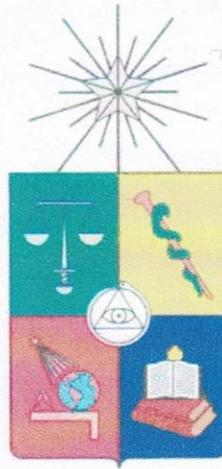


UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
Escuela de Tecnología Médica



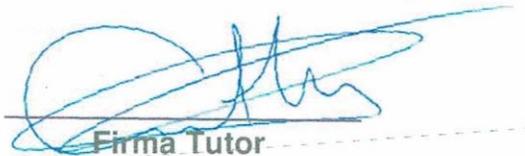
TESIS PROFESIONAL

Para optar al título profesional de Tecnólogo Médico con mención

“Evaluación comparativa de modelos prediseñados de Inteligencia artificial, basados en imágenes de OCT, para la detección de actividad inflamatoria intraocular”

Alumno: Pedro Nicolás Alonso Torrejon Gutiérrez

**Tutor: Dr. Cristhian Urzúa S.
Cotutora: Dra. Loreto Cuitiño T.**



Firma Tutor



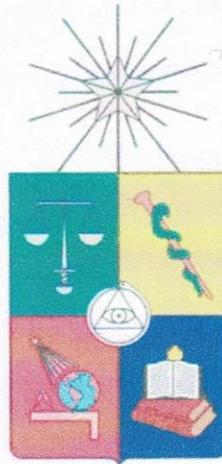
Firma Cotutora

Fecha: 14 de diciembre de 2022



Departamento
de Tecnología Médica
FACULTAD DE MEDICINA
UNIVERSIDAD DE CHILE

UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
Escuela de Tecnología Médica



TESIS PROFESIONAL

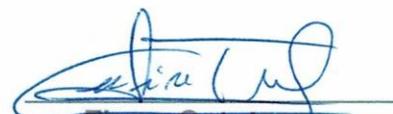
Para optar al título profesional de Tecnólogo Médico con mención

“Evaluación comparativa de modelos prediseñados de Inteligencia artificial, basados en imágenes de OCT, para la detección de actividad inflamatoria intraocular”

Alumno: Pedro Nicolás Alonso Torrejon Gutiérrez

Tutor: Dr. Cristhian Urzúa S.
Cotutora: Dra. Loreto Cuitiño T.


Firma Tutor


Firma Cotutora

Fecha: 14 de diciembre de 2022

Dedicatoria:

El presente escrito va dedicado a toda mi familia por las palabras de aliento en cada momento difícil, y en especial a mis padres por su incansable apoyo e infinito cariño, por ser pilar y motor en cada etapa de mi vida, por darme las herramientas para enfrentarme al mundo y moldearme en la persona que soy hoy.

Pedro Torrejon G.

Agradecimientos:

Un profundo agradecimiento a todo el equipo del Laboratorio de Enfermedades Autoinmunes Oculares y Sistémicas (LEAOS), por el buen recibimiento y ayuda, en especial a Fabian Vega por su ayuda corrigiendo este escrito.

A mis tutores Dr. Cristhian Urzúa y la Dra. Loreto Cuitiño, por la motivación y confianza depositados en mí.

ACTA CURSO TRABAJO DE INVESTIGACIÓN

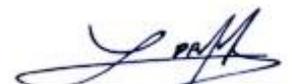
El Sr. Pedro Nicolas Alonso Torrejon Gutierrez, estudiante de Tecnología Médica con Mención en Oftalmología y Optometría, cumpliendo con los requisitos establecidos en el plan de estudio, realizó durante el décimo semestre de la carrera, la Tesis Profesional titulada: *“Evaluación comparativa de modelos prediseñados de inteligencia artificial, basados en imágenes de OCT, para la detección de actividad inflamatoria intraocular”* dirigida por el Dr. Cristian Urzua y la Dra. Loreto Cuitiño, ambos del Laboratorio de enfermedades autoinmunes oculares y sistémicas, Hospital Clínico de la Universidad de Chile.

La Escuela de Tecnología Médica designó para su corrección una Comisión integrada por: Prof. TM Hernán Torres MPH y la Prof. Katherine Anabalón ambos académicos del Departamento de Tecnología Médica, Facultad de Medicina de la Universidad de Chile.

La calificación obtenida se detalla a continuación:

Corrector 1 Hernán Torres	6.55	25%
Corrector 2 Katherine Anabalón	6.92	25%
Tutor(es) Guía: Cristian Urzúa y Loreto Cuitiño		
Evaluación intermedia	7.00	25%
Nota final tutor	7.00	25%
Nota final tesis profesional	6.87	

En consecuencia el estudiante Pedro Nicolas Alonso Torrejon Gutierrez aprueba satisfactoriamente la asignatura.



Prof. Iván Plaza Rosales
Coordinador(a) curso
Trabajo de Investigación



Prof. Hernán Torres.
PEC curso
Trabajo de Investigación

Tabla de contenido

Resumen:	1
Introducción:	2
Uveítis	2
OCT y su uso en la evaluación clínica de la uveítis	4
Inteligencia artificial y sus aplicaciones clínicas en uveítis.	7
Objetivos:	14
General:	14
Específicos:.	14
Hipótesis:	14
Materiales y métodos:	15
Diseño experimental:	15
Población:	15
Lugar de realización:	15
Tamaño de Muestra:	15
Criterios de inclusión y exclusión:	16
Diseño Metodológico:	16
Estrategia de entrenamiento de modelos de DL:	16
Recolección de datos:	17
Adquisición, selección y preprocesamiento de imágenes:	18
Entrenamiento de los modelos y optimización del LR:	18
Análisis comparativo del desempeño de los modelos:	19
Resultados:	20
Características demográficas y clínicas:	20
Optimización del LR para cada red convolucional según accuracy	21
Entrenamiento y validación final de los modelos	24
Análisis comparativo del desempeño de los modelos generados	25
Discusión:	30
Conclusiones:	35
Bibliografía:	36
Anexos:	41

Resumen:

La uveítis es la inflamación de la capa media del globo ocular, la úvea, pudiendo afectar la integridad y la función de estructuras adyacentes. El uso de imágenes de tomografía de coherencia óptica (OCT) junto al desarrollo del *machine learning* tiene el potencial de discriminar de forma automatizada y objetiva la presencia o ausencia de inflamación ocular en pacientes con uveítis. El objetivo de esta investigación es producir un modelo de clasificación utilizado para reconocer la presencia de inflamación intraocular en pacientes con uveítis a través OCT y evaluar su desempeño.

Se utilizaron imágenes de pacientes con uveítis adquiridas mediante OCT clasificadas según su estado de inflamación intraocular en activas o inactivas por un especialista. Posteriormente, las imágenes fueron utilizadas para entrenar modelos de clasificación generados por 4 arquitecturas de *Deep Learning* públicamente disponibles y se compararon los desempeños de cada uno.

Fue posible reentrenar varios modelos clasificadores de imágenes para su uso en el análisis de imágenes OCT, resultando en valores de *accuracy* mayores al 80%, pero solo uno (ResNet) superó el 95% de efectividad en la evaluación de la inflamación intraocular en pacientes con uveítis. Se espera que esta herramienta pueda instaurarse como un apoyo dentro la práctica clínica para el diagnóstico de la uveítis y como un primer estándar a superar al generar nuevos clasificadores en el futuro.

Introducción:

Uveítis

El término uveítis se refiere a un grupo de entidades clínicas con diversas etiologías y manifestaciones clínicas que tienen como factor común la inflamación intraocular presente en la úvea, la capa vascular intermedia del ojo compuesta por el iris, el cuerpo ciliar y la coroides ⁽¹⁾. Sin embargo, el término se aplica en la práctica clínica de forma más extensa, abarcando un conjunto de enfermedades que afectan tanto al tracto uveal como a las estructuras intraoculares adyacentes, las cuales incluyen el vítreo, retina, nervio óptico y vasos sanguíneos ⁽²⁾.

Esta patología es el tipo más común de enfermedad inflamatoria ocular y es una causa importante de ceguera y discapacidad visual tanto en países desarrollados como en vías de desarrollo ⁽³⁾. Su epidemiología es variable, dependiendo de la región donde se estudie. Se estima que la uveítis tiene una incidencia entre 17-52 casos por 100.000 habitantes al año, con una prevalencia de 38-714 casos por 100.000 habitantes⁽⁴⁾. Si bien puede aparecer a cualquier edad, se manifiesta predominantemente entre los 20 y 59 años, afectando principalmente adultos en edad laboral⁽³⁾.

En cuanto a su etiología, puede estar asociada a diversas causas, ya que puede tener su origen en la manifestación de trastornos autoinmunes o infecciones, también puede ocurrir secundario a la ingesta de medicamentos o exposición a toxinas; asimismo, puede ocurrir sin una causa conocida (idiopática)⁽⁵⁾. Cabe destacar dentro de las causas de uveítis la enfermedad de Vogt-Koyanagi-Harada (VKH), al ser la causa de uveítis con mayor presencia dentro de nuestro país, siendo superado solo

por las uveítis idiopáticas⁽⁶⁾, siendo responsable de aproximadamente el 17% de las uveítis no infecciosas en Chile, una prevalencia 2 a 3 veces mayor que en Estados Unidos o países europeos ⁽⁷⁾.VKH corresponde a una enfermedad autoinmune, sistémica, caracterizada por presentar una inflamación bilateral que afecta a todas las estructuras intraoculares, frecuentemente asociado con desprendimientos de retina y también con diversas manifestaciones extraoculares (neurológicas, auditivas y dermatológicas), manifestada principalmente en pacientes de sexo femenino⁽⁸⁾.

La clasificación de la uveítis ha sido consensuada por un grupo de especialistas, conocido como el sistema SUN (*Standardization of Uveitis Nomenclature*)⁽⁹⁾, la cual recientemente ha estandarizado los criterios de clasificación de las diversas entidades correspondientes y sus presentaciones clínicas ⁽¹⁰⁾. Es así como la uveítis puede ser clasificada según la ubicación anatómica de la inflamación (anterior, posterior, intermedia o panuveítis), forma de aparición (repentina o insidiosa), duración (limitada o persistente) y curso clínico (aguda, recurrente o crónica) ⁽⁹⁾. Para la aplicación de este sistema se establecieron parámetros cuantificables incluyendo la presencia de células infiltrantes y *flare* en cámara anterior y *haze* vítreo para medir la presencia de actividad inflamatoria ^(11,12). Otro esfuerzo que merece distinguirse es el del *Disease Activity Index for Patients with Uveitis* (UVEDAI), el cual añadió otros parámetros, ya que los criterios SUN solo contemplaban la presencia de inflamación de cámara anterior y grado de *haze* vítreo⁽¹³⁾. Los parámetros clínicos que se utilizan hoy en día para la evaluación de la uveítis, avaladas por los estudios previamente mencionadas, son los siguientes:

- Presencia de células o *flare* (proteínas) en la cámara anterior, debido a que es el procedimiento estándar para la detección de inflamación en el segmento anterior

del ojo⁽¹⁴⁾.

- Puntaje mediante oftalmoscopia binocular indirecta (BIO por sus siglas en inglés *Binocular Indirect Ophthalmoscopy*), que consiste en una escala de 6 puntos que estima la claridad del humor vítreo utilizando un oftalmoscopio indirecto. Se recomienda esta escala, ya que es la forma de mayor aceptación para medir la inflamación en la cavidad vítrea⁽¹⁴⁾.
- Signos de inflamación de polo posterior son la presencia de vasculitis, alteraciones inflamatorias en la coroides y/o retina, o presencia de papilitis⁽¹⁴⁾.

A pesar del importante avance que implicó la instauración de un sistema universal de evaluación de la uveítis, este no está libre de falencias. En particular, los criterios son subjetivos y altamente dependientes de la experiencia del examinador, lo que puede traer como consecuencias diferencias en la evaluación clínica, y por ende en el diagnóstico y el tratamiento⁽¹⁵⁾. Por otro lado, existen casos de uveítis idiopáticas de los cuales se desconoce su causa, y al no estar caracterizada la enfermedad su evaluación y tratamiento es desafiante.

Por estas razones, existe una necesidad de desarrollar nuevas estrategias de evaluación de la uveítis que sean objetivas y confiables, estableciendo la uveítis como una patología clave en el área investigativa.

OCT y su uso en la evaluación clínica de la uveítis

Una tecnología que ha tenido una gran relevancia en la evaluación de la uveítis es la tomografía de coherencia óptica (OCT, por sus siglas en inglés). Desde su introducción a las prácticas clínicas hace ya más de dos décadas y debido a su constante evolución ha cambiado drásticamente el entendimiento y manejo de

muchas condiciones oculares⁽¹⁶⁾. OCT es una tecnología que nos permite obtener imágenes de alta resolución de las microestructuras oculares en vivo con una resolución mayor a cualquier otra tecnología no invasiva, usada en la especialidad de oftalmología ⁽¹⁶⁾. Esto se logra mediante el uso de luz cercana al espectro infrarrojo, la que impacta en los tejidos oculares y se reflejan de distintas formas; la diferencia en la reflexión de la luz de diferente longitud de onda no puede ser medida directamente. Para lograr esto, se utiliza un interferómetro que produce una medida de referencia a través de la división del haz de luz, dirigiendo una parte al tejido a visualizar y la otra parte hacia un brazo de referencia con una longitud de onda conocida⁽¹⁷⁾.

Al utilizar este equipo se pueden utilizar distintas técnicas para poder visualizar las diferentes estructuras presentes al interior del ojo. Para ver en mayor detalle la cavidad vítrea, se utiliza la técnica EVI-OCT (*Enhanced Vitreous Imaging*), que consiste en dirigir la luz infrarroja del OCT, de manera habitual en retina, para luego cambiar el enfoque agregando +2 dioptrías. Esto permite una mejor visualización de las estructuras presentes en la cavidad vítrea, como también de la interfaz vitreoretinal⁽¹⁸⁾. Para la observación de la zona coroidea, se han utilizado longitudes de onda mayores a las utilizadas convencionalmente, lo que permite una penetración de la luz hacia zonas más profundas del ojo mejorando la imagen. Junto con esto, se puede lograr una imagen aún mejor utilizando la técnica EDI-OCT (*Enhanced Depth imaging*) que consiste en situar el OCT más cerca del ojo, lo que genera una imagen invertida y ubica la coroides más cerca de la zona de mayor nitidez, dando como resultado una mayor resolución de los detalles de la coroides⁽¹⁹⁾.

La uveítis representa un fenómeno de inflamación intraocular que promueve cambios en el suministro de sangre, los componentes celulares y el entorno químico de la zona afectada ⁽¹⁶⁾. Estas modificaciones, tanto a nivel estructural como funcional, pueden tener como consecuencia cambios transitorios o permanentes en los tejidos afectados, lo que puede estudiarse mediante OCT ⁽¹⁷⁾. Utilizando un solo *scan* de OCT sobre una lesión inflamatoria es posible evaluar la profundidad de la alteración, estructuras oculares afectadas y sus patrones de evolución, proporcionando ayuda en relación con los diagnósticos diferenciales, a la predicción de los resultados del tratamiento y el seguimiento de la patología ⁽¹⁶⁾.

Cada tipo de uveítis tiene hallazgos característicos al OCT, dependiendo de las estructuras oculares involucradas y el tipo de proceso inflamatorio o infeccioso involucrado ⁽¹⁷⁾. Aun así, algunos de los cambios transitorios o permanentes en la fisiología y/o anatomía del ojo ocurren independiente de la entidad causante de la uveítis ⁽¹⁶⁾. El aumento en la permeabilidad vascular es una de las características clave al momento de observar inflamación. Esto resulta en el escape de proteínas y la migración de células inflamatorias desde el espacio intravascular hacia la zona inflamada ⁽¹⁷⁾. En los ojos, estos cambios son visibles clínicamente a través de la presencia de células inflamatorias y turbiedad en la cámara anterior y en la cavidad vítrea ⁽²⁰⁾, parámetros que son captados como parte de los criterios SUN para la clasificación de la uveítis. Estas células inflamatorias que flotan en la cavidad vítrea son visibles como puntos hiperreflectivos en las imágenes tomadas con la técnica EVI-OCT ⁽¹⁸⁾. El *haze* vítreo también es detectable mediante OCT, como un aumento en el brillo de la cavidad vítrea ⁽²¹⁾. El aumento del grosor coroideo también puede ser detectado mediante OCT en la mayor parte de los casos de uveítis, incluyendo

aquellos que afectan principalmente el segmento anterior ⁽²¹⁾, esto es una medición indirecta del aumento de flujo sanguíneo dentro de la coroides inducido por la inflamación y puede ser monitoreado para evaluar el estado inflamatorio a través del tiempo ⁽¹⁶⁾. La inflamación de la cabeza del nervio óptico también es común en uveítis y la posterior hinchazón del disco óptico puede ser fácilmente evaluada y monitorizada mediante OCT ⁽²³⁾. Dado que las alteraciones detectables por OCT aquí descritas presentan una íntima asociación con la inflamación, suelen utilizarse durante la evaluación de la uveítis para corroborar observaciones clínicas o, incluso, detectar inflamación inadvertida por el especialista, denominada inflamación subclínica ⁽¹⁶⁾.

Inteligencia artificial y sus aplicaciones clínicas en uveítis.

Otra tecnología que ha tomado relevancia en los últimos tiempos en el área de la salud en general es la llamada Inteligencia Artificial (IA), la cual se ha ido incorporando en gran parte de los procesos de atención clínica con el objetivo de mejorarlos. De este modo, se han podido acortar los tiempos de diagnóstico y lograr una mayor precisión en estos, abriendo el camino a una mejor atención médica en general.

La Inteligencia Artificial corresponde a una rama de las ciencias de la computación que incluye conceptos muy transversales relacionados con la lógica y el aprendizaje ⁽²⁴⁾. Se trata, por lo tanto, de diseñar herramientas informáticas que simulen procesos de inteligencia humana que incluyen el aprendizaje, el razonamiento y la autocorrección ⁽²⁴⁾.

Un subcampo de estudio de la Inteligencia artificial es el *Machine Learning* (ML), dedicado al desarrollo de modelos computacionales de aprendizaje automatizado ⁽²⁵⁾. Esta disciplina puede ser comprendida de dos formas. Por un lado, se refiere a la

habilidad de una máquina para mejorar su desempeño basado en resultados anteriores. Por otra parte, se puede entender como el estudio de cómo crear computadoras que puedan aprender desde experiencias anteriores y modificar su actividad basándose en ese aprendizaje, diferenciándose así de las computadoras tradicionales cuya actividad no cambia a menos que un programador lo indique explícitamente ⁽²⁵⁾. Luego, como subenfoco del ML aparece el *Deep Learning* (DL), que tiene sus bases en el funcionamiento del sistema neurológico humano, con redes ordenadas de manera jerárquica, con unidades de proceso o “neuronas” artificiales especializadas en la identificación de características específicas en imágenes, sonidos o agrupaciones de datos ⁽²⁶⁾. De este mismo concepto nacen las redes convolucionales, sistemas basados en DL capaces de aprender diferentes niveles de abstracción mediante múltiples capas de “neuronas” organizadas para reconocer características específicas. Los procesamientos más simples realizados por la red neuronal ocurren en las primeras capas (p. ej., identificación de bordes o colores), luego las diversas combinaciones de estos en las capas medias, para finalmente en la última capa conseguir identificar qué es exactamente lo que se está viendo ⁽²⁷⁾.

Anteriormente, se ha aplicado este tipo de modelos a distintas áreas de la oftalmología. Por ejemplo, se ha usado para clasificar casos de glaucoma ⁽²⁸⁾, en predicción de la curvatura corneal y astigmatismo en implantes intracorneales en queratocono ⁽²⁹⁾, detección de retinopatía y edema macular diabético ⁽³⁰⁾, para establecer criterios de clasificación para distintos tipos de uveítis ^(11,12) entre otros.

Existen tres fases principales en la generación un modelo clasificador de imágenes, a saber, entrenamiento, prueba y validación ⁽³¹⁾.

En la primera fase (entrenamiento) se enseña al modelo a reconocer patrones

característicos utilizando pautas de entrada, conocidas como datos de entrenamiento que pueden variar según la naturaleza de los datos y la necesidad a cubrir, por lo tanto, pueden caer en distintas clasificaciones ⁽³¹⁾.

1. Aprendizaje supervisado: cuando se tiene un conjunto de entrenamiento formado por datos etiquetados, es decir, entradas de ejemplo y el resultado que se espera recibir por parte del modelo de clasificación, así este aprende una regla general que asigna las entradas y los resultados ⁽³²⁾.
2. Aprendizaje no supervisado: el conjunto de entrenamiento no posee ninguna etiqueta, la principal tarea del modelo de clasificación es encontrar las soluciones por sí misma, en otras palabras, patrones o estructuras que agrupen los datos no etiquetados ⁽³²⁾.
3. Aprendizaje por refuerzo: en este caso, los datos de entrenamiento se dan sólo como retroalimentación, en forma de recompensa o castigo, dándole una guía al sistema de los resultados esperados, sin clasificarlos de manera explícita ⁽³²⁾.
4. Aprendizaje por transferencia o *transfer learning*: consiste en la transferencia del aprendizaje logrado por un modelo generado para una aplicación determinada a otro con el fin de obtener soluciones para un problema diferente, pero de similar naturaleza ⁽³²⁾.

Además, es importante definir los hiperparámetros a utilizar para esta fase, los cuales consisten en variables que nos permiten optimizar el desempeño de cada modelo durante el entrenamiento⁽³³⁾. A continuación, se explicarán algunos de estos hiperparámetros y su importancia en el entrenamiento de modelos predictivos mediante DL.

Batch size o tamaño de lote: se define como la cantidad de muestras que se utilizaran para cada iteración del entrenamiento⁽³⁴⁾. Para definir este parámetro se debe considerar un balance, ya que, si se utiliza un *batch size* muy pequeño, significa que el clasificador tiene en memoria una baja cantidad de datos, lo que acelera el entrenamiento, pero se corre el riesgo de que se pasen por alto características que pudiesen ser valiosas para la predicción. Por el contrario, de ser muy grande, es más probable que se tengan en cuenta todas las características importantes a la hora de aprender, pero tiene un impacto negativo en los tiempos de entrenamiento y los requerimientos de memoria RAM ⁽³⁴⁾.

Epochs o épocas: consisten en la cantidad de veces que la totalidad de datos pasaran por el clasificador con el fin de ajustarlo. Se llega a un punto donde se logra el valor mínimo de error para cada modelo en concreto, y a partir de ahí además de consumir tiempo y recursos computacionales innecesarios, se corre el riesgo de producir un “sobreajuste”. Un sobreajuste (u *overfitting* en inglés) es el fenómeno en el cual el modelo aprende de manera sobresaliente de los datos de entrenamiento, pero que no es capaz de producir la respuesta apropiada frente a datos nuevos. Por otro lado, un número muy bajo de épocas puede provocar que el modelo no aprenda lo suficiente⁽³³⁾.

Learning Rate (LR) o tasa de aprendizaje: es la velocidad de cambio con la cual se va aproximando al error mínimo al ir pasando las “*epochs*”⁽³³⁾. Se debe considerar que si esta es muy pequeña, el entrenamiento tomaría demasiado tiempo en llegar a su óptimo desempeño, mientras que con una tasa demasiado grande no logra reducir el error del entrenamiento de manera eficiente⁽³³⁾.

El objetivo de la fase de entrenamiento es aprender de los datos entregados para luego predecir los resultados en grupos de datos no vistos anteriormente. Dado que en la segunda fase (prueba) el error de prueba no puede estimarse porque los datos no están etiquetados o debido a que los resultados de las muestras de prueba son desconocidos ⁽³¹⁾, es que se utiliza la tercera fase (validación) para evaluar el rendimiento del modelo entrenado. En esta última etapa, los datos de validación proporcionan una evaluación no sesgada del clasificador ⁽³¹⁾.

Para la fase de validación se usan diversas estrategias para obtener métricas del desempeño del proceso de aprendizaje, siendo una de las más utilizadas las matrices de confusión. Las matrices de confusión se utilizan como un método para evaluar los errores en un modelo de clasificación. Dependiendo de la cantidad de resultados a analizar existen dos tipos de matrices: binarias, cuando solo existen dos posibles resultados, y multiclase cuando la cantidad de resultados posibles es mayor a dos ⁽³²⁾. Las matrices de confusión binarias se construyen mediante tablas de contingencia 2x2, donde, por norma general, las columnas representan los valores reales de la variable objetivo y las filas representan los valores entregados por el modelo al intentar clasificar los datos recibidos (anexo 1). Estas tablas resumen cuatro tipos de resultados, a saber:

- Verdadero positivo (TP, del inglés *True Positive*) cuando el valor real es positivo y el modelo predice un valor positivo.
- Verdadero negativo (TN, del inglés *True Negative*) cuando el valor real es negativo y el modelo predice un valor negativo.
- Falso positivo (FP, del inglés *False Positive*) cuando el valor real es negativo,

pero el modelo predice un valor positivo.

- Falso negativo (FN, del inglés *False Negative*) cuando el valor real es positivo, pero el modelo predice un valor negativo ⁽³²⁾.

De estos últimos valores se calculan métricas necesarias para conocer el desempeño del modelo utilizado:

1. Accuracy: Es la métrica más básica para medir el rendimiento de los modelos clasificadores, representa la cantidad de datos correctamente clasificados (TP+TN), dentro del total de datos obtenidos (TP+TN+FP+FN)⁽²⁸⁾.
2. Recall: También conocido como sensibilidad o tasa de acierto, representa la cantidad de datos correctamente evaluados como positivos (TP) dentro del número total real de datos positivos (TP+FN) ⁽²⁸⁾. *Recall* es entonces una medida de la efectividad en incluir elementos relevantes dentro de los datos escogidos. Es posible alcanzar un *Recall* del cien por ciento clasificando todos los datos (positivos y negativos) dentro de la categoría positiva, lo cual anularía el propósito de los sistemas de clasificación en sí mismos; aun así, la habilidad de alcanzar un *Recall* alto es claramente deseable ⁽²⁸⁾. Su fórmula se detalla en **anexo 2**.
3. Precision: También conocido como valor predictivo positivo, representa la proporción entre las predicciones positivas correctamente clasificadas (TP) y el número total de predicciones positivas realizadas (TP+FP) ⁽²⁸⁾. *Precision* es por lo tanto una medida de la pureza en la capacidad de clasificación, de la efectividad en excluir elementos que no son relevantes ⁽³²⁾. Se obtiene de la fórmula precisada en el **anexo 3**.
4. F1-Score: Corresponde a la media armónica entre *Precision* y *Recall*. Evalúa por lo tanto el nivel total de desempeño del modelo al tener en cuenta tanto la

capacidad de selección y la pureza de esta. El valor del F1-Score varía entre 0 y 1, siendo los valores más cercanos a 1 los que indican un mejor desempeño del modelo ⁽²⁸⁾. Se obtiene con la fórmula explicitada en **anexo 4**.

5. ZeroR: es el método de clasificación más simple, el cual toma como positivos a aquellos elementos a clasificar que pertenezcan a la categoría mayoritaria, ignorando cualquier otro predictor. Aunque ZeroR no tiene ningún alcance predictivo, es útil para generar un desempeño *baseline* o de línea base, que sirve como punto de comparación para el resto de las métricas de desempeño ⁽²⁸⁾. Se genera mediante una matriz de confusión, suponiendo que el modelo predijo solo resultados positivos, luego se compara el desempeño de “F1-Score” de la matriz “ZeroR” con el entregado en la realidad por el modelo.

Dada la necesidad que existe de herramientas objetivas y certeras de clasificación de la actividad de la uveítis, se propone utilizar las herramientas de adquisición y análisis de imágenes disponibles para originar procedimientos que faciliten el diagnóstico, pero sin perder de vista una alta precisión en los resultados. Por lo tanto, este trabajo tiene como propósito, por un lado, generar instancias de evaluación que independiente del examinador tengan un rendimiento óptimo en la detección de actividad inflamatoria ocular, y, por otro lado, generar un estándar de comparación para herramientas futuras de detección.

Para lograr estos objetivos, se propone el uso del entrenamiento y evaluación de un modelo de clasificación del estado inflamatorio intraocular entrenado utilizando aprendizaje por transferencia y supervisado, en base a imágenes de OCT para la evaluación de pacientes con uveítis.

Objetivos:

General:

Comparar el desempeño de cuatro modelos prediseñados de clasificación basado en DL utilizados para reconocer la presencia de inflamación intraocular en imágenes recolectadas mediante OCT.

Específicos:.

- Obtener Imágenes de OCT de pacientes con inflamación intraocular, en estado activo y/o inactivo utilizando los protocolos EDI y EVI-OCT.
- Entrenar y validar los modelos de clasificación para la detección de actividad inflamatoria intraocular.
- Evaluar y comparar el desempeño de los modelos generados

Hipótesis:

El uso de DL permite la clasificación de imágenes de OCT según el estado de actividad inflamatoria intraocular en pacientes con uveítis con parámetros de desempeño de alto estándar.

Materiales y métodos:

Diseño experimental:

Estudio de tipo retrospectivo, de alcance exploratorio.

Población:

Pacientes con diagnóstico y sospecha de uveítis activa o inactiva atendidos en los centros hospitalarios HCUCH y HFB entre los meses de septiembre de 2018 a octubre de 2022.

Lugar de realización:

- Laboratorio de Enfermedades Autoinmunes Oculares y Sistémicas (LEAOS), Departamento de Oftalmología, Facultad de Medicina, Hospital Clínico Universidad de Chile. (HCUCH)
- Hospital Metropolitano Occidente Félix Bulnes. (HFB)

Tamaño de Muestra:

2347 imágenes de OCT de 75 pacientes distintos capturadas entre los meses de septiembre de 2018 a octubre de 2022 mediante los protocolos EDI y/o EVI. Cada una de las imágenes obtenidas en distintas atenciones se utilizó como dato independiente, de igual manera se evaluó cada ojo como independiente para los fines de este estudio.

Criterios de inclusión y exclusión:

Inclusión:

- Pacientes con diagnóstico de uveítis, independiente de su etiología, que presenten ficha en los hospitales HCUCH y/o HFB
- Registro de imágenes de OCT centrada en fóvea tomada con los protocolos EVI-OCT hasta 7 días después de la atención oftalmológica.
- Registro de imágenes de OCT centrada en fóvea tomada con los protocolos EDI-OCT hasta 7 días después de la atención oftalmológica.

Exclusión:

- Pacientes sin imágenes OCT centradas en fóvea.
- Pacientes con imágenes de OCT tomadas luego de 7 días de la atención oftalmológica registrada en ficha clínica.
- Pacientes con otras patologías oftalmológicas asociadas, distintas a uveítis.

Diseño Metodológico:

Estrategia de entrenamiento de modelos de DL:

Para el presente estudio se utilizaron modelos prediseñados para la clasificación de imágenes disponibles de manera gratuita y para uso público en la plataforma Matlab. Estos modelos fueron construidos utilizando arquitecturas basadas en DL, específicamente, en las redes neuronales convolucionales. Al ser modelos previamente entrenados se utilizó *Transfer Learning* para utilizar los aprendizajes anteriores de cada modelo en cuanto a detección de características, para luego utilizar aprendizaje supervisado para su uso en imágenes de OCT.

Debido a que no se tiene la posibilidad de estimar de manera previa qué modelo de clasificación tendrá el mejor desempeño dados los datos recopilados, se evaluaron cuatro modelos clasificatorios con diversas arquitecturas internas y luego se compararon entre sí evaluando su desempeño predictivo. Para esto, se modificaron las capas encargadas de la clasificación final de cada imagen, para adaptarlas a las etiquetas del set de imágenes utilizado, es decir, para que luego del análisis clasifique cada imagen como “activa” o “inactiva”.

Los modelos elegidos fueron “GoogLeNet”, “ResNet”, “EfficientNet” y “ShuffleNet”, ya que presentan el mismo tamaño de *input*, por lo tanto, reciben la imagen con la misma cantidad de distorsión al redimensionarla. El entrenamiento de los modelos fue realizado en un computador que cuenta con las siguientes características: Procesador 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz, GPU: NVIDIA GeForce RTX 3050 Ti Laptop GPU, Entorno: Windows 11.

Recolección de datos:

Se realizó un análisis retrospectivo de fichas clínicas de sujetos diagnosticados con uveítis atendidos en el Hospital Clínico Universidad de Chile u Hospital Félix Bulnes, entre los meses de septiembre de 2018 y octubre de 2022, evaluados por un único médico oftalmólogo especializado en úvea. Los signos inflamatorios observados se cuantificaron y registraron bajo los criterios establecidos por los grupos SUN y UVEDAI, como fue descrito previamente. Las imágenes de OCT fueron analizadas por el especialista en úvea, quien clasificó las imágenes según la actividad inflamatoria detectada en “activo” o “inactivo”. Los datos clínicos obtenidos se analizaron como datos independientes para cada una de las visitas correspondientes y la clasificación

de las imágenes según actividad inflamatoria fueron registrados en una base de datos creada en una planilla de Microsoft Excel.

Adquisición, selección y preprocesamiento de imágenes:

Las imágenes de OCT de sujetos con sospecha de uveítis activa e inactiva o previamente diagnosticada fueron obtenidas mediante el uso del equipo OCT2 Spectralis de la compañía Heidelberg, haciendo uso de los protocolos EVI y EDI-OCT descritos previamente. Las imágenes se extrajeron manteniendo la integridad de la escala de píxeles, luego fueron redimensionadas al tamaño de entrada necesario para los modelos utilizados (224x224 píxeles). Cada paciente contaba con entre cuatro y seis mediciones por ojo, tomada en distintas secciones transversales, todas centradas en fovea. Se consideraron las imágenes de cada visita como un dato independiente y se seleccionaron aquellas en que la nitidez de la imagen permitiese un análisis correcto de esta y no hubiera artefactos u opacidades impidiendo la visualización de las estructuras a analizar. Se descartaron 48 imágenes por estas razones, quedando un total de 2299 imágenes a utilizar. Debido a la diferencia de dimensiones y de características a analizar, se consideraron las imágenes EDI y EVI como dos *datasets* por separado con análisis independientes.

Entrenamiento de los modelos y optimización del LR:

Se tuvieron en consideración diferentes criterios al momento de definir la configuración experimental a utilizar. En primer lugar, se definió la división de cada *dataset* en dos subconjuntos: un subconjunto de entrenamiento correspondiente al 60% de las imágenes escogidas de manera aleatoria y un subconjunto de validación

correspondiente al 40% restante de las imágenes. Cada uno de estos subconjuntos fueron cargados en lotes o *batches* al modelo. Para elegir el tamaño de cada lote se tuvo en consideración el impacto de este (poder predictivo vs demanda de tiempo y recursos computacionales), de este modo, el tamaño máximo de lote posible de procesar fue 15 para las imágenes EDI y 17 para imágenes EVI. En segundo lugar, se definió el número de épocas o *epochs* que se utilizaron como criterio de finalización del entrenamiento. Para este estudio en particular, se utilizaron 10 *epochs* con 45 iteraciones para EDI y 43 iteraciones para EVI.

Para la determinación del LR óptimo se realizaron pruebas de entrenamiento de ajuste con diferentes valores entre 0.01 y 0.0001 con los lotes y *epochs* mencionados previamente, usando cada una de las cuatro redes convolucionales propuestas. Para cada red convolucional se eligió el LR que produjo un mayor *accuracy* para la etapa de entrenamiento.

Análisis comparativo del desempeño de los modelos:

Posterior al entrenamiento, se obtiene la matriz de confusión resumen de la clasificación producida por cada modelo y se confeccionan adicionalmente las respectivas matrices de confusión "ZeroR". Se analizó el desempeño de cada clasificador utilizado mediante las métricas de Accuracy, Recall, Precision, F1-Score de la matriz de confusión respectiva al modelo. En paralelo, se compararon las métricas precision y F1-Score con las obtenidas mediante las matrices ZeroR respectivas.

Resultados:

Características demográficas y clínicas:

Se reclutó un total de 75 pacientes con uveítis, de los cuales el 88% corresponden a personas de sexo femenino (**Tabla 1**). El promedio de edad de los pacientes es de 36.2 años, con un rango entre 21 y 68 años. Al estratificar la edad de los pacientes por décadas, el grupo más representado en el grupo en estudio es de la los 30 hasta los 40 años.

Dentro de las causas de uveítis de los pacientes enrolados, la más común fue VKH (n = 65, 86.67%) seguida de las uveítis idiopáticas (n = 9, 12.00%), y se incluyó un único caso de uveítis infecciosa causada por *Bartonella henselae* (1.33%). La mayoría de los datos recolectados corresponden a pacientes que presentaban panuveítis (n =65, 86.67%) en comparación a los casos de uveítis intermedia (n = 4, 5.33%) y posterior (n = 6, 8.00%). No se encontraron pacientes con imágenes de OCT que tuvieran historial de uveítis exclusivamente a nivel de cámara anterior.

Se recolectaron un total de 2299 imágenes de OCT de polo posterior (Tabla 2), 1175 imágenes correspondientes al protocolo EDI-OCT siendo el grupo mayoritario el de clasificación “Inactivo” (n = 711, 60.51%) por parte de los clínicos tratantes, en contraste a la clasificación “Activo” (n = 464, 39.49%). Mientras que las imágenes correspondientes al protocolo EVI-OCT fueron 1124 en total, siendo de igual manera mayoritariamente “Inactivos” (n = 655, 58.27%), en comparación con los “Activos” (n = 469, 41.73%).

Tabla 1. Características demográficas y clínicas de los pacientes en estudio.

Sexo	n (%)
Femenino	66 (88%)
Masculino	9 (12%)
Rango etario	
[20 - 30[19 (25.33%)
[30 - 40[23 (30.67%)
[40 - 50[19 (25.33%)
[50 - 60[10 (13.33%)
[60 - 70[4 (5.33%)
Etiología	
VKH	65 (86.67%)
Idiopática	9 (12.00%)
<i>B. henselae</i>	1 (1.33%)
Localización anatómica	
Intermedia	4 (5.33%)
Posterior	6 (8.00%)
Panuveítis	65 (86.67%)

Tabla 2. Clasificación de imágenes adquiridas según actividad inflamatoria, protocolos EDI y EVI.

	EDI-OCT	EVI-OCT
Activos	464 (39.49%)	469 (41.73%)
Inactivos	711 (60.51%)	655 (58.27%)

Optimización del LR para cada red convolucional según accuracy

Cada proceso realizado para optimizar los modelos entregó gráficos de cambio de la *accuracy* del entrenamiento y la validación (**Figura 1**, anexo 5-12), los cuales fueron resumidos en los gráficos posteriores.

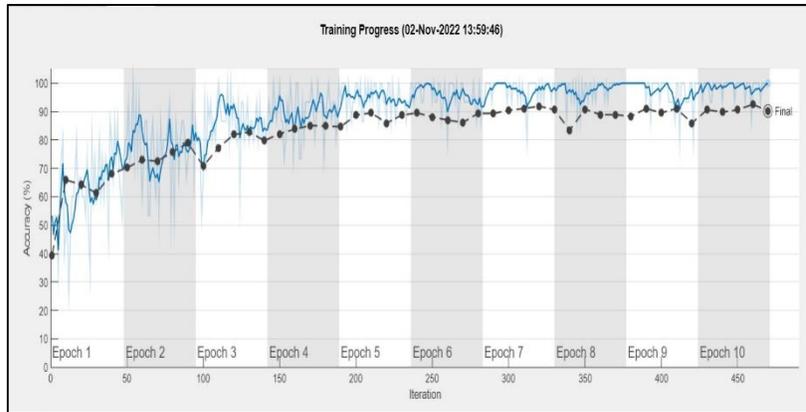


Figura 1: Evolución de *accuracy* en función del número de iteraciones. Se presenta el *accuracy* producido por GoogLeNet en cada iteración estructuradas en 10 *epochs* para las imágenes de EDI-OCT para el modelo entrenado. La línea azul representa el resultado obtenido durante la fase de entrenamiento, y en negro los resultados para el lote de validación.

Se observo que los modelos GoogLeNet, ResNet, y ShuffleNet tienen el peak de *accuracy* a una tasa de aprendizaje de 0.001, independiente del protocolo de imagen mientras que el modelo EfficientNet alcanza su punto más alto con un LR de 0.01 (*accuracy* 96.22%) Para fines comparativos se debe mantener el aprendizaje igual para todos los modelos, por lo tanto, se eligió la tasa que optimice a la mayoría (LR = 0.001) (**Figuras 2 y 3**).

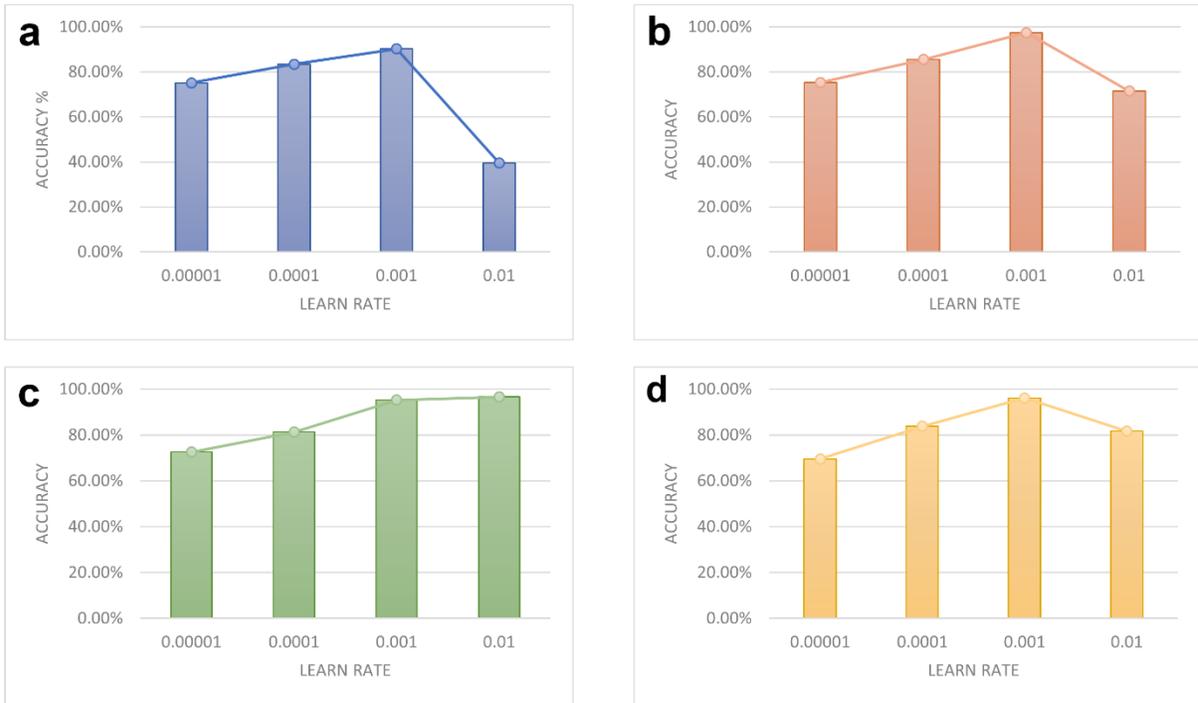


Figura 2. *Accuracy* en función del LR para el análisis de imágenes EDI-OCT. Se evaluó el *accuracy* para diferentes valores de LR (0.00001, 0.0001, 0.001 y 0.01) empleando las redes convolucionales GoogLeNet (a), ResNet (b), EfficientNet (c) y ShuffleNet (d). Se obtuvo que los modelos produjeron un mejor desempeño en términos de *accuracy* con un LR de 0.001, excepto para EfficientNet, que mostró una pequeña mejora a un valor de 0.01.

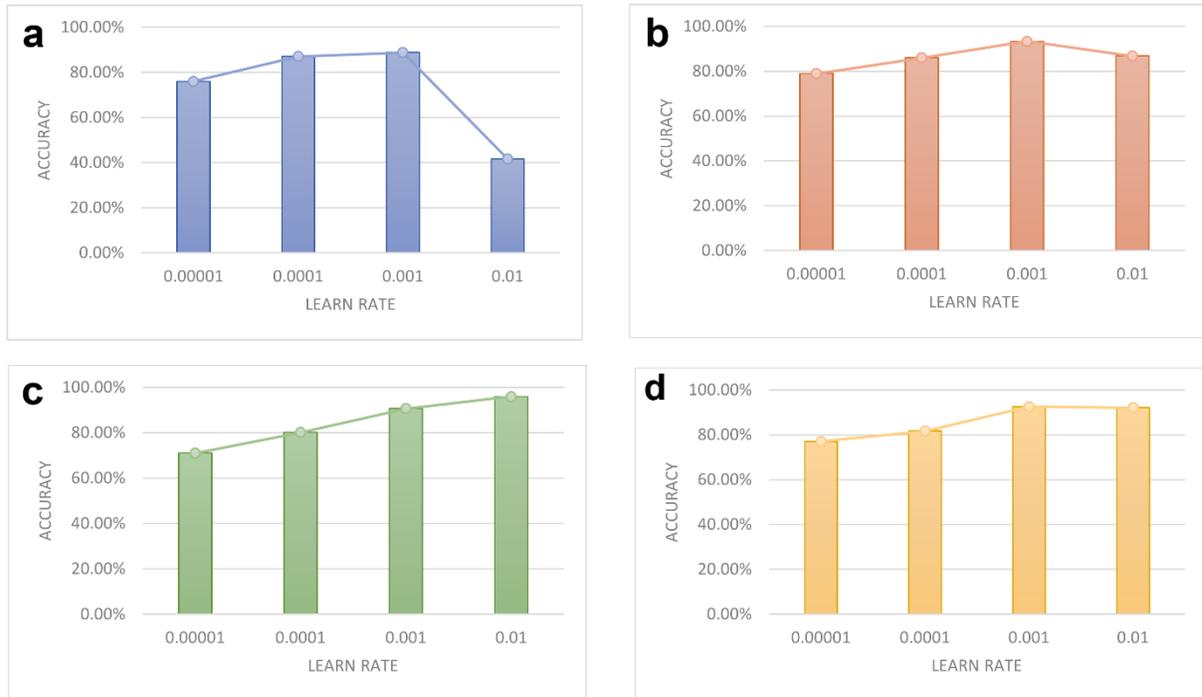


Figura 3. *Accuracy* en función del LR para el análisis de imágenes EVI-OCT. Se evaluó el *accuracy* para diferentes valores de LR (0.00001, 0.0001, 0.001 y 0.01) empleando las redes convolucionales GoogLeNet (a), ResNet (b), EfficientNet (c) y ShuffleNet (d). Se obtuvo que los modelos produjeron un mejor desempeño en términos de *accuracy* con un LR de 0.001, excepto para EfficientNet, que mostró una mejora a un valor de 0.01.

Entrenamiento y validación final de los modelos

Una vez completada la optimización, se procedió a la evaluación comparativa de todas las redes, obteniendo como resultado el *accuracy* de validación para cada clasificador (**Tabla 2**), siendo ResNet el modelo con mejor desempeño para ambos protocolos observados (97.234% EDI, 96.962% EVI). EfficientNet (95.106% EDI, 90.585% EVI) y

ShuffleNet (92.128% EDI, 91.094% EVI) obtuvieron rendimientos por sobre el 90% de efectividad, por otra parte, GoogLeNet (85.319% EDI, 85.496% EVI) tuvo el menor rendimiento comparativo de todos las redes puestas a prueba, para ambos protocolos.

Tabla 2. *Accuracy* de los modelos obtenidos durante la fase de entrenamiento para la clasificación de imágenes de EDI- y EVI-OCT.

Modelo	<i>Accuracy (%)</i>	
	EDI-OCT (%)	EVI-OCT (%)
GoogLeNet	85.3191	85.4962
ResNet	97.2340	96.9621
EfficientNet	95.1064	90.5852
ShuffleNet	92.1277	91.0941

Análisis comparativo del desempeño de los modelos generados

De cada modelo entrenado se obtuvo una matriz de confusión binaria (**Figuras 4 y 5**), de las cuales se calcularon las métricas de desempeño para su comparación. En primera instancia se compararon las métricas *precision* y *F1-Score* obtenidas con las correspondientes matrices, con aquellas desprendidas de las matrices ZeroR (**Tabla 3**). Todos los modelos superan el desempeño de línea base definido por “ZeroR”, indicando que todos superan un estándar de rendimiento mínimo.

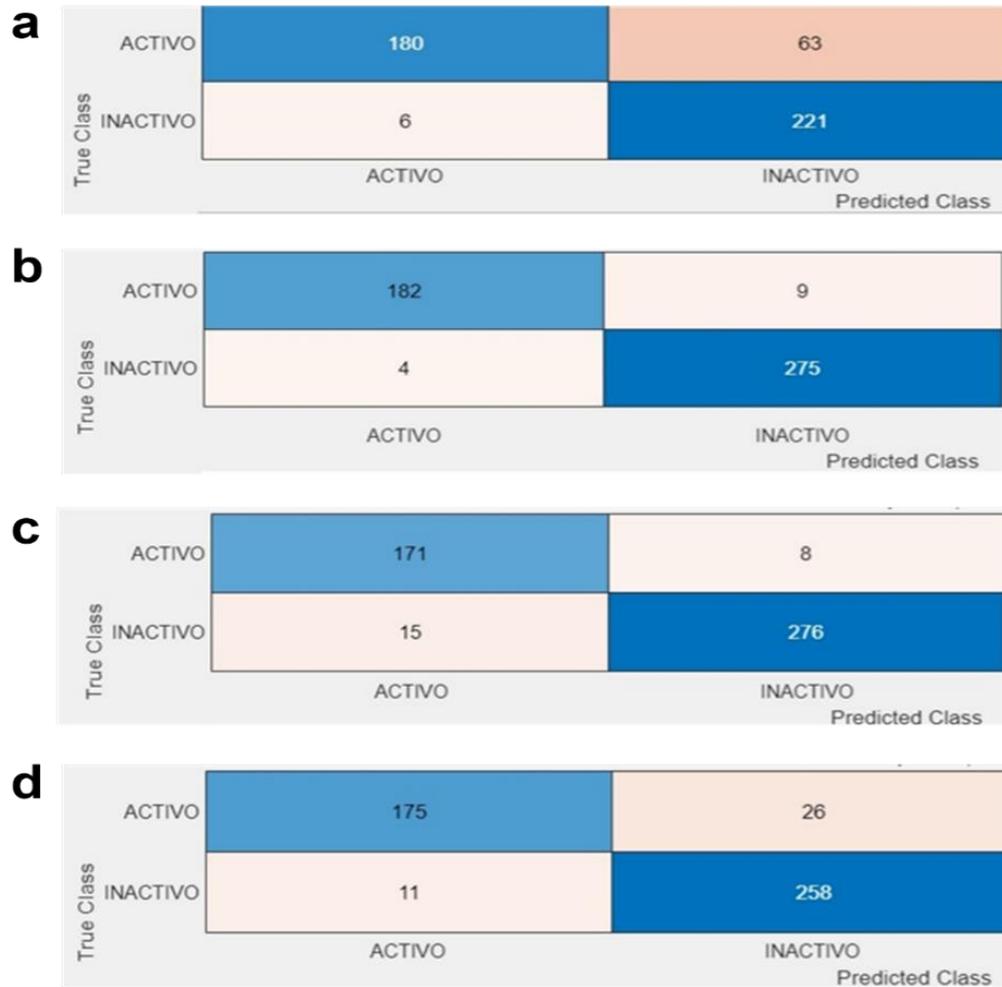


Figura 4. Matrices de confusión resultantes del proceso de clasificación de imágenes de EDI-OCT del lote de validación. Los resultados de la clasificación se resumen en estas matrices de confusión: **a**, GoogLeNet; **b**, ResNet; **c**, EfficientNet; **d**, ShuffleNet.

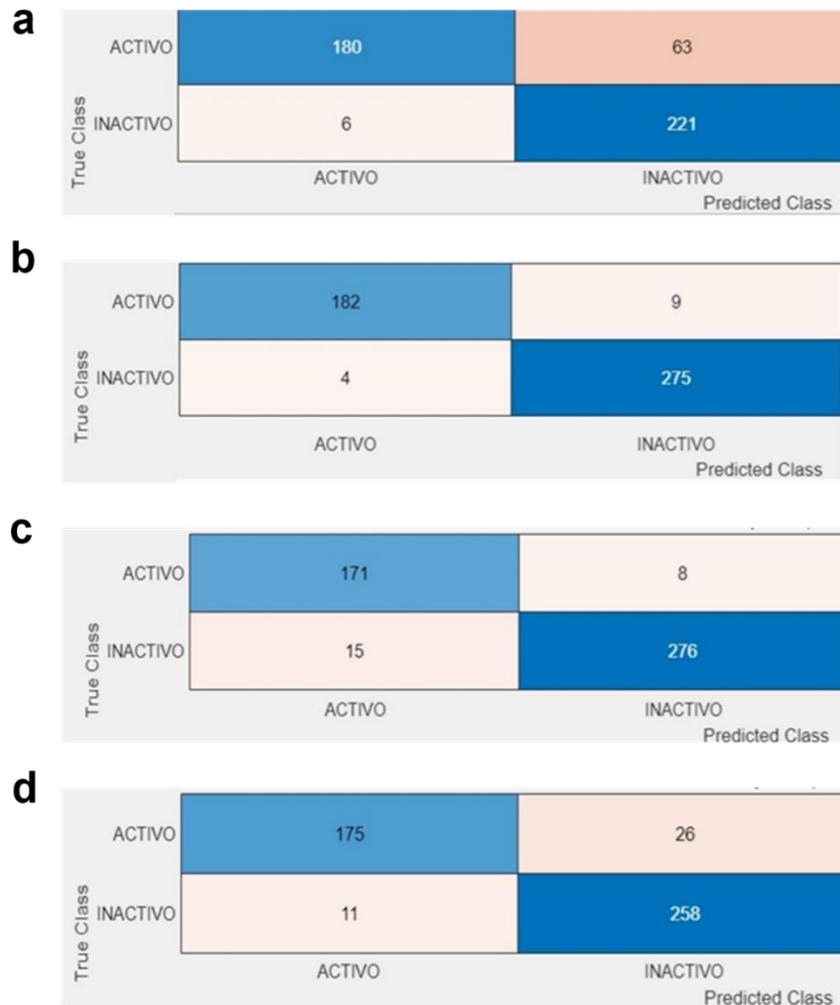


Figura 5. Matrices de confusión resultantes del proceso de clasificación de imágenes de EVI-OCT del lote de validación. Los resultados de la clasificación se resumen en estas matrices de confusión: a, GoogLeNet; b, ResNet; c, EfficientNet; d, ShuffleNet.

Tabla 3. Valores de Precision y F1-Score según matriz de confusión y matriz ZeroR para cada modelo clasificador obtenidos para los protocolos EDI- y EVI-OCT.

	Modelo	Matriz confusión validación		Matriz ZeroR	
		Precision	F1-Score	Precision	F1-Score
EDI-OCT	GoogLeNet	96.77%	83.92%	51.70%	68.16%
	ResNet	97.85%	96.55%	40.64%	57.79%
	EfficientNet	91.94%	93.70%	38.09%	55.16%
	ShuffleNet	94.09%	90.44%	42.70%	59.91%
EVI-OCT	GoogLeNet	95.12%	84.55%	52.16%	68.56%
	ResNet	96.95%	96.07%	42.49%	59.64%
	EfficientNet	84.76%	88.25%	38.42%	55.51%
	ShuffleNet	94.51%	89.86%	46.19%	63.19%

Posteriormente, se compararon las métricas de desempeño respectivas entre sí (**Tabla 4**), resultando el modelo ResNet como el clasificador con mejores resultados para ambos protocolos, teniendo el mayor *Accuracy* (97.23% EDI, 96.69% EVI), *Recall* (95.29% EDI, 95.21% EVI), *Precision* (97.85% EDI, 96.95% EVI) y *F1-Score* (96.55% EDI, 96.07% EVI). Cabe destacar también que fue el único modelo en lograr mantener el desempeño total (*F1-score*), por sobre el 95% de rendimiento para ambos protocolos, mientras que GoogLeNet (83.92% EDI, 84.55% EVI), EfficientNet (93.70% EDI, 88.25% EVI) y ShuffleNet (90.44% EDI, 89.86% EVI) se mantuvieron por debajo de este estándar.

Tabla 4. Métricas de desempeño obtenidos en la validación de los modelos clasificadores para los protocolos EDI- y EVI-OCT.

	Modelo	Accuracy	Recall	Precision	F1-Score
EDI-OCT	GoogLeNet	85.32%	74.07%	96.77%	83.92%
	ResNet	97.23%	95.29%	97.85%	96.55%
	EfficientNet	95.11%	95.93%	91.94%	93.70%
	ShuffleNet	92.13%	87.06%	94.09%	90.44%
EVI-OCT	GoogLeNet	85.50%	76.10%	95.12%	84.55%
	ResNet	96.69%	95.21%	96.95%	96.07%
	EfficientNet	90.59%	92.05%	84.76%	88.25%
	ShuffleNet	91.09%	85.64%	94.51%	89.86%

Discusión:

En esta investigación se logró comprobar que efectivamente es posible alcanzar un alto desempeño predictivo en la detección de actividad inflamatoria intraocular utilizando modelos basados entrenados mediante DL y redes neuronales convolucionales.

Debido a la gran variabilidad de las manifestaciones clínicas entre pacientes con uveítis, no se puso restricción dentro de los criterios de inclusión o exclusión en cuanto a la localización anatómica de la uveítis. Esto significa que imágenes de pacientes con uveítis anterior, sin hallazgos clínicos en polo posterior, pudieran ser incorporadas en el entrenamiento de los modelos de clasificación automatizada. No obstante, estudios anteriores determinaron el engrosamiento del grosor coroideo subfoveal⁽³⁴⁾, de la mácula y de la capa de fibras nerviosas de la retina⁽³⁵⁾ detectables mediante OCT en pacientes con uveítis de polo anterior. Esto sugiere que el entrenamiento con imágenes del polo posterior de pacientes con uveítis anterior también podría ser de utilidad, sin embargo, no fue posible obtener imágenes de pacientes con diagnóstico de uveítis restringida al polo anterior. Esto debe ser tomado en cuenta, ya que implica que los modelos predictivos generados en este trabajo están entrenados y validados solo para imágenes de uveítis intermedia, posterior y panuveítis, por lo que su uso para análisis de imágenes de pacientes con uveítis anterior podría producir errores importantes.

En base a los datos demográficos se pudo observar un predominio, en el rango etario de 30 a 39 años, seguido de cerca por el rango de 20-29 y 40-49, lo que concuerda con observaciones previas de la predominancia de la uveítis entre adultos en edad

laboral. También se vio una predominancia del género femenino sobre el masculino y de la etiología VKH por sobre el resto, lo que concuerda con lo descrito previamente: el VKH, una enfermedad que afecta principalmente a mujeres es la enfermedad más frecuente entre los casos de uveítis de causa conocida en Chile⁽⁷⁾.

En cuanto a la optimización del proceso de aprendizaje, se puede notar que el LR óptimo (el que logra un mayor *accuracy*) para la mayoría de estos modelos se encuentra cercana a 0.001, a excepción del clasificador Efficient-Net, el cual encuentra su punto óptimo más cercano a 0.01. Esto concuerda con lo descrito previamente en la literatura: no existe en la actualidad una forma certera de predecir el LR óptimo y lo adecuado es determinarlo de forma empírica⁽³⁶⁾. Es de importancia identificar de antemano qué criterios se aplicarán para la elección del LR. Para este trabajo, el criterio de elección fue el *accuracy*, una métrica del grado del desempeño del modelo en todas las clases, es decir, qué tan bueno es identificando correctamente tanto a pacientes activos como inactivos.

Para realizar la comparación entre los distintos modelos generados, la plataforma utilizada exige mantener un LR igual para todas las redes neuronales, por lo tanto, se mantuvo una variación que optimizara a la mayoría de ellos. A pesar de aquello, se lograron valores de *Accuracy* por sobre el 90% para 3 modelos (ResNet, EfficientNet y ShuffleNet) tanto para las imágenes de EDI-OCT como EVI-OCT. De todos modos, queda la incertidumbre de si se pudiera producir un clasificador con mejores rendimientos finales con otros valores de LR.

Con respecto a las métricas ZeroR, si bien no tienen ningún alcance predictivo, si establecen una línea de base a superar para determinar su utilidad para la tarea designada. Se observó que todas las redes superan esta barrera en cuanto a los

criterios Precision y F1-Score, por lo que se concluye que todas tienen al menos un grado mínimo de utilidad predictiva.

Dado que una superación de las métricas de desempeño a las del criterio ZeroR no aseguran un valor de clasificación clínica suficiente, se procedió a determinar y comparar las distintas métricas planteadas. Se observó que solo con ResNet se produjeron modelos con un *accuracy* mayor al 95% para los protocolos EDI- y EVI-OCT, lo que indica un muy buen desempeño. Sin embargo, se debe reconocer que, si bien el *accuracy* es útil para una comparación general del desempeño entre modelos, esta métrica carece de profundidad, ya que solo considera predicciones correctas sobre el total, independientemente de la categoría de clasificación. El análisis basado solo en este parámetro podría sobreestimar la utilidad de los clasificadores, por lo que se realizó una comparativa de los valores Recall, Precision y F1 – Score para lograr una evaluación más acabada.

Para poder ser aplicable en la práctica clínica, se deben establecer estándares mínimos a los parámetros de desempeño obtenido con una herramienta clínica de clasificación o *screening* para asegurar su utilidad y fiabilidad. Dado que “*Recall*” está enfocado en minimizar la chance de pasar por alto los verdaderos positivos, es necesario que este parámetro se mantenga mayor o igual al 95% para reducir al mínimo posible esta chance. La importancia de esta métrica resulta de sus implicancias clínicas: un bajo *Recall* implica que pacientes con uveítis activa sean clasificados como inactivos, interfiriendo con su evaluación clínica y su correcto tratamiento. Sumado a esto, considerando que el foco de la métrica “*precision*” se encuentra en evitar los falsos positivos, es importante mantenerla lo más alta posible para tener un clasificador confiable. No obstante, el impacto de este criterio no es tan

significativo como el del *Recall*, ya que el riesgo que implica clasificar un paciente con uveítis inactiva como activa no es tal como el del caso opuesto. Idealmente, se espera un modelo que logre un buen desempeño en ambos aspectos, lo cual puede evaluarse a través del *accuracy* y F1–Score, los cuales estiman el desempeño en la correcta clasificación tanto de pacientes activos como inactivos.

Establecido esto, cabe destacar que el modelo ResNet es el único de los cuatro clasificadores estudiados que logra un desempeño mayor o igual al 95% en cada uno de estos estándares, corroborando su potencial utilidad para la práctica clínica. Si bien estos son resultados prometedores, se deben considerar las limitaciones del presente estudio y, por tanto, a la posible utilidad clínica de esta nueva herramienta. Este estudio constituye una prueba de concepto de que las aplicaciones de ML permiten establecer un método objetivo y efectivo para la clasificación de la actividad inflamatoria intraocular a partir del análisis de imágenes de OCT. Este estudio utilizó imágenes que en ocasiones se repitieron (imágenes tomadas durante distintas visitas al centro de atención), independiente de su estado inflamatorio, pudiendo haber una sobrerrepresentación de algunos datos de acuerdo a la evolución del paciente. Por ejemplo, un paciente que no respondiera bien a la terapia requeriría de controles médicos frecuentemente, aumentando las posibilidades de adquirir imágenes OCT del mismo ojo en estado activo. En contraste, pacientes con uveítis de larga data con un manejo adecuado de la enfermedad y que acudan a los servicios de oftalmología participantes por motivos de control periódico, producen un aumento de la probabilidad de adquirir imágenes del mismo ojo sin inflamación activa. De este modo, es posible que se haya producido un sobreajuste de los modelos a raíz del método de selección de imágenes. Por otro lado, la calidad de la información contenida en las

imágenes OCT dependen de los protocolos de captura de imágenes, las cualidades del equipo OCT y de la pericia del Tecnólogo Médico que ejecuta la examinación. Por lo anterior, una validación de este método para su uso y distribución en distintos centros de salud requerirá de un estudio multicéntrico, con la participación de distintos Tecnólogos Médicos y maquinaria OCT a modo de asegurar que los resultados son consistentes a pesar de estas variables.

Conclusiones:

La presente investigación tenía como propósito principal propósito, entrenar, validar y comparar modelos de clasificación con el fin de generar instancias dentro de la actividad clínica que mantengan un alto desempeño en la detección de actividad inflamatoria ocular sin importar el examinador, y también proporcionar un estándar de comparación para herramientas futuras.

Se logró generar un modelo mediante el uso de la red convolucional ResNet capaz de mantener un desempeño adecuado para la tarea de clasificación de imágenes OCT según el estado de inflamación intraocular, confirmando la hipótesis. La importancia de definir adecuadamente los parámetros de entrenamiento de un método de DL fue evidenciada al variar los valores de LR, permitiendo lograr un mayor desempeño independiente de la red convolucional empleada. Esto permitió obtener valores satisfactorios para cada una de las métricas de desempeño evaluadas.

Este trabajo constituye una parte de esencial en el desarrollo de la tecnología de la clasificación automática de imágenes de OCT de pacientes con uveítis. Su utilidad clínica deberá ser confirmado mediante estudios multicéntricos para su validación final y la masificación de su uso.

Bibliografía:

1. Zeghidi H, LeHoang P, Bodaghi B. (2020) Uveítis. EMC - Tratado de Medicina. Vol 24, edición 1, Pag 1-8
2. Díaz-Valle D., Méndez R., Arriola P., Cuiña R., Ariño M. (2008) Enfermedades sistémicas no infecciosas y uveítis. *Anales Sis San Navarra*; Vol 3 (Suppl 3), pág 97-110.
3. Wakefield D, Chang J. (2005) Epidemiology of Uveitis. *International Ophthalmology Clinics*. Vol 45, edición 2, pág 1-13.
4. Tsirouki, T., Dastiridou, A., Symeonidis, C., Tounakaki, O., Brazitikou, I., Kalogeropoulos, C., & Androudi, S. (2016). A Focus on the Epidemiology of Uveitis. *Ocular Immunology and Inflammation*. Vol 26, edición 1, pág 2-16
5. Barisani-Asenbauer, T., Maca, S. M., Mejdoubi, L., Emminger, W., Machold, K., & Auer, H. (2012). Uveitis- a rare disease often associated with systemic diseases and infections- a systematic review of 2619 patients. *Orphanet Journal of Rare Diseases*. Vol 7, edición 1, pág 57
6. Vega, F., Elizalde, N., Valenzuela, R., Urzua, C., & Cuitiño, L. (2022). miRNAs: una nueva mirada de la enfermedad de Vogt-Koyanagi-Harada. *Revista Hospital Clínico Universidad De Chile*. Vol. 3, edicion 2, pág 108-119.
7. Liberman, P., Gauro, F., Berger, O., & Urzua, C. A. (2014). Causes of Uveitis in a Tertiary Center in Chile: A Cross-sectional Retrospective Review. *Ocular Immunology and Inflammation*, 23(4)Standardization of Uveitis Nomenclature for Reporting Clinical Data. Results of the First International Workshop. (2005). *American Journal of Ophthalmology*. Vol 23, edicion 4, pág 339-345
8. Development of Classification Criteria for the Uveitides. (2021). *American*

Journal of Ophthalmology. Vol 228, pág 96-105

9. The Standardization of Uveitis Nomenclature (SUN) Working Group (2021), Classification Criteria for Juvenile Idiopathic Arthritis–Associated Chronic Anterior Uveitis. *American Journal of Ophthalmology*. Vol 52, edicion 3, pág 259-265
10. The Standardization of Uveitis Nomenclature (SUN) Working Group (2021), Classification Criteria for Tubulointerstitial Nephritis With Uveitis Syndrome. *American Journal of Ophthalmology*. Vol 52, edicion 3, pág 259-265
11. Pato, E., Martin-Martinez, M. A., Castelló, A., Méndez-Fernandez, R., Muñoz-Fernández, S., Cordero-Coma, M., Diaz-Valle, D. (2016). Development of an activity disease score in patients with uveitis (UVEDAI). *Rheumatology International*. Vol 37, edicion 4, pág 647-656
12. abs D, Nussenblatt R, Rosenbaum J et al (2005) Standardization of uveitis nomenclature for reporting clinical data. Results of the first international workshop. *Am J Ophthalmol*. Vol 140, edicion 3, pág 509-516.
13. Kempen, J. H., Ganesh, S. K., Sangwan, V. S., & Rathinam, S. R. (2008). Interobserver Agreement in Grading Activity and Site of Inflammation in Eyes of Patients with Uveitis. *American Journal of Ophthalmology*. Vol 146, edicion 6, pág 813-818.
14. Invernizzi A, Cozzi M, Staurenghi G. (2019) Optical coherence tomography and optical coherence tomography angiography in uveitis: A review. *Clinical & Experimental Ophthalmology*. Vol 47, edición 3, pág 357-371.
15. Onal S, Tugal-Tutkun I, Neri P, P Herbort C. (2013) Optical coherence

- tomography imaging in uveitis. *International Ophthalmology*. Vol 34, edición 2, pág 401–435.
16. Schmitt, J. M. (1999). Optical coherence tomography (OCT): a review. *IEEE Journal of Selected Topics in Quantum Electronics*. Vol 5, edición 4, pág 1205 - 1215
 17. Laviers, H., & Zambarakji, H. (2014). Enhanced depth imaging-OCT of the choroid: a review of the current literature. *Graefe's Archive for Clinical and Experimental Ophthalmology*. Vol 252, edición 12, pág 1871-1883.
 18. Invernizzi A, Marchi S, Aldigeri R, Mastrofilippo V, Viscogliosi F, Soldani A. (2017) Objective Quantification of Anterior Chamber Inflammation. *Ophthalmology*. Vol 124, edición 11, pág 1670-1677
 19. 10. Keane P, Karampelas M, Sim D, Sadda S, Tufail A, Sen H (2014). Objective Measurement of Vitreous Inflammation Using Optical Coherence Tomography. *Ophthalmology*. Vol 121, edición 9, Pág 1706-1714.
 20. Gabriel M, Kruger R, Shams-Mafi F, Hermann B, Zabihian B, Schmetterer L. (2017) Mapping Retinal and Choroidal Thickness in Unilateral Nongranulomatous Acute Anterior Uveitis Using Three-Dimensional 1060-nm Optical Coherence Tomography. *Investigative Ophthalmology & Visual Science*. Vol 58, edición 11, pág 4778-4783.
 21. Moore D, Jaffe G, Asrani S. (2015) Retinal Nerve Fiber Layer Thickness Measurements: Uveitis, A Major Confounding Factor. *Ophthalmology*. Vol 122, edición 3, pág 511-517
 22. Avila-Tomás J, Mayer-Pujadas M, Quesada-Varela V (2020). La inteligencia artificial y sus aplicaciones en medicina I: introducción antecedentes a la IA

- y robótica. Atención Primaria. Vol 52, edición 10, Páginas 778-784.
23. Yao X, Liu Y. (2013) Machine Learning. Search Methodologies. pag 477-517.
 24. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, Vol 521, edicion 7553, pág 436-44.
 25. Tulbure, A.-A., Tulbure, A.-A., & Dulf, E.-H. (2021). A review on modern defect detection models using DCNNs – Deep convolutional neural networks. Journal of Advanced Research. Vol 35, Pag 33-48
 26. Silva F, Vidotti V, Cremasco F, Dias M, Gomi E, Costa V. (2013) Sensitivity and specificity of machine learning classifiers for glaucoma diagnosis using Spectral Domain OCT and standard automated perimetry. Arquivos Brasileiros de Oftalmologia. Vol 76, edicion 3, pág 170-4.
 27. Valdés-Mas M, Martín-Guerrero J, Rupérez M, Pastor F, Dualde C, Monserrat C (2014). A new approach based on Machine Learning for predicting corneal curvature (K1) and astigmatism in patients with keratoconus after intracorneal ring implantation. Computer Methods and Programs in Biomedicine. Vol 116, edición 1, pág 39-47
 28. Gulshan V, Peng L, Coram M, Stumpe M, Wu D, Narayanaswamy A (2016). Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. Vol 316, edición 22 pág 2402-2410.
 29. Tharwat A. (2020) Classification assessment methods. Applied Computing and Informatics. Vol 17, edición 1, pag 168-192.
 30. Zhang X. Machine Learning. (2020) A Matrix Algebra Approach to Artificial

Intelligence. Pag 3-54.

31. Zhang Y, Zhang H, Cai J, Yang B. (2014) A Weighted Voting Classifier Based on Differential Evolution. *Abstract and Applied Analysis*. Vol 2014
32. Kandel, I., & Castelli, M. (2020). The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. *ICT Express*. Vol 6, edición 4, pág 312-315.
33. Masters D, Luschi C, (2018) Revisiting Small Batch Training for Deep Neural Networks, Cornell University arXiv.
34. Kim, M., Choi, S. Y., & Park, Y.-H. (2017). Analysis of choroidal and central foveal thicknesses in acute anterior uveitis by enhanced-depth imaging optical coherence tomography. *BMC Ophthalmology*. Vol17, pag 225-232.
35. Shulman S, Goldenberg D, Habet-Wilner Z, Goldstein M, Neudorfer M. (2012) Optical coherence tomography characteristics of eyes with acute anterior uveitis. *ISR Med Assoc*. Vol 14, edicion 9, pag 543-546.
36. Smith L. (2017), Cyclical Learning Rates for Training Neural Networks, Cornell University arXiv.

Anexos:

1.

		True/Actual Class	
		Positive (P)	Negative (N)
Predicted Class	True (T)	True Positive (TP)	False Positive (FP)
	False (F)	False Negative (FN)	True Negative (TN)

Matriz de confusión binaria, Tharwat A. Classification assessment methods. Applied Computing and Informatics. 2020;17(1):168-192.

2.

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

Ecuación de Recall Tharwat A. Classification assessment methods. Applied Computing and Informatics. 2020;17(1):168-192.

3.

$$PPV = \text{Precision} = \frac{TP}{FP + TP}$$

Ecuación de Precision Tharwat A. Classification assessment methods. Applied Computing and Informatics. 2020;17(1):168-192.

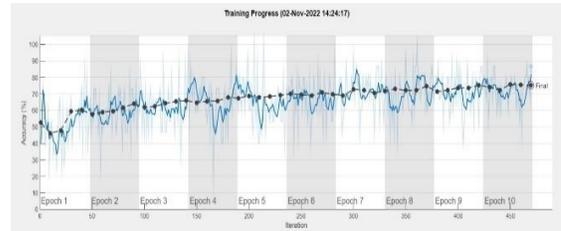
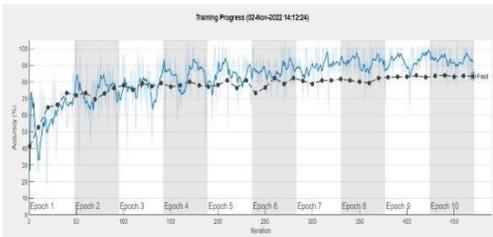
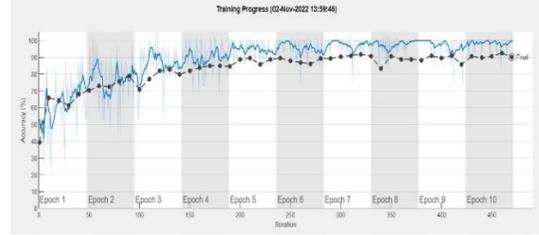
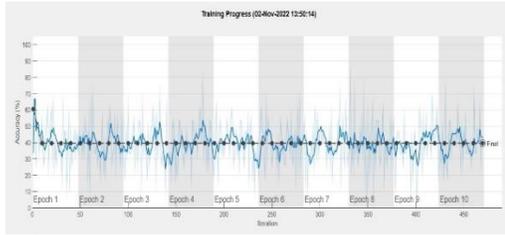
4.

$$F\text{-measure} = \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$$

Ecuación F1-Score Tharwat A. Classification assessment methods. Applied

computing and Informatics. 2020;17(1):168-192.

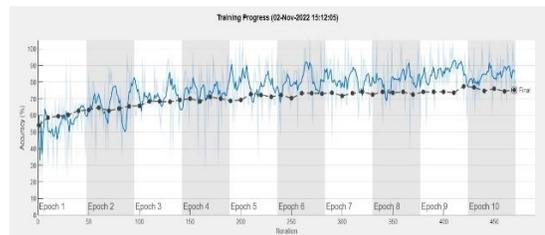
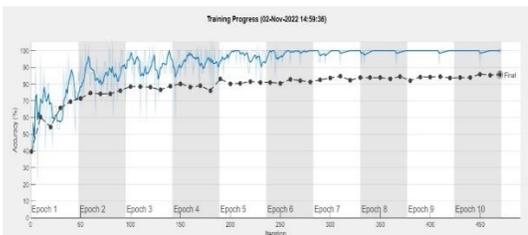
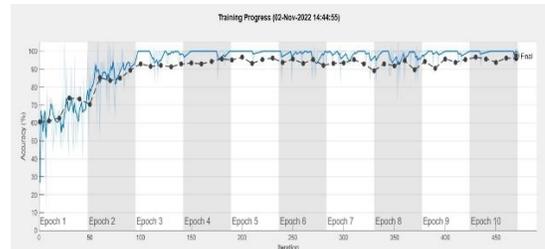
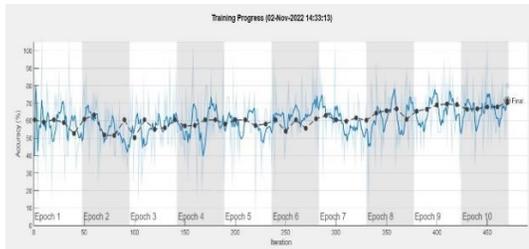
5.



Gráficos, “Entrenamiento-validación” Protocolo EDI, GoogLe-Net, Learn Rates:

1e-2; 1e-3; 1e-4; 1e-5.

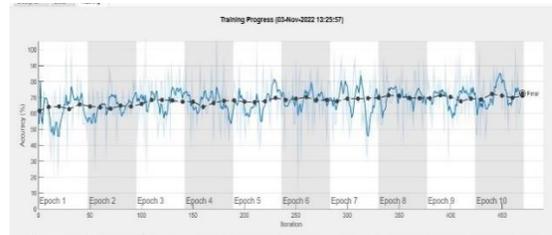
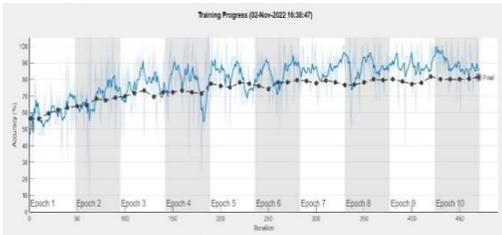
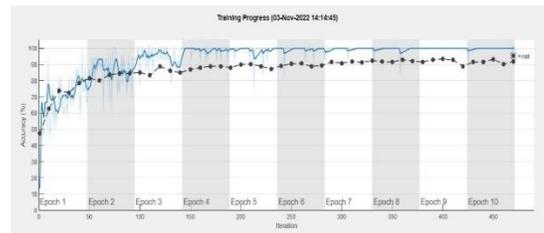
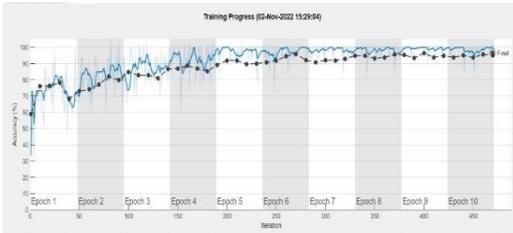
6.



Gráficos, “Entrenamiento-validación” Protocolo EDI, Res-Net, Learn Rates: 1e-2;

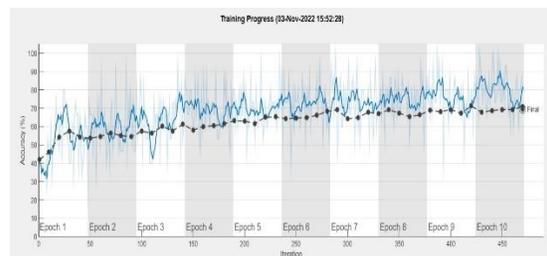
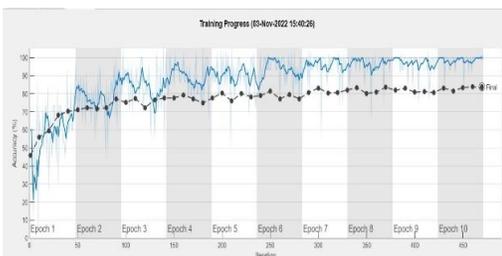
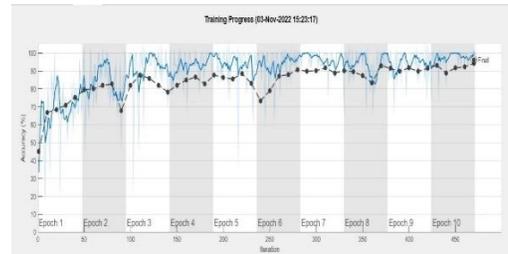
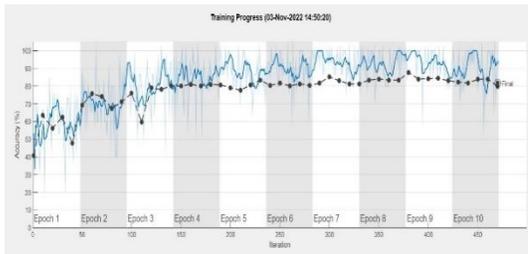
1e-3; 1e-4; 1e-5.

7.



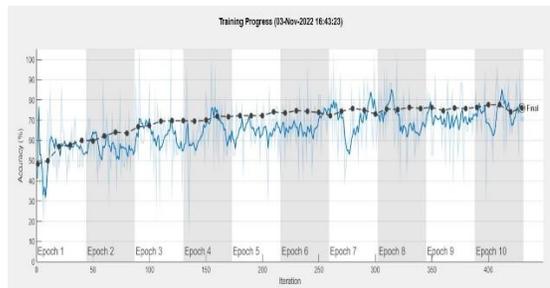
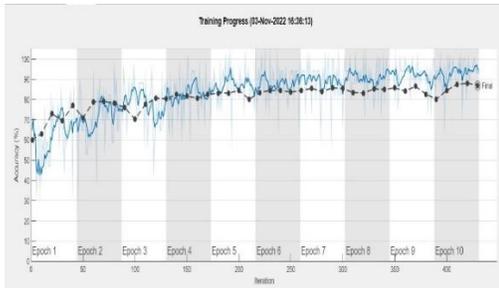
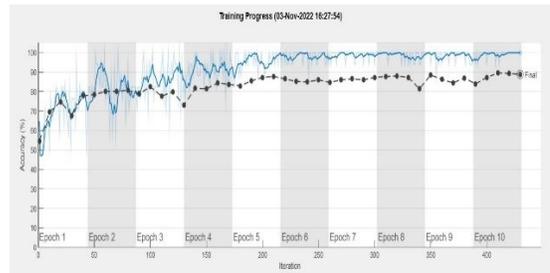
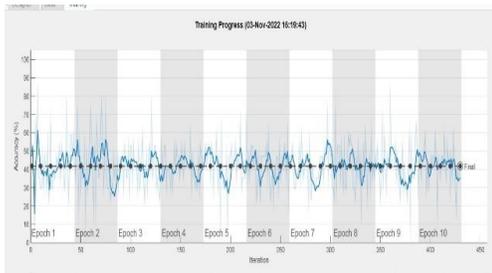
Gráficos, “Entrenamiento-validación” Protocolo EDI, Efficient-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

8.



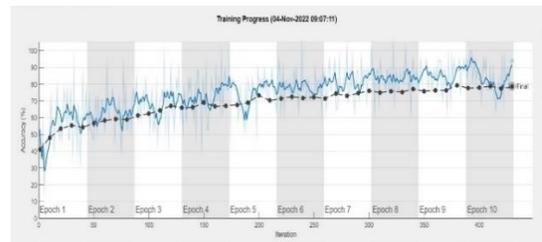
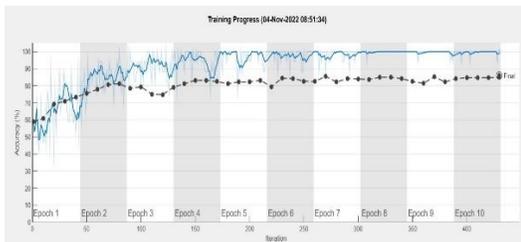
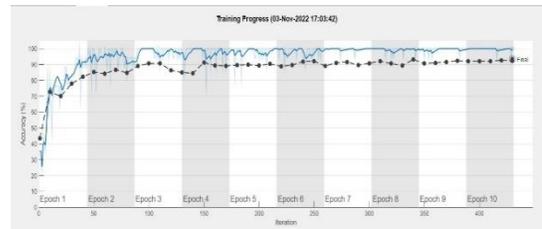
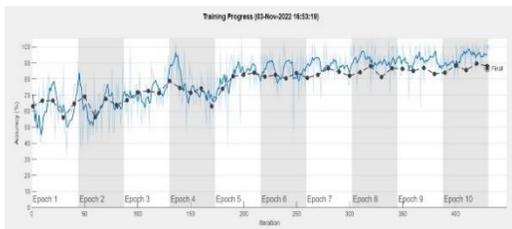
Gráficos, “Entrenamiento-validación” Protocolo EDI, Shuffle-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

9.



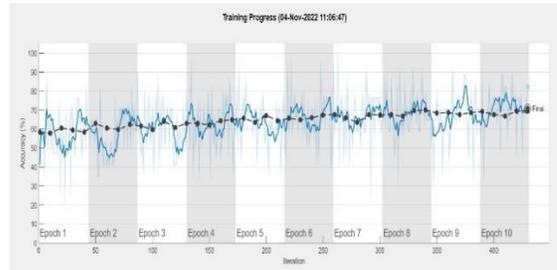
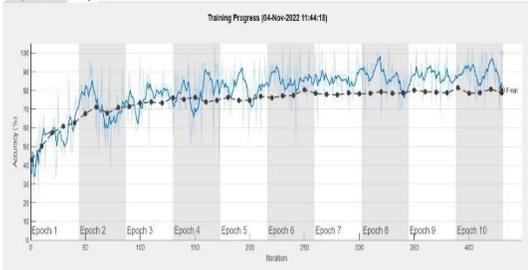
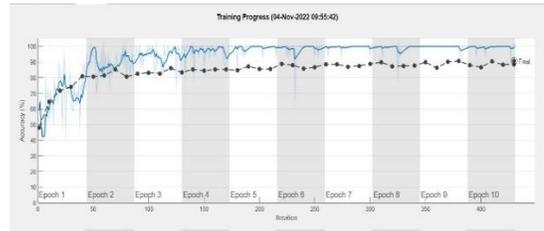
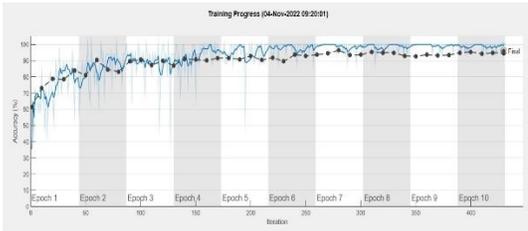
Gráficos, “Entrenamiento-validación” Protocolo EVI, Google-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

10.



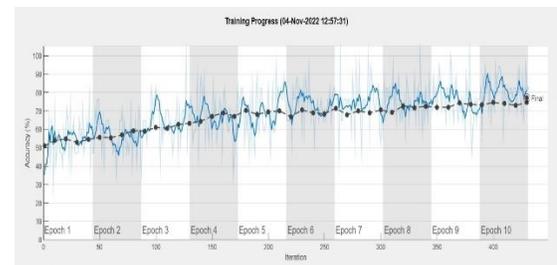
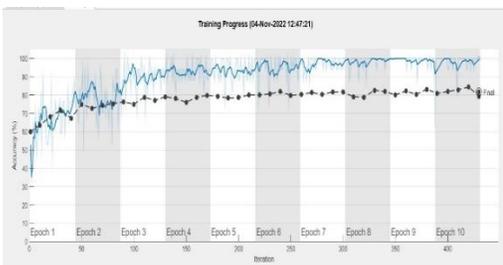
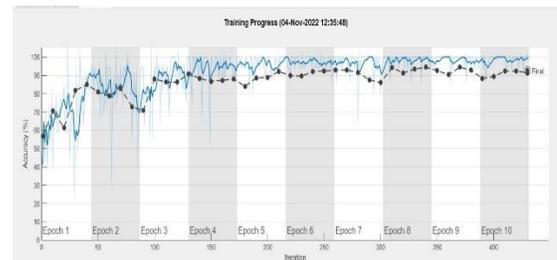
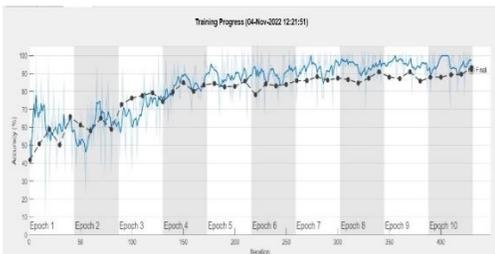
Gráficos, “Entrenamiento-validación” Protocolo EVI, Res-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

11.



Gráficos, “Entrenamiento-validación” Protocolo EVI, Efficient-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

12.



Gráficos, “Entrenamiento-validación” Protocolo EVI, Shuffle-Net, Learn Rates: 1e-2; 1e-3; 1e-4; 1e-5.

Santiago, 06 de abril de 2022

Señora
Dra. Loreto Cuitiño Tride
Investigadora responsable
Hospital Clínico de la Universidad de Chile
Presente

Ref.: "Análisis computacional para el diagnóstico temprano de la Degeneración Macular Relacionada a la Edad".

Estimada Dra. Cuitiño:

A través de la presente, acuso recibo de su carta con fecha 28 de febrero de 2022, en la que solicita aprobación de cambios al estudio detallado en la referencia, que consisten en modificar del título del proyecto con el objetivo de ampliar el espectro de enfermedades a estudiar. Por lo cual pasará a llamarse: "Análisis computacional para el diagnóstico temprano de patologías de retina externa y uveítis".

Se tuvo a la vista:

- Documento de Consentimiento Informado Voluntarios Sanos, Versión 2, fechado 28 de febrero de 2022.
- Documento de Consentimiento Informado Pacientes con Patologías de Retina y Uveítis, Versión 2, fechado 28 de febrero de 2022.
- Revocación del Consentimiento Informado, Versión 2, fechado 28 de febrero de 2022.

El Comité Ético Científico, otorga la aprobación ética de los cambios mencionados, en sesión del 06 de abril de 2022.

Saludos cordiales,



06-04-2022

Dra. Lucía Teresa Massardo Vega
Presidenta del Comité Ético Científico
Hospital Clínico de la Universidad de Chile

N° 13/22

ACTA DE EXTENSIÓN DE APROBACIÓN N° 23

Santiago 7 de septiembre de 2022.

El Comité Ético Científico o de Investigación del Hospital Clínico de la Universidad de Chile, integrado por los siguientes miembros:

Dra. Teresa Massardo Vega. Especialista en Medicina Nuclear. Presidenta
Sr. Mauricio Venegas Santos. Bioquímico. Vicepresidente.
Dra. Aída Verónica Araya Quintanilla. Médico Endocrinóloga. Secretaria de actas.
Dr. Sergio Valenzuela Puchulu. Médico Gineco-Obstetra. Integrante.
Dr. Juan Carlos Prieto Domínguez. Médico Farmacología Clínica. Integrante.
Dra. Ana María Madrid Silva. Gastroenteróloga. Integrante.
Dra. Gloria López Stewart. Médico Endocrinóloga. Integrante.
Dr. Juan Jorge Silva Solís. Médico Cirujano. Integrante.
Dr. Melchor Lemp Miranda. Médico Neurocirujano. Integrante.
Sra. Ginette Zúñiga Navarrete. Miembro de la comunidad.

Ha revisado el proyecto **“Análisis computacional para el diagnóstico temprano de patologías de retina externa y uveítis”** cuyo investigador principal es el **Dr. Cristian Urzúa Salinas de la Facultad de Medicina. Participa como investigadora responsable de proyecto, la Dra. Loreto Cuitiño Tride del Servicio de Oftalmología.** El proyecto se está desarrollando en este mismo servicio del Hospital Clínico de la Universidad de Chile.

Para estos efectos se tuvo a la vista:

- Informe de Avance de Proyecto, fechado 25 de agosto de 2022.

No encontrándose objeciones, el Comité otorga la Aprobación Ética para la continuación del Estudio propuesto, en sesión ordinaria del 7 de septiembre de 2022, la que se extiende por un plazo de 12 meses a contar de esta fecha.

Saluda atentamente a Ud.



07-09-2022

Dra. Teresa Massardo Vega
Presidenta del Comité Ético Científico
Hospital Clínico de la Universidad de Chile

**DOCUMENTO DE CONSENTIMIENTO INFORMADO PARA PACIENTES CON
PATOLOGÍAS DE RETINA EXTERNA Y UVEÍTIS**

TÍTULO DEL PROYECTO: Análisis computacional para el diagnóstico temprano de patologías de retina externa y uveítis

Nombre del Investigador principal: **Cristhian Urzúa Salinas**
Teléfono: 562 2978 8866
Email: **leaos.uchile@gmail.com**
Servicio: Oftalmología
Institución: **Hospital Clínico Universidad de Chile.**
Santos Dumont 999, 1° Piso Sector E. Independencia (Santiago).

Le estamos invitando a participar como parte del grupo de pacientes de patologías de retina y uveítis en el proyecto de investigación "Análisis computacional para el diagnóstico temprano de patologías de retina externa y uveítis", un proyecto que podría mejorar la detección y diagnóstico temprano de diversas enfermedades que afectan al ojo.

Antes de tomar la decisión de participar lea atentamente este documento.

Introducción: Las patologías de retina externa, siendo las más comunes la degeneración macular asociada a la edad y la maculopatía generada por medicamentos, y la uveítis son causa de ceguera en personas de todas las edades. El mal funcionamiento de la retina o la inflamación intraocular, llamada uveítis, puede provocar daños irreparables si su diagnóstico y tratamientos son tardíos. Los hallazgos de características importantes para su diagnóstico son llevados a cabo por el médico oftalmólogo, junto a exámenes oftalmológicos como angiografía retinal y tomografía de coherencia óptica (OCT), que son básicamente fotografías del interior del ojo.

Objetivos: Este estudio tiene como objetivo crear una herramienta computacional que ayude al médico oftalmólogo a encontrar y diagnosticar tempranamente estas enfermedades y así evitar que siga dañando a las estructuras oculares, logrando una mejor respuesta al tratamiento.

Procedimiento: Se le realizará una sesión de OCT, con los tecnólogos médicos y estudiantes de Tecnología Médica donde se realizarán varias tomas fotográficas en uno o ambos ojos. La duración de la prueba es entre 5 a 15 minutos. Durante la sesión, ud. deberá estar sentado frente al equipo. No se aplicará ningún tipo de anestesia ni medicación sedante. Se le pedirá que mire con ojos bien abiertos un estímulo fijado cuando se vaya a realizar la toma

(fotografía). La OCT es una prueba que no es molesta para el paciente, no requiere contacto con el ojo, es rápida y carece de efectos secundarios.

Riesgos: Esta intervención no tiene riesgos para Ud., ya que el examen que se le hará (OCT) sólo saca fotografías en alta definición de manera no invasiva, por lo que no produciría complicaciones adicionales ni alteraría del curso natural de la enfermedad.

Costos: Los exámenes oftalmológicos serán aportados por el Hospital Clínico Universidad de Chile, sin costo alguno para Ud. durante el desarrollo de este proyecto. Por lo tanto, su participación no le representará gastos adicionales.

Beneficios: Este estudio significará un gran progreso para la detección y tratamiento de las patologías de retina externa y uveítis, que será útil especialmente para los futuros pacientes que las desarrollen. Además, su participación le traerá los siguientes beneficios:

- Atención con especialista oftalmólogo para evaluar el estado de su retina y úvea.
- Detección y derivación, sólo si corresponde, a otro especialista oftalmólogo por presencia de enfermedad ocular distintas a las estudiadas, y que no se encuentre con tratamiento o control.
- Quedar ingresado en la fuente de datos del equipo de OCT, del Hospital Clínico Universidad de Chile, que le podría permitir el seguimiento de la evolución de su enfermedad ocular.

Alternativas: Si Ud. decide no participar en esta investigación recibirá la atención y tratamiento que se aplica habitualmente, el que no necesariamente contempla el uso del equipo de OCT.

Compensación: Ud. no recibirá ninguna compensación económica por su participación en el estudio.

Confidencialidad: Toda la información derivada de su participación en este estudio será conservada en forma de estricta confidencialidad, lo que incluye el acceso de los investigadores o agencias supervisoras de la investigación. Cualquier publicación o comunicación científica de los resultados de la investigación será completamente anónima.

Voluntariedad: Su participación en esta investigación es totalmente voluntaria y se puede retirar en cualquier momento comunicándolo al investigador y a su médico tratante, sin que ello signifique modificaciones en la atención y tratamientos habituales de su enfermedad. De igual manera su médico tratante o el investigador podrán determinar su retiro del estudio si consideran que esa decisión va en su beneficio.

Complicaciones: En el improbable caso de que Ud. presente complicaciones directamente por el uso del equipo de OCT, Ud. recibirá el tratamiento médico completo de dicha complicación y sin costo alguno para Ud. o su sistema previsional, siendo este costo financiado por el equipo de investigadores. **Esto no incluye las complicaciones propias de su enfermedad y de su curso natural.**

Derechos del participante: Usted recibirá una copia de este documento firmado. Si Ud. requiere cualquier otra información sobre su participación en este estudio o sus derechos, puede comunicarse con:

Investigador:

Cristhian Urzúa Salinas

Teléfono: 562 2978 8866

Email: leaos.uchile@gmail.com

Santos Dumont 999, 1° Piso Sector E. Independencia (Santiago).

Comité Ético Científico del Hospital Clínico Universidad de Chile:

Teléfono: 562 2978 9008

Email: comiteetica@hcuch.cl

Santos Dumont 999, 4° Piso Sector D. Independencia (Santiago).

Conclusión: Después de haber recibido y comprendido la información de este documento y de haber podido aclarar todas mis dudas, otorgo mi consentimiento libre y voluntario para participar en el proyecto "Análisis computacional para el diagnóstico temprano de patologías de retina externa y uveítis"

