

UNIVERSIDAD PRIVADA ANTENOR ORREGO
FACULTAD DE MEDICINA HUMANA
PROGRAMA DE ESTUDIOS DE MEDICINA HUMANA



TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE MEDICO CIRUJANO

“ALGORITMO DEL SISTEMA DE APRENDIZAJE PROFUNDO USANDO
RETINOGRAFÍA COMO MÉTODO DE DIAGNÓSTICO PRECÓZ DE NEUROPATÍA
ÓPTICA GLAUCOMATOSA BASADO EN LA RELACIÓN COPA/DISCO Y LA REGLA
ISNT”

Área de investigación:
Enfermedades no transmisibles

Autora :
Br. Castillo Castillo, Diana Edith

Jurado Evaluador:

Presidente: Pomatanta Plasencia, Jose Luis

Secretario: Hashimoto Carrasco, Victor Humberto

Vocal: Cisneros Gómez, Carlos Augusto

Asesor:
Barba Chirinos, Carlos Enrique
Código Orcid: <https://orcid.org/0000-0003-0583-199X>

Trujillo – Perú
2018

Fecha de sustentación: 2022/11/07

DEDICATORIA

Dedico el presente trabajo de investigación a mis padres y hermanos, por brindarme su apoyo incondicional, amor y fe en que sí podía lograr esta importante meta en mi vida profesional y personal.

AGRADECIMIENTOS

Agradezco a Dios por su protección, amor y
por haberme permitido cumplir esta meta
profesional.

A la Universidad Privada Antenor Orrego, mi
casa de estudios y a mis maestros por los
años de aprendizaje y experiencias vividas.

A mi asesor, el Dr. Carlos Enrique Barba Chirinos
por su apoyo en todo momento de
la realización de mi trabajo de investigación,
sin ello no hubiera logrado esto.

TABLA DE CONTENIDOS

RESUMEN.....	5
I. INTRODUCCIÓN.....	7
II. MATERIALES Y MÉTODOS.....	12
III. RESULTADOS.	15
IV. DISCUSIÓN.....	17
V. CONCLUSIONES.	22
VI. REFERENCIAS BIBLIOGRÁFICAS.....	23
VII. TABLAS Y GRÁFICOS.....	29
VIII. ANEXOS.....	35

RESUMEN

Introducción:

La neuropatía óptica glaucomatosa es considerada mundialmente una de las primeras causas de ceguera, y la primera causa de forma irreversible, ésta se produce por daño del nervio óptico del ojo.

Objetivo:

Determinar la precisión diagnóstica del algoritmo del Sistema de Aprendizaje Profundo (Deep Learning) para detectar la neuropatía óptica glaucomatosa utilizando la retinografía.

Métodos:

Se realizó una revisión sistemática y metaanálisis, en las siguientes bases de datos: Pubmed, Scopus, Web of Science, Ovid-Medline, Ovid-Embase. Se aplicaron los criterios de selección en dos fases respectivamente: primero por título y resumen y luego a texto completo. Luego de esto, se extrajeron las características más relevantes de cada estudio y se realizó el análisis de sesgo utilizando el QUADAS-2, para finalmente meta analizar los resultados usando el software Stata v14.

Resultados:

Se obtuvieron 300 resultados al realizar las búsquedas en las 5 bases de datos, incluyéndose finalmente 8. En el análisis de sesgo, el dominio más afectado en la mayoría de estudios, fue el de la selección de pacientes; seguido por el de la prueba índice. Finalmente, al combinar los resultados se obtuvo una sensibilidad de 0.90 con un IC al 95% que va de 0.86 a 0.94 y una especificidad que va de 0.93 con un IC al 95% que va de 0.90 a 0.96.

Conclusiones:

El uso del sistema de aprendizaje profundo para la detección de GON ofrece una sensibilidad y especificidad aceptable respecto a la capacidad de detección de oftalmólogos especializados.

Palabras clave: aprendizaje profundo, glaucoma, sensibilidad, especificidad.

ABSTRACT

Introduction:

Glaucomatous optic neuropathy is considered one of the leading causes of blindness worldwide, and the leading cause irreversibly, it is caused by damage to the optic nerve of the eye.

Objective:

To determine the diagnostic accuracy of the Deep Learning System algorithm to detect glaucomatous optic neuropathy using retinography.

Methods:

A systematic review and meta-analysis was carried out in the following databases: Pubmed, Scopus, Web of Science, Ovid-Medline, Ovid-Embase. The selection criteria were applied in two phases respectively: first by title and abstract and then by full text. After this, the most relevant characteristics of each study were extracted and the bias analysis was performed using QUADAS-2, to finally meta-analyze the results using Stata v14 software.

Results:

300 results were obtained when searching the 5 databases, finally including 8. In the bias analysis, the most affected domain in most studies was patient selection; followed by that of the index test. Finally, when combining the results, a sensitivity of 0.90 with a 95% CI ranging from 0.86 to 0.94 and a specificity ranging from 0.93 with a 95% CI ranging from 0.90 to 0.96 were obtained.

Conclusions:

The use of deep learning systems for the detection of GON offers acceptable sensitivity and specificity with respect to detection capacity of specialized ophthalmologists.

Keywords: Deep learning, glaucoma, sensitivity, specificity.

I. INTRODUCCIÓN

Glaucoma es un conjunto de neuropatías ópticas progresivas e irreversibles que provocan daño estructural en la cabeza del nervio óptico, conduciendo a una merma severa del campo de visualización y por último a ceguera, por ello se le considera mundialmente la primera causa de ceguera irreversible. En el 2014, la prevalencia en todo el mundo fue de 64,3 millones en las personas entre 40 y 80 años y un estudio predijo que, para el 2020, 79,6 millones de personas padecerían de glaucoma, que 11,1 millones de personas quedarían ciegas por esta enfermedad, y que para el 2040 aumentará a 111,8 millones de casos de glaucoma. (1,2)

Los factores de riesgo implicados son; la edad, género, etnia, antecedentes familiares con neuropatía óptica glaucomatosa, diabéticos, hipertensos, miopía alta y lesión ocular.(3–6)

La neuropatía óptica glaucomatosa (NOG) produce daño al nervio óptico por incremento de la presión intraocular (PIO), esto puede deberse al acúmulo del humor acuoso por una falla en el sistema de drenaje a través de la malla trabecular.(7) Una de las causas es el mecanismo isquémico o compresivo, que al incrementar la presión interna del ojo genera degeneración progresiva de la cabeza del nervio óptico con pérdida de las células ganglionares de la retina, adelgazamiento de la capa de las fibras neuroretinianas y aumento de la excavación papilar en el disco óptico. En un estudio se realizó tensiones equivalentes de Von Mises y se observó que los tejidos más afectados por el aumento de la presión intraocular son la lámina cribosa y la esclerótica, jugándose un papel muy importante en la biomecánica de esta enfermedad. (8) El espesor corneal central se correlaciona con la histéresis corneal (CH) y ambos son indicadores de la progresión del glaucoma ya que muestran asociación con la deformación del nervio óptico por aumento de la PIO, por ello la disminución de la CH se asocia con la enfermedad, mientras que su elevación indicará bajo riesgo de glaucoma.(9,10)

La inteligencia artificial es una ciencia que simula el comportamiento

de las computadoras, que tiene como subconjunto al aprendizaje automático; que se encarga de configurar algoritmos informáticos para reconocer patrones en los datos utilizando técnicas tradicionales, sin la necesidad de que dichos datos o reconocimiento sea dictado por un humano. A uno de esos algoritmos que utilizan el aprendizaje de representación pertenece el aprendizaje profundo (Deep Learning), en dónde aprenden representaciones o características de los datos automáticamente, siempre que se les brinden suficientes datos. El aprendizaje profundo (AP) es un tipo de red neuronal artificial conformada por varias capas de neuronas artificiales interconectadas, en dónde los datos se alimentan a la red utilizando datos mucho más complejos dónde se obtiene un mejor rendimiento, con ello observando y haciendo modificaciones al modelo y repitiendo ese proceso con el fin de lograr el resultado deseado. Existen 3 tipos de aprendizaje; el que es supervisado, el no supervisado y el semi supervisado, sin embargo todos requieren de un requisito importante que es un gran conjunto de datos para el entrenamiento, pudiendo ser miles o millones de retinografías.(11) La gran mayoría de casos llegan a consulta en etapas avanzadas como consecuencia de la sobredemanda de pacientes y el bajo número de especialistas, aumentando así la incidencia anualmente, es por ello que con el paso del tiempo se han venido creando estos algoritmos gracias a la inteligencia artificial para ayudar a disminuirla no solo a través de un diagnóstico precoz, sino también para el seguimiento de esta enfermedad y a la vez poder disminuir la carga de consultas con el especialista en glaucoma, descartando casos poco probables, ya que este algoritmo estará diseñado gracias al aprendizaje profundo para que las retinografías pasen por un proceso y con ello se puedan clasificar solo a los pacientes susceptibles a glaucoma para que sean derivados con el especialista. Para diagnosticar esta enfermedad, se requieren exámenes costosos y sobre todo tiempo tanto del paciente como del especialista. Los pacientes con glaucoma pueden acelerar la pérdida gradual del nervio óptico, por lo que su evaluación de los cambios tempranos en imágenes cualitativas (retinografía) y cuantitativas

(Tomografía de coherencia óptica (TCO)) pueden ayudar a prevenir cambios en el campo visual. (12)

Las retinografías son imágenes proyectivas y a color de la superficie interna del ojo y es una técnica no invasiva, sencilla de realizar y la más económica. Cuenta con una clasificación binaria, de un fondo de ojo normal y otro referenciable y en el algoritmo basado en el procesamiento de las retinografías; se analizarán las estructuras con apariencia de sospecha de glaucoma, presencia de atrofia peripapilar y apariencia ahuecada o "ventosas" de la cabeza del nervio óptico asociado a merma de axones de las correspondientes células ganglionares de la retina, teniendo en cuenta que la academia americana de oftalmología define al sospechoso de glaucoma a aquél que presente factores de riesgo o características clínicas como: 1) Aspecto sospechoso de la cabeza del nervio óptico y capas de fibras nerviosas, tales como el acrecentamiento o la asimetría de la relación copa/disco (CDR) en eje vertical, afinamiento del anillo neuroretiniano, daño de la capa de las fibras nerviosas y hemorragia del disco (en astilla) 2) Longitud visual sospechos de perjuicio glaucomatoso sin la existencia de señales clínicas o de otras afecciones, como; escalón nasal, escotoma paracentral, defecto arqueado, altitudinal y aumento de la desviación media 3) Campo visual, fibras nerviosas y cabeza del nervio óptico apropiados consistente con hipertensión ocular (13).

La evaluación a través de la retinografía de debe hacer minuciosamente abordando en el tamaño de la copa y excentricidad, con ello la relación copa/disco, también el color del borde neuroretiniano y los cambios estructurales antes mencionados, que por medio de una serie de procesos nos permitirá con el algoritmo, poder llegar un método de diagnóstico más preciso. (14). A través de la retinografía también se puede evaluar el grosor del anillo neuroretiniano, que, en una persona sana, éste va a ser más grueso en la región inferior del nervio, seguido por la superior, nasal y temporal, esta secuencia se le va a denominar la regla ISNT, al no cumplir esta regla se determinará riesgo de glaucoma y ésta también sirve para determinar la progresión de la enfermedad.(15)

Para que este sistema sea exitoso, se recopilaron datos en la historia clínica y examen clínico de imágenes de fondo de ojo glaucomatoso y no glaucomatoso en ambos ojos, el cual es capacitado con el objetivo de minimizar la pérdida de probabilidades en las muestras de entrenamiento y así posteriormente aplicarlo a conjuntos de datos de validación. Otra forma de obtener las muestras fue a través de la recolección manual. La investigación y capacitación sobre el sistema de aprendizaje profundo en el fondo de ojo, es un método con una sensibilidad del 95.6% y especificidad del 92%., que a diferencia de muchos exámenes métricos convencionales existentes, también nos permite obtener mejores resultados minimizando costos, sin embargo este sistema también tiene controversias en saber cuál precisión diagnóstica es mejor, ya que, encontramos que otros estudios como ESPERANZA, DRISHTI-GS, VGG19TL (S:87.01% y E:89.01%), PEGASUS tuvieron mayor sensibilidad y especificidad por parte del software, además de una lectura de imagen <10 segundos. (16–20) y estudios que mostraron que el oftalmólogo tiene misma sensibilidad y mayor especificidad que el programa, teniendo mayor incidencia de falsos positivos y falsos negativos por miopía.(21,22)

El diagnóstico por aprendizaje profundo, suele ser muy beneficioso para la detección temprana de la neuropatía óptica glaucomatosa, ya que la escasez de profesionales oftalmólogos capacitados, la alta demanda de pacientes y el alto costo de los exámenes, ha conllevado a un incremento de ceguera por no detectarse a tiempo (23,24) por lo que este trabajo resaltaría la importancia del sistema de aprendizaje profundo para detectar esta enfermedad. Por lo tanto, el objetivo de este estudio es la precisión diagnóstica del algoritmo del sistema de aprendizaje profundo para detectar neuropatía óptica glaucomatosa utilizando la retinografía.

Enunciado del problema:

¿Cuál es la precisión diagnóstica del algoritmo del sistema de aprendizaje profundo para detectar neuropatía óptica glaucomatosa utilizando la retinografía?

Objetivos:**General:**

Determinar la precisión diagnóstica del Algoritmo del Sistema de aprendizaje profundo para detectar neuropatía óptica glaucomatosa utilizando la retinografía.

Específicos:

- Determinar la sensibilidad del Algoritmo del Sistema de aprendizaje profundo para detectar neuropatía óptica glaucomatosa utilizando la retinografía.
- Determinar la especificidad del Algoritmo del Sistema de aprendizaje profundo para detectar neuropatía óptica glaucomatosa utilizando la retinografía.

Hipótesis:

Hipótesis nula (H1):

El algoritmo del sistema de aprendizaje profundo usando la retinografía si tiene buena sensibilidad y especificidad para detectar neuropatía óptica glaucomatosa.

Hipótesis alterna (H0):

El algoritmo del sistema de aprendizaje profundo usando la retinografía no tiene buena sensibilidad y especificidad para detectar neuropatía óptica glaucomatosa.

II. MATERIALES Y MÉTODOS

DISEÑO DEL ESTUDIO

Revisión sistemática y metaanálisis.

CRITERIOS DE SELECCIÓN

Criterios de inclusión:

- Tipo de estudio: estudios observacionales (prueba diagnóstica) o ensayos aleatorizados.
- Población: retinografía en color totalmente anonimizadas
- Intervención: sistema de aprendizaje profundo
- Resultado: detectar neuropatía óptica glaucomatosa

Criterios de exclusión:

- Estudios de reporte de casos, cartas al editor, resúmenes de congreso, revisiones sistemáticas o narrativas, comentarios.
- Retinografía de enfermedades asociadas, que afecten los resultados de las pruebas: Retinopatía diabética e hipertensiva, miopía alta
- Retinografía y campo visual no óptimos

PROCEDIMIENTOS Y TÉCNICAS

Estrategia de búsqueda:

Se realizó la estrategia de búsqueda (Anexo 1) y se complementó con términos adicionales tomados de otros trabajos similares y relacionados al tema.

Con esta base, se consultó las siguientes bases de datos: pubmed, Scopus, Web-Science, Ovid-Medline, Ovid-Embase. Se adaptó cada una de éstas a la estrategia de búsqueda. Luego de ubicar aquellos estudios que se incluyeron, también se buscaron en sus referencias otros estudios que puedan ser incluidos.

Manejo de datos obtenidos:

Posteriormente fueron descargados los metadatos de los estudios incluidos para ser vinculados en un único fichero. Se empleará el software Rayyan QCRI para hallar los análisis que se encuentran duplicados y los procedimientos de elección.

Selección de estudios:

Se realizó en el Rayyan QCRI a través de dos fases. La correspondiente revisión de acuerdo al título-resumen y otra a todo el texto. En la fase primaria, la revisora hizo la elección conforme a los elementos de exclusión como de pertinente inclusión. La segunda fase se revisó a texto completo aquellos trabajos que quedaron incluidos de la selección anterior.

Extracción de datos:

La extracción se hizo independientemente por la autora. Los datos extraídos fueron del tipo: metadatos (autoría, año de publicación, país, tipo de estudio, características de prueba diagnóstica, cuando los valores no estuvieron disponibles también extrajimos manualmente y calculamos valores para sensibilidad, especificidad, las razones de probabilidad y curva ROC.).

Análisis de riesgo de sesgo:

Esto se hizo de forma independiente. Se usó estudios de precisión diagnóstica 2(QUADAS-2) para estimar la calidad metodológica de los estudios de prueba diagnóstica incluidos(25)

Análisis estadístico y síntesis de datos:

El análisis de datos se realizó: (1) cálculo de las medidas de precisión de la prueba (sensibilidad, la especificidad) del total de estudios y por subgrupos (2) prueba de heterogeneidad. Para el primer punto, se calculó la sensibilidad y especificidad realizando un metaanálisis de proporciones utilizando distribución binomial exacta. Los subgrupos fueron en función a la arquitectura de base

que usó el sistema de aprendizaje profundo en cada estudio. Para la heterogeneidad se usó el estadístico I² para determinar la heterogeneidad en los estudios, hay heterogeneidad si el I² > 30%. Estos datos fueron obtenidos con el programa Stata v14.

Aspectos éticos:

Este estudio de Revisión Sistemática, consta de datos públicos y acceso abierto a las redes, con pacientes anonimizados, por ello, según la declaración de Helsinki, no violaremos el párrafo 21 de los principios básicos .(26)

III. RESULTADOS

Búsqueda y selección de estudios

Se obtuvieron un total de 300 artículos procedentes de las bases de datos: Pubmed (n=37), Scopus (n=33), Web of Science (n=21), Ovid Embase (n=24) y Ovid Medline (n=185) que, tras la eliminación de 74 estudios duplicados, inició con la fase de selección por título y resumen, excluyendo 200 artículos y luego 18, considerando a los excluidos por texto completo (Anexo 2) y los que no pudieron ser recuperados en su versión completa (27–43)), para finalmente quedarse con 8 artículos como resultado de la selección final que entrarían en la revisión (Gráfico. 1)

Extracción de datos

Fueron seleccionadas las características más relevantes de cada estudio: fuente de datos de la muestra, donde se especifica la procedencia de las retinografías, también se precisa la forma de obtención de la muestra (si era aleatoria o consecutiva). Se menciona también los criterios de calidad para las muestras, si es que el estudio definió y los parámetros que consideraron para calidad adecuada. Otra columna menciona las características de la prueba índice, si esta se encuentra disponible o si fue desarrollada y validada dentro del estudio. En el mismo sentido, en la columna de prueba de referencia se mencionan los protocolos que siguieron en cada estudio para conducir de la manera más objetiva posible la revisión por parte de los especialistas. Finalmente, también se encuentran las columnas que especifican las categorías que usó el sistema de aprendizaje profundo y los parámetros de sensibilidad y especificidad que se calcularon en cada estudio. (Tabla 1)

Análisis de riesgo de sesgo

Se llevó a cabo teniendo en cuenta los 3 dominios del instrumento QUADAS-2: Selección de pacientes, prueba índice, prueba de referencia, considerando también las preocupaciones para su

aplicación. Así mismo para cada estudio se evaluó también la sección de flujos y tiempos. El dominio más afectado en la mayoría de estudios, fue el de la selección de pacientes; seguido por el de la prueba índice. (Gráfico 2) Los detalles del análisis de sesgo por cada estudio, se especifican en el Anexo 3.

Síntesis de datos

Para la ejecución de los metaanálisis se tuvo en cuenta 3 subgrupos, según la arquitectura de Aprendizaje Profundo empleada para cada caso: InceptionV-3, Resnet y Pegasus. La sensibilidad y especificidad para cada uno, fue calculada (Gráfico 3) y finalmente al combinarse todos los estudios se alcanzó una sensibilidad de 0.90 con un IC al 95% que va de 0.86 a 0.94 y una especificidad que va de 0.93 con un IC al 95% que va de 0.90 a 0.96.

IV. DISCUSIÓN

Los 8 estudios incluidos mostraron muy buenos parámetros de agudeza diagnóstica individualmente, mientras que al combinar los resultados se encontró que los sistemas de aprendizaje profundo (AP) tienen una probabilidad del 90% de detectar la enfermedad cuando esta esté presente y del 93% de dar un diagnóstico negativo cuando la enfermedad esté realmente ausente. Así mismo, los artículos incluidos también respondieron diversas preguntas que involucraban directa o indirectamente los sistemas de AP: uno de los estudios comparó la precisión diagnóstica de los optómetras respecto a oftalmólogos y sistemas de AP(44), otro señaló los casos que deberían volver a ser evaluados por especialistas en glaucoma (45), otro permitió tener una clasificación de los casos de glaucoma según severidad (46), un estudio puso a prueba diversos perfiles de arquitectura de AP en la detección de glaucoma para ver cuál era mejor(47), y finalmente también hubo un estudio que utilizó el AP para distinguir la neuropatía no glaucomatosa con palidez discal del glaucoma(14).

Es preciso mencionar también sobre los estudios incluidos que desarrollaron y validaron con indicadores de agudeza diagnóstica dentro del mismo estudio (14,47) pues consideramos que pudo haber permitido la introducción de sesgo a nivel de las pruebas índice, al no tener definido un umbral previo y esto conlleve a una sobreestimación de sus resultados, debido a un sobreajuste de su capacidad predictiva (48), esto a diferencia de los estudios que trabajaron sobre sistemas de AP existentes para detección de glaucoma (Pegasus) y en el estudio llevaron a cabo su validación (44,49), o aquellos que tuvieron dos fases claramente diferenciadas en las que abordaron en primer lugar, el desarrollo del sistema y en segundo lugar la validación del mismo (45,46,50,51), incluso con imágenes distintas a las usadas en la primera fase.

Desde el descubrimiento de los rayos X y su potencial uso en la medicina, las imágenes diagnósticas han cobrado especial

relevancia en la práctica clínica hasta hoy resultan indispensables, no solo para el diagnóstico, sino también para la caracterización, estadificación, evaluación, seguimiento de la enfermedad, orientación de los procesos de intervención. En tal sentido, la cantidad de imágenes también fue aumentando, desde unas cuantas imágenes 2D a cientos de imágenes en 3D hasta llegar a miles de imágenes dinámicas en 4D. En tal sentido, con este creciente flujo de información para interpretar, resulta difícil que el médico mantenga su eficiencia en la adecuada interpretación y procesamiento de toda la información disponible en las imágenes para mejorar la precisión y la atención del paciente.(52) Es a partir de esta realidad, que surge la necesidad de aplicar técnicas computacionales basadas en aprendizaje autónomo para ayudar al médico en el procesamiento de información proveniente de imágenes médicas. (52) En la Universidad de Chicago, se llevó a cabo el desarrollo sistemático de técnicas de aprendizaje autónomo para el análisis de imágenes médicas (53), posterior a esto hubo interés en desarrollar estos diagnósticos asistidos por computadora como una segunda opinión para la interpretación de imágenes, como el caso de Chan et al (54) y su sistema para la detección de microcalcificaciones en mamografías. Sin embargo, el aprendizaje autónomo convencional por sí solo, empezaba a mostrar serias limitaciones: siendo la principal de ellas que dependía directamente de la experiencia de quienes indican las características de la imagen que pueden representar las diversas categorías de la variable que se está buscando representar, así como las formulaciones matemáticas que subyacen al proceso y son las responsables de traducir las características de la imagen a valores numéricos y de esta forma sean utilizadas como variables predictoras de entrada para un clasificador, formándose un modelo predictivo ajustando los pesos de las diversas características en función de las propiedades estadísticas de un conjunto de muestras de entrenamiento para estimar la probabilidad de que una imagen pertenezca a uno de los estados(52). En este sentido, las

principales limitaciones del aprendizaje autónomo convencional son: el desarrollador humano no puede traducir los patrones complejos de la enfermedad en un número finito de descriptores de características, incluso si ha visto una gran cantidad de casos de la población de pacientes(52) y en los casos en los que esto se lograra, las características diseñadas por un humano, también pueden tener dificultades para ser robustas frente a las grandes variaciones de patrones normales y anormales en la población(52). De esta forma, el rendimiento del sistema CAD desarrollado a menudo está limitado en su poder de discriminación o generalización.(52)

De esta forma, surge el interés por el AP como el método de aprendizaje autónomo con más aplicaciones en la actualidad. Pues a diferencia del método anterior, este usa una compleja red neuronal multicapa que transforma la información de entrada en múltiples niveles de abstracción(52,55). En la actualidad las redes neuronales convolucionales profundas (RNCP) son las más utilizadas, pues con un conjunto de datos de entrenamiento lo suficientemente grande puede aprender a extraer automáticamente características relevantes de las muestras de entrenamiento para una tarea determinada ajustando iterativamente sus pesos, por lo tanto, descubre representaciones de características a través del entrenamiento y no requiere características diseñadas manualmente como entrada (52,55) En este sentido, respecto al aprendizaje autónomo convencional, si el AP se entrena con una muestra lo suficientemente grande se espera que las características diseñadas por sus redes neuronales sean superiores a las diseñadas manualmente por un humano ya que puede tomar de referencia miles a millones de casos que incluso los especialistas humanos con más años de experiencia, no lograrían ver en sus vidas. (52) El primer indicio de aplicaciones de la aplicación de las RNCP data del año 1993 en el que Lo et al (56) quien las empleó para la detección de nódulos pulmonares en radiografías de tórax y microcalcificaciones en radiografías y

destacaron el potencial uso de estas técnicas en un escenario clínico. Trabajos similares, para microcalcificaciones mamarias, se siguieron desarrollando en los años siguientes tanto para detección como para clasificación (57–61), aunque estas primeras RNCP no eran muy profundas (en cantidad de capas) se demostró la capacidad de reconocimiento de imágenes médicas(52).

Esto mejoraría con el transcurso del tiempo y el desarrollo de técnicas de entrenamiento para la RNCP(52) tales como: el entrenamiento previo no supervisado por capas seguido de un ajuste fino supervisado (62), el uso de la unidad lineal rectificadora(63), abandono para reducir sobre ajuste (64) y la normalización por lotes (65) permitiendo el entrenamiento para redes neuronales con más capas y millones de pesos, como en un inicio lo fue “AlexNet” que constaba de 5 capas convolucionales y 3 capas completamente conectadas con 60 millones de pesos, que lograron clasificar más de 1000 tipos de objetos cotidianos demostrando la capacidad de reconocimiento de múltiples capas de una estructura profunda(52). Sin embargo, una creciente desventaja de las RNCP conforme se volvían más profundas, era la degradación del rendimiento que aumentaba el nivel de error conforme se agregaban más capas, por lo que surge el aprendizaje residual o ResNet, el cual tenía por ventaja de proponer conexiones de salto o atajos para las redes convolucionales, que permitan superar el problema de la degradación (66) pudiendo agregar de esta forma muchas más capas (de 110 a 152) sin que el rendimiento se vea afectado. (52)

En ese sentido, dado el creciente aumento de las publicaciones sobre AP que hemos tenido en los últimos años, hay estudios que han buscado evaluar su presencia dentro de la medicina, como el de Aggarwal et al (67) en el 2021, encontró que el principal interés actual del AP está sobre la oftalmología (detección de retinopatía diabética y degeneración macular, además de glaucoma) con 82 estudios, enfermedades mamarias (cáncer de mama) también con 82 estudios y 115 sobre aplicaciones en neumología (nódulos

pulmonares o cáncer de pulmón). Sin embargo, un punto que destacan es la elevada heterogeneidad que hay dentro de los estudios, lo que discrepa con lo reportado por otra revisión(68), que evalúa nuestra misma pregunta de investigación, pero con notables diferencias en su desarrollo a lo planteado en nuestro estudio: a parte del hecho que encuentran 2 estudios menos, al realizar su meta análisis, calcularon el área bajo la curva (AUC) para el sistema de AP y lo compararon con el obtenido por los oftalmólogos, lo que probablemente les permitiese haber obtenido bajos niveles de heterogeneidad, sin embargo se ve restringido al no presentar estimaciones para la sensibilidad y especificidad.

En el caso de nuestro estudio, los niveles de heterogeneidad elevados, se pudo haber debido a las grandes diferencias que existieron en las metodologías de cada estudio, a pesar de que todos respondían a una misma pregunta de investigación: en primer lugar, en la Tabla 1, se evidencia la gran variabilidad que pudo haber dentro de las pruebas de referencia, en especial en todos los mecanismos que implementaron algunos estudios(31,45,46,50,51) para lograr una evaluación objetiva que sin embargo pudieron variar bastante entre sí. También cabe variabilidad introducida por las 3 versiones diferentes de arquitectura para el sistema de AP, tanto de los que validaban programas existentes como de los que desarrollaban uno dentro del estudio. Por último, otro punto importante de variabilidad pudo estar dado por los criterios de calidad tan rigurosos o explícitos de algunos estudios respecto a otros en los que no implementaban ninguno.

La principal limitación del estudio, está dada por la aún poca cantidad de investigaciones disponibles que haya, con metodologías homogéneas o estandarizadas, probablemente por lo aún bastante nuevo que es este enfoque; sin embargo, estos resultados permitirían la aplicación programas de despistaje para prevención de la ceguera a gran escala, de la misma forma que se ha realizado ya para detección de cáncer de mama(58–60).

V. CONCLUSIONES

- El uso del sistema de aprendizaje profundo para la detección de GON ofrece una sensibilidad adecuada respecto a la capacidad de detección de oftalmólogos especializados.
- El uso del sistema de aprendizaje profundo para la detección de GON ofrece una sensibilidad adecuada respecto a la capacidad de detección de oftalmólogos especializados.

REFERENCIAS BIBLIOGRÁFICAS

1. Glaucoma - American Family Physician [Internet]. [cited 2020 Jul 21]. Available from: <https://www.aafp.org/afp/2016/0415/p668.html>
2. Yebra JC, Mora NF. Glaucoma y población en 2020. *Gac Optom Óptica Oftálmica*. 2016;(513):40–5.
3. Quigley HA. 21st century glaucoma care. *Eye Lond Engl*. 2019;33(2):254–60.
4. Aging and ocular tissue stiffness in glaucoma. - Abstract - Europe PMC [Internet]. [cited 2020 Jul 21]. Available from: <https://europepmc.org/article/med/28666629>
5. Emerging Concepts in Glaucoma and Review of the Literature - PubMed [Internet]. [cited 2020 Jul 21]. Available from: <https://pubmed.ncbi.nlm.nih.gov/27125182/>
6. Giaquinta Aranda A, Fernández Araque A, Curbelo Rodriguez R, Rojo Aragues A. Glaucoma y antioxidantes: revisión sistemática. *Rev Mex Oftalmol*. 2017 May 1;91(3):112–21.
7. Heijl A, Bengtsson B, Hyman L, Leske MC. Natural History of Open-Angle Glaucoma. *Ophthalmology*. 2009 Dec 1;116(12):2271–6.
8. Lascaratos G, Chau KY, Zhu H, Gkotsi D, King R, Gout I, et al. Resistance to the most common optic neuropathy is associated with systemic mitochondrial efficiency. *Neurobiol Dis*. 2015 Oct 1;82:78–85.
9. Evangelho K, Mogilevskaya M, Losada-Barragan M, Vargas-Sanchez JK. Pathophysiology of primary open-angle glaucoma from a neuroinflammatory and neurotoxicity perspective: a review of the literature. *Int Ophthalmol*. 2019 Jan;39(1):259–71.
10. Liang L, Zhang R, He LY. Corneal hysteresis and glaucoma. *Int Ophthalmol*. 2019 Aug;39(8):1909–16.
11. Thompson AC, Jammal AA, Medeiros FA. A Review of Deep Learning for Screening, Diagnosis, and Detection of Glaucoma Progression. *Transl Vis Sci Technol*. 2020 Jul 22;9(2):42.
12. Yung ES, Chang EL, Moster MR. The Use of Optic Disc and Retinal Nerve Fiber Imaging in Detecting Structural Damage Before the Onset of Functional Loss. *Adv Ophthalmol Optom*. 2017 Aug;2(1):243–60.
13. Castañeda Díez R, Jiménez Román J, Iriarte Barbosa MJ. Concepto de sospecha de glaucoma de ángulo abierto: definición, diagnóstico y tratamiento. *Rev Mex Oftalmol*. 2014 Oct 1;88(4):153–60.
14. Yang HK, Kim YJ, Sung JY, Kim DH, Kim KG, Hwang JM. Efficacy for Differentiating Nonglaucomatous Versus Glaucomatous Optic Neuropathy Using Deep Learning Systems. *Am J Ophthalmol*. 2020 Aug 1;216:140–6.
15. Diaz-Pinto A, Morales S, Naranjo V, Alcocer P, Lanzagorta-Aresti A. Diagnóstico Automático del Glaucoma a través de la Segmentación y Análisis de la Copa Óptica Usando Imágenes de Fondo de Ojo. In 2016.

16. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology*. 2018;125(8):1199–206.
17. Chakrabarty L, Joshi GD, Chakravarty A, Raman GV, 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.
18. Gómez-Valverde JJ, Antón A, Fatti G, Liefers B, Herranz A, Santos A, et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. *Biomed Opt Express*. 2019 Feb 1;10(2):892–913.
19. Al-Aswad LA, Kapoor R, Chu CK, Walters S, Gong D, Garg A, et al. Evaluation of a Deep Learning System For Identifying Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *J Glaucoma*. 2019;28(12):1029–34.
20. Pérez Molina EC, León Veitía L. La fotografía de fondo de ojo como método de diagnóstico en el glaucoma. *Medicentro Villa Clara*. 2017;3–10.
21. 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 Dec 1;137(12):1353–60.
22. Bojikian KD, Lee CS, Lee AY. Finding Glaucoma in Color Fundus Photographs Using Deep Learning. *JAMA Ophthalmol*. 2019 Sep 12;
23. An Adaptive Threshold Based Image Processing Technique for Improved Glaucoma Detection and Classification - PubMed [Internet]. [cited 2020 Jul 3]. Available from: <https://pubmed.ncbi.nlm.nih.gov/26321351/>
24. 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*. 2019;126(12):1627–39.
25. Ciapponi A. QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica. 2015;5.
26. Issue Information-Declaration of Helsinki. *J Bone Miner Res*. 2017;32(9):BM i-BM ii.
27. Norouzifard M, Nemati A, Abdul-Rahman A, GholamHosseini H, Klette R, Chang C.-Y., et al. A Comparison of Transfer Learning Techniques, Deep Convolutional Neural Network and Multilayer Neural Network Methods for the Diagnosis of Glaucomatous Optic Neuropathy [Internet]. *Commun. Comput. Info. Sci.* ["School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology (AUT), Auckland, New Zealand", "School of Engineering and Technology, University of Washington, Tacoma, United States", "Department of Ophthalmology, Counties Manukau DHB, Auckland, New Zealand"]: Springer Verlag; 2019. 627–635 p. (23rd International Computer Symposium, ICS 2018; vol. 1013). Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85069675819&doi=10.1007%2f978-981-13-9190-3_69&partnerID=40&md5=c3dfafaac83c8bbc655acc2ef7b60275

28. He M, Li Z, Keel S, Chang R, He M, Li Z, et al. A deep learning system for detecting glaucomatous optic neuropathy and age-related macular degeneration based on color fundus photographs. *Invest Ophthalmol Vis Sci*. 2018;59(9).
29. Muramatsu C. Diagnosis of Glaucoma on Retinal Fundus Images Using Deep Learning: Detection of Nerve Fiber Layer Defect and Optic Disc Analysis [Internet]. *Adv. Exp. Med. Biol.* Faculty of Data Science, Shiga University, Hikone, Japan: Springer; 2020. 121–132 p. (Advances in Experimental Medicine and Biology; vol. 1213). Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85079082855&doi=10.1007%2f978-3-030-33128-3_8&partnerID=40&md5=31947be82d04b8c1a952a104a213d5f4
30. Sun L. Considerations When Using a Deep Learning System to Diagnose Glaucomatous Optic Neuropathy. *JAMA Ophthalmol*. 2020;138(4):419–20.
31. Al-Aswad L., Kapoor R., Chu C., Walters S., Gong D., Garg A., et al. Evaluation of the pegasus deep learning system for identifying glaucomatous optic neuropathy based on color fundus photographs. *Invest Ophthalmol Vis Sci* [Internet]. 2019;60(9). Available from: <https://iovs.arvojournals.org/article.aspx?articleid=2746173>
32. Sponsel WE. Integrating numerical indices of structure and function to optimize diagnostic sensitivity and specificity in screening for glaucoma. *Ophthalmic Epidemiol*. 2005;12(3):163–6.
33. MacCormick I, Williams B, Zheng Y, Li K, Al-Bander B, Czanner S, et al. Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile. *PLOS ONE*. 2019;14(1).
34. Montesano G, Bryan SR, Crabb DP, Fogagnolo P, Oddone F, McKendrick AM, et al. A Comparison between the Compass Fundus Perimeter and the Humphrey Field Analyzer. *Ophthalmology*. 2019;126(2):242–51.
35. Kim YK, Yoo BW, Kim HC, Park KH. Automated Detection of Hemifield Difference across Horizontal Raphe on Ganglion Cell--Inner Plexiform Layer Thickness Map. *Ophthalmology*. 2015;122(11):2252–60.
36. Ko Y.-C., Wey S.-Y., Lee C.-Y., Liu C.J.-L. Automated glaucoma detection based on deep convolutional neural network. *Invest Ophthalmol Vis Sci* [Internet]. 2018;59(9). Available from: <https://iovs.arvojournals.org/article.aspx?articleid=2691779>
37. Oh JE, Yang HK, Kim KG, Hwang JM. Automatic computer-aided diagnosis of retinal nerve fiber layer defects using fundus photographs in optic neuropathy. *Invest Ophthalmol Vis Sci*. 2015;56(5):2872–9.
38. Zapata M.A., Royo-Fibla D., Font O., Vela J.I., Marcantonio I., Moya-Sanchez E.U., et al. Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. *Clin Ophthalmol*. 2020;14:419–29.
39. Grewal DS, Jain R, Grewal SPS, Rihani V. Artificial neural network-based

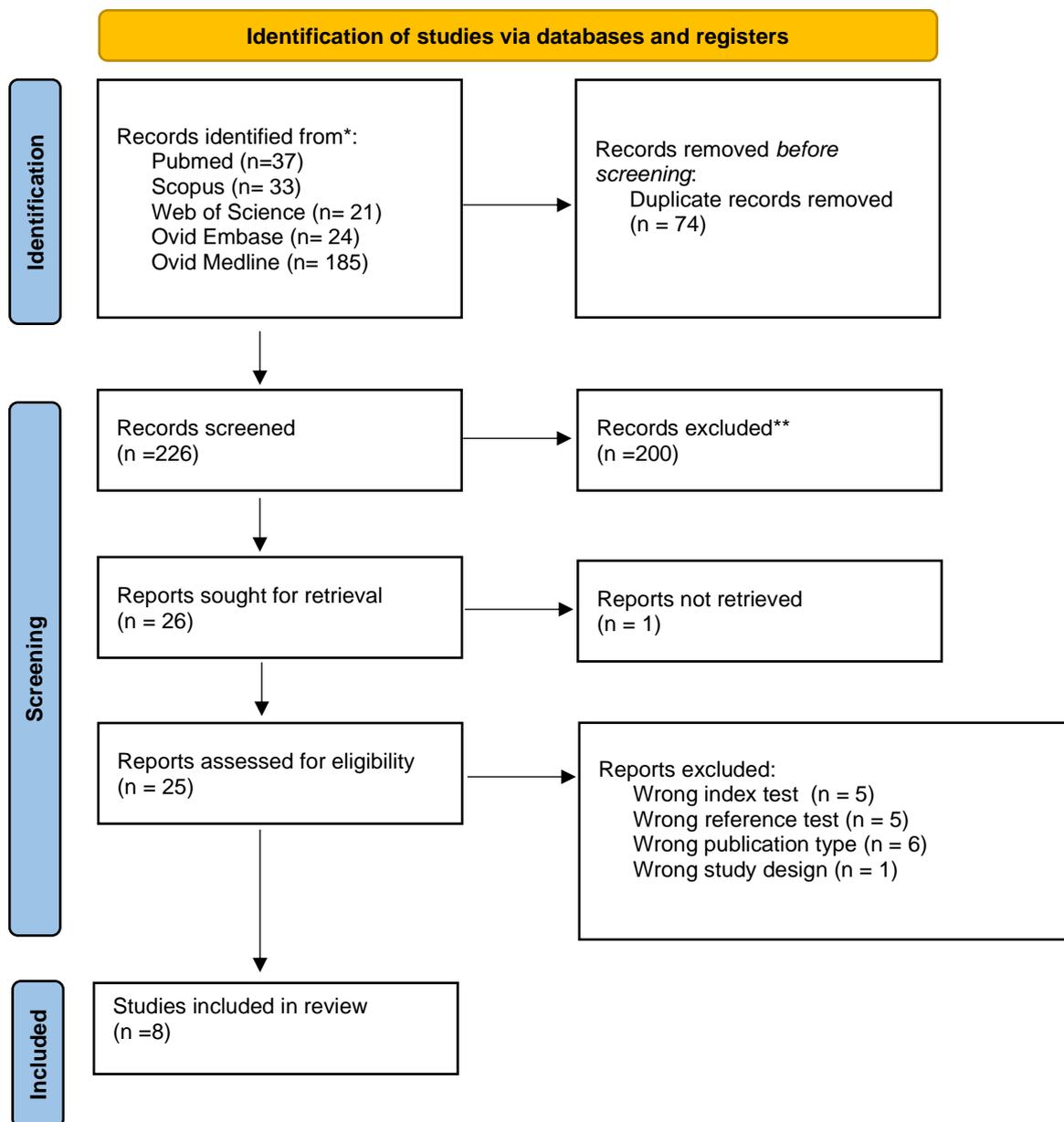
- glaucoma diagnosis using retinal nerve fiber layer analysis. *Eur J Ophthalmol*. 2008;18(6):915–21.
40. Ko YC, Wey SY, Chen WT, Chang YF, Chen MJ, Chiou SH, et al. Deep learning assisted detection of glaucomatous optic neuropathy and potential designs for a generalizable model. *PloS One*. 2020;15(5):e0233079.
 41. 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*. 2019;126(4):513–21.
 42. Jammal AA, Thompson AC, Mariottoni EB, Berchuck SI, Urata CN, Estrela T, et al. Human Versus Machine: Comparing a Deep Learning Algorithm to Human Gratings for Detecting Glaucoma on Fundus Photographs. *Am J Ophthalmol*. 2020;211(3):123–31.
 43. Li B, Chen H, Zhang B, Yuan M, Jin X, Lei B, et al. Development and evaluation of a deep learning model for the detection of multiple fundus diseases based on colour fundus photography. *Br J Ophthalmol* [Internet]. 2021 [cited 3AD Jan 1]; Available from: <https://pubmed.ncbi.nlm.nih.gov/33785508/>
 44. Rogers TW, Jaccard N, Carbonaro F, Lemij HG, Vermeer KA, Reus NJ, et al. Evaluation of an AI system for the automated detection of glaucoma from stereoscopic optic disc photographs: the European Optic Disc Assessment Study. *Eye*. 2019 Nov 1;33(11):1791–7.
 45. 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*. 2019 Dec 1;126(12):1627–39.
 46. Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a Deep Learning System for Detecting Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *Ophthalmology*. 2018 Aug 1;125(8):1199–206.
 47. 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 Nov 12;8(1):16685.
 48. Agustin Ciapponi. QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica. *Evid Actual En Práctica Ambulatoria* [Internet]. 2015 Apr 1 [cited 2022 Mar 28];18(1). Available from: <https://www.evidencia.org/index.php/Evidencia/article/view/6341>
 49. Al-Aswad LA, Kapoor R, Chu CK, Walters S, Gong D, Garg A, et al. Evaluation of a Deep Learning System For Identifying Glaucomatous Optic Neuropathy Based on Color Fundus Photographs. *J Glaucoma* [Internet]. 2019;28(12). Available from: https://journals.lww.com/glaucomajournal/Fulltext/2019/12000/Evaluation_of_a_Deep_Learning_System_For.2.aspx
 50. 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 Dec 1;137(12):1353–60.

51. Li F, Yan L, Wang Y, Shi J ping, Chen H, Zhang X, et al. Deep learning-based automated detection of glaucomatous optic neuropathy on color fundus photographs. *Graefes Arch Clin Exp Ophthalmol*. 2020;258:851–67.
52. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol*. 2020;1213:3–21.
53. Chen CM, Chou YH, Tagawa N, Do Y. Computer-aided detection and diagnosis in medical imaging. *Comput Math Methods Med*. 2013;2013.
54. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. *Med Phys*. 1987 Aug;14(4):538–48.
55. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015 May 28;521(7553):436–44.
56. Lo SCB, Chan HP, Lin JS, Li H, Freedman MT, Mun SK. Artificial convolution neural network for medical image pattern recognition. *Neural Netw*. 1995;8(7–8):1201–14.
57. Chan H, Lo S, Helvie M, Goodsitt M, Cheng S, Adler D. Recognition of mammographic microcalcifications with artificial neural network. *Radiology*. 1993;189(P):318.
58. Chan H, Sahiner B, Lo S, Helvie M, Petrick N, Adler D, et al. Computer-aided diagnosis in mammography: detection of masses by artificial neural network. *Med Phys*. 1994;21:875–6.
59. Chan H, Lo SB, Sahiner B, Lam KL, Helvie MA. Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Med Phys*. 1995;22(10):1555–67.
60. Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, et al. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans Med Imaging*. 1996;15(5):598–610.
61. Wei D, Sahiner B, Chan HP, Petrick N. Detection of masses on mammograms using a convolution neural network. In *IEEE*; 1995. p. 3483–6.
62. Erhan D, Courville A, Bengio Y, Vincent P. Why does unsupervised pre-training help deep learning? In *JMLR Workshop and Conference Proceedings*; 2010. p. 201–8.
63. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In 2010.
64. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(1):1929–58.
65. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *PMLR*; 2015. p. 448–56.

66. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In 2016. p. 770–8.
67. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Digit Med.* 2021 Apr 7;4(1):65.
68. Buisson M, Navel V, Labbé A, Watson SL, Baker JS, Murtagh P, et al. Deep learning versus ophthalmologists for screening for glaucoma on fundus examination: A systematic review and meta-analysis. *Clin Experiment Ophthalmol.* 2021 Dec;49(9):1027–38.

VI. TABLAS Y GRÁFICOS

Gráfico 1.



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: <http://www.prisma-statement.org/>

Tabla 1

Autor (año)	Fuente de datos de la muestra	Forma de obtención de la muestra	Muestra (n)	Criterios de calidad para las muestras	Prueba índice	Prueba de referencia	Categorías de diagnóstico por el sistema de aprendizaje profundo	Sensibilidad Especificidad
Liu (2019)	Base de datos en línea de la Alianza China para el Estudio del Glaucoma (Chinese Glaucoma Study Alliance)	Imágenes seleccionadas al azar de una base de datos en línea	28569 imágenes de fondo de ojo obtenidas de ambos ojos de 1092 pacientes	Control de calidad de imágenes incluidas: (1) la imagen no contenía una resolución severa reducciones o artefactos significativos; (2) el campo de imagen incluido toda la cabeza del nervio óptico y la mácula; (3) la iluminación era aceptable (es decir, ni demasiado oscuro ni demasiado claro); (4) la imagen era suficientemente enfocado para clasificar la cabeza del	Sistema de Aprendizaje profundo basado en ResNet desarrollado en la primera fase del mismo estudio	Un primer control de calidad realizado por estudiantes de medicina, una calificación inicial realizada por oftalmólogos que recibieron entrenamiento previo de manera independiente y por duplicado, y finalmente con la participación de especialistas en glaucoma con más de 10 años de experiencia para resolver discrepancias.	Glaucoma confirmado y glaucoma Probable Glaucoma improbable	S: 96.2% E: 97.7%.

				nervio óptico y la retina capa de fibras nerviosas				
Li (2020)	Imágenes obtenidas junto al historial médico de pacientes de 2 hospitales de Shanghái entre junio de 2014 y julio de 2018.	Estudio transversal con muestras tomadas de cohortes	3481 imágenes de fondo de ojo obtenidas de 249 pacientes	Criterios de calidad para las imágenes incluidas: 1) Mala calidad: Artefactos severos, Insuficiente inclusión del entorno, sobreexposición del disco óptico, no se pueden identificar los vasos dentro de 1 DD del margen del disco óptico, $\geq 50\%$ del área está oscurecida 2) Mala ubicación si no es posible ver el disco óptico completamente	Sistema de aprendizaje profundo basado en ResNet101 durante la primera fase del estudio	En una fase inicial, todas las imágenes obtenidas procedían de evaluaciones conducidas por oftalmólogos. Dentro del estudio, en un primer momento cada imagen fue asignada aleatoriamente a 4 oftalmólogos por 3 veces, en sesiones espaciadas por al menos una semana. En la segunda revisión se llevó la misma dinámica por especialistas en glaucoma. Las discrepancias serían por un comité conformado por todos los médicos que participaron en el estudio.	Ojo normal Glaucoma confirmado*	S: 97.4% E: 92.5%
Christopher (2018)	Imágenes obtenidas de estudios longitudinales en curso: Estudio de Evaluación de	Estudio transversal con muestras tomadas de cohortes	14822 imágenes de fondo de ojo obtenidas de 2329 participantes.	No especificados.	Sistema de Aprendizaje Profundo basado en ResNet50 desarrollado y puesto a prueba en la	En una fase inicial fue revisada por dos evaluadores independientemente y cegados, ante discrepancias acudieron a un especialista experimentado.	Glaucoma (leve, moderado, severo) * Sin glaucoma	S: 84% E: 83%

	Ascendencia Africana y Glaucoma y la Universidad de California: Innovaciones diagnósticas en el estudio del glaucoma.				misma fase del estudio			
Al-Aswad (2019)	Imágenes obtenidas de la Base de datos en línea de imágenes de fondo de retina para análisis e investigación de glaucoma (ORIGA)	Estudio de casos y controles tomando muestras aleatorias de estudios de cohorte	100 imágenes aleatorias, de las cuales 50 tenían diagnóstico de glaucoma y 50 no. De estas se seleccionaron y repitieron 10	Todas las imágenes incluidas fueron seleccionadas para ser graduables y tener baja presencia de artefactos.	Pegasus Sistema de Aprendizaje profundo basado en ResNet50 validado previamente	Las imágenes fueron revisadas entre 6 evaluadores: 4 oftalmólogos (2 especialistas en glaucoma y 2 en proceso de especialización) y 2 residentes de segundo y tercer año respectivamente, cegados al diagnóstico clínico de los pacientes y a la existencia de imágenes duplicadas.	Glaucoma Sin Glaucoma	S: 83.7% E: 88.2%
Phene (2019)	Imágenes obtenidas de diversas bases de datos en Estados Unidos, 3 hospitales oftalmológicos de la India y el Biobanco	Estudio transversal tomando muestras aleatorias de las bases de datos seleccionadas	1195 imágenes de la misma cantidad de pacientes	Los especialistas evaluaron cada imagen, teniendo en cuenta que todas las imágenes deben permitir la identificación de los elementos de	Sistema de aprendizaje profundo basado en Inception-V3 durante la primera fase del estudio	12 oftalmólogos en proceso de especialización para glaucoma, distribuidos en paneles de 3 mediante el siguiente proceso: cada imagen es evaluada por algún miembro del panel, luego en una segunda ronda cegada esta clasificación es verificada	Glaucoma referible (alto riesgo, alta sospecha) Sin Glaucoma (no glaucoma, sospecha)	S: 80.0% E: 90.2%

	de Reino Unido.			un fondo de ojo y no deben presentar cualquier artefacto que se preste para confusión. Las imágenes no calificables, no formaban parte del estudio		por cualquiera de los miembros del panel, hasta 2 veces por cada miembro del panel durante la segunda ronda, a no ser que se alcance el consenso antes. Si a la segunda ronda hay discrepancias, se guía según la mayoría.	de bajo riesgo)	
Li (2018)	Imágenes obtenidas de la base de datos en línea LabelMe que contiene más de 200 000 imágenes de fondos de ojo a color recolectadas de diversos entornos clínicos de China	Estudio transversal tomando muestras aleatorias de las bases de datos seleccionadas	39745 imágenes seleccionadas aleatoriamente, quedándose finalmente con 8000	Se seleccionaron todas aquellas imágenes que muestren el disco óptico visible. Aquellas consideradas como de mala calidad y mala ubicación fueron excluidas	Sistema de aprendizaje profundo basado en Inception-V3 durante la primera fase del estudio	Se invitaron a 27 oftalmólogos a participar del proceso de entrenamiento, en el que se les repartieron 4 grupos de 100 imágenes, seleccionándose finalmente a 21 que lograron mantener un índice de concordancia k mayor a 0.75 en todos los casos. Luego, fue asignada una imagen a cada evaluador y secuencialmente a otros evaluadores hasta lograr 3 calificaciones consistentes.	Glaucoma Sin Glaucoma	S: 95.6% E: 92.0%
Rogers (2019)	Imágenes obtenidas del Ensayo Europeo de Evaluación del Disco óptico en el	Estudio de casos y controles tomando muestras aleatorias de estudios	94 imágenes que incluían a 40 controles sanos, 48 pacientes con glaucoma y 6 con	No especifica	Pegasus Sistema de Aprendizaje profundo basado en ResNet50 validado	La referencia fue tomada a partir de 243 oftalmólogos especialistas de 11 países.	Glaucoma Sin Glaucoma	S: 80.9% E: 86.2%

	2010	de cohorte	hipertensión ocular seleccionando una sola imagen por paciente		previamente			
Yang (2020)	Imágenes obtenidas del del sistema de archivo y comunicación de imágenes de Hospital Bundang de la Universidad Nacional de Seúl.	Estudio de transversal a partir de un diseño de casos y controles	3815 imágenes que incluían 2883 ojos sanos, 486 glaucomas y 446 ojos con neuropatía	Fueron excluidos: Sujetos con imágenes retinianas indistinguibles relacionadas con la opacidad de los medios, como opacidad de la córnea, cataratas y anomalías.	Sistema de Aprendizaje Profundo basado en ResNet50 desarrollado y puesto a prueba en la misma fase del estudio	Dos especialistas en neuro oftalmología y corroborado por una evaluación de campo visual y tomografía de coherencia óptica	Glaucoma Sin Glaucoma Neuropatía óptica no glaucomatosa	S: 92.5% E: 99.5%

* se calculó sensibilidad para cada grado de severidad, no obstante, fue considerada la general

Gráfico 2.

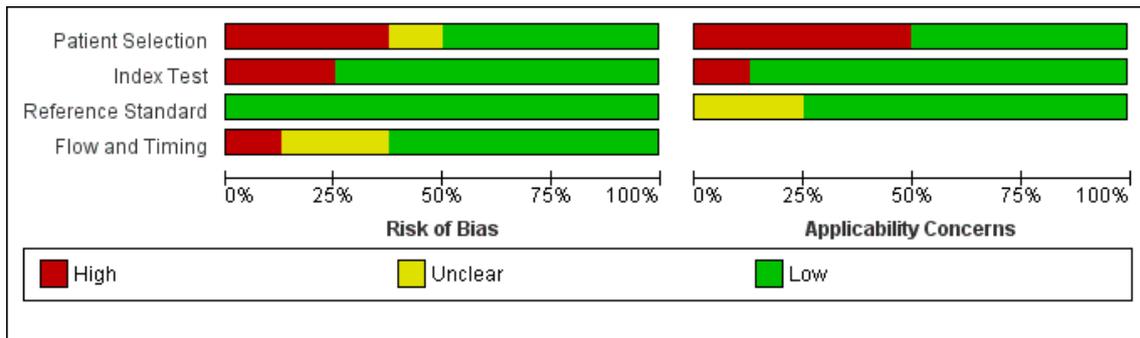
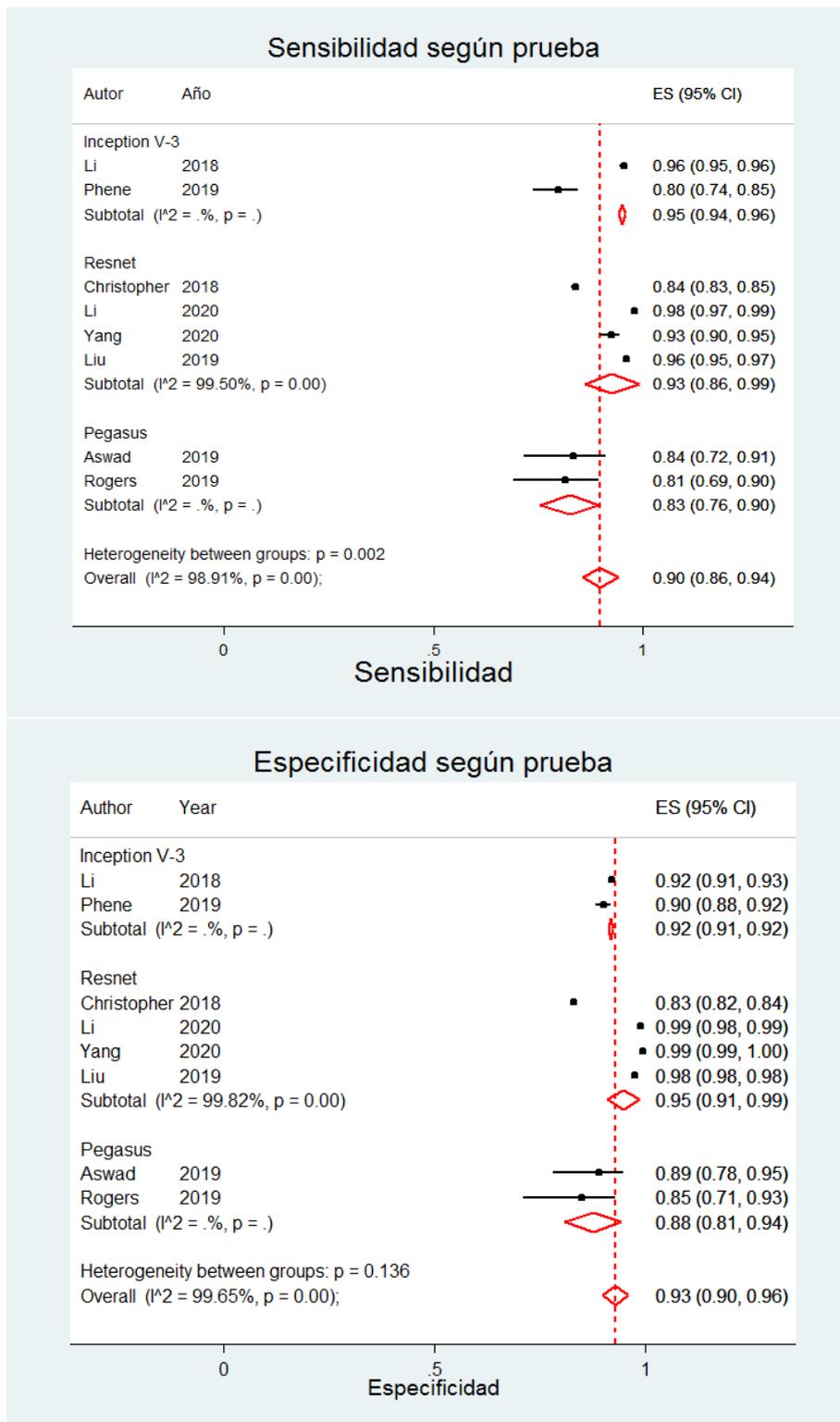


Gráfico 3.



	<p>Fields] AND "nerve"[All Fields]) AND "diseases"[All Fields])) OR "optic nerve diseases"[All Fields]) OR (("optic"[All Fields] AND "disk"[All Fields]) AND "disorder"[All Fields]))) AND (((("optic nerve diseases"[MeSH Terms] OR ("optic"[All Fields] AND "nerve"[All Fields]) AND "diseases"[All Fields])) OR "optic nerve diseases"[All Fields]) OR ("optic"[All Fields] AND "disc"[All Fields]) AND "disorders"[All Fields])) OR "optic disc disorders"[All Fields]) AND (((("optic nerve diseases"[MeSH Terms] OR ("optic"[All Fields] AND "nerve"[All Fields]) AND "diseases"[All Fields])) OR "optic nerve diseases"[All Fields]) OR ("disc"[All Fields] AND "disorder"[All Fields]) AND "optic"[All Fields]))) AND (((("optic nerve diseases"[MeSH Terms] OR ("optic"[All Fields] AND "nerve"[All Fields]) AND "diseases"[All Fields])) OR "optic nerve diseases"[All Fields]) OR ("disorder"[All Fields] AND "optic"[All Fields]) AND "disc"[All Fields])) OR "disorder optic disc"[All Fields]) AND (((("optic nerve diseases"[MeSH Terms] OR ("optic"[All Fields] AND "nerve"[All Fields]) AND "diseases"[All Fields])) OR "optic nerve diseases"[All Fields]) OR ("optic"[All Fields] AND "disc"[All Fields]) AND "disorder"[All Fields]))))</p> <p>#3= #1 AND #2</p>
SCOPUS	<p>#1 TITLE-ABS-KEY("Deep Learning System" OR "Deep Learning")</p> <p>#2 TITLE-ABS-KEY("Optic Nerve Disease" OR "Cranial Nerve II Disorder" OR "Optic Neuropathy" OR "Neuropathy, Optic" OR "Optic Neuropathies" OR "Second Cranial Nerve Diseases" OR "Neural-Optical Lesion" OR "Lesion, Neural-Optical" OR "Neural Optical Lesion" OR "Neural-Optical Lesions" OR "Cranial Nerve II Diseases" OR "Glaucomatous Optic Neuropathy")</p> <p>#3= #1 AND #2</p>
WEB OF SCIENCE	<p>#1 TS=("Deep Learning System" OR "Deep Learning")</p> <p>#2 TS=("Optic Nerve Disease" OR "Cranial Nerve II Disorder" OR "Optic Neuropathy" OR "Neuropathy, Optic" OR "Optic Neuropathies" OR "Second Cranial Nerve Diseases" OR "Neural-Optical Lesion" OR "Lesion, Neural-Optical" OR "Neural Optical Lesion" OR "Neural-Optical Lesions" OR "Cranial Nerve II Diseases" OR "Glaucomatous Optic Neuropathy")</p> <p>#3= #1 AND #2</p>
Ovid-Medline	<p>#1 Image Interpretation, Computer-Assisted/ or Deep Learning/ or Artificial Intelligence/ or "Deep Learning System".mp. or Algorithms/ or Neural Networks, Computer/</p> <p>#2 Optic Nerve Diseases/ or Glaucomatous Optic Neuropathy.mp.</p> <p>#3= #1 AND #2</p>

Ovid-	#1 deep learning/ or artificial neural network/ or Deep Learning System.mp.
Embase	#2 Glaucomatous Optic Neuropathy.mp. or glaucomatous optic neuropathy/
	#3= #1 AND #2

ANEXO 2

Estudio	Motivo de exclusión
Norouzifard (2019) He (2018) Sun (2020) Muramatsu (2020) Al-Aswad (2019) Sponsel (2005)	Tipo de publicación incorrecta
Keel (2019)	Tipo de estudio incorrecto
MacCormick (2019) Montesano (2019) Kim (2015) Ko (2018) Oh (2015)	Prueba índice incorrecta
Zapata (2020) Grewal (2008) Ko (2020) Medeiros (2019) Jammal (2020)	Prueba de referencia incorrecta
Li (2021)	Artículo no disponible a texto completo

ANEXO 3

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Aswad 2019	⊖	+	+	⊖	⊖	+	+
Christopher 2018	?	⊖	+	+	⊖	⊖	?
Li 2018	+	+	+	+	+	+	+
Li 2020	+	+	+	+	+	+	+
Liu 2019	+	+	+	+	⊖	+	+
Phene 2019	+	+	+	+	+	+	+
Rogers 2019	⊖	+	+	?	⊖	+	?
Yang 2020	⊖	⊖	+	?	+	+	+

⊖	High	?	Unclear	+	Low
---	-------------	---	----------------	---	------------