



DETERMINACION DE LA CONCORDANCIA DEL DAÑO DEL NERVIO OPTICO
ENTRE UN GLAUCOMATOLOGO Y UN ALGORITMO DE APRENDIZAJE

Vanessa Carpio Rosso Delgado, MD

Trabajo presentado como requisito para optar por el título de Oftalmólogo

Bogotá D.C- Colombia

2020

DETERMINACION DE LA CONCORDANCIA DEL DAÑO DEL NERVIO OPTICO
ENTRE UN GLAUCOMATOLOGO Y UN ALGORITMO DE APRENDIZAJE

Vanessa Carpio Rosso Delgado, MD

Tutores

Sandra Belalcazar Rey, MD

Shirley Rosenstiehl, MD

Escuela de Medicina y Ciencias de la Salud

Oftalmología

Universidad del Rosario

Bogotá D.C- Colombia

2020

Identificación del proyecto

Institución académica: Universidad del Rosario

Dependencia: Escuela de Medicina y Ciencias de la Salud

Título de la investigación: Determinación de la concordancia del daño del nervio óptico entre un glaucomatólogo y un algoritmo de aprendizaje

Instituciones participantes: Fundación Oftalmológica Nacional, Universidad del Rosario

Tipo de investigación: Estudio de concordancia

Investigador principal: Sandra Belalcazar Rey, MD. Oscar Julian Perdomo, PhD.

Investigadores asociados: Vanessa Carpio Rosso Delgado, Hernán Andrés Ríos , Shirley Rosenstiehl Colon

Asesor clínico o temático: Sandra Belalcazar Rey, MD

Asesor metodológico: Shirley Rosenstiehl, MD

“La Universidad del Rosario no se hace responsable de los conceptos emitidos por los investigadores en su trabajo, solo velará por el rigor científico, metodológico y ético del mismo en aras de la búsqueda de la verdad y la justicia”.

Agradecimientos

Contenido

1. Introducción	8
1.1. Planteamiento del Problema	8
1.2. <u>Justificación</u>	<u>9</u>
2. Marco Teórico	10
3. Pregunta de investigación	15
4. Objetivos	15
4.1. <i>Objetivo general</i>	15
4.2. <i>Objetivos específicos</i>	15
5. Metodología	16
5.1. Tipo y diseño de estudio.....	16
5.2. <i>Población</i>	16
5.3. <i>Tamaño de la muestra</i>	17
5.4. Criterios de selección.....	17
5.4.1. <i>Criterios de inclusión</i>	17
5.4.2. <i>Criterios de exclusión</i>	17
5.5. <i>Variables</i>	17
5.6. <i>Plan de análisis</i>	18
5.7. Control de sesgo y error.....	19
5.8. Porceso de recolección.....	19
6. Aspectos éticos	19
7. Administración del proyecto	20
7.1. <i>Cronograma</i>	20
7.2. <i>Presupuesto</i>	21
8. Resultados	21
9. Discusión	22
10. Referencias	25

Resumen

Propósito: Determinar la concordancia entre la interpretación de las fotos a color de polo posterior de un especialista en glaucoma y un algoritmo de aprendizaje no supervisado para determinar el daño del nervio óptico.

Metodología: Se realizó un estudio de concordancia diagnóstica entre la interpretación de las fotos a color de polo posterior de un especialista en glaucoma y un algoritmo de aprendizaje no supervisado con respecto a la identificación del daño del nervio óptico según el sistema de clasificación de Armaly y usando el coeficiente de kappa de Cohen.

Resultados:

El algoritmo de aprendizaje no supervisado evaluó 689 fotos a color de polo posterior, clasificadas como con nervio óptico sano (sin daño) y con daño leve, moderado y severo. Posteriormente un clasificador K-means, agrupó las características extraídas en los cuatro grupos mencionados y se obtuvo un coeficiente kappa de Cohen de 0.037. Cuando se clasificaron las imágenes en dos grupos, sanos y con daño, se evidenció un estadístico kappa para la clasificación dicotómica de 0.03.

Conclusión:

El Algoritmo de aprendizaje no supervisado usado para la clasificación de daño del nervio óptico en fotos a color de polo posterior, mostró una mala concordancia con la realizada por el especialista en glaucoma según el sistema de clasificación de Armaly.

Palabras claves:

Algoritmo de aprendizaje no supervisado, fotos a color de polo posterior, daño del nervio optico.

Abstract

Purpose: To determine the concordance between an Unsupervised Learning Algorithm and eye fundus color photos interpretation by a specialist for the identification of the optic disc damage.

Methodology: A concordance study between an Unsupervised Learning Algorithm and a glaucoma specialist was made. The Cohen's kappa coefficient was calculated for identification of the optic disc damage in eye fundus color photos and were assessed according to Armaly's cup/disc ratio classification.

Results: The Unsupervised Learning Algorithm evaluated 689 color optic disc images classified as: healthy (no damage), mild, moderate and severe damage. A k-means classifier clustered the extracted features in four groups and obtained a Cohen's kappa coefficient of 0.037. While classifying the images in two groups: Healthy and with damage, we found a Cohen's kappa coefficient of 0.03.

Conclusion: The Unsupervised Learning Algorithm for the classification of optic disc damage on color fundus photos showed a bad concordance with the one done by the glaucoma specialist, using Armaly's cup/disc ratio classification.

Key Word: Unsupervised Learning Algorithm, color fundus photos, optic disc damage.

1. Introducción

1.1. Planteamiento del Problema

El Glaucoma es la segunda causa de ceguera a nivel global (1) y es la causa principal de ceguera irreversible en el mundo. Se ha estimado que el número total de personas con glaucoma en el año 2020 es de 64.3 millones y este se incrementará a 111.8 millones para el año 2040 (2). Por otro lado Cook y colaboradores reportaron que para el año 2020, 5.9 millones de personas quedarán ciegas debido al glaucoma de ángulo abierto y que el glaucoma de ángulo cerrado producirá ceguera en 5.3 millones de personas (3).

El pobre conocimiento sobre la salud visual por parte de la población y el limitado acceso a servicios de salud constituyen verdaderas barreras para tener una evaluación oftalmológica adecuada, lo que conduce a un diagnóstico tardío de la enfermedad (4) incrementando la probabilidad de ceguera en los pacientes. Hattenhauer y colaboradores mostraron que la probabilidad acumulada de glaucoma asociada a ceguera durante 20 años es del 27% en la población general (5) y siendo el Glaucoma una de las principales causas de pérdida de visión es importante detectar la enfermedad en un estadio temprano (6).

Algunos estudios evidencian que hay un porcentaje considerable de glaucoma no diagnosticado debido a la falta de una herramienta simple, precisa y costo efectiva para realizar el tamizaje del glaucoma (12). En el trabajo de Los Ángeles Latino Eye Study, se evaluaron 6357 pacientes latinos en el estado de Los Ángeles y se mostró una prevalencia de Glaucoma Primario de Angulo Abierto del 4.74%, de los cuales el 75% de los casos no habían sido previamente diagnosticados (7). De acuerdo al proyecto Vision Evaluation and Research (VER) la prevalencia de glaucoma en 4774 pacientes fue del 1.97%, de los cuales tan solo un 38% había sido diagnosticado antes de este estudio (8). En el Baltimore Eye Survey se encontró una prevalencia de 3.73% de Glaucoma Primario de Angulo Abierto en 5308 pacientes, de los cuales el 46% de los casos no habían sido diagnosticados previamente (10). Por otro lado se ha visto que en comparación con la población blanca no hispana, los hispanos y la población afroamericana tienen un riesgo aumentado de presentar glaucoma no diagnosticado debido a un difícil acceso a la evaluación oftalmológica (7).

Algunos estudios han demostrado que la medición de la presión intraocular y el campo visual no son suficientemente específicos o sensibles para ser utilizados como herramientas de tamizaje (9)(10) y por otra parte el uso de la tomografía de coherencia óptica es de alto costo para realizar este tipo de medición.

La evaluación del nervio óptico ha sido uno de los métodos más ampliamente usados en la valoración de los pacientes con glaucoma y las fotos de esta estructura han probado ser herramientas muy útiles para su documentación (11). Esta valoración puede ser realizada por un profesional entrenado pero ello no es útil para realizar un tamizaje en masa. Por ello el desarrollo de sistemas de aprendizaje profundo supervisado y no supervisado para la evaluación y caracterización de fotografías de nervio óptico podrían ser herramientas útiles en el tamizaje de glaucoma en poblaciones con dificultad para el acceso a una atención especializada.

1.2. Justificación

El glaucoma es una neuropatía óptica degenerativa y constituye la segunda causa de ceguera en el mundo, representando el 12,3% de la ceguera total (12). La enfermedad se caracteriza por alteraciones estructurales específicas del nervio óptico con consecuentes alteraciones funcionales detectadas en el campo visual (13). En esta patología se produce el daño de las células ganglionares y sus axones, lo cual está relacionado con el aumento de la presión intraocular (3), manifestándose en un adelgazamiento del anillo neuroretiniano y el correspondiente aumento de la excavación del nervio óptico, características que pueden ser evidenciadas en las fotos a color de polo posterior del ojo (14).

La evaluación del nervio óptico por tanto, juega un papel esencial en la detección y manejo del glaucoma (3). Las fotos a color de polo posterior siguen siendo el método de elección para la documentación de la apariencia del nervio óptico (15) y su uso para el tamizaje del glaucoma podría ser más eficiente que la medición de la presión intraocular (16).

No existe una cura definitiva para el glaucoma, pero su diagnóstico y tratamiento temprano pueden prevenir la pérdida de la visión. Estudios en India y otros países han demostrado que entre el 50% y 90% de los casos de glaucoma permanecen sin ser detectados en la población (17)(18). Sumado a lo anterior el proceso de diagnóstico estándar es costoso y de difícil

acceso en la mayoría de países en desarrollo y como respuesta a ello se han realizado algunos esfuerzos en la detección automática del glaucoma en una etapa temprana con diferentes sistemas automatizados (19).

En el último cuarto de siglo, el aprendizaje profundo se ha convertido en uno de los aspectos más resaltantes de la revolución tecnológica informativa y precisamente la medicina es una de sus áreas de mayor aplicación. Existen dos tipos de aprendizaje profundo, el supervisado y el no supervisado, este último es uno de los territorios más fascinantes de esta tecnología con el que se analizan grandes bases de datos y se descubren patrones subyacentes sin la necesidad de dar instrucciones previas (20).

En glaucoma se está usando para identificar la excavación de nervios ópticos en fotos a color de polo posterior, para analizar campos visuales y tomografías de coherencia óptica y se está proyectando como una herramienta útil para el tamizaje y evaluación de la progresión de esta patología (21).

2. Marco Teórico

2.1. Patología del Glaucoma

El glaucoma es una neuropatía óptica degenerativa crónica que afecta específicamente a las células ganglionares y sus axones, dando lugar a un adelgazamiento progresivo de la capa de fibras nerviosas, lo cual se acompaña de un adelgazamiento del anillo neuroretiniano y un aumento difuso o focal de la excavación, desplazamiento y variación del calibre de los vasos retinianos, hemorragias del nervio óptico, defectos de la capa de fibras nerviosas y atrofia peripapilar (22).

2.2. Repercusión del Glaucoma sobre la salud visual

Esta patología es la causa principal de ceguera irreversible a nivel global. En el año 2010 se estimó que 60.5 millones de personas fueron diagnosticadas con glaucoma, en el 2013 el número total de personas con este diagnóstico fue de 64.3 millones y para el año 2020 se estimó que habría un incremento del 18.3% a 76 millones y del 74% a 111.8 millones de personas en el 2040 (13).

Según una revisión realizada por Flaxman y colaboradores en la cual se analizaron 288 estudios con sujetos de 98 países, se observó que 216.6 millones de personas presentaron discapacidad visual moderada a severa en el 2015, siendo el glaucoma responsable de 4 millones de casos en promedio. Así mismo se estimó que para el año 2020, 237 millones de personas presentarían una discapacidad visual moderada a severa y que casi el 2% se debería al glaucoma; finalmente se calculó que 38.5 millones de individuos quedarían ciegos y que 3.2 millones de ellos lo serían debido al glaucoma (16).

De acuerdo a Quigley y colaboradores se estima que en el 2020 habrán 79.6 millones de personas con el diagnóstico de glaucoma y que el 74% corresponderá a un Glaucoma Primario de Angulo Abierto. De igual modo se estima que la ceguera bilateral estará presente en 5.9 millones de pacientes con Glaucoma Primario de Angulo Abierto y en 5.3 millones de individuos con Cierre Angular Primario. Un aspecto interesante es que el 12 % de la población global en el 2010 con el diagnóstico de Glaucoma Primario de Angulo Abierto correspondió a la comunidad latino americana y que para el año 2020 esta proporción ascenderá a 12.9% (23). Tham y colaboradores estimaron que para el año 2040 habrán 111.8 millones de personas con glaucoma y que dicha enfermedad afectará desproporcionadamente a la población asiática (predominando en ella el Glaucoma Primario de Angulo cerrado) y del África (predominando el Glaucoma Primario de Angulo Abierto) (13). Finalmente según la revisión realizada por Bourne y colaboradores desde el año 1990 hasta el 2010, el número de individuos con ceguera e incapacidad visual incrementó en 0.8 millones y 2.3 millones respectivamente (24).

2.3. Métodos de evaluación del Glaucoma

En la evaluación de los pacientes con glaucoma, la medida de la presión intraocular es de suma importancia ya que es uno de los principales factores de riesgo para desarrollar esta enfermedad. Sin embargo hay casos en los cuales esta patología puede aparecer con rangos de presión intraocular normal (25), se ha evidenciado que en el 17% al 50% de los pacientes con glaucoma la presión intraocular inicial puede ser menor a 21 mmHg (26). Es por ello que su medición no predice de manera precisa y consistente en que pacientes se desarrollará el daño del nervio óptico y es que la medida de la presión intraocular puede estar limitada

por algunos factores como la dificultad para detectar fluctuaciones diurnas (27). Leske y colaboradores encontraron que la sensibilidad de la medición de la presión intraocular para detectar glaucoma fue del 97% y la especificidad del 64% (28), quedando sin identificar un gran porcentaje de falsos negativos.

Con respecto al campo visual automatizado, los defectos de dicho examen deben cumplir las siguientes características: asimetría a lo largo de la línea media horizontal (en casos leves a moderados), ubicación en la zona periférica media (en casos leves a moderados), reproducibilidad en al menos dos ocasiones, defectos no explicados por otra enfermedad y que el examen se considere una representación válida del nivel funcional del sujeto (25). Uno de los problemas de este examen es el grado de fluctuación entre las diferentes pruebas, no siendo inusual que el área defectuosa del campo visual de un paciente fluctúe varios decibeles inclusive en pacientes confiables. Por ello es de suma importancia corroborar los cambios del campo visual con la apariencia del nervio óptico (26).

En un estudio realizado por Quigley y colaboradores se hizo una evaluación histológica del número de fibras nerviosas en 5 discos ópticos de sujetos sanos y 3 discos ópticos de sujetos con glaucoma, lo cuales tenían un campo visual normal; obteniendo como resultado un número promedio de 964 mil axones en los discos ópticos sanos en comparación con 744 mil axones presentes en los discos ópticos con daño glaucomatoso (29), evidenciando que hasta un 33% de las fibras nerviosas se pueden perder a pesar de tener un buen campo visual.

Por otro lado el uso de la tomografía de coherencia óptica es de alto costo y se ha reportado que presenta un sensibilidad y especificidad diagnóstica similar a las fotos a color de polo posterior (30).

Una evaluación cuidadosa del nervio óptico usualmente puede identificar el daño glaucomatoso temprano antes de que se produzca una alteración funcional. La apariencia del nervio óptico se puede usar como un monitor sensible de daño, particularmente en los estadios tempranos de la enfermedad (19). Las fotos del nervio óptico han probado ser herramientas muy útiles para su documentación permitiendo una evaluación mas minuciosa y su comparación con otras fotografías pueden ayudar a determinar la progresión del daño glaucomatoso (21)

Se han realizado algunos estudios para determinar la exactitud, sensibilidad y especificidad en la identificación del daño glaucomatoso en fotos a color de polo posterior. La Agencia

para la Investigación de la Asistencia y Calidad Médica evaluó múltiples técnicas de tamizaje incluyendo fotos a color de nervio óptico y encontró una sensibilidad del 65 al 77% y una especificidad del 59 al 98% en la identificación de glaucoma (31). Por otro lado, el Ensayo Europeo para la Evaluación del Disco Óptico encontró una exactitud del 80% en la identificación de discos ópticos sanos y con glaucoma (30).

Por lo anterior, la evaluación del nervio óptico para identificar su daño y extensión es de gran importancia.

Desde 1960, la clasificación cuantitativa más comúnmente usada para determinar el daño del nervio óptico ha sido la proporción excavación/área de disco de Armaly. En ella se describe el diámetro de la copa como un porcentaje del diámetro total del disco. Por tanto esta proporción, especialmente la vertical, representa una medida significativa en la cuantificación de la neuropatía óptica glaucomatosa (3) y es precisamente esta técnica de medición la utilizada en nuestro trabajo.

2.4. Inteligencia Artificial en el campo del Glaucoma

La inteligencia artificial es el desarrollo de sistemas computacionales capaces de realizar tareas imitando la inteligencia humana (10). El aprendizaje automático “le da a la computadora la habilidad de aprender sin ser explícitamente programada”. Usando esta tecnología, el computador es programado para optimizar su desempeño usando datos de “experiencias pasadas” (11). Estos algoritmos pueden identificar patrones de manera autónoma, ajustar su respuesta a los datos y predecir resultados (10).

En las redes neuronales artificiales el procesamiento de datos se basa en el peso de las conexiones neuronales y su entrenamiento depende de la información usada para tal fin: estas contienen múltiples “capas escondidas” entre los nodos de entrada y salida y básicamente existen dos tipos de redes neuronales, la supervisada y la no supervisada, esta última es uno de los aspectos más fascinantes de esta tecnología con la que se analizan grandes bases de datos y se descubren patrones subyacentes sin la necesidad de dar instrucciones. En la aproximación del Aprendizaje Profundo (Deep Learning) se usa una cascada de redes neuronales artificiales de múltiples capas, por tanto este puede ser visto como una optimización de las redes neuronales artificiales convencionales (10).

Una de las áreas de mayor aplicación de aprendizaje profundo es la medicina, donde los datos que deben ser analizados son complejos, el número de pacientes es extenso y los resultados deben ser claros (11).

Dentro de las áreas que pueden hacer uso de esta tecnología está el campo de la retina. Por ejemplo la red de aprendizaje profundo se ha aplicado en un software que analiza tomografías de coherencia óptica de mácula para diferenciar ojos sanos de ojos con degeneración macular relacionada con la edad. En un estudio realizado por Lee y colaboradores se reportó una sensibilidad del 92.64% y una especificidad del 93.69% para detectar esta patología usando una base de datos de 2.6 millones de imágenes de OCT (32). En otro estudio realizado por Gargeya y colaboradores se encontró una sensibilidad del 94% y una especificidad del 98% para detectar retinopatía diabética en 75.137 fotos a color de polo posterior de pacientes diabéticos (33).

El aprendizaje automático también se ha usado en el diagnóstico de catarata y en la optimización de la selección del poder del lente intraocular a implantar, en el análisis de la topografía corneal y en el desarrollo de herramientas diagnósticas para la retinopatía diabética (11)

Otro campo que se beneficia en gran medida del aprendizaje automático es el área de glaucoma. Los sistemas de aprendizaje están siendo utilizados para identificar excavaciones de discos ópticos glaucomatosos en fotos a color de polo posterior y analizar campos visuales y tomografías de coherencia óptica de nervios ópticos de pacientes con esta patología. Christopher y colaboradores realizaron un estudio para determinar la capacidad de un sistema supervisado de aprendizaje profundo para identificar nervios ópticos glaucomatosos en 14.822 fotos a color de polo posterior, obteniendo una sensibilidad del 88% y una especificidad del 95% para detectar daño moderado a severo en el nervio óptico (34). En otro estudio realizado por Medeiros y colaboradores, se cuantificó el daño glaucomatoso en 32.820 fotos a color de polo posterior usando datos cuantitativos del OCT Spectral Domain (SD OCT) para entrenar un algoritmo de aprendizaje profundo, con una predicción del grosor promedio de la capa de fibras nerviosas de las fotos de 83.3 ± 14.5 μm y grosor promedio de la capa de fibras nerviosas del SD OCT de 82.5 ± 16.8 , con lo cual se obtuvo un fuerte correlación entre ambos valores (35)

Por todo lo ya explicado, es evidente que los sistemas de aprendizaje profundo están avanzando a grandes pasos en el área de la medicina, ayudando en la clasificación, segmentación y detección objetiva de las imágenes oftalmológicas y proyectándose como una herramienta muy útil en el tamizaje con el fin de brindar apoyo en la prevención de patologías de alto costo social y económico.

3. Pregunta de investigación

¿Cuál es la concordancia entre la interpretación de las fotos a color de polo posterior y un algoritmo de aprendizaje no supervisado para determinar el daño del nervio óptico?

4. Objetivos

4.1. Objetivo general

Determinar la concordancia entre la interpretación de fotos a color de polo posterior y un algoritmo de aprendizaje no supervisado para determinar el daño del nervio óptico.

4.2. Objetivos específicos

4.2.1. Determinar la concordancia entre la interpretación de fotos a color de polo posterior y un algoritmo de aprendizaje no supervisado para clasificar los nervios ópticos según el grado de daño: sin daño, con daño leve, moderado o severo.

4.2.2. Determinar la concordancia entre la interpretación de fotos a color de polo posterior y un algoritmo de aprendizaje no supervisado para clasificar los nervios ópticos como sin daño (sanos) y con daño.

5. Metodología

5.1. Tipo y diseño de estudio:

Se realizó un estudio de concordancia entre un algoritmo de aprendizaje no supervisado y un especialista en glaucoma.

Se desarrolló una arquitectura que combina internamente dos modelos. El primer modelo es conocido como una red neuronal convolucional (CNN) llamada VGG con 16 capas de profundidad que es pre entrenada con un conjunto de imágenes de dominio natural (no médicas) de *ImageNet* ([http:// www.image-net.org/](http://www.image-net.org/)) con el fin de extraer las características del nervio óptico. Específicamente la capa 15 tiene una longitud de 4096 valores entre los que se encuentran: bordes, formas, texturas, objetos y clases. Todo ello con el objetivo de determinar si un nervio óptico no presenta daño (sano) o presenta daño leve, moderado o severo (36)

El segundo modelo esta compuesto por un algoritmo de agrupamiento llamado K-means que agrupa las características de los 4096 valores según la mayor similitud entre éstas y separa las que presentan la mayor diferencia.

Posteriormente se programó un modelo para que determine 4 grupos de acuerdo a la clasificación de daño asignada por el especialista en glaucoma.

5.2. Población

- *Población de referencia*: Fotos a color de polo posterior.
- *Población objetivo*: Fotos a color de polo posterior de 20 grados obtenidas de la base de datos de la Fundación Oftalmológica Nacional y de las bases de datos públicas disponibles on line de imágenes retinianas para la evaluación del nervio óptico Drishti - GS y RIM – ONE.

5.3. *Tamaño de la muestra*

- Se realizó un muestreo no probabilístico por conveniencia, seleccionando de forma consecutiva, fotos a color de polo posterior de la base de datos de la Fundación Oftalmológica Nacional, las cuales fueron tomadas con el Angiógrafo Zeiss FF 450 plus IR y de las bases de datos públicas disponibles on line de imágenes retinianas para la evaluación del nervio óptico Drishti – GS (37) y RIM – ONE (38), las cuales fueron facilitadas por un investigador asociado (Ingeniero Oscar Perdomo, PhD) y que proveen de imágenes de polo posterior tanto del nervio óptico como de la mácula. Se seleccionaron las fotos a color de polo posterior de 20 grados con evidencia de la totalidad del nervio óptico y se descartaron aquellas que presentaban una mala calidad de la imagen, obteniendo 77 fotos a color de polo posterior de la base de datos de la Fundación Oftalmológica Nacional y 612 de las bases de datos públicas con un total de 689 fotos.

5.4. *Criterios de selección*

5.4.1. *Criterios de inclusión*

Fotos a color de polo posterior de 20 grados, con evidencia de la totalidad del nervio óptico en la imagen.

5.4.2. *Criterios de exclusión*

Inadecuada calidad de la imagen.

5.5. *Variables*

Tabla 1. Definición de variables

VARIABLE	DEFINICIÓN OPERATIVA	TIPO DE VARIABLE	ESCALA
-----------------	-----------------------------	-------------------------	---------------

Presencia de daño del nervio óptico detectado por un especialista en glaucoma	Proporción excavación/disco aumentada	Cualitativa nominal dicotómica	0: Sin daño 1: Daño
Presencia de daño del nervio óptico detectado por un algoritmo de aprendizaje no supervisado	Proporción excavación/disco aumentada	Cualitativa nominal dicotómica	0: Sin daño 1: Daño
Clasificación de daño del nervio óptico por un especialista en glaucoma	Resultado de la prueba realizada por el especialista	Ordinal	0. Sin daño 1: Daño leve 2: Daño moderado 3: Daño severo
Clasificación de daño del nervio óptico por algoritmo de aprendizaje no supervisado	Resultado de la prueba realizada por el especialista	Ordinal	0. Sin daño 1: Daño leve 2: Daño moderado 3: Daño severo

5.6. *Plan de análisis*

Se determinó la concordancia entre el Algoritmo de aprendizaje no supervisado y el especialista en glaucoma con respecto a la identificación de los nervios ópticos sin daño (sanos) y con daño y en el reconocimiento del daño leve, moderado y severo usando el coeficiente Kappa de Cohen.

5.7. Control de sesgo y error

La evaluación de la base de datos por parte del sistema de aprendizaje no supervisado fue llevado a cabo por una sola persona, con el fin de evitar su manipulación y variabilidad en el procesamiento de la información.

5.8. Proceso de recolección de la información

Los datos obtenidos por los investigadores, tomados de las fotos a color de polo posterior, se registraron en una base de datos electrónica en Excel que incluía las variables de nervio óptico sin daño y con daño leve, moderado y severo.

La recolección de datos estuvo a cargo del especialista en glaucoma. Se verificó que todas las imágenes cumplieran con los criterios de inclusión.

6. Aspectos éticos

El estudio se realizó dentro de los principios éticos para las investigaciones médicas en seres humanos según la Declaración de Helsinki - 59^a Asamblea General, Seúl, Corea, Octubre 2008 (39)

Se tuvo 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” (40)

La presente investigación es clasificada dentro de la categoría 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.

Se limitará el acceso de los instrumentos de investigación únicamente a los investigadores según Artículo 8 de la Resolución 008430 de 1993 del Ministerio de Salud.

Será responsabilidad de los investigadores el guardar con absoluta reserva la información contenida en las historias clínicas y a cumplir con la normatividad vigente en cuanto al manejo de la misma reglamentados en los siguientes: Ley 100 de 1993, Ley 23 de 1981, Decreto 3380 de 1981, Resolución 008430 de 1993 y Decreto 1995 de 1999.

Todos los integrantes del grupo de investigación 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 mantendrá absoluta confidencialidad y se preservará el buen nombre institucional profesional.

El estudio se realizará con un manejo estadístico imparcial y responsable.

No existe ningún conflicto de interés por parte de los autores del estudio que deba declararse.

7. Administración del proyecto

7.1. Cronograma

ETAPAS Y ACTIVIDADES													
	1	2	3	4	5	6	7	8	9	10	11	12	
Preparación y aprobación del proyecto	X												
Sometimiento y aprobación CEI	X												
Análisis de la base de datos							X						
Análisis de resultados							X						

Informe de resultados y entrega final de investigación										X
--	--	--	--	--	--	--	--	--	--	---

7.2. Presupuesto

	VALOR EN PESOS COLOMBIANOS
<i>Comité de ética</i>	\$3.800.000
<i>Análisis de Datos</i>	\$2.000.000
<i>Gastos Administrativos</i>	\$1.000.000
<i>Publicación</i>	\$1.500.000
TOTAL	\$8.300.000

8. Resultados

El algoritmo de aprendizaje no supervisado evaluó 612 fotos a color de polo posterior de las bases de datos públicas Drishti – GS y RIM – ONE y 77 fotos a color de polo posterior de la base de datos de la Fundación Oftalmológica Nacional, las cuales fueron clasificadas como con nervio óptico sin daño (sano) y con daño leve, moderado y severo según el especialista en glaucoma.

Posteriormente un clasificador K – means agrupó las características extraídas de los cuatro grupos mencionados y al comparar la interpretación realizada por el especialista y la propuesta por el modelo, se observó una concordancia en la identificación de 212 nervios ópticos sin daño, 10 nervios ópticos con daño leve, 21 nervios ópticos con daño moderado y 17 nervios ópticos con daño severo.

Los resultados se describen en la Tabla 1.

TABLA 1. Matrix de Confusión de cuatro clases

Observación del Especialista						
Observación del Modelo		Sin daño	Daño leve	Moderado	Severo	Total
	Sin daño	212	53	127	104	496
	Daño leve	24	10	12	20	66
	Moderado	25	9	21	15	70
	Severo	25	4	11	17	57
	Total	286	76	171	156	689

Al determinar el rendimiento del algoritmo en la clasificación de estas 4 clases: nervio óptico sin daño, con daño leve, moderado y severo, se obtuvo una exactitud de 37.73 % y al comparar la interpretación del especialista y del algoritmo se obtuvo una concordancia medida con el coeficiente kappa de Cohen de 0.037.

Por otro lado al analizar la concordancia para dos clases: nervios ópticos sin daño (sanos) y nervios ópticos con daño (leve, moderado y severo), se observó que el algoritmo fué capaz de identificar 212 nervios ópticos sin daño y 119 nervios ópticos con daño. Obteniendo una exactitud de 48.04% y un estadístico kappa para la clasificación dicotómica de 0.03.

Los resultados se describen en la Tabla 2.

TABLA 2. Matrix de Confusión de dos clases (con daño y sin daño)

Observación del Especialista				
Observación del Modelo		Sin daño	Con daño	Total
	Sin daño	212	284	496
	Con daño	74	119	193
	Total	286	403	689

9. Discusión

En el presente estudio se desarrolló una arquitectura que combinó internamente dos modelos, un primer modelo de red neuronal convolucional llamado VGG, entrenado con un conjunto de imágenes de dominio natural con el fin de extraer las características del nervio óptico y un segundo modelo compuesto por un algoritmo de agrupamiento llamado K-means, que agrupó estos caracteres según su mayor similitud para clasificar el nervio óptico como sano, con daño leve, moderado o severo.

Estudios previos ya han descrito el uso de la inteligencia artificial en la valoración de fotos a color de polo posterior para determinar el daño del nervio óptico (43). Ting y col. propusieron un algoritmo de aprendizaje profundo para realizar el tamizaje de glaucoma en 125,189 fotos a color de polo posterior de pacientes diabéticos, las cuales fueron clasificadas subjetivamente por los oftalmólogos como “referibles” durante la fase de entrenamiento, obteniendo en la fase de prueba un área bajo la curva (AUC) de 0.942 con una sensibilidad del 96.4% y especificidad del 87.2% (41). En otro estudio Li y col. clasificaron también de manera subjetiva 48,116 fotos a color de polo posterior como “referibles o no” para glaucoma, obteniendo un área bajo la curva, sensibilidad y especificidad de 0.986, 95.6% y 92% respectivamente (42). Medeiros y col. desarrollaron un sistema más complejo de identificación de daño del nervio óptico, en el que un algoritmo de aprendizaje profundo fue entrenado con fotos a color de polo posterior que fueron clasificadas de acuerdo al grosor de la capa de fibras nerviosas medida por un tomógrafo de coherencia óptica Spectral Domain (SD-OCT) obteniendo una buena área bajo la curva para discriminar nervios ópticos sanos de los glaucomatosos (35). En estos trabajos se puede usar el aprendizaje profundo supervisado como herramienta para la clasificación de daño de esta estructura o algoritmos de aprendizaje no supervisado como el presente estudio. Al realizar una búsqueda sistemática de estos trabajos se encontró que su uso es menos frecuente que el aprendizaje profundo supervisado en el que se desarrolla una etapa de entrenamiento y otra de prueba y que sobre todo estos intentan identificar patrones de daño en el campo visual o en la tomografía de coherencia óptica de nervio óptico (44,45). En este estudio se realizó una clasificación subjetiva del daño del nervio óptico por parte de un especialista en glaucoma y

posteriormente se analizó la concordancia con un algoritmo de aprendizaje no supervisado, obteniendo un coeficiente de Kappa de Cohen de 0.037 al clasificar los nervios ópticos como sin daño, con daño leve, moderado y severo. Así mismo, se obtuvo un estadístico Kappa para la clasificación dicotómica de nervios ópticos con y sin daño de 0.03, mostrando una mala concordancia en ambos casos. También se analizó el rendimiento del algoritmo en la clasificación de las 4 clases (nervio óptico sin daño, con daño leve, moderado y severo) y la dicotómica (nervio óptico sin y con daño) obteniendo una exactitud de 37.73 % y 48.04% respectivamente, con lo cual se evidenció un bajo rendimiento del algoritmo. Esto podría explicarse por la limitada longitud de valores (4096) correspondientes a la capa 15 de la red convolucional VGG, la cual permite el agrupamiento de las características con mayor similitud y la exclusión de las que presentan mayores diferencias y con la que finalmente se identifican los nervios ópticos sin daño y con daño leve, moderado y severo. Por otra parte también se ha descrito que los algoritmos de aprendizaje no supervisado requieren de grandes bases de datos para poder identificar patrones de daño subyacentes y que la información suministrada debe ser variada en sus clases (10); en este trabajo se usaron 689 fotos a color de polo posterior y el número de fotos correspondiente a cada clase no fué uniforme, presentando la mayoría de los nervios ópticos un daño moderado, con lo cual podría explicarse también la inadecuada concordancia entre el especialista en glaucoma y el algoritmo.

Dentro de las limitaciones en este estudio se puede citar el limitado número de fotos a color de polo posterior que fueron analizadas y que de acuerdo a lo anteriormente descrito podría haber dificultado la identificación de cada clase (nervio óptico sin daño, con daño leve, moderado y severo) por parte del algoritmo de aprendizaje no supervisado.

Se recomienda ampliar el número de la muestra de fotos a color de polo posterior con el fin de identificar todos los patrones de daño y obtener clases más uniformes.

En conclusión, el algoritmo de aprendizaje no supervisado no presentó una adecuada concordancia en la clasificación del daño del nervio óptico en las fotos a color de polo posterior. Es posible que se requiera una mejor selección de las variables de entrenamiento para lograr el desarrollo de un algoritmo más eficiente. Actualmente los sistemas de inteligencia artificial están dando buenos resultados con respecto a clasificación,

segmentación y detección objetiva de las imágenes oftalmológicas, proyectándose como una herramienta muy útil en el tamizaje de patologías como el glaucoma.

10. Referencias

1. Shaikh Y, Yu F, Coleman AL. Burden of undetected and untreated glaucoma in the United States. *Am J Ophthalmol* [Internet]. 2014;158(6):1121-1129.e1. Available from: <http://dx.doi.org/10.1016/j.ajo.2014.08.023>
2. 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. Available from: <http://dx.doi.org/10.1016/j.ophtha.2014.05.013>
3. Henderer JD. Disc damage likelihood scale. *Br J Ophthalmol*. 2006;90(4):395–6.
4. Shickle D, Todkill D, Chisholm C, Rughani S, Griffin M, Cassels-Brown A, et al. Addressing inequalities in eye health with subsidies and increased fees for General Ophthalmic Services in socio-economically deprived communities: A sensitivity analysis. *Public Health*. 2015;129(2):131–7.
5. Hattenhauer MG, Johnson DH, Ing HH, Herman DC, Hodge DO, Yawn BP, et al. The probability of blindness from open-angle glaucoma. *Ophthalmology*. 1998;105(11):2099–104.
6. Stevens GA, White RA, Flaxman SR, Price H, Jonas JB, Keeffe J, et al. Global prevalence of vision impairment and blindness: Magnitude and temporal trends, 1990-2010. *Ophthalmology*. 2013;120(12):2377–84.
7. Varma R, Ying-Lai M, Francis BA, Nguyen BBT, Deneen J, Wilson MR, et al. Prevalence of open-angle glaucoma and ocular hypertension in Latinos: The Los Angeles Latino Eye Study. *Ophthalmology*. 2004;111(8):1439–48.
8. Quigley HA, West SK, Rodriguez J, Munoz B, Klein R, Snyder R. The prevalence of glaucoma in a population-based study of Hispanic subjects: Proyecto VER. *Arch Ophthalmol*. 2001;119(12):1819–26.

9. Caprioli J. Clinical evaluation of the optic nerve in glaucoma. *Trans Am Ophthalmol Soc.* 1994;92:589–641.
10. Gordon MO, Torri V, Miglior S, Beiser JA, Floriani I, Miller JP, et al. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology.* 2007;114(1):10-19.e2.
11. Myers JS, Fudemberg S LD. Evolution of optic nerve photography for glaucoma screening: a review. *Clin Exp Ophthalmol.* 2018;13(3):287–8.
12. Chakrabarty L, Joshi GD, Chakravarty A, Raman G V., Krishnadas SR, Sivaswamy J. Automated Detection of Glaucoma from Topographic Features of the Optic Nerve Head in Color Fundus Photographs. *J Glaucoma.* 2016;25(7):590–7.
13. Cook C, Foster P. Epidemiology of glaucoma: What’s new? *Can J Ophthalmol.* 2012;47(3):223–6.
14. Bock R, Meier J, Nyúl LG, Hornegger J, Michelson G. Glaucoma risk index: Automated glaucoma detection from color fundus images. *Med Image Anal.* 2010;14(3):471–81.
15. Spaeth GL, Reddy SC. Imaging of the optic disk in caring for patients with glaucoma: Ophthalmoscopy and photography remain the gold standard. *Surv Ophthalmol [Internet].* 2014;59(4):454–8. Available from: <http://dx.doi.org/10.1016/j.survophthal.2013.10.004>
16. Hasanreisoglu M, Priel E, Naveh L, Lusky M, Weinberger D, Benjamini Y, et al. Screening for glaucoma with stereo disc photopgraphy. *J Glaucoma.* 1995;22(3):238–42.
17. Vijaya L, George R, Paul PG, Baskaran M, Arvind H, Raju P, et al. Prevalence of open-angle glaucoma in a rural south Indian population. *Investig Ophthalmol Vis Sci.* 2005;46(12):4461–7.
18. Ramakrishnan R, Nirmalan PK, Krishnadas R, Thulasiraj RD, Tielsch JM, Katz J, et al. Glaucoma in a rural population of Southern India: The Aravind Comprehensive Eye Survey. *Ophthalmology.* 2003;110(8):1484–90.
19. Haleem MS, Han L, van Hemert J, Li B. Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review. *Comput Med Imaging Graph [Internet].* 2013;37(7–8):581–96. Available from:

<http://dx.doi.org/10.1016/j.compmedimag.2013.09.005>

20. Rahimy E. Deep learning applications in ophthalmology. *Curr Opin Ophthalmol*. 2018;29(3):254–60.
21. Lee A, Taylor P, Kalpathy-Cramer J. Machine Learning Has Arrived! *Ophthalmology*. 2017;124(12):1726–8.
22. Kotowski J, Wollstein G IH. Imaging of the optic nerve and retinal nerve fiber layer: an essential part of glaucoma diagnosis and monitoring. *Surv Ophthalmol*. 2014;23(1):1–7.
23. Quigley H, Broman AT. The number of people with glaucoma worldwide in 2010 and 2020. *Br J Ophthalmol*. 2006;90(3):262–7.
24. Bourne RRA, Taylor HR, Flaxman SR, Keeffe J, Leasher J, Naidoo K, et al. Number of people blind or visually impaired by glaucoma worldwide and in world regions 1990 - 2010: A meta-analysis. *PLoS One*. 2016;11(10):1–16.
25. Foster PJ, Buhrmann R, Quigley HA, Johnson GJ. The definition and classification of glaucoma in prevalence surveys. *Br J Ophthalmol*. 2002;86(2):238–42.
26. Caprioli J, Prum B, Zeyen T. Comparison of methods to evaluate the optic nerve head and nerve fiber layer for glaucomatous change. *Am J Ophthalmol*. 1996;121(6):659–67.
27. Liu JHK, Zhang X, Kripke DF, Weinreb RN. Twenty-four-hour intraocular pressure pattern associated with early glaucomatous changes. *Investig Ophthalmol Vis Sci*. 2003;44(4):1586–90.
28. Leske MC. Distribution of Intraocular Pressure. *Arch Ophthalmol*. 1997;115(8):1051.
29. Wall M, Stanek KE, Chauhan BC. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Investig Ophthalmol Vis Sci*. 1996;37(3).
30. Reus NJ, Lemij HG, Garway-Heath DF, Airaksinen PJ, Anton A, Bron AM, et al. Clinical Assessment of Stereoscopic Optic Disc Photographs for Glaucoma: The European Optic Disc Assessment Trial. *Ophthalmology* [Internet]. 2010;117(4):717–23. Available from: <http://dx.doi.org/10.1016/j.ophtha.2009.09.026>
31. Ervin A, Boland M, Myrowitz E, Prince J, Hawkins B, Vollenweider D, et al.

Treatment for Glaucoma: Comparative Effectiveness. Comparative Effectiveness Review. AHRQ Publ No 12-EHC038-EF [Internet]. 2012;Review No.(60):443.

Available from:

www.effectivehealthcare.ahrq.gov/reports/final.cfm.%5Cnwww.effectivehealthcare.ahrq.gov/reports/final.cfm

32. Lee CS, Baughman DM, Lee AY. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. *Kidney Int Reports* [Internet]. 2017;1(4):322–7. Available from: <http://dx.doi.org/10.1016/j.oret.2016.12.009>
33. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* [Internet]. 2017;124(7):962–9. Available from: <http://dx.doi.org/10.1016/j.ophtha.2017.02.008>
34. Christopher M, Belghith A, Bowd C, Proudfoot JA, Goldbaum MH, Weinreb RN, et al. Performance of Deep Learning Architectures and Transfer Learning for Detecting Glaucomatous Optic Neuropathy in Fundus Photographs. *Sci Rep*. 2018;8(1):1–13.
35. Medeiros FA, Jammal AA, Thompson AC. From Machine to Machine: An OCT-Trained Deep Learning Algorithm for Objective Quantification of Glaucomatous Damage in Fundus Photographs. *Ophthalmology* [Internet]. 2019;126(4):513–21. Available from: <https://doi.org/10.1016/j.ophtha.2018.12.033>
36. Simonyan K ZA. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Int Conf Learn Represent*. 2015;75(6):398–406.
37. Sivaswamy J, Gopal SRK, Joshi D, Jain M, Syed U, Hospital AE. DRISHTI-GS : Retinal image dataset for optic nerve head segmentation. 2014;53–6.
38. Fumero F, Alayon S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. RIM-ONE: An open retinal image database for optic nerve evaluation. *Proc - IEEE Symp Comput Med Syst*. 2011;2–7.
39. Association WM. WorldMedical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects. *JAMA - J Am Med Assoc*. 2013;310.
40. Colombia R de. Ministerio de Salud. Resolucion Numero 8430 de 1993. RN 008430. 2012;32(4):471–3.

41. 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 - J Am Med Assoc.* 2017;318(22):2211–23.
42. 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* [Internet]. 2018;125(8):1199–206. Available from: <https://doi.org/10.1016/j.ophtha.2018.01.023>
43. Phene S, Dunn RC, Hammel N, Liu Y, Krause J, Kitade N, et al. Deep Learning and Glaucoma Specialists: The Relative Importance of Optic Disc Features to Predict Glaucoma Referral in Fundus Photographs. *Ophthalmology* [Internet]. 2019;126(12):1627–39. Available from: <https://doi.org/10.1016/j.ophtha.2019.07.024>
44. Christopher M, Belghith A, Weinreb RN, Bowd C, Goldbaum MH, Saunders LJ, et al. Retinal nerve fiber layer features identified by unsupervised machine learning on optical coherence tomography scans predict glaucoma progression. *Investig Ophthalmol Vis Sci.* 2018;59(7):2748–56.
45. Goldbaum MH, Sample PA, Zhang Z, Chan K, Lee T, Boden C, et al. Using Unsupervised Learning with Independent Component Analysis to Identify Patterns of Glaucomatous Visual Field Defects. *Invest Ophthalmol.* 2005;46(10):3676–83.