887	0,092	0,903	0,124	0,902	0,086
Random Forest	F	80_9	0,914	0,123	0,907	0,115	0,937	0,059	0,871	0,150
Decisión Tree	F	80_80	0,908	0,075	0,924	0,050	0,924	0,077	0,912	0,066
Decisión Tree	F	80_9	0,730	0,165	0,765	0,172	0,783	0,098	0,798	0,150

Se presentan en orden decreciente los resultados de cada clasificador, base de datos, seleccionando los resultados del vector F y dos métodos de balance de clases 80_80 y 80_9. Destacados en negrita los resultados obtenidos por la entrada de datos MGLG que utiliza la información de la posición macular y lesiones *ground truth*.

6.4.1.- Entrada de datos

Al graficar los resultados obtenidos por los clasificadores *Support Vector Machine* y *Logistic Regression* se observa de manera general que sus rendimientos decrecen según se utilice la entrada de datos MGLG, UMLG, MGLD y UMLD. Adicionalmente se corrobora que ambos clasificadores alcanzan los mejores desempeños al utilizar los vectores F y E, tal como se puede observar en la figura 13.

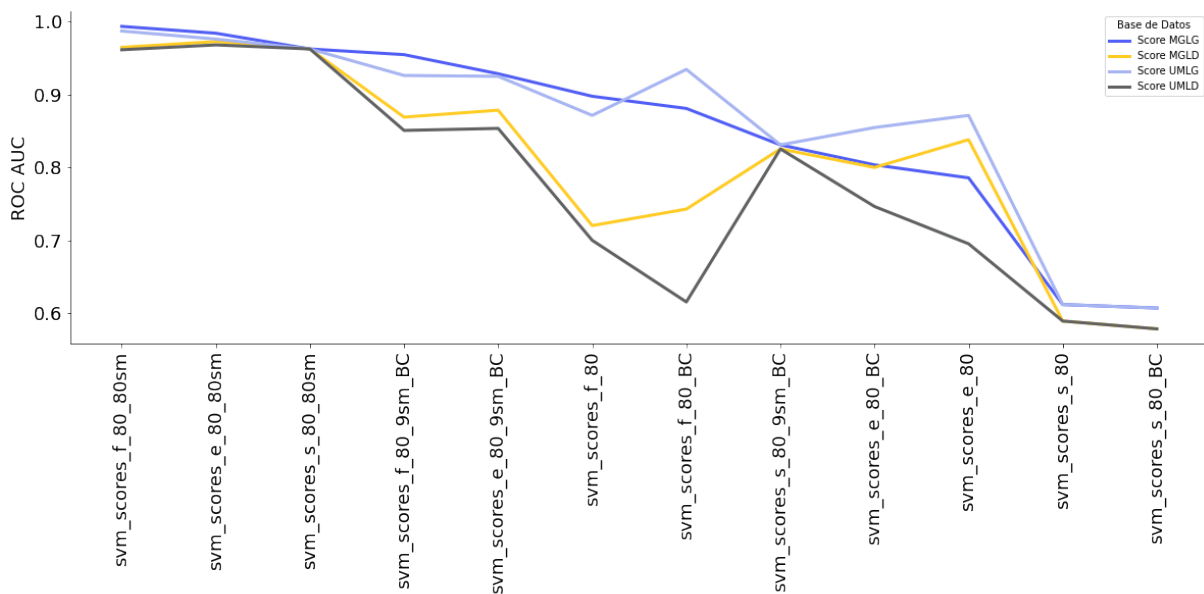


Figura 13. Comparación de los rendimientos alcanzados por el clasificador *Support Vector Machine* en cada entrada de datos utilizando validación cruzada (ROC AUC). Se presentan en orden decreciente los resultados de cada clasificador y su combinación con cada método de balance y vector utilizado. Se representa en líneas de colores según el rendimiento alcanzado por cada entrada de datos: en azul para MGLG, amarillo para MGLD, celeste para UMLG y gris para UMLD. Las siglas utilizadas representan los componentes de cada combinación de datos: *Support Vector Machine* (svm), vector F(f), balance 80 (80), balance 80_BC (80_BC), balance 80_80 (80_80sm) y balance 80_9(80_9sm).

6.4.2.- Vectores de características

Si comparamos nuevamente la información proveniente de la posición macular (vector “E”) y la información general de las lesiones retinales sin considerar la posición macular (vector “S”) utilizada en la determinación del estado macular, los resultados de las métricas de validación cruzada sobre el conjunto de datos MGLG permiten corroborar que los algoritmos *Logistic Regression* y *Support Vector Machine* como los clasificadores con mejor desempeño. Estos resultados utilizaron la información de la posición macular, alcanzando un ROC AUC de 0.9904 y 0.9841 respectivamente (Ver anexo tabla 5, Resultados ROC AUC).

Los mejores rendimientos fueron alcanzados por los clasificadores *Logistic Regression*, *Support Vector Machine* y *Random Forest*, utilizando la información proveniente de la posición macular. Adicionalmente es posible observar que independientemente del vector utilizado, al procesar la entrada de datos (MGLG), los mejores resultados fueron

alcanzados utilizando los métodos de balance 80_80 y 80_9. El desempeño más bajo correspondió al clasificador *Decision Tree* utilizando el vector S, mediante el método de balance de clases 80_BC (Ver figura 14).

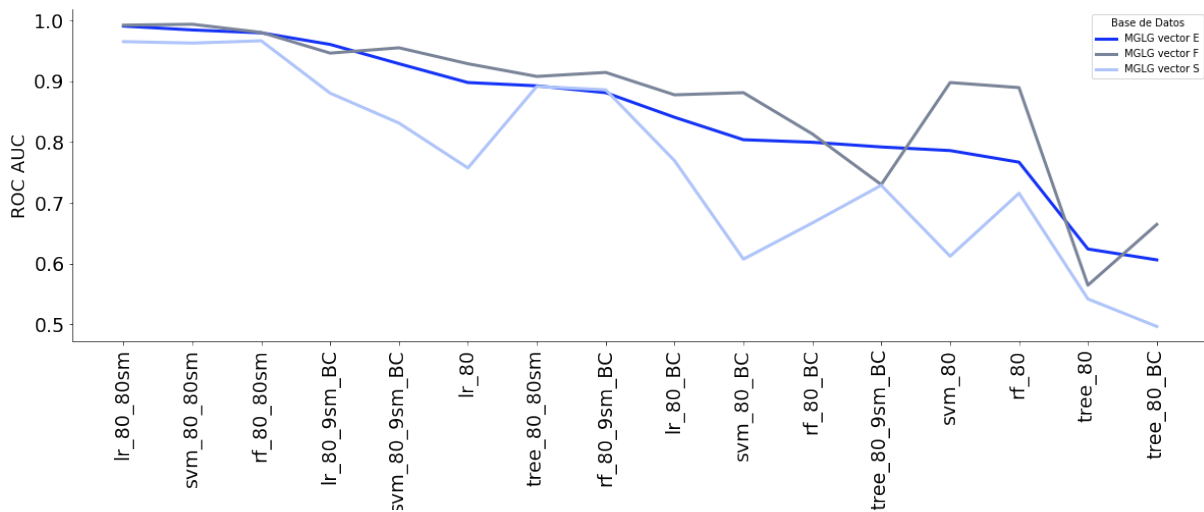


Figura 14. Resultados de los rendimientos utilizando validación cruzada (ROC AUC) utilizando la entrada de datos MGLG. Se presentan en orden decreciente los resultados de cada clasificador, base de datos, balance y vector utilizado. Se representa en líneas de colores según el rendimiento alcanzado por cada entrada de datos: vector F en gris, vector E en azul y vector S en celeste. Las siglas utilizadas representan los componentes de cada combinación de datos: Support Vector Machine (svm), vector F(f), balance 80 (80), balance 80_BC (80_BC), balance 80_80 (80_80sm) y balance 80_9(80_9sm).

6.4.3.- Comparación de los rendimientos alcanzados por cada clasificador

Al analizar el desempeño global alcanzado por cada clasificador, los resultados obtenidos a partir del vector F y el método de balance 80_80, se observa un comportamiento relativamente similar entre los clasificadores *Support Vector Machine* (0.993), *Logistic Regression* (0.992) y *Random Forest* (0.980), rendimientos muy por encima de lo alcanzado por el clasificador *Decision Tree* (0.907).

Al procesar la base de datos MGLG y ordenar los resultados alcanzados de manera decreciente, los clasificadores se comportan de la siguiente manera: *Support Vector Machine* (0.993), *Logistic Regression* (0.992), *Random Forest* (0.980) y *Decision Tree* (0.907). Al comparar los rendimientos obtenidos procesando la entrada de datos UMLD

se observa una disminución global y cambios en las posiciones de los clasificadores con mejor rendimiento: *Logistic Regression* (0.976), *Random Forest* (0.962), *Support Vector Machine* (0.961) y *Decision Tree* (0.912) (Ver figura 15).

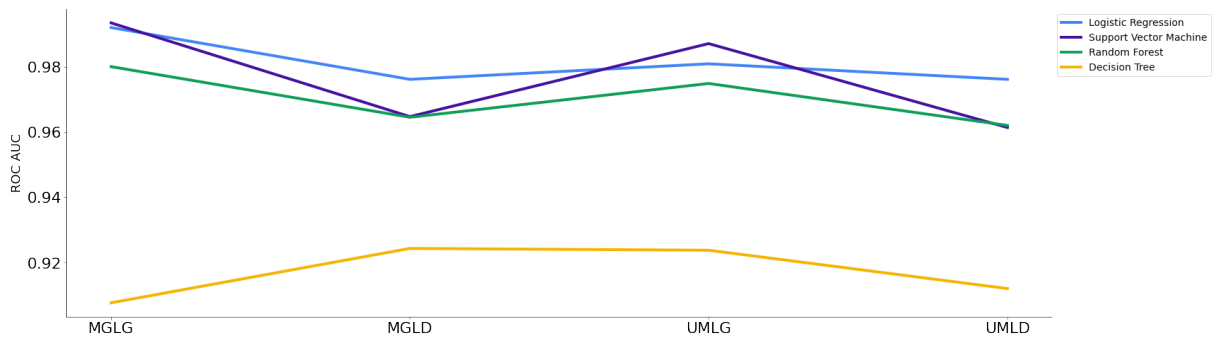


Figura 15. Rendimiento de cada clasificador utilizando el método de balance 80_80 mediante validación cruzada (ROC AUC). Se presentan en orden decreciente los resultados de cada clasificador y base de datos, utilizando el vector F. Se representa el rendimiento alcanzado por cada clasificador diferenciado en líneas de colores: *Logistic Regression* en celeste, *Support Vector Machine* en morado, *Random Forest* en verde y *Decision Tree* en amarillo.

Al analizar el comportamiento de los vectores E, S y F procesando las entradas de datos MGLG y UMLD, se observa que los vectores E y F se comportan de manera similar logrando el mejor desempeño general, sin embargo al utilizar los datos provenientes de la posición macular *U-net* y las lesiones localizadas por el sistema DART, los resultados del clasificador *Logistic Regression* muestran que el vector F se comporta de manera más estable. Al utilizar el clasificador *Support Vector Machine* es el vector E el que logra el mejor resultado al analizar UMLD, tal como se muestra en la figura 16.

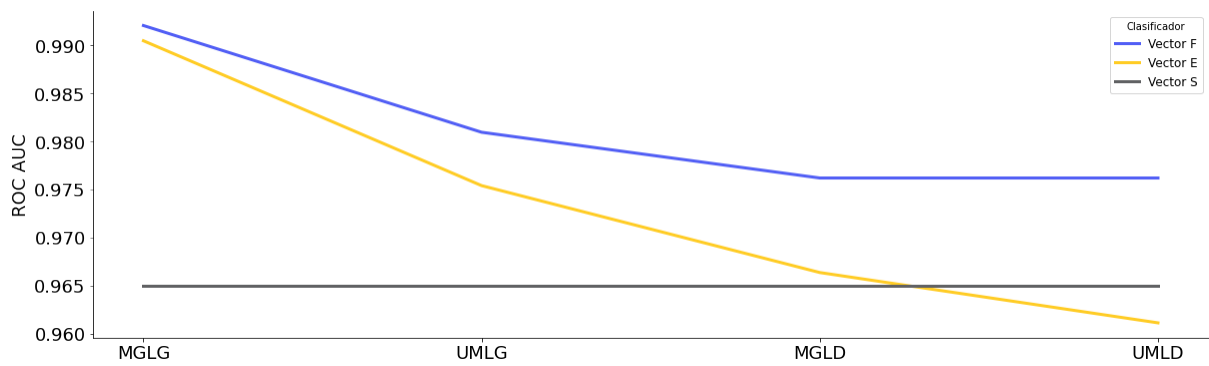


Figura 16. Rendimiento obtenido por el clasificador *Logistic Regression* en cada vector utilizando el origen de las coordenadas de la posición macular (*ground truth* o *U-net*) y las lesiones (*ground truth* o *DART*). Al analizar el comportamiento del clasificador *Logistic Regression* utilizando el vector F se observa que éste presenta el mejor rendimiento para la métrica ROC AUC. Al comparar los resultados originados en MGLG (mácula y lesiones *ground truth*), su desempeño disminuye a medida que clasifica los datos originados en el procesamiento de UMLG, MGLD y UMLD, datos obtenidos a partir del método de balance 80_80.

7.- Discusión

La RD constituye la tercera causa de ceguera a nivel mundial, pero es la primera en afectar a personas en edad productiva en países en vías de desarrollo [1]. Se estima que un tercio de las personas con *Diabetes Mellitus* desarrollará RD y de éstas la tercera parte requerirá de tratamiento oftalmológico debido a alteraciones maculares (MD). La ceguera causada por RD es prevenible en un 80% de los casos, por lo que las recomendaciones de la Organización Mundial de la Salud (OMS), instan a reforzar las políticas de prevención de la discapacidad visual evitable, incluida la ceguera [41]. Si bien dichas directrices sanitarias son consideradas e incorporadas a nivel nacional, diversos estudios demuestran que actualmente existe “un déficit significativo de oftalmólogos en los países en vías de desarrollo” [27], generando una necesidad urgente de mejorar la detección y prevención de la RD. Debe tenerse en cuenta que importantes lagunas en la disponibilidad de recursos oftalmológicos suponen un gran obstáculo. A nivel nacional la realidad no es distinta, según datos obtenidos del DEIS, durante el año 2018, tan solo un 37,5% de los pacientes derivados pudieron acceder a su evaluación anual de fondo de ojo.

En Chile el sistema de Diagnóstico Automatizado de Retinografías Telemáticas (DART) utiliza en uno de sus módulos estrategias de tamizaje automatizado. Si bien este sistema contribuye a la pesquisa de pacientes con sospecha de RD, es necesario señalar que el módulo de inteligencia artificial del sistema DART no incorpora en su algoritmo de clasificación la determinación de MD, por lo que deja afuera un importante factor de categorización de la RD.

En este trabajo de tesis se relacionan las lesiones retinales centrales con respecto al área macular, incorporando un clasificador binario de MD contribuyendo así a la clasificación automática de RD realizada actualmente, con el fin de priorizar la evaluación de los casos más agudos de esta enfermedad en el módulo de teleinforme. El clasificador desarrollado en este trabajo fue construido procesando retinografías pertenecientes a bases de datos públicas, localizando la posición macular de manera automática e identificando las lesiones retinales presentes en la retina a partir de la información entregada por el algoritmo DART.

Los resultados obtenidos permiten establecer que la integración de la información de la posición macular y la presencia de lesiones retinales constituyen los factores preponderantes en la tarea de determinación de MD. Por otro lado, en la tarea de estimación automática de la posición macular, se obtuvieron resultados concordantes con los reportados en la literatura, alcanzando desempeños cercanos a un 98% de precisión utilizando una red tipo *U-net*. Finalmente, al implementar un clasificador de maculopatía relacionando la detección automática de la posición macular y las lesiones retinales estimadas por DART, los algoritmos *Logistic Regression* y *Support Vector Machine* lograron alcanzar un rendimiento ROC AUC de 0.9761 y 0.9680 respectivamente, demostrando de esta manera la factibilidad de identificar MD automáticamente.

La discusión del presente trabajo será presentada y analizada siguiendo el orden de los objetivos planteados.

7.1.- Integración de la información de la posición macular y la presencia de lesiones retinales en la determinación de MD.

Si bien la definición clínica de MD implica por sí misma la localización de la posición macular, dentro del flujo de los modelos de clasificación automatizada este antecedente no necesariamente es considerado un elemento indispensable para llevar a cabo la tarea de clasificación de MD, la mayoría de los estudios analizados consideran que el etiquetado de presencia o ausencia de MD constituye un dato suficiente para resolver el problema de clasificación.

Con la finalidad de validar la contribución de la posición macular en la determinación de maculopatía en pacientes con RD, se diseñaron dos vectores construidos en base a la utilización a información de la posición macular y la distancia a las lesiones retinales, basándose en el supuesto de que si el clasificador utilizaba esta información y alcanzaba los mejores resultados, entonces su uso y valor predictivo quedaban confirmados.

Los resultados obtenidos por los vectores (E y S) sobre la base de datos MGLG corroboran esta premisa: los clasificadores que utilizaron como dato de entrada la información proveniente del vector E alcanzaron los rendimientos más altos: *Logistic*

Regression y *Decision Tree* (validación simple 100% en *Precision*, *Recall* y *F1-Score*) y *Logistic Regression* y *Support Vector Machine* (validación cruzada ROC AUC 0.99), resultados concordantes con los reportados por Chua *et al* [40] y Johny *et al* [41] (100% en *Precision*, *Recall* y *F1-Score*), Giancardo *et al* (AUC 0.94) y Singh *et al* (AUC 0.9654)[30].

7.2.- Determinación de la posición macular.

Los resultados alcanzados en este objetivo permitieron seleccionar al método de aprendizaje de máquinas (*U-net*) como el algoritmo que alcanzó el mejor desempeño en la localización automática de la posición macular. En términos de implementación y comparación del presente trabajo la selección del mejor algoritmo se basó en el análisis de la distancia euclidiana y el error medio acumulado entre la posición real y la estimada por cada método. No obstante, con la finalidad de llevar a cabo un análisis comparativo con los resultados encontrados en otros trabajos fue necesario tomar ciertas consideraciones al respecto de los criterios utilizados para reportar la exactitud alcanzada en la localización de la posición macular: Fleming *et al* [42] y Niemeijer *et al* [43] utilizaron el criterio 1R, es decir determinaron que la posición macular estaba correctamente localizada si la estimación propuesta estaba dentro de una distancia 1 diámetro de disco (DD), por otro lado, Tobin *et al* [44] y Yu *et al* [45] utilizan los criterios 2R y $\frac{1}{2}$ R como medidas de referencia, respectivamente (es decir, 2DD y $\frac{1}{2}$ DD). Finalmente Gegundez-Arias *et al* [4] reportan de manera más exhaustiva sus resultados en base a criterios de $\frac{1}{8}$ R, $\frac{1}{4}$ R, $\frac{1}{2}$ R, 1R y 2R. Debemos considerar que cada uno de estos resultados puede verse modificado en base a la escala con la cual se determinó el tamaño del diámetro de disco o la distancia umbral en píxeles a la cual la estimación de la posición macular era clasificada como correcta, Medhi *et al* [6] definen una distancia de 50 píxeles desde la posición macular en imágenes escaladas a 768 x 576 píxeles, considerando que esta reducción responde a la necesidad de normalizar los distintos tamaños de las retinografías utilizadas en su trabajo. Por otro lado Gegundez-Arias *et al* consideran una ecuación variable para determinar la proporción del tamaño del DD considerando las distintas resoluciones de las retinografías procesadas.

Otro aspecto a considerar y que ha sido muy poco discutido en la literatura fue la ausencia de un criterio unificado que permitiese una correcta comparación y validación de resultados entre estudios similares, encontrándose una gran variabilidad en el número de imágenes y criterios utilizadas para la determinación y validación de los resultados alcanzados. A modo de ejemplo cabe citar algunas publicaciones que utilizaron bases de datos con rangos muy variables en su composición, encontrando trabajos que utilizaron 35 retinografías Li *et al* [36], 345 retinografías, Tobin *et al* [44] hasta 1200 retinografías en el estudio de Gegundez-Arias *et al* [4]. En el presente trabajo se utilizaron 318 imágenes en la fase de validación pertenecientes a la base de datos MESSIDOR, correspondiendo al 30 % de los datos que no fueron utilizados en las fases de entrenamiento ni validación de los algoritmos de estimación de la posición macular.

Teniendo en cuenta los criterios mencionados anteriormente, es posible reportar un 98% de precisión en la detección de la posición macular utilizando el criterio 1 R (considerando $R=100$ pixeles) estimando correctamente la posición macular en 314 retinografías (de un total de 318 imágenes analizadas), alcanzando resultados similares a los reportados por los autores Gegundez-Arias *et al* [4](96.99%) y Jhonson *et al* (98.29 %) [46].

Finalmente, utilizando la red generativa adversaria tipo U-net, la estimación de la posición macular tuvo un desempeño muy superior a los otros métodos de procesamiento de imágenes citados en el presente trabajo. Se podría mencionar dentro de sus ventajas que su implementación y entrenamiento de la red tipo *U-net* requirió un tiempo considerablemente menor de ajuste. No obstante, su desventaja radica en el número significativamente alto de retinografías necesarias para su entrenamiento y el tiempo requerido para su etiquetado.

7.3.- Clasificación automatizada de maculopatía.

Debido al gran número de combinaciones resultantes y la alta variabilidad en los resultados de la validación simple, es que se seleccionaron dos métodos cuya combinación permitió un análisis exploratorio más simple, pero al mismo tiempo más robusto: la validación cruzada de K-iteraciones y ROC AUC. La ventaja de contar con

un solo valor consolidado permite graficar el rendimiento alcanzado por cada clasificador en cada escenario posible, permitiendo identificar y cuantificar la influencia de los otros componentes involucrados en los resultados obtenidos, tales como entradas de datos, métodos de balance, vectores de características y comportamiento propio de cada clasificador.

7.3.1- Clasificación de MD utilizando información proveniente del sistema DART

Al comparar los resultados alcanzados clasificando el riesgo de presentar MD mediante la estimación de la posición macular de manera automatizada (UMLG) y la localización de lesiones obtenida a partir del sistema DART (MGLD) se observa un descenso mayor en el rendimiento en MGLD. En el caso del clasificador *Support Vector Machine* el rendimiento varía desde un ROC AUC de 0.95 (MGLG) a un 0.86 (MGLD) atribuible a las diferencias en la precisión alcanzada en la detección de lesiones utilizando el sistema DART, por otro lado utilizando la estimación de la posición macular de manera automatizada (UMLG) se evidencia un menor descenso (ROC AUC de 0.92 en UMLG) en comparación al rendimiento alcanzado mediante la utilización de los datos provenientes de DART. Debido a que la información procesada en el clasificador automatizado de MD es dependiente de la correcta localización tanto de lesiones retinales como de la posición macular, se desprende que para alcanzar un mejor desempeño, es necesario mejorar el proceso de detección de lesiones obtenida mediante el procesamiento de los *heatmaps* proporcionados por DART. Las dificultades encontradas con este ajuste se relacionan a la ausencia de información correspondiente al *ground truth* de la localización de lesiones de las bases de datos utilizadas, problema analizado en “Consideraciones en la elaboración de bases de datos utilizadas en oftalmología” en el apartado anexos.

7.3.2.- Clasificación de MD utilizando la determinación automática de la posición macular, localización de lesiones basadas en el sistema DART y un clasificador automático

El rendimiento alcanzado por clasificador de MD, integrando la información de la detección automatizada de la posición macular, la localización de las lesiones DART y el clasificador automático de MD (ROC AUC 0.97 utilizando *Logistic Regression*) permitieron alcanzar resultados similares a los reportados en la literatura por autores como Giancardo *et al* [47] y Singh *et al* [30], demostrando la factibilidad de identificar MD automáticamente. Adicionalmente, es posible inferir que si se disminuye el error asociado a la determinación de lesiones y se mejora la estimación de la posición macular automatizada, es posible aumentar la precisión alcanzada por el clasificador automático de MD. En base a los antecedentes expuestos, es posible aceptar la hipótesis que establece que la determinación de la ubicación del área macular y su relación con las lesiones retinales detectadas previamente, incrementará la precisión diagnóstica del sistema de clasificación automatizada.

La detección de MD constituye un antecedente esencial en la evaluación temprana de RD, entregando información fundamental en la evaluación del pronóstico visual del paciente con RD. Debido a que el desarrollo de la MD está directamente relacionada con la RD, la detección temprana de ambas patologías es parte de los criterios de derivación actualmente implementados, por lo tanto aquellos algoritmos que realizan tamizaje de RD y que no detectan a la MD como una entidad aparte, pueden dar cabida a una inadecuada derivación de pacientes, obteniendo resultados por debajo del nivel óptimo requerido para la clasificación de pacientes con RD y MD [48], no obstante en diversas publicaciones es posible evidenciar que la detección de MD es caracterizada como un problema anexo y de menor relevancia dentro de la clasificación de RD. En un programa estándar de tamizaje de RD [49], es necesario detectar y clasificar tanto la retinopatía como la maculopatía [50]. Bajo estas consideraciones se desprende que un clasificador automatizado diseñado para ser utilizado como herramienta de tamizaje debe considerar la evaluación de ambas patologías.

La integración de estrategias de tamizaje automatizado en la detección temprana de MD permiten complementar el trabajo realizado por el sistema de tamizaje basado en

inteligencia artificial DART, entregando información relevante en el pronóstico diferencial de los pacientes con riesgo aumentado de pérdida visual, permitiendo mejorar los tiempos de espera, mejorando el acceso a la atención pública en salud y contribuyendo a alcanzar las metas sanitarias propuestas a nivel nacional.

8.- Conclusiones

En el presente trabajo de investigación se construyó, evaluó y validó la utilización de un clasificador de maculopatía diabética. Para ello se relacionó la información de la localización de las lesiones retinales centrales con respecto al área macular, generando así una herramienta complementaria al sistema de detección automatizada de RD DART.

La construcción de vectores de características elaborados a partir de la información conjunta entre las lesiones y la región macular, validó la utilización de la posición macular como un factor esencial en la determinación de MD aportando un mayor grado de interpretabilidad al modelo desarrollado. Adicionalmente el presente trabajo permite comprobar que la combinación de estos vectores y la implementación del balance de clases en las fases de entrenamiento incrementa los rendimientos obtenidos por los clasificadores de MD, corroborando de esta manera la importancia de la utilización de la información obtenida a partir de la localización macular.

La incorporación del clasificador de maculopatía diabética desarrollado en esta tesis permitirá mejorar el flujo de trabajo actualmente utilizado, identificando los casos con mayor riesgo de pérdida visual entre los pacientes que presentan RD, priorizando su atención y tratamiento oportuno, optimizando los recursos disponibles, disminuyendo los tiempos de espera e impactando positivamente en la calidad de vida de los pacientes con RD.

9.- Bibliografía

- [1] F. Verdaguer, J. T., Martínez G, F., & Barría V, «Guía Clínica Retinopatía Diabética». 2016. Accedido: dic. 15, 2021. [En línea]. Disponible en: <http://www.icoph.org/dynamic/attachments/resources/guiaclinicaretinopatiadiabetica2016.pdf> (última visita 15 dic 2021)
- [2] E. Treatment y D. Retinopathy, «Early Treatment Diabetic Retinopathy Study Design and Baseline Patient Characteristics: ETDRS Report Number 7», *Ophthalmology*, vol. 98, n.º 5, pp. 741-756, 1991, doi: 10.1016/S0161-6420(13)38009-9.
- [3] Z. Chen, S. S. Zhang, y H. M. Zhu, «Analysis of international clinical diabetic retinopathy disease severity scale», *Int. J. Ophthalmol.*, vol. 11, n.º 8, pp. 1394-1401, 2011, doi: 10.3969/j.issn.1672-5123.2011.08.025.
- [4] M. E. Gegundez-Arias, D. Marin, J. M. Bravo, y A. Suero, «Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques», *Comput. Med. Imaging Graph.*, vol. 37, n.º 5-6, pp. 386-393, jul. 2013, doi: 10.1016/j.compmedimag.2013.06.002.
- [5] R. Jalal, L. Kovacs, B. Harangi, B. Nagy, T. Peto, y A. Hajdu, «Combining algorithms for automatic detection of optic disc and macula in fundus images», *Comput. Vis. Image Underst.*, vol. 116, n.º 1, pp. 138-145, 2012, doi: 10.1016/j.cviu.2011.09.001.
- [6] J. P. Medhi y S. Dandapat, «An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images», *Comput. Biol. Med.*, vol. 74, pp. 30-44, jul. 2016, doi: 10.1016/j.combiomed.2016.04.007.
- [7] Ministerio de Salud (Minsal), «Guía de Práctica Clínica Retinopatía Diabética». 2017. Accedido: dic. 15, 2021. [En línea]. Disponible en: https://diprece.minsal.cl/wp-content/uploads/2018/06/RE_GPC-Retinopatia-Diabetica_2017v2.pdf (última visita 15 dic 2021)
- [8] P. L. Kaufman y A. Alm, *Adler Fisiología Del Ojo: Aplicación Clínica*. Harcourt Brace De España, 2003. [En línea]. Disponible en: <https://books.google.cl/books?id=01mHDarbV9UC>
- [9] T. D. Duane, W. Tasman, y E. A. Jaeger, *Duane's Clinical Ophthalmology*. J. B. Lippincott Company, 1976. [En línea]. Disponible en: <https://books.google.cl/books?id=6NohAQAAMAAJ>
- [10] A. Sopharak, B. Uyyanonvara, S. Barman, y T. H. Williamson, «Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods», vol. 32, pp. 720-727, 2008, doi: 10.1016/j.compmedimag.2008.08.009.
- [11] M. D. Abramoff, M. K. Garvin, y M. Sonka, «Retinal Imaging and Image Analysis», *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169-208, 2010, doi: 10.1109/RBME.2010.2084567.
- [12] Ministerio de Salud (Minsal), «Guía Clínica Retinopatía Diabética». pp. 1-32, 2010. Accedido: dic. 15, 2021. [En línea]. Disponible en: <https://www.minsal.cl/portal/url/item/7222754637ab8646e04001011f014e64.pdf> (última visita 15 dic 2021)
- [13] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, y P. F. Sharp, «Automated Assessment of Diabetic Retinal Image Quality Based on Clarity and Field Definition», *Investig. Ophthalmology Vis. Sci.*, vol. 47, n.º 3, p. 1120, mar. 2006, doi: 10.1167/iovs.05-1155.
- [14] A. Anaya-Isaza, L. Mera-Jiménez, y M. Zequera-Díaz, «An overview of deep learning in medical imaging», *Inform. Med. Unlocked*, vol. 26, p. 100723, 2021, doi: <https://doi.org/10.1016/j.imu.2021.100723>.

- [15] F. Tobar, «Aprendizaje De Máquinas». Universidad de Chile, 2020. Accedido: dic. 15, 2021. [En línea]. Disponible en: github.com/GAMES-UChile/Curso-Aprendizaje-de-Maquinas
- [16] D. E. Rumelhart, G. E. Hinton, y R. J. Williams, «Learning representations by back-propagating errors», *Nature*, vol. 323, n.º 6088, pp. 533-536, 1986, doi: 10.1038/323533a0.
- [17] Y. LeCun, P. Haffner, L. Bottou, y Y. Bengio, «Object Recognition with Gradient-Based Learning», en *Shape, Contour and Grouping in Computer Vision*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 319-345. doi: 10.1007/3-540-46805-6_19.
- [18] S. Muhammad, A. Muhammad, M. Adnan, Q. Muhammad, A. Majdi, y M. K. Khan, «Medical Image Analysis using Convolutional Neural Networks A Review», *J. Med. Syst.*, pp. 1-13, 2018.
- [19] I. Goodfellow *et al.*, «Generative Adversarial Nets», en *Advances in Neural Information Processing Systems*, 2014, vol. 27. [En línea]. Disponible en: <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
- [20] W. Weng y X. Zhu, «UNet: Convolutional Networks for Biomedical Image Segmentation», *IEEE Access*, vol. 9, pp. 16591-16603, 2021, doi: 10.1109/ACCESS.2021.3053408.
- [21] P. Isola, J. Y. Zhu, T. Zhou, y A. A. Efros, «Image-to-image translation with conditional adversarial networks», *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognit. CVPR 2017*, vol. 2017-Janua, pp. 5967-5976, 2017, doi: 10.1109/CVPR.2017.632.
- [22] U. Ishtiaq, S. Abdul Kareem, E. R. M. F. Abdullah, G. Mujtaba, R. Jahangir, y H. Y. Ghafoor, «Diabetic retinopathy detection through artificial intelligent techniques: a review and open issues», *Multimed. Tools Appl.*, 2019, doi: 10.1007/s11042-018-7044-8.
- [23] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan, y K. L. Zimmerman, «Review of Medical Decision Support and Machine-Learning Methods», *Vet. Pathol.*, vol. 56, n.º 4, pp. 512-525, jul. 2019, doi: 10.1177/0300985819829524.
- [24] H. T. James G, Witten D, *An Introduction to Statistical Learning: with Applications in R.* Springer Science & Business Media, 2013.
- [25] V. A. Dev y M. R. Eden, «Gradient Boosted Decision Trees for Lithology Classification», en *Proceedings of the 9th International Conference on Foundations of Computer-Aided Process Design*, vol. 47, S. G. Muñoz, C. D. Laird, y M. J. Realff, Eds. Elsevier, 2019, pp. 113-118. doi: <https://doi.org/10.1016/B978-0-12-818597-1.50019-9>.
- [26] L. Breiman, «Random Forests», *Mach. Learn.*, vol. 45, n.º 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.
- [27] J. T. Arenas-Cavalli, I. Abarca, M. Rojas-Contreras, F. Bernuy, y R. Donoso, «Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system», *Eye Basingstoke*, 2021, doi: 10.1038/s41433-020-01366-0.
- [28] M. M. Fraz, M. Badar, A. W. Malik, y S. A. Barman, «Computational Methods for Exudates Detection and Macular Edema Estimation in Retinal Images: A Survey», *Arch. Comput. Methods Eng.*, vol. 26, n.º 4, pp. 1193-1220, 2019, doi: 10.1007/s11831-018-9281-4.
- [29] T. Fushiki, «Estimation of prediction error by using K-fold cross-validation», *Stat. Comput.*, vol. 21, n.º 2, pp. 137-146, abr. 2011, doi: 10.1007/s11222-009-9153-8.
- [30] S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, y V. Lakshminarayanan, «Ophthalmic diagnosis using deep learning with fundus images – A critical review», *Artif. Intell. Med.*, vol. 102, n.º November 2019, p. 101758, 2020, doi: 10.1016/j.artmed.2019.101758.
- [31] W. Samek, T. Wiegand, y K.-R. Müller, «Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models», 2017, [En línea]. Disponible en: <http://arxiv.org/abs/1708.08296>

- [32] F. Doshi-Velez y B. Kim, «Towards A Rigorous Science of Interpretable Machine Learning», n.º MI, pp. 1-13, 2017.
- [33] J. T. Arenas-Cavalli, S. A. Ríos, M. Pola, y R. Donoso, «A Web-based Platform for Automated Diabetic Retinopathy Screening», *Procedia Comput. Sci.*, vol. 60, pp. 557-563, 2015, doi: 10.1016/j.procs.2015.08.179.
- [34] D. Retinopathy, «MESSIDOR Methods for Evaluating Segmentation and Indexing techniques Dedicated to Retinal Ophthalmology 1 . Issue and Context», *Communication*, 2004.
- [35] P. Porwal *et al.*, «Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research», *Data*, vol. 3, n.º 3, pp. 1-8, 2018, doi: 10.3390/data3030025.
- [36] Z. Li *et al.*, «An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs», *Diabetes Care*, vol. 41, p. dc180147, 2018, doi: 10.2337/dc18-0147.
- [37] V. Gulshan *et al.*, «Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs», vol. 94043, 2019, doi: 10.1001/jama.2016.17216.
- [38] S. Sekhar, W. Al-Nuaimy, y A. Nandi, «Automated Localisation of Optic Disk and Fovea in Retinal Fundus Images», p. 5, 2008.
- [39] M. Lalonde, M. Beaulieu, y L. Gagnon, «Fast and robust optic disc detection using pyramidal decomposition and Hausdorff-based template matching», *IEEE Trans. Med. Imaging*, vol. 20, n.º 11, pp. 1193-1200, nov. 2001, doi: 10.1109/42.963823.
- [40] T. Petsatodis, A. Diamantis, y G. Syrcos, «A Complete Algorithm for Automatic Human Recognition based on Retina Vascular Network Characteristics», 2004.
- [41] Organización Mundial de la Salud, *Salud ocular universal: un plan de acción mundial para 2014-2019*. Ginebra: Organización Mundial de la Salud, 2013. Accedido: nov. 22, 2021. [En línea]. Disponible en: <https://apps.who.int/iris/handle/10665/105956>
- [42] A. D. Fleming, K. A. Goatman, S. Philip, J. A. Olson, y P. F. Sharp, «Automatic detection of retinal anatomy to assist diabetic retinopathy screening», *Phys. Med. Biol.*, vol. 52, n.º 2, pp. 331-345, 2007, doi: 10.1088/0031-9155/52/2/002.
- [43] M. Niemeijer, M. D. Abramoff, y B. Van Ginneken, «Segmentation of the optic disc, macula and vascular arch in fundus photographs», *IEEE Trans. Med. Imaging*, vol. 26, n.º 1, pp. 116-127, 2007, doi: 10.1109/TMI.2006.885336.
- [44] K. W. Tobin, E. Chaum, V. Priya Govindasamy, y T. P. Karnowski, «Detection of anatomic structures in human retinal imagery», *IEEE Trans. Med. Imaging*, vol. 26, n.º 12, pp. 1729-1739, 2007, doi: 10.1109/TMI.2007.902801.
- [45] P. S. H Yu, S Barrigab, C Agurto, S Echegaray, M Pattichis, G Zamora, W Baumanc, «Fast Localization of Optic Disc and Fovea in Retinal Images for Eye Disease Screening». 1994.
- [46] R. J. Chalakkal, W. H. Abdulla, y S. S. Thulaseedharan, «Automatic detection and segmentation of optic disc and fovea in retinal images», *IET Image Processing*, vol. 12, n.º 11, pp. 2100-2110, 2018. doi: 10.1049/iet-ipr.2018.5666.
- [47] L. Giancardo *et al.*, «Exudate-based diabetic macular edema detection in fundus images using publicly available datasets», *Med. Image Anal.*, vol. 16, n.º 1, pp. 216-226, 2012, doi: <https://doi.org/10.1016/j.media.2011.07.004>.
- [48] M. R. Mathews y S. M. Anzar, «A comprehensive review on automated systems for severity grading of diabetic retinopathy and macular edema», *International Journal of Imaging Systems and Technology*, vol. 31, n.º 4, pp. 2093-2122, 2021. doi: 10.1002/ima.22574.
- [49] A. Chu, D. Squirrell, A. M. Phillips, y E. Vaghefi, «Essentials of a Robust Deep Learning

- System for Diabetic Retinopathy Screening: A Systematic Literature Review», *J. Ophthalmol.*, vol. 2020, pp. 1-11, nov. 2020, doi: 10.1155/2020/8841927.
- [50] P. H. Scanlon, «The English National Screening Programme for diabetic retinopathy 2003–2016», *Acta Diabetologica*, vol. 54, n.º 6. pp. 515-525, 2017. doi: 10.1007/s00592-017-0974-1.
- [51] J. A. Gliner, G. A. Morgan, N. L. Leech, J. A. Gliner, y G. A. Morgan, «Going deeper with convolutions», *Res. Methods Appl. Settings*, pp. 319-338, 2021, doi: 10.4324/9781410605337-29.
- [52] J. Krause *et al.*, «Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy», *Ophthalmology*, vol. 125, n.º 8, pp. 1264-1272, 2018, doi: 10.1016/j.optha.2018.01.034.
- [53] M. Banko y E. Brill, «Scaling to very very large corpora for natural language disambiguation», pp. 26-33, 2001, doi: 10.3115/1073012.1073017.

10.- Anexos

Tabla 1 : Resumen de los métodos de detección o graduación automática para edema macular utilizando retinografías digitales.

Author	Database	Method and Classifiers	Outcomes and Comments
Fleming et al (2007)	Private (14406)	Morphological image processing, exudate location.	Accuracy – 99.2% (NCSME) Accuracy – 97.3% (CSME)
Nayak et al (2009)	Private (350)	Matched correlation and neural Network	Sensitivity – 95.40% Specificity – 100%
Siddalingaswamy et al (2010)	Private (148)	Histogram thresholding, positional constraints and clustering	Sensitivity – 95.60% Specificity – 96.15%
Lim et al (2011)	MESSIDOR (88)	CLAHE and Watershed transform, exudate location	Sensitivity –80.90% Specificity –90.20% Accuracy – 85.20%
Ang et al (2011)	Private (90)	Mathematical morphology and neural network	Sensitivity – 90% Specificity – 100% Accuracy – 96.67%
Hunter et al (2011)	Private (1000)	Morphological features of exudates, intensity, colour, texture	Sensitivity – 97% Specificity – 80%
Akram et al (2012)	MESSIDOR (1200)	Exudate segmentation- Gabor Filter, Otus thresholding and Hough Transform-SVM classifier	Sensitivity –92.60% Specificity –97.80% Accuracy – 97.30%
Giancardo et al (2012)	HEI-MED and MESSIDOR (1200)	Wavelet transform, Kirsch edge detection, colour, and support vector machine	AUC – 0.94
Deepak and Sivaswamy (2012)	HEI-MED, MESSIDOR, MESSIDOR(IR), DIARETDB1 (590)	Motion patterns, Gaussian data description	Highest AUC – 0.99
Kumar et al (2012)	DRIVE	Hard Exudates were detected using an Extreme learning machine classifier (ELM)	Accuracy 98% Sensitivity 99.5% Specificity 85–98%
Alipour et al (2012)	Private (75)	Curvelet and Foveal Avascular Zone (FAZ) size	Sensitivity – 93% Specificity – 86%
Medhi and Dandapat (2012)	DRIVE, DIARETDB1, and HRF (174)	Top hat filtering, thresholding, and exudate location	Sensitivity – 97.5% Specificity – 98.7%
Tariq et al (2012)	MESSIDOR and STARE (1281)	Morphological features of exudates, Gabor filter, thresholding, texture and Gaussian mixture model	Accuracy – 97.30% (MESSIDOR), Accuracy – 97.89% (STARE)
Punnolil et al (2013)	DRIVE, DIARETDB1, STARE (251)	Morphological features of exudates, texture, and SVM	Sensitivity –96.89% Specificity –97.15%
Tariq et al (2013)	MESSIDOR and STARE (1281)	Gabor filter, Otus thresholding and support vector machine	Accuracy – 97.20% (MESSIDOR), Accuracy – 97.53% (STARE)
Zaidi et al (2013)	MESSIDOR (1200)	Gabor filter, morphological image processing, thresholding,	Sensitivity – 93.9%

		Bayesian classifier and exudate location	Specificity – 95.8% Accuracy – 94.1%
Medhi et al (2014)	DRIVE, DIARETDB1, ARIA (50)	Morphological image processing and exudate location	Sensitivity – 96%
Chowriappa et al (2013)	Private (90)	Texture and ensemble classifier	Accuracy – 96.70%
Chua et al (2013)	Private (300)	Texture and fuzzy classifier	Sensitivity – 100% Specificity – 100% Accuracy – 86.67%
Baby and Chandy et al (2013)	MESSIDOR (1200)	Gaussian data description, Dual- tree complex wavelet transform, and divergence	Precision rate –78.23%
Naguib et al (2013)	MESSIDOR (100)	Morphological image processing, EHD and distance measure	Precision rate –79.2%
Mookiah et al (2015)	Private and MESSIDOR (600)	Higher-order statistic, Naive Bayes, and Support vector	Accuracy – 95.56% (Private)
Ibrahim et al (2015)	Private (300)	Entropies, Fuzzy Sugeno, discrete wavelet transform, and neuro-fuzzy interference	Accuracy – 95.93% (MESSIDOR) Accuracy – 98.55%
Kunwar et al (2015)	MESSIDOR (1200)	Texture features, SVM classifier	Sensitivity-91% Specificity-75% Accuracy-80%
Saha et al (2016)	Private	The CFP from multiple visits of the patients were compared	Longitudinal evaluation of DME patients require accurate image registration
Johny et al (2016)	Several public datasets- details not provided	The detection of the lesion requires the localization and masking of the OD, detection of fovea, ROI extraction and motion pattern generation	High sensitivity of 100% Specificity of 74%
Lim et al (2017)	MESSIDOR	Local Binary pattern -extract gray level features and SVM classifier	Sensitivity –80% Specificity –70%
Punniyamoorthy (2018)	DIARETDB1 CFP database	Image processing techniques to detect the OD, exudates and the presence of macular oedema	exudate's detection: sensitivity 96.07%, selectivity 97.36% Accuracy 96.62% macular oedema: Sensitivity 97.75% selectivity 100% Accuracy 98.86%
Khojasteh (2018)	DIARETDB1 and e-Ophtha	Convolution neural networks used for automatic analysis of fundus images to detect signs of diabetic retinopathy using two approaches: patch and image-based analysis	Sensitivity was 0.96, 0.84 and 0.85 for detection of exudates, haemorrhages and microaneurysms, respectively when considering patch- based analysis. Overall accuracy for DIARETDB1 was 97.3% and 86.6% for e-Ophtha
Varadarajan (2020)	Dataset of 7072 images, 6039 were used for training, and 1033 for validation. Another dataset, EyePACS-DME-990 macula- centred images were also tested	Deep learning model used to predict Center involved DME using CFP	ROC–AUC of 0.89 (95% CI:0.87–0.91) corresponding to 85% sensitivity at 80%

Khojasteh (2019)	DIARETDB1 e-Ophtha	Compared multiple deep learning methods, and both supervised and unsupervised classifiers for improving the performance of automatic exudate detection, i.e., CNNs, pre-trained Residual Networks (ResNet-50) and Discriminative Restricted Boltzmann Machines	ResNet-50 with Support Vector Machines outperformed other networks with an accuracy and sensitivity of 98% and 0.99
Singh (2020)	IDRiD MESSIDOR	DME Net Algorithm which is built using Hierarchical Ensemble of CNNs (HE-CNN) for DME diagnosis	Accuracy of 96.12% Sensitivity of 96.32% Specificity of 95.84% F - 1 score of 0.9609 AUC 0.9654 MESSIDOR and IDRiD dataset
Auccahuasi (2020)	DIARETDB1 906 images	Alternative for the automatic recognition of hard exudates in retinal images using CNN	Sensitivity 0.92% Specificity-0.93%
Sulaiman (2020)	IDRiD	Deep learning- based DME-grading model is proposed for automatic DME grading of retinal fundus images	Accuracy 68% Precision 66% Recall 68% F1-score 65%
Alaguselvi et al (2020)	DIARETDB1	Used matched filter for analysis of CFP and tested using public dataset to detect micro-aneurysms, haemorrhages and exudates	They showed that compared with others, they had better results.
Hacisoftaoglu et al (2020)	Smartphone based images	ResNet50 network. The model was trained on smartphone-based synthetic images to explore the DR detection accuracy of smartphone- based retinal imaging systems and report AUC = 0.9978 on and independent test dataset	They overcame the problem with deep- learning regarding the size of the image dataset
Theera et al (2021)	DIARETDB1	Morphological techniques coupled with machine learning classifiers trained using patches	Morphological techniques coupled with machine learning is suitable for analysis of CFP to detect lesion
Quershi et al (2021)	EyePACS 54,000 images- self annotated	They report an automatic recognition of the DR stage is proposed based on a new multi-layer architecture of active deep learning (ADL).	They show the potential of using deep-learning for recognising the stages of DR using a public dataset, EyePACS. However, the annotation were performed by the authors and the details are not available
Chalakkala et al (2021)	MESSIDOR	Report a machine learning approach to identify the CSME cases from the other ME eyes using pre-trained deep-learning. Their work is to identify and overcome data- imbalance problems, and the need for large datasets for deep-learning	They report the ability of the pre-trained network to perform the feature analysis, the ability to work with imbalanced datasets and their investigation in the role of segmentation of fovea prior to the analysis.

Fuente: "A review of methods for automatic detection of macular edema"

Tabla 2: Resultados alcanzados por diversos estudios en la tarea de localización de la posición macular

Estudio	Método	Base de Datos Utilizada	Resultados obtenidos
Tobin <i>et al</i>		MESSIDOR	Acc: 76.00 %
Chin <i>et al</i>		MESSIDOR	Acc: 72.90 %
Akram <i>et al</i>		MESSIDOR	Acc: 97.20 %
Akram <i>et al</i>		HEIMED	Acc: 98.22 %
Huang <i>et al</i>	Cascade Neural Network	IDRiD	Acc: 99.03 %
Huang <i>et al</i>	Cascade Neural Network	MESSIDOR	Acc: 99.25 %
Li <i>et al</i>	Parabola fitting	Local, 35	Sensitivity : 100 %
Fleming <i>et al</i>	Visual characteristics of blood vessels, fovea, and OD	Local, 1056	Acc: 96.50 %
Tobin <i>et al</i>	Characteristics of blood vessels in relation to OD position	Local, 345	Acc.:92.50 %
Niemeijer <i>et al</i>	Cost function and a point distribution model	Local, 499	Acc.: 94.40 %
Niemeijer <i>et al</i>	k-NN regressor	Local, 500	Acc.: 96.80 %
Welfer <i>et al</i>	Selection of ROI and morphology	DRIVE	Acc.: 100 %
Qureshi <i>et al</i>	Combining the prediction of multiple algorithms	DIARETDB0	Acc.: 96.79 %
Giachetti <i>et al</i>	Fast radial symmetry transform	MESSIDOR	Acc.: 99.10 %
Gegundez-Arias <i>et al</i>	Priori known anatomical features and thresholding Priori known anatomical features and thresholding	MESSIDOR	Acc.: 96.92 %
Aquino <i>et al</i>	Visual and anatomical macula and OD feature-based method	MESSIDOR	Acc.: 98.24 %
Al-Bander <i>et al</i>	Multiscale sequential convolutional neural networks	MESSIDOR	Acc.: 96.60 %

Se listan de manera descriptiva los componentes de cada base de datos utilizada en la evaluación del rendimiento alcanzado en la detección de localización macular. Fuente: "Biomedical Signal Processing and Control Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc"

Tabla 3: Resumen de los métodos de localización automática de la posición macular utilizando retinografías digitales

Method	Accuracy %	Database used
Sinthanayothin <i>et al</i>	80.40	local
Sagar <i>et al.</i>	96.00	DRIVE, STARE
Niemeijer <i>et al</i>	96.80	local
Li and Chutatape	100.0	local
Xu <i>et al</i>	88.10	local
Welfer <i>et al</i>	92.13	DIARETDB1
Singh <i>et al</i>	96.61	DRIVE, STARE, local
Al-Bander <i>et al</i>	96.60	MESSIDOR
Jhonson <i>et al</i>	98.29	DRIVE, DIARETDB1,MESSIDOR, UoARD

Fuente: "Automatic detection and segmentation of optic disc and fovea in retinal images"

Tabla 4: Resultados validación cruzada

Clasificador	Vector	Balance	Score MGLG	std MGLG	Score MGLD	std MGLD	Score UMLG	std UMLG	Score UMLD	std UMLD
Support Vector Machine	F	80_80	0,993	0,013	0,965	0,048	0,987	0,017	0,961	0,047
Logistic Regression	F	80_80	0,992	0,013	0,976	0,029	0,981	0,023	0,976	0,022
Logistic Regression	E	80_80	0,990	0,013	0,966	0,043	0,975	0,024	0,961	0,051
Support Vector Machine	E	80_80	0,984	0,021	0,973	0,032	0,976	0,026	0,968	0,043
Random Forest	F	80_80	0,980	0,037	0,965	0,057	0,975	0,043	0,962	0,052
Random Forest	E	80_80	0,979	0,034	0,958	0,066	0,974	0,042	0,952	0,049
Random Forest	S	80_80	0,966	0,031	0,966	0,031	0,966	0,031	0,966	0,031
Logistic Regression	S	80_80	0,965	0,053	0,965	0,053	0,965	0,053	0,965	0,053
Support Vector Machine	S	80_80	0,963	0,062	0,963	0,062	0,963	0,062	0,963	0,062
Logistic Regression	E	80_9	0,960	0,069	0,881	0,156	0,878	0,135	0,885	0,158
Support Vector Machine	F	80_9	0,955	0,062	0,869	0,146	0,926	0,092	0,851	0,164

Logistic Regression	F	80_9	0,946	0,074	0,887	0,092	0,903	0,124	0,902	0,086
Support Vector Machine	E	80_9	0,929	0,111	0,879	0,166	0,925	0,104	0,854	0,222
Logistic Regression	F	80	0,929	0,115	0,755	0,201	0,875	0,146	0,818	0,201
Random Forest	F	80_9	0,914	0,123	0,907	0,115	0,937	0,059	0,871	0,150
Decisión Tree	F	80_80	0,908	0,075	0,924	0,050	0,924	0,077	0,912	0,066
Support Vector Machine	F	80	0,898	0,213	0,720	0,335	0,871	0,207	0,700	0,303
Logistic Regression	E	80	0,898	0,113	0,714	0,367	0,775	0,294	0,471	0,424
Decisión Tree	E	80_80	0,892	0,090	0,885	0,058	0,833	0,083	0,913	0,090
Decisión Tree	S	80_80	0,891	0,103	0,884	0,098	0,891	0,103	0,884	0,098
Random Forest	F	80	0,889	0,150	0,795	0,207	0,883	0,167	0,662	0,360
Random Forest	S	80_9	0,886	0,108	0,882	0,113	0,886	0,108	0,882	0,113
Support Vector Machine	F	80_BC	0,881	0,115	0,743	0,291	0,935	0,093	0,615	0,338
Random Forest	E	80_9	0,881	0,132	0,857	0,164	0,894	0,117	0,810	0,174
Logistic Regression	S	80_9	0,880	0,140	0,880	0,140	0,880	0,140	0,880	0,140
Logistic Regression	F	80_BC	0,877	0,175	0,792	0,235	0,820	0,224	0,795	0,186
Logistic Regression	E	80_BC	0,840	0,162	0,710	0,313	0,726	0,250	0,640	0,399
Support Vector Machine	S	80_9	0,831	0,151	0,825	0,158	0,831	0,151	0,825	0,158
Random Forest	F	80_BC	0,813	0,131	0,782	0,197	0,843	0,165	0,791	0,254
Support Vector Machine	E	80_BC	0,804	0,213	0,800	0,333	0,855	0,181	0,746	0,312
Random Forest	E	80_BC	0,799	0,191	0,804	0,192	0,825	0,130	0,679	0,318
Decisión Tree	E	80_9	0,792	0,183	0,812	0,139	0,788	0,172	0,706	0,126
Support Vector Machine	E	80	0,786	0,258	0,838	0,254	0,871	0,196	0,695	0,309
Logistic Regression	S	80_BC	0,769	0,213	0,769	0,213	0,769	0,213	0,769	0,213

Random Forest	E	80	0,767	0,242	0,790	0,197	0,719	0,255	0,661	0,322
Logistic Regression	S	80	0,757	0,279	0,757	0,279	0,757	0,279	0,757	0,279
Decisión Tree	F	80_9	0,730	0,165	0,765	0,172	0,783	0,098	0,798	0,150
Decisión Tree	S	80_9	0,729	0,164	0,721	0,156	0,729	0,164	0,721	0,156
Random Forest	S	80	0,715	0,126	0,730	0,117	0,715	0,126	0,730	0,117
Random Forest	S	80_BC	0,667	0,192	0,652	0,196	0,667	0,192	0,652	0,196
Decisión Tree	F	80_BC	0,664	0,248	0,599	0,255	0,700	0,261	0,702	0,274
Decisión Tree	E	80	0,624	0,212	0,604	0,219	0,560	0,229	0,539	0,182
Support Vector Machine	S	80	0,612	0,246	0,589	0,269	0,612	0,246	0,589	0,269
Support Vector Machine	S	80_BC	0,607	0,233	0,579	0,276	0,607	0,233	0,579	0,276
Decisión Tree	E	80_BC	0,606	0,249	0,505	0,176	0,614	0,256	0,563	0,204
Decisión Tree	F	80	0,564	0,203	0,638	0,182	0,714	0,252	0,599	0,223
Decisión Tree	S	80	0,542	0,234	0,535	0,220	0,542	0,234	0,535	0,220
Decisión Tree	S	80_BC	0,496	0,169	0,489	0,171	0,496	0,169	0,489	0,171

Cada columna presenta los resultados alcanzados por cada clasificador, vector, método de balance utilizado. Los resultados de los clasificadores están ordenados de manera decreciente de acuerdo a los desempeños obtenidos procesando la base de datos MGLG.

Ejemplos de clasificación de MD

A continuación se presentan algunos ejemplos de clasificaciones correctas e incorrectas: i) retinografías sin MD (Ver figura 1) y ii) retinografías con MD (Ver figuras 2, 3 y 4).

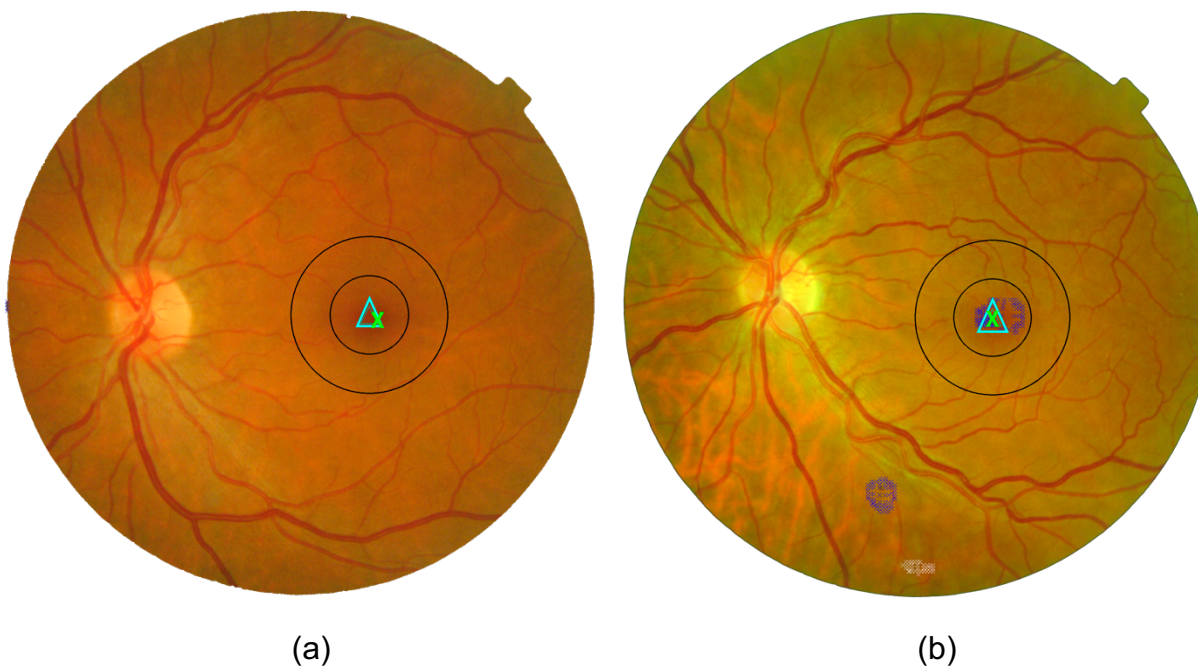


Figura 1. Ejemplos de clasificación de retinografías sin maculopatía diabética. En ambos casos se trata de retinas sin lesiones y máculas sanas, es decir sin diagnóstico de MD. En la figura (a) una imagen clasificada correctamente, tanto la estimación del centro macular por la *U-net*, sin ninguna lesión detectada por el algoritmo DART. La figura (b) también se trata de una retina sana, pero clasificada con MD debido a las lesiones maculares incorrectamente determinadas por el algoritmo DART, también existen errores en la determinación de algunas lesiones rojas y brillantes en la periferia retinal. Las lesiones rojas son representadas con marcas azules, las lesiones brillantes con marcas blancas sobre la retinografía, la posición macular real representada por una cruz verde y la estimada por la *U-net* con un triángulo calipso, círculos concéntricos representan distancia de 100 y 200 píxeles desde el centro estimado por la *U-net*. Imágenes tomadas desde la base de datos MESSIDOR y procesadas por los algoritmos *U-net* y DART.

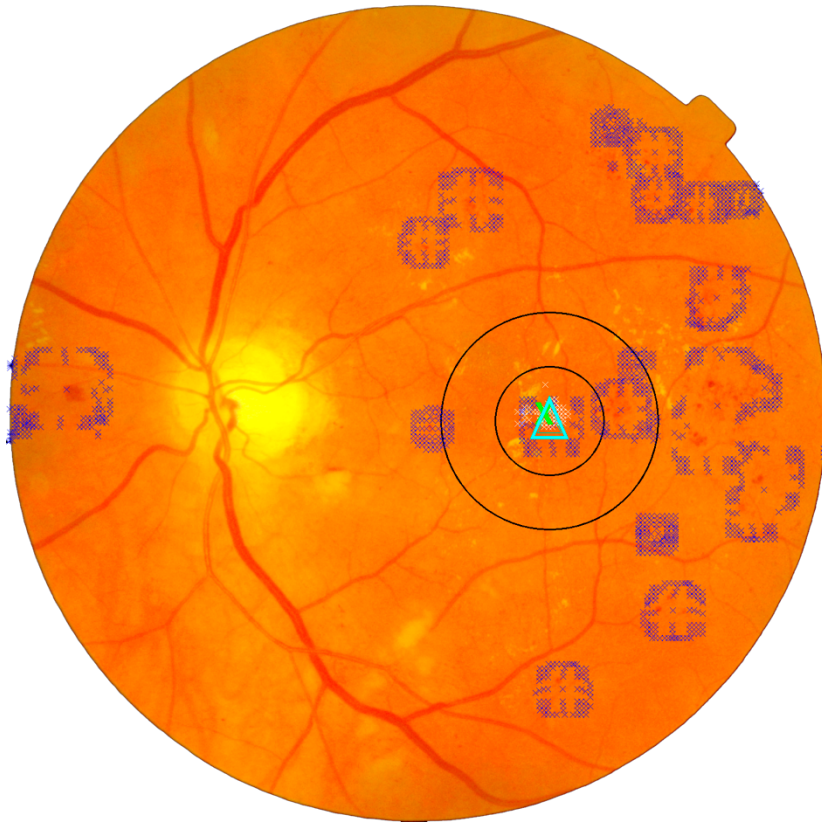


Figura 2. Retinografía correctamente clasificada con presencia de MD. Al analizar el comportamiento del clasificador en retinografías con diagnóstico de RD, es posible observar: posición macular determinada correctamente por la *U-net*, lesiones brillantes y rojas en general correctamente determinadas por el algoritmo DART. Las lesiones rojas son representadas con marcas azules, las lesiones brillantes con marcas blancas sobre la retinografía, la posición macular real representada por una cruz verde y la estimada por la *U-net* con un triángulo calipso, círculos concéntricos representan distancia de 100 y 200 pixeles desde el centro estimado por la *U-net*. Imágenes tomadas desde la base de datos MESSIDOR y procesadas por los algoritmos *U-net* y DART.

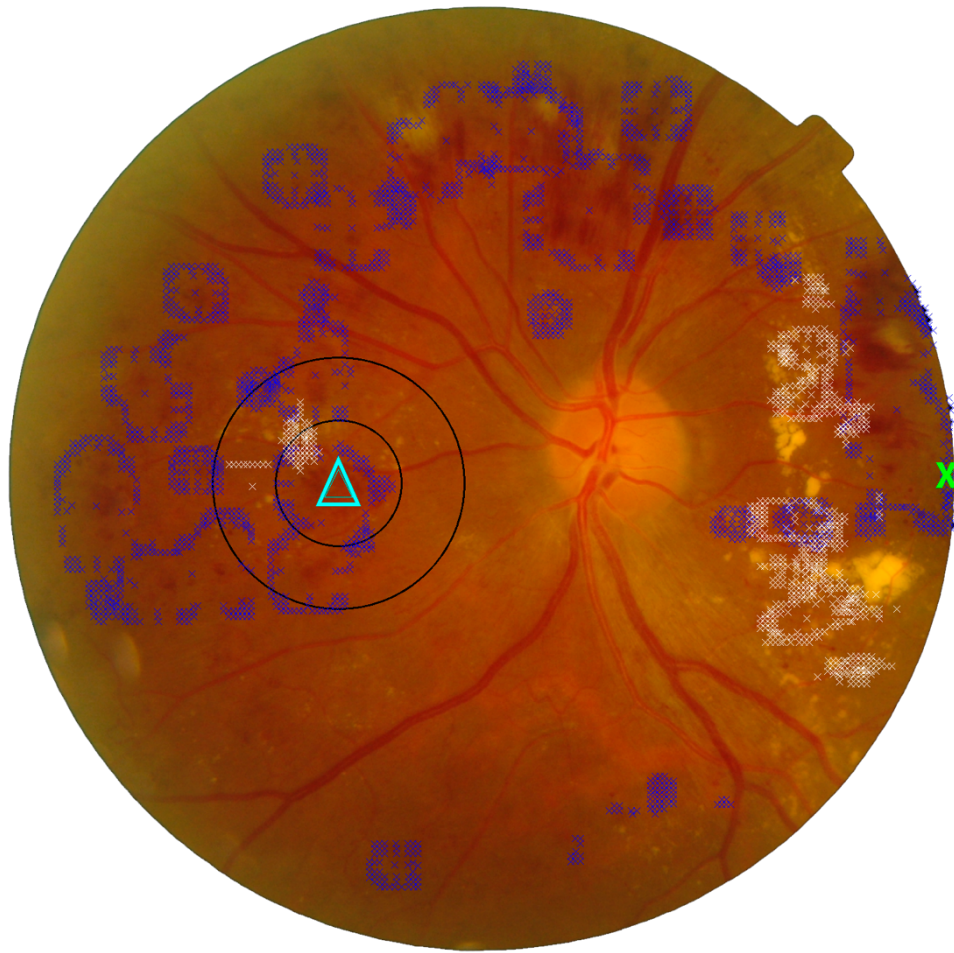


Figura 3. Retinografía incorrectamente clasificada con presencia de MD. A pesar de que la estimación de MD es correcta, la información utilizada por el clasificador se basó en una localización de la posición macular errada . Se observa que las lesiones brillantes y rojas están en general correctamente determinadas por el algoritmo DART, pero la posición macular está incorrectamente localizada por el algoritmo *U-net*. Las lesiones rojas son representadas con marcas azules, las lesiones brillantes con marcas blancas sobre la retinografía, la posición macular real representada por una cruz verde y la estimada por la *U-net* con un triángulo calipso, círculos concéntricos representan distancia de 100 y 200 pixeles desde el centro estimado por la *U-net*. Imágenes tomadas desde la base de datos MESSIDOR y procesadas por los algoritmos *U-net* y DART.

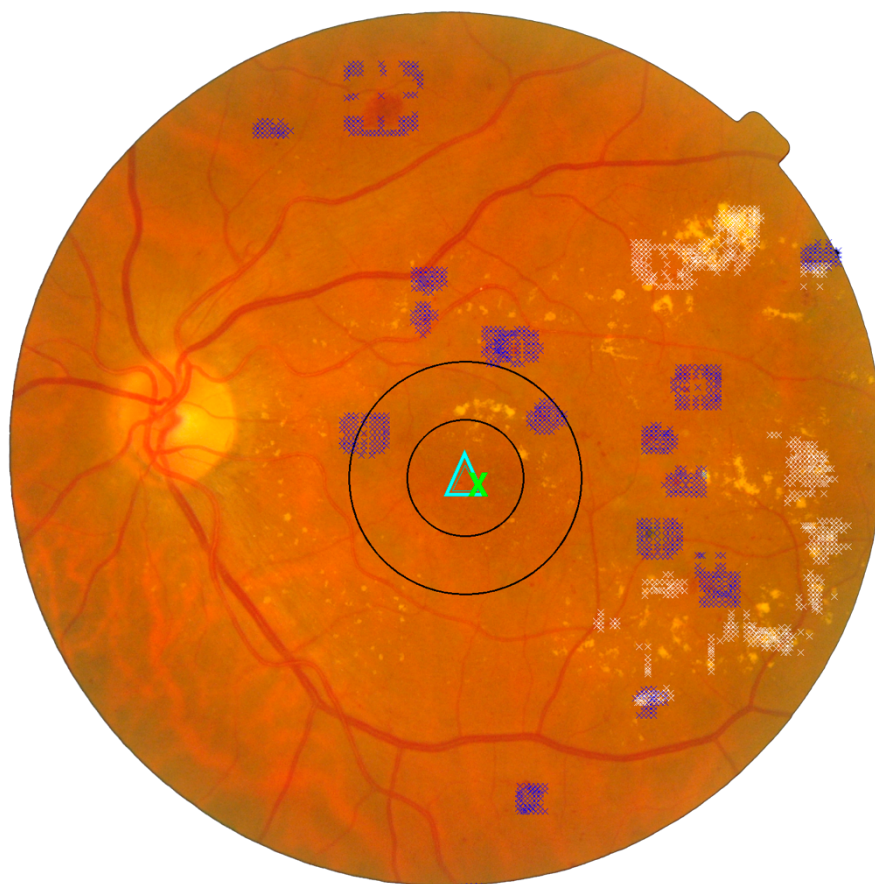


Figura 4. Retinografía incorrectamente clasificada con ausencia de MD. La posición macular está correctamente estimada por la *U-net*, pero las lesiones brillantes centrales no fueron detectadas por el algoritmo DART, entregando información incompleta para la adecuada clasificación de MD. Las lesiones rojas son representadas con marcas azules, las lesiones brillantes con marcas blancas sobre la retinografía, la posición macular real representada por una cruz verde y la estimada por la *U-net* con un triángulo calipso, círculos concéntricos representan distancia de 100 y 200 píxeles desde el centro estimado por la *U-net*. Imágenes tomadas desde la base de datos MESSIDOR y procesadas por los algoritmos *U-net* y DART.

Análisis complementario de los resultados del coeficiente correlación de Pearson

Al determinar el coeficiente de correlación de Pearson, el grado de asociación estimado se modifica y mejora al incorporar algún método de balance de datos, esto puede ser explicado,

debido al bajo número de casos disponibles de pacientes sanos, escenario en donde unos pocos casos fuera de rango (*outliers*) tienen una mayor repercusión en la determinación de la pendiente resultante. Al incorporar casos sin MD cambia este desbalance, corrigiendo la pendiente obtenida (o covarianza) y por tanto el valor obtenido del coeficiente de correlación (ver figuras 5 y 6).

Es posible diferenciar un alto coeficiente de correlación de Pearson de uno bajo, si se comparan las medias obtenidas para cada clase de interés, si la media alcanzada entre cada clase (ausencia o presencia de MD) presenta una gran diferencia en sus valores, el coeficiente resultante es alto (ver figura 6). Del mismo modo si la diferencia entre medias es menor, el coeficiente obtenido es bajo, incluso si se agregan casos que equilibran el número de casos, tal como se observa en las figuras 7 y 8.

Coeficiente Correlación de Pearson -0.3966
base de datos desbalanceada

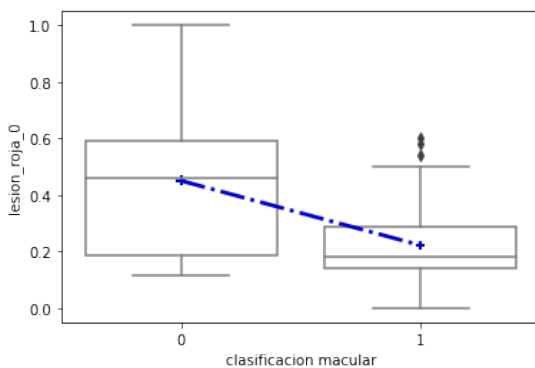


Figura 5. Número de lesiones rojas localizadas a una distancia menor a un radio de 200 píxeles desde la mácula.

Coeficiente Correlación de Pearson -0.8815
base de datos balanceada

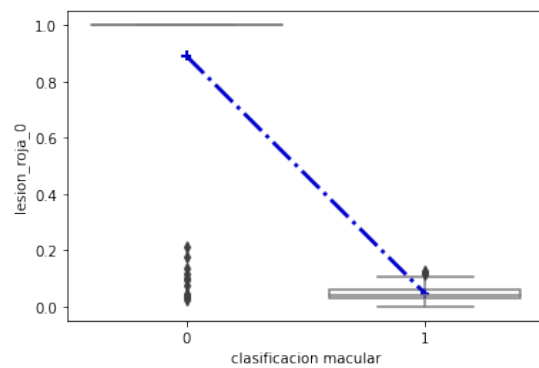


Figura 6. Número de lesiones rojas localizadas a una distancia menor a un radio de 200 píxeles desde la mácula.

Coefficiente Correlación de Pearson 0.1074
base de datos desbalanceada

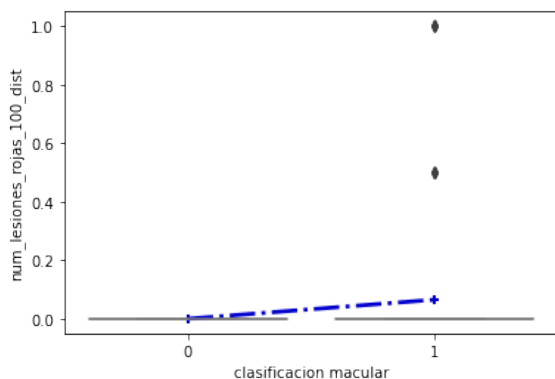


Figura 7. Número de lesiones rojas localizadas a una distancia menor a un radio de 100 píxeles desde la mácula.

Coefficiente Correlación de Pearson 0.2146
base de datos desbalanceada

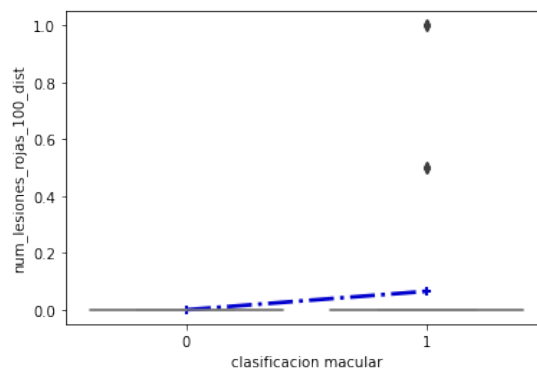


Figura 8. Número de lesiones rojas localizadas a una distancia menor a un radio de 100 píxeles desde la mácula.

Se representa en cada boxplot la distribución de un par de características representativas, diferenciadas en el resultado de su coeficiente es obtenido y si se aplicó métodos de balance de datos. Se destaca mediante una línea azul la diferencia obtenidas entre sus correspondientes medias para cada clase de interés (ausencia o presencia de MD). Se representa en el eje de las ordenadas el atributo correspondiente y en el eje de las abscisas la clasificación de maculopatía (0= ausencia de MD, 1 = presencia de MD) tal como se observa en las figuras 5 a la figura 8.

Tabla 5: Determinación del Coeficiente de Correlación de Pearson en base al vector F utilizando cada característica y su variación al utilizar balance de clases.

Característica	clasificación macular 80_80	clasificación macular 80_BC	clasificación macular 80_9	clasificación macular 80
Centroide del total de lesiones brillantes. (S)	-0.888556	-0.423411	-0.698720	-0.423411
Centroide del total de lesiones rojas. (S)	-0.881814	-0.296186	-0.676827	-0.296186
Lesión roja más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.881594	-0.442684	-0.676534	-0.442684
Lesión brillante más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.879311	-0.396637	-0.669041	-0.396637
Lesión más cercana a la mácula dentro de un radio de 200 píxeles (distancia en píxeles). (E)	-0.876452	-0.445612	-0.659662	-0.445612
Número de lesiones rojas localizadas a una distancia menor a un radio de 100 píxeles desde la mácula. (E)	0.214629	0.107423	0.139379	0.107423
Número de lesiones brillantes localizadas a una distancia menor a un radio de 100 píxeles desde la mácula. (E)	0.365432	0.174165	0.235033	0.174165
Número de lesiones rojas localizadas a una distancia menor a un radio de 200 píxeles desde la mácula. (E)	0.497790	0.195155	0.303098	0.195155
Número de lesiones brillantes localizadas a una distancia menor a un radio de 200 píxeles desde la mácula (E)	0.527206	0.271221	0.358072	0.271221
Número total de lesiones localizadas a una distancia menor a un radio de 200 píxeles desde la mácula. (E)	0.573545	0.294679	0.392729	0.294679
Número total de lesiones retinales brillantes. (S)	0.733855	0.144292	0.429689	0.144292
Porcentaje: Lesiones brillantes vs total de lesiones. (S)	0.770360	-0.268746	0.375878	-0.268746
Número total de lesiones retinales. (S)	0.809361	0.294661	0.559714	0.294661
Número total de lesiones retinales rojas. (S)	0.823329	0.396313	0.608462	0.396313

Porcentaje: Lesiones rojas vs total de lesiones. (S)	0.854190	0.268746	0.616066	0.268746
Presencia de lesión macular dentro de un radio de 200 píxeles (0=ausencia,1=presencia) (E)	0.874599	0.390189	0.673392	0.390189

La tabla presenta los resultados obtenidos por el vector F, compuesto por los elementos concatenados de los vectores E y S, ordenados de manera decreciente según su correlación sea positiva, cercana a cero y negativo. Se puede evidenciar la diferencia de resultados entre las columnas GT_80 y GT_80_80ms, el coeficiente de Pearson mejora de una correlación media a una correlación alta al utilizar balance de las clases.

Metodología: balance de clases

Debido al desbalance en el número de casos que componen la base de datos IDRiD y que está compuesta por 80 retinografías con presencia de lesiones (69 con riesgo aumentado de maculopatía y 11 sin maculopatía), es que fue necesario implementar métodos de balance que permitieran incrementar de manera sintética los ejemplos disponibles de las clases minoritarias, balanceando los ejemplos disponibles sin alterar su distribución. Estos casos corresponden a la clase etiquetada como sin riesgo de desarrollar maculopatía: es decir, retinografías sin lesiones en el área macular pero con lesiones en el área periférica y casos sin ningún tipo de lesión y que conformaron los casos correspondientes a retinas sanas.

Para lograr este objetivo se utilizaron tres técnicas de balance de datos:

i) aumentando la clase minoritaria (*Synthetic Minority Oversampling Technique SMOTE*), ii) disminuyendo en cierto porcentaje la clase mayoritaria (*Undersampling Technique*) y iii) incorporando casos de retinografías con máculas sin lesiones a partir de los otros subconjuntos de datos de IDRiD.

Este balance de clases se llevó a cabo con el propósito de representar los distintos escenarios clínicos posibles en las bases de datos a analizar. Tales escenarios se describen a continuación: i) retinografías con lesiones periféricas sin lesiones maculares (RD sin maculopatía), ii) retinografías con lesiones periféricas con lesiones maculares (RD y maculopatía), iii) retinografías sin lesiones periféricas y sin lesiones maculares (sano, sin maculopatía).

Análisis de resultados: influencia del balance de datos

Al evaluar el comportamiento obtenido por cada método de balance comparando los resultados obtenidos por el clasificador *Logistic Regression* procesando la entrada de datos MGLG, los mejores rendimientos son alcanzados por el método de balance 80_80, seguido del balance 80_9, 80 y 80_BC. Si comparamos los rendimientos alcanzados entre las entradas UMLG y MGLD, es decir la variación en el desempeño alcanzado al obtener los datos de la posición macular de manera automática (*U-net*) y la localización de las lesiones de manera automática (DART) observamos que hay un descenso mayor al procesar los datos obtenidos por DART. Al utilizar la entrada de datos UMLD, la diferencia entre los rendimientos alcanzados se incrementa, probablemente como el resultado de la sumatoria de variaciones entre la estimación de la posición macular y la localización de lesiones de manera automática. Por otro lado, al comparar los resultados obtenidos en las entradas MGLG y UMLD y la variación experimentada entre los métodos de balance 80_80 y 80_9 la disminución es notoriamente menor al contrastarlos con los métodos 80 y 80_BC. El menor rendimiento alcanzado se observa en la entrada UMLD utilizando el método de balance 80_BC (ver figuras 9 y 10).

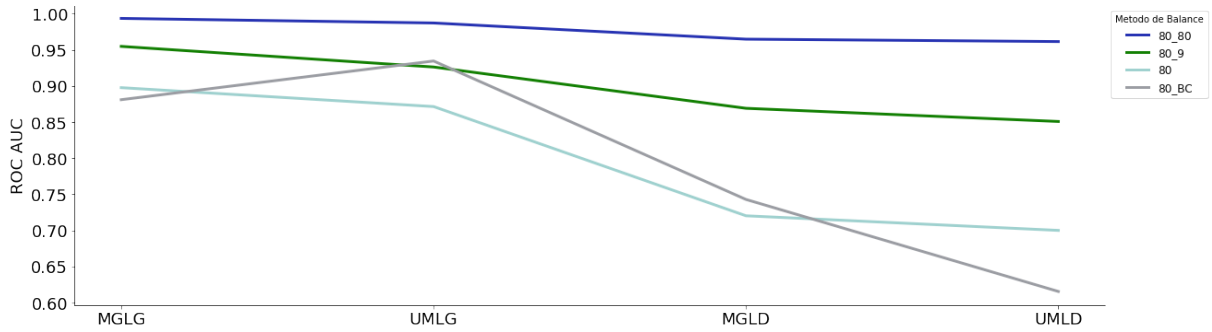


Figura 9. Rendimiento de cada método de balance de clases utilizando el clasificador *Logistic Regression* y el vector F mediante validación cruzada (ROC AUC). Se presentan en orden decreciente los resultados de cada clasificador, base de datos, balance y vector utilizado. Se representa en líneas de colores según el rendimiento alcanzado por cada entrada de datos: balance 80_80 morado, balance 80_9 en verde, balance 80 en celeste y balance 80_BC en gris.

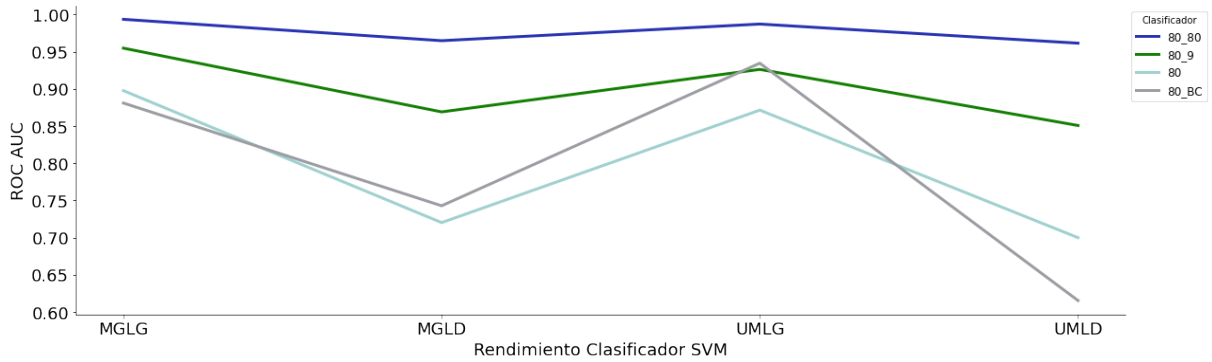


Figura 10. Rendimiento de cada método de balance de clases utilizando el clasificador *Support Vector Machine* y el vector F mediante validación cruzada (ROC AUC). Se presentan en orden decreciente los resultados de cada clasificador, base de datos, balance y vector utilizado. Se representa en líneas de colores según el rendimiento alcanzado por cada entrada de datos: balance 80_80 en morado, balance 80_9 en verde, balance 80 en celeste y balance 80_BC en gris.

Por otro lado, al procesar las entradas de datos MGLG y UMLD y comparar el desempeño alcanzado utilizando el método de balance 80_9, Logistic Regression es el que experimenta menor variabilidad. Al procesar las entradas UMLG, MGLD y UMLD tanto *Support Vector Machine* (0.850) como *Random Forest* (0.901) se comportan de manera similar. Nuevamente el clasificador *Decision Tree* (0.797) presenta el desempeño más bajo entre todos los clasificadores (Ver figura 11).

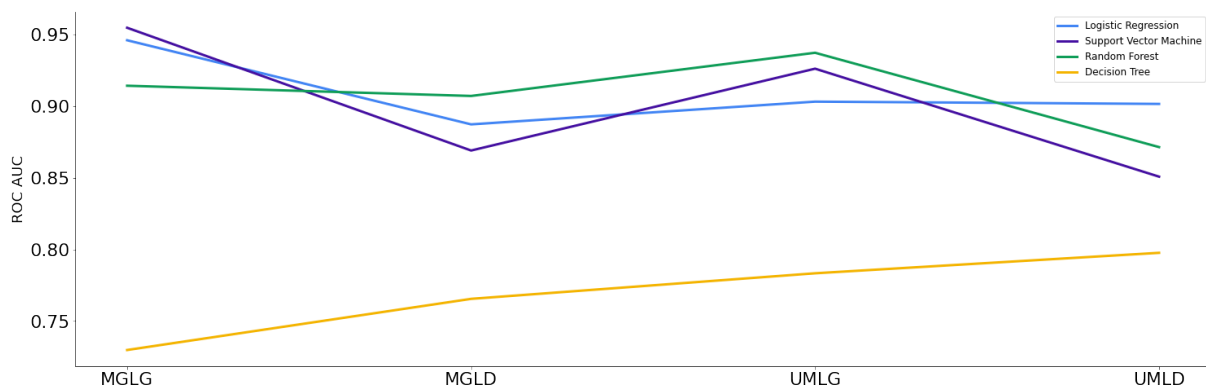


Figura 11. Rendimiento de cada clasificador utilizando el método de balance 80_9 mediante validación cruzada (ROC AUC). Se presentan en orden decreciente los resultados de cada base de datos y balance 80_9, utilizando el vector F. Se representa el rendimiento alcanzado por cada clasificador diferenciado en líneas de colores: *Logistic Regression* en celeste, *Support Vector Machine* en morado, *Random Forest* en verde y *Decision Tree* en amarillo.

Análisis de las métricas obtenidas en los distintos escenarios

En esta etapa se llevó a cabo la comparación del comportamiento de cada clasificador, entrada de datos, vector de características y método de balance utilizado. Su finalidad fue seleccionar los mejores clasificadores de MD y estimar su comportamiento, utilizando los datos provenientes de la localización automatizada de la posición macular y las lesiones *ground truth*, contrastando sus resultados con los obtenidos por la determinación de la posición macular *U-net* y las lesiones localizadas por el sistema DART.

Entrada de datos.

La base de datos MGLG fue la que alcanzó los mejores resultados en la clasificación de MD, debido principalmente a que la entrada de datos se construyó a partir de los valores de referencia etiquetados manualmente por expertos. Es posible observar un leve descenso en el rendimiento obtenido por la base de datos UMLG (posición macular *U-net* y lesiones *ground truth*), seguido de un mayor descenso al utilizar los datos provenientes de la localización de lesiones por el sistema DART (MGLD) producto de la diferencia en la precisión alcanzada en la estimación de la posición macular por parte de la *U-net* (objetivo 2) y/o la localización de lesiones al utilizar los datos provenientes del algoritmo DART (objetivo 3). Al combinar la información proveniente de la *U-net* y las lesiones estimadas por DART con la posición macular *ground truth* el descenso en los resultados alcanzados es mayor.

Método de balance.

Para solucionar el desbalance de clases fue necesario ajustar la cantidad de datos representativos de cada clase de interés. Existe evidencia en la literatura que sostiene que mejorar el desequilibrio de clases permite alcanzar mejores resultados [22], [51]. Adicionalmente bajo un escenario clínico real es posible lograr representar de mejor manera la distribución de clases esperada. El balance de clases fue realizado utilizando el conjunto de datos de entrenamiento y no el de validación, estrategia que permite una mejor precisión en la evaluación de la clasificación.

La implementación de 4 técnicas de balance permitió confirmar el impacto positivo de este parámetro en los desempeños obtenidos por los clasificadores seleccionados. A este respecto, cabe señalar que si bien los resultados alcanzados por los dos mejores métodos de balance se diferenciaron tan solo en un 5%, ambos presentaban importantes diferencias en la complejidad del método de balance implementado. Dependiendo del objetivo que se persigue, los resultados obtenidos por ambos métodos de balance, aportan información de gran valor: la base de datos 80_80 entrega un balance simple y abstracto, ideal para la evaluación de los otros componentes, como por ejemplo el rol de los vectores de características en la validación de la integración de la posición macular y la base de datos 80_9 entrega información que permite inferir

el rendimiento alcanzado por los clasificadores en un escenario clínico más complejo y realista.

La utilización de diversos métodos de balance que constituye uno de los aportes de esta tesis, permite corroborar el incremento en el rendimiento general de los clasificadores utilizados, entregando al mismo tiempo criterios de utilidad con los cuales seleccionar el método más adecuado para llevar a cabo el balance.

Vectores de características.

Uno de los principales aportes de esta tesis, consistió en la construcción de los vectores de características E y S, ya que permitieron establecer si sus componentes utilizaban o no la información correspondiente a la integración de la posición macular. Adicionalmente, la composición de cada vector permitió identificar cuales de estos se comportaron como factores predominantes en la tarea de clasificación de MD. Los resultados obtenidos permitieron validar la localización y utilización de la posición macular en etapas posteriores del presente trabajo (objetivo específico 2) y al mismo tiempo determinar un rendimiento referencial (MGLG) con el cual comparar el desempeño alcanzado por cada clasificador mediante el procesamiento de cada entrada de datos, método de balance o vectores de características.

A diferencia de los criterios utilizados en la literatura, en el presente trabajo de tesis, la elaboración de cada componente de características combinó los atributos, lesiones rojas (microaneurismas y hemorragias) y lesiones brillantes (exudados duros y blandos) y se analizó el peso de cada atributo mediante el coeficiente de correlación de Pearson. Estos hallazgos abren la puerta a la reevaluación de los atributos utilizados tradicionalmente para llevar a cabo la tarea de detección automatizada de MD, permitiendo la exploración de nuevos vectores de características con un mayor rango de componentes. El rol de los vectores implementados aportan un mayor grado de interpretabilidad del modelo, haciendo posible identificar atributos contrastables, verificables y comprensibles tanto para los desarrolladores como para los usuarios finales de estas herramientas.

Clasificadores.

Las diferencias en rendimiento individual para cada clasificador pueden ser explicadas en base al correcto procesamiento de datos de entrada imprecisos, al manejo eficiente del sobreajuste, a una muestra de datos muy pequeña o con desbalance de clases. En el caso de *Logistic Regression* la variación en el desempeño entre MGLG y UMLD es menos pronunciada, probablemente debido a que este clasificador permite el filtrado del ruido entre los datos y evita generar modelos demasiado complejos, tendiendo a generalizar. Por su parte *Random Forest* y *Support Vector Machine*, se comportan de manera similar y con un desempeño levemente menor al alcanzado por el clasificador *Logistic Regression*. Si bien en la literatura uno de los algoritmos más usados en clasificación de RD y con mejor desempeño es *Support Vector Machine* [22], en el presente trabajo este clasificador no fue el que alcanzó el mejor desempeño, resultados que pueden explicarse en parte debido a la baja cantidad de datos utilizado en la etapa de validación. Por otro lado, los bajos rendimientos obtenidos por el algoritmo *Decision Tree* pueden ser explicados sobre la base de las características propias del clasificador y a su tendencia al sobreajuste.

Consideraciones en la elaboración de bases de datos utilizadas en oftalmología

La calidad de la información con la que cuenta la base de datos seleccionada resulta ser un factor fundamental para el desempeño alcanzado, indistintamente del clasificador utilizado. Esto es particularmente relevante, dado que muchas de las bases de datos disponibles actualmente presentan problemas en la distribución de clases o en la cantidad de información disponible con la cual realizar el entrenamiento de los clasificadores de MD. Las estrategias implementadas en el presente trabajo tuvieron el objetivo de disminuir el efecto generado por el desbalance de clases. Asimismo la información disponible para etiquetar a cada base de datos es insuficiente, el trabajo de Krause *et al* [52] concluye que una “adecuada determinación de la información *ground truth* es un aspecto crítico en el desarrollo de algoritmos de inteligencia artificial clínicamente aplicables en la detección o tamizaje de patologías retinales”. Por consiguiente, para el desarrollo de algoritmos de clasificación aplicables en el ámbito clínico es fundamental contar con bases de datos que cumplan ciertos estándares

mínimos necesarios con las cuales desarrollar, entrenar y validar a los diversos clasificadores de RD. Autores como Banko *et al* [53] sugieren el desarrollo de corpus más robustos necesarios para llevar a cabo el entrenamiento de los clasificadores. Los argumentos anteriormente expuestos permiten afirmar que el contar con bases de datos de mejor calidad resulta ser un aspecto fundamental para el desarrollo de algoritmos cuya finalidad es convertirse en herramientas de apoyo clínico oftalmológicos.

Por los argumentos anteriormente expuestos es que se infiere que una base de datos oftalmológica debiese contar con ciertas características mínimas necesarias para su adecuada construcción, entre las que podemos mencionar: i) contar con el etiquetado del diagnóstico tanto RD como MD, ii) incluir los criterios de clasificación utilizado para determinar el diagnóstico, iii) localizar y segmentar los hitos anatómicos y lesiones presentes iv) utilizar indicadores de calidad retinográfica, detallando si es que hay presencia de artefactos, v) incluir imágenes provenientes de diferentes equipos retinográficos y vi) contar con un gran número de imágenes representativas de la población en la cual el clasificador será aplicado. Finalmente para llevar a cabo de manera más estandarizada la correcta estimación y comparación de los resultados alcanzados, el establecimiento de indicadores de desempeño claves (KPI), es un aspecto fundamental debido a la alta variabilidad de métricas utilizadas, cantidad de casos y balances de clases presentes en las muestras de validación observadas entre los diversos estudios publicados.