

# Deep Learning y enfermedades oculares

Sergio Jacobo Zavaleta <sup>1</sup>, Walter Zavaleta-Gavidia <sup>2</sup>, Víctor Zavaleta-Gavidia <sup>3</sup>

## Resumen

Debido al avance tecnológico digital y la disponibilidad de grandes volúmenes de imágenes clínicas, la inteligencia artificial y específicamente el deep learning como subcampo de estudio, han empezado a revolucionar el diagnóstico asistido por computadora. Sobre todo, en especialidades médicas como la oftalmología que depende del análisis de imágenes como herramienta diagnóstica. Después de una breve descripción relacionada a los criterios de entrenamiento y validación de modelos en deep learning, se describe diversos sistemas destinados a tareas de tamizaje, diagnóstico y pronóstico de las enfermedades oculares más relevantes causantes de ceguera irreversible, como glaucoma, degeneración macular relacionada con la edad y retinopatía diabética. De esta revisión puede inferirse que cada vez más sistemas con deep learning entrenados para tareas específicas están obteniendo mejores resultados diagnósticos por imagen, muchas veces igualando o superando al obtenido por médicos expertos y entrenados.

## Palabras Clave

Deep Learning — Glaucoma — Retinopatía Diabética

<sup>1</sup> Ingeniero Mecatrónico de la Universidad Nacional de Trujillo, Trujillo, Perú

<sup>2</sup> Biólogo, Maestrando en Biotecnología y Bioingeniería, Universidad Nacional de Trujillo, Perú

<sup>3</sup> Médico Cirujano, Maestro en Salud Pública, Docente de la Facultad de Medicina de la Universidad Nacional de Cajamarca, Perú

\*Correspondencia: vzavaletag@unc.edu.pe

## Abstract

Due to digital technological advances and the availability of large image dataset, artificial intelligence, specifically deep learning as a subfield of study, has begun to revolutionize computer-assisted diagnosis. Above all, in medical specialties such as ophthalmology that relies on image-based analysis as a diagnostic tool. After a brief description related to the training criteria and validation of deep learning models, several deep learning systems are described for screening, diagnosis and, disease prognosis tasks in three of the most relevant eye diseases that cause irreversible blindness, such as glaucoma, age-related macular degeneration and diabetic retinopathy. From this review, it can be inferred that more and more deep learning systems are trained for specific tasks to obtain better performances, often equaling or surpassing those obtained by experts and trained graders.

## Keywords

Deep Learning — Glaucoma — Diabetic retinopathy

## 1. Introducción

En 2021, la Organización Mundial de la Salud estimó que aproximadamente 1 300 millones de personas viven con alguna forma de deficiencia visual a nivel global. Destacando como principales causas, los errores de refracción no corregidos y cataratas, sobre todo en el adulto mayor [1]. A nivel nacional, las causas de discapacidad visual y ceguera no son diferente. A inicios de la década pasada, los problemas visuales constituyeron la segunda causa de discapacidad causado principalmente por catarata [2]. Para 2020 en salud ocular, la Encuesta Demográfica de Salud Familiar reportó que solo el 16.4% de los adultos mayores a nivel nacional recibieron diagnóstico por cataratas, siendo cerca del 30% mayores de 70 años. Con mayor prevalencia en la Costa, seguido por la Sierra y la Selva [3]. Sumado a ello, la creciente población, la limitada atención de la salud y el reducido número de especialistas oculares han derivado en un sistema de atención médica sobrecargado, dificultando diagnósticos rápidos y oportunos.

Por otro lado, el avance tecnológico a nivel mundial en los últimos años ha sido el resultado de las innovaciones digitales en información y comunicación que han llevando a revolucionar la industria moderna, incluido el sector salud, gracias al Internet de las Cosas (*Internet of Things*, IoT), desarrollo de redes inalámbricas de quinta generación (5G), seguridad digital en *Blockchain* y el fortalecimiento de la inteligencia artificial (IA), en sus formas de aprendizaje máquina (*machine learning*, ML) y aprendizaje profundo (subcampo del ML conocido como *deep learning*, DL) [4, 5]. Esta última definida incluso como la cuarta revolución industrial en la década pasada [6], desde su primera formulación en 1956 y formulada por McCarthy como aquella tecnología capaz de imitar el comportamiento humano [7]. Desde entonces, se ha conseguido altos niveles de exactitud y precisión en múltiples áreas del procesamiento de lenguaje natural, reconocimiento de voz y visión computacional [8].

Del mismo modo, el DL ha impulsado el análisis y

diagnóstico por imágenes de diversas enfermedades como la tuberculosis a partir de imágenes de Rayos X [9, 10], melanomas malignos en fotografías de piel [11] e identificación de metástasis para cáncer de mama a partir de muestras de tejido [12]. Por su parte, las imágenes oculares basadas en fotografía de fondo de ojo y campos visuales han permitido la implementación de soluciones diversas en tamizajes (*screening*) automáticos y diagnóstico de enfermedades oculares relacionadas con Retinopatía diabética (RD) [13, 14, 15, 16], glaucoma [17], degeneración macular relacionado con la edad (DMRE) [18, 19]. Que a nivel hospitalario, estas soluciones supondrían mejorar la detección temprana de las enfermedades oculares frente a un diagnóstico convencional, sujeto a diversos factores como el aumento significativo del volumen de imágenes a procesar y procedimientos extensos. Y sobre todo, con resultados económicos prometedores como herramienta diagnóstica automática o semiautomática [20].

Por consecuencia, esta revisión describe los recientes avances en IA para el tamizaje, diagnóstico y pronóstico de las principales enfermedades oculares con daños irreversibles basados en el procesamiento de imágenes utilizando DL.

## 2. Como funciona un sistema DL

Los sistemas de DL se basan en el concepto de una red neuronal artificial (RNA), compuesta por múltiples entradas (capa) y salidas. Todas las capas intermedias si existen se conocen como capas ocultas y se encargan de extraer diferentes características para los datos de entrada, permitiendo a la red aprender y completar tareas complejas. Entre las diversas arquitecturas (secuencia de capas) de DL, destacan las redes neuronales convolucionales (CNN) por su mayor rendimiento en el análisis de datos a partir de imágenes, pero sobre todo por la inclusión de capas especiales para operaciones matemáticas de filtrado conocido como convolución [22]. En general, un sistema de DL puede entenderse como el procedimiento de entrenamiento de modelos algorítmicos con datos etiquetados (de ser posible), por ejemplo, fotografías de fondo de ojo, clasificadas manualmente por un especialista en base a la presencia de alguna anomalía de interés y que también pueden utilizarse para asignar etiquetas automáticamente a nuevos datos. No obstante, estos algoritmos se diferencian de los métodos tradicionales de ML, por que las características específicas o indicios de una enfermedad (región de interés) pueden aprenderse directamente de las propias imágenes [23].

La Figura 1 describe la implementación de un sistema de DL para detección de enfermedades oculares como la RD, edema macular diabético (EMD) y glaucoma. Además, es necesario mencionar que los términos modelo y sistema fueron usados de manera indistinta a lo largo del trabajo.

## 3. Medidas de desempeño de los Algoritmos de DL

Muchos problemas de DL resuelven problemas de clasificación apoyados en técnicas estadísticas a la hora de evaluar el rendimiento del mismo, destacando el uso de la matriz

de confusión, para el cálculo de tres parámetros principales denominados: exactitud, sensibilidad y especificidad. Específicamente, en el caso de predicción de clases (por ejemplo, sano y no sano) dependerá para los datos encontrar el límite óptimo entre clases. En problemas de ML, usualmente se utiliza la exactitud, mientras que en DL, se tiene en cuenta el porcentaje de clasificación y la matriz de confusión, como el de una clasificación binaria en la Tabla 1.

**Tabla 1.** Matriz de confusión utilizada en la evaluación del modelo de DL

Predicción del modelo	Verdadero diagnóstico	
	Sano	Enfermo
Positivo	VP <sup>a</sup>	FN <sup>b</sup>
Negativo	FP <sup>c</sup>	VN <sup>d</sup>

<sup>a</sup>: Verdadero positivo, <sup>b</sup>: Falso negativo, <sup>c</sup>: Falso positivo, <sup>d</sup>: Verdadero negativo.

Estas medidas y otros estadístico como el valor predictivo positivo (VPP) y valor predictivo negativo (VPN) se calculan de la forma,

$$\text{Exactitud} = \frac{VP + VN}{VP + FP + VN + FN} \quad (1)$$

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (2)$$

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (3)$$

$$\text{VPP} = \frac{VP}{VP + FP} \quad (4)$$

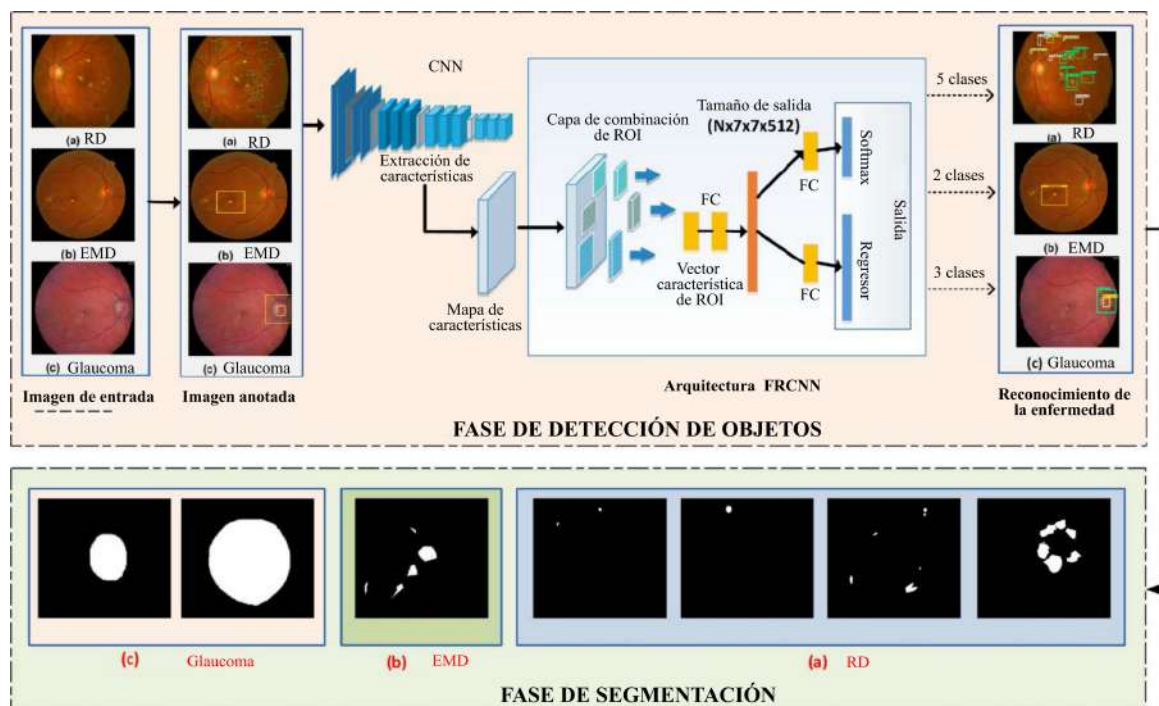
$$\text{VPN} = \frac{VN}{VN + FN} \quad (5)$$

Entre medidas más técnicas destacan la curva ROC (gráfico entre la tasas de VP Y FP), el área bajo la curva (AUC), que significa la probabilidad de que un modelo clasifique más positivos que negativo, y el coeficiente DICE, para la similitud entre dos muestras.

## 4. Datasets para diagnóstico asistido por computadora

En la actualidad, existen diversas bases de datos disponibles en la web conocidos como datasets. Entre los datos de imágenes para análisis de enfermedades oculares previamente diagnósticos por expertos médicos destacan,

- AREDS/AREDS2.[24]. A partir del Estudio de Enfermedades Oculares Relacionadas con la Edad en DMRE y catarata.
- LabelMe[25]. Imágenes de fondo de ojo en color para diagnóstico de RD, DMRE, glaucoma y otras.
- Messidor[26]. Imágenes de fondo de ojo en color para diagnóstico de RD.



**Figura 1.** Arquitectura del método para detección de retinopatía diabética (RD), edema macular diabético (EMD) y glaucoma. En la fase de detección de objetos, la agrupación de las regiones de interés (ROI) de la enfermedad en cuestión, se detecta utilizando una red neuronal convolucional rápida (FRCNN). En la fase de segmentación, las regiones detectadas de las tres enfermedades fueron segmentadas mediante un modelo de agrupación k-means con lógica difusa (fuzzy k-means clustering). Donde; CNN, red neuronal convolucional; FC, capas completamente conectas; Tomado de [21].

- DUKE Dataset. Imágenes de tomografía de coherencia óptica (TCO) para DMRE relacionada a la edad.
- EyePACS. Imágenes de fondo de ojo de alta resolución para la detección de la RD.

## 5. Glaucoma

El glaucoma es un grupo heterogéneo de trastornos neurales degenerativos caracterizados por la pérdida progresiva de las células ganglionares de la retina y sus axones [27]. A nivel mundial, es la principal causa de ceguera irreversible con una prevalencia de 3,5 % entre personas de 40 a 80 años y que para 2040, se estima 111,8 millones de personas con la enfermedad [28]. Actualmente, los tratamientos disponibles no pueden revertir los daños y que en la mayoría de los casos, es una afección crónica que requiere un tratamiento de por vida. Sin embargo, un diagnóstico y tratamiento temprano pueden prevenir su progresión.

A nivel de tamizaje, los modelos de DL detectan y cuantifican el daño glaucomatoso utilizando fotografías de fondo de ojo en color [29]. Así, Pascal et al [30] utilizando un modelo de multitarea con optimizadores independientes, obtuvieron un rendimiento del AUC de hasta  $92,61 \pm 0,38$  y equiparable al de expertos capacitados.

A nivel de diagnóstico, Hemelings et al [31] entrenaron y evaluaron modelos utilizando imágenes de fondo de ojo sometidos a un recorte radial proporcional al tamaño total de la imagen (de 10 a 60 %), centrado en la cabeza del nervio óptico (ONH). De forma análoga, se aplicó el recorte inverso correspondiente a la periferia de la imagen. Como

resultado, se destaca la detección de glaucoma a partir de regiones de imágenes de fondo de ojo fuera de la ONH con un rendimiento significativo del AUC de hasta 0,88 (IC 95 %: 0,85–0,90) y un valor de 0,37 (IC 95 %: 0,35–0,40) para el coeficiente de determinación ( $R^2$ ) en la relación copa-disco (configuración más extrema de corte al 60 % de ONH). A contrario, del obtenido de las imágenes originales con un AUC de 0,94 (IC 95 %: 0,92–0,96) y  $R^2$  al 0,77 (IC 95 %: 0,77–0,79).

De forma análoga, se desarrolló un sistema de clasificación detallado (FGGDL) para evaluar y comparar la pérdida de campo visual frente al criterio oftalmológico. Para ello, se utilizaron campos visuales tomados con equipos especializados (*Octopus perimeters* and *Humphrey Field Analyzer*) y recolectados de hospitales y una base de datos pública. El FGGDL logró una alta precisión entre 85 al 90 %, con AUC de 0,93 y 0,90 para los datos de Humphrey Field Analyzer y Octopus perimeters, respectivamente. Los resultados fueron significativamente superiores ( $p < 0,01$ ) al diagnóstico de estudiantes de medicina y cercano ( $p = 0,614$ ) al de médicos oftalmólogos [32].

Por su parte, Sulot et al [33] utilizaron imágenes de ONH, capturadas con oftalmoscopia láser de barrido (SLO) durante la obtención de imágenes TCO estándar del segmento posterior, para la clasificación de la enfermedad a partir del espesor de la capa de fibras nerviosas de la retina (RNFL) mediante el uso de un modelo con arquitectura CNN. Utilizando los datos de 227 personas (105 pacientes y 122 de control), una validación cruzada del modelo alcanzó un rendimiento de hasta 96,3 % de exactitud, superando a los especialistas y valorando la información clínica almace-

nada en imágenes SLO.

Por otro lado, el diagnóstico de la progresión de glaucoma sigue siendo un desafío considerable en la práctica clínica. Entre las pruebas utilizadas para evaluar el cambio temporal, destacan la tomografía de coherencia óptica de dominio espectral (SDOCT) y perimetría automatizada estándar (SAP) que sufren una considerable variabilidad test-retest, lo que dificulta discriminar el cambio real de la variabilidad. Además, no existe consenso sobre criterios específicos para diagnosticar el campo visual o posible progresión estructural. Tal falta de consenso ha obstaculizado el progreso en el campo oftalmológico dificultando la comparación de los resultados en diferentes enfoques [34].

A pesar de ello, Mariottoni et al desarrollaron un algoritmo capaz de relacionar el daño en las capas RNFL vistas en SDOCT, con la pérdida funcional en la perimetría SAP. Utilizando dos grupos iguales aleatorios de 26 499 imágenes para SAP y SDOCT bajo el entrenamiento y validación de un modelo CNN, se consiguió predecir los 52 umbrales de sensibilidad (24-2 SAP) a partir del grosor de la RNFL, con un coeficiente de correlación de 0,60 ( $P < 0,001$ ) con los valores medidos de SAP y un error absoluto promedio de 4,25 dB. Finalmente, se mejoró la comprensión de cómo las pérdidas de SDOCT se traducen en daños detectables de SAP para una intervención rápida ante signos tempranos de neuropatía óptica glaucomatosa [35].

## 6. Degeneración macular relacionada con la edad

La degeneración macular relacionada a la edad (DMRE) es una enfermedad multifactorial y otra de las causas más frecuente de pérdida severa de visión en ambos ojos en personas mayores de 60 años. El aumento de la esperanza de vida en la población sugiere un aumento exponencial de esta patología, ya que el riesgo por DMRE aumenta con la edad. Entre sus causas además del envejecimiento, responde a factores de riesgo genético y ambientales [36]. Aproximadamente, el 8,7 % de la población mundial tiene DMRE con una proyección de 288 millones pacientes para 2040 [37].

De acuerdo con la Academia Americana de Oftalmología, la enfermedad puede clasificarse en cinco etapas, comenzado por la ausencia de ella, temprana, intermedia y tardía [38]. De allí, la recomendación de una revisión periódica cada dos años para pacientes en etapa intermedia. La DMRE temprana incluye signos clínicos como drusas y anomalías del epitelio pigmentario de la retina, mientras que la DMRE avanzada presenta atrofia neovascular (también llamada DMRE húmeda o exudativa) y atrofia geográfica central (DMRE seca o no exudativa) [22].

A nivel de tamizaje, Gonzalez-Gonzalo et al validaron el sistema comercial RetCAD basado en DL para la detección automática conjunta de RD y DMRE a partir del entrenamiento con un dataset de 600 imágenes de fondo de color para enfermedades oculares (referibles o no). Para la etapa de entrenamiento, cada imagen fue calificada por un oftalmólogo experimentado para establecer el estándar de referencia (RS) y por cuatro observadores independientes. Posteriormente, la validación conjunta se consiguió utili-

zando los datasets Messidor (1 200 imágenes para DR) y AREDS (133 821 imágenes para DMRE). Para la detección del DMRE, el sistema logró un rendimiento de hasta 94,9 % (sensibilidad de 91,8 % y especificidad de 87,5 %) y un AUC de 92,7 %, mientras el desempeño humano promedio alcanzó una sensibilidad de 76,5 % y especificidad al 96,1 %. Resultados comparables al de expertos oftalmólogos, demostrando así una posible superioridad del DL como herramienta de tamizaje rápido y fiable [39].

En diagnóstico, Peng2018 et al consiguieron una clasificación automática de la enfermedad según la escala de gravedad simplificada (0-5) de AREDS utilizando el modelo DeepSeeNet e imágenes de fondo de ojo a color (58 402 imágenes de entrenamiento y 900 de prueba). Para ello, el modelo fue capaz de detectar factores de riesgo individuales (tamaño de las drusas y anomalías pigmentarias) para cada ojo y asignarles una puntuación. El mejor desempeño se consiguió en una clasificación multiclase (exactitud = 67.1 %; kappa = 0,558) frente a especialistas de retina (exactitud = 59.9 %; kappa = 0,467), con valores AUC en la detección de drusas grandes (0.94), anomalías pigmentarias (0.93) y DMRE tardía (0.97). Con un rendimiento inferior en la detección de DMRE tardía (exactitud 96.7 %) frente a especialistas (97.3 %). Concluyendo una vez más en el potencial de modelos de DL como ayuda para mejorar los procesos de toma de decisiones clínicas en la detección temprana de DMRE y predicción del riesgo tardío [23].

Por su parte, Ting et al. entrenaron un sistema utilizando 108 558 imágenes de retina (fóvea y disco óptico) provenientes de 38 189 pacientes diabéticos del programa *Singapore National Diabetic Retinopathy Screening Program* (SIDRP) durante los años 2010 y 2013 [16]. El sistema consiguió un sensibilidad de 96.4 % y especificidad de 93.2 % luego de probarse con 35 948 imágenes, a pesar de la ausencia de pacientes e imágenes específicas (etapa intermedia o peor de la enfermedad) relacionadas con DMRE. En otros estudios, Burlina et al, reportaron un sistema con una exactitud diagnóstica entre 88.4 % y 91.6 %, que a diferencia de la anterior, existió una presegmentación de la región macular a priori para entrenar y probar el modelo (120 656 imágenes de AREDS) [18]. Del mismo modo, Grassmann et al reportaron una sensibilidad de 84.2 % y especificidad de 94.3 % en la detección de cualquier DMRE [19].

En 2019, Keel et al desarrollaron un sistema para la detección de atrofia neurovascular temprana a partir del entrenamiento con 56 113 imágenes de retina y validado externamente mediante 86 162 imágenes más, alcanzando una sensibilidad y especificidad de 96.7 % y 96.4 % respectivamente [40].

En pronóstico, Peng et al. demostraron el uso del DL para el análisis de supervivencia en la predicción de probabilidades en la progresión de DMRE tardía utilizando más de 80 000 imágenes alojadas en AREDS y AREDS2. Mediante la validación con un dataset de 601 participantes, se alcanzó una alta precisión pronóstica (estadística C de 5 años; IC 95 %: 86.2-86.6) al superar notablemente a los especialistas en retina bajo dos estándares clínicos (81,3; IC 95 %: 81,1-81,5) y (82,0; IC 95 %: 81,8-82,3), respectivamente. Sobre todo, al pronosticar algún riesgo superior al

50 % [41].

Recientemente, Ganjdanesh et al. propusieron dos modelos secuenciales de predicción. El primero, para una clasificación binaria de imágenes de fondo de ojo (avanzado o no de la enfermedad) denominado LONGL-Net, y el segundo, denominado *temporal-correlation-structure guided Generative Adversarial Network* (GAN) para aprendizaje con los datos anteriores y pronosticar la condición del paciente a largo plazo (reconocimiento del tamaño de drusas y severidad en las anomalías pigmentarias) a partir de imágenes de AREDS y del UK Biobank (en etapa de validación). En este último, obteniendo una exactitud promedio de 90.5 % y sensibilidad de 0,797 al calificar 300 imágenes [42].

## 7. Retinopatía diabética

La diabetes, uno de los grandes problemas de salud, alcanzó una prevalencia mundial de 9.3 % en 2019 y se estima que seguirá aumentando hasta un 10.9 % para el 2045 [43]. Muchos de los programas de tamizaje en retinopatía diabética (RD) se basan en las ventajas que ofrece la telemedicina con fotografía de fondo de ojo al minimizar las barreras geográficas y aumentar la cobertura remota de tamizajes, como los implementados a gran escala en los países de Reino Unido, Singapur, China y Los Estados Unidos [44]. Tales son los esfuerzos por uniformizar y promover su adopción a nivel internacional que se evidencia en la reciente guía clínica en telemedicina para RD [45].

Vale la pena mencionar que los grandes programas oftalmológicos se consideran una gran oportunidad para la recolección y creación de grandes bases de datos de imágenes cuidadosamente etiquetadas por especialistas, que más tarde pueden convertirse en extensos datasets para entrenamiento de modelos de DL. Tal es caso del Programa Nacional de Tamizaje para Telemedicina en Singapur o SIDRP, que fue capaz de realizar fotografías de retina de hasta 200 000 personas con diabetes en 18 centros hospitalarios para su eventual evaluación por médicos entrenados en 2016 [46]. Desde entonces, diversos proyectos han intentando integrarlo con DL y convertirse en el primer programa nacional de lectura autónoma para tamizaje de RD [16, 47]. A nivel comercial, destacan las compañías norteamericanas *Intelligent Retinal Imaging Systems*, *Eyenuk* y *IDx*, donde esta última inventado por *IDx-DR* en 2018, es el primer sistema de IA para detección temprana y automática de RD aprobada por la Administración Norteamericana de Alimentos y Medicamentos [48]. A partir de los resultados alcanzados en un estudio observacional prospectivo con una sensibilidad de 87.2 % y especificidad de 90.7 % [13].

A nivel de tamizaje, Gargeya y Leng reportaron el desarrollo de un modelo robusto para tamizaje automático a partir del entrenamiento de imágenes de fondo públicamente disponibles (75 137 imágenes previamente clasificadas por un panel de expertos) y una clasificación binaria (sanos y con RD). Para la validación externa, se utilizaron los dataset MESSIDOR 2 y E-Ophtha. Utilizando un dataset local, el modelo alcanzó un valor AUC de 0.97 con una sensibilidad y especificidad de 94 % y 98 % respectivamente, mientras que para los datasets externos, se consiguieron valores AUC inferiores de 0.94 y 0.95 respectivamente [49].

Li et al desarrollaron un sistema para la detección automática de los signos más comunes de la enfermedad como hemorragias de retina utilizando un base de datos de 16 827 fotografías de fondo (11 339 pacientes) como parte de la alianza médica Chinese Medical Alliance for Artificial Intelligence (CMAAI) [50]. Con un sensibilidad y especificidad mayores al 96 %, el sistema demostró un gran potencial para la detección a partir del uso de imágenes de fondo de ojo ultra ancha *ultra-widefield fundus image* (UWF), al ser imágenes cinco veces más grandes que las de fondo de ojo tradicional.

Nazir et al presentaron un modelo de localización y segmentación automatizado basado en un tipo de red CNN rápida basado en regiones de interés (FRCNN) con agrupamiento k-means con lógica difusa (FKM) para RD, edema macular diabético y glaucoma de acuerdo a la Figura 1. Para ello, se entrenaron y segmentaron imágenes (datasets Diaretdb1, MESSIDOR, ORIGA, DR-HAGIS y HRF) previamente etiquetadas (región cuadrangular) mediante una asignación in situ del problema. En la etapa de localización, el modelo logró una precisión media promedio de 0.945, 0.943 y 0.941 para RD, EMD y glaucoma, respectivamente. Mientras en segmentación, se logró una precisión de 0,952, 0,958 y 0,9526 para las regiones RD, EMD y glaucoma, respectivamente [21].

Por su parte, Dai et al desarrollaron el sistema DeepDR para detección de las etapas tempranas y tardías de RD. Este sistema, entrenado para la medición de imágenes en tiempo real, detección de lesiones y clasificación a partir de 466 247 imágenes de fondo de ojo provenientes de 121 342 pacientes diabético, fue validado con ayuda de un dataset local con 200 136 imágenes de fondo y otras 209 322 más provenientes de datasets externos. Como resultado, el valor AUC para casos particulares de detección de microaneurismas, manchas algodonosas, exudados duros y hemorragias fueron de 0,901, 0,941, 0,954 y 0,967, respectivamente. Y que a la hora de clasificar la RD en leve, moderada, severa y proliferativa se alcanzaron valores de AUC iguales a 0,943, 0,955, 0,960 y 0,972, respectivamente. En validaciones externas, se consiguió valores eficientes de AUC de 0,916 a 0,970 al momento de la clasificación [51].

Por otro lado, Guo et al examinaron la eficacia de un modelo para cuantificar el área sin perfusión (NPA) en angiografía de TCO (en el complejo vascular superficial). El modelo se entrenó en 978 escaneos volumétricos (un total 137 pacientes con amplio rango de RD y 26 sanos). Entre sus resultados se calcularon la media del score F1 (métrica de ML basado en la precisión y sensibilidad del modelo) para la detección de severidades de RD en exploraciones de escaneo nasal, macular y temporal de 0.78, 0.82 y 0.78, respectivamente. Por otro lado, el área avascular extrafoveal en la exploración macular mostró la mejor sensibilidad con un 54 % para diferenciar a las personas con diabetes del grupo sano, mientras que la exploración en angiografía de TCO mostró una sensibilidad significativamente mayor que las exploraciones maculares para detectar ojos con RD al 66 %, RD referible (grado intermedio) al 63 % y RD grave al 62 % [52].

Finalmente, Ruamviboonsuk et al propusieron evaluar

el desempeño real y viabilidad un sistema como parte de un estudio prospectivo de cohortes en el sistema de atención médica de Tailandia a partir de fotografías de fondo de ojo tomadas de 7 651 pacientes. El sistema alcanzó una exactitud de 94,7 % (IC 95 %: 93,0–96,2), sensibilidad de 91,4 % (IC 95 %: 87,1–95,0) y especificidad de 95,4 % (IC 95 %: 94,1–96,7). Por su parte, los especialistas en retina alcanzaron una precisión de 93,5 % (IC 95 %: 91,7–95,0;  $p=0,17$ ), sensibilidad de 84,8 % (IC 95 %: 79,4–90,0;  $p=0,024$ ) y especificidad al 95,5 % (IC 95 %: 94,1–96,7;  $p=0,98$ ). Para valores de VPP, el modelo obtuvo 0,79 (IC 95 %: 0,74–0,84) en comparación con 0,76 (IC 95 %: 0,70–0,81) de los especialistas. Mientras, el VPN para el sistema fue de 0,95 (IC 95 %: 0,93–0,97) frente al 0,92 (IC 95 %: 0,89–0,95) de los especialistas. El sistema fue capaz de ofrecer un tamizaje en tiempo real similar a la de los especialistas en retina [53].

## 8. Conclusiones

Los esfuerzos por desarrollar sistemas automáticos o semiautomáticos en la detección, tamizaje y diagnóstico de enfermedades oculares causantes de ceguera irreversible, han encontrado gran potencial en las aplicaciones basadas en modelos de deep learning. Sobre todo, debido a los avances en telemedicina y el acceso a grandes bases de datos que han acelerado el diagnóstico por imagen, principalmente a partir de fotografías de fondo de ojo y tomografía de coherencia óptica. Con resultados de gran precisión llegando a ser comparables o superar al rendimiento de médicos expertos y entrenados en lo mismo. Con grandes posibilidades de diagnóstico en tiempo real al permitir mayor cantidad de análisis que métodos convencionales, a un menor costo económico y tiempo.

En consecuencia, los modelos de inteligencia artificial basados en deep learning representan nuevas y potentes herramientas para perfeccionar la medicina porque son potencialmente fiables no solo para diagnóstico, sino también, en todos los niveles de atención en salud para prevención, tamizaje y pronóstico de enfermedades oculares. Que llevado a un contexto nacional, la posible integración de imágenes clínicas en base a datos públicos para entrenamiento de modelos en inteligencia artificial, formación de equipos multidisciplinarios (especialmente expertos médicos y científicos de datos), definición de flujos de trabajo e infraestructura, podrían acelerar la creación de un sistema integral en la prevención de enfermedades oculares. Aún así, todavía hay un largo camino por recorrer hasta que su modelo computacional sea de uso corriente en la práctica clínica cotidiana.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Referencias

- [1] Organización Mundial de la Salud. Ceguera y discapacidad visual; 2021. <https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
- [2] Ministerio de Salud. Problemas visuales son la segunda causa de discapacidad en el país; 2012. <https://www.gob.pe/institucion/minsa/noticias/34794-problemas-visuales-son-la-segunda-causa-de-discapacidad-en-el-pais>.
- [3] Instituto Nacional de Estadística e Informática. Enfermedades No Transmisibles y Transmisibles, Peru 2020. Lima: Instituto Nacional de Estadística e Informática; 2020.
- [4] Ting DSW, Carin L, Dzau V, Wong TY. Digital Technology and COVID-19. *Nature Medicine*. 2020 Apr;26(4):459-61.
- [5] Ting DSW, Lin H, Ruamviboonsuk P, Wong TY, Sim DA. Artificial Intelligence, the Internet of Things, and Virtual Clinics: Ophthalmology at the Digital Translation Forefront. *The Lancet Digital Health*. 2020 Jan;2(1):e8-9.
- [6] Forum WE. The Fourth Industrial Revolution: What It Means, How to Respond; 2016.
- [7] McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*. 1995;(27):1-13.
- [8] LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature*. 2015 May;521(7553):436-44.
- [9] Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*. 2017 Aug;284(2):574-82.
- [10] Ting DSW, Yi PH, Hui F. Clinical Applicability of Deep Learning System in Detecting Tuberculosis with Chest Radiography. *Radiology*. 2018 Feb;286(2):729-31.
- [11] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*. 2017 Feb;542(7639):115-8.
- [12] Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. 2017 Dec;318(22):2199.
- [13] Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal Trial of an Autonomous AI-based Diagnostic System for Detection of Diabetic Retinopathy in Primary Care Offices. *npj Digital Medicine*. 2018 Dec;1(1):39.
- [14] Gulshan V, Rajan RP, Widner K, Wu D, Wubbels P, Rhodes T, et al. Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India. *JAMA Ophthalmology*. 2019 Sep;137(9):987.
- [15] Ruamviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep Learning versus Human Graders for Classifying Diabetic Retinopathy Severity in a Nationwide Screening Program. *npj Digital Medicine*. 2019 Dec;2(1):25.

- [16] Ting DSW, Cheung CYL, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017 Dec;318(22):2211.
- [17] Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology*. 2018 Aug;125(8):1199-206.
- [18] Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmology*. 2017 Nov;135(11):1170.
- [19] Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, et al. A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography. *Ophthalmology*. 2018 Sep;125(9):1410-20.
- [20] Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, et al. Artificial Intelligence for Teleophthalmology-Based Diabetic Retinopathy Screening in a National Programme: An Economic Analysis Modelling Study. *The Lancet Digital Health*. 2020 May;2(5):e240-9.
- [21] Nazir T, Irtaza A, Javed A, Malik H, Hussain D, Naqvi RA. Retinal Image Analysis for Diabetes-Based Eye Disease Detection Using Deep Learning. *Applied Sciences*. 2020 Jan;10(18):6185.
- [22] Dong L, Yang Q, Zhang RH, Wei WB. Artificial Intelligence for the Detection of Age-Related Macular Degeneration in Color Fundus Photographs: A Systematic Review and Meta-Analysis. *eClinicalMedicine*. 2021 May;35.
- [23] Peng Y, Dharssi S, Chen Q, Keenan T, Agrón E, Wong W, et al. DeepSeeNet: A Deep Learning Model for Automated Classification of Patient-based Age-related Macular Degeneration Severity from Color Fundus Photographs. *Ophthalmology*. 2018 Nov;126.
- [24] Studies ARED. Age Related Eye Disease Studies (AREDS); 2020. <https://areds.org/>.
- [25] LabelMe. Label Me; 2022. <https://www.labelme.org/>.
- [26] Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, et al. FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. *Image Analysis & Stereology*. 2014 Aug;33(3):231-4.
- [27] Vrabec JP, Levin LA. The Neurobiology of Cell Death in Glaucoma. *Eye*. 2007 Dec;21(1):S11-4.
- [28] Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global Prevalence of Glaucoma and Projections of Glaucoma Burden through 2040. *Ophthalmology*. 2014 Nov;121(11):2081-90.
- [29] Kang JM, Tanna AP. Glaucoma. *The Medical Clinics of North America*. 2021 May;105(3):493-510.
- [30] Pascal L, Perdomo O, Bost X, Huet B, Otálora S, Zuluaga MA. Multi-Task Deep Learning for Glaucoma Detection from Color Fundus Images. *Scientific Reports*. 2022 Feb.
- [31] Hemelings R, Elen B, Barbosa-Breda J, Blaschko MB, De Boever P, Stalmans I. Deep Learning on Fundus Images Detects Glaucoma beyond the Optic Disc. *Scientific Reports*. 2021 Dec;11(1):20313.
- [32] Huang X, Jin K, Zhu J, Xue Y, Si K, Zhang C, et al. A Structure-Related Fine-Grained Deep Learning System With Diversity Data for Universal Glaucoma Visual Field Grading. *Frontiers in Medicine*. 2022;9.
- [33] Sułot D, Alonso-Caneiro D, Ksieniewicz P, Krzyzanowska-Berkowska P, Iskander DR. Glaucoma Classification Based on Scanning Laser Ophthalmoscopic Images Using a Deep Learning Ensemble Method. *PLOS ONE*. 2021 Jun;16(6):e0252339.
- [34] Thompson AC, Jammal AA, Medeiros FA. A Review of Deep Learning for Screening, Diagnosis, and Detection of Glaucoma Progression. *Translational Vision Science & Technology*. 2020 Jul;9(2):42.
- [35] Mariottoni EB, Datta S, Dov D, Jammal AA, Berchuck S, Tavares I, et al. A Deep Learning-Based Mapping of Structure to Function in Glaucoma. *Investigative Ophthalmology & Visual Science*. 2020 Jun;61(7):3213.
- [36] Juan Verdaguer T. Degeneración macular relacionada a la edad. *Revista Médica Clínica Las Condes*. 2010 Nov;21(6):949-55.
- [37] Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, et al. Global Prevalence of Age-Related Macular Degeneration and Disease Burden Projection for 2020 and 2040: A Systematic Review and Meta-Analysis. *The Lancet Global Health*. 2014 Feb;2(2):e106-16.
- [38] Chew EY, Clemons TE, Agrón E, Launer LJ, Grodstein F, Bernstein PS. Effect of Omega-3 Fatty Acids, Lutein/Zeaxanthin, or Other Nutrient Supplementation on Cognitive Function: The AREDS2 Randomized Clinical Trial. *JAMA*. 2015 Aug;314(8):791.
- [39] González-Gonzalo C, Sánchez-Gutiérrez V, Hernández-Martínez P, Contreras I, Lechanteur YT, Domanian A, et al. Evaluation of a Deep Learning System for the Joint Automated Detection of Diabetic Retinopathy and Age-Related Macular Degeneration. *Acta Ophthalmologica*. 2020;98(4):368-77.
- [40] Keel S, Li Z, Scheetz J, Robman L, Phung J, Makeyeva G, et al. Development and Validation of a Deep-learning Algorithm for the Detection of Neovascular Age-related Macular Degeneration from Colour Fundus Photographs. *Clinical & Experimental Ophthalmology*. 2019 Nov;47(8):1009-18.
- [41] Peng Y, Keenan TD, Chen Q, Agrón E, Allot A, Wong WT, et al. Predicting Risk of Late Age-Related Ma-

cular Degeneration Using Deep Learning. *npj Digital Medicine*. 2020 Aug;3(1):1-10.

- [42] Ganjdanesh A, Zhang J, Chew EY, Ding Y, Huang H, Chen W. LONGL-Net: Temporal Correlation Structure Guided Deep Learning Model to Predict Longitudinal Age-Related Macular Degeneration Severity. *PNAS Nexus*. 2022 Mar;1(1):pgab003.
- [43] Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and Regional Diabetes Prevalence Estimates for 2019 and Projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th Edition. *Diabetes Research and Clinical Practice*. 2019 Nov;157:107843.
- [44] Salongcay RP, Silva PS. The Role of Teleophthalmology in the Management of Diabetic Retinopathy. *Asia-Pacific Journal of Ophthalmology*. 2019.
- [45] Horton MB, Brady CJ, Cavallerano J, Abramoff M, Barker G, Chiang MF, et al. Practice Guidelines for Ocular Telehealth-Diabetic Retinopathy, Third Edition. *Telemedicine and e-Health*. 2020 Apr;26(4):495-543.
- [46] Nguyen HV, Tan GSW, Tapp RJ, Mital S, Ting DSW, Wong HT, et al. Cost-Effectiveness of a National Telemedicine Diabetic Retinopathy Screening Program in Singapore. *Ophthalmology*. 2016 Dec;123(12):2571-80.
- [47] Bellemo V, Lim G, Rim TH, Tan GSW, Cheung CY, Sadda S, et al. Artificial Intelligence Screening for Diabetic Retinopathy: The Real-World Emerging Application. *Current Diabetes Reports*. 2019 Sep;19(9):72.
- [48] Diagnostics D. Digital Diagnostics. AI the Right Way; 2022. <https://www.digitaldiagnostics.com/products/eye-disease/idx-dr/>.
- [49] Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*. 2017 Jul;124(7):962-9.
- [50] Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Development and Evaluation of a Deep Learning System for Screening Retinal Hemorrhage Based on Ultra-Widefield Fundus Images. *Translational Vision Science & Technology*. 2020 Jan;9(2):3.
- [51] Dai L, Wu L, Li H, Cai C, Wu Q, Kong H, et al. A Deep Learning System for Detecting Diabetic Retinopathy across the Disease Spectrum. *Nature Communications*. 2021 May;12(1):3242.
- [52] Guo Y, Hormel TT, Gao L, You Q, Wang B, Flaxel CJ, et al. Quantification of Nonperfusion Area in Montaged Widefield OCT Angiography Using Deep Learning in Diabetic Retinopathy. *Ophthalmology Science*. 2021 Jun;1(2).
- [53] Ruamviboonsuk P, Tiwari R, Sayres R, Nganthavee V, Hemarat K, Kongprayoon A, et al. Real-Time Diabetic Retinopathy Screening by Deep Learning in a Multisite National Screening Programme: A Prospective Interventional Cohort Study. *The Lancet Digital Health*. 2022 Apr;4(4):e235-44.