



Pontificia Universidad
JAVERIANA
Cali

**Identificación de la actividad de la toxoplasmosis ocular mediante distintas redes
neuronales convolucionales**

Germán Alberto Mejía Salgado (8985212)

Juliana Andrea Olarte Zuluaga (8985068)

Juan David Raigoso (8985069)

*Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos*

Director(a)

Hernán Darío Vargas Cardona, ING, PhD

Codirector(a)

Alejandra de-la-Torre, MD, PhD

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI, FEBRERO 03 DE 2025

TABLA DE CONTENIDO

1	INTRODUCCIÓN.....	1
2	DEFINICIÓN DEL PROBLEMA	2
2.1	PLANTEAMIENTO DEL PROBLEMA	2
2.2	FORMULACIÓN DEL PROBLEMA.....	3
2.2.1.	Pregunta de investigación	3
2.3	JUSTIFICACIÓN.....	4
3	OBJETIVOS DEL PROYECTO	5
3.1	OBJETIVO GENERAL.....	5
3.2	OBJETIVOS ESPECÍFICOS.....	5
5.	MARCO TEÓRICO Y ANTECEDENTES	6
5.1.	MARCO TEÓRICO	6
5.1.1.	Toxoplasmosis Ocular	6
5.1.2.	Inteligencia Artificial.....	13
5.1.3.	Evaluación de modelos de aprendizaje profundo.....	31
5.2.	ANTECEDENTES.....	21
5.2.1.	Antecedentes de uso de CNN en oftalmología	21
5.2.2.	Antecedentes en uveítis y toxoplasmosis ocular	22
6.	Gestión de una base de datos.....	23
6.1.	Metodología.....	23
6.1.1.	Recopilación de Datos.....	23
6.1.2.	Criterios de Calidad de imagen	24
6.1.3.	Proceso de etiquetado por especialistas	24
6.1.4.	Organización de las Imágenes	25
6.2.	Resultados.....	26
7.	Entrenamiento de las arquitecturas	26
7.1.	Metodología.....	26
7.1.2.	Aumento de Datos (Data Augmentation)	27
7.1.4.	Ajuste de Hiperparámetros y Optimización	28
7.1.5.	Construcción de arquitecturas propias.....	29
7.2.	Resultados.....	33
7.2.1.	ResNetV2.....	33
7.2.2.	VGG16.....	33
7.2.3.	InceptionV3.....	34
8.	Evaluación de los modelos.....	41

8.1.	<i>Metodología</i>	41
8.2.	<i>Resultados</i>	41
9.	DISCUSIÓN	43
10.	LIMITACIONES	44
11.	CONCLUSIONES Y TRABAJOS FUTUROS	45
11.1.	<i>CONCLUSIONES</i>	45
11.2.	<i>TRABAJOS FUTUROS</i>	45
12.	CRONOGRAMA	46
13.	REFERENCIAS BIBLIOGRÁFICAS	49
14.	Anexo 1: Glosario médico	55
15.	Anexo 2: Experimentos realizados para la selección de hiperparámetros	58

LISTA DE FIGURAS

Figura 1: Ciclo de vida del <i>Toxoplasma gondii</i> .	6
Figura 2: Lesiones típicas de toxoplasmosis ocular	9
Figura 3: Lesiones atípicas de toxoplasmosis ocular	10
Figura 4: Arquitectura de las redes convolucionales	16
Figura 5: Arquitectura de la red Inception Resnet V2.	17
Figura 6: Arquitectura de la red VVG-16	18
Figura 7: Arquitectura de la red InceptionV3	19
Figura 8: Matriz de confusión	31
Figura 9: Proceso de gestión de la base de datos	25
Figura 10: Organización de la base de datos utilizada	26
Figura 11: Exactitud, pérdida y área bajo la curva de ResNetV2	35
Figura 12: Exactitud, pérdida y área bajo la curva deVGG16	36
Figura 13: Exactitud, pérdida y área bajo la curva de InceptionV3	37
Figura 14: Exactitud, pérdida y área bajo la curva de la arquitectura número uno	39
Figura 15: Exactitud, pérdida y área bajo la curva de la arquitectura propia número dos	40

LISTA DE TABLAS

Tabla 1: Resultados arquitecturas de transfer learning seleccionadas.....	35
Tabla 2: Resultados arquitecturas propias realizadas.....	39
Tabla 3: Resumen de cinco iteraciones de los modelos elegidos para ResNetV2, VGG16 e InceptionV3.....	43

LISTA DE ANEXOS

Anexo 1: Glosario médico.

Anexo 2: Experimentos realizados para la selección de hiperparámetros.

1 INTRODUCCIÓN

La toxoplasmosis ocular (TO) es una enfermedad inflamatoria causada por la infección del protozoo *Toxoplasma gondii* (Tg), un parásito intracelular que afecta principalmente la retina y la coroides, generando lesiones retinocoroideas que pueden llevar a la ceguera si no son tratadas oportunamente [1]. El Tg es capaz de infectar a una amplia variedad de mamíferos, incluidos los humanos, siendo la infección una de las causas principales de uveítis posterior en el mundo. Se estima que más del 30% de la población mundial ha sido infectada por este parásito, con variaciones significativas en la prevalencia dependiendo de la región geográfica [2]. En América Latina, y específicamente en Colombia, se han registrado tasas de seroprevalencia de hasta el 47%, lo que resalta la magnitud del problema en esta región [2], [3], [4].

El diagnóstico de la TO sigue siendo un reto. A menudo se basa en la clínica y en la observación de lesiones retinocoroideas típicas. Puede apoyarse en pruebas complementarias, como la serología y la reacción en cadena de la polimerasa (PCR). Sin embargo, estas últimas presentan limitaciones en cuanto a sensibilidad y especificidad [5]. La recurrencia de la infección es frecuente, ya que los quistes tisulares del parásito pueden permanecer latentes y reactivarse, provocando nuevas lesiones [1],[2]. Esta situación subraya la necesidad de métodos diagnósticos que permitan identificar de manera más oportuna las lesiones activas, diferenciándolas de las inactivas, para mejorar el manejo terapéutico y prevenir el deterioro visual.

Este estudio tuvo como objetivo identificar automáticamente la actividad de la TO mediante CNN entrenadas con imágenes de fondo de ojo previamente etiquetadas por expertos en activas e inactivas. Se utilizaron arquitecturas de CNN previamente conocidas mediante técnicas de transfer learning y arquitecturas de CNN construidas manualmente. Adicionalmente, para cada modelo se realizó un ajuste de hiperparámetros incluyendo el ajuste del optimizador, tasa de entrenamiento, número de épocas, dropout, función de activación y tamaño de lote. Finalmente todos los modelos fueron evaluados mediante técnicas de validación Hold-Out aleatoria con semilla.

2 DEFINICIÓN DEL PROBLEMA

La TO es una infección causada por el parásito Tg, un organismo intracelular apicomplejo capaz de infectar a una amplia gama de especies, incluidos los humanos [6]. La transmisión de este parásito se facilita principalmente a través del consumo de agua no tratada, frutas y verduras sin lavar, y carnes mal cocidas [7]. Una vez dentro del huésped, el Tg tiene una afinidad particular por el sistema nervioso central y la retina, donde puede causar toxoplasmosis cerebral y ocular, respectivamente [7]. En el ojo, la infección activa se manifiesta en forma de lesiones retinocoroideas blanco-amarillentas que, aunque tratables, dejan cicatrices que comprometen permanentemente la visión [6]. Los quistes retinianos residuales representa una amenaza latente de recurrencia y deterioro visual adicional [2].

La severidad de la TO varía considerablemente, influenciada por factores genéticos del parásito y la respuesta inmune del huésped [8], [9]. En Colombia, la situación es particularmente grave debido a la alta seroprevalencia del Tg, que alcanza hasta un 47% en la población general [1], una creciente incidencia [10] y la presencia de cepas atípicas y altamente virulentas de Tg que son más comunes en el hemisferio sur (I, III, atípicas) [11], [12]. Estas cepas han demostrado ser capaces de alterar la respuesta inmune del huésped de una respuesta TH1 a una TH2, exacerbando la severidad de la infección y sus consecuencias clínicas [8].

2.1 PLANTEAMIENTO DEL PROBLEMA

La toxoplasmosis es una gran problemática en salud en Colombia, por la alta seroprevalencia del Tg contribuyendo a muchos de los casos de ceguera, particularmente entre los niños [1], [13]. La enfermedad es responsable del 37.8% de los casos de ceguera legal infantil [13]. No obstante, en adultos también provoca una gran pérdida visual y afección de la calidad de vida [3]. Esto subraya la necesidad urgente de mejorar las estrategias de diagnóstico y tratamiento.

El diagnóstico de la TO es clínico, tiene una sensibilidad entre el 71-96% y especificidad del 58-100%) entre especialistas de uveítis [14]. Los hallazgos clínicos se apoyan de los resultados serológicos y de procedimientos invasivos como la PCR [15], [16]. Sin embargo, estas ayudas diagnósticas tienen limitaciones dado la persistencia de marcadores serológicos y la variabilidad en la sensibilidad de las pruebas moleculares [15], [16]. La serología puede ser engañosa, con

una IgM que permanece positiva meses después de la infección, dificultando la confirmación de casos recientes [16], [17]. La PCR, aunque específica, tiene un rendimiento variable y a menudo insatisfactorio, especialmente en pacientes inmunocompetentes [15].

La inteligencia artificial (IA) ha emergido como una herramienta potencialmente revolucionaria en el campo del diagnóstico médico. Estudios preliminares utilizando algoritmos de machine learning y CNNs han demostrado una capacidad diagnóstica en la identificación de la TO, con niveles de exactitud (accuracy), sensibilidad y especificidad que superan los métodos tradicionales [18], [19], [20]. Sin embargo, aún no se ha explorado suficientemente la capacidad de la IA para evaluar la actividad de las lesiones retinocoroideas entre activas e inactivas, un factor crítico que podría influir significativamente en el manejo de la enfermedad. Nuestro proyecto tuvo como objetivo abordar esta laguna crítica en la identificación automática de la actividad de la TO mediante CNNs entrenadas con imágenes de fondo de ojo. Este enfoque tiene el potencial de servir como un valioso complemento a la experiencia clínica, optimizando así la toma de decisiones médicas en el abordaje de pacientes con TO.

2.2 FORMULACIÓN DEL PROBLEMA

En el presente proyecto se abordó si era posible identificar automáticamente la actividad de la TO mediante el entrenamiento de CNN's utilizando imágenes de fondo de ojo. Adicionalmente, se planteó cómo gestionar una base de datos compuesta por imágenes de fondo de ojo etiquetadas según la actividad de la TO, activa y no activa. También se exploraron cuáles algoritmos o arquitecturas de CNN se podían implementar para la detección de toxoplasmosis a partir de estas imágenes, considerando el fenómeno de sobreajuste. Adicionalmente, se investigaron las métricas de clasificación más adecuadas para evaluar el rendimiento de las arquitecturas de CNN implementadas.

2.2.1. Pregunta de investigación

¿Es posible identificar automáticamente la actividad de la toxoplasmosis mediante el entrenamiento de CNNs con imágenes de fondo de ojo?

2.3 JUSTIFICACIÓN

La TO representa una de las principales causas de discapacidad visual en regiones como Colombia [13], donde la seroprevalencia del parásito Tg es particularmente alta [4]. A pesar de los avances en el diagnóstico clínico, la identificación de la actividad de las lesiones retinocoroideas sigue siendo un desafío significativo debido a la variabilidad en la presentación clínica [6] y las limitaciones de las pruebas serológicas y moleculares existentes [9], [21], [22]. La implementación de CNN para la clasificación automática de imágenes de fondo de ojo ofrece una solución innovadora que puede mejorar la capacidad diagnóstica y la toma de decisiones terapéuticas, especialmente en áreas con acceso limitado a especialistas en oftalmología. Este proyecto no solo tiene el potencial de apoyar el manejo clínico de la TO, sino que también podría servir como un modelo para la aplicación de IA en otras enfermedades inflamatorias e infecciosas oculares, marcando un avance significativo en la integración de tecnologías emergentes en la práctica médica.

3 OBJETIVOS DEL PROYECTO

3.1 OBJETIVO GENERAL

Identificar automáticamente la actividad de la TO mediante CNN entrenadas con imágenes de fondo de ojo.

3.2 OBJETIVOS ESPECÍFICOS

- Gestionar una base de datos compuesta por imágenes de fondo de ojo que sean etiquetadas por un especialista en dos categorías: toxoplasmosis activa y no activa.
- Realizar el entrenamiento de arquitecturas de CNN con imágenes de fondo de ojo para identificar la actividad de la toxoplasmosis aplicando estrategias para reducir el sobreajuste.
- Evaluar el rendimiento de las técnicas de clasificación usando métricas establecidas en el estado del arte para clasificación.

5. MARCO TEÓRICO Y ANTECEDENTES

5.1. MARCO TEÓRICO

5.1.1. Toxoplasmosis Ocular

5.1.1.1. Contexto Epidemiológico y Ciclo de Vida del Parásito

La TO se destaca como la etiología predominante de uveítis posterior y es causada por la infección del protozoo *Tg*, perteneciente a la familia Apicomplexa [3]. Este parásito posee la notable capacidad de infectar una amplia variedad de células en mamíferos y aves. Su ciclo de vida comprende tres fases distintas: **(i) el taquizoíto**, que se caracteriza por una rápida multiplicación y es prevalente durante las etapas agudas de la infección en el huésped intermediario, como los humanos; **(ii) el bradizoíto**, que se replica más lentamente y es típico de la fase crónica, formando quistes tisulares; y **(iii) el esporozoíto**, que se desarrolla únicamente en el huésped definitivo, los felinos, y se excreta en los oocistos a través de las heces (**Figura 1**) [3].

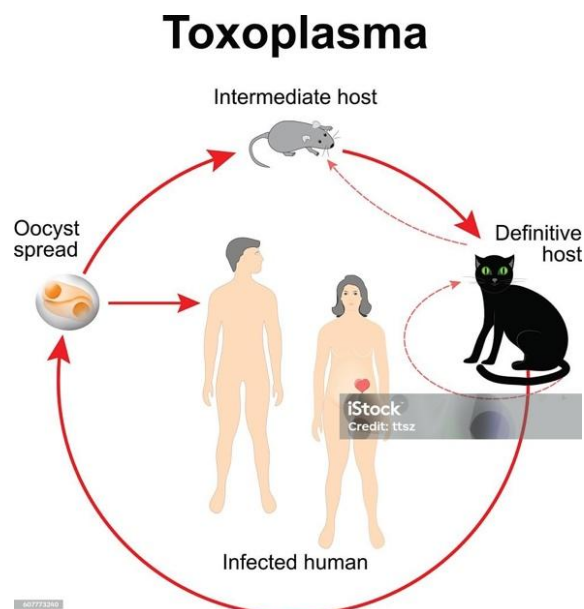


Figura 1: Ciclo de vida del *Toxoplasma gondii*.

Tomado de Istock [23].

La transmisión de Tg puede ser congénita, a través de la transmisión vertical durante el embarazo, o adquirida después del nacimiento. Clínicamente, la manifestación más frecuente es la retinocoroiditis, que suele ser unilateral y unifocal, y comúnmente se acompaña de vitreítis [24]. La variabilidad en la presentación clínica puede depender de si la infección es congénita o adquirida. Los principales vectores de infección incluyen la ingestión de alimentos y agua contaminados con ooquistes de origen felino o carne que contiene quistes tisulares [6]. La TO es reconocida como un problema de salud pública significativo debido a su potencial para causar discapacidad visual en individuos en su etapa productiva de la vida [2].

5.1.1.2. Prevalencia e Incidencia Global

La seroprevalencia de Tg varía geográficamente, con estimaciones que indican una seropositividad de entre el 12% y el 43% en países de bajos ingresos, como aquellos en América del Sur, América Central, el Caribe y África [25], [26], [27], [28]. Estas cifras son considerablemente más altas que las reportadas en países de altos ingresos, como los de Europa y Estados Unidos, donde la prevalencia oscila entre el 1% y el 39% [29], [30], [31]. En Colombia, la seroprevalencia en la población general se reporta en un 47.1% [1], [4].

Es importante destacar que, aunque la seropositividad para Tg muestra una prevalencia e incidencia más elevadas en países suramericanos en comparación con otras regiones del mundo, solamente el 6% de los pacientes seropositivos desarrollan lesiones retinocoroideas, es decir, TO. [1] Un estudio reciente por Cifuentes y colaboradores estima una incidencia de solamente 42 casos de TO por cada 100.000 pacientes con toxoplasmosis [10].

Estas estadísticas convierten a la TO en una enfermedad poco conocida entre los oftalmólogos; un estudio realizado en 2009 en el contexto colombiano indica que la uveítis representa solo el 0.006% de todas las consultas oftalmológicas [25]. Esto se correlaciona con la baja prevalencia e incidencia de uveítis posterior en Colombia (8.06 y 7.66 por 100,000 habitantes) [32] y el hecho de que no todos los casos de uveítis posterior son secundarios a TO. En Colombia, del total de pacientes con uveítis, el 27.73% resultan ser idiopáticas y apenas el 25.34% corresponden a TO. [33] Por lo tanto, la TO no es un diagnóstico frecuente, incluso para un especialista en oftalmología, y generalmente requiere la intervención de un especialista en uveítis.

5.1.1.3. Evaluación Clínica y Diagnóstico Diferencial

El diagnóstico clínico de la TO ha reportado tener un rendimiento variable, con sensibilidades entre el 71-96% y especificidades entre 58-100% entre especialistas de uveítis [14]. Estas cifras demuestran la necesidad de apoyar el diagnóstico clínico con marcadores serológicos, imagen multimodal y procedimientos invasivos como la PCR que se explican más adelante. El diagnóstico clínico se fundamenta en la evaluación oftalmológica detallada, donde se identifican síntomas como disminución de la agudeza visual, visión borrosa, y en casos menos frecuentes, defectos en los campos visuales y alteraciones en la percepción de colores [15]. Estos hallazgos clínicos deben correlacionarse con lesiones identificadas en el fondo de ojo que se explican a continuación.

5.1.1.4. Lesiones típicas en toxoplasmosis congénita

En la toxoplasmosis congénita, las lesiones retinocoroideas se presentan en aproximadamente el 80% de los casos durante los primeros años de vida. Estas pueden manifestarse como lesiones exudativas activas o como cicatrices con pigmentación variable en estados inactivos, con una prevalencia mayor en el polo posterior y afectación macular en comparación con las infecciones adquiridas, así como una incidencia más alta de bilateralidad [34]. Un patrón distintivo de las lesiones congénitas es la cicatriz en forma de "rueda de carreta" en la región macular, caracterizada por una destrucción tisular central con pigmentación centripeta en sus bordes.

5.1.1.5. Lesiones típicas en toxoplasmosis adquirida

En el caso de infecciones adquiridas, la retinocoroiditis se manifiesta en hasta un 20% de los pacientes con primoinfección [35]. Se caracteriza por exudados superficiales blancos-amarillentos o grisáceos, de morfología redondeada o irregular, de tamaño variable y con bordes difusos rodeados de edema retiniano. Estas lesiones suelen resolverse espontáneamente en individuos inmunocompetentes en cicatrices retinocoroideas, independientemente del tratamiento, al igual que pueden reactivarse en cualquier momento (**Figura 2**).

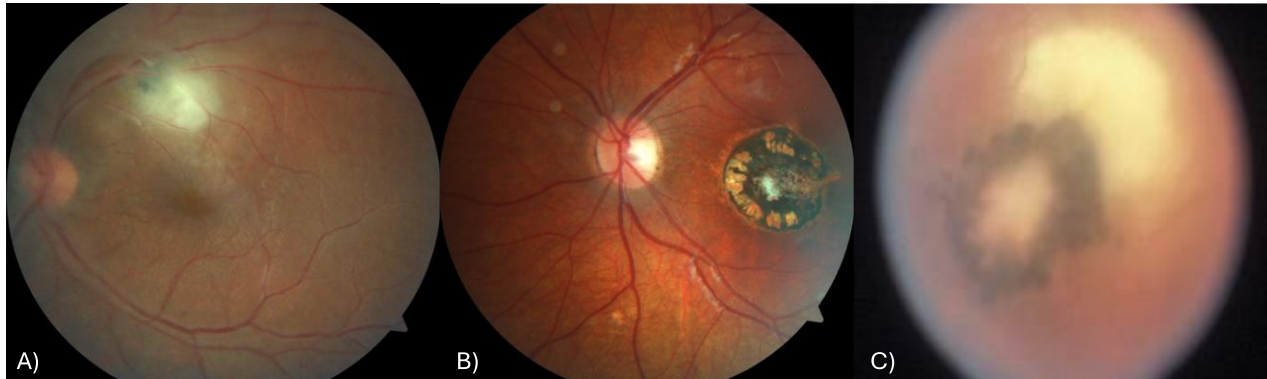


Figura 2: Lesiones típicas de toxoplasmosis ocular

Fotos de fondo de ojo con A) lesión blanco-amarillenta de $1 \frac{1}{2}$ diámetro de disco (DD) en arcada temporal superior secundaria a toxoplasmosis ocular (TO). B) Cicatriz retinocoroidea macular de 2DD inactiva C) Cicatriz retinocoroidea con bordes hiperpigmentados y centro atrófico con lesión satélite en borde temporal superior que indica una nueva lesión activa, Tomado de la presentación de de-la-Torre A en las grandes rondas del Grupo Internacional de Estudio de la Uveítis (IUSG) [36].

5.1.1.6. Lesiones atípicas de toxoplasmosis ocular

Aunque la TO se presenta típicamente como una retinocoroiditis focal con vitreítis asociada, también puede manifestarse de maneras atípicas, lo que puede dificultar su diagnóstico y tratamiento [37]. Las presentaciones atípicas incluyen lesiones extensas de retinitis necrotizante, múltiples focos de inflamación o retinitis periférica, que no se ajustan a la imagen clínica clásica de la enfermedad [38]. Estas formas atípicas pueden estar asociadas con una respuesta inmunológica alterada, como ocurre en pacientes inmunocomprometidos [9], donde la enfermedad puede progresar rápidamente y afectar grandes áreas de la retina, lo que aumenta el riesgo de ceguera [2].

Entre las presentaciones atípicas, una de las más destacadas es la denominada toxoplasmosis retiniana externa puntiforme. Esta variante, que se caracteriza por pequeñas lesiones multifocales localizadas en las capas profundas de la retina, es menos común y se asocia tanto a infecciones congénitas como postnatales [39]. A diferencia de las formas clásicas de TO, en la toxoplasmosis retiniana externa puntiforme la inflamación vítrea suele ser leve o ausente, y las lesiones pueden confundirse con otras patologías oculares como los síndromes de puntos blancos. Debido a su

presentación inusual, el diagnóstico requiere una evaluación detallada con técnicas de imagen multimodal, incluyendo la tomografía de coherencia óptica y la autofluorescencia [40].

Otra presentación atípica de la TO es la neuroretinitis, en la cual el nervio óptico se ve afectado de manera significativa, con edema del disco óptico y exudados duros que forman un patrón estrellado en la mácula [38]. Aunque esta forma es más comúnmente asociada con infecciones por Bartonella, también ha sido documentada en casos de TO. En estos casos, la neuroretinitis puede presentarse como una manifestación aislada o en combinación con una retinocoroiditis típica [38]. Esta presentación requiere un enfoque terapéutico intensivo, dado el riesgo elevado de daño visual permanente [2] (**Figura 3**).

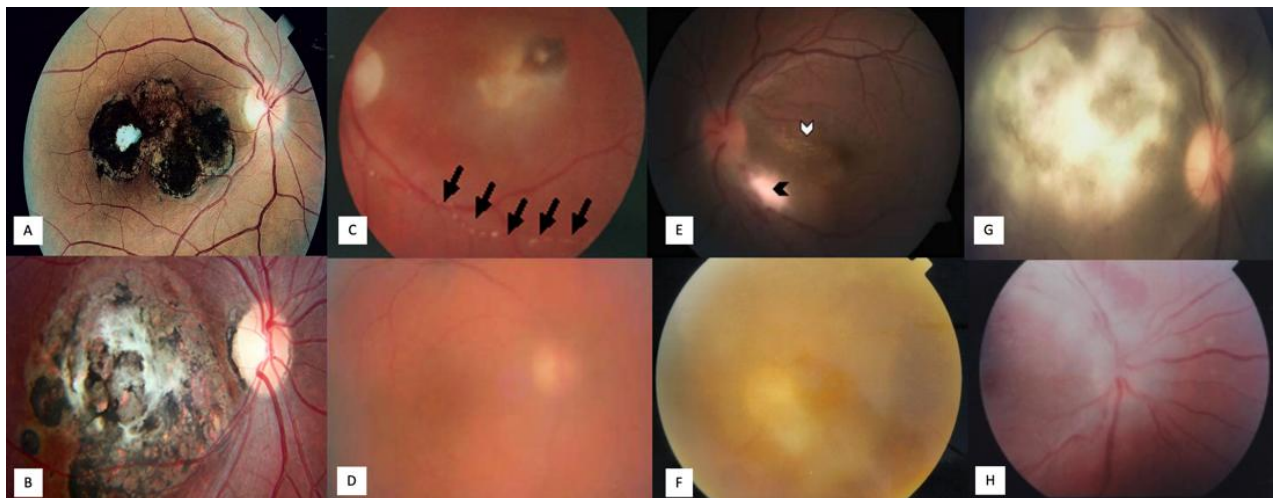


Figura 3: Lesiones atípicas de toxoplasmosis ocular

A) Cicatrices retinocoroideas maculares extensas, B) Fenómeno de Kyrieleis, D) Vitreítis, E) Neuroretinitis, F) Signo del faro en la niebla sugestivo de lesiones activas, G) Coriorretinitis extensa multifocal, H) Papilitis. Tomado de la presentación de de-la-Torre A en las grandes rondas del Grupo Internacional de Estudio de la Uveítis (IUSG) [36].

5.1.1.7. Reactivación de la Enfermedad

Durante la reactivación de la TO, las lesiones satélites constituyen el signo clínico más característico (**Figura 2C**). Sin embargo, en situaciones de inflamación intensa, la visualización de las lesiones puede verse obstaculizada por una vitreítis pronunciada, dando lugar a un fenómeno conocido como "faro en la niebla" (**Figura 3F**) [41]. Además, pueden presentarse otros signos como edema macular, precipitados inflamatorios en la hialoides posterior y afectación vascular, esta última evidenciada por hemorragias o vasculitis, también denominada arteriolitis

de Kyrieleis (**Figura 3C**), ya sea adyacente o distante a la lesión exudativa primaria [41].

5.1.1.8. Exámenes diagnósticos adicionales

Diversas técnicas se han desarrollado para la detección de Tg en tejidos humanos, abarcando desde evaluaciones histopatológicas hasta análisis serológicos. Entre estas, se encuentran la tinción de Sabin-Feldman, la inmunofluorescencia indirecta (IFI), el ensayo por inmunoabsorción ligado a enzimas (ELISA), los ensayos de quimioluminiscencia, la técnica de aglutinación modificada (MAT), la prueba de avidéz de IgG en suero, la PCR y la medición de producción de anticuerpos intraoculares con el coeficiente de Goldman Witmer [15], [17], [37]. Estos últimos estudios aunque no se realizan en todos los casos se consideran como las pruebas más específicas tanto para el diagnóstico como para la definición de actividad de TO [37]. No obstante, su efectividad es variable reportando para la PCR una sensibilidad de 2.3–75% y una especificidad de 75-100% [21], [37], [42], [43], [44], [45] y para el coeficiente de Goldman Witmer de 59% que alcanza el 71% cuando se suma con otras técnicas como el Western blot [5], [22]; lo que confirma que tanto el diagnóstico como la determinación de la actividad de la TO es un verdadero reto diagnóstico

De todos estos la serología es el más ampliamente utilizado. En el que la fase aguda de la infección, los anticuerpos séricos IgM e IgG contra Tg son indicativos de la actividad del parásito. Los niveles de IgM se elevan desde la primera semana y pueden persistir hasta 6 a 9 meses después de la infección aguda, en ocasiones hasta por 2 años. Por otro lado, los niveles bajos de IgG en suero sugieren una infección inactiva. Sin embargo, en pacientes inmunosuprimidos, la respuesta de IgG puede ser atenuada, requiriendo pruebas con mayor especificidad clínica [15], [46]. La decisión diagnóstica final de la TO y la clasificación de su actividad depende de una combinación de factores y pruebas en donde la sospecha y el criterio clínico es la más importante [14].

5.1.1.9. Tratamiento de la toxoplasmosis ocular

El tratamiento de la TO es desafiante debido a la ausencia de un protocolo terapéutico estandarizado y a la incapacidad de las terapias actuales para erradicar los bradizoitos enquistados en los tejidos retinianos [37]. Los oftalmólogos utilizan una amplia gama de regímenes terapéuticos, ajustando los tratamientos según las características clínicas del paciente y la gravedad de la infección. Una revisión sistemática de la literatura y una encuesta a expertos en uveítis han identificado 24 regímenes distintos empleados para el manejo de la TO. Sin embargo, la efectividad de estos tratamientos sigue siendo objeto de controversia, y ningún enfoque ha demostrado ser claramente superior en cuanto a prevenir las recurrencias o mejorar los resultados visuales a largo plazo [47]. Algunos estudios soportan el uso de trimetoprim/sulfametoxazol (TMP-SMX) por ser la terapia más costo efectiva en comparación con otros esquemas terapéuticos [48].

Los regímenes terapéuticos más comúnmente empleados incluyen una combinación de pirimetamina, sulfadiazina y ácido fólico. La pirimetamina, que inhibe la síntesis de ácido fólico en Tg, se administra inicialmente en dosis de 100 mg el primer día, seguido de una dosis de mantenimiento de 25 a 75 mg diarios durante 4 a 6 semanas, o hasta que la lesión se haya inactivado completamente [37]. La sulfadiazina se administra en dosis de 4 g diarios, dividida en cuatro dosis, y el ácido fólico es añadido para prevenir la toxicidad hematológica asociada con la pirimetamina [49]. Este régimen, a pesar de ser efectivo en la eliminación de los taquizoitos, tiene limitaciones significativas debido a los efectos secundarios sistémicos que pueden incluir náuseas, vómitos, supresión medular y erupciones cutáneas [37], [49].

Para aquellos pacientes que no toleran la sulfadiazina, el trimetoprim/sulfametoxazol (TMP-SMX) es una alternativa eficaz. Este régimen se administra en dosis de 160 mg/800 mg dos veces al día durante 4 a 6 semanas. Aunque TMP-SMX tiene un perfil de efectos secundarios más favorable, su acción está igualmente limitada a los taquizoitos, sin efecto sobre los bradizoitos enquistados, lo que no evita las recurrencias a largo plazo [37], [49].

El uso de corticosteroides orales, como la prednisolona, es común en casos de inflamación severa, particularmente cuando las lesiones afectan zonas de alto riesgo como la mácula o el nervio óptico. La administración de prednisolona se inicia generalmente 48 horas después del

comienzo de la terapia antimicrobiana, a dosis de 0,5 a 1 mg/kg/día, y se reduce de forma gradual para evitar recaídas [49]. Este enfoque ayuda a controlar la respuesta inflamatoria sin potenciar la replicación de los taquizoitos, pero también se asocia con riesgos, como el aumento de la presión intraocular y el desarrollo de cataratas [49].

En los últimos años, se ha comenzado a explorar el uso de tratamientos intravítreos, como la administración de clindamicina y dexametasona intravítreas, para minimizar los efectos secundarios sistémicos. La clindamicina intravítrea (1 mg/0,1 mL) en combinación con dexametasona (400 µg/0,1 mL) puede ser una opción en casos donde la terapia sistémica esté contraindicada o en pacientes con efectos adversos graves [37]. Sin embargo, este tratamiento no debe emplearse como monoterapia, ya que es insuficiente para controlar la replicación de los taquizoitos a nivel sistémico [37].

Por último, la profilaxis con trimetoprim/sulfametoxazol tres veces por semana se recomienda para prevenir recurrencias en pacientes con lesiones en la mácula o en el nervio óptico que amenacen la visión, o en aquellos con reactivaciones frecuentes [50]. Su uso adecuado ha mostrado la capacidad de reducción de recurrencias en un 82% durante el primer año y 87% durante el segundo año [2].

5.1.2. Inteligencia Artificial

La IA se define como la capacidad de un sistema informático, ya sea un ordenador, una red de ordenadores o una red de robots controlados por ordenadores, para llevar a cabo tareas que típicamente requieren inteligencia humana [51]. Esta disciplina de la informática se enfoca en la simulación de comportamientos inteligentes, permitiendo a las máquinas procesar información y realizar actividades como el aprendizaje, la planificación, la resolución de problemas, el razonamiento y la identificación de patrones, de manera similar a como lo haría la mente humana. La IA no solo replica la capacidad cognitiva del ser humano para entender y razonar de forma compleja, sino que también automatiza estos procesos, mejorando su eficiencia y exactitud (accuracy) mediante variaciones en los datos y algoritmos [51].

5.1.2.1. Aprendizaje supervisado

El aprendizaje supervisado es una técnica de aprendizaje automático que utiliza conjuntos de datos etiquetados para entrenar un modelo capaz de predecir resultados específicos [52]. Este enfoque está inspirado en el proceso de aprendizaje humano y es ampliamente utilizado en el aprendizaje automático. Se entrena al modelo utilizando conjuntos de datos etiquetados, ya sea de forma continua o discreta, y se requiere una supervisión adecuada a través de un modelo de entrenamiento que se ajuste a las características del conjunto de datos. El aprendizaje supervisado se utiliza comúnmente en tareas de regresión y clasificación, donde su objetivo es predecir con exactitud los resultados mediante el uso de datos previamente conocidos [52]. Así, el aprendizaje supervisado parte del conocimiento de un conjunto de entrenamiento que incluye datos compuestos por variables explicativas o características y la variable respuesta. El objetivo es desarrollar modelos predictivos que, una vez entrenados, puedan predecir la variable respuesta de nuevos datos, utilizando únicamente sus características explicativas [53].

5.1.2.2. Tipos de aprendizaje supervisado

El aprendizaje supervisado se aplica en diversos problemas, como la regresión y la clasificación. Dentro de los modelos de regresión se incluyen la regresión lineal, la regresión logística, y el *Support Vector Regression* (SVR), cuyo objetivo es predecir un valor numérico continuo basado en una o más variables explicativas [54], [55]. Estos modelos establecen una relación matemática entre las variables independientes y la variable dependiente. Por otro lado, los modelos de clasificación asignan entradas a categorías discretas. Un modelo de clasificación aprende a predecir la clase de una nueva observación basándose en ejemplos previos con categorías conocidas. Algunos de los modelos más comunes incluyen las Máquinas de Soporte Vectorial, los Árboles de Decisión, el algoritmo K-Nearest Neighbors y el clasificador Naive Bayes [54], [55].

5.1.2.3. Aplicaciones del aprendizaje supervisado en Medicina

El aprendizaje supervisado ha demostrado ser una herramienta poderosa en la medicina, especialmente en el pronóstico y diagnóstico de enfermedades como el cáncer. Por ejemplo, Duo et al. utilizaron modelos de regresión logística, árboles aleatorios y máquinas de vectores de soporte (SVM) para desarrollar un modelo predictivo de la recurrencia del cáncer de mama, logrando una mayor exactitud gracias a la inclusión de datos clínicos y genómicos [56]. De manera similar, Deng et al. aplicaron estos mismos enfoques para el pronóstico del cáncer de próstata, mientras que Yawei et al. desarrollaron modelos predictivos para la detección temprana del cáncer de pulmón [57], [58]. Además, Sorayaie et al. utilizaron algoritmos de clasificación, como SVM, árboles de decisión, bosques aleatorios y XGBoost, para predecir la supervivencia de pacientes con cáncer de ovario, obteniendo predicciones con una exactitud superior al 80% [59].

5.1.2.4. Aprendizaje profundo

El aprendizaje profundo, una rama del aprendizaje automático, se basa en el uso de múltiples capas de redes neuronales para extraer características clave de grandes volúmenes de datos. Técnicas como las redes neuronales convolucionales, recurrentes y profundas son fundamentales en el procesamiento de datos complejos. Estas redes imitan la estructura del cerebro humano, permitiendo el procesamiento y la clasificación de datos a través de múltiples capas de nodos interconectados [60], [61].

5.1.2.5. Redes neuronales convolucionales

Las redes neuronales convolucionales, también conocidas como CNN, son algoritmos que forman parte del ámbito del aprendizaje automático. Su función principal radica en procesar imágenes como entrada, identificar una serie de características dentro de cada imagen y, basándose en estas características, tienen la capacidad de distinguir unas imágenes de otras, pudiendo realizar tareas de clasificación [62].

Las CNN representan un ejemplo de una arquitectura de red neuronal especializada que incorpora conocimiento acerca de la invarianza de formas bidimensionales mediante el uso de patrones de conexión local y restricciones en los pesos [63]. Su inspiración proviene de la

organización de la corteza visual de los animales [64]. En esta organización, la respuesta de cada neurona se describe matemáticamente como una convolución. Cada neurona en la corteza visual responde a estímulos externos dentro de una región espacial específica conocida como campo receptivo. Las CNN constituyen una variante de las redes neuronales profundas, originada a partir de la adaptación del perceptrón multicapa. Se distinguen por su uso de capas convolucionales y submuestreo [65].

Desde una perspectiva conceptual, se puede entender una red neuronal convolucional como un filtro en movimiento que recorre las imágenes recibidas, interactuando a una proximidad predefinida con las neuronas. Estas redes, en su mayoría, se inician con una imagen y constan de capas alternadas de convolución y submuestreo (pooling) para realizar predicciones, proporcionando una probabilidad de acierto o un porcentaje de similitud. La dimensión de entrada en las CNN varía según si la imagen recibida es a color (tridimensional) o en blanco y negro (bidimensional). La adaptación del núcleo de la red (las capas ocultas) se ajusta según este número de dimensiones, siendo este un aspecto crucial [66].

La convolución, realizada de manera repetida en esta capa, busca extraer características de alto nivel de la entrada. En las primeras capas, se identifican características de bajo nivel, mientras que capas adicionales combinan estas para lograr características de alto nivel. Posteriormente, la fase de pooling, que sigue a la convolución, reduce las dimensiones espaciales de la entrada mediante submuestreo o resumen estadístico. Este proceso conlleva ventajas como la reducción de recursos computacionales, la identificación de características clave y la eliminación de información redundante para evitar el sobreajuste en la salida (**Figura 4**) [67].

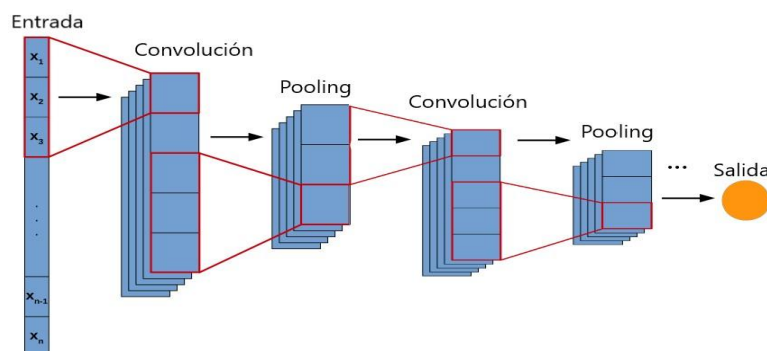


Figura 4: Arquitectura de las redes convolucionales

Tomado de Lopez-Santader y colaboradores [67].

5.1.2.6. ResNetV2

La ResNetV2 es una arquitectura de red neuronal convolucional que se basa en la familia de arquitecturas Inception, pero introduce la incorporación de conexiones residuales [68]. A diferencia de las arquitecturas tradicionales de Inception, que emplean la concatenación de filtros para combinar características, ResNetV2 utiliza estas conexiones residuales para facilitar el flujo de la información a lo largo de la red. Estas conexiones permiten que la información salte a través de múltiples capas, lo que reduce la posibilidad de pérdida de gradiente y mejora tanto la eficiencia del entrenamiento como la capacidad de la red para aprender características más complejas. Este enfoque no solo acelera el proceso de entrenamiento, sino que también mejora el rendimiento general de la red en tareas de clasificación de imágenes (**Figura 5**) [68].

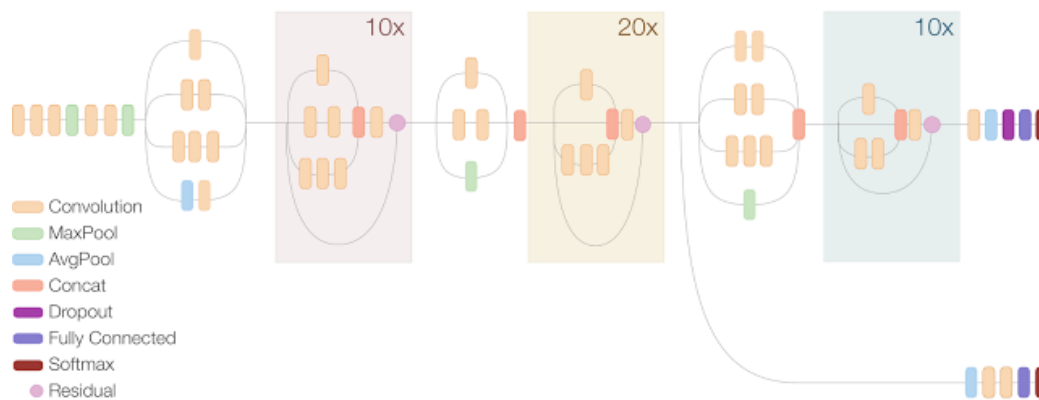


Figura 5: Arquitectura de la red Inception Resnet V2.

Tomado de Github.com [69]

5.1.2.7. VGG16

La VGG16 es una red neuronal convolucional ampliamente utilizada para la clasificación de imágenes, compuesta por 16 capas de neuronas artificiales que procesan las imágenes de manera gradual, lo que mejora la exactitud en la clasificación. Esta arquitectura se caracteriza por emplear capas convolucionales con filtros de tamaño 3x3 y un paso (stride) de 1, manteniendo un *padding* que preserve las dimensiones de la imagen [70], [71]. Además, utiliza capas de max

pooling con filtros de 2×2 y un paso de 2 para reducir las dimensiones de las representaciones intermedias sin perder información relevante. Este patrón de capas de convolución seguidas por capas de max pooling se repite a lo largo de toda la red. Finalmente, la VGG16 incluye dos capas completamente conectadas antes de la capa *softmax*, que genera la salida final encargada de asignar la clase a la imagen procesada (**Figura 6**) [70], [71].

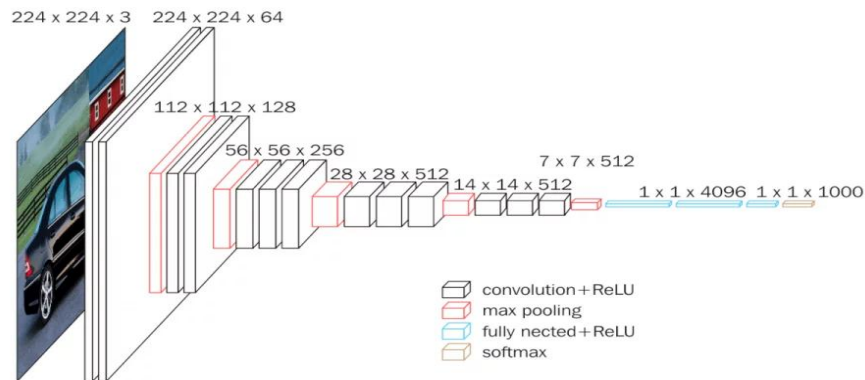


Figura 6: Arquitectura de la red VGG-16

Tomado de Hassan Muneeb [72].

5.1.2.8. InceptionV3

La red neuronal **InceptionV3** emplea módulos denominados *Inception*, los cuales aplican de manera simultánea múltiples filtros de diferentes tamaños a una misma entrada, a través de capas convolucionales y de pooling. Este enfoque permite a la red detectar patrones a distintas escalas y combinar los resultados en la salida del módulo, lo que mejora su capacidad para reconocer características complejas en las imágenes. Aunque el diseño de *Inception* incrementa el número de parámetros y la carga computacional del modelo, la red incorpora técnicas de optimización, como la factorización de convoluciones, que mejoran la eficiencia sin comprometer la exactitud. Esto permite que **InceptionV3** mantenga un rendimiento competitivo en tareas de clasificación de imágenes, incluso con un mayor número de parámetros (**Figura 7**) [73], [74].

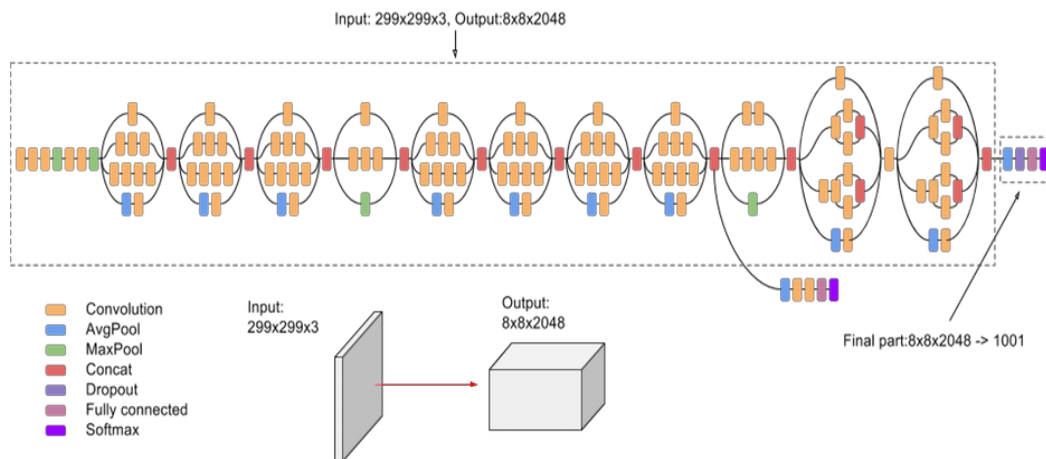


Figura 7: Arquitectura de la red InceptionV3

Tomado de Szegedy y colaboradores [75]

5.1.2.9. Aplicaciones del Aprendizaje Profundo en Medicina

El aprendizaje profundo ha sido ampliamente utilizado en medicina para la clasificación de imágenes en diversas enfermedades, facilitando un mejor diagnóstico para los especialistas de la salud. Ghulam et al. aplicaron CNNs para la clasificación del cáncer de mama utilizando imágenes de mamografías e histopatología, aprovechando su capacidad para extraer características complejas de las imágenes médicas. Se utilizaron tanto arquitecturas preentrenadas, como VGG y ResNet, así como redes personalizadas diseñadas específicamente para esta tarea, logrando altos niveles de exactitud y área bajo la curva (AUC) en la diferenciación entre casos de cáncer y no cáncer [76].

Asimismo, Xue et al. desarrollaron una red neuronal completamente convolucional llamada F2 FCN, que fue utilizada para la segmentación de tumores cerebrales a nivel de píxeles. Este modelo implementó un sistema de membrana de hipérgrafo para ejecutar múltiples redes en paralelo, logrando un rendimiento superior en comparación con otras redes de última generación como U-Net y FR FCN, especialmente en los conjuntos de datos BRATS2015 y BRATS2013 [77].

Operaciones técnicas fundamentales en las redes neuronales convolucionales

Las operaciones técnicas fundamentales en las CNNs son esenciales para el análisis y procesamiento de imágenes. La convolución es la operación principal en estas redes, donde se aplica un filtro (o kernel) sobre una imagen para extraer características locales, tales como bordes y texturas [63]. En el caso de las imágenes de fondo de ojo, esta técnica permite identificar patrones importantes, como aquellos que podrían indicar la actividad de la toxoplasmosis ocular (TO). Además, la convolución ayuda a reducir la dimensionalidad de las imágenes sin perder información relevante [63].

El pooling, particularmente el max pooling, se utiliza para reducir la dimensionalidad de las imágenes seleccionando el valor máximo dentro de una ventana específica [78]. Esto conserva las características más importantes y permite disminuir el tamaño de los mapas de características, facilitando su análisis y manteniendo los elementos clave para la clasificación de imágenes médicas, como las de fondo de ojo [78].

Las funciones de activación introducen no linealidad en el modelo, permitiendo que este capture relaciones complejas en los datos [63]. La función ReLU (Rectified Linear Unit) es una de las más utilizadas en CNN, ya que activa solo las neuronas con valores positivos, lo que mejora tanto la eficiencia computacional como la capacidad de las redes para identificar características significativas, como las presentes en lesiones activas e inactivas de TO [63].

La normalización por lotes (batch normalization) estandariza las entradas de cada capa, ajustándolas para que tengan una media cercana a cero y una varianza unitaria [78]. Este proceso estabiliza y acelera el entrenamiento, reduciendo el riesgo de sobreajuste, lo cual es especialmente útil cuando se trabaja con imágenes médicas de fondo de ojo que pueden tener variaciones en la intensidad de los píxeles [78].

La desactivación aleatoria (dropout) es una técnica de regularización que apaga de forma aleatoria un porcentaje de las neuronas durante el entrenamiento [63]. Esto ayuda a prevenir el sobreajuste y es crucial en redes profundas como las CNN, donde las imágenes de fondo de ojo son limitadas, permitiendo que el modelo generalice mejor [63].

El proceso de flattening convierte los mapas de características bidimensionales extraídos por las capas convolucionales en un vector unidimensional, facilitando la conexión con las capas

totalmente conectadas, que son responsables de la clasificación final. En el contexto de TO, este proceso es esencial para transformar las características extraídas de las imágenes en probabilidades clasificables [79].

Las capas totalmente conectadas toman las decisiones finales en la clasificación, procesando las características extraídas para predecir la probabilidad de que una lesión sea activa o inactiva en TO [78].

La capa softmax convierte las salidas de la red en probabilidades, asignando una probabilidad a cada clase. En el caso de TO, esta capa asigna probabilidades a las clases activa e inactiva, permitiendo hacer predicciones interpretables en términos probabilísticos [78].

El preprocesamiento de datos incluye técnicas de normalización y aumento de datos (data augmentation), como rotaciones y desplazamientos, lo que aumenta la cantidad de datos y reduce el riesgo de sobreajuste. Estas técnicas son fundamentales para asegurar la diversidad y mejorar el rendimiento del modelo en imágenes de fondo de ojo [79].

Finalmente, SMOTE es utilizado para equilibrar clases desbalanceadas mediante la creación de ejemplos sintéticos de la clase minoritaria, mejorando la sensibilidad del modelo en la detección de casos activos de TO [79].

5.2. ANTECEDENTES

5.2.1. Antecedentes de uso de CNN en oftalmología

La integración IA en la medicina ha marcado un hito transformador, particularmente en la oftalmología, una especialidad que se beneficia enormemente de las tecnologías de imagen. La naturaleza ambulatoria de la oftalmología, junto con la prevalencia de diagnósticos basados en imágenes, hace que esta especialidad sea un campo importante para la adopción de IA, con el objetivo de optimizar la atención al paciente [82].

Las CNN han demostrado ser herramientas valiosas en el análisis de imágenes de fondo de ojo. En 2020, Shaban et al. presentaron un modelo de CNN para analizar imágenes de fondo de ojo,

logrando diferenciar entre controles sanos, retinopatía diabética moderada y desprendimiento de retina severo con una exactitud (accuracy) y sensibilidad de aproximadamente 88-89% [83]. Asimismo, Brown et al. en 2018 validaron un algoritmo de CNN para diagnosticar la retinopatía del prematuro con una exactitud (accuracy) comparable o superior a la evaluación humana experta [84].

5.2.2. Antecedentes en uveítis y toxoplasmosis ocular

La IA ha demostrado un impacto prometedor en el diagnóstico y manejo de la uveítis [85]. Un estudio evaluó la capacidad diagnóstica y las recomendaciones de manejo proporcionadas por ChatGPT en pacientes con uveítis, comparándolas con las de oftalmólogos expertos. Los resultados mostraron que ChatGPT alcanzó una capacidad del 64% en la identificación del diagnóstico correcto en su primer intento, mejorando un 8% cuando se incluyeron diagnósticos diferenciales adicionales [86]. En comparación, los especialistas humanos obtuvieron una exactitud que varió entre el 60% y el 92% en los mismos escenarios. Además, ChatGPT mostró una concordancia del 91% con las recomendaciones de tratamiento proporcionadas por los expertos en los casos correctamente diagnosticados [86].

En cuanto al uso de análisis de imagen y uveítis, Du y colaboradores desarrollaron un sistema de aprendizaje profundo para la detección de retinitis por citomegalovirus mediante imágenes de fondo de ojo de campo ultra amplio. Este sistema alcanzó un AUC de 0.945 en la validación interna y demostró una excelente capacidad de clasificación en la validación externa, con una exactitud del 90%, una sensibilidad del 96.15% y una especificidad del 81.25%. En la validación prospectiva, el sistema mostró una exactitud del 94%, con una sensibilidad del 95.24% y una especificidad del 93.10%, superando los métodos tradicionales en la identificación de lesiones activas de retinitis por citomegalovirus [87].

En el contexto de la TO, se ha evaluado el potencial de las CNN para la identificación de lesiones de TO en distintos estudios Chakravarthy y colaboradores en 2019 realizaron por primera vez detección automática de lesiones retinocoroideas de TO en imágenes de fondo de ojo utilizando CNN [88]. Los investigadores utilizaron un modelo basado en la arquitectura VGG16 para clasificar imágenes en dos categorías: sanas y con toxoplasmosis. Se entrenó el modelo con un conjunto de datos que incluía imágenes etiquetadas por especialistas, logrando

exactitudes entre 0.81-0.85, sensibilidades entre 0.75-0.90 y especificidades entre 0.69-0.94 de acuerdo con la variación del muestreo entre los datos (30/70 vs. 50/50 vs. 70/30) demostrando el potencial de las CNN en el diagnóstico de la TO [88].

De manera similar en 2021, Parra y colaboradores utilizaron una arquitectura ResNet18 para el diagnóstico automático de la TO a partir de imágenes de fondo de ojo [19]. Utilizando un conjunto de datos compuesto por 160 imágenes capturadas en el Hospital de Clínicas en Asunción, Paraguay, Obteniendo con un optimizador tipo gradiente estocástico descendente, una tasa de optimización de $1e^{-2}$ y métodos de submuestreo una sensibilidad del 94% y una especificidad del 93%, afirmando la utilidad en el diagnóstico que tienen las arquitecturas de CNN en el diagnóstico de TO [19].

Recientemente en 2024 Milad y colaboradores evaluaron la efectividad del aprendizaje automático automatizado (AutoML) en comparación con modelos diseñados por expertos para la detección y localización de lesiones de TO utilizando imágenes de fondo de ojo [89]. Los autores utilizaron un conjunto de datos de 304 imágenes etiquetadas, donde los modelos AutoML fueron creados por oftalmólogos sin experiencia en programación. El modelo binario de AutoML mostró un área bajo la curva de precisión-recall (AuPRC) de 0.945, con una sensibilidad del 100%, especificidad del 83% y precisión del 93.5%, resultados comparables y superiores a los modelos diseñados por expertos previamente mencionados [19], [88]. Además, se desarrolló un modelo de detección de objetos para identificar visualmente las lesiones de TO, que alcanzó una precisión del 93.3% y un recall del 56%. Los autores concluyeron que las plataformas AutoML permiten a los clínicos crear modelos de inteligencia artificial efectivos sin conocimientos de programación, lo que democratiza el acceso a estas tecnologías en la práctica clínica [89].

6. Gestión de una base de datos

6.1. Metodología

6.1.1. Recopilación de Datos

Se obtuvieron imágenes de fondo de ojo de pacientes con diagnóstico confirmado de TO a partir de los registros clínicos de un consultorio privado de una especialista en uveítis. El diagnóstico

de TO se basó en distintos criterios según la presentación clínica. En los casos típicos, se determinó la presencia de cicatrices retinocoroideas y lesiones satélites características de la TO sin necesidad de realizar serología. Para las lesiones blanco-amarillentas aisladas, se exigió la confirmación mediante serología IgM positiva. En presentaciones atípicas, se recurrió a la serología para descartar diagnósticos alternativos; dado la alta seroprevalencia de Tg en Colombia [4] no se confirmó el diagnóstico exclusivamente con serología positiva sino que se requirió el apoyo con PCR de humor acuoso. Si no se disponía de PCR de humor acuoso, la confirmación del diagnóstico final se basó en las características clínicas, el curso de la enfermedad y la respuesta al tratamiento. Las imágenes obtenidas incluyeron tanto casos de TO activa, definidas por la presencia de lesiones satélites o lesiones únicas con serología IgM positiva, como casos de TO inactiva, caracterizados por cicatrices retinocoroideas sin signos de actividad inflamatoria o vitreítis.

6.1.2. Criterios de Calidad de imagen

Para garantizar la calidad de las imágenes incluidas en el estudio, se establecieron ciertos requisitos. Solo se incluyeron imágenes originales obtenidas directamente con equipos de alta calidad, como cámaras ZEISS Visucam y Optos. Se descartaron aquellas imágenes que presentaban baja resolución (capturas de pantalla, escaneo de fotos originales), sobreexposición, subexposición o artefactos que impedían una correcta visualización de las estructuras oculares. Sin embargo, dado que el objetivo principal era identificar la actividad de las lesiones, las imágenes en las que la visibilidad del fondo de ojo estaba comprometida por vitreítis, un signo de actividad inflamatoria, fueron incluidas en el análisis.

6.1.3. Proceso de etiquetado por especialistas

Las imágenes recopiladas fueron evaluadas inicialmente por un grupo de cuatro expertos, quienes clasificaron cada imagen como TO activa o inactiva, basándose en las características clínicas observadas en las mismas. En los casos en los que no fue posible llegar a una clasificación concluyente con base en la imagen, se decidió catalogarlas como indeterminadas. Posteriormente, la Dra. Alejandra de-la-Torre, realizó una segunda revisión de todas las imágenes de manera cegada, es decir, sin conocer las decisiones previas de los expertos. En los casos en

los que la evaluación inicial de los expertos coincidía con la revisión de la Dra. De-la-Torre, se consideró que el diagnóstico era definitivo. Cuando hubo discordancia entre la primera clasificación y la revisión de la Dra. de la Torre, las imágenes fueron discutidas en reuniones de consenso entre los autores hasta alcanzar una decisión final.

La Dra. Alejandra De la Torre es una oftalmóloga con amplia experiencia en inmunología ocular y uveítis. Obtuvo su doctorado en Medicina Traslacional, con un enfoque en la inmunopatología y la virulencia de la TO, en el Institut de Parasitologie et de Pathologie Tropical de la Université de Strasbourg, Francia. Además, cuenta con un doctorado en Ciencias Biomédicas otorgado por la Universidad del Quindío, Colombia. Ha sido reconocida mundialmente como la experta número uno en TO a nivel mundial en 2020 por AD Scientific Index [90], con numerosos premios y menciones y más de 200 artículos en revistas indexadas. Actualmente, se desempeña como profesora y jefa de la Unidad de Inmunología en la Universidad del Rosario en Bogotá.

6.1.4. Organización de las Imágenes

Una vez completado el proceso de doble revisión y tras aplicar los criterios de calidad establecidos (**Figura 9**), las imágenes se organizaron en carpetas separadas en Google Drive, clasificadas en los grupos de TO activa e inactiva, con el objetivo de proceder a su análisis y procesamiento en las fases subsiguientes del estudio (**Figura 10**).

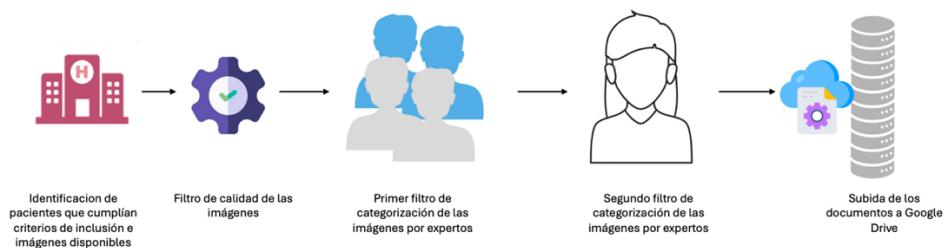


Figura 8: Gestión de la base de datos

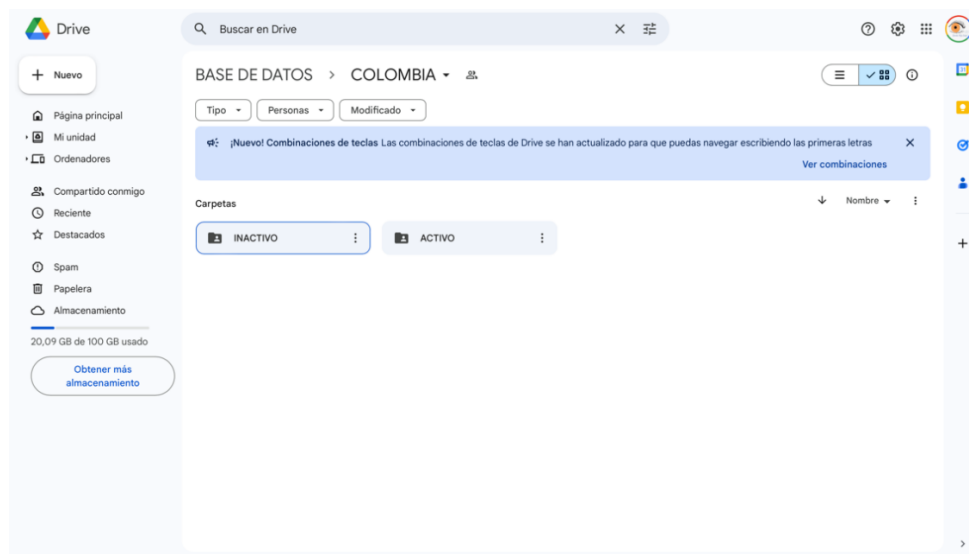


Figura 9: Organización de la base de datos utilizada

6.2. Resultados

De las 1.185 imágenes provenientes de 130 pacientes inicialmente identificadas como posibles candidatas, se eliminaron 573 imágenes (48.4%) debido a problemas de baja calidad que impedían una adecuada visualización de las estructuras del fondo de ojo, tales como mácula, nervio óptico y estructuras vasculares. Tras el proceso de filtrado, se obtuvieron 612 imágenes aptas para la clasificación por parte de los especialistas. En una primera instancia, el 67.3% (412/612) de las imágenes fueron clasificadas como activas o inactivas. Posteriormente, la Dra. Alejandra de la Torre evaluó el conjunto completo y logró clasificar el 82.0% (502/612) de las imágenes. La concordancia entre los oftalmólogos del primer filtro y la Dra. de la Torre fue del 69% para las imágenes inactivas y del 75% para las activas. Finalmente, se logró clasificar el 85% (520/612) del total de imágenes, de las cuales 184 fueron etiquetadas como activas y 359 como inactivas.

7. Entrenamiento de las arquitecturas de aprendizaje profundo

7.1. Metodología

7.1.1. Preprocesamiento de Imágenes

El análisis y preprocesamiento se llevaron a cabo en Google Colaboratory [91], utilizando una variedad de librerías, como numpy, OpenCV (cv2), tensorflow y keras, para manipular las imágenes y preparar los datos para su uso en los modelos de aprendizaje profundo.

Cada imagen fue redimensionada a un tamaño estándar de 224x224x3 píxeles para garantizar la uniformidad en la entrada de los modelos, ya que estas dimensiones son las que mejor se ajustan a las arquitecturas de CNN utilizadas [92]. Las imágenes fueron transformadas a matrices de tensores mediante la función `img_to_array` de Keras y luego normalizadas dividiendo los valores de los píxeles entre 255, convirtiéndolos a un rango de [0, 1]. Esta normalización es esencial para asegurar que los modelos de redes neuronales converjan de manera más eficiente durante el entrenamiento [92].

Se clasificaron las imágenes en dos categorías: "activo" (etiquetadas como 0) e "inactivo" (etiquetadas como 1). Se empleó la función `train_test_split` de Scikit-learn para dividir los datos en un 80% para entrenamiento y un 20% para prueba, asegurando la separación correcta de los datos y evitando fugas entre los conjuntos de entrenamiento y prueba. Para garantizar la reproducibilidad de los resultados, se fijó una semilla aleatoria con un valor de 42.

7.1.2. Aumento de Datos (Data Augmentation)

Para mitigar el riesgo de sobreajuste y mejorar la capacidad de generalización del modelo, se implementó un esquema de aumento de datos mediante la clase `ImageDataGenerator` de Keras. Este proceso permitió aumentar de manera artificial el tamaño del conjunto de datos mediante la generación de nuevas imágenes basadas en transformaciones aleatorias aplicadas a las imágenes originales. Las transformaciones incluyeron rotaciones de hasta 10 grados, desplazamientos horizontales y verticales de hasta un 15%, zoom de hasta un 15%, y la inversión horizontal de las imágenes. Estas técnicas ayudaron a exponer al modelo a una mayor diversidad de variaciones de las imágenes durante el entrenamiento, mejorando así su robustez frente a nuevas muestras no vistas.

7.1.3. Modelos de Transfer Learning

Para aprovechar la capacidad de aprendizaje profundo de CNN previamente entrenadas en grandes conjuntos de datos, se utilizaron tres arquitecturas preentrenadas en el conjunto de datos ImageNet: InceptionV3, VGG16 y ResNetV2. Estas arquitecturas han sido ampliamente probadas en tareas de clasificación de imágenes médicas y han demostrado ser efectivas para la identificación de patologías en imágenes de fondo de ojo [19], [88], [93]. El modelo VGG16 fue seleccionado específicamente por su efectividad comprobada en estudios previos de clasificación de TO, como el trabajo de Parra y colaboradores [88].

El proceso de *transfer learning* implicó cargar las arquitecturas preentrenadas y reemplazar las capas de salida de los modelos originales con una nueva capa completamente conectada (*fully connected*) para la clasificación binaria, adaptada a las categorías de TO activa e inactiva. Las capas iniciales de las arquitecturas preentrenadas se mantuvieron congeladas durante las primeras fases del entrenamiento, lo que permitió que el modelo conservara las características aprendidas en ImageNet, la base de datos compuesta por 3.2 millones de imágenes con la que inicialmente fue entrenada [57], mientras las capas superiores ajustaban los pesos a las características específicas de las imágenes de TO.

7.1.4. Ajuste de Hiperparámetros y Optimización

Durante el desarrollo del modelo, se realizaron múltiples experimentos ajustando diversos hiperparámetros para optimizar el rendimiento de las CNN. Se exploraron varios optimizadores, incluyendo Adam (Adaptive Moment Estimation), SGD (Gradiente Descendente Estocástico), RMSprop (Root Mean Square Propagation), Adagrad (Adaptive Gradient Algorithm), AdaDelta y Nadam (Nesterov-accelerated Adaptive Moment Estimation). Asimismo, se ajustaron tasas de aprendizaje (con valores que variaban desde $1e^{-2}$ hasta $5e^{-3}$, el tamaño de los lotes (*batch size*), el número de épocas, y se probó con diferentes funciones de activación, como ReLU en las capas intermedias y sigmoide en la capa de salida.

La determinación de los hiperparámetros en nuestros modelos se realizó mediante un enfoque empírico basado en experimentación sistemática (ensayo y error), respaldado por referencias en la literatura. En cada experimento, ajustamos distintos valores de los hiperparámetros y analizamos su impacto en el rendimiento del modelo en términos de métricas como la precisión, la pérdida y la matriz de confusión. Para establecer valores iniciales, revisamos estudios previos

sobre modelos similares basados en arquitecturas CNN como VGG, ResNet e Inception, y tomamos como referencia configuraciones óptimas reportadas en la literatura científica. A partir de estos valores base, realizamos pruebas iterativas, ajustando parámetros mencionados en el párrafo anterior.

Para evitar el sobreajuste, se implementaron técnicas de early stopping, que detienen el entrenamiento cuando las métricas de validación dejan de mejorar tras un número determinado de épocas, y reducción de la tasa de aprendizaje, que ajusta dinámicamente la tasa de aprendizaje cuando las métricas de validación no mejoran. Además, se implementaron técnicas de **balanceo de clases**, como el uso de `class_weight` y SMOTE, para equilibrar las clases dado que el conjunto de datos tenía una mayor prevalencia de imágenes de TO inactiva.

Para la implementamos SMOTE para abordar el desbalance de clases aplicándolo a las imágenes aplanadas (convertidas a vectores 1D) y posteriormente restaurándolas a su forma original. Somos conscientes de que SMOTE no preserva la estructura espacial de las imágenes, por lo que complementamos esta estrategia con **Data Augmentation**, aplicando transformaciones como rotaciones, desplazamientos y cambios de escala para mejorar la variabilidad del conjunto de datos sintético. Esto nos permitió generar un conjunto de entrenamiento más balanceado y diverso, evaluando su impacto en la clasificación con modelos como VGG16, ResNetV2 e InceptionV3.

Para evaluar nuestros modelos, empleamos la validación Hold-Out con una división 80/20, asegurando que el conjunto de prueba fuera independiente del entrenamiento y evitando fugas de información. Optamos por este método debido a su eficiencia computacional y simplicidad, permitiendo entrenar el modelo con la mayoría de los datos mientras evaluamos su capacidad de generalización. Para garantizar reproducibilidad y estabilidad, aplicamos una selección aleatoria con una semilla fija (`random_state=42`), asegurando condiciones equitativas en la comparación de modelos y evitando variabilidad en los resultados. Además, utilizamos `stratify=labels` para mantener la proporción de clases en la partición de datos, asegurando una evaluación más equilibrada y representativa del problema.

7.1.5. Construcción de arquitecturas propias

Además de las arquitecturas preentrenadas utilizadas mediante transfer learning, se diseñaron dos arquitecturas de redes neuronales convolucionales específicas para la tarea de clasificación.

Arquitectura 1

Esta arquitectura inicia con una capa de entrada que recibe imágenes con dimensiones $224 \times 224 \times 3$ (correspondiente a altura, anchura y los tres canales de color RGB). A continuación, se implementa una secuencia de capas convolucionales y de pooling. La primera capa convolucional (Conv2D) utiliza 64 filtros de tamaño 3×3 , produciendo una salida de dimensiones $222 \times 222 \times 64$. Luego, se aplica una capa de max pooling con un filtro de 2×2 y un stride de 2, lo que reduce las dimensiones espaciales a $111 \times 111 \times 64$. Para mejorar la estabilidad del entrenamiento, se introduce una capa de normalización por lotes (BatchNormalization). La segunda capa convolucional utiliza 128 filtros de tamaño 3×3 , generando una salida de $109 \times 109 \times 128$, seguida por una capa de pooling que reduce las dimensiones a $54 \times 54 \times 128$. Este patrón se repite a lo largo de la red, incrementando el número de filtros en las capas convolucionales subsiguientes (256 y 512 filtros en las siguientes capas), mientras que las capas de pooling van reduciendo las dimensiones espaciales. Finalmente, la salida de la última capa convolucional se aplanan (Flatten) para formar un vector de tamaño 4608, el cual se alimenta a una capa totalmente conectada (fully connected) con 4096 unidades y activación ReLU. Se añade una capa de dropout con una tasa del 50% para evitar el sobreajuste. Luego, se aplica otra capa totalmente conectada de 4096 unidades antes de llegar a la capa final de una sola unidad, diseñada para la clasificación binaria entre TO activa e inactiva.

Arquitectura 2

La segunda arquitectura sigue un esquema similar, comenzando también con una capa de entrada para imágenes de tamaño $224 \times 224 \times 3$. La primera capa convolucional utiliza 32 filtros de 3×3 , produciendo una salida de $222 \times 222 \times 32$, seguida por una capa de pooling que reduce las dimensiones a $111 \times 111 \times 32$. Esta es seguida por una capa de normalización por lotes para estabilizar las activaciones. La segunda capa convolucional aplica 64 filtros de 3×3 , seguida por una capa de pooling que reduce las dimensiones a $54 \times 54 \times 64$. Nuevamente, se aplica normalización por lotes. A medida que la red avanza, el número de filtros aumenta: la tercera capa convolucional utiliza 128 filtros y reduce las dimensiones a $26 \times 26 \times 128$ mediante pooling, mientras que la cuarta capa convolucional emplea 256 filtros, reduciendo las dimensiones a $12 \times 12 \times 256$. Después de aplicar normalización por lotes, la salida de la cuarta capa se aplanan,

formando un vector de tamaño 36864, que se conecta a una capa totalmente conectada con 512 unidades y activación ReLU. Para evitar el sobreajuste, se utiliza una capa de dropout que apaga el 50% de las unidades. La red concluye con otra capa totalmente conectada de 512 unidades antes de la capa final de una unidad para la clasificación binaria.

7.1.6. Evaluación de modelos de aprendizaje profundo

7.1.6.1. Matriz de confusión

La matriz de confusión es una herramienta útil para evaluar el rendimiento de las CNN, permitiendo identificar en una muestra de los datos, el número de veces que el modelo ha clasificado correcta e incorrectamente cada clase (**Figura 8**). [67]

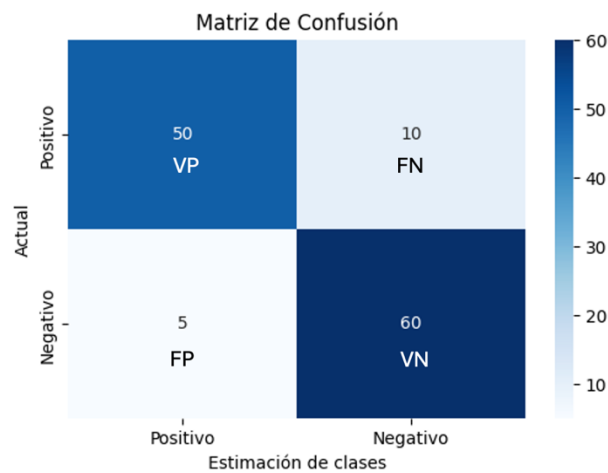


Figura 10: Matriz de confusión

VN: Verdaderos Negativos: Datos de la clase negativa clasificados de manera correcta, FP: Falsos Positivos: Datos de la clase negativa clasificados de manera incorrecta como positivos, FN: Falsos Negativos: Datos de la clase positiva clasificados de manera incorrecta como negativos, VP: Verdaderos Positivos: Datos de la clase positiva clasificados de manera correcta.

A partir de la matriz de confusión se pueden obtener otras métricas de evaluación como la sensibilidad (Capacidad del modelo para identificar correctamente los casos positivos), la especificidad (Capacidad del modelo para identificar correctamente los casos negativos), la precisión (Proporción de positivos predichos correctamente entre todos los predichos como positivos) el F1 (Media armónica de la precisión y la sensibilidad) y la exactitud (accuracy) (proporción de predicciones correctas entre el total de casos) [67], [80].

La exactitud (accuracy) mide el porcentaje de predicciones correctas en comparación con todas las predicciones realizadas [81]. En el caso de las CNN aplicadas a la detección de toxoplasmosis ocular, esta métrica indica la proporción de imágenes clasificadas correctamente como activas o inactivas. La fórmula de exactitud es:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

- TP (True Positives): Verdaderos positivos, casos correctamente clasificados como positivos.
- TN (True Negatives): Verdaderos negativos, casos correctamente clasificados como negativos.
- FP (False Positives): Falsos positivos, casos incorrectamente clasificados como positivos.
- FN (False Negatives): Falsos negativos, casos incorrectamente clasificados como negativos [81].

La sensibilidad mide la capacidad del modelo para identificar correctamente los casos positivos [81], es decir, la proporción de verdaderos positivos entre todos los casos positivos reales. La fórmula de sensibilidad es:

$$\frac{TP}{TP + FN}$$

La especificidad mide la proporción de casos negativos correctamente identificados [81], siendo especialmente útil cuando es crítico evitar falsos positivos, como en los diagnósticos médicos. La fórmula de especificidad es:

$$\frac{TN}{TN + FP}$$

La precisión evalúa cuántas de las predicciones positivas fueron correctas [81]. Es decir, mide la exactitud de los casos positivos predichos por el modelo. La fórmula de precisión es:

$$\frac{TP}{TP + FP}$$

El F1 Score es la media armónica entre la precisión y la sensibilidad [81], proporcionando un equilibrio entre ambas métricas. La fórmula es:

$$2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$$

El Macro Average calcula la métrica (como precisión o sensibilidad) para cada clase y luego promedia los resultados, dando igual peso a cada clase. Esta métrica es particularmente útil en conjuntos de datos desbalanceados, ya que trata todas las clases por igual sin importar su tamaño [81]. Por otro lado, el Micro Average agrega los valores a nivel global antes de calcular la métrica, considerando todos los verdaderos positivos, falsos positivos y falsos negativos de todas las clases, proporcionando una visión más equilibrada del rendimiento global del modelo [81].

7.2. Resultados

7.2.1. ResNetV2

Los mejores resultados se obtuvieron al descongelar todas las capas del modelo y entrenarlo durante 10 épocas. Se utilizó un tamaño de lote de 32, una tasa de aprendizaje de $1e^{-4}$ y el optimizador Adam, además de aplicar la técnica de balanceo de clases SMOTE para corregir el desbalance en el conjunto de datos. Esta produjo una precisión del 97% y una sensibilidad del 70% en la clase activa, y una precisión del 82% y sensibilidad del 98% en la clase inactiva. El valor F1 para la clase activa fue de 0.81, mientras que para la clase inactiva fue de 0.90 con una exactitud general del 86% (**Tabla 1**). El AUC fue de 0.93, las gráficas de exactitud (accuracy) y pérdida mostraron estabilidad en la exactitud (accuracy) de entrenamiento, mientras que en la validación presentaron fluctuaciones, sugiriendo un posible sobreajuste (**Figura 11**).

7.2.2. VGG16

En el caso de la arquitectura VGG16, los mejores resultados se alcanzaron sin descongelar las capas preentrenadas, manteniendo las primeras capas congeladas y entrenando el modelo durante 50 épocas. El tamaño de lote fue de 32, la tasa de aprendizaje se fijó en $1e^{-6}$, y se empleó

también el optimizador Adam con la técnica SMOTE para balancear las clases. Este modelo logró una precisión del 61% y una sensibilidad del 83% en la clase activa, y una precisión del 90% y sensibilidad del 76% en la clase inactiva. El valor F1 para la clase activa fue de 0.70 y para la clase inactiva fue de 0.82 con una exactitud general del 77% (**Tabla 1**). El AUC alcanzado fue de 0.90, Las gráficas de exactitud (accuracy) y pérdida mostraron un aumento y disminución respectivamente más constante en los conjuntos de entrenamiento y validación que el ResNetV2. Sin embargo, aún es evidente algo de sobreajuste. (**Figura 12**).

7.2.3. InceptionV3

Para la arquitectura InceptionV3, se obtuvo el mejor rendimiento descongelando todas las capas y entrenando el modelo durante 10 épocas. Se utilizó un tamaño de lote de 32, una tasa de aprendizaje de $1e^{-6}$, y el optimizador Adam junto con la técnica SMOTE (**Tabla 1**). Esta resultó en una precisión del 88% y una sensibilidad del 62% para la clase activa, mientras que para la clase inactiva se obtuvo una precisión del 78% y sensibilidad del 94%. El valor F1 fue de 0.73 para la clase activa y de 0.85 para la clase inactiva con una exactitud (accuracy) del 80%. El AUC fue de 0.82. Sin embargo, se documentó un evidente sobreajuste en las curvas de exactitud (accuracy) y pérdida (**Figura 13**).

Tabla 1: Resultados arquitecturas de transfer learning seleccionadas

Métrica	Clase	ResNetV2	VGG16	InceptionV3
Precisión	Activa	0.97	0.61	0.88
	Inactiva	0.82	0.90	0.78
Sensibilidad	Activa	0.70	0.83	0.62
	Inactiva	0.98	0.76	0.94
F1	Activa	0.81	0.70	0.73
	Inactiva	0.90	0.82	0.85
Exactitud General (Accuracy)		0.86	0.77	0.80
Área bajo la curva		0.93	0.90	0.82

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Sensibilidad (Recall):** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

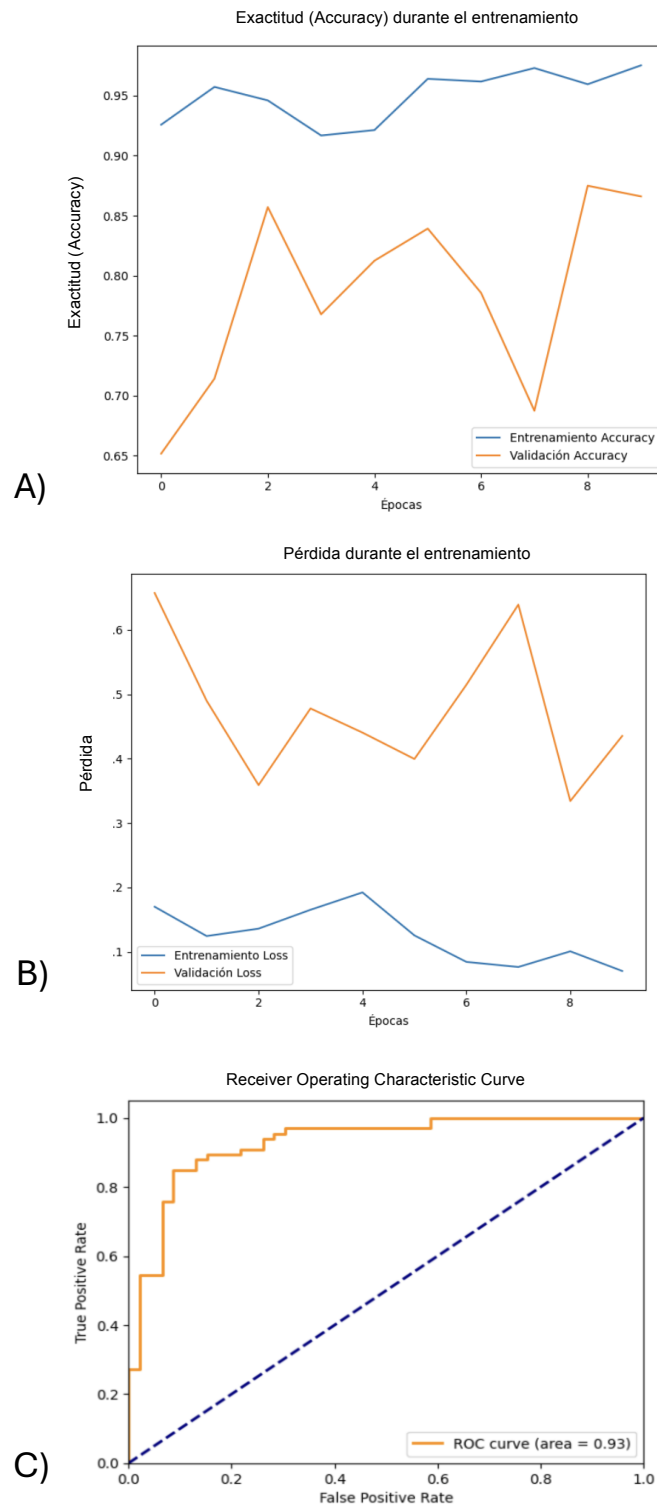


Figura 11: Exactitud (accuracy), pérdida y área bajo la curva de ResNetV2

A) Curva de exactitud a lo largo de las épocas, B) Curva de pérdida a lo largo de las épocas C) área bajo la curva y curva ROC (receiver operating characteristic curve).

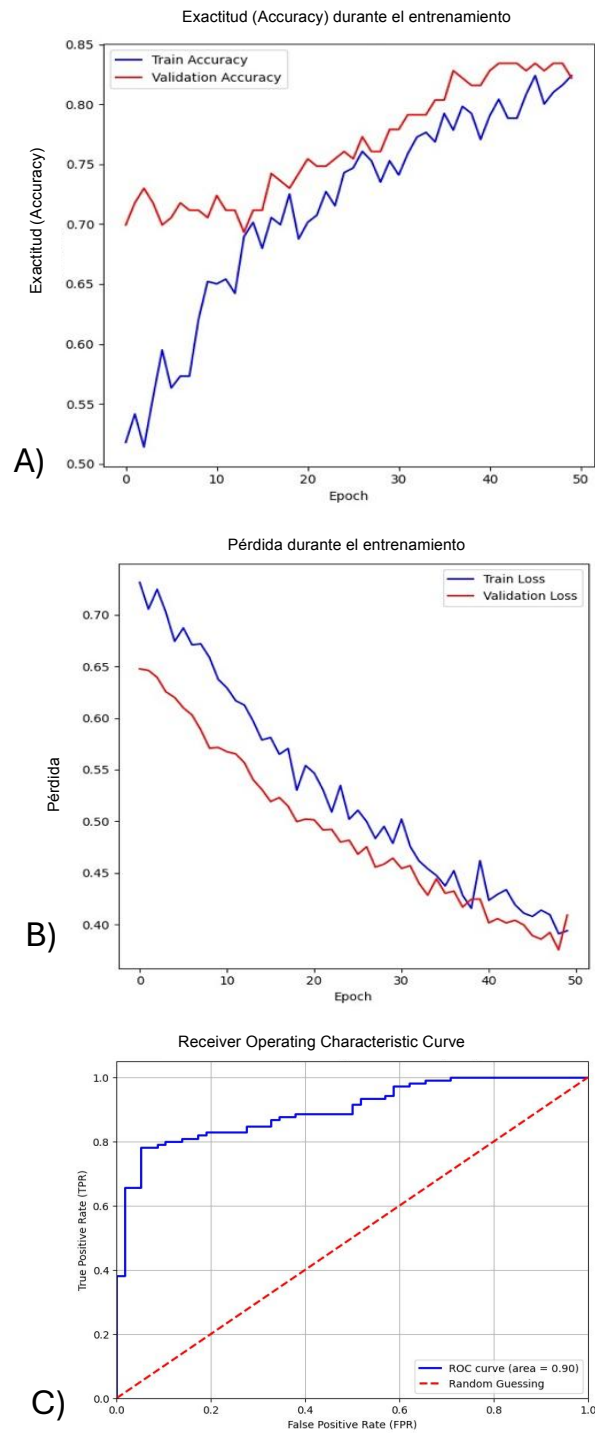


Figura 12: Exactitud (accuracy), pérdida y área bajo la curva deVGG16

A) Curva de exactitud a lo largo de las épocas, B) Curva de pérdida a lo largo de las épocas C) área bajo la curva y curva ROC (receiver operating characteristic curve)

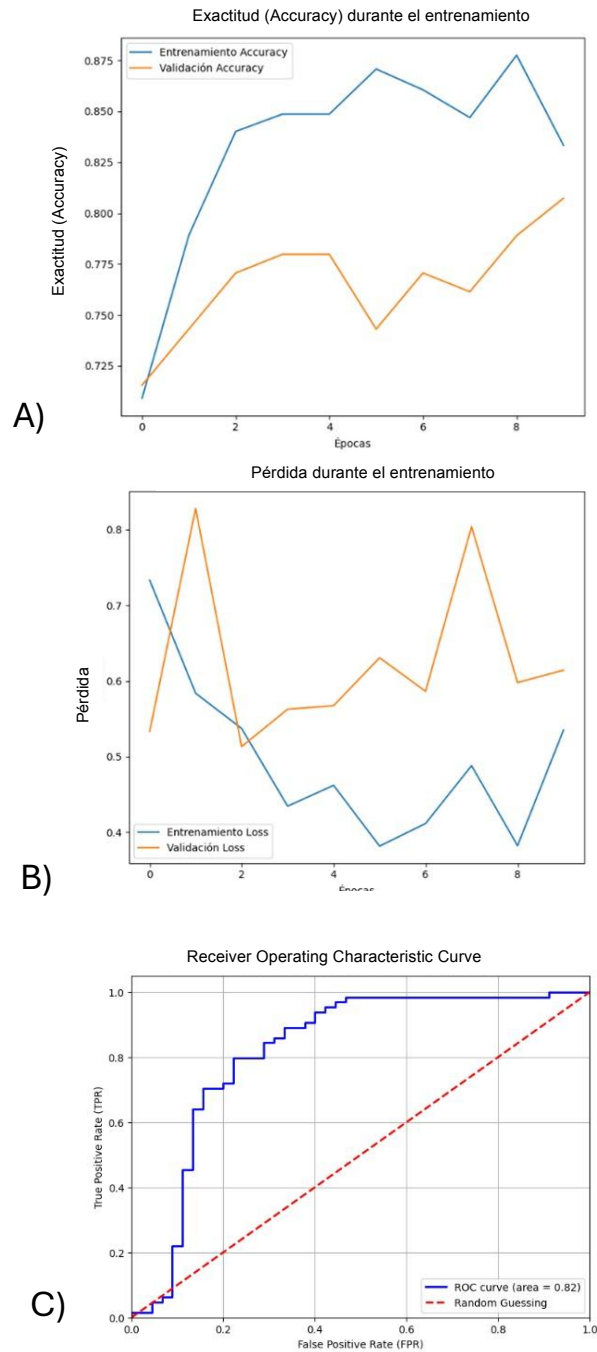


Figura 13: Exactitud (accuracy), pérdida y área bajo la curva de InceptionV3

A) Curva de exactitud (accuracy) a lo largo de las épocas, B) Curva de pérdida a lo largo de las épocas C) área bajo la curva y curva ROC (receiver operating characteristic curve).

La arquitectura propia 1 obtuvo una precisión del 95% para la clase activa, con una sensibilidad significativamente más baja del 39%, lo que resultó en un valor F1 de 0.55. En cuanto a la clase inactiva, la precisión fue del 70% y la sensibilidad del 98%, con un valor F1 de 0.82. La exactitud general fue del 74%, mientras que el AUC alcanzó un valor de 0.68. Las curvas de exactitud (accuracy) y pérdida mostraron una fluctuación considerable en la exactitud (accuracy) y pérdida de validación, indicando un notable sobreajuste del modelo (**Tabla 2 e Figura 14**).

Por otro lado, la arquitectura propia 2 mostró una mejora en el balance de las métricas, obteniendo una precisión del 85% y una sensibilidad del 76% para la clase activa, con un valor F1 de 0.80. Para la clase inactiva, la precisión fue del 84% y la sensibilidad del 91%, con un valor F1 de 0.87. La exactitud general fue del 84% y el AUC fue de 0.72. Las curvas de pérdida y exactitud (accuracy) reflejaron una mayor estabilidad en comparación con la arquitectura propia 1, aunque se observó también cierta fluctuación en la exactitud (accuracy) y pérdida de validación a medida que avanzaban las épocas, lo que sugiere sobreajuste (**Tabla 2 e Figura 15**).

Tabla 2: Resultados de arquitecturas propias realizadas

Métrica	Clase	Arquitectura Propia 1	Arquitectura Propia 2
Precisión	Activa	0.95	0.85
	Inactiva	0.70	0.84
Sensibilidad	Activa	0.39	0.76
	Inactiva	0.98	0.91
F1	Activa	0.55	0.80
	Inactiva	0.82	0.87
Exactitud General (Accuracy)		0.74	0.84
Área bajo la curva		0.68	0.72

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Sensibilidad (Recall):** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

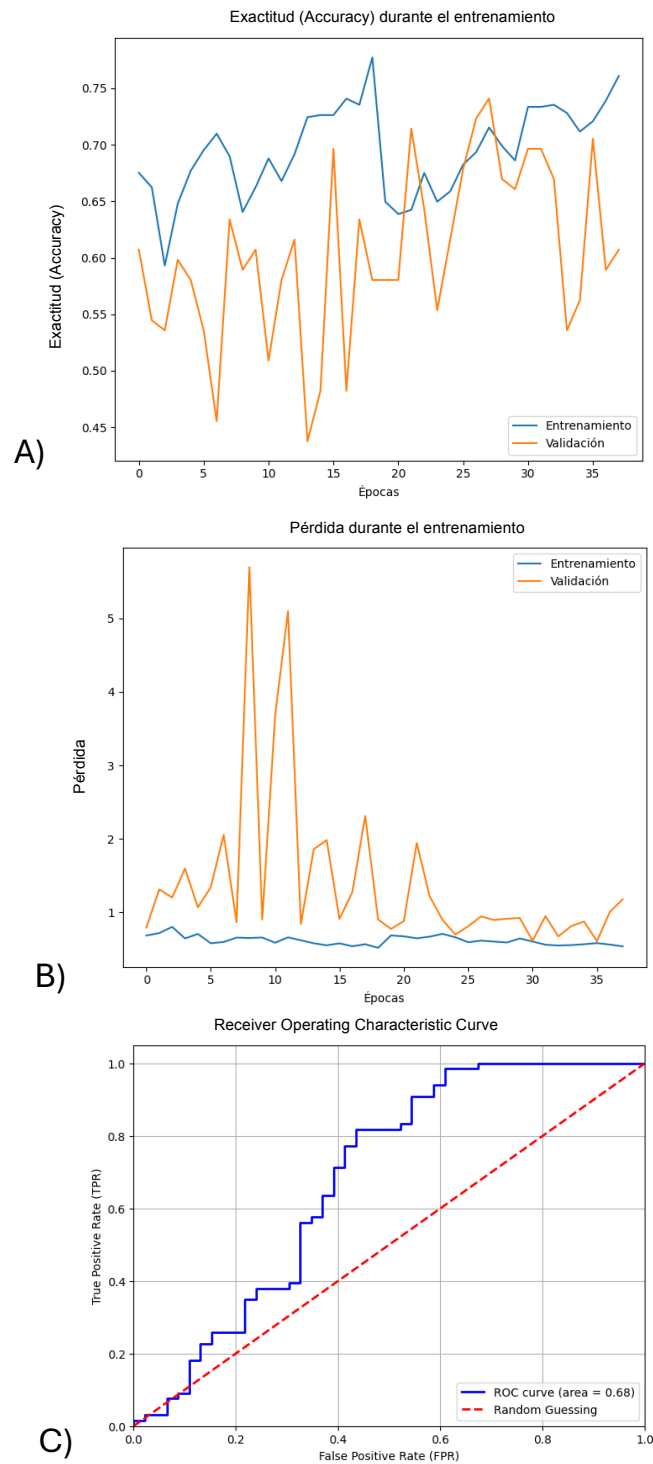


Figura 14: Exactitud (accuracy), pérdida y área bajo la curva de la arquitectura propia número uno

A) Curva de exactitud (accuracy) a lo largo de las épocas, B) Curva de pérdida a lo largo de las épocas C) área bajo la curva y curva ROC (receiver operating characteristic curve).

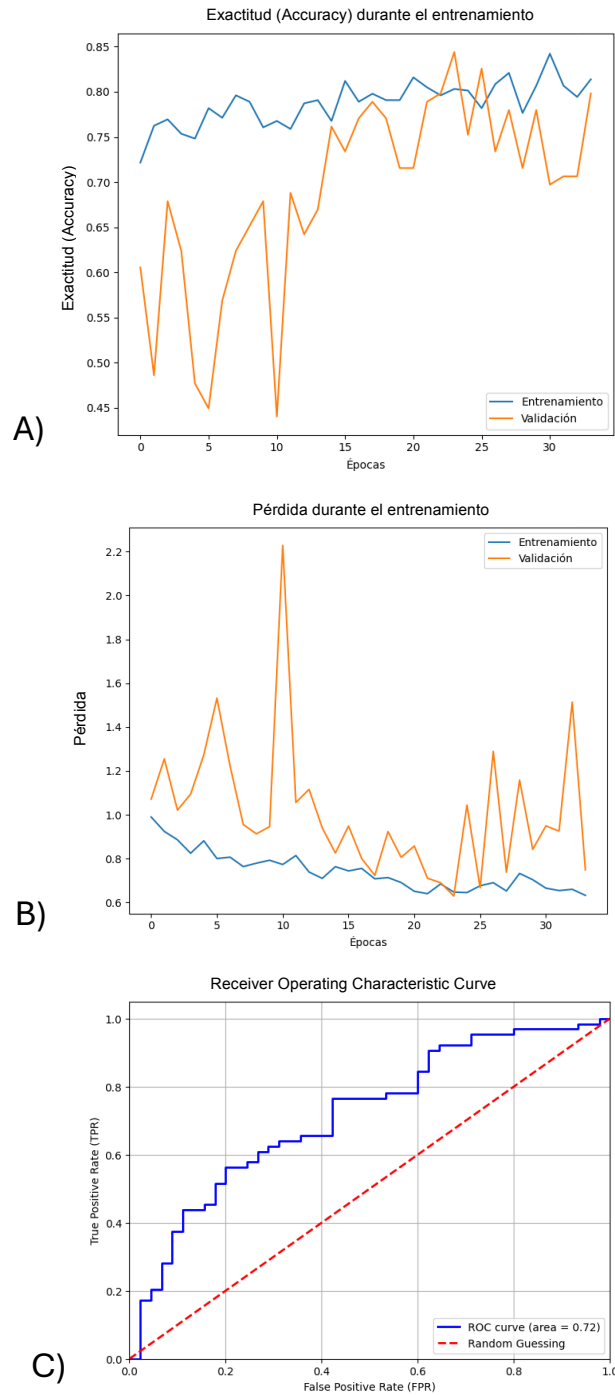


Figura 15: Exactitud (accuracy), pérdida y área bajo la curva de la arquitectura propia número dos

A) Curva de exactitud (accuracy) a lo largo de las épocas, B) Curva de pérdida a lo largo de las épocas C) área bajo la curva y curva ROC (receiver operating characteristic curve).

8. Evaluación de los modelos

8.1. Metodología

Se seleccionaron las siguientes métricas para evaluar el rendimiento de los modelos por cada clase: sensibilidad (recall), especificidad, precisión, y F1-score, además de la exactitud general (accuracy). Estas métricas fueron escogidas debido a su uso común en la evaluación de modelos de IA y su relevancia en la toma de decisiones en ambientes clínicos [94], [95]. El modelo seleccionado como “óptimo” para cada arquitectura fue aquel que mostró la sensibilidad más alta para la detección de casos activos ya que, en el contexto clínico el impacto de clasificar erróneamente un caso activo como inactivo es significativamente mayor que el de clasificar erróneamente un caso inactivo.

Una vez identificado el mejor modelo en términos de las métricas, se procedió a evaluar la reproducibilidad de los modelos de transfer learning, ya que obtuvieron mejores resultados en comparación con las arquitecturas propias. Para estas evaluaciones, se desactivó la semilla aleatoria, configurándola en “None”, y se ejecutó en 5 iteraciones independientes. A partir de estas iteraciones, se calculó la media y la desviación estándar para cada métrica de evaluación, garantizando que los resultados obtenidos fueran robustos y consistentes en diferentes ejecuciones, minimizando el efecto de variaciones aleatorias en el entrenamiento.

8.2. Resultados

La **Tabla 3** resume las medias y desviaciones estándar de las métricas de evaluación para los tres modelos de transfer learning utilizados. Entre estos, el modelo con mayor sensibilidad para los casos activos fue ResNetV2 (0.90 ± 0.06). Por lo tanto, este fue el modelo seleccionado. Aunque el modelo VGG16 presentó una mayor exactitud global (accuracy), alcanzando 0.88 ± 0.01 en comparación con ResNetV2 (0.81 ± 0.02) e InceptionV3 (0.80 ± 0.03), su sensibilidad en la clase activa fue inferior (0.80 ± 0.04).

Selección del modelo definitivo

Para cada arquitectura se seleccionó el modelo que priorizaba tanto la exactitud (*accuracy*) como la sensibilidad (*recall*) en la identificación de casos activos de toxoplasmosis ocular. Esto se debe a que, en un entorno clínico, el impacto de clasificar erróneamente un caso activo como inactivo es significativamente mayor que clasificar incorrectamente un caso inactivo. El modelo seleccionado fue aquel que mostró el mejor ajuste de hiperparámetros en términos de estas métricas clave. Posteriormente, se desactivó la semilla aleatoria para las pruebas finales, configurándola en “None”, y se ejecutó el modelo en 5 iteraciones independientes para evaluar su reproducibilidad. A cada métrica de evaluación se le calculó la media y la desviación estándar, con el fin de garantizar que los resultados fueran robustos y consistentes en múltiples ejecuciones.

Tabla 3: Resumen de cinco iteraciones de los modelos elegidos para ResNetV2, VGG16 e InceptionV3

ResNetV2									
Métrica	precisión		Sensibilidad (recall)		f1_score		Especificidad		Exactitud (accuracy)
	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	
Media	0.74	0.93	0.90	0.77	0.77	0.84	0.69	0.91	0.81
DS	0.08	0.04	0.06	0.07	0.02	0.03	0.08	0.03	0.02
VGG 16									
Métrica	precisión		Sensibilidad (recall)		f1_score		Especificidad		Exactitud (accuracy)
	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	
Media	0.84	0.90	0.80	0.928	0.82	0.91	0.92	0.80	0.88
DS	0.04	0.02	0.043	0.02	0.02	0.01	0.02	0.04	0.01
InceptionV3									
Métrica	precisión		Sensibilidad (recall)		f1_score		Especificidad		Exactitud (accuracy)
	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	Activo	Inactivo	
Media	0.74	0.83	0.66	0.88	0.70	0.85	0.76	0.73	0.807
DS	0.04	0.04	0.08	0.02	0.04	0.02	0.07	0.08	0.032

DS: Desviación estándar. **Precisión:** Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

9. DISCUSIÓN

A pesar de los avances en la detección serológica, imagen multimodal y las técnicas de diagnóstico molecular, el diagnóstico de la TO aún depende en gran medida de la interpretación clínica de las fotografías del fondo de ojo por parte de los oftalmólogos [6]. Stanford MR et al. documentaron una sensibilidad que varía entre el 71% y el 96% y una especificidad entre el 58% y el 100% para la interpretación clínica de las imágenes del fondo de ojo por especialistas en uveítis [14]. Sin embargo, la alta carga de pérdida visual causada por la TO, la escasez de especialistas en uveítis y los largos tiempos de espera resaltan la necesidad de herramientas que permitan un diagnóstico correcto sin la intervención de dichos expertos [2], [3], [96]. En estos escenarios, la inteligencia IA puede mejorar significativamente la exactitud en el diagnóstico de la TO. No obstante, la evaluación por un especialista en uveítis sigue siendo esencial, ya que las imágenes por sí solas no pueden tener en cuenta otros factores cruciales en el diagnóstico y tratamiento, como las complicaciones secundarias a la uveítis, la inflamación en la cámara anterior y los efectos adversos de los medicamentos.

Este estudio demuestra la aplicación de modelos de IA, como las CNN, en la clasificación de la actividad inflamatoria en lesiones retinocoroideas causadas por la TO. Estos hallazgos indican que la IA puede brindar una capacidad de clasificación comparable con la de un experto. Además, abre horizontes de su potencial uso para ayudar a toma de decisiones clínicas ya que cuando se encuentra una lesión activa, se debe iniciar tratamiento para evitar extensión de la lesión y pérdida visual

El alto rendimiento de los modelos anteriormente realizados y este subraya su potencial para integrarse en la práctica clínica rutinaria, permitiendo decisiones diagnósticas y terapéuticas más exactas. Esto es especialmente beneficioso en regiones con una escasez crítica de oftalmólogos y especialistas en uveítis, como en África y algunos países de América Latina [4], [97], [98]. En estas regiones, la prevalencia y gravedad de la TO son mayores debido a la presencia de cepas atípicas y más virulentas, a diferencia del hemisferio norte, donde las cepas de tipo II son más comunes [11], [12]. Esto exacerba el riesgo de malos resultados en los pacientes con TO, lo que

puede llevar a una discapacidad visual grave o ceguera [2]. La implementación de estas herramientas impulsadas por IA democratiza el acceso a servicios de diagnóstico a nivel experto en áreas desatendidas, lo que potencialmente reduce la alta carga de la enfermedad en estas regiones [99].

10. LIMITACIONES

Este estudio presenta varias limitaciones que deben ser consideradas. En primer lugar, los modelos de *transfer learning*, como VGG y ResNet, requieren grandes conjuntos de datos para alcanzar una alta exactitud. Estos modelos se entrenaron originalmente en la base de datos ImageNet, que incluye aproximadamente 1.2 millones de imágenes distribuidas en 1000 clases, y contienen entre 25.6 y 138 millones de parámetros [100], [101]. Aunque el *transfer learning* permite el uso de conjuntos de datos más pequeños, esto aumenta el riesgo de sobreajuste [102], lo que limita la capacidad de generalización de los modelos a nuevas poblaciones de pacientes; recopilación de una mayor cantidad de imágenes puede llevar a mejores resultados con menos niveles de sobreajuste.

Además, el estudio incluyó únicamente fotografías de pacientes atendidos en el consultorio de una especialista en uveítis. Este factor puede introducir un sesgo de selección, ya que los pacientes referidos a esta consulta suelen presentar cuadros clínicos más severos y difíciles de tratar. Como resultado, la muestra puede no ser representativa de todos los pacientes con TO, especialmente aquellos con presentaciones más leves de la enfermedad.

Finalmente, solo se incluyeron imágenes de alta calidad capturadas con equipos especializados, lo que puede limitar la aplicabilidad de los resultados a situaciones clínicas donde se disponga de imágenes de menor calidad como lo pueden ser las áreas donde no se cuentan con especialistas en uveítis. La interpretación de los resultados en el contexto de imágenes de baja resolución debe hacerse con precaución, ya que el rendimiento de los modelos en estas condiciones no ha sido evaluado.

11. CONCLUSIONES Y TRABAJOS FUTUROS

11.1. CONCLUSIONES

Sobre la gestión de la base de datos: se logró gestionar de manera efectiva una base de datos compuesta por 612 imágenes de fondo de ojo, etiquetadas según su actividad (activa o inactiva), lo que permitió contar con un conjunto de datos para el entrenamiento de los modelos de CNN. Si bien el tamaño del conjunto de datos podría considerarse limitado en comparación con la cantidad de imágenes con las que fueron entrenados estos modelos, fue adecuado para los objetivos de este proyecto, dada la prevalencia de la enfermedad. El proceso de clasificación se realizó en dos fases, garantizando la consistencia entre los expertos en oftalmología.

Con relación al entrenamiento de las arquitecturas CNN, se permitió la identificación automática de la actividad de la TO. Las técnicas utilizadas, como el balanceo de datos mediante SMOTE y el aumento de datos, ayudaron a reducir el sobreajuste y mejorar la capacidad de generalización del modelo. De las arquitecturas probadas, ResNetV2 mostró los mejores resultados, con una alta sensibilidad para la clase activa del 90%, lo que demuestra el potencial de esta técnica en la identificación de lesiones activas.

Finalmente, sobre la evaluación de las técnicas de clasificación: las métricas obtenidas (sensibilidad, especificidad, precisión, F1 score y exactitud (accuracy)) confirmaron que las CNN preentrenadas, combinadas con las técnicas de optimización implementadas, son capaces de clasificar correctamente las imágenes de toxoplasmosis ocular. Estos resultados se alinean con el estado del arte, lo que respalda la viabilidad de la IA como herramienta de apoyo en el diagnóstico de esta enfermedad y de su actividad.

11.2. TRABAJOS FUTUROS

Dadas las limitaciones mencionadas, es crucial llevar a cabo estudios futuros utilizando bases de datos más grandes y provenientes de múltiples centros para mejorar la generalización de los modelos y reducir el riesgo de sobreajuste. Además, resulta fundamental validar los resultados obtenidos en poblaciones europeas, quienes exhiben características clínicas diferentes de la TO, atribuibles a un predominio de cepas menos virulentas (Tipo II) y una respuesta inmunitaria

TH1 que conduce a la producción de interferón gamma. Estas respuestas generan lesiones más pequeñas y periféricas. La validación podría realizarse mediante aprendizaje federado, lo cual permitiría ajustar los modelos de redes neuronales convolucionales a las particularidades clínicas de las poblaciones.

12. CRONOGRAMA

Mayo 2024

- Semana 1: Modificación tasa de aprendizaje modelo VGG16.
- Semana 2: Modificación número de épocas y tamaño de lote modelo VGG16.
- Semana 3: Aplicación técnicas de balanceo modelo VGG16.
- Semana 4: Modificación funciones de activación y optimizadores modelo VGG16.

Junio 2024

- Semana 1: Modificación tasa de aprendizaje modelo InceptionV3.
- Semana 2: Modificación número de épocas y tamaño de lote modelo InceptionV3.
- Semana 3: Aplicación técnicas de balanceo modelo InceptionV3.
- Semana 4: Modificación funciones de activación y optimizadores modelo InceptionV3.

Julio 2024

- Semana 1: Modificación tasa de aprendizaje modelo InceptionResNetV2.
- Semana 2: Modificación número de épocas y tamaño de lote modelo InceptionResNetV2.
- Semana 3: Aplicación técnicas de balanceo modelo InceptionResNetV2.
- Semana 4: Modificación funciones de activación y optimizadores modelo InceptionResNetV2.

Agosto 2024

- Semana 1: Pruebas de ajuste modelo VGG16
- Semana 2: Pruebas de ajuste modelo VGG16
- Semana 3: Pruebas de ajuste modelo InceptionV3.
- Semana 4: Pruebas de ajuste modelo InceptionV3.

Septiembre 2024

- Semana 1: Pruebas de ajuste modelo InceptionResNetV2.
- Semana 2: Pruebas de ajuste modelo InceptionResNetV2.
- Semana 3: Selección de métricas para los 3 modelos.

- Semana 4: Prueba de reproducibilidad del modelo VGG16.

-

Octubre 2024

- Semana 1: Prueba de reproducibilidad del modelo InceptionV3.

- Semana 2: Prueba de reproducibilidad del modelo InceptionResNetV2.

- Semana 3: Comparación de las métricas de los 3 modelos.

- Semana 4: Selección del modelo final a utilizar.

Noviembre 2024

- Semana 1: Realización de redes neuronales convolucionales propias.

- Semana 2: Ajustes de redes neuronales convolucionales propias.

- Semana 3: Ajustes de redes neuronales convolucionales propias.

- Semana 4: Selección final mejores modelos de arquitecturas propias.

Diciembre 2024

- Semana 1: Presentación final

- Semana 2: Presentación final

- Semana 3: Presentación final

- Semana 4: Presentación final

Periodo	2023	2024											
	Dic	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sept	Oct	Nov	Dic
Recopilación de Datos	x	x											
Preprocesamiento de Datos			x										
Aseguramiento de la Calidad de los Datos			x										
Selección de				x	x	x							

Arquitecturas de CNN													
Desarrollo y Con Figura del Modelo:				x	x	x	x						
Entrenamiento y Validación del Modelo						x	x	x					
Pruebas y Ajustes del Modelo								x	x				
Selección de Métricas de Evaluación									x				
Análisis de Rendimiento									x				
Interpretación y Documentación de Resultados									x	x			
Revisión y mejora										x	x	x	
Presentación final													x

CNN: Redes neuronales convolucionales

13. REFERENCIAS BIBLIOGRÁFICAS

- [1] A. de-la-Torre, G. González, J. Díaz-Ramírez, y J. E. Gómez-Marín, «Screening by Ophthalmoscopy for Toxoplasma Retinochoroiditis in Colombia», *American Journal of Ophthalmology*, vol. 143, n.º 2, pp. 354-356, 2007, doi: 10.1016/j.ajo.2006.09.048.
- [2] C. Cifuentes-González *et al.*, «Risk factors for recurrences and visual impairment in patients with ocular toxoplasmosis: A systematic review and meta-analysis», *PLOS ONE*, vol. 18, n.º 4, p. e0283845, abr. 2023, doi: 10.1371/journal.pone.0283845.
- [3] A. de-la-Torre, G. González-López, J. M. Montoya-Gutiérrez, V. Marín-Arango, y J. E. Gómez-Marín, «Quality of Life Assessment in Ocular Toxoplasmosis in a Colombian Population», *Ocular Immunology and Inflammation*, vol. 19, n.º 4, pp. 262-266, ago. 2011, doi: 10.3109/09273948.2011.582220.
- [4] J. Gómez-Marín, «Toxoplasmosis: Un problema de Salud Pública en Colombia», *Rev. Salud Pública*, vol. 4, n.º 2, pp. 7-10, 2002.
- [5] G. Bou, M. S. Figueroa, P. Martí-Belda, E. Navas, y A. Guerrero, «Value of PCR for Detection of *Toxoplasma gondii* in Aqueous Humor and Blood Samples from Immunocompetent Patients with Ocular Toxoplasmosis», *J Clin Microbiol*, vol. 37, n.º 11, pp. 3465-3468, 1999, doi: 10.1128/JCM.37.11.3465-3468.1999.
- [6] E. J. H. Goh *et al.*, «Ocular Toxoplasmosis», *Ocular Immunology and Inflammation*, vol. 31, n.º 7, pp. 1342-1361, ago. 2023, doi: 10.1080/09273948.2022.2117705.
- [7] M. Attias, D. E. Teixeira, M. Benchimol, R. C. Vommaro, P. H. Crepaldi, y W. De Souza, «The life-cycle of *Toxoplasma gondii* reviewed using animations», *Parasites & Vectors*, vol. 13, n.º 1, p. 588, nov. 2020, doi: 10.1186/s13071-020-04445-z.
- [8] A. de-la-Torre *et al.*, «Severe South American Ocular Toxoplasmosis Is Associated with Decreased Ifn- γ /Il-17a and Increased Il-6/Il-13 Intraocular Levels», *PLOS Neglected Tropical Diseases*, vol. 7, n.º 11, p. e2541, nov. 2013, doi: 10.1371/journal.pntd.0002541.
- [9] D. Kalogeropoulos *et al.*, «Pathophysiological Aspects of Ocular Toxoplasmosis: Host-parasite Interactions», *Ocular Immunology and Inflammation*, vol. 30, n.º 3, pp. 560-569, 2022, doi: 10.1080/09273948.2021.1922706.
- [10] C. Cifuentes-González *et al.*, «Colombian Ocular Infectious Epidemiology Study (COIES): Ocular Toxoplasmosis Incidence and Sociodemographic Characterization, 2015-2019», *International Journal of Infectious Diseases*, vol. 117, pp. 349-355, abr. 2022, doi: 10.1016/j.ijid.2022.02.028.
- [11] C. Alvarez *et al.*, «Striking Divergence in *Toxoplasma ROP16* Nucleotide Sequences From Human and Meat Samples», *J Infect Dis.*, vol. 211, n.º 12, pp. 2006-2013, 2015, doi: 10.1093/infdis/jiu833.
- [12] V. Sánchez, A. de-la-Torre, y J. E. Gómez-Marín, «Characterization of ROP18 alleles in human toxoplasmosis», *Parasitology International*, vol. 63, n.º 2, pp. 463-469, 2014, doi: 10.1016/j.parint.2013.10.012.
- [13] D. C. Dabian y F. P. Moyano, «Prevalencia y causas de ceguera y discapacidad visual en Colombia», *Ciencia y Tecnología para la Salud Visual y Ocular*, vol. 18, n.º 2, pp. 21-30, mar. 2020, doi: <https://doi.org/10.19052/sv.vol18.iss2.3>.
- [14] M. R. Stanford, «Reliability of expert interpretation of retinal photographs for the diagnosis of toxoplasma retinochoroiditis», *British Journal of Ophthalmology*, vol. 86, n.º 6, pp. 636-639, 2002, doi: 10.1136/bjo.86.6.636.
- [15] C. Ozgonul y C. G. Besirli, «Recent Developments in the Diagnosis and Treatment of Ocular Toxoplasmosis», *Ophthalmic Research*, vol. 57, n.º 1, pp. 1-12, oct. 2016, doi: 10.1159/000449169.

- [16] M. Previato *et al.*, «A Brazilian report using serological and molecular diagnosis to monitoring acute ocular toxoplasmosis», *BMC Research Notes*, vol. 8, n.º 1, p. 746, dic. 2015, doi: 10.1186/s13104-015-1650-6.
- [17] K. Zhang, G. Lin, Y. Han, y J. Li, «Serological diagnosis of toxoplasmosis and standardization», *Clinica Chimica Acta*, vol. 461, pp. 83-89, oct. 2016, doi: 10.1016/j.cca.2016.07.018.
- [18] A. D. Chakravarthy *et al.*, «An Approach Towards Automatic Detection of Toxoplasmosis using Fundus Images», en *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, oct. 2019, pp. 710-717. doi: 10.1109/BIBE.2019.00134.
- [19] R. Parra *et al.*, «Automatic Diagnosis of Ocular Toxoplasmosis from Fundus Images with Residual Neural Networks», en *Public Health and Informatics*, IOS Press, 2021, pp. 173-177. doi: 10.3233/SHTI210143.
- [20] M. Hasanreisoglu *et al.*, «Ocular Toxoplasmosis Lesion Detection on Fundus Photograph using a Deep Learning Model», *Investigative Ophthalmology & Visual Science*, vol. 61, n.º 7, p. 1627, jun. 2020.
- [21] J. Garweg, M. Boehnke, y F. Koerner, «Restricted applicability of the polymerase chain reaction for the diagnosis of ocular toxoplasmosis», *Ger J Ophthalmol*, vol. 5, n.º 2, pp. 104-108, 1996.
- [22] F. Robert-Gangneux, P. Binisti, D. Antonetti, J. Dupouy-Camet, A. Brezin, y H. Yera, «Usefulness of Immunoblotting and Goldmann-Witmer Coefficient for Biological Diagnosis of Toxoplasmic Retinochoroiditis», *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 23, n.º 1, pp. 34-38, 2004, doi: 10.1007/s10096-003-1048-6.
- [23] Istock, «Ciclo de vida Toxoplasma gondii». [En línea]. Disponible en: [<https://www.istockphoto.com/es/vector/ciclo-de-vida-de-toxoplasma-gondii-gm607773240-104224773>]
- [24] M. Harrell y P. E. Carvounis, «Current Treatment of Toxoplasma Retinochoroiditis: An Evidence-Based Review», *Journal of Ophthalmology*, vol. 2014, p. e273506, ago. 2014, doi: 10.1155/2014/273506.
- [25] A. De-la-Torre, C. A. López-Castillo, J. C. Rueda, R. D. Mantilla, J. E. Gómez-Marín, y J.-M. Anaya, «Clinical patterns of uveitis in two ophthalmology centres in Bogota, Colombia», *Clinical & Experimental Ophthalmology*, vol. 37, n.º 5, pp. 458-466, 2009, doi: 10.1111/j.1442-9071.2009.02082.x.
- [26] M. J. Runday *et al.*, «Aetiology of uveitis in Sierra Leone, west Africa.», *British Journal of Ophthalmology*, vol. 80, n.º 11, pp. 956-961, nov. 1996, doi: 10.1136/bjo.80.11.956.
- [27] D. Das *et al.*, «Pattern of uveitis in North East India: A tertiary eye care center study», *Indian Journal of Ophthalmology*, vol. 57, n.º 2, p. 144, abr. 2009, doi: 10.4103/0301-4738.45506.
- [28] D. Gonzalez Fernandez, H. Nascimento, C. Nascimento, C. Muccioli, y R. Belfort, «Uveitis in São Paulo, Brazil: 1053 New Patients in 15 Months», *Ocular Immunology and Inflammation*, vol. 25, n.º 3, pp. 382-387, 2017, doi: 10.3109/09273948.2015.1132741.
- [29] L.-X. Z. Yang Zheng y L.-X. Z. Yang Zheng, «Clinical patterns and characteristics of uveitis in a secondary hospital in southern China», *qwer*, vol. 8, n.º 2, pp. 337-341, abr. 2015, doi: 10.3980/j.issn.2222-3959.2015.02.22.
- [30] B. Bodaghi *et al.*, «Chronic Severe Uveitis: Etiology and Visual Outcome in 927 Patients from a Single Center», *Medicine*, vol. 80, n.º 4, p. 263, jul. 2001.
- [31] P. T. Merrill, J. Kim, T. A. Cox, C. C. Betor, R. M. McCallum, y G. J. Jaffe, «Uveitis in the southeastern United States», *Current Eye Research*, vol. 16, n.º 9, pp. 865-874, 1997, doi: 10.1076/ceyr.16.9.865.5048.
- [32] G. Florez-Esparza *et al.*, «Colombian Ocular Inflammatory Diseases Epidemiology Study (COIDES): Prevalence, Incidence, and Sociodemographic Characterization of Uveitis in

Colombia, 2015-2020», *Ocular Immunology and Inflammation*, pp. 1-7, 2023, doi: 10.1080/09273948.2023.2279681.

[33] A. de-la-Torre *et al.*, «Epidemiology, clinical features, and classification of 3,404 patients with uveitis: Colombian Uveitis Multicenter Study (COL-UVEA)», *Graefes Arch Clin Exp Ophthalmol*, 2024, doi: 10.1007/s00417-024-06422-z.

[34] D. V. Vasconcelos-Santos *et al.*, «Congenital Toxoplasmosis in Southeastern Brazil: Results of Early Ophthalmologic Examination of a Large Cohort of Neonates», *Ophthalmology*, vol. 116, n.º 11, pp. 2199-2205.e1, nov. 2009, doi: 10.1016/j.ophtha.2009.04.042.

[35] T. E. F. Arantes *et al.*, «Ocular Involvement Following Postnatally Acquired *Toxoplasma gondii* Infection in Southern Brazil: A 28-Year Experience», *American Journal of Ophthalmology*, vol. 159, n.º 6, pp. 1002-1012.e2, jun. 2015, doi: 10.1016/j.ajo.2015.02.015.

[36] «Webinar Grand Round». Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <https://www.youtube.com/watch?v=5N4ABEjyEhY>.

[37] D. Kalogeropoulos *et al.*, «Ocular toxoplasmosis: a review of the current diagnostic and therapeutic approaches», *Int Ophthalmol*, vol. 42, n.º 1, pp. 295-321, ene. 2022, doi: 10.1007/s10792-021-01994-9.

[38] J. E. Gomez-Marin y A. de-la-Torre, «Ocular disease due to *Toxoplasma gondii*», en *Toxoplasma gondii*, Elsevier, 2020, pp. 229-291. doi: 10.1016/B978-0-12-815041-2.00005-0.

[39] A. Kelgaonkar, A. Khalsa, S. K. Padhy, U. C. Behera, T. R. Padhi, y S. Basu, «Punctate Inner Retinal Toxoplasmosis: Case Series and Review of Literature», *Ocular Immunology and Inflammation*, vol. 30, n.º 3, pp. 546-555, 2022, doi: 10.1080/09273948.2021.1980815.

[40] C. Brandão-de-Resende *et al.*, «Clinical and Multimodal Imaging Findings and Risk Factors for Ocular Involvement in a Presumed Waterborne Toxoplasmosis Outbreak, Brazil1», *Emerg. Infect. Dis.*, vol. 26, n.º 12, pp. 2922-2932, nov. 2020, doi: 10.3201/eid2612.200227.

[41] J. G. Montoya y J. S. Remington, «Toxoplasmic Chorioretinitis in the Setting of Acute Acquired Toxoplasmosis», *Clinical Infectious Diseases*, vol. 23, n.º 2, pp. 277-282, ago. 1996, doi: 10.1093/clinids/23.2.277.

[42] F. D. Verbraak *et al.*, «Serological and polymerase chain reaction-based analysis of aqueous humour samples in patients with AIDS and necrotizing retinitis», *AIDS*, vol. 10, n.º 10, pp. 1091-1099, 1996.

[43] F. Robert, T. Ouatas, P. Blanche, C. Tourte-Schaefer, D. Sicard, y J. Dupouy-Camet, «[Retrospective evaluation of the detection of *Toxoplasma gondii* by polymerase chain reaction in AIDS patients]», *Presse Med*, vol. 25, n.º 11, pp. 541-545, 1996.

[44] A. P. Brézin *et al.*, «Analysis of aqueous humor in ocular toxoplasmosis», *N Engl J Med*, vol. 324, n.º 10, p. 699, 1991.

[45] F. Aouizerate *et al.*, «Detection of *Toxoplasma gondii* in aqueous humour by the polymerase chain reaction.», *British Journal of Ophthalmology*, vol. 77, n.º 2, pp. 107-109, 1993, doi: 10.1136/bjo.77.2.107.

[46] M. Previato *et al.*, «A Brazilian report using serological and molecular diagnosis to monitoring acute ocular toxoplasmosis», *BMC Research Notes*, vol. 8, n.º 1, p. 746, dic. 2015, doi: 10.1186/s13104-015-1650-6.

[47] J. E. Feliciano-Alfonso *et al.*, «Safety and efficacy of different antibiotic regimens in patients with ocular toxoplasmosis: systematic review and meta-analysis», *Syst Rev*, vol. 10, n.º 1, p. 206, 2021, doi: 10.1186/s13643-021-01758-7.

[48] V. Álvarez-García, L. Rubio-Romero, M. A. Maldonado, M. Gómez-Suárez, y A. de-la-Torre, «Ocular toxoplasmosis in immunocompetent adults: current cost-effectiveness of four treatment regimens in Colombia», *Heliyon*, vol. 7, n.º 11, p. e08265, 2021, doi: 10.1016/j.heliyon.2021.e08265.

[49] A. de-la-Torre, M. Stanford, A. Curi, G. J. Jaffe, y J. E. Gomez-Marin, «Therapy for Ocular

- Toxoplasmosis», *Ocular Immunology and Inflammation*, vol. 19, n.º 5, pp. 314-320, 2011, doi: 10.3109/09273948.2011.608915.
- [50] R. Kopec, G. D. Caro, E. Chapnick, M. Ghitan, y N. Saffra, «Prophylaxis for Ocular Toxoplasmosis», *Clinical Infectious Diseases*, vol. 37, n.º 10, pp. e147-e148, 2003, doi: 10.1086/379125.
- [51] J. C. Omil, «Inteligencia artificial ¿Dr. Jekyll o Mr. Hyde?», *Mercados y Negocios*, n.º 40, pp. 5-22, 2019.
- [52] R. Verma, V. Nagar, y S. Mahapatra, «Introduction to Supervised Learning», en *Data Analytics in Bioinformatics*, 1.ª ed., R. Satpathy, T. Choudhury, S. Satpathy, S. N. Mohanty, y X. Zhang, Eds., Wiley, 2021, pp. 1-34. doi: 10.1002/9781119785620.ch1.
- [53] G. Valenzuela González, «Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones», *Supervised Learning: Methods, Properties and Applications*, 2022, Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <https://riuma.uma.es/xmlui/handle/10630/25147>
- [54] G. James, *An Introduction to Statistical Learning*. 2021. Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <https://link.springer.com/book/10.1007/978-1-0716-1418-1>
- [55] C. Matthieu, *Machine Learning Techniques for Multimedia*. 2008. Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <https://link.springer.com/book/10.1007/978-3-540-75171-7>
- [56] D. Zuo, L. Yang, Y. Jin, H. Qi, Y. Liu, y L. Ren, «Machine learning-based models for the prediction of breast cancer recurrence risk», *BMC Med Inform Decis Mak*, vol. 23, n.º 1, p. 276, 2023, doi: 10.1186/s12911-023-02377-z.
- [57] X. Deng *et al.*, «Machine learning model for the prediction of prostate cancer in patients with low prostate-specific antigen levels: A multicenter retrospective analysis», *Front. Oncol.*, vol. 12, p. 985940, 2022, doi: 10.3389/fonc.2022.985940.
- [58] Y. Li, X. Wu, P. Yang, G. Jiang, y Y. Luo, «Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis», *Genomics, Proteomics & Bioinformatics*, vol. 20, n.º 5, pp. 850-866, 2022, doi: 10.1016/j.gpb.2022.11.003.
- [59] A. Sorayaie Azar *et al.*, «Application of machine learning techniques for predicting survival in ovarian cancer», *BMC Medical Informatics and Decision Making*, vol. 22, n.º 1, p. 345, 2022, doi: 10.1186/s12911-022-02087-y.
- [60] R. K. Mishra, G. Y. S. Reddy, y H. Pathak, «The Understanding of Deep Learning: A Comprehensive Review», *Mathematical Problems in Engineering*, vol. 2021, p. e5548884, abr. 2021, doi: 10.1155/2021/5548884.
- [61] R. Indrakumari, T. Poongodi, y K. Singh, «Introduction to Deep Learning», en *Advanced Deep Learning for Engineers and Scientists: A Practical Approach*, K. B. Prakash, R. Kannan, S. A. Alexander, y G. R. Kanagachidambaresan, Eds., Cham: Springer International Publishing, 2021, pp. 1-22. doi: 10.1007/978-3-030-66519-7_1.
- [62] A. L. Moreno Díaz, «Análisis comparativo de arquitecturas de redes neuronales para la clasificación de imágenes.», Universidad Internacional de La Rioja (UNIR), Madrid. [En línea]. Disponible en: <https://reunir.unir.net/bitstream/handle/123456789/10008/Moreno%20D%C3%ADaz-Alejo%2C%20Lara.pdf?sequence=1&isAllowed=y>
- [63] Y. Lecun, L. Bottou, Y. Bengio, y P. Haffner, «Gradient-based learning applied to document recognition», *Proceedings of the IEEE*, vol. 86, n.º 11, pp. 2278-2324, nov. 1998, doi: 10.1109/5.726791.
- [64] D. H. Hubel y T. N. Wiesel, «Receptive fields and functional architecture of monkey striate cortex», *J Physiol*, vol. 195, n.º 1, pp. 215-243, mar. 1968, doi: 10.1113/jphysiol.1968.sp008455.
- [65] M. A. Lopez Pacheco, «Redes neuronales convolucionales para el modelado de sistemas no lineales con aplicación al monitoreo de daño estructural.», CENTRO DE INVESTIGACIÓN Y

- DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL, 2021. [En línea]. Disponible en: <https://www.ctrl.cinvestav.mx/~yuw/pdf/DoTesMLopez.pdf>
- [66] B. Armada Carrión, «Utilización de técnicas de Machine Learning para mejorar la detección de infracción de marcas», 2021, Accedido: 4 de diciembre de 2023. [En línea]. Disponible en: <https://repositorio.comillas.edu/xmlui/handle/11531/46947>
- [67] D. A. López Santander, «Arquitectura de aprendizaje profundo usando CNNs y RNNs para la clasificación de la enfermedad de Parkinson y Huntington a partir de señales de voz», 2022, Accedido: 4 de diciembre de 2023. [En línea]. Disponible en: <https://bibliotecadigital.udea.edu.co/handle/10495/33159>
- [68] S. M. Chavez Leyton, «Detección de Deepfakes», 2021, Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <http://hdl.handle.net/1992/55660>
- [69] «Arquitectura ResNetV2». [En línea]. Disponible en: <https://jabocken.github.io/ML2017Fall/docs/models/>
- [70] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, y K.-W. Huang, «An Improved VGG16 Model for Pneumonia Image Classification», *Applied Sciences*, vol. 11, n.º 23, p. 11185, 2021, doi: 10.3390/app112311185.
- [71] J. Verdeguez Gómez, «Redes neuronales para la clasificación y segmentación de imágenes médicas», 2021, Accedido: 7 de septiembre de 2024. [En línea]. Disponible en: <https://riunet.upv.es/handle/10251/158739>
- [72] M. Hassan, «Arquitectura de la red VVG-16». [En línea]. Disponible en: <https://neurohive.io/en/popular-networks/vgg16/>
- [73] C. Quintero *et al.*, «Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitorio Participativo», 2518-6841, 2018, doi: 10.18502/keg.v3i1.1462.
- [74] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, y Z. Wojna, «Rethinking the Inception Architecture for Computer Vision», 2015, doi: 10.48550/ARXIV.1512.00567.
- [75] Szegedy, «Inception-v3». [En línea]. Disponible en: <https://paperswithcode.com/method/inception-v3>
- [76] G. Murtaza *et al.*, «Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges», *Artif Intell Rev*, vol. 53, n.º 3, pp. 1655-1720, 2020, doi: 10.1007/s10462-019-09716-5.
- [77] J. Xue *et al.*, «Hypergraph membrane system based F2 fully convolutional neural network for brain tumor segmentation», *Applied Soft Computing*, vol. 94, p. 106454, 2020, doi: 10.1016/j.asoc.2020.106454.
- [78] O. A. Montesinos López, A. Montesinos López, y J. Crossa, «Convolutional Neural Networks», en *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Cham: Springer International Publishing, 2022, pp. 533-577. doi: 10.1007/978-3-030-89010-0_13.
- [79] K. Simonyan y A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition», 2014, *arXiv*. doi: 10.48550/ARXIV.1409.1556.
- [80] C. Sweeney, E. Ennis, M. Mulvenna, R. Bond, y S. O'Neill, «How Machine Learning Classification Accuracy Changes in a Happiness Dataset with Different Demographic Groups», *Computers*, vol. 11, n.º 5, Art. n.º 5, may 2022, doi: 10.3390/computers11050083.
- [81] M. Grandini, E. Bagli, y G. Visani, «Metrics for Multi-Class Classification: an Overview», 2020, *arXiv*. doi: 10.48550/ARXIV.2008.05756.
- [82] R. Kapoor, S. P. Walters, y L. A. Al-Aswad, «The current state of artificial intelligence in ophthalmology», *Survey of Ophthalmology*, vol. 64, n.º 2, pp. 233-240, mar. 2019, doi: 10.1016/j.survophthal.2018.09.002.
- [83] M. Shaban *et al.*, «A convolutional neural network for the screening and staging of diabetic retinopathy», *PLOS ONE*, vol. 15, n.º 6, p. e0233514, jun. 2020, doi:

10.1371/journal.pone.0233514.

- [84] J. M. Brown *et al.*, «Automated Diagnosis of Plus Disease in Retinopathy of Prematurity Using Deep Convolutional Neural Networks», *JAMA Ophthalmology*, vol. 136, n.º 7, pp. 803-810, jul. 2018, doi: 10.1001/jamaophthalmol.2018.1934.
- [85] W. Rojas-Carabali *et al.*, «Managing a patient with uveitis in the era of artificial intelligence: Current approaches, emerging trends, and future perspectives», *Asia-Pacific Journal of Ophthalmology*, vol. 13, n.º 4, p. 100082, 2024, doi: 10.1016/j.apjo.2024.100082.
- [86] W. Rojas-Carabali *et al.*, «Evaluating the Diagnostic Accuracy and Management Recommendations of ChatGPT in Uveitis», *Ocular Immunology and Inflammation*, vol. 32, n.º 8, pp. 1526-1531, 2024, doi: 10.1080/09273948.2023.2253471.
- [87] K. Du *et al.*, «Deep learning system for screening AIDS-related cytomegalovirus retinitis with ultra-wide-field fundus images», *Heliyon*, vol. 10, n.º 10, p. e30881, 2024, doi: 10.1016/j.heliyon.2024.e30881.
- [88] A. D. Chakravarthy *et al.*, «An Approach Towards Automatic Detection of Toxoplasmosis using Fundus Images», en *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2019, pp. 710-717. doi: 10.1109/BIBE.2019.00134.
- [89] D. Milad, F. Antaki, A. Bernstein, S. Touma, y R. Duval, «Automated Machine Learning versus Expert-Designed Models in Ocular Toxoplasmosis: Detection and Lesion Localization Using Fundus Images», *Ocular Immunology and Inflammation*, pp. 1-7, 2024, doi: 10.1080/09273948.2024.2319281.
- [90] Caracol Noticias, «Médica colombiana es reconocida como la número uno del mundo en toxoplasmosis ocular». [En línea]. Disponible en: <https://www.noticiascaracol.com/colombia/medica-colombiana-es-reconocida-como-la-numero-uno-del-mundo-en-toxoplasmosis-ocular#:~:text=La%20médica%20colombiana%20Ligia%20Alejandra,su%20aporte%20a%20la%20ciencia.>
- [91] «Google Colaboratory». [En línea]. Disponible en: <https://colab.research.google.com>
- [92] A. Ramos-Michel, M. Pérez-Cisneros, E. Cuevas, y D. Zaldivar, «Image Classification with Convolutional Neural Networks», en *Metaheuristics in Machine Learning: Theory and Applications*, vol. 967, D. Oliva, E. H. Houssein, y S. Hinojosa, Eds., Cham: Springer International Publishing, 2021, pp. 445-473. doi: 10.1007/978-3-030-70542-8_18.
- [93] Y. Pan *et al.*, «Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases», *Front. Physiol.*, vol. 14, p. 1126780, 2023, doi: 10.3389/fphys.2023.1126780.
- [94] O. Rainio, J. Teuho, y R. Klén, «Evaluation metrics and statistical tests for machine learning», *Sci Rep*, vol. 14, n.º 1, p. 6086, 2024, doi: 10.1038/s41598-024-56706-x.
- [95] H. M y S. M.N, «A Review on Evaluation Metrics for Data Classification Evaluations», *IJDKP*, vol. 5, n.º 2, pp. 01-11, 2015, doi: 10.5121/ijdkp.2015.5201.
- [96] A. Villalobos-Pérez *et al.*, «Referral Process in Patients with Uveitis: A Challenge in the Health System», *OPHTH*, vol. Volume 15, pp. 1-10, ene. 2021, doi: 10.2147/OPHTH.S287766.
- [97] H. Hong, O. J. Mújica, J. Anaya, V. C. Lansingh, E. López, y J. C. Silva, «The Challenge of Universal Eye Health in Latin America: distributive inequality of ophthalmologists in 14 countries», *BMJ Open*, vol. 6, n.º 11, p. e012819, 2016, doi: 10.1136/bmjopen-2016-012819.
- [98] J. G. Shantha *et al.*, «Development of a screening eye clinic for Ebola virus disease survivors: Lessons learned and rapid implementation at ELWA Hospital in Monrovia, Liberia 2015», *PLoS Negl Trop Dis*, vol. 13, n.º 3, p. e0007209, 2019, doi: 10.1371/journal.pntd.0007209.
- [99] A. J. Thirunavukarasu *et al.*, «Democratizing Artificial Intelligence Imaging Analysis With Automated Machine Learning: Tutorial», *J Med Internet Res*, vol. 25, p. e49949, 2023, doi: 10.2196/49949.

- [100] W. J. Zhang, G. Yang, Y. Lin, C. Ji, y M. M. Gupta, «On Definition of Deep Learning», en *2018 World Automation Congress (WAC)*, jun. 2018, pp. 1-5. doi: 10.23919/WAC.2018.8430387.
- [101] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, y T. Ganslandt, «Transfer learning for medical image classification: a literature review», *BMC Med Imaging*, vol. 22, n.º 1, p. 69, 2022, doi: 10.1186/s12880-022-00793-7.
- [102] J. Demšar y B. Zupan, «Hands-on training about overfitting», *PLoS Comput Biol*, vol. 17, n.º 3, p. e1008671, mar. 2021, doi: 10.1371/journal.pcbi.1008671.

14. Anexo 1: Glosario médico

- **Protozoo:** Los protozoos son microorganismos unicelulares que pueden vivir y multiplicarse dentro de los organismos humanos, causando infecciones. A menudo se transmiten a través de agua o alimentos contaminados. Un ejemplo común de un protozoo es *Toxoplasma gondii*, el causante de la toxoplasmosis, una infección que puede afectar los ojos y otras partes del cuerpo.
- **Retinocoroiditis:** Es la inflamación simultánea de la retina y la coroides, dos capas del ojo que son esenciales para la visión. La retina convierte la luz en señales que el cerebro puede interpretar, mientras que la coroides es una capa rica en vasos sanguíneos que nutre la retina. La retinocoroiditis, que puede ser causada por infecciones como la toxoplasmosis, puede provocar daño visual permanente si no se trata adecuadamente.
- **Vitreítis:** Es la inflamación del humor vítreo, una sustancia gelatinosa que llena el centro del ojo y le da su forma. Cuando el humor vítreo se inflama, puede volverse turbio, lo que afecta la claridad de la visión. La vitreítis puede estar relacionada con infecciones o enfermedades autoinmunes.
- **Mácula:** Es la parte central de la retina responsable de la visión detallada y de alta resolución. Esta pequeña área es esencial para actividades como leer, escribir, y reconocer rostros. Las enfermedades que afectan la mácula, como la degeneración macular, pueden llevar a una pérdida significativa de la visión central.
- **Exudados:** Son acumulaciones de proteínas, lípidos y otros fluidos que se filtran desde los vasos sanguíneos hacia los tejidos circundantes durante procesos inflamatorios o infecciosos. En la retina, los exudados suelen aparecer como manchas blanquecinas y son un signo de daño a los vasos sanguíneos, común en enfermedades como la retinopatía diabética o la toxoplasmosis ocular.
- **Edema:** Se refiere a la acumulación anormal de líquido en los tejidos del cuerpo, lo que provoca hinchazón. En el contexto ocular, el edema macular ocurre cuando se acumula

líquido en la mácula, lo que puede distorsionar la visión y llevar a una pérdida visual si no se trata.

- **Necrosis:** Es la muerte de las células o tejidos en una parte del cuerpo debido a una falta de flujo sanguíneo, infección u otro daño. En el ojo, la necrosis retiniana puede ser causada por infecciones graves, como la toxoplasmosis, y puede resultar en la pérdida irreversible de visión en las áreas afectadas.
- **Papilitis:** Es una inflamación del nervio óptico en la parte que se conecta con el ojo (la papila óptica). Esta condición puede causar visión borrosa o pérdida de visión, así como dolor al mover el ojo. La papilitis puede ser un signo de una infección o una enfermedad autoinmune.
- **Vasculitis:** Es la inflamación de los vasos sanguíneos, lo que puede alterar el flujo normal de sangre a los tejidos. En el ojo, la vasculitis puede afectar los vasos de la retina, lo que resulta en daños a la visión. Las vasculitis oculares pueden ser causadas por infecciones o por enfermedades inflamatorias.
- **Arteriolitis:** Es la inflamación de las arteriolas, que son los vasos sanguíneos más pequeños del cuerpo. En el ojo, la arteriolitis puede restringir el flujo sanguíneo hacia la retina, lo que provoca daños en los tejidos y contribuye a problemas visuales como la retinocoroiditis.
- **Quimioluminiscencia:** Es una técnica de laboratorio que mide la luz emitida durante una reacción química. Se utiliza en análisis médicos para detectar la presencia de ciertos compuestos en el cuerpo, como anticuerpos o marcadores de enfermedades, mediante la medición de la luz que emiten cuando reaccionan con un químico específico.
- **IgG:** Es uno de los principales tipos de anticuerpos que circulan en la sangre y otros fluidos del cuerpo. Estos anticuerpos son parte de la respuesta inmune del cuerpo a las infecciones, ayudando a identificar y neutralizar patógenos como bacterias y virus. Los niveles de IgG son una indicación de una infección pasada o una inmunización, y proporcionan inmunidad a largo plazo.
- **IgM:** Es otro tipo de anticuerpo que el cuerpo produce al inicio de una infección. Los niveles de IgM aumentan rápidamente tras la exposición a un patógeno y son una indicación de infección reciente. A medida que la infección progresa, los niveles de IgM disminuyen y son reemplazados por IgG, que proporciona una defensa más duradera.
- **Reacción en cadena de la polimerasa (PCR):** La PCR es una técnica de laboratorio

utilizada para amplificar pequeñas cantidades de material genético, lo que permite detectar la presencia de ADN o ARN de virus, bacterias, u otros microorganismos en una muestra. En el caso de las infecciones oculares, la PCR puede utilizarse para analizar muestras de humor acuoso (el líquido que llena la parte frontal del ojo) o vítreo, lo que permite diagnosticar infecciones como la toxoplasmosis ocular. El procedimiento implica la extracción cuidadosa de una pequeña cantidad de este líquido del ojo, que luego se somete a múltiples ciclos de calentamiento y enfriamiento para replicar el material genético del microorganismo, permitiendo su identificación precisa.

15. Anexo 2: Experimentos realizados para la selección de hiperparámetros

Tabla suplementaria 1: Modificaciones arquitectura ResNetV2 de acuerdo con el número de épocas.

# de épocas	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
10	0.79	0.67	0.73	0.79	0.88	0.83	0.79
20	0.95	0.46	0.62	0.72	0.98	0.83	0.76
30	1.00	0.37	0.54	0.69	1.00	0.82	0.74
40	0.89	0.67	0.77	0.81	0.94	0.87	0.83
50	0.84	0.70	0.76	0.81	0.91	0.86	0.82

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 2: Modificaciones arquitectura ResNetV2 de acuerdo con el uso del descongelamiento de capas.

Descongelamiento de capas	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
Sin descongelamiento	0.79	0.67	0.73	0.79	0.88	0.83	0.79
Con descongelamiento	0.91	0.87	0.89	0.91	0.94	0.93	0.91

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 3: Modificaciones arquitectura ResNetV2 de acuerdo con modificaciones en el tamaño de lote.

Tamaño de lote	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
64	0.79	0.67	0.73	0.79	0.88	0.83	0.79
32	0.70	0.61	0.65	0.75	0.82	0.78	0.73
128	0.92	0.52	0.67	0.74	0.97	0.84	0.78.
256	1.00	0.15	0.26	0.63	1.00	0.77	0.65

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 4: Modificaciones arquitectura ResNetV2 de acuerdo con la tasa de aprendizaje.

Tasa de aprendizaje	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
1e-2	0.79	0.67	0.73	0.79	0.88	0.83	0.79
1e-3	0.94	0.37	0.53	0.69	0.98	0.81	0.73
1e-4	0.63	0.67	0.70	0.76	0.73	0.74	0.70
1e-5	0.41	0.24	0.30	0.59	0.76	0.66	0.54
5e-4	0.90	0.41	0.57	0.70	0.97	0.82	0.74
5e-3	0.88	0.15	0.26	0.62	0.98	0.76	0.64

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 5: Modificaciones arquitectura ResNetV2 de acuerdo con el uso de Synthetic Minority Over-sampling Technique.

	Activo			Inactivo			
SMOTE	Precisión	Recall	F1	Precisión	Recall	F1	Exactitud General (Accuracy)
Sin SMOTE	0.79	0.67	0.73	0.79	0.88	0.83	0.79
Con SMOTE	0.90	0.57	0.69	0.76	0.95	0.85	0.79

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. **SMOTE:** Synthetic Minority Over-sampling Technique.

Tabla suplementaria 6: Modificaciones arquitectura ResNetV2 de acuerdo con el uso de peso de clases.

	Activo			Inactivo			
Class weight (CW)	Precisión	Recall	F1	Precisión	Recall	F1	Exactitud General (Accuracy)
Sin CW	0.79	0.67	0.73	0.79	0.88	0.83	0.79
Con CW	1.00	0.22	0.36	0.65	1.00	0.79	0.67

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. **CW:** Class-weight (peso de clases).

Tabla suplementaria 7: Modificaciones arquitectura ResNetV2 de acuerdo con el uso de la desactivación aleatoria.

Dropout	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
0.2	0.79	0.67	0.73	0.79	0.88	0.83	0.79
0.3	1.00	0.39	0.56	0.70	1.00	0.92	0.75
0.4	0.95	0.41	0.58	0.71	0.98	0.82	0.75
0.5	1.00	0.28	0.44	0.67	1.00	0.80	0.70

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. **Dropout:** Desactivación aleatoria.

Tabla suplementaria 8: Modificaciones arquitectura ResNetV2 de acuerdo con el optimizador

Optimizador	Activo			Inactivo			Exactitud General (Accuracy)
	Precisión	Recall	F1	Precisión	Recall	F1	
Adam	0.79	0.67	0.73	0.79	0.88	0.83	0.79
SGD	1.00	0.15	0.26	0.63	1.00	0.77	0.65
RMSprop	0.93	0.30	0.46	0.67	0.98	0.80	0.70
Adagrad	0.79	0.57	0.66	0.75	0.89	0.81	0.75
Adadelta	0.45	0.48	0.46	0.62	0.59	0.60	0.54
Nadam	0.95	0.39	0.55	0.70	0.98	0.82	0.74

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. SMOTE: Synthetic Minority Over-sampling Technique. **Adam:** Adaptive Moment Estimation. **SGD:** Stochastic Gradient Descent. **RMSprop:** Root Mean Square Propagation. **Adagrad:** Adaptive Gradient Algorithm. **Adadelta:** Adaptive Delta. **Nadam:** Nesterov-accelerated Adaptive Moment Estimation.

Tabla suplementaria 9: Modificaciones arquitectura ResNetV2 de acuerdo con el optimizador

Modelo	Función de activación	Épocas	Optimizador	Clase	Precisión	Recall	F1	Exactitud General (Accuracy)	AUC
General	ReLU	10	Adam	Activo	0.94	0.69	0.79	0.85	0.94
				Inactivo	0.82	0.97	0.89		
Modelo General con Submuestreo	ReLU	10	Adam	Activo	0.94	0.71	0.81	0.86	0.95
				Inactivo	0.83	0.97	0.89		
Modelo General con Submuestreo	ELU	15	Nadam	Activo	0.86	0.69	0.77	0.82	0.92
				Inactivo	0.81	0.92	0.86		
Modelo General con Submuestreo	SELU	10	Nadam	0.80	0.80	0.80	0.80	0.83	0.91
				0.86	0.86	0.86	0.86		

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. **ReLU:** Rectified Linear Unit. **ELU:** Exponential Linear Unit. **SELU:** Scaled Exponential Linear Unit. **Adam:** Adaptive Moment Estimation. **Nadam:** Nesterov-accelerated Adaptive Moment Estimation **AUC:** Area under the ROC curve (área bajo la curva).

Tabla suplementaria 10: Modificaciones arquitectura VGG16 de acuerdo con el optimizador

Modelo	Función de activación	Épocas	Optimizador	Clase	Precisión	Recall	F1	Exactitud General (Accuracy)	AUC
VGG16 SGG	ReLU	15	SGD	Activo	0.43	0.98	0.60	0.45	0.79
				Inactivo	0.86	0.09	0.17		
VGG16 DATA augmentati on	ReLU	20	Nadam	Activo	0.88	0.62	0.73	0.80	0.90
				Inactivo	0.78	0.94	0.85		
VGG16 DATA augmentati on Sin congelar	ReLU	20	Nadam	Activo	0.87	0.73	0.80	0.84	0.94
				Inactivo	0.83	0.92	0.87		

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. **ReLU:** Rectified Linear Unit. **SGD:** Stochastic Gradient Descent. **Adam:** Adaptive Moment Estimation. **Nadam:** Nesterov-accelerated Adaptive Moment Estimation. **AUC:** Area under the ROC curve (área bajo la curva).

Tabla suplementaria 11: Modificaciones arquitectura VGG16 con 10 épocas sin descongelamiento.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.90	0.42	0.58	0.70	0.97	0.82	0.74

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 12: Modificaciones arquitectura VGG16 con 20 épocas sin descongelamiento.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.91	0.47	0.62	0.72	0.97	0.83	0.76

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 13: Modificaciones arquitectura VGG16 con 10 épocas con descongelamiento.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.41	1.00	0.58	0.00	0.00	0.00	0.41

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 14: Modificaciones arquitectura VGG16 con 10 épocas con Synthetic Minority Over-sampling Technique.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.88	0.62	0.73	0.78	0.94	0.85	0.80

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 15: Modificaciones arquitectura VGG16 con 20 épocas con Synthetic Minority Over-sampling Technique.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.81	0.67	0.73	0.79	0.89	0.84	0.79

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total.

Tabla suplementaria 16: Modificaciones arquitectura VGG16 con 10 épocas con peso de clases.

Activo			Inactivo			Exactitud General (Accuracy)
Precisión	Recall	F1	Precisión	Recall	F1	
0.76	0.56	0.64	0.74	0.88	0.80	0.74

Precisión: Proporción de predicciones correctas entre todos los casos predichos como positivos. **Recall:** Proporción de casos positivos reales que fueron correctamente identificados. **F1:** Media armónica entre precisión y recall, útil cuando los datos están desbalanceados. **Exactitud (Overall Accuracy):** Porcentaje de predicciones correctas en total. SMOTE: Synthetic Minority Over-sampling Technique