



---

**Universidad de Valladolid**

FACULTAD DE CIENCIAS

# **Grado en Óptica y Optometría**

MEMORIA TRABAJO FIN DE GRADO TITULADO

Screening visual con inteligencia artificial  
aplicada en imágenes pupilares

Presentado por Claudia Plaza Cimas

Tutelado por Irene Sánchez Pavón y Roberto Román Díez

Tipo de TFG:  Revisión  Investigación

En Valladolid a, 25 de mayo de 2025

## ÍNDICE

RESUMEN .....	3
ABSTRACT .....	4
1. INTRODUCCIÓN .....	5
2. MATERIAL Y MÉTODO .....	8
2.1. Diseño del estudio .....	8
2.2 Material .....	8
2.3 Metodología .....	8
2.3.1. Etiquetado de las imágenes .....	9
2.3.2 Entrenamiento de la red neuronal .....	10
2.4 Análisis estadístico .....	12
3. RESULTADOS .....	13
3.1. Conjunto de entrenamiento .....	13
3.2. Conjunto de validación .....	14
4. DISCUSIÓN .....	16
4.1. Limitaciones .....	17
5. CONCLUSIONES .....	18
6. BIBLIOGRAFÍA .....	19

## RESUMEN

**Introducción:** Las disfunciones visuales en la infancia, como la anisometropía, la ambliopía o el estrabismo, pueden tener consecuencias irreversibles si no se detectan a tiempo. Sin embargo, las consultas de pediatría no disponen del material ni del tiempo suficiente para poder realizar un correcto screening visual, lo cual impide la detección temprana de patologías importantes que pueden pasar desapercibidas. En este contexto, la fotorrefracción es un método rápido, no invasivo y útil para estimar el estado refractivo del ojo mediante un análisis del reflejo rojo pupilar.

**Material y método:** Se ha llevado a cabo un estudio piloto, comparativo y no invasivo, con 32 voluntarios de entre 18 y 45 años, seleccionados de forma no sistemática, y se han aplicado los criterios de inclusión y exclusión con el fin de garantizar la validez del estudio. Se ha realizado una exploración optométrica completa y se han obtenido imágenes del reflejo rojo de la pupila mediante una cámara seleccionada para fotorrefracción. Las imágenes han sido etiquetadas manualmente según el tipo de error refractivo (miopía, emetropía o hipermetropía), en función de la localización del reflejo pupilar. Se ha entrenado una red neuronal convolucional utilizando la herramienta "Teachable Machine 2.0" para clasificar automáticamente las imágenes de los ojos, evaluando el rendimiento mediante un análisis estadístico.

**Resultados:** El modelo ha obtenido un rendimiento del 99,2% de precisión sobre el conjunto de entrenamiento, con una media ponderada del 99,5% al considerar el nivel de confianza. Ha clasificado correctamente el 100% de las imágenes de ojos miopes y emétropes y el 97,6% de las imágenes de ojos hipermetropes, y ha acertado las tres imágenes de un mismo paciente en el 97,6% de los casos. En el conjunto de validación, la precisión ha descendido a un 63,6%, aumentando al 69% al considerar el nivel de confianza. El modelo ha mostrado mayor precisión en la clasificación de ojos miopes (88,9%) y menor en hipermetropes (50%) y emétropes (44,4%). A nivel de paciente, ha clasificado correctamente las tres imágenes en el 45,5% de los casos.

**Conclusiones:** Los resultados han mostrado una buena capacidad de aprendizaje del modelo, pero también una generalización limitada, lo que ha resaltado la necesidad de ampliar la base de datos para mejorar su utilidad como herramienta de screening visual, especialmente en población pediátrica.

**Palabras clave:** fotorrefracción, screening visual, inteligencia artificial.

## ABSTRACT

**Introduction:** Visual dysfunctions in childhood, such as anisometropia, amblyopia or strabismus, can have irreversible consequences if not detected in time. However, paediatricians consultations do not have enough material or time to carry out a correct visual screening, which prevents the early detection of important pathologies that may go unnoticed. In this context, photorefraction is a fast, non-invasive and useful method to estimate the refractive error of the eye by analyzing the red pupillary reflex.

**Material and Method:** A pilot, comparative, non-invasive study was carried out with 32 volunteers between 18 and 45 years, selected non-systematically and inclusion and exclusion criteria were applied to guarantee the validity of the study. A complete optometric examination was performed, and images of the red pupillary reflex were obtained using a camera selected for photorefraction. The images were manually labelled according to the type of refractive error (myopia, emmetropia or hyperopia), based on the location of the pupillary reflex. A convolutional neural network was trained using the "Teachable Machine 2.0" tool to automatically classify the eyes, evaluating the performance by statistical analysis.

**Results:** The model performed 99,2% accurate over the training set, with a weighted average of 99,5% when considering the confidence level. It correctly classified 100% of the images of myopic and emmetropic eyes and 97,6% of the images of hyperopic eyes and got all three images of the same patient correct in 97,6% of the cases. In the validation set, accuracy decreased to 63,6%, increasing to 69% when considering the confidence level. The model showed higher accuracy in the classification of myopic eyes (88,9%) and lower accuracy in hyperopic (50%) and emmetropic (44,4%). At the patient level, it correctly classified all three images in 45,5% of the cases.

**Conclusions:** The results have shown a good learning ability of the model, but also limited generalization, which has highlighted the need to expand the database to improve its usefulness as a visual screening tool, especially in paediatric population.

**Keywords:** photorefraction, visual screening, artificial intelligence.

## 1. INTRODUCCIÓN

Las disfunciones visuales en niños son de gran relevancia ya que puede ocasionar alteraciones irreversibles en el sistema visual si se ocasionan dentro del periodo de desarrollo. Generalmente, las consultas de pediatría no disponen del material ni del tiempo suficiente para poder realizar un correcto screening visual a los niños, lo cual impide la detección de patologías importantes que pueden pasar desapercibidas. Además, ciertas pruebas que realizan habitualmente los pediatras, como los test de agudeza visual con optotipos, requieren de la colaboración del niño. Esto puede resultar complicado en ciertas ocasiones o edades, requiriendo en muchos casos más tiempo del disponible.

También es importante destacar que, aunque los niños tengan una agudeza visual adecuada a su edad, en un momento concreto (como podría ser la revisión de los 3-4 años que suelen realizar los pediatras) no es garantía de un correcto desarrollo visual. Por ello, es recomendable realizar visitas de seguimiento. El pediatra, basándose en los escasos datos visuales que ha podido obtener debido a la carencia de material que ayude al diagnóstico visual pediátrico, se ve obligado en muchas ocasiones a remitir al servicio de oftalmología como medida preventiva. Esto provoca el aumento de remisiones no justificadas al servicio de oftalmología y la demora y aumento de las listas de espera.

Existen varios factores que pueden alterar el desarrollo visual de manera irreversible como por ejemplo las anisometropías, la ambliopía o el estrabismo. La ametropía se define como aquella alteración en el poder refractivo del ojo, en la que, sin acomodar, el punto conjugado de la retina no coincide con el infinito.<sup>1</sup> Esto provoca que la imagen de un objeto que está situado en el infinito óptico se forme por delante o por detrás de la retina, generando visión borrosa. Clínicamente, existen tres tipos de ametropía: miopía (el ojo tiene un exceso de potencia), hipermetropía (el ojo tiene un defecto de potencia) y astigmatismo.<sup>1</sup> La emetropía, en cambio, se define como el estado refractivo del ojo en el que el punto remoto se sitúa en el infinito, lo que permite que la imagen procedente de un objeto situado en el infinito óptico se forme en la retina.<sup>1</sup>

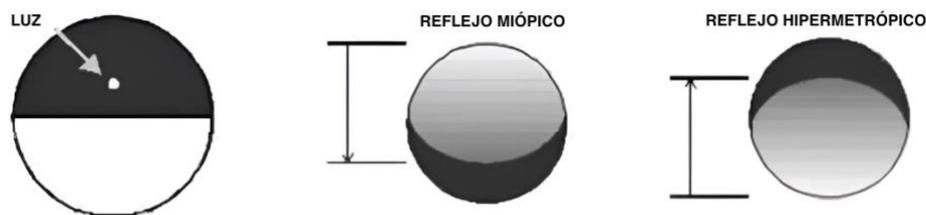
La ametropía más común en la actualidad es la miopía, siendo una de las principales causas de discapacidad visual reversible y ceguera en el mundo en la edad adulta. La prevalencia de la miopía sigue aumentando y se prevé que para 2050, el 50% de la población mundial será miope.<sup>2</sup> Si la miopía se desarrolla antes de los 6 años, su progresión puede incrementar las posibilidades de la aparición de alteraciones retinianas en la edad adulta.<sup>3</sup> La detección precoz de la miopía, permite plantear ciertos métodos que pueden frenar su avance y por tanto que pueden evitar que en la edad adulta el paciente presente miopía magna.<sup>3,4</sup> De ahí, la importancia de realizar un correcto y efectivo screening visual. Sin embargo, la miopía es el error refractivo que con menor probabilidad puede producir ambliopía.

La ambliopía se define como la disminución de la visión sin precedente patológico, siendo el resultado de un desarrollo deficiente del sistema visual durante la primera década de vida. La presencia de estímulos anómalos durante este periodo provoca un incorrecto desarrollo de los centros visuales del córtex visual produciendo una disminución de la agudeza visual, unilateral o bilateral con ausencia de estereopsis, que puede ser irreversible si no se corrige antes

de que finalice la primera década de la vida.<sup>5,6,7</sup> Estos estímulos anómalos suelen ser principalmente: errores refractivos no corregidos, tanto moderados (a partir de 1,50 D) como elevados (principalmente hipermetropía y/o astigmatismo); diferencias entre ambos ojos (anisometropías de 1,00 D ya que pueden ser suficientes para cursar con ambliopía); o la presencia de un estrabismo, entre otras alteraciones. La prevalencia de la ambliopía varía en función de la población y el estudio consultado, estimándose que afecta entre el 2 y el 4% de la población a partir de la segunda década de la vida, entre el 3 y el 4% de los niños en edad preescolar y entre el 2 y el 7% de los niños en edad escolar.<sup>8</sup> Se estima que a nivel mundial afecta entre el 1% y el 4%.<sup>9</sup>

El sistema visual necesita de una correcta estimulación para poder desarrollarse adecuadamente, con la formación de imágenes retinianas correctamente enfocadas y nítidas en el área macular. El periodo de mayor maduración del sistema visual se encuentra entre los primeros seis años de vida, periodo tras el cual la plasticidad neurovisual disminuye.<sup>5,10</sup> Existe un periodo crítico en el que cualquier alteración tendrá importantes consecuencias sobre el sistema visual del adulto. Cuanto más temprano se produzca, mayor será esta alteración. Por otro lado, cuanto antes se detecte el factor ambliogénico que está impidiendo el desarrollo visual normal del niño y se aplique el tratamiento adecuado dentro del periodo de plasticidad cerebral, más aumenta la probabilidad de recuperar una agudeza visual óptima y una fusión estable con estereopsis. Por ello, es realmente importante establecer un programa de screening visual pediátrico eficaz, en el que se realicen revisiones visuales a niños en edades tempranas, prestando especial atención a aquellos niños que posean antecedentes familiares de ambliopía o de posibles factores ambliogénicos.<sup>11,12</sup>

Un método rápido, sencillo y no invasivo que permite para estimar el estado refractivo del ojo es la fotorrefracción.<sup>10,13,14</sup> Esta técnica consiste en iluminar el ojo con una fuente luminosa y capturar la imagen del reflejo retiniano. El reflejo rojo que emerge a través de la pupila presenta características que varían en función de la ametropía (Figura 1).



**Figura 1.** Esquema del principio de funcionamiento de la fotorrefracción. La dirección de la sombra reflejada en la pupila depende del defecto refractivo.

La eficacia de los fotorrefractómetros, que analizan el reflejo rojo en las fotos capturadas, se ha probado para la detección de anisometropía, hipermetropía, miopía y astigmatismo con una precisión suficiente para su uso como técnica de screening, especialmente en niños.<sup>9,10</sup> El análisis de estas fotos podría realizarse a través de las técnicas más novedosas que existen, que actualmente incluyen dentro de lo que se conoce como inteligencia artificial.

Por lo tanto, la necesidad de disponer de técnicas de screening rápidos y accesibles como la fotorrefracción, que proporcionen una evaluación directa con métodos rápidos y precisos, motiva el objetivo principal de este trabajo: realizar una prueba de concepto de fotorrefracción con una cámara tradicional, y analizar las imágenes del reflejo mediante técnicas de inteligencia artificial conocidas como Machine Learning. En concreto redes neuronales convolucionales, que han demostrado ser capaces de identificar patrones y características, incluyendo el error refractivo, en imágenes de fotorrefracción.<sup>14</sup>

## 2. MATERIAL Y MÉTODO

### 2.1 Diseño del estudio

Se ha realizado un estudio piloto comparativo y no invasivo, con pacientes de entre 18 y 45 años, seleccionados de forma no sistemática. Todos aquellos voluntarios que cumplían los criterios de inclusión y aceptaron participar en el estudio, tras ser informados del procedimiento y su naturaleza no invasiva, fueron incluidos.

Se incluyeron al estudio aquellos pacientes emétopes, amétopes e incluso con ambliopía o estrabismo y se excluyeron del estudio a aquellos pacientes que presentaban cualquier tipo de patología ocular en la que se viera afectada la retina y que impidiera la correcta visión y su desarrollo, a excepción de los anteriormente citados. También se excluyeron a los que presentaban alteraciones en la transparencia de los medios oculares que dificultara la obtención de una refracción objetiva, así como a los pacientes con alguna patología sistémica con afectación ocular. Tras aplicar estos criterios, participaron 32 pacientes.

### 2.2 Material

El equipamiento empleado es el que habitualmente permite realizar una exploración optométrica habitual, en la que se incluyen distintas fases de evaluación. Los materiales empleados para cada procedimiento fueron los siguientes:

- Refracción: Este procedimiento permite determinar la corrección óptica del paciente, obteniendo esfera, cilindro y eje de ambos ojos. Se realizó tanto con el método objetivo como con el método subjetivo. Para ello, se utilizó un foróptero (Phorovist200, Rodenstock) y un retinoscopio (Beta 200, Heine).
- Revisión del polo anterior y posterior: En esta fase se evaluaron diversas estructuras oculares como la córnea, la conjuntiva, el cristalino y la retina, entre otras. Para su observación, se utilizó una lámpara de hendidura (SL-2E, TOPCON) y un oftalmoscopio (Beta 200, Heine).
- Frontofocómetro: Este dispositivo se empleó para medir la potencia de la corrección de aquellos pacientes que usaran lentes.
- Evaluación del sistema binocular: Este procedimiento evalúa la capacidad de ambos ojos para trabajar de manera conjunta y efectiva. Se incluyó el test de estereopsis (TNO, OO Tech Lameris), así como la cuantificación de la foria o estrabismo, realizado mediante el cover test.
- Fotografía: Para capturar las imágenes se utilizó una cámara Nikon D5200, específicamente configurada para obtener el reflejo de ojos rojos. Para garantizar la aparición de este reflejo, se desactivó la opción que evita que aparezca este efecto.

### 2.3 Metodología

Se explicó detalladamente a los pacientes la naturaleza del estudio, resolviendo todas las preguntas necesarias y proporcionando la hoja de información. Una vez firmado el consentimiento informado, se realizaron las pruebas del estudio.

En primer lugar, se llevó a cabo una exploración optométrica completa, que incluyó la toma de la agudeza visual tanto monocular como binocular, la medición de la estereopsis (utilizando específicamente el test TNO), la realización del cover test, la obtención de la refracción (esfera, cilindro y eje) mediante dos métodos: el objetivo (retinoscopia) y el subjetivo (con la colaboración del paciente). Además, se evaluaron la motilidad ocular extrínseca e intrínseca, y se realizó la revisión de los polos anterior y posterior del ojo.

Posteriormente, se realizaron tres fotografías utilizando la cámara seleccionada para fotorrefracción, colocándola a la distancia focal más cercana que permitiera obtener el reflejo rojo necesario (2.85 m). Las fotografías fueron recortadas para mostrar únicamente los ojos, garantizando la confidencialidad del paciente. Para poder obtener las imágenes monoculares, se recortaron las fotografías binoculares, separando ambos ojos. De cada paciente se obtuvieron 3 fotografías monoculares del ojo derecho y 3 fotografías monoculares del ojo izquierdo, sumando un total de 6 fotografías por paciente.

En cuanto a la distribución de los 32 pacientes que participaron en el estudio, se optó por considerar cada ojo como una unidad de análisis independiente, tratándolos como si correspondieran a pacientes distintos, obteniendo por ello un total de 64 pacientes. Esta decisión metodológica permite aumentar el tamaño muestral disponible y mejorar la capacidad del modelo para aprender y generalizar a partir de un mayor número de muestras.

### 2.3.1 Etiquetado de las imágenes

Esta etapa consistió en etiquetar manualmente las imágenes correspondientes a cada paciente, asignando a cada una de ellas una etiqueta que las clasifica según diversas categorías. Estas categorías se definieron en función de la refracción del paciente, diferenciándose tres categorías generales:

1. Ojos miopes: Esta categoría incluyó a los pacientes con una refracción de  $-0,50$  D o inferior. En las imágenes de esta categoría, el reflejo rojo se presenta de forma significativa más pronunciado en la parte superior de la pupila (Figura 2a).

2. Ojos emétopes: Esta categoría incluyó a los pacientes con una refracción comprendida entre  $-0,50$  D a  $+0,75$  D. En las imágenes de esta categoría, el reflejo rojo se distribuye de manera homogénea a través de toda la pupila (Figura 2b).

3. Ojos hipermétropes: Esta categoría incluyó a los pacientes con una refracción mayor a  $+0,75$  D. En las imágenes de esta categoría, el reflejo rojo se presenta de forma significativa más pronunciado en la parte inferior de la pupila (Figura 2c).



(a) Ojo miope

(b) Ojo emétope

(c) Ojo hipermétrope

**Figura 2.** Ejemplo de la proyección de la luz en las diferentes categorías.

Este procedimiento de clasificación de imágenes se realizó de manera monocular, cada ojo por separado.

### 2.3.2 Entrenamiento de la red neuronal

Una vez se completó el etiquetado de las imágenes, estas se dividieron en dos conjuntos: entrenamiento y validación. El conjunto de entrenamiento se utiliza para entrenar una red neuronal, mientras que el de validación se emplea posteriormente para cuantificar el rendimiento del modelo entrenado.

Una red neuronal es un modelo matemático compuesto por capas de unidades conectadas entre sí, capaz de aprender representaciones a partir de datos mediante la optimización de pesos sinápticos. Una red neuronal convolucional es una arquitectura especializada de red neuronal que utiliza operaciones de convolución para extraer automáticamente características espaciales locales, siendo especialmente eficaz en el procesamiento de datos estructurados en forma de imágenes.

En este caso se ha utilizado y entrenado una red neuronal convolucional como modelo que sea capaz de clasificar las imágenes de los ojos según la ametropía del paciente de forma automática. Para obtener el modelo se ha utilizado la herramienta de Google “Teachable Machine 2.0”. El acceso a esta herramienta es totalmente gratuito y solo precisa de conexión a internet. Se puede acceder mediante el siguiente enlace: <https://teachablemachine.withgoogle.com/>

La interfaz de “Teachable Machine 2.0” (Figura 3) está dividida en tres secciones principales: las imágenes del conjunto de entrenamiento se seleccionan en la parte izquierda, los parámetros de entrenamiento se escogen en la zona central y pulsando “preparar modelos” se inicia el entrenamiento. Una vez finalizado éste se puede exportar desde la parte derecha.

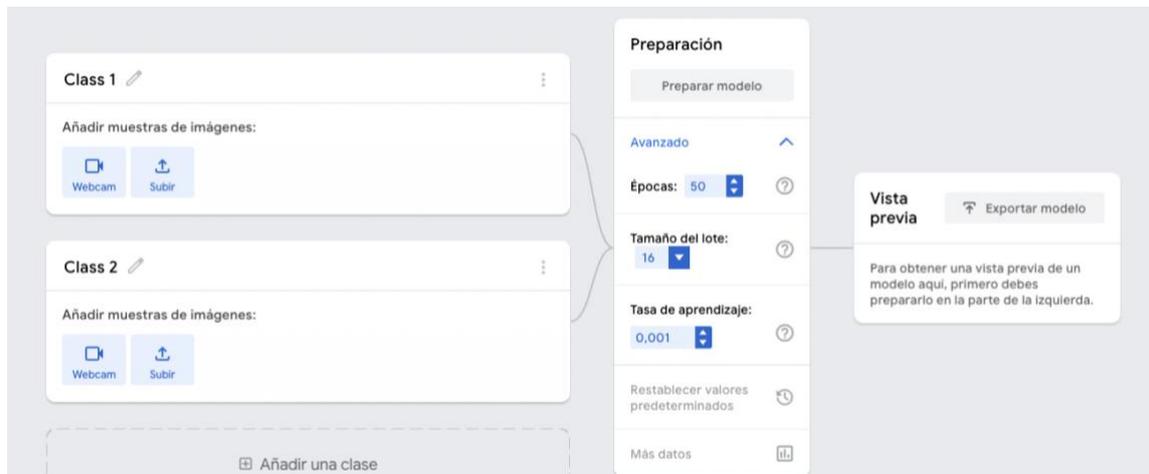


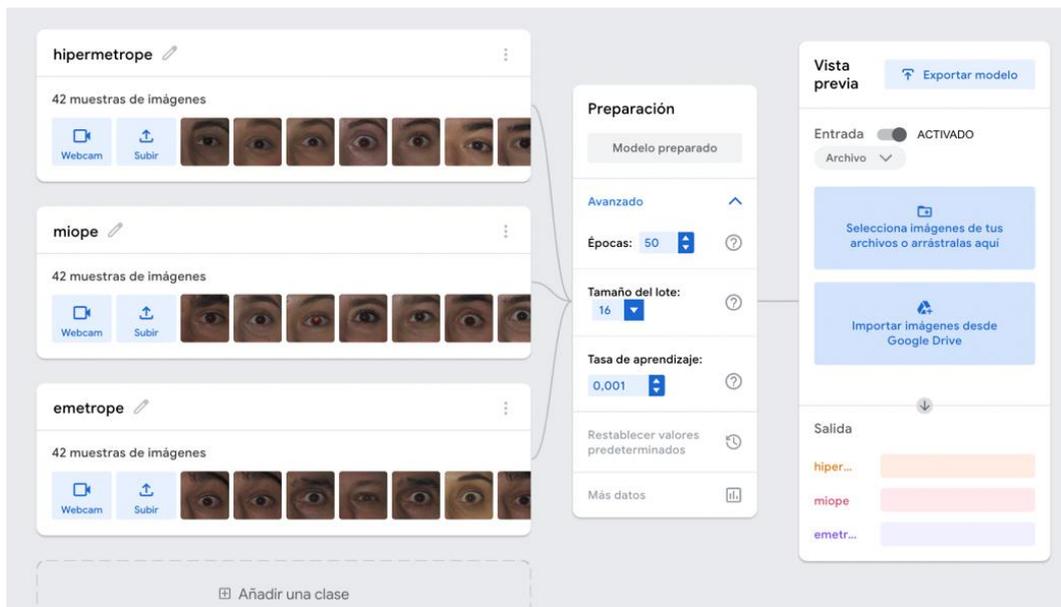
Figura 3. Interfaz de “Teachable Machine 2.0”.

Para este modelo se han definido tres clases de imágenes correspondientes a ojos hipermétropes, ojos miopes y ojos emétopes. A partir del total de imágenes obtenidas en el estudio, se realizó una división estándar en los dos conjuntos mencionados de entrenamiento y validación.

Se contó con un total de 18 pacientes hipermétropes, 23 pacientes miopes y 23 pacientes emétopes. Dado que el grupo de hipermétropes era menos numeroso, se decidió igualar el número de pacientes utilizados para el conjunto de entrenamiento en las tres clases, tomando 14 pacientes por clase (80% de los pacientes hipermétropes) de forma aleatoria con un código programado en lenguaje Python. De este modo, se garantiza que el modelo reciba un número equilibrado de ejemplos por categoría en la fase de entrenamiento, lo que contribuye a evitar sesgos en la clasificación.

Cada paciente aporta tres imágenes, por lo que el total de imágenes utilizadas en el conjunto de entrenamiento fue de 42 por clase, sumándose un total de 126 imágenes. Para la validación del modelo se utilizaron las imágenes restantes: 27 imágenes de ojos miopes (correspondiente a 9 pacientes), 27 imágenes de ojos emétopes (9 pacientes) y 12 imágenes de ojos hipermétropes (4 pacientes), siendo un total de 66 imágenes.

Las imágenes se procesaron utilizando el formato estándar requerido por el modelo, es decir, imágenes a color con una resolución de 224x224 píxeles. Las imágenes seleccionadas para el entrenamiento se subieron a la interfaz “Teachable Machine 2.0” ya con las tres clases definidas.



**Figura 4.** Interfaz de “Teachable Machine 2.0” con las clases e imágenes definidas.

“Teachable Machine 2.0” permite ajustar tres parámetros de entrenamiento, de los cuales se seleccionaron los valores por defecto:

- **Épocas:** El entrenamiento de una red neuronal es un proceso iterativo debido a la no linealidad de esos modelos. Cuando el modelo de preparación ha procesado todas las muestras del conjunto de datos de preparación una vez, se dice que se ha completado una época, o iteración completada. Por ejemplo, si se establecen 50 épocas, el modelo procesará 50 veces todo el conjunto de datos de preparación. Por lo general, cuanto mayor es el número, mejor aprenderá el modelo a predecir los datos. Se recomienda ajustar este número hasta que el modelo ofrezca buenos resultados predictivos. Sin embargo, un valor muy elevado podría causar el fenómeno de sobreajuste (también

conocido como *overfitting*): el modelo se centra en aprender demasiado bien la clasificación de las imágenes de entrenamiento, pero empeora su desempeño en imágenes nuevas (que no formaban parte del conjunto de entrenamiento). Por lo tanto, es importante elegir este valor cuidadosamente. Su valor por defecto es 50.

- **Tamaño del lote:** Procesar todas las imágenes a la vez en una época es un proceso costoso computacionalmente, por lo que las imágenes que se procesan en una época se suelen dividir en lotes de imágenes más pequeños. Un lote es un conjunto de muestras que se usan en una iteración de preparación. Por ejemplo, si hay 80 imágenes y el tamaño de lote es de 16, el conjunto se dividirá en 5 lotes de 16 imágenes. Una vez procesados los 5 lotes, se completa una época. Su valor por defecto es 16.
- **Tasa de aprendizaje:** Es un valor numérico entre 0 y 1 que regula cuanto cambia el modelo con cada iteración durante el entrenamiento. La tasa de aprendizaje limita la cantidad de información que puede aprender el modelo en cada iteración. Este parámetro se emplea para evitar el fenómeno de sobreajuste. Sin embargo, una tasa de aprendizaje muy pequeña hará que el modelo aprenda muy lentamente, aumentando el coste computacional. Se trata de un parámetro que debe ajustarse correctamente, pequeños cambios pueden afectar enormemente a la eficacia del modelo. Su valor por defecto es 0,001.

Una vez completados todos los apartados, se ejecutó el entrenamiento (opción “preparar modelo”), y se obtuvo y exportó el modelo resultante. Este modelo generado se utilizó para realizar predicciones sobre el conjunto de imágenes de entrenamiento y de validación. Finalmente se generó un fichero con los resultados obtenidos, que permitió analizar el rendimiento del modelo sobre los datos.

## 2.4 Análisis estadístico

Una vez realizadas todas las medidas, los datos clínicos se recogieron en una hoja de Microsoft Office Excel 2013 (V16.77-23091003) para su posterior análisis estadístico. En esta hoja de cálculo se juntaron las predicciones del modelo entrenado sobre cada imagen del conjunto de validación con los datos clínicos obtenidos para cada ojo. Lo mismo se hizo para el conjunto de entrenamiento. Es importante remarcar que el modelo no solo da una predicción para una foto dada (miope, hipermetrope o emétrope), sino que también devuelve un nivel de confianza, entre 0 y 100%, que cuantifica el grado de confianza con el que se hace esa predicción. Este nivel de confianza también se añadió a la hoja de cálculo. Los resultados obtenidos se compararon con los datos clínicos que se obtuvieron.

Con toda esta información, se ha analizado en la siguiente sección la sensibilidad y especificidad mediante tablas cruzadas (tablas de contingencia). Los conceptos de sensibilidad y especificidad permiten valorar la validez de la prueba diagnóstica.

Los índices estadísticos que se calcularon fueron la media aritmética, la media ponderada para cuantificar la exactitud y la precisión del modelo.

### 3. RESULTADOS

Los resultados del modelo se dividen en dos bloques: el rendimiento sobre el conjunto de entrenamiento y sobre el conjunto de validación. En ambos casos, se analiza la precisión de las predicciones y la confianza asociada a las mismas. Además, se incluye un análisis por paciente.

#### 3.1 Conjunto de entrenamiento

Como se puede observar en la Tabla 1, el modelo clasificó correctamente el 100% de las imágenes de ojos miopes, sin confundir ninguna con otra clase. Lo mismo ocurrió con las imágenes de ojos emétopes, donde el modelo obtuvo un 100% de acierto, clasificando correctamente las imágenes. Sin embargo, en las imágenes de los ojos hipermétropes, el modelo clasificó correctamente el 97,62% de ellas, interpretando como miope el 2,38%. En conjunto, el modelo acertó 125 imágenes, lo que representa un 99,2% de aciertos en las imágenes utilizadas para el entrenamiento.

		Predicción		
		Miope	Emétrope	Hipermétrope
Real	Miope	42	0	0
	Emétrope	0	42	0
	Hipermétrope	1	0	41

**Tabla 1.** Matriz de confusión del conjunto de entrenamiento.

Considerando el nivel de confianza con el que el modelo hace estas predicciones (a través de la media ponderada), el porcentaje global de aciertos se eleva a 99,5% (Tabla 2).

Clase	Media ponderada de aciertos (%)
Miope	100%
Emétrope	100%
Hipermétrope	98,6%
Todas	99,5%

**Tabla 2.** Precisión ponderada por clase en el conjunto de entrenamiento.

Los aciertos ponderados correspondientes a las imágenes de los ojos hipermétropes presentan un porcentaje ligeramente superior en comparación con la precisión no ponderada. Este incremento se debe a que el único error cometido por el modelo se produce en una imagen clasificada como ojo miope, con un nivel de confianza del 57%. Este valor es relativamente bajo e indica que el modelo no estaba completamente seguro de su predicción, lo que justifica que, al considerar la ponderación por confianza, el impacto de este error se vea atenuado en el cálculo global de aciertos.

Además, resulta relevante evaluar el rendimiento del modelo a nivel de cada paciente. Cada sujeto en el conjunto de entrenamiento está representado por tres imágenes, lo cual permite comprobar no solo la precisión puntual sino también su consistencia al clasificar distintas imágenes de un mismo paciente.

Como se observa en la Tabla 3, el modelo logró clasificar correctamente las tres imágenes en el 97,62% de los casos, y solo en un 2,38% de los pacientes cometió un error clasificando correctamente dos de las tres imágenes.

N.º de imágenes correctas	N.º de pacientes
0	0 (0%)
1	0 (0%)
2	1 (2,4%)
3	41 (97,6%)

**Tabla 3.** Análisis de aciertos por paciente del conjunto de entrenamiento.

### 3.2 Conjunto de validación

Los resultados obtenidos en el conjunto de entrenamiento, como es de esperar, presentan un elevado rendimiento dado que el modelo ha sido ajustado específicamente a estos datos. No obstante, el rendimiento real del modelo debe evaluarse sobre datos no utilizados durante el proceso de entrenamiento, como son los pertenecientes al conjunto de validación.

		Predicción		
		Miope	Emétrope	Hipermétrope
Real	Miope	24	1	2
	Emétrope	8	12	7
	Hipermétrope	6	0	6

**Tabla 4.** Matriz de confusión del conjunto de validación.

En la Tabla 4, se observa que de las 27 imágenes de ojos miopes que hay, el modelo clasificó correctamente el 88,9% de ellas, el 7,4% las interpretó como ojos hipermétropes y el 3,7% como emétopes. En el caso de las imágenes de ojos emétopes, el modelo acertó 44,4% de ellas, identificando el 29,6% como miopes y el 25,9% como hipermétropes. En el caso de los ojos hipermétropes, el modelo acertó el 50%, identificando el otro 50% como ojos miopes. En conjunto, el modelo acertó 42 imágenes, lo que representa un 63,6% de aciertos en las imágenes utilizadas para la validación.

Considerando el nivel de confianza con el que el modelo hace estas predicciones (a través de la media ponderada), el porcentaje global de aciertos se eleva a 69% (Tabla 5).

Los aciertos ponderados correspondientes a las imágenes de los ojos hipermétropes el resultado es de un 53,5% siendo ligeramente superior en comparación con la precisión no ponderada.

<b>Clase</b>	<b>Media ponderada de aciertos (%)</b>
<b>Miope</b>	91,9%
<b>Emétrope</b>	51,1%
<b>Hipermétrope</b>	53,5%
<b>Todas</b>	69,0%

**Tabla 5.** Precisión ponderada por clase.

Al evaluar el rendimiento del modelo por paciente (Tabla 6), se puede observar que en un 22,7% no acierta ninguna de las tres imágenes, en un 9,1% acierta una sola imagen de las tres presentadas, en otro 22,7% acierta dos imágenes y en el 45,5% acierta las tres imágenes presentadas.

<b>N.º de imágenes correctas</b>	<b>N.º de pacientes</b>
<b>0</b>	5 (22,7%)
<b>1</b>	2 (9,1%)
<b>2</b>	5 (22,7%)
<b>3</b>	10 (45,5%)

**Tabla 6.** Análisis de aciertos por paciente.

Aparece un caso llamativo dentro de los pacientes hipermétropes, con un paciente con una hipermetropía elevada de +7,00 D. En este caso concreto, el modelo identificó correctamente como ojo hipermétrope solo una de las tres imágenes con una confianza del 71%. Sin embargo, las otras dos imágenes fueron clasificadas erróneamente como ojos miopes, con niveles de confianza elevados, del 85% y 98%. Este resultado es peculiar debido a la magnitud de la hipermetropía. Esto puede deberse a la escasa representación de casos de hipermetropías elevadas en el conjunto de entrenamiento del modelo.

## 4. DISCUSIÓN

Dado que las consultas de pediatría no suelen disponer del material ni del tiempo suficiente para poder realizar un correcto screening visual a los niños, la inteligencia artificial se presenta como una herramienta de bajo coste y de fácil utilización para establecer un programa de screening visual pediátrico eficaz y rápido. El objetivo principal de este estudio es valorar la posibilidad de realizar un screening visual a partir del análisis con inteligencia artificial de imágenes pupilares, centrándose en el reflejo rojo característico de cada tipo de ametropía.

Los resultados obtenidos en el conjunto de validación muestran que la red neuronal convolucional entrenada ha logrado una exactitud notablemente inferior a la del conjunto de entrenamiento. Pese a ello, se sitúa en un rango cercano al umbral mínimo que suele considerarse aceptable para herramientas de apoyo al diagnóstico. Esto sugiere que el modelo presenta una limitada capacidad de generalización, siendo una característica frecuente en modelos entrenados con bases de datos pequeñas o poco representativas.

El rendimiento de los modelos generados por la inteligencia artificial depende de la base de datos empleada durante el entrenamiento, tanto por la cantidad o la calidad de los datos. Cuando se entrenan algoritmos con pequeñas bases de datos, menos de mil imágenes, la eficacia del modelo puede verse comprometida, especialmente si los datos no están bien codificados o etiquetados.<sup>15</sup> En estos casos, se recurre a la técnica de generación de datos artificiales, aumentando la muestra generando variaciones en los datos originales. Sin embargo, se pueden alcanzar resultados razonables aún con la relativa falta de datos.<sup>15</sup>

La literatura científica apoya esta observación. Estudios como los de Lin et al. y Malik et al. mencionan que el rendimiento de los modelos de Machine Learning en el ámbito de la oftalmología mejora significativamente cuando se entrenan con bases de datos clínicas amplias, diversas y bien etiquetadas.<sup>16,17</sup> En este estudio realizado, el número limitado de imágenes ha podido contribuir en una menor precisión durante el proceso de validación.

Este planteamiento también coincide con los hallazgos de Fageeri et al. y Achiron et al., quienes demostraron la capacidad de modelos basados en algoritmos como árboles de decisión para clasificar errores refractivos y predecir resultados quirúrgicos con alta precisión.<sup>18,19</sup> El uso de inteligencia artificial en este trabajo, al igual que en los estudios mencionados anteriormente, demuestra su potencial como herramienta de apoyo en la toma de decisiones clínicas, especialmente cuando se dispone de recursos limitados.

El método de adquisición de datos, basado en fotorrefracción, tiene ventajas prácticas importantes: permite obtener imágenes con dispositivos sencillos y no invasivos, sin necesidad de una colaboración activa del paciente, lo que lo convierte en una opción especialmente útil en población pediátrica.

La mayor precisión obtenida en la clasificación de las imágenes de pacientes con miopía coincide con estudios como el de Lin et al., que aplicaron modelos de Machine Learning para predecir el desarrollo de miopía en edad escolar, resaltando la utilidad de estos algoritmos para el diagnóstico precoz y el control de su progresión.<sup>16</sup>

#### **4.1 Limitaciones**

Una de las principales limitaciones de este estudio es el tamaño muestral. La escasa variedad de refracciones dentro de cada categoría dificultó el entrenamiento del modelo, especialmente en los casos de ojos hipermétropes con valores elevados de refracción. Esta limitación afecta a la capacidad de generalización del modelo, destacando la necesidad de aumentar y equilibrar el tamaño muestral para mejorar el rendimiento y fiabilidad del modelo en la detección de los diferentes errores refractivos.

El estudio se realizó exclusivamente en población adulta sana, a pesar de que este dispositivo se pretende utilizar en niños como herramienta de detección o screening de ambliopía, dado que es la principal causa de pérdida visual en la infancia. Por ello, resulta fundamental validar el modelo en población pediátrica en futuros estudios, con el fin de garantizar su aplicabilidad y eficacia en este contexto.

## 5. CONCLUSIONES

En este trabajo se ha propuesto un método para reconocer la ametropía de un paciente mediante inteligencia artificial aplicada en imágenes pupilares capturadas por una cámara convencional. Se ha entrenado un modelo para reconocer si el paciente es miope, hipermetrope o emетроpe; este modelo ha sido evaluado y se han extraído las siguientes conclusiones:

El modelo ha obtenido una exactitud del 69% sobre el conjunto de validación. Aunque la exactitud es notablemente inferior a la del conjunto de entrenamiento, sigue situándose en un rango cercano al umbral mínimo que suele considerarse aceptable para herramientas de apoyo al diagnóstico.

El análisis global de los resultados muestra una mayor precisión en pacientes miopes, en comparación con los emétropes o hipermétropes. Esto podría deberse a una mayor consistencia en el patrón del reflejo rojo asociado a esta ametropía, lo cual facilita su identificación por parte del modelo. Por el contrario, los errores más frecuentes se dan en las categorías de hipermetropía y emetropía.

En resumen, los resultados muestran un muy buen rendimiento del modelo sobre el conjunto de entrenamiento, lo que indica que presenta una buena capacidad de aprendizaje. Sin embargo, la menor exactitud sobre el conjunto de validación indica una capacidad de generalización aún limitada, destacando la necesidad de ampliar la base de datos. Esto permitiría aumentar la sensibilidad y la capacidad de discriminación del modelo entre categorías refractivas, mejorando así su utilidad como técnica de screening visual, especialmente en población pediátrica.

## 6. BIBLIOGRAFÍA

1. Martín R, Vecilla G. Ametropías y emetropización. En: Martín R, Vecilla G. Manual de Optometría; Madrid: Editorial Médica Panamericana; 2018: Pag. 63-69.
2. Han X, Liu C, Chen Y, He M. Myopia prediction: a systematic review. *Eye (Lond)*. 2022;36:921-929.
3. Baird PN, Saw SM, Lanca C, Guggenheim JA, Smith Iii EL, Zhou X, Matsui KO, Wu PC, Sankaridurg P, Chia A, Rosman M, Lamoureux EL, Man R, He M. Myopia. *Nat Rev Dis Primers*. 2020;6:99.
4. Cho P, Tan Q. Myopia and orthokeratology for myopia control. *Clin Exp Optom*. 2019;102:364-377.
5. Martín R, Vecilla G. Desarrollo visual. Ambliopía. Nistagmus. En: Martín R, Vecilla G. Manual de Optometría; Madrid: Editorial Médica Panamericana; 2018: Pag. 377-408.
6. Friendly DS. Amblyopia: definition, classification, diagnosis, and management considerations for pediatricians, family physicians, and general practitioners. *Pediatr Clin North Am*. 1987;34:1389-1401.
7. Webber AL, Wood J. Amblyopia: prevalence, natural history, functional effects and treatment. *Clin Exp Optom*. 2005;88:365-375.
8. Cecilia E, Negreira P, Herbanía Y, Rodríguez L. Rehabilitación visual en niños ambliopes. *Rev Cuba Oftalmol*. 2009;22.
9. Birch EE. Amblyopia and binocular vision. *Prog Retin Eye Res*. 2012;33:67-84.
10. Sánchez I, Ortiz S, Martín R, de Juan V. Advantages, limitations, and diagnostic accuracy of photoscreeners in early detection of amblyopia: a review. *Clin Ophthalmol*. 2016;10:1365-1373.
11. American Optometric Association. Care of the patient with amblyopia. <https://www.aoa.org/AOA/Documents/Practice%20Management/Clinical%20Guidelines/Consensus-based%20guidelines/Care%20of%20Patient%20with%20Amblyopia.pdf> (23 de enero 2025).
12. Park S. H. Current Management of Childhood Amblyopia. *Korean J Ophthalmol*. 2019;33:557-568.
13. Yan XR, Jiao WZ, Li ZW, Xu WW, Li FJ, Wang LH. Performance of the Plusoptix A09 photoscreener in detecting amblyopia risk factors in Chinese children attending an eye clinic. *PLoS One*. 2015;10.
14. Xu D, Ding S, Zheng T, Zhu X, Gu Z, Ye B, Fu W. Deep learning for predicting refractive error from multiple photorefractive images. *Biomed Eng Online*. 2022;21:55.
15. Chollet F. En: Chollet F. Deep learning with python; Shelter Island, NY: Manning Publications Company; 2017.
16. Lin H, Long E, Ding X, Diao H, Chen Z, Liu R, Huang J, Cai J, Xu S, Zhang X, Wang D, Chen K, Yu T, Wu D, Zhao X, Liu Z, Wu X, Jiang Y, Yang X, Cui D, Liu W, Zheng Y, Luo L, Wang H, Chan CC, Morgan IG, He M, Liu Y. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study. *PLoS Med*. 2018;15.
17. Malik S, Kanwal N, Asghar MN, Sadiq MAA, Karamat I, Fleury M. Data driven approach for eye disease classification with machine learning. *Appl Sci*. 2019;9.
18. Fageeri SO, Ahmed SMM, Almubarak SA, Mu'azu AA. Eye refractive error classification using machine learning techniques. En 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE). Khartoum, Sudan, 2017. pp.1-6.
19. Achiron A, Gur Z, Aviv U, Hilely A, Mimouni M, Karmona L, Rokach L, Kaiserman I. Predicting refractive surgery outcome: Machine learning approach with big data. *J Refract Surg*. 2017;33:592-597.