



Rendimiento de un algoritmo para el diagnóstico de glaucoma basado en la relación excavación/disco del nervio óptico

Autores: Sandra Belalcázar, Shirley Rosenstiehl, Jaime Andrés Suárez-Garavito, María A. Martínez-Ceballos, Claudia Rosa Carvajal, Leonardo Flórez-Valencia

Trabajo presentado como requisito para optar por el
título de Oftalmólogo

Bogotá - Colombia

2022

Rendimiento de un algoritmo para el diagnóstico de glaucoma basado en
la relación excavación/disco del nervio óptico

Presentado por

Jaime Andrés Suárez Garavito

Tutores

Sandra Belalcázar Rey

Shirley Margarita Rosenstiehl Colón

Escuela de Medicina y Ciencias de la Salud

Oftalmología

Universidad del Rosario

Bogotá-Colombia

2022

Identificación del proyecto

Institución académica: Universidad del Rosario

Dependencia: Escuela de Medicina y Ciencias de la Salud

Título de la in

vestigación:

Rendimiento de un algoritmo para el diagnóstico de glaucoma basado en la relación excavación/disco en fotografías de nervio óptico

Instituciones participantes: Fundación Oftalmológica Nacional – Fundonal, Pontificia Universidad Javeriana

Tipo de investigación: Corte Transversal

Investigador principal: Sandra Belalcázar Rey

Investigadores asociados: Shirley Margarita Rosenstiehl Colón, Claudia Carvajal, Jaime Andrés Suárez-Garavito, María Alejandra Martínez-Ceballos, Leonardo Flórez-Valencia

Asesor clínico o temático: Sandra Belalcázar Rey

Asesor metodológico: Shirley Margarita Rosenstiehl Colón

Contenido

1. Introducción	7
1.1 <i>Planteamiento del problema</i>	7
1.2 <i>Justificación</i>	8
2. Marco Teórico	9
3. Pregunta de investigación	11
4. Objetivos	11
4.1 <i>Objetivo general</i>	11
4.2 <i>Objetivos específicos</i>	12
5. Formulación de hipótesis	12
6. Metodología	12
6.1 <i>Tipo y diseño de estudio</i>	12
6.2 <i>Población y muestra</i>	14
6.3 <i>Criterios de inclusión y exclusión</i>	14
6.3.1 <i>Criterios de inclusión:</i>	14
6.3.2 <i>Criterios de exclusión</i>	14
6.4 <i>Tamaño de muestra</i>	14
6.5 <i>Muestreo</i>	15
7. Resultados	16
8. Discusión	18
9. Conclusiones	20
10. Figuras	23
11. Tablas	27
12. Referencias	38

Resumen

Introducción

Los algoritmos para tamizaje de glaucoma determinan la presencia o ausencia de enfermedad basados en la relación excavación/disco (Clasificación de Armaly). Sin embargo, se podrían pasar por alto hallazgos sugestivos de daño del nervio óptico clasificándolo incorrectamente. La escala de probabilidad de daño del disco de Spaeth puede ser útil identificando cambios tempranos de neuropatía óptica glaucomatosa.

Objetivo

Determinar el rendimiento de un algoritmo para tamizaje de glaucoma analizando fotografías de nervio óptico.

Métodos

Estudio transversal de prueba diagnóstica. Se diseñó un algoritmo para identificar características sugestivas de daño en imágenes del nervio óptico. Se clasificaron las imágenes según las escalas de Armaly y Spaeth. El algoritmo segmenta el nervio y la excavación mediante análisis de tono, saturación e iluminación; extrae los contornos con el método de Otsu de múltiples umbrales y cada contorno es representado mediante series de Fourier, alimentando una red neuronal de dos capas.

Resultados

La concordancia del algoritmo con el criterio del especialista fue de 80% (p menor a 0.05) para determinar la relación excavación/disco, 91% (p menor a 0.05) para la escala DDLS de Spaeth, 92% (p menor a 0.05) para la escala DDLS de Spaeth modificada, y de 99% (p menor a 0.05) para la clasificación de severidad de daño por glaucoma. El criterio de los dos especialistas coincidió en 94% (p menor a 0.05) para la clasificación de severidad de daño por glaucoma, 40% para la relación excavación/disco, en 12,4% para la escala DDLS de Spaeth y en 12,5% para la escala DDLS de Spaeth modificada.

Conclusiones

Se propone un algoritmo de inteligencia artificial basado en una red neuronal de 2 capas que alcanza alta precisión en resultados preliminares para clasificar riesgo de glaucoma usando las escalas de Armaly, DDLS de Spaeth y la clasificación de severidad de daño por glaucoma.

Palabras clave

Inteligencia artificial, Oftalmología, Glaucoma, Nervio óptico, Análisis de Fourier, Redes neurales de la computación

Abstract

Introduction

Glaucoma screening algorithms determine the presence or absence of disease based on the cup/disc ratio (Armaly's classification). However, findings suggestive of optic nerve damage may be missed and misclassified as healthy in affected eyes. Spaeth's Disc Damage Likelihood Scale helps to identify early changes of glaucomatous optic neuropathy.

Objective

To determine the performance of an algorithm for glaucoma screening based on Armaly's classification and Spaeth's Disc Damage Likelihood Scale by analyzing optic nerve images.

Methods

Cross-sectional diagnostic test study. An algorithm was designed to identify findings suggestive of optic nerve damage and not only the presence or absence of glaucoma. Optic nerve photos were classified using Armaly's classification and Spaeth's scale. The algorithm segments the optic nerve and cup by analyzing hue, saturation, and lighting values, then extracts contours using Otsu's method with multiple thresholds. Each contour is represented using Fourier series. All the information feeds a two-layer neural network.

Results

The agreement of the algorithm with the specialist's criteria was 80% (p less than 0.05) for determining the cup/disc ratio, 91% (p less than 0.05) for Spaeth's Disc Damage Likelihood Scale (DDLS), 92% (p less than 0.05) Spaeth's modified DDLS and 99% (p less than 0.05) for the glaucoma damage classification. The agreement between the two specialists was 94% (p less than 0.05) for the glaucoma damage classification, 40% for the excavation/disc ratio, 12,4% for Spaeth's DDLS and 12.5% for Spaeth's modified DDLS.

Conclusions

We propose an algorithm based on a 2-layer neural network that in preliminary results achieves high accuracy to estimate glaucoma risk based on Armaly's classification, Spaeth' original and modified DDLS scales and glaucoma damage classification.

Keywords

Artificial intelligence, Ophthalmology, Glaucoma, Optic nerve, Fourier Analysis, Neural Networks

1. Introducción

1.1 Planteamiento del problema

Las ciencias médicas avanzan de forma rápida y constante, el uso de nuevas tecnologías como herramientas de apoyo en el campo de la medicina, se ha constituido como parte importante en los procesos de diagnóstico, manejo y seguimiento de patologías (1–3). La información proveniente de estas herramientas, así como la posibilidad de almacenarla, categorizarla y analizarla (Big Data), se convierte en una fuente valiosa de conocimiento, que puede contribuir a la precisión en el diagnóstico de patologías prevenibles o retrasar su aparición y progresión si son detectadas tempranamente (4). Por este motivo la integración transdisciplinar entre la informática (inteligencia artificial) y la medicina facilitan la posibilidad que estos instrumentos se apliquen cotidianamente en el quehacer de los médicos e investigadores (5).

La inteligencia artificial (IA) es un término que se refiere a la realización de tareas por parte de una red neuronal computarizada, con una mínima intervención del ser humano (6); a partir de sistemas que reúnen toda la información disponible, permitiendo la extracción de datos aplicables, entre los cuales se establecen relaciones que culminan en la creación de algoritmos para diagnosticar y tratar condiciones a partir de modelos de aprendizaje (7) (*machine learning – deep learning*). Dentro de las ventajas que suponen este tipo de sistemas se encuentran una alta precisión diagnóstica, lo que disminuye las posibilidades de error que se derivan del ser humano por factores como el cansancio, la falta de conocimiento o entrenamiento, la variabilidad inter-observador, entre otras; así como la posibilidad de procesar rápidamente grandes cantidades de información, optimizando el uso de recursos limitados a un bajo costo. De esta manera emergiendo como herramientas con un gran potencial si se aplican de manera automatizada a enfermedades de alto impacto y prevalencia(8). La relevancia que tienen los sistemas basados en IA se debe a la poca disponibilidad de especialistas y supra-especialistas en algunas regiones del país, así como los costos y el tiempo requeridos para el entrenamiento de profesionales de la salud, que alcancen el nivel de formación necesario para la identificación de patologías como el glaucoma (6,8,9).

En el área de la oftalmología, la IA ha tenido gran alcance en los últimos años, especialmente en el diagnóstico basado en imágenes (10). Ejemplo de ello es la identificación de glaucoma, su asociación con los cambios en el nervio óptico y hallazgos en el campo visual; el diagnóstico de la retinopatía diabética proliferativa y no proliferativa; así como la detección de la degeneración macular relacionada con la edad (1,11–13). Se destaca el estudio de estas patologías oculares, pues se posicionan como las primeras causas de ceguera irreversible a nivel mundial y su diagnóstico es primordialmente clínico, basado en características propias que son evaluadas por el observador experto, que en este caso es el oftalmólogo(2).

El glaucoma se constituye como una neuropatía óptica y es la primera causa de ceguera prevenible y la segunda causa de ceguera a nivel mundial, con 76 millones de personas estimadas con este diagnóstico en el 2020 (14,15). La prevalencia de la enfermedad en Latinoamérica se estima entre 4,51% - 5,6% según la población estudiada (15,16),

posicionándose como una de las más elevadas después de Asia y África (14). Se espera que para el 2040 existan en el mundo 111.8 millones de personas con glaucoma (14). El glaucoma primario de ángulo abierto es la principal etiología, estimándose su prevalencia alrededor de 3 de cada 4 casos de glaucoma (7,14,17–19). Dentro de los factores de riesgo para el desarrollo de la enfermedad se encuentran la hipertensión ocular, antecedente heredo-familiar, aumento o asimetría de la excavación papilar, edad mayor a los 60 años, ascendencia africana o hispana, hipertensión arterial, miopía y migraña (20). Se constituye como una enfermedad que repercute directamente en la calidad de vida de los pacientes, esto derivado de la cronicidad, progresión y potencial ceguera propias de la enfermedad, siendo esta última el factor más perjudicial para el paciente.

Para un adecuado tratamiento del Glaucoma en el que se busca disminuir o detener su progresión, el paso inicial es la identificación del paciente con esta enfermedad o con sospecha de la misma, teniendo en cuenta que un diagnóstico y tratamiento temprano tiene consecuencias favorables en el control de la enfermedad y la pérdida de función visual (4,9,21). Sin embargo, dentro de las limitaciones para el diagnóstico temprano del Glaucoma se encuentra el hecho de que los pacientes en muchas ocasiones no presenten síntomas hasta etapas avanzadas de la enfermedad, así como la poca disponibilidad de especialistas en Oftalmología o supraespecialistas en Glaucoma, tanto en Colombia como en otros países en vía de desarrollo o sub-desarrollados; situación que se hace aún más evidente en las zonas rurales.

Buscando mejorar la capacidad de diagnosticar pacientes con Glaucoma, se pueden llegar a utilizar sistemas computacionales basados en inteligencia artificial para analizar imágenes del nervio óptico identificando su excavación, con el fin de estudiar al paciente que pueda requerir una valoración más detallada por el especialista. Este escenario proporciona grandes posibilidades como aquella de realizar tamizajes para la identificación temprana de pacientes con Glaucoma, y para su implementación en zonas rurales donde el acceso a un especialista con los conocimientos y habilidades requeridas para la valoración de estos pacientes es más difícil. En este escenario, el advenimiento y la aplicación de la IA resulta una herramienta de apoyo importante para la atención en salud visual.

1.2 Justificación

Se ha descrito que sólo el 55% de los pacientes con glaucoma han sido diagnosticados (18). El modelo de Ladapo y cols., estima que para diagnosticar a un paciente con glaucoma sería necesario tamizar 58 personas(22). Por lo anterior, resulta favorable el desarrollo de estrategias de tamizaje y detección temprana de la neuropatía óptica glaucomatosa, que permitan optimizar el tiempo y el trabajo requerido, reduciendo costos directos e indirectos. Sería ideal que estas estrategias puedan ser aplicadas en diferentes escenarios, especialmente en los países en vía de desarrollo y en las zonas rurales con una disponibilidad limitada de equipos oftalmológicos y de especialistas en glaucoma.

El propósito de este estudio es desarrollar un algoritmo computacional basado en inteligencia artificial, y evaluarlo como una posible herramienta para el tamizaje de glaucoma en la población colombiana, basado en las características del nervio óptico y la relación excavación/disco, según el análisis de fotografías a color del nervio óptico.

2. Marco Teórico

GLAUCOMA

El glaucoma es una neuropatía óptica en la que se genera daño a la capa de fibras nerviosas de la retina que genera pérdida visual irreversible que inicia desde el campo visual periférico hacia la visión central (23).

Se caracteriza clínicamente por el hallazgo de una relación excavación/disco aumentada en el fondo de ojo del paciente con enfermedad avanzada, y se asocia a cifras elevadas de presión intraocular, mayores a 21 mmHg. Sin embargo hay casos de Glaucoma que presentan presión intraocular normal, así como también existen pacientes sanos con una relación excavación/disco aumentada; por lo que para el diagnóstico del Glaucoma se requieren distintos estudios (tomografía de coherencia óptica de nervio óptico, campos visuales, tonometría, paquimetría, fotografías a color del nervio óptico) así como su seguimiento en el tiempo con el fin de evaluar si hay cambios o progresión de los hallazgos encontrados en la consulta inicial (23,24)

El glaucoma se puede clasificar según su etiología y la configuración del ángulo de la cámara anterior de la siguiente manera (23,24)

- Glaucoma Primario de Ángulo Abierto: Neuropatía óptica en pacientes sin una etiología asociada y sin evidencia de cierre angular según la gonioscopia.
- Glaucoma Primario de Ángulo Cerrado: Neuropatía óptica que se asocia a elevación de la presión intraocular por una obstrucción del drenaje del humor acuoso en un paciente con cierre del ángulo iridocorneal.
- Glaucoma Secundario: Neuropatía óptica asociada a elevación de la presión intraocular por otra condición o patología oftalmológica.
- Sospecha de Glaucoma: Pacientes con elevación de la presión intraocular mayor a 21 mmHg, un disco óptico con excavación o características sugestivas de cambios por Glaucoma, asimetría mayor a 0.2 en la relación excavación disco entre un ojo y el otro, o defectos en la campimetría que sean sugestivos de glaucoma.

Para evaluar las características clínicas del nervio óptico, se han desarrollado escalas y clasificaciones, como la desarrollada por Armaly y cols., que corresponde a la medición de la relación excavación/disco y que es ampliamente usada en el contexto clínico (25). Posteriormente, Spaeth y cols., desarrollaron la escala de probabilidad de daño del disco óptico (DDLs), que tiene en cuenta las características del anillo neuroretiniano de acuerdo con el tamaño del disco óptico y que permite la identificación de daño temprano del nervio óptico (26). Algunos autores han modificado la numeración de la clasificación de Spaeth y otros la han utilizado para caracterizar la severidad del daño de forma cualitativa pero más objetiva (27,28).

INTELIGENCIA ARTIFICIAL:

La inteligencia artificial comprende el diseño de sistemas que son capaces de tener un comportamiento que se considera inteligente. Estos sistemas toman una serie de datos para extraer características claves que les permiten reconocer patrones, identificar imágenes o tomar decisiones, según el análisis realizado con base en la información previamente aprendida (1,29) Existen distintas aplicaciones de inteligencia artificial actualmente, incluyendo aplicaciones en el campo de la medicina y la oftalmología. Las técnicas usadas para diseñar estos sistemas de inteligencia artificial han avanzado en el tiempo, permitiendo la creación de sistemas más complejos que logran extraer e integrar información proveniente de distintas fuentes, como diferentes estudios diagnósticos que ofrecen información funcional y estructural de un órgano para llegar a un diagnóstico con niveles de sensibilidad y especificidad que alcanzan aquellos del médico, y en algunos casos, lo superan (30,31) En el contexto de la IA, existen diferentes tipos de redes neuronales, cuyas características se resumen en la **tabla 1**.

INTELIGENCIA ARTIFICIAL EN EL DIAGNOSTICO DE GLAUCOMA:

Se realizó una búsqueda en la base de datos Pubmed, con los términos “(Artificial Intelligence) AND (Glaucoma)”, se aplicaron los filtros: estudios en seres humanos y artículos de los últimos 5 años, se incluyeron estudios observacionales, analíticos y algunas revisiones de la literatura. Los resultados arrojados en total fueron 87 artículos, de los cuales se eligieron 26 de acuerdo con el título y resumen.

La mayoría de las investigaciones adelantadas en IA se han desarrollado en países industrializados como Estados Unidos, Alemania, España, Japón, centrándose primordialmente en establecer el rendimiento de la IA como prueba diagnóstica, comparada con la capacidad del experto oftalmólogo en la detección de cambios a nivel de nervio óptico. En cuanto a Latinoamérica, con esta búsqueda se encontró un estudio desarrollado en Brasil por Cerentini y col., en el que se evidenció una precisión diagnóstica de la red neuronal que alcanza el 90% (32)

Los tamaños de las muestras varían en número, siendo el más pequeño el de Zilly y col. con 50 imágenes, reconociendo este aspecto como limitante, pues se requieren tamaños de muestras más grandes para llegar a una conclusión extrapolable en la aplicación de la estrategia de redes neuronales que ellos proponen (33) Este tamaño muestral contrasta con aquellas investigaciones que incluyeron el análisis de bases de datos públicas, contando con una mayor cantidad de información. Se resalta el trabajo adelantado por Park y col., en el que se propuso crear una base de datos propia con pacientes de Korea durante 6 meses, evaluando inicialmente la concordancia diagnóstica en diferentes patologías del polo posterior entre 28 oftalmólogos, esto mediante un sistema de diagnóstico paso a paso, para posteriormente alimentar un sistema integrado de machine learning, modelo que puede ser replicado y mejorado para aumentar el rendimiento de las redes neuronales convolucionales (34).

Es de resaltar que todos los estudios realizados en los últimos 5 años representan un grado de validez global de esta herramienta diagnóstica, con niveles de sensibilidad y especificidad

superiores al 80%. En el caso del estudio de Díaz-Pinto y col., en el que tenían como objetivo demostrar la validez de la red neuronal Image-Net como herramienta diagnóstica a partir de la información registrada en 5 bases de datos, encontraron una alta validez global de la prueba, demostrada por un área bajo la curva de 0.93 y 0.85 para sensibilidad y especificidad respectivamente (35). En la **tabla 2** se resumen los resultados de algunos estudios encontrados en la revisión bibliográfica.

Sin embargo, se debe tener en cuenta que existen algunas limitaciones al momento de su aplicación, como las diferencias poblacionales y operativas; entre ellas se mencionan la raza, el estado de refracción (miopías altas), el grado de severidad del glaucoma, o el uso de diferentes cámaras fotográficas con resoluciones variables. Para el adecuado control de estos sesgos, resulta necesario aumentar la disponibilidad de información respecto a estas características (19) Teniendo en cuenta el control de estos factores de confusión, y el entrenamiento correcto de algoritmos de inteligencia artificial a través de fotografías de fondo de ojo de pacientes con glaucoma de ángulo abierto, se puede llegar a obtener una mayor sensibilidad y una especificidad comparable respecto a la de oftalmólogos expertos (1)

La mayoría de los algoritmos de IA se basan en la clasificación de Armaly, que describe la relación excavación/disco. Esta clasificación es ampliamente utilizada en el ejercicio clínico por ser fácil y rápida de aplicar, sin embargo, presenta limitaciones como la subjetividad y una moderada variabilidad inter e intra-observador (27).

Varias estrategias de perfeccionamiento en la identificación de características del nervio óptico y de la relación excavación/disco han sido desarrolladas, entre ellas, el mapa térmico del nervio óptico a partir de fotografías de fondo de ojo, el uso de mapas de bits, la delimitación y extracción de segmentos del nervio óptico a través de una representación por píxeles, la extracción de segmentos específicos del nervio óptico evitando el ruido en las imágenes, identificación de las excavaciones extrayendo estructuras vasculares, detección automática del nervio óptico y la excavación, así como la inclusión de información complementaria proveniente de otras herramientas diagnósticas como la tomografía óptica coherente y el campo visual 24-2 (1,3,8,12,20,36,37)

3. Pregunta de investigación

¿Cuál es la concordancia diagnóstica de un algoritmo computacional basado en inteligencia artificial para el diagnóstico de glaucoma, que analiza fotografías de nervio óptico; con el diagnóstico clínico del médico especialista en Oftalmología?

4. Objetivos

4.1 Objetivo general

Desarrollar un algoritmo basado en inteligencia artificial y evaluar su concordancia diagnóstica con el criterio del especialista en oftalmología al analizar fotografías de nervio óptico, basando

su análisis en la relación excavación/disco, la escala de probabilidad de daño del disco (DDLDS) de Spaeth, la escala de Spaeth modificada y la clasificación de severidad de daño por glaucoma.

4.2 *Objetivos específicos*

1. Diseñar un algoritmo computacional de forma interdisciplinar con el área de ingeniería de sistemas, que sea capaz de segmentar y analizar imágenes del nervio óptico
2. Implementar el algoritmo para enfrentarlo al análisis de fotografías del nervio óptico
3. Caracterizar clínica y cualitativamente los datos obtenidos de las fotografías de nervio óptico para determinar los grupos de entrenamiento para el algoritmo.
4. Comparar el rendimiento del algoritmo en términos de concordancia diagnóstica con la caracterización de las fotografías realizada por los expertos

5. **Formulación de hipótesis**

Un algoritmo de inteligencia artificial entrenado en la segmentación y análisis de la relación excavación/disco en fotografías del nervio óptico para el diagnóstico de glaucoma, así como en las clasificaciones DDLDS de Spaeth, Spaeth modificada y la clasificación de severidad de daño por glaucoma; puede alcanzar una concordancia diagnóstica similar con el criterio del especialista en oftalmología.

H0: índice de concordancia diagnóstica es de 0

Ha: Índice de concordancia diagnóstica es mayor de 0

6. **Metodología**

6.1 *Tipo y diseño de estudio*

Estudio transversal analítico y de concordancia diagnóstica para determinar el rendimiento de una red neuronal superficial densa en la identificación de daño del nervio óptico, basado en la relación excavación/disco (Armaly), la escala de probabilidad de daño del disco (DDLDS) de Spaeth, la escala de Spaeth modificada y la clasificación de severidad de daño por glaucoma.

Para el estudio se diseñó y entrenó una red neuronal con una serie de fotos ya analizadas por los especialistas, en las que se determinó el valor de la relación excavación/disco con una escala numérica entre 0.1-0.9, aproximada a un decimal, así como su correspondiente clasificación .

Desarrollo, diseño y prueba del algoritmo

Dos oftalmólogos (SR y SB) revisaron 310 imágenes de nervios ópticos, sobre las cuales se aplicaron los criterios de inclusión y exclusión. Se realizó la medición y clasificación de 166 imágenes de nervios ópticos. Se registraron los datos de forma independiente y ciega en una base de datos en Excel™ V.16.51. Las imágenes fueron analizadas y delimitadas en el aplicativo GIMP V.2.99.6, el cual permite delinear el contorno del nervio óptico y su excavación, para definir

el límite del anillo neural. Una representación de la medición de las imágenes se puede observar en la **figura 1**. La información obtenida y clasificada se utilizó para el entrenamiento del algoritmo y para evaluar la precisión de éste.

El algoritmo se desarrolló usando la metodología de programación extrema (XP: eXtreme Programming). Esta metodología es considerada ágil y de buena aceptación en ambientes de investigación científica. Las características de dicha metodología son:

1. Desarrollo iterativo e incremental: se hacen pequeños esfuerzos (*sprints*) con avances demostrables rápidamente a los interesados en el proyecto (*stakeholders*).
2. Frecuente integración del (de los) programador(es) con los usuarios. Para este proyecto se propone una periodicidad de 15 días entre reuniones de entrega/presentación.
3. Pruebas unitarias continuas, escritas antes de la codificación. Esto garantiza la calidad numérica de la implementación (es decir, alta cobertura del código/baja relación de *bugs* por cada línea de código).
4. Refactorización del código: reescribir ciertas partes del código para aumentar su legibilidad sin modificar su comportamiento.
5. Simplicidad en el código.

El algoritmo tiene tres pasos principales: primero, el nervio se segmenta mediante un análisis de saturación de tono; en segundo lugar, la excavación del disco se extrae enmascarando la imagen alrededor del nervio y analizando el valor de iluminación y finalmente, ambos contornos segmentados se convierten a su representación de la serie de Fourier para alimentar una red neuronal simple y densa de dos capas para clasificar el caso. Dicho algoritmo se basa en gran medida en nuestra hipótesis que el nervio óptico y los vasos pueden segmentarse fácilmente al procesar los datos contenidos en los canales de color, iluminación y saturación de tono (en el sentido de la informática, un algoritmo sencillo tiene tiempos de ejecución rápidos; estos algoritmos también se conocen por tener baja complejidad temporal). Esto permite realizar el paso de clasificación con algoritmos de aprendizaje automático bastante simples, en términos de ejecución y entrenamiento.

Para el entrenamiento del algoritmo, los oftalmólogos revisaron la base de datos de imágenes, obteniendo el grupo final después de aplicar los criterios de inclusión y exclusión. Determinaron según su criterio clínico el valor de la relación excavación/disco (Clasificación de Armaly); luego delinearon el contorno del nervio óptico y de su excavación, midiendo los diámetros horizontales y verticales de ambos contornos. Posteriormente se calculó el cociente entre anillo neural/disco óptico y con esta información se categorizaron las imágenes según las clasificaciones DDLS de Spaeth, Spaeth modificada, y clasificación de severidad de daño por glaucoma (que finalmente clasifica a las imágenes dentro de las categorías normal, en riesgo, daño por glaucoma o discapacidad por glaucoma). Con esta información se construyó la base de datos en Excel, con la que se alimentó el algoritmo en la fase de entrenamiento, complementando las imágenes medidas con las que fue entrenado.

Posterior al proceso de entrenamiento se realiza la validación del sistema para confirmar que el algoritmo computacional sea capaz de identificar la relación excavación/disco con el mismo paquete de imágenes que se utilizaron para su entrenamiento.

Finalmente se realiza la validación del sistema, en la que se presentan a la red neuronal una serie de imágenes del nervio óptico diferentes a las utilizadas para su entrenamiento, para que realice una estimación de la relación excavación/disco, que sea comparada con aquella realizada por los especialistas.

Análisis estadístico

El desarrollo, entrenamiento y validación del algoritmo fue realizado utilizando lenguaje Python™. El acuerdo entre los dos observadores fue determinado por medio de una matriz de confusión, en la cual se calculó el Score f1.

6.2 Población y muestra

Fotografías de fondo de ojo a color, de pacientes que han asistido a la Fundación Oftalmológica Nacional.

6.3 Criterios de inclusión y exclusión de las imágenes seleccionadas en el estudio

6.3.1 Criterios de inclusión:

- Fotografías de fondo de ojo a color, de pacientes que han asistido a la Fundación Oftalmológica Nacional, en las cuales se observe la totalidad del nervio óptico.
- Fotografías correspondientes a pacientes mayores de 18 años.
- Las fotografías debían contar con las siguientes características: buena resolución, ausencia de artefactos, campo de trabajo de 30°, que incluya toda la cabeza del nervio óptico, iluminación aceptable (ni oscura, ni sobreexpuesta), enfoque suficiente que permita la visualización óptima del contorno del nervio óptico, el anillo neural y la excavación.

6.3.2 Criterios de exclusión:

- Opacidad de medios que impida la visualización de detalles del nervio óptico como leucomas, opacidades corneales, catarata hipermadura, hemorragia vítrea, entre otros.
- Coexistencia o antecedentes de otras patologías del nervio óptico como papilitis, neuritis óptica, neuropatía óptica isquémica anterior, entre otras.
- Fotografías correspondientes a pacientes con diagnóstico de Glaucoma secundario.
- Fotografías correspondientes a pacientes con diagnóstico previo de Glaucoma.
- Antecedente de oclusiones vasculares retinianas.

6.4 Tamaño de muestra

Las redes neuronales mejoran su rendimiento a medida que reciban mayor cantidad de información durante su entrenamiento, por lo cual entre mayor sea el número de imágenes utilizadas para el aprendizaje de la red neuronal, mejor será su capacidad de estimar la relación excavación/disco.

En el presente estudio se estima que un 60% de las imágenes disponibles sean utilizadas para el entrenamiento de la red neuronal, 20% para la prueba y 20% para la validación del algoritmo.

Para el tamaño muestral se dispondrá de 702 fotos de nervio óptico para determinar según los criterios de inclusión y exclusión, cuáles son aptas para ser seleccionadas en el estudio.

6.5 Muestreo

Se recolectaron 702 fotos a color de fondo de ojo de la base de imágenes correspondientes a 702 ojos de pacientes que han asistido a la Fundación Oftalmológica Nacional. De estas imágenes, 392 correspondían a pacientes con diagnóstico previo de glaucoma, motivo por el cual no fueron tenidas en cuenta. Sobre las 310 imágenes restantes se aplicaron los demás criterios de inclusión y exclusión, obteniendo en total 166 imágenes aptas para el estudio.

En cuanto a posibles sesgos del estudio, al aplicar el puntaje QUADAS-2 (0-14) se obtiene un bajo riesgo de sesgo en 3 dominios y riesgo de sesgo intermedio en el primer dominio (ver anexo) (38).

6.6 Consideraciones Éticas

El estudio se realizó dentro de los principios éticos para las investigaciones médicas. Se clasifica como una investigación sin riesgo, teniendo en cuenta las regulaciones locales del Ministerio de Salud de Colombia Resolución 8430 de 1993 en lo concerniente al Capítulo I “De los aspectos éticos de la investigación en seres humanos (39):

ARTICULO 11. *Para efectos de este reglamento las investigaciones se clasifican en las siguientes categorías: a) Investigación sin riesgo: Son estudios que emplean técnicas y métodos de investigación documental retrospectivos y aquellos en los que no se realiza ninguna intervención o modificación intencionada de las variables biológicas, fisiológicas, psicológicas o sociales de los individuos que participan en el estudio, entre los que se consideran: revisión de historias clínicas, entrevistas, cuestionarios y otros en los que no se le identifique ni se traten aspectos sensitivos de su conducta.*

ARTÍCULO 8 *En las investigaciones en seres humanos se protegerá la privacidad del individuo, sujeto de investigación, identificándolo solo cuando los resultados lo requieran y éste lo autorice”.*

Todos los integrantes del grupo de investigación estuvieron y estarán prestos a dar información sobre el estudio a entes organizados, aprobados e interesados en conocerlo siempre y cuando sean de índole académica y científica, preservando la exactitud de los resultados y haciendo referencia a datos globales y no a pacientes o instituciones en particular. Se mantuvo y mantendrá absoluta confidencialidad y se preservará el buen nombre institucional profesional. El estudio se realizó con un manejo estadístico imparcial y responsable.

Este estudio cuenta con aprobación del comité de ética del centro de referencia en oftalmología CEI057/2020 acta No. 194 del 11 de marzo de 2020.

7. Resultados

El desarrollo del algoritmo de análisis de imágenes de fondo de ojo incluyó los siguientes pasos:

Segmentación de la imagen

El esquema general de la segmentación del nervio óptico se detalla en la **figura 2**.

Este algoritmo se basa en dos hipótesis que se sustentan en la configuración de nuestro sistema de captura de imágenes: el tinte dominante rojo de la imagen y la alta iluminación del nervio óptico.

A su vez, este sistema se compone de una imagen de entrada que corresponde a $F(x,y)$ como una matriz de 1476 columnas por 1187 filas en formato RGB de 8 bits por canal. Por otra parte, se compone de una imagen de salida (contorno del nervio óptico) que se representa por una una secuencia de números complejos donde:

$$Z=[z_0,z_1,z_{-1},z_2,z_{-2},z_3,z_{-3},z_4,z_{-4}]$$

Extracción de canales

Siguiendo nuestras hipótesis sobre la imagen de entrada, esta es transformada en sus canales tinte (*hue*), saturación (*saturation*) e iluminación (*lightness*). Estos canales representan numéricamente la percepción humana de los colores, lo que nos permite usar más fácilmente las intensidades de cada píxel.

El canal de tinte se compone de valores representados en grados en el rango $[0,360]$, lo que nos permite priorizar los píxeles cercanos al rojo. En la **figura 3** se muestra la interpretación de este canal.

El tinte $H(x,y)$ representa los colores puros, de hecho, cada 60° se representa uno de los colores básicos en los canales aditivos RGB y CMY. El color rojo tiene asociado un ángulo de 0° , entonces, cuando se usa ese valor para extraer el color puro de cada píxel de la imagen se puede ver que en su mayoría tienen colores naranja-rojo-púrpura (**figura 3**).

El canal de iluminación $L(x,y)$ está definido en el rango adimensional $[0,1]$, donde 0 representa el negro y 1 es el blanco; todos los otros valores representan qué tan oscuro (en el rango $(0,0.5)$) o qué tan claro (en el rango $[0.5,1)$) es el color. Particularmente, los colores que perceptualmente son muy claros y con un tinte evidente, tienen valores cercanos 0.5.

El algoritmo de conversión de RGB (canales red-green-blue) a HSL (canales hue-saturation-lightness) es bien conocido en la literatura. Una buena descripción formal de éste se puede encontrar en el estudio de Agoston y cols. (40).

Realce del nervio

Con la imagen de tinte $H(x,y)$ e iluminación $L(x,y)$ calculada como se presentó antes, se define el filtro de realce como: Este filtro gaussiano produce números que están en el rango $[0,1]$ y prioriza dos tipos de valores, alineados con nuestras dos hipótesis:

- Los tintes rojos, en un rango de confianza de $\approx 30^\circ$ (entre naranja y púrpura);
- Las iluminaciones de colores evidentes (alrededor del rango $[0.3,0.7]$).

Explicado de una forma más intuitiva: cuando un píxel tiene “buena” iluminación y es cercano al color rojo, el filtro produce un valor cercano a 1; en caso contrario los valores serán cercanos a 0. Un ejemplo del resultado de este filtro se muestra en la **figura 4**.

Binarización Kappa- Sigma y componente más grande

La binarización kappa-sigma calcula un umbral para separar dos regiones en la imagen: una región “no usable” de otra “usable”. Esto se logra analizando el histograma de la imagen $N(x,y)$ para detectar dos modos en él: el modo “usable” y el modo “no usable”. El modo usable es aquel que es más compacto en el histograma. Este algoritmo se describe en el artículo de Lehmann (41).

Este valor es usado para binarizar la imagen, a partir de la cual se extrae la región conectada más grande $B(x,y)$, que corresponde al nervio óptico. Un ejemplo de este resultado se muestra en la **figura 4**.

Casco convexo y coeficiente de Fourier

A partir de la máscara $B(x,y)$ se extraen las coordenadas de los puntos que tienen valores no nulos (pertenecientes a la máscara). Estas coordenadas se usan para calcular el casco convexo de la máscara (25). El casco convexo es el polígono de área mínima que envuelve esos puntos. Ese polígono $CH = [p_i]$ es una secuencia de puntos bidimensionales, a partir de los cuales se calculan los coeficientes de Fourier donde:

$Z=[z_0, z_1, z_{-1}, z_2, z_{-2}, z_3, z_{-3}, z_4, z_{-4}]$.

El coeficiente z_0 representa el centro del contorno y cada pareja de coeficientes z_i y z_{-i} representa una elipse que gira alrededor del centro del contorno; esos datos son suficientes para representar de una forma compacta el polígono.

Caracterización de la imagen

Una vez segmentado, el caso se representa con la pareja de secuencias $[O,E]$ que tienen los datos para representar el contorno de nervio óptico y de la excavación, respectivamente. Esto resulta en 18 números complejos (36 números reales).

Para la caracterización (etiquetado) de la imagen se usan esos 36 números reales para alimentar una red neuronal densa que tiene las siguientes características: Dos capas activadas con la función ReLU de 72 neuronas la primera y 18 la segunda. Una capa de salida activada con la función softmax de 10 neuronas de salida, una por cada etiqueta de salida.

Entrenamiento de la red

La red se entrenó siguiendo el procedimiento:

1. Sobre un grupo de 166 imágenes resultantes después de aplicar los criterios de inclusión y exclusión, dos oftalmólogos trazaron los contornos del nervio óptico y la excavación.
2. Cada uno de esos trazos fue convertido a su representación en coeficientes de Fourier.
3. Esa base de 166 casos fue aumentada a 60.000 casos: cada contorno de Fourier fue escalado (con un factor entre 0.9 y 1.1) y desfasado aleatoriamente.
4. La base de datos final fue dividida aleatoriamente en tres conjuntos: 70% de entrenamiento, 20% de prueba y 10% de validación.

5. Cada caso fue normalizado respecto al radio máximo del coeficiente fundamental del contorno del nervio óptico.
6. La red fue entrenada usando un tamaño de grupo (batch) de 32, tasa de aprendizaje 0.01, sin regularización.
7. Los dos algoritmos se implementaron en C++ (estándar 2017) usando las librerías ITK, Eigen3 y CGAL. El entrenamiento de la red neuronal se hizo con la librería TensorFlow sobre el lenguaje Python (versión 3.9).

La **tabla 3** muestra la concordancia diagnóstica del algoritmo comparado con el estándar de oro en el presente trabajo que fue el criterio del especialista según las clasificaciones de Armaly (relación excavación/disco), la escala DDLS de Spaeth, Spaeth modificada y la clasificación de severidad de daño por glaucoma. Se muestra la concordancia de los grupos de imágenes de entrenamiento, prueba y validación del algoritmo. Estos resultados fueron calculados con el "F1-score".

La tabla 4. Presenta además las características de tamaño, diámetro, clasificación de Armaly, Clasificación DDLS de Spaeth, clasificación DDLS de Spaeth modificada y la clasificación de severidad de daño por glaucoma.

Implementación del algoritmo

El algoritmo propuesto alcanza una concordancia del 80% (p-valor del F1-Score menor a 0.05) en la determinación de la relación excavación/disco (Armaly) y de un 92% (p-valor F1-Score menor a 0.05) en la clasificación DDLS de Spaeth, a comparación del criterio del especialista.

El criterio de los dos especialistas coincidió en 94% (p-valor f1-score menor a 0.05) de los casos para severidad del daño, en 40% para la clasificación de Armaly, en 12,4% para la clasificación DDLS de Spaeth y en 12,5% para la clasificación DDLS de Spaeth modificada.

8. Discusión

La mayoría de las investigaciones adelantadas en IA se han desarrollado en países industrializados como Estados Unidos, Alemania, España y Japón, centrándose primordialmente en establecer el rendimiento de la IA como prueba diagnóstica, comparada con la capacidad del experto oftalmólogo en la detección de cambios a nivel de nervio óptico. En Latinoamérica está publicado un estudio desarrollado en Brasil por Cerentini y col., en el que se evidenció una precisión diagnóstica de la red del 90% (32). Resultados similares a los encontrados en nuestro estudio.

Estos resultados, aunque preliminares por la falta de casos reales (60.000 mil imágenes a partir de 166 fotos a color del nervio óptico), son alentadores. La concordancia del algoritmo en el grupo de entrenamiento en las cuatro categorías está por encima del 95%, que podría indicar un problema de sobre ajuste (*over-fitting*); por esto se evalúa la precisión sobre el conjunto de casos que no se tuvieron en cuenta durante el entrenamiento. Los resultados en prueba y validación son buenos para la clasificación de Severidad de daño por glaucoma, para la escala DDLS de Spaeth y la escala DDLS de Spaeth modificada; observación soportada en los valores de

concordancia que son comparables con aquellos del grupo de entrenamiento. En el caso de la medida de la relación excavación/disco (Armaly), la concordancia alcanzada en el entrenamiento fue de un 96%, sin embargo, en la validación disminuyó a un 80%, tal vez por la subjetividad propia de dicha estimación clínica (**tabla 3**).

Es de resaltar que todos los estudios realizados en los últimos 5 años representan un grado de validez global de esta herramienta diagnóstica, con niveles de sensibilidad y especificidad superiores al 80%. En el caso del estudio de Díaz-Pinto y cols., en el que tenían como objetivo demostrar la validez de la red neuronal Image-Net como herramienta diagnóstica a partir de la información registrada en 5 bases de datos, encontraron una alta validez global de la prueba, demostrada por un área bajo la curva de 0.93 y 0.85 para sensibilidad y especificidad respectivamente (35), valores que se asemejan a los encontrados en nuestro estudio.

Como se mencionó previamente, estos resultados son alentadores, pero requieren de más pruebas; particularmente sobre una base de casos clínicos reales, variados, etiquetados por un grupo más grande de expertos; a diferencia de este estudio en el cual se utilizaron casos aumentados mediante el escalamiento y el desfaseamiento aleatorizado de cada contorno de Fourier. Además, se requiere hacer una validación del algoritmo “en campo”, donde expertos puedan alimentar matrices de confusión con pacientes nuevos atendidos en consulta.

Una de las limitaciones de la clasificación de Armaly es la subjetividad de su resultado, consideramos que la rigurosidad y detalle de los parámetros en la clasificación de DDLS y DDLS modificada, permitiría un mayor acuerdo en el resultado y por lo tanto mayor reproducibilidad del método. Observamos, en el caso de la clasificación de Armaly, que la concordancia entre los resultados de los dos oftalmólogos es baja para esta clasificación (40%), resultados acordes a lo descrito en otros estudios dada la subjetividad de esta estimación (27). En cuanto a la clasificación de DDLS y DDLS modificada, el observador 1 (SR) tiende a estimar los casos en un nivel más grave (por ejemplo, observador 2, etiqueta "nivel 4" y observador 1 etiqueta "nivel 5"). Si se modifican artificialmente las estimaciones de SR un nivel hacia abajo, la clasificación resultante tiene una concordancia del 96% entre los dos observadores. Además, las dos clasificaciones de DDLS son coherentes: tanto SR como SB tienen tendencia a evaluar de la misma forma en ambas escalas. La clasificación de severidad fue la misma entre los dos observadores en un 94% de los casos.

El uso de herramientas automatizadas basadas en sistemas computacionales e inteligencia artificial se ha convertido en una alternativa prometedora para identificar tempranamente y de forma más eficiente a los pacientes con ciertas enfermedades. La aplicabilidad clínica de estas herramientas es particularmente significativa para enfermedades como el glaucoma, en donde un diagnóstico y tratamiento tempranos permiten disminuir la progresión de la enfermedad, así como la pérdida irreversible de función visual que pueden sufrir estos pacientes.

Quigley y cols. reportan que solo alrededor de un 55% de los pacientes con glaucoma han sido diagnosticados (18) y el modelo descrito por Ladapo y cols. estima que, para diagnosticar a 1 paciente con glaucoma, sería necesario tamizar 58 pacientes (22). Los sistemas de salud ocupan a la mayoría de los oftalmólogos en los procesos de diagnóstico y tratamiento, limitando los programas de promoción y prevención en salud visual para patologías como el glaucoma.

Teniendo en cuenta que el daño causado por el glaucoma es irreversible, es primordial buscar estrategias de tamizaje que sean costo efectivas y que puedan ser aplicadas a grupos poblacionales en diferentes escenarios, particularmente en zonas rurales y países en desarrollo donde la disponibilidad de equipos avanzados y de supraespecialistas en glaucoma, es limitada. El escenario ideal permitiría aplicar este tipo de herramientas en dispositivos como celulares inteligentes que se encuentran ampliamente distribuidos y que puedan ser usados por personal de la salud como una herramienta de ayuda diagnóstica, independientemente de la disponibilidad de especialistas, internet, o equipos de diagnóstico avanzado.

El entrenamiento del algoritmo con la clasificación de DDLS, DDLS modificado y clasificación de severidad basada en el DDLS, introduce información de medición más objetiva del daño de los nervios ópticos, teniendo en cuenta la proporción excavación/disco, el grosor del anillo neural en su punto más delgado y el tamaño del nervio óptico. El algoritmo podrá de esta manera, identificar daño temprano por glaucoma de forma más eficiente. Dentro de la literatura relacionada con IA, no hay una combinación de estas clasificaciones de daño de nervio óptico, para entrenamiento de algoritmos de diagnóstico sobre fotografías de fondo de ojo.

9. Conclusiones

Los algoritmos basados en IA permiten el análisis automatizado de datos clínicos, con la ventaja de utilizar de manera más eficiente recursos limitados como el tiempo y la disponibilidad de personal capacitado. La mayoría de los algoritmos desarrollados se basan en redes neuronales convolucionales. Proponemos un nuevo modelo de análisis de datos con resultados preliminares que muestran una alta concordancia entre el algoritmo y el criterio del especialista, para identificar daño del nervio óptico por glaucoma, según la relación excavación/disco, la escala DDLS de Spaeth, Spaeth modificada y clasificación de la severidad del daño.

Como resultados de este estudio, se encontró que la segmentación de las imágenes se logra completar adecuadamente al tener en cuenta los canales de la imagen que contienen la información de tono, saturación, color e iluminación. De esta manera es posible programar un algoritmo de inteligencia artificial con una programación sencilla, que sea entrenado con estas imágenes, su segmentación y su análisis según distintas clasificaciones, que permiten un estudio más detallado del nervio óptico y sus características clínicas de forma objetiva. Se encontró una alta concordancia diagnóstica entre el algoritmo y el criterio del especialista, para los resultados de la clasificación de severidad, Spaeth y Spaeth modificada.

Se encontró que la clasificación de Armaly asignada por los dos oftalmólogos, coincidió únicamente en 40% de las imágenes, mientras que la clasificación de severidad fue la misma entre los dos observadores en un 94% de los casos. Estos resultados son concordantes con los obtenidos por el algoritmo, donde en la fase de prueba y validación se obtiene una concordancia respecto al criterio del especialista, del 80% para la relación excavación/disco (clasificación de Armaly), 91% para la escala DDLS de Spaeth, 92% para la clasificación DDLS de Spaeth modificada y 99% para la clasificación de severidad de daño por glaucoma.

Dentro de las fortalezas de este estudio se encuentra el uso de escalas de clasificación (DDL de Spaeth original y modificada, clasificación de severidad de daño por glaucoma) que tienen en cuenta distintos parámetros y que permiten categorizar las imágenes de forma menos subjetiva que lo que permitiría el utilizar únicamente la clasificación de Armaly valorando la relación excavación/disco. Se tuvieron en cuenta criterios de inclusión y exclusión que ayudan a reducir sesgos asociados a otras patologías oculares. Se utilizaron fotografías tomadas con la misma cámara de fondo de ojo, con la misma configuración de captura en cuanto a tamaño y resolución de imagen. Es de resaltar que con la segmentación lograda en las imágenes y con el tipo de programación utilizado para el algoritmo, el entrenamiento de este se puede realizar en tiempos cortos, a bajo costo y sin requerir capacidades de cómputo muy altas, como si lo requiere el entrenamiento de algoritmos basados en redes neurales convolucionales. Esto permitiría continuar alimentando el algoritmo y mejorando su entrenamiento desde cualquier lugar con el que se cuente con cualquier equipo de cómputo moderno, e inclusive desde dispositivos móviles.

Dentro de las limitaciones del presente estudio se encuentran el limitado número de oftalmólogos, pues entre más observadores clasifiquen las imágenes, más completo el entrenamiento del algoritmo; el limitado número de imágenes reales con las que se realiza el entrenamiento del algoritmo y el limitado número de imágenes resultantes en los grupos de prueba y validación. Otra limitación es aquella que concierne respecto a la calidad de las imágenes tomadas en un contexto clínico real, donde en ocasiones puede ser más difícil lograr obtener una imagen nítida, bien enfocada, con una adecuada visualización del nervio óptico; en cuyo caso el rendimiento del algoritmo sería diferente y el análisis alcanzado de menor confiabilidad.

Estudios adicionales podrían brindar información respecto al rendimiento del algoritmo en un entorno clínico con pacientes reales, así como evaluar el rendimiento del algoritmo al ampliar el conjunto de imágenes usadas para el entrenamiento. Algunos de los retos que implica el desarrollar herramientas basadas en inteligencia artificial en el ámbito de la salud incluyen el lograr que su implementación logre llegar a las comunidades menos favorecidas y que los conjuntos de entrenamiento tengan en cuenta una población diversa para minimizar posibles sesgos en el análisis de la información. Este estudio muestra resultados prometedores y puede ser un punto de partida para el desarrollo de herramientas basadas en inteligencia artificial cuya implementación sea sencilla, rápida y de bajo costo que sean más fáciles de llevar a lugares apartados y con menos recursos. Otros estudios posteriores podrían brindar más información, teniendo en cuenta que ya está construida la base de la programación del algoritmo, al ser alimentados con una base de datos y de imágenes más amplia y diversa, con imágenes que sean analizadas por un grupo más amplio de oftalmólogos.

10. Figuras

Figura 1

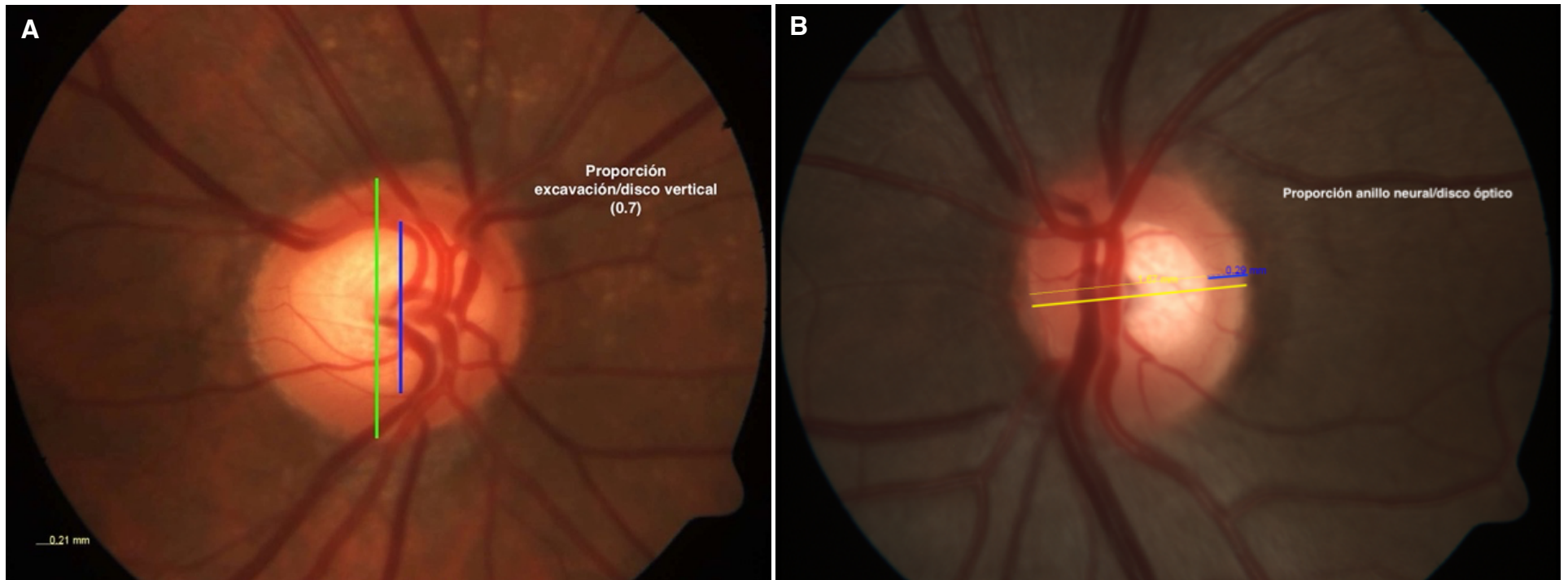


Figura 1A:

Representación de la clasificación de Armaly: Proporción excavación/disco vertical

Figura 1B:

Representación de la proporción anillo neural/disco óptico donde la línea azul muestra la parte más delgada del anillo neuroretiniano. Sobre esta medición se basa la clasificación DDLS de Spaeth.

Figura 2

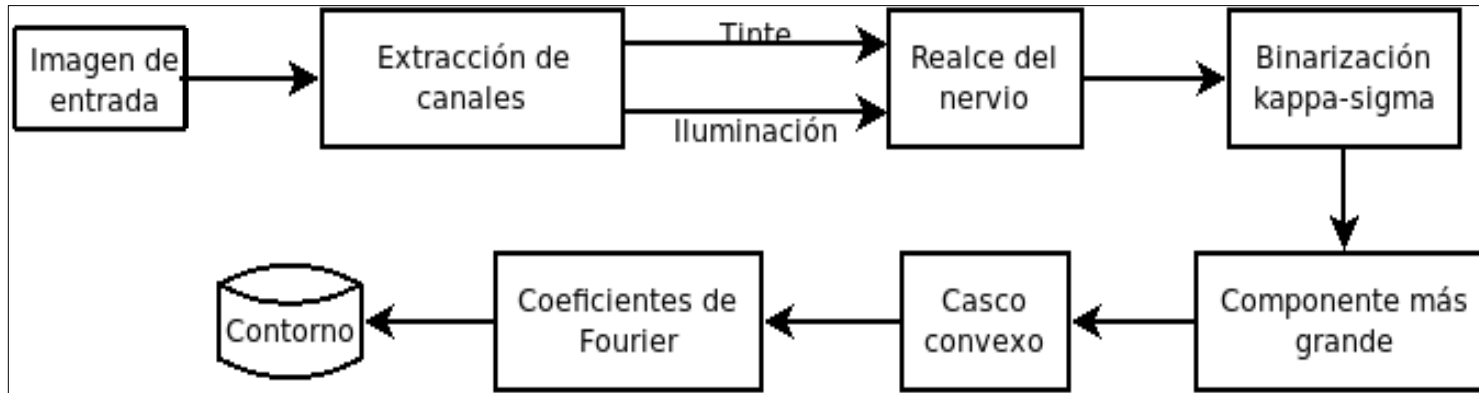


Figura 2:

Esquema general de la segmentación del nervio óptico

Figura 3

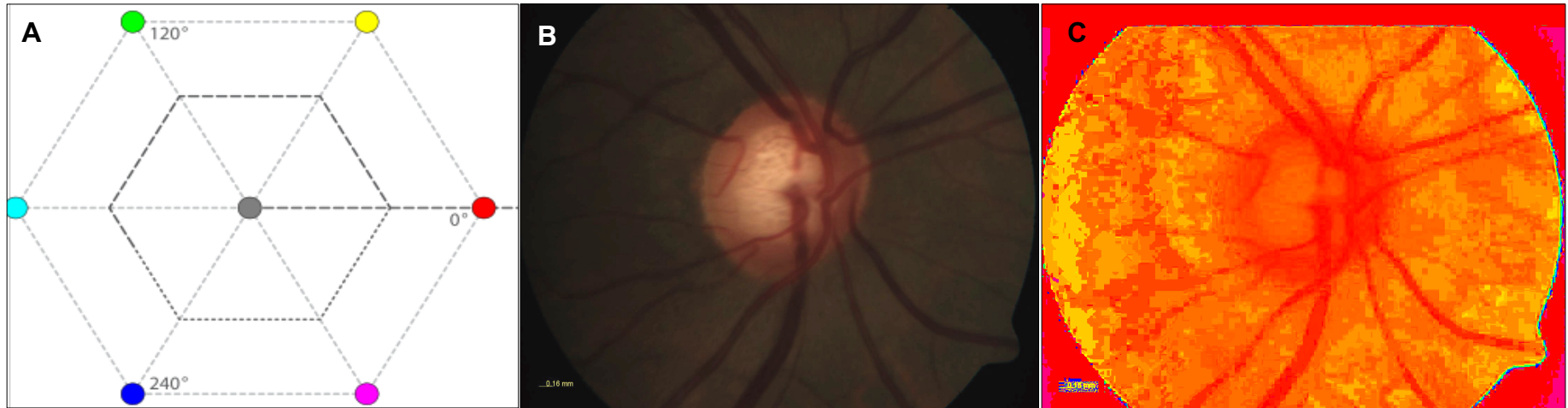


Figura 3:

Interpretación del canal del tinte. A: Hexágono de apoyo para interpretar los tintes de una imagen como ángulos. El rojo corresponde a 0° . B: Imagen de fondo de ojo original. C: Imagen del tinte de cada pixel (tinte dominante de la imagen).

Figura 4



Figura 4:

A: Fotografía de nervio óptico. B: Resultado de nuestro algoritmo de realce del nervio óptico. C: En verde, los pixeles que hacen parte de la región que corresponde al nervio óptico.

11. Tablas

Tabla 1. Características de los tipos de redes en inteligencia artificial

Redes superficiales	Redes profundas
Basadas en operaciones tensoriales sencillas	
Requieren de un conocimiento experto del contexto para su entrenamiento	
Fáciles de construir	Difíciles de construir
Sirven en contextos donde se tiene conocimiento del modelo subyacente de los datos	Sirven en contextos donde no se conoce muy bien el modelo subyacente de los datos
Funcionan en ambientes de vectores de características pequeñas (decenas de datos)	Funcionan en ambientes de vectores de características grandes (miles de datos)
En general, el entrenamiento es corto (del orden de algunas horas), entonces se pueden hacer muchos entrenamientos a bajo costo	El entrenamiento puede ser largo (puede llegar a las semanas), entonces es más costoso
Normalmente no requiere de mucho poder de cómputo (entrenamiento y uso)	Normalmente requiere de un poder de cómputo importante (principalmente en entrenamiento)

Tabla 2. Resumen de algunos de los estudios encontrados en la base de datos Pubmed

Autor	Año	Tipo de IA	n	Resultados
Phan, S	2019	DCNN	3312	1. Tres DCNN reportaron AUC de 0.9 o más.
				2. El tamaño de la imagen no afectó la capacidad discriminativa.
				3. El mapa de calor mostró que el área del disco óptico era la más importante para la discriminación de glaucoma
Thompson, A	2019	Algoritmos de aprendizaje profundo	9282	1. Las predicciones de AP de apertura de la membrana de Bruch en las fotografías de disco óptico se correlacionaron altamente con los valores observados en el OCT-SD.
				2. El AUC para discriminar ojos sanos de ojos con glaucoma con las predicciones del AP y la apertura de la membrana de Bruch por OCT-SD fue 0,945 (IC: 0.874–0.980) y 0.933 (IC: 0.856–0.975).
Park, SJ	2018	NA	71560	Disponibilidad de un conjunto de datos a gran escala con un alto nivel de información, que puede utilizarse en el desarrollo futuro de algoritmos basados en aprendizaje automático para la identificación de condiciones anormales y el sistema de soporte de decisiones clínicas.
Cerentini, A	2017	RNA	724	1. Precisión del 90% en la base de datos de fondos de alta resolución, RIM-ONE r1 precisión del 94,2%, RIM-ONE r2 precisión del 86,2% RIM-ONE r3 precisión del 86,4%.
				2. La combinación de las 3 bases de datos resultó en una precisión del 87,6%.
				3. Se desarrolló un algoritmo para la detección de la región de interés.

Haleem, S	2017	NA	3000	1. Se desarrolló un modelo de clasificación del nervio óptico y excavación en función de la clasificación por píxeles.
				2. Crear un modelo dinámico para la identificación del área de interés que no se limita a una figura circular o elíptica.
Singh, A	2016	Red neuronal convolucional.	63	1. El algoritmo de redes neuronales y <i>random forest</i> tuvieron una precisión del 94,7%. Con una sensibilidad y especificidad del 100% y 87.5% para las redes neuronales y 90.9% y 100% para el modelo de <i>random forest</i> . KN-N Y SVM los de mejor rendimiento para el análisis vectorial.
Zilly, J	2016	Red neuronal convolucional.	50	Los resultados obtenidos pueden ser aplicados cuando la disponibilidad de datos es menor, cuando se compara con las redes neuronales convolucionales tradicionales.
Bajwa, MN	2019	1ª etapa: CNN	67	Precisión del 100% en 4 sets de datos para localizar el nervio óptico
		2ª etapa: DCNN		El AUC para la clasificación como Glaucoma 0.874, mejoría relativa del 2.7% respecto a resultados previos en el set ORIGA 0.874 mejoría relativa del 2.7% respecto a los mejores resultados previos en el set ORIGA
Diaz-Pinto, A	2019	CNN	1707	El promedio del AUC 0.9605 IC95% (95.92-97.07%). Usando arquitectura Xception: Sensibilidad promedio de 0.9346 Especificidad promedio 0.8580.
Shankaranarayana, SM	2019	CNN	650	La técnica usada para estimar la profundidad del Nervio Óptico supera a las anteriormente usadas en imágenes individuales del fondo de ojo en la base de datos INSPIRE.
MacCormick, IJC	2019	Algoritmo probabilístico espacial	809	El algoritmo espacial identifica los discos sanos y con glaucoma con una precisión del 99.6% en validaciones internas y del 91%.

Ahn, JM	2018	Clasificación logística simple y CNN	1542	<p>El modelo de clasificación logística tuvo una precisión: Entrenamiento 82.9% Validación 79.9% Prueba 77.2%</p> <p>La CNN logró una precisión y AUC de: Entrenamiento 92.2% - 0.98 Validación 88.6% - 0.95 Prueba 87.9% - 0.94</p> <p>El modelo de <i>GoogleNet Inception v3</i> con transferencia de aprendizaje logró una precisión y AUC de: Entrenamiento 99.7% - 0.99 Validación 87.7% - 0.95 Prueba 84.5% - 0.93</p>
Christopher M, Zangwill LM	2018	Evaluaron múltiples modelos	14822	<p>El mejor modelo alcanzó: AUC 0.91 para diferenciar ojos sanos y con glaucoma AUC 0.97 para identificar pérdida funcional moderada-severa AUC 0.89 para identificar pérdida funcional leve</p> <p>Sensibilidad 88% Especificidad 95%</p> <p>Para detectar glaucoma moderado-severo</p>
Mitra, A	2018	CNN	11200	<p>El método propuesto alcanzó una precisión del 99.05% en la base de datos Kaggle, y del 98.78% en la base de datos MESSIDOR para detectar la región de interés.</p> <p>Detección de la región de interés 0.0045 segundos con una latencia de 25</p>

				milisegundos, lo que supera los mejores tiempos y no requiere de intervenciones "manuales".
Shibata, N	2018	DCNN	3132	El algoritmo de aprendizaje profundo alcanzó un rendimiento diagnóstico significativamente mayor que los residentes de oftalmología. El AUC para Resnet fue de 96.5% (IC 95%: 93.5-99.6%) El AUC para los 3 residentes estuvo entre 72.6-91.2%
Fu, H	2018	NA	8109	Tienen en cuenta 4 flujos profundos en diferentes niveles y módulos para el análisis de las imágenes: 1.Flujo de imagen global 2.Red guiada por segmentación 3.Flujo de región de disco local 4.Flujo de transformación de disco polar
Li, Z	2018	Sistema de aprendizaje profundo	48116	AUC de 0.986 Sensibilidad 95.6% Especificidad 92%. Falsos negativos en pacientes con miopía alta, retinopatía diabética, y degeneración macular asociada a la edad. Falsos positivos en aumento de la excavación fisiológica del nervio óptico y miopía.
Omodaka, K	2017	DCNN	163	Precisión de la red neuronal 87.8%.

Tabla 3 Concordancia diagnóstica del algoritmo entrenado según cada categoría de daño del nervio óptico para cada escala

Categoría	Entrenamiento	Prueba y validación
Severidad	0.9872	0.9958
Armaly	0.9653	0.8006
Spaeth	0.9569	0.9182
Spaeth modificada	0.9798	0.9235

Tabla 4. Descripción de características del nervio óptico y escalas

Variable	N (%)
Tamaño	
Pequeño	20 (17.39)
Mediano	89 (77.39)
Grande	6 (5.22)
Escala de Spaeth (DDLS) Modificada	
Estadio 2	17 (9.24)
Estadio 3	132 (71.74)
Estadio 4	29 (15.76)
Estadio 5	2 (1.09)
Estadio 7a	4 (2.17)
Severidad del Glaucoma	
At risk	148 (80.88)
Glaucoma damage	31 (16.94)
Glaucoma disability	4 (2.18)
Escala Disc Damage Likelihood Scale (DDLS) de Spaeth	
0b - Sin daño	15 (8.77)
Estadio 1	123 (71.93)
Estadio 2	28 (16.37)
Estadio 3	2 (1.17)
Estadio 4	2 (1.17)
Estadio 5	1 (0.58)
Variable	Promedio +/- D. E
Diámetro total	1.67 +/- 0.21
Clasificación de Spaeth modificada (DDLS modificada)	4.17 +/- 0.78

Tabla 5. Variables de estudio

	Variable	Definición operativa	Tipo de variable	Escala
1	Tamaño	Tamaño del disco óptico	Cualitativa ordinal	1: Pequeño 2: Mediano 3: Grande
2	Diámetro	Recta que une dos puntos de una circunferencia, pasando por su centro	Cuantitativa continua	mm
3	Radio	Segmento lineal que une un punto cualquiera de la circunferencia o de la superficie de una esfera con su centro	Cuantitativa continua	mm
4	Cociente	Razón entre el radio y diámetro total del disco óptico	Cuantitativa continua	NA
5	Clasificación de Armaly	Relación entre la excavación del disco óptico y el tamaño del disco óptico (25)	Cuantitativa continua	Resultado de la clasificación

6	Clasificación de Spaeth (DDLs)	Escala de probabilidad de daño del nervio óptico causado por glaucoma basado en la apariencia del ANR y ajustado al diámetro del disco óptico (26)	Cualitativa ordinal	0: 0a - Sin daño 1: 0b - Sin daño 2: Estadio 1 3: Estadio 2 4: Estadio 3 5: Estadio 4 6: Estadio 5 7: Estadio 6 8: Estadio 7a 9: Estadio 7b
7	Clasificación de Spaeth modificada (DDLs modificada)	Estadio del glaucoma de acuerdo con las características clínicas y o paraclínicas propuestas por la guía canadiense de glaucoma (24)	Cualitativa ordinal	1: Estadio 1 2: Estadio 2 3: Estadio 3 4: Estadio 4 5: Estadio 5 6: Estadio 6 7: Estadio 7 8: Estadio 8 9: Estadio 9 10: Estadio 10

8	Severidad del daño	Nomograma de la escala de probabilidad de daño del disco.	Cualitativa ordinal	Normal: Estadio 0a y 0b En riesgo: Estadio 1 y 2 Daño por glaucoma: 3, 4 y 5 Discapacidad por glaucoma: 6 y 7
---	--------------------	---	---------------------	--

NA: No aplica, ANR: Anillo neuroretiniano

ANEXO:

Evaluación QUADAS-2

Dominio	Evaluación	Observaciones
Selección de pacientes	Riesgo de sesgo intermedio	Se incluyeron pacientes de forma consecutiva
Prueba índice	Bajo riesgo de sesgo	Fueron interpretados los resultados de la prueba índice sin conocimiento de los resultados de la de referencia
Prueba de referencia	Bajo riesgo de sesgo	Fueron interpretados los resultados de la prueba de referencia sin conocimiento de los resultados de la prueba índice
Flujos y tiempos	Bajo riesgo de sesgo	Fue aplicada en todos los individuos la misma prueba de referencia

12. Referencias

1. Zheng C, Johnson T v., Garg A, Boland M v. Artificial intelligence in glaucoma. *Curr Opin Ophthalmol*. 2019;30(2):97–103.
2. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. *Clin Exp Ophthalmol*. 2019;47(1):128–39.
3. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. *British Journal of Ophthalmology*. 2019;103(2):167–75.
4. Coleman AL, Miglior S. Risk factors for glaucoma onset and progression. *Surv Ophthalmol* [Internet]. noviembre de 2008;53 Suppl1:S3-10. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/19038621>
5. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems [Internet]. 2018. Disponible en: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>
6. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism* [Internet]. 2017;69:S36–40. Disponible en: <http://dx.doi.org/10.1016/j.metabol.2017.01.011>
7. Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli M v, Das A, Jonas JB, et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob Health*. septiembre de 2017;5(9):e888–97.
8. Du XL, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol*. 2018;11(9):1555–61.
9. Park I, Gale J, Skalicky SE. Health Economic Analysis in Glaucoma. *J Glaucoma* [Internet]. abril de 2020;29(4):304–11. Disponible en: <https://journals.lww.com/10.1097/IJG.0000000000001462>
10. Liu H, Li L, Wormstone IM, Qiao C, Zhang C, Liu P, et al. Development and Validation of a Deep Learning System to Detect Glaucomatous Optic Neuropathy Using Fundus Photographs. *JAMA Ophthalmol*. 2019;137(12):1353–60.
11. Phan S, Satoh S, Yoda Y, Kashiwagi K, Oshika T, Hasegawa T, et al. Evaluation of deep convolutional neural networks for glaucoma detection. *Jpn J Ophthalmol*. 2019;63(3):276–83.
12. Smits DJ, Elze T, Wang H, Pasquale LR. Machine Learning in the Detection of the Glaucomatous Disc and Visual Field. *Semin Ophthalmol* [Internet]. 2019;34(4):232–42. Disponible en: <https://doi.org/10.1080/08820538.2019.1620801>
13. Biggerstaff KS, Lin A. Glaucoma and Quality of Life. *Int Ophthalmol Clin*. 2018;58(3):11–22.
14. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: A systematic review and meta-analysis.

- Ophthalmology [Internet]. 2014;121(11):2081–90. Disponible en: <http://dx.doi.org/10.1016/j.ophtha.2014.05.013>
15. Rivera CE, Cantor E, Castillo A, Martinez A, Newball L, Rueda JC, et al. Prevalence of Primary Open Angle Glaucoma among Patients with Diagnosis of Systemic Hypertension and Diabetes Mellitus: The Colombian Glaucoma Study. *Open J Ophthalmol*. 2020;10(02):99–114.
 16. Wurster P, Harris A, Gonzalez AC, Adjei S, Verticchio Vercellin A, Mathew S, et al. Risk factors for open-angle glaucoma in persons of latin american descent. *J Glaucoma*. 2020;29(3):217–25.
 17. Pascolini D, Mariotti SP. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*. mayo de 2012;96(5):614–8.
 18. Quigley HA. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*. marzo de 2006;90(3):262–7.
 19. Kapetanakis V v, Chan MPY, Foster PJ, Cook DG, Owen CG, Rudnicka AR. Global variations and time trends in the prevalence of primary open angle glaucoma (POAG): a systematic review and meta-analysis. *British Journal of Ophthalmology*. enero de 2016;100(1):86–93.
 20. FB VB, J JR. Guía latinoamericana de glaucoma primario de ángulo abierto. 2019. 1–93 p.
 21. Gordon MO, Beiser JA, Brandt JD, Heuer DK, Higginbotham EJ, Johnson CA, et al. The Ocular Hypertension Treatment Study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol* [Internet]. el 1 de junio de 2002;120(6):714–20; discussion 829-30. Disponible en: <http://archophth.jamanetwork.com/article.aspx?doi=10.1001/archophth.120.6.714>
 22. Ladapo JA. Projected Clinical Outcomes of Glaucoma Screening in African American Individuals. *Archives of Ophthalmology* [Internet]. el 1 de marzo de 2012;130(3):365. Disponible en: <http://archophth.jamanetwork.com/article.aspx?doi=10.1001/archophthalmol.2011.1224>
 23. Foster PJ. The definition and classification of glaucoma in prevalence surveys. *British Journal of Ophthalmology* [Internet]. el 1 de febrero de 2002;86(2):238–42. Disponible en: <https://bjo.bmj.com/lookup/doi/10.1136/bjo.86.2.238>
 24. Canadian Ophthalmological Society evidence-based clinical practice guidelines for the management of glaucoma in the adult eye / Guide factuel de pratique clinique de la Société canadienne d'ophtalmologie pour la gestion du glaucome chez l'adulte. *Canadian Journal of Ophthalmology* [Internet]. 2009;44(S1):S1–93. Disponible en: <http://article.pubs.nrc-cnrc.gc.ca/RPAS/rpv?hm=Hlnit&calyLang=eng&journal=cjo&volume=44&afpf=cjo44s1.pdf>
 25. Armaly MF. The Cup/Disc Ratio. *Archives of Ophthalmology* [Internet]. el 1 de agosto de 1969;82(2):191. Disponible en: <http://archophth.jamanetwork.com/article.aspx?doi=10.1001/archophth.1969.00990020193008>
 26. Spaeth GL, Henderer J, Liu C, Kesen M, Altangerel U, Bayer A, et al. The disc damage likelihood scale: reproducibility of a new method of estimating the amount of optic nerve damage caused by glaucoma. *Trans Am Ophthalmol Soc* [Internet]. 2002;100:181–5; discussion 185-6. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/12545692>

27. Kara-José AC, Melo LAS, Esporcatte BLB, Endo ATNH, Leite MT, Tavares IM. The disc damage likelihood scale: Diagnostic accuracy and correlations with cup-to-disc ratio, structural tests and standard automated perimetry. Acott TS, editor. *PLoS One* [Internet]. el 20 de julio de 2017;12(7):e0181428. Disponible en: <https://dx.plos.org/10.1371/journal.pone.0181428>
28. Pahlitzsch M, Torun N, Erb C, Bruenner J, Maier AK, Gonnermann J, et al. Significance of the disc damage likelihood scale objectively measured by a non-mydriatic fundus camera in preperimetric glaucoma. *Clinical Ophthalmology* [Internet]. noviembre de 2015;2147. Disponible en: <https://www.dovepress.com/significance-of-the-disc-damage-likelihood-scale-objectively-measured--peer-reviewed-article-OPHTH>
29. Lu W, Tong Y, Yu Y, Xing Y, Chen C, Shen Y. Applications of Artificial Intelligence in Ophthalmology: General Overview. *J Ophthalmol* [Internet]. el 19 de noviembre de 2018;2018:1–15. Disponible en: <https://www.hindawi.com/journals/joph/2018/5278196/>
30. 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* [Internet]. el 2 de febrero de 2017;542(7639):115–8. Disponible en: <http://www.nature.com/articles/nature21056>
31. Bejnordi BE, Zuidhof G, Balkenhol M, Hermsen M, Bult P, van Ginneken B, et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging* [Internet]. el 14 de diciembre de 2017;4(04):1. Disponible en: <https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-4/issue-04/044504/Context-aware-stacked-convolutional-neural-networks-for-classification-of-breast/10.1117/1.JMI.4.4.044504.full>
32. Cerentini A, Welfer D, d’Ornellas MC, Haygert CJP, Dotto GN. Automatic identification of glaucoma using deep learning methods. *Stud Health Technol Inform.* 2017;245:318–21.
33. Zilly J, Buhmann JM, Mahapatra D. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Computerized Medical Imaging and Graphics* [Internet]. 2017;55:28–41. Disponible en: <http://dx.doi.org/10.1016/j.compmedimag.2016.07.012>
34. Park SJ, Shin JY, Kim S, Son J, Jung KH, Park KH. A novel fundus image reading tool for efficient generation of a multi-dimensional categorical image database for machine learning algorithm training. *J Korean Med Sci.* 2018;33(43):1–12.
35. Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. CNNs for automatic glaucoma assessment using fundus images: An extensive validation. *Biomed Eng Online* [Internet]. 2019;18(1):1–19. Disponible en: <https://doi.org/10.1186/s12938-019-0649-y>
36. Chaurasia AK, Greatbatch CJ, Hewitt AW. Diagnostic Accuracy of Artificial Intelligence in Glaucoma Screening and Clinical Practice. *J Glaucoma.* mayo de 2022;31(5):285–99.
37. Ittoop SM, Jaccard N, Lanouette G, Kahook MY. The Role of Artificial Intelligence in the Diagnosis and Management of Glaucoma. *J Glaucoma.* marzo de 2022;31(3):137–46.

38. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* el 10 de diciembre de 2003;3(1):25.
39. Resolución número 8430 de 1993, Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Colombia:
<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/RESOLUCION-8430-DE-1993.PDF>; 1993.
40. Agoston MK. *Computer Graphics and Geometric Modeling* [Internet]. London: Springer-Verlag; 2005. 908 p. Disponible en: <http://link.springer.com/10.1007/b138805>
41. Lehmann G. Kappa Sigma Clipping. *Insight J.* 2006;July-Decem.