



---

**Universidad de Valladolid**

FACULTAD DE CIENCIAS

# Grado en Óptica y Optometría

MEMORIA TRABAJO DE FIN DE GRADO TITULADO

Detección de patología con inteligencia artificial  
aplicada en imágenes de retina II

Presentado por María Sáez López de Foronda

Tutelado por: Irene Sánchez y Roberto Román

Tipo de TFG: Investigación

En Valladolid a, 27 de mayo de 2025

## ÍNDICE

<b>RESUMEN .....</b>	<b>3</b>
<b>ABSTRACT.....</b>	<b>4</b>
<b>1. INTRODUCCIÓN .....</b>	<b>5</b>
<b>2. MATERIAL Y METODO.....</b>	<b>8</b>
<b>2.1 Procedimiento para la obtención de imágenes .....</b>	<b>8</b>
<b>2.2 Fuentes de obtención de las imágenes .....</b>	<b>9</b>
<b>2.3 Clasificación de las imágenes .....</b>	<b>10</b>
<b>2.4 Entrenamiento del modelo mediante Teachable Machine.....</b>	<b>13</b>
<b>2.5 Sensibilidad y Especificidad en la Clasificación de Imágenes ....</b>	<b>14</b>
<b>3. RESULTADOS .....</b>	<b>15</b>
<b>3.1 Visualización de resultados del modelo .....</b>	<b>15</b>
<b>3.2 Validación del modelo entrenado .....</b>	<b>16</b>
<b>3.3 Media ponderada de sensibilidad y especificidad .....</b>	<b>16</b>
<b>4. DISCUSIÓN .....</b>	<b>18</b>
<b>5. CONCLUSIONES .....</b>	<b>20</b>
<b>6. BIBLIOGRAFÍA .....</b>	<b>21</b>

## RESUMEN

**Objetivo.** El objetivo principal de este trabajo fue desarrollar un modelo de inteligencia artificial (IA), basado en redes neuronales, capaz de clasificar retinografías como imágenes de ojos sanos o patológicos. Esta herramienta pretende facilitar el cribado automatizado desde atención primaria y mejorar la derivación temprana de pacientes hacia atención especializada.

**Material y método.** Para entrenar el modelo, se recopilaron imágenes de fondo de ojo de acceso libre, procedentes de distintos repositorios digitales, clasificadas en dos categorías: sanos y patológicos (con diversas enfermedades retinianas). Se seleccionaron 180 imágenes para el entrenamiento (90 sanas y 90 patológicas) y 360 para la validación (91 sanas y 269 patológicas). El modelo se entrenó utilizando la herramienta Teachable Machine de Google, empleando los parámetros predeterminados. La evaluación se realizó mediante el análisis de sensibilidad, especificidad y media ponderada de acierto, calculada con el nivel de confianza de cada predicción.

**Resultados.** El modelo final obtuvo una sensibilidad del 100 % y una especificidad del 93,4 %, siendo capaz de detectar correctamente todas las imágenes patológicas, aunque clasificó erróneamente 6 ojos sanos como patológicos (falsos positivos). Además, se calculó una media ponderada de acierto del 99,49 %, lo que refleja que las predicciones correctas fueron realizadas con un alto grado de confianza. Estos resultados confirman la solidez del modelo en el conjunto de datos utilizado.

**Conclusiones.** El modelo desarrollado demuestra un rendimiento muy alto en la detección de alteraciones retinianas a partir de retinografías, siendo especialmente eficaz como herramienta de cribado. Su alta sensibilidad garantiza que no se pasen por alto casos patológicos, aunque conlleve un leve aumento de falsos positivos, lo cual es aceptable en un entorno clínico preventivo. Cabe destacar que las imágenes patológicas empleadas mostraban signos clínicos relativamente evidentes, lo que puede facilitar la clasificación; estudios previos indican que el rendimiento suele disminuir ante signos sutiles o fases iniciales. Aunque el modelo no identifica patologías concretas, su aplicación general lo hace útil en atención primaria. Para futuras mejoras, se propone ampliar el conjunto de datos con casos más complejos y probar su rendimiento en contextos clínicos reales.

## ABSTRACT

**Objective.** The main objective of this project was to develop an artificial intelligence (AI) model based on neural networks capable of classifying fundus images as either healthy or pathological. This tool aims to facilitate automated screening in primary care and improve early referral of patients to specialized care.

**Materials and Methods.** To train the model, publicly available fundus images were collected from various online repositories and classified into two categories: healthy and pathological (with various retinal diseases). A total of 180 images were selected for training (90 healthy and 90 pathological) and 360 for validation (91 healthy and 269 pathological). The model was trained using Google's Teachable Machine tool with default parameters. Evaluation was carried out by analyzing sensitivity, specificity, and weighted average accuracy, calculated based on the confidence level of each prediction.

**Results.** The final model achieved a sensitivity of 100% and a specificity of 93.4%, successfully detecting all pathological images, although it incorrectly classified 6 healthy eyes as pathological (false positives). In addition, a weighted average accuracy of 99.49% was calculated, indicating a high degree of confidence in the correct predictions. These results confirm the robustness of the model on the dataset used.

**Conclusions.** The developed model shows very high performance in detecting retinal abnormalities from fundus images, proving to be especially effective as a screening tool. Its high sensitivity ensures that no pathological cases are missed, although it results in a slight increase in false positives, which is acceptable in a preventive clinical setting. It is worth noting that the pathological images used presented relatively evident clinical signs, which may facilitate classification; previous studies suggest that performance tends to decrease when dealing with subtle signs or early disease stages. Although the model does not identify specific pathologies, its general application makes it useful in primary care. For future improvements, it is proposed to expand the dataset with more complex cases and to test its performance in real clinical contexts.

## 1. INTRODUCCIÓN

A nivel mundial, la pérdida de visión causada por distintas afecciones oculares es un problema de salud pública significativo, en especial cuando deriva en discapacidad visual o ceguera. La pérdida de agudeza visual (AV) es una de las manifestaciones clínicas más comunes en estas condiciones. Se estima que alrededor de 2200 millones de personas padecen problemas de visión y, en el 50% de los casos, la pérdida de visión podría haberse evitado. [1] La limitación visual depende de muchos factores, como la disponibilidad de intervenciones y tratamientos, y el acceso a la rehabilitación visual. En España, según las últimas estimaciones de la Encuesta Nacional de Salud, el 13,4% de la población mayor de 15 años padece algún tipo de limitación visual, de la cual el 11,6% presenta una limitación visual moderada y el 1,9% una pérdida visual grave o completa. [1]

Existen diversas enfermedades oculares que pueden provocar ceguera o pérdida de la visión y están asociadas a múltiples factores. Entre estos, el envejecimiento y el aumento de enfermedades crónicas como la diabetes mellitus (DM) son los principales factores de riesgo. En Europa, y particularmente en España, las principales enfermedades oftalmológicas que causan pérdida de visión y ceguera son el glaucoma, la retinopatía diabética (RD), el edema macular diabético (EMD), la degeneración macular asociada a la edad (DMAE) y la alta miopía (MH). [1]

En España, según los datos más recientes de la Organización Nacional de Ciegos Españoles (ONCE), la degeneración macular y otras maculopatías afectan al 11,2% de los afiliados, mientras que el glaucoma representa el 5,6%, y la RD el 4,2%. [2] La miopía magna es, con diferencia, la patología más común, afectando al 21,0% de los afiliados. La prevalencia de estas enfermedades se ve intensificada en personas mayores de 65 años, grupo que representa el 47,8% de los afiliados a la ONCE. Estos datos reflejan una clara relación entre estas enfermedades y la afectación de la retina, ya que la mayoría compromete directa o indirectamente esta estructura ocular. [2]

Muchas de las enfermedades que afectan a la retina pueden diagnosticarse mediante la observación del fondo de ojo, lo que resalta la importancia de la retinografía como herramienta diagnóstica. Dado que algunas de estas patologías son asintomáticas en fases iniciales, como el glaucoma o la RD, la detección precoz es clave para evitar la progresión a una ceguera irreversible. [1]

La retinografía es una técnica de imagen no invasiva que permite capturar imágenes del fondo de ojo para detectar alteraciones. La retinografía no midriática destaca por no requerir midriasis, facilitando evaluaciones rápidas y cómodas para el paciente. Además, actualmente muchos retinógrafos son prácticamente automáticos, por lo que la realización de la prueba se puede realizar sin grandes conocimientos técnicos en la mayoría de los casos. Entre los dispositivos más avanzados, algunos permiten capturar imágenes con pupilas de hasta 1,5 mm de diámetro (como el Easyscan de Iop-tics, desarrollado por I-optics, Países Bajos [4]), o alcanzar un campo de visión de hasta 200° utilizando

la tecnología *Scanning Laser Ophthalmoscopy* (SLO), como el Optomap de OPTOS PCL. [3] Tradicionalmente, para facilitar la observación del fondo de ojo, especialmente en retina periférica, se recomienda la dilatación pupilar. [3] A pesar de estas limitaciones, la retinografía no midriática sigue siendo una herramienta muy útil para la detección precoz y el seguimiento de enfermedades que afectan a estructuras clave como la mácula y el nervio óptico, ya que estas se localizan en la zona central de la retina, que es precisamente la región que capturan este tipo de dispositivos.

En atención primaria, aunque la realización de retinografías se ha facilitado bastante gracias a los avances en la automatización de los dispositivos, sigue siendo necesario disponer de personal especializado y correctamente entrenado para su interpretación. Diversas investigaciones han demostrado que la inteligencia artificial (IA) puede ser una herramienta eficaz para asistir en la detección, seguimiento y análisis de enfermedades retinianas comunes, como la degeneración macular asociada a la edad y la retinopatía diabética, mediante la interpretación rápida y cuantitativa de imágenes obtenidas con la retinografía. Estas aplicaciones pueden facilitar el cribado de pacientes, predecir la progresión de la enfermedad e incluso evaluar la respuesta al tratamiento, ofreciendo así un apoyo diagnóstico relevante en contextos donde el acceso a especialistas es limitado. [5]

En el campo de la retinopatía diabética (RD), ya se han desarrollado múltiples trabajos que demuestran que la inteligencia artificial (IA) aplicada al análisis de retinografías es una herramienta eficaz y validada para su detección. Se encontró una revisión sistemática con metaanálisis que analizó el rendimiento de algoritmos de aprendizaje profundo (*deep learning*) en la detección de RD, concluyendo que estos modelos presentan una alta sensibilidad y especificidad, lo que podría contribuir a reducir los errores diagnósticos y mejorar la eficiencia en el cribado. [6] De la misma forma, se localizó un metaanálisis que examinó la precisión diagnóstica de diversos algoritmos de aprendizaje automático (*machine learning*), recalcando su buen rendimiento, pero también señalando limitaciones metodológicas, como la falta de validación externa o el uso de muestras poco representativas, que pueden comprometer la precisión del modelo en situaciones clínicas reales. [7] Por último, se halló otra revisión sistemática que también confirmó la eficacia de los modelos de IA para detectar RD en imágenes de fondo de ojo, recomendando su implementación clínica por su capacidad para ahorrar tiempo, reducir costes y facilitar la detección temprana en entornos clínicos. [8] Además, universidades como la de Valladolid han desarrollado modelos basados en aprendizaje profundo específicamente orientados a esta enfermedad, con resultados prometedores en entornos clínicos reales. [9,10]

Estos antecedentes muestran que la aplicación de IA en la detección de retinopatía diabética no es solo teórica, sino que ya está siendo implementada con éxito en distintos contextos. Además, diversos estudios han explorado la utilidad del aprendizaje profundo en otras patologías retinianas. Por ejemplo, se han desarrollado sistemas capaces de predecir la aparición y progresión del glaucoma utilizando fotografías del fondo de ojo con muy buenos resultados en sensibilidad y especificidad. [11,12] Asimismo, en el caso de la DMAE, los modelos de IA han demostrado eficacia en la identificación de biomarcadores de progresión, lo que permite una derivación más rápida a tratamiento y una mejor gestión clínica. [13]

Sin embargo, la mayoría de estos estudios se han centrado en modelos entrenados para una única patología y en condiciones controladas. En cambio, este trabajo propone un modelo más general que clasifica las retinografías en dos grandes grupos: ojos sanos y ojos con alguna alteración, sin necesidad de especificar la patología exacta. Esta estrategia busca facilitar el cribado inicial desde atención primaria, permitiendo detectar casos sospechosos de forma automatizada.

Todo esto motiva el objetivo principal de este TFG, el cual es desarrollar un algoritmo de IA (red neuronal) capaz de clasificar retinografías en ojos sanos o patológicos para facilitar la remisión temprana desde atención primaria.

## 2. MATERIAL Y METODO

### 2.1 Procedimiento para la obtención de imágenes

En este trabajo, la técnica empleada para la adquisición de imágenes ha sido la retinografía no midriática, una herramienta cada vez más extendida en entornos clínicos gracias a los avances tecnológicos recientes, que han mejorado tanto su facilidad de uso como la calidad de imagen obtenida. [14]

Esta técnica permite obtener fotografías del fondo de ojo sin necesidad de dilatar la pupila farmacológicamente, facilitando su uso especialmente en atención primaria o en pacientes sensibles a la midriasis. El paciente se sienta frente a la cámara en una sala con iluminación ambiental atenuada, dirige la mirada hacia una luz de fijación, y el dispositivo utiliza luz infrarroja para enfocar la zona central de la retina, donde se encuentran la mácula y el nervio óptico. La luz infrarroja no provoca constricción pupilar, lo cual permite enfocar sin alterar el tamaño pupilar. Una vez enfocado, se utiliza un flash de baja intensidad para capturar la imagen, reduciendo el riesgo de constricción pupilar prolongada y permitiendo capturas consecutivas con un intervalo recomendado de 30 a 60 segundos. [14]

Además, las cámaras digitales actuales incorporan mecanismos de alineación y enfoque automático, interfaz de usuario simplificada y opciones para exportar las imágenes a registros médicos electrónicos o redes clínicas, lo que ha contribuido a su adopción creciente en la práctica sanitaria diaria. [14]

El principio de funcionamiento de esta cámara de fondo de ojo, también conocida como retinógrafo, está representado esquemáticamente en la Figura 1. Para obtener las mencionadas fotografías del fondo del ojo, se dispone de dos fuentes de luz: L1 para fotografías con flash y L2 para observación, donde un espejo abatible (E1) selecciona una u otra. Los condensadores C1 proyectan el haz de luz sobre F, donde se puede situar un filtro coloreado si se desea, controlando la iluminación con un diafragma. El elemento C2 proyecta, por medio del espejo E2, el haz de iluminación sobre el centro plateado de E3, de tal manera que la luz forma un círculo de difusión en la retina del sujeto, por medio del objetivo (Ob). La luz difundida por la retina es recogida por Ob y proyectada mediante el sistema P sobre la placa fotográfica o sobre el plano focal del ocular (Oc), eligiendo una u otra según la posición del espejo E4, que está conectado mecánicamente a E1. Es de notar que el espejo E3 solo tiene espejada su parte central, por lo que el haz de iluminación (líneas de trazos) se refleja en él, pero el haz de observación (líneas continuas) pasa en su mayor parte. El elemento B es proyectado por parte del sistema C2 sobre el objetivo, siendo visto por el sujeto como un punto luminoso que sirve como blanco de fijación (B es una varilla metálica con una bola en su punta). Variando la posición de B, se obliga al sujeto a girar su ojo para seguirlo, presentando distintas partes de la retina a observación. [15]

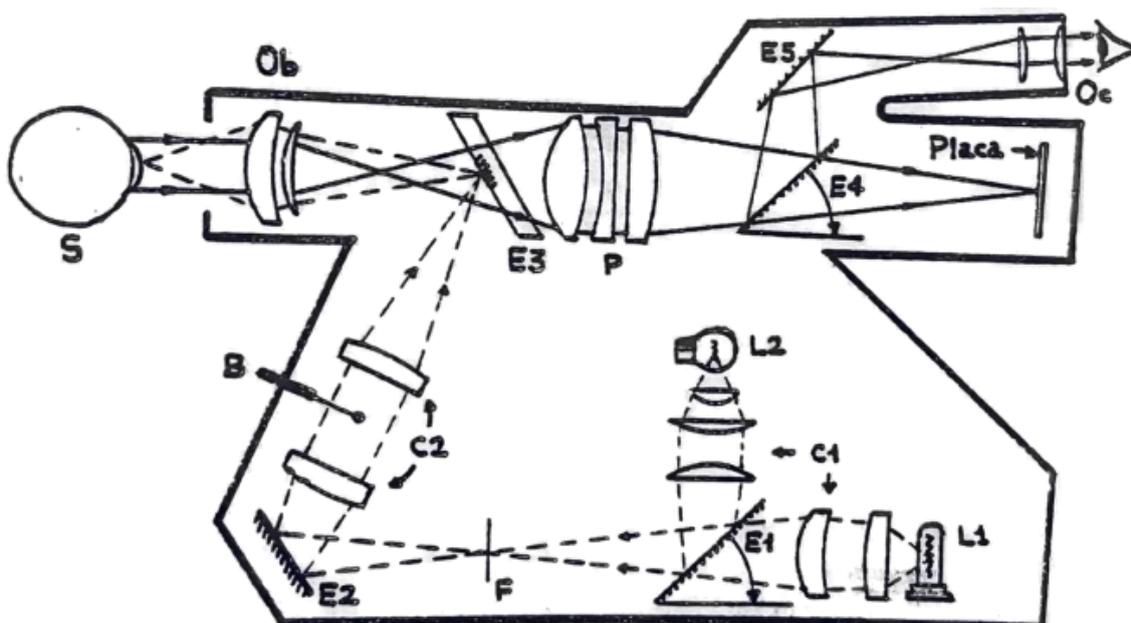


Figura 1. Representación esquemática de la cámara de fondo de ojo. Imagen extraída de: Marcén Grasa J, Antó Roca J. Óptica instrumental. En: Servei Publicacions; 1986. Capítulo 24, apartado 6, subapartado 3, p. 285.[15]

## 2.2 Fuentes de obtención de las imágenes

Para el desarrollo de este trabajo, se realizó una búsqueda preliminar de repositorios de imágenes de fondo de ojo de acceso libre, asegurando que las imágenes utilizadas contaran con la licencia adecuada para su uso y garantizando su calidad. Se identificaron inicialmente ocho repositorios, de los cuales se seleccionaron cuatro principales en función del número de imágenes disponibles y su relevancia para este trabajo.

A continuación, se describen los repositorios usados:

- **Dataset from fundus images for the study of diabetic retinopathy:** Se utilizó la carpeta subclasificada como "No DR signs", que contenía 711 imágenes de ojos sanos. Se realizó una revisión por parte de la tutora de este trabajo para asegurar la calidad y la ausencia de patologías. Finalmente se seleccionaron 181 imágenes para su uso en este estudio. [16]

[16]

<https://zenodo.org/records/4647952>

- **STARE Database:** Este repositorio cuenta con 402 imágenes clasificadas en 13 patologías distintas, las cuales se recogen en la Tabla 1. La clasificación de las imágenes se realizó utilizando la tabla de diagnósticos proporcionada por el repositorio. Las imágenes están clasificadas por patologías; algunas imágenes cuentan con más de una patología. [17,18]

<https://cecas.clemson.edu/~ahoover/stare/>

- **DR HAGIS Database:** Este repositorio contiene 39 imágenes de alta resolución obtenidas de un programa de cribado de retinopatía diabética

en el Reino Unido. Se utilizaron 10 imágenes ya clasificadas con la enfermedad de Degeneración Macular Asociada a la Edad (DMAE). [19]

<https://personalpages.manchester.ac.uk/staff/niall.p.mcloughlin/>

- **High-Resolution Fundus (HRF) Image Database:** Este repositorio proporciona imágenes de alta resolución clasificadas en tres categorías: ojos sanos, pacientes con retinopatía diabética y pacientes con glaucoma. Se seleccionaron 25 imágenes correspondientes a la categoría de glaucoma. [20]

<https://www5.cs.fau.de/research/data/fundus-images/>

Todas las imágenes utilizadas en este estudio son de acceso libre y su selección se realizó meticulosamente para garantizar su calidad.

### 2.3 Clasificación de las imágenes

Inicialmente, se consideró la posibilidad de obtener un modelo de inteligencia artificial que identificara la patología específica presente en cada imagen. Sin embargo, debido a la variabilidad en el número de imágenes disponibles de cada patología, se optó por una clasificación binaria: ojos sanos y ojos con alguna patología.

Con este fin, se separaron todas las imágenes de ojos sanos seleccionadas previamente en dos carpetas: "SANOS" y "PATOLOGICOS". La carpeta "PATOLOGICOS" incluía imágenes de las 15 patologías identificadas en la Tabla 1. En la Tabla 1 se mantuvieron las abreviaturas descritas por la base de datos *STARE Database*. [17]

Para mostrar visualmente las diferentes imágenes de las categorías utilizadas en este trabajo, se ha creado la Figura 2, que recoge una imagen representativa por cada una de las 16 clases de fondo de ojo. La imagen etiquetada con el número 0 corresponde a un ojo sano, mientras que las imágenes numeradas del 1 al 15 representan casos con diferentes patologías.

Dado que la patología con menor número de imágenes disponibles era la Oclusión de la Arteria Retiniana en Rama (BRAO), con 7 imágenes (ver Tabla 1), se utilizó este número como referencia para equilibrar el conjunto de datos. De esta categoría se aplicó un criterio de división del 90% para entrenamiento y el 10% para validación, resultando en 6 imágenes para entrenamiento. Para un entrenamiento equilibrado del modelo, se seleccionó el mismo número de imágenes de cada patología en el entrenamiento. Por tanto, se seleccionaron 6 imágenes de cada patología para entrenar el modelo, dando un total de 90 imágenes. La selección de las imágenes se realizó de forma aleatoria utilizando la función "random" de la librería "numpy" del lenguaje "Python". En cuanto a las imágenes del conjunto sanos, se seleccionaron de forma aleatoria 90 imágenes para balancear el entrenamiento y realizarlo con el mismo número de imágenes sanas como patológicas. Las imágenes no seleccionadas para el entrenamiento se reservaron para la validación del modelo.

Finalmente, se estructuró el conjunto de datos en dos carpetas principales: "Train" y "Test". Cada una de estas carpetas contenía dos subcarpetas: "SANOS" y "PATOLOGICOS". La distribución de las imágenes fue la siguiente:

- **Train:**
  - SANOS: 90 imágenes
  - PATOLOGICOS: 90 imágenes
- **Test:**
  - SANOS: 91 imágenes
  - PATOLOGICOS: 269 imágenes

Número de Diagnóstico	Diagnóstico	Abreviatura	Número de imágenes disponibles
1	Émbolos de Hollenhorst	Émbolos	13
2	Oclusión de la Arteria Retiniana en Rama	BRAO	7
3	Oclusión de la Arteria Cilio-Retiniana	CRAO	9
4	Oclusión de la Vena Retiniana en Rama	BRVO	11
5	Oclusión de la Vena Central de la Retina	CRVO	25
6	Oclusión de la Hemi-Vena Central de la Retina	Hemi-CRVO	10
7	Retinopatía Diabética No Proliferativa	BDR/NPDR	70
8	Retinopatía Diabética Proliferativa	PDR	26
9	Retinopatía Arterioesclerótica	ASR	35
10	Retinopatía Hipertensiva	HTR	37
11	Enfermedad de Coat's		13
12	Macroaneurisma		8
13	Neovascularización Coroidea	CNV	61
14	Degeneración Macular Asociada a la Edad	DMAE	13
15	Glaucoma		25

Tabla 1. Clasificación de patologías utilizadas en el estudio

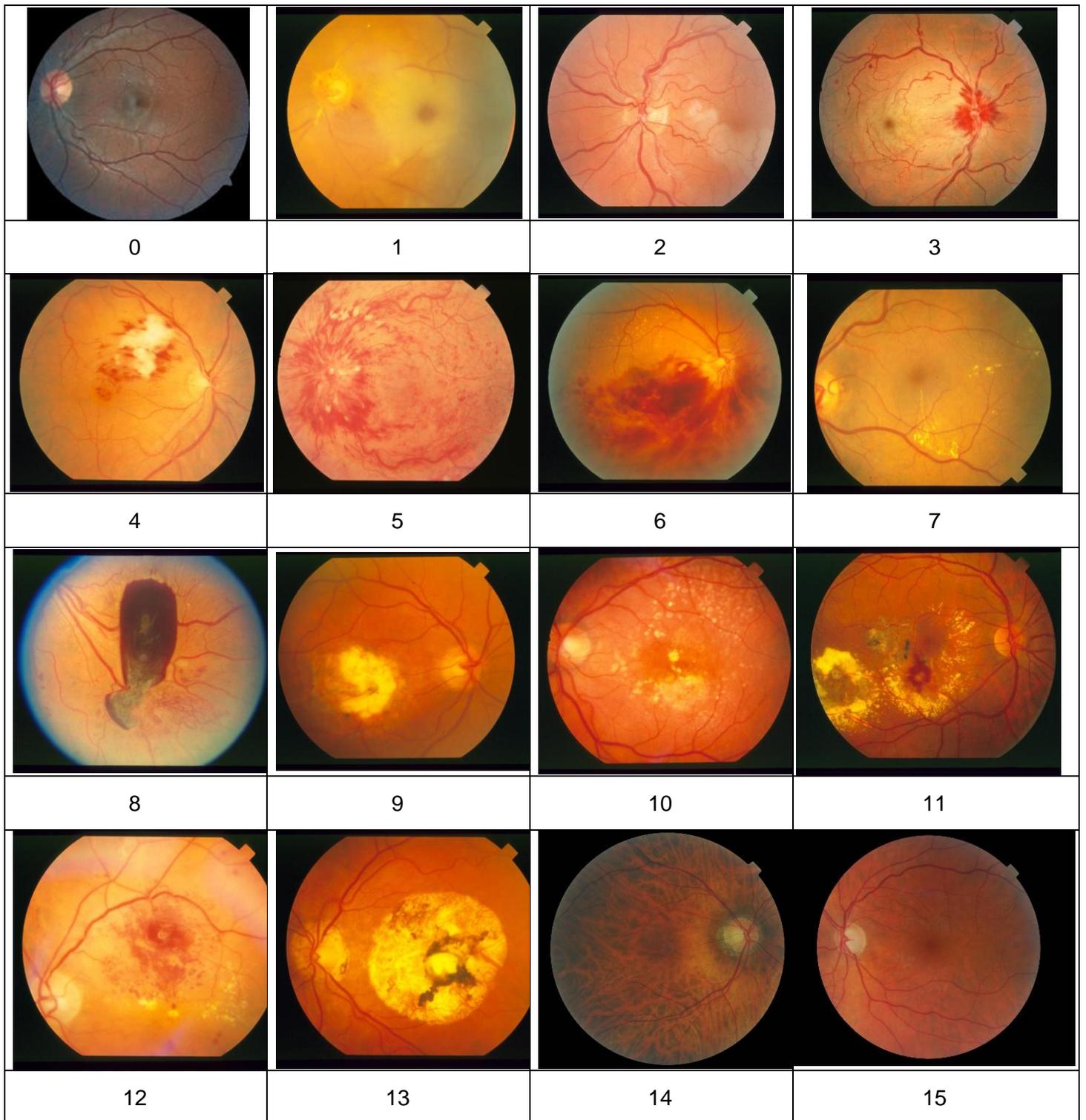


Figura 2. Muestra las imágenes representativas de las 16 clases de fondo de ojo, donde la imagen etiquetada con el número 0 corresponde a un ojo sano, y las imágenes del 1 al 15 representan distintos casos patológicos, clasificados numéricamente en la Tabla 1.

## 2.4 Entrenamiento del modelo mediante Teachable Machine

En este trabajo se usó como referencia un trabajo de fin de grado realizado en la Universidad de Valladolid por uno de los alumnos del grado en Física. [21] Para la detección de imágenes patológicas frente a imágenes de ojos sanos, se entrenó un modelo de inteligencia artificial utilizando la herramienta en línea Teachable Machine, desarrollada por Google. Esta plataforma permite construir modelos basados en IA de clasificación de imágenes de forma accesible y sin necesidad de conocimientos avanzados en programación. El acceso es gratuito y está disponible en el siguiente enlace: <https://teachablemachine.withgoogle.com/train/image>.

Una vez organizadas las imágenes en dos categorías principales (ojos sanos y ojos con patologías), se entrenó el modelo con la herramienta mencionada. A cada clase se le asignaron las imágenes correspondientes, respetando la división previamente establecida: 90 imágenes de cada clase para el entrenamiento.

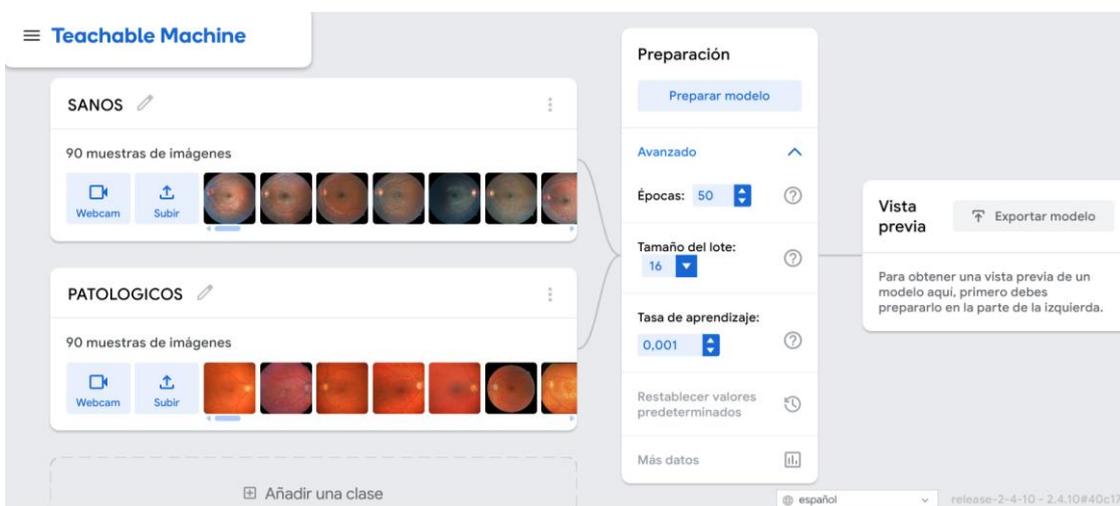


Figura 3. Captura de pantalla del proceso de creación del modelo en Teachable Machine.

En la Figura 3 se muestra la interfaz de entrenamiento del modelo, donde se aprecia que se ha utilizado un total de 90 imágenes correspondientes a ojos sanos y otras 90 imágenes de ojos patológicos. Además, pueden observarse los parámetros predeterminados utilizados durante el entrenamiento, como el número de épocas, el tamaño del lote y la tasa de aprendizaje.

Durante el proceso de entrenamiento, se utilizaron los parámetros por defecto recomendados por la plataforma:

- **Número de épocas:** Se estableció en 50. Este parámetro indica la cantidad de veces que el modelo es reentrenado de forma iterativa con todas las imágenes de entrenamiento. Un número adecuado de épocas permite que el modelo aprenda patrones significativos sin memorizar los datos (evitando el sobreajuste). [22]
- **Tamaño del lote:** Se fijó en 16. Este parámetro define cuántas imágenes se procesan por cada iteración, es decir, cuántas imágenes se introducen en el modelo de una sola vez para que aprenda a partir de ellas antes de pasar al siguiente grupo. En este caso, las 180 imágenes se dividen en

lotes de 16 imágenes. Se mantuvo el valor predeterminado al no tener una influencia crítica en los resultados para un problema de clasificación binaria como el presente. [22]

- **Tasa de aprendizaje:** Se utilizó el valor estándar de 0.001. Esta tasa determina la magnitud de los ajustes que realiza el modelo en cada iteración durante el proceso de entrenamiento. Si es demasiado baja, el aprendizaje se vuelve lento; si es demasiado alta, el modelo puede no converger correctamente. [22]

Al finalizar el entrenamiento, la herramienta Teachable Machine devolvió los coeficientes del modelo entrenado en un fichero de extensión “.h5”, junto a unas instrucciones detalladas para poder ejecutar el modelo en lenguaje “Python” con la librería “Keras/Tensorflow”. Para evaluar el rendimiento del modelo entrenado, se ejecutó sobre todas las imágenes disponibles, tanto de entrenamiento como de validación (Test). Para cada imagen el modelo devolvió una predicción (Sano o Patológico) y un valor de la confianza, en tanto por ciento, con la que estima el modelo esa predicción.

Al exportar el modelo, se obtuvo un archivo que permite integrarlo y ejecutarlo externamente. Para su implementación y ejecución se utilizó el entorno de programación Anaconda, desde donde se cargó el modelo y se automatizó la evaluación de las imágenes mediante Python 3.10.

## 2.5 Sensibilidad y Especificidad en la Clasificación de Imágenes

Para validar el modelo entrenado se van a usar dos conceptos: sensibilidad y especificidad. Estos dos parámetros son cruciales para evaluar el rendimiento del modelo entrenado de inteligencia artificial y su capacidad para detectar si el ojo es sano o patológico. [23]

- **Sensibilidad:** Este término se refiere a la capacidad del modelo para identificar correctamente a los pacientes que tienen la enfermedad. Es decir, es la proporción de casos verdaderos positivos, en otras palabras, aquellos ojos enfermos que el modelo es capaz de detectar como patológicos. Tener una alta sensibilidad es fundamental en este tipo de modelos porque lo que realmente interesa es no pasar por alto ningún ojo enfermo. Si el modelo es muy sensible, se minimiza el riesgo de no remitir a pacientes que necesita atención especializada. Por ejemplo, si un paciente tiene una patología ocular grave y el modelo no la detecta (falso negativo), este paciente podría quedar sin tratamiento adecuado, lo que podría empeorar su condición. Por lo tanto, es preferible equivocarse en la detección de patologías (aunque esto signifique tener algunos falsos positivos) que dejar a una persona sin el diagnóstico necesario. [23]
- **Especificidad:** Este concepto hace referencia a la capacidad del modelo para identificar correctamente a los pacientes que no tienen la enfermedad; es decir, los ojos sanos. La especificidad es importante porque ayuda a evitar que pacientes sin ninguna patología sean etiquetados erróneamente como enfermos (falsos positivos). Sin embargo, en el contexto de la atención primaria, es más aceptable que la especificidad sea algo más baja, ya que lo primordial es no dejar escapar a aquellos pacientes con enfermedades oculares. [23]

### 3. RESULTADOS

Una vez finalizado el entrenamiento del modelo mediante Teachable Machine, se procedió a su validación utilizando las imágenes reservadas en el conjunto de prueba ("Test").

#### 3.1 Visualización de resultados del modelo

En este apartado se muestra un ejemplo visual del funcionamiento del modelo de IA aplicado a las imágenes de fondo de ojo. El modelo da una predicción (clase: patológico o sano) junto con un índice de confianza, que representa el nivel de confianza asociado a esa predicción. Además, se compara dicha predicción con la clase real previamente asignada, para comprobar si el resultado ha sido correcto.

La Tabla 3 muestra dos ejemplos concretos:

- En el primer caso, el modelo clasifica correctamente la imagen 10\_im0007.jpg como patológica, coincidiendo con la clase real. El índice de confianza es muy alto (99,99%), y el modelo acierta (valor 1 en la columna final; ver Figura 4).
- En el segundo caso, la imagen 87.jpg es sana según la clasificación real, pero el modelo la predice como patológica con un índice de confianza del 52,70%. Como la predicción es incorrecta, el valor final es 0(ver Figura 5).

Estos ejemplos ayudan a entender cómo interpreta el modelo las imágenes en función de su grado de certeza y su precisión diagnóstica.

NOMBRE_IMAGEN	CLASE_REAL	PREDICCIÓN	CONFIANZA (%)	¿ACIERTO?
10_im0007.jpg	Patológicos	Patológicos	99,99	1
87.jpg	Sanos	Patológicos	52,70	0

Tabla 2. Ejemplos de predicciones del modelo con índice de confianza



Figura 4. Imagen 10\_im0007.jpg: caso patológico correctamente clasificado por el modelo.



Figura 5. Imagen 87.jpg: caso sano clasificado erróneamente como patológico.

### 3.2 Validación del modelo entrenado

Se entrenó un primer modelo que logró clasificar correctamente el 100% de las imágenes del conjunto de prueba (Test), obteniendo una sensibilidad y especificidad del 100%, sin falsos positivos ni falsos negativos. Debido a la perfección de los resultados, se decidió repetir el entrenamiento para comprobar la robustez del modelo. En este segundo entrenamiento, se volvieron a utilizar las 270 imágenes patológicas y 91 imágenes sanas reservadas en el conjunto de Test. El modelo clasificó correctamente las 270 imágenes patológicas, pero cometió el error de clasificar 6 imágenes sanas como patológicas. Los siguientes resultados fueron los obtenidos con el modelo final, que es el que se ha utilizado en este trabajo. Estos resultados quedan resumidos en la matriz de confusión generada que se presenta en la Tabla 2.

	Predichos SANOS	Predichos PATOLOGICOS
Reales SANOS	85	6
Reales PATOLOGICOS	0	270

Tabla 3. Matriz de confusión del Modelo 2

A partir de esta matriz, se obtuvieron los siguientes parámetros clave para el modelo:

- Sensibilidad: 1.00 (100%)
- Especificidad: 0.934 (93.4%)

El valor de especificidad fue calcula con la siguiente ecuación (1):

$$Especificidad = \frac{Verdaderos\ Negarivos(VN)}{VN + Falsos\ Positivos(FP)} \quad (1)$$

Siendo en este caso:

- **Verdaderos Negativos (VN)** = 85 (imágenes sanas correctamente clasificadas como sanas)
- **Falsos Positivos (FP)** = 6 (imágenes sanas clasificadas erróneamente como patológicas)

Con la ecuación (1) se obtuvo el valor de especificidad del 93,4%.

### 3.3 Media ponderada de sensibilidad y especificidad

Para evaluar con mayor precisión el rendimiento global del modelo, se calculó la media ponderada de acierto. Este cálculo tiene en cuenta que la confianza de las predicciones del modelo varía, por lo que las predicciones con menor confianza deberían tener menor peso en el resultado final.

La media ponderada no solo considera si el modelo clasificó correctamente cada imagen (con un valor de 1 si acertó y 0 si falló), sino también el nivel de confianza que el modelo asignó a cada predicción, expresado como un número entre 0 y 1. Esto significa que los aciertos con alta confianza tienen más peso que aquellos con menor confianza.

La ecuación (2) utilizada para calcular la media ponderada es la siguiente:

$$\text{Media ponderada} = \frac{\sum x_i \cdot \omega_i}{\sum \omega_i} \times 100\% \quad (2)$$

Donde:

- $x_i$ , es 1 si el modelo acertó en la imagen, y 0 si falló.
- $w_i$ , es la confianza que el modelo tiene en esa predicción, expresada en decimal (por ejemplo, 0.99 para 99%).

El resultado fue una media ponderada de acierto de aproximadamente 99,49%.

Este valor indica que el modelo no solo acertó en la mayoría de los casos, sino que además lo hizo con una alta confianza media, lo que confirma la fiabilidad del sistema para la clasificación de imágenes en este conjunto de datos.

## 4. DISCUSIÓN

Durante el desarrollo de este trabajo, se entrenaron dos modelos de inteligencia artificial utilizando la herramienta *Teachable Machine*. El primero de ellos ofreció resultados perfectos: clasificó correctamente todas las imágenes del conjunto de prueba, sin errores. Aunque estos resultados fueron sorprendentes en un primer momento, también generaron dudas sobre su fiabilidad real. En el contexto de modelos de aprendizaje automático, un rendimiento del 100 % suele indicar posibles problemas de sobreajuste o falta de representatividad en los datos utilizados para la validación.

Por este motivo, se procedió a realizar un segundo entrenamiento del modelo, para comprobar si estos resultados eran reproducibles. En esta segunda ejecución, el modelo mostró un rendimiento más realista: obtuvo una sensibilidad del 100 % y una especificidad del 93,4 %. Es decir, detectó todas las imágenes patológicas sin clasificar erróneamente ningún ojo enfermo como sano, aunque cometió errores al identificar como patológicos algunos ojos sanos (falsos positivos). Desde una perspectiva clínica, este comportamiento es aceptable y deseable en entornos de cribado, donde se prioriza la sensibilidad para evitar pasar por alto casos que podrían requerir atención urgente o especializada.

No obstante, cabe señalar que tanto las imágenes patológicas como las sanas utilizadas para entrenar y validar el modelo eran, en su mayoría, casos con signos clínicos evidentes. Es decir, las imágenes patológicas presentaban signos evidentes de enfermedad y las sanas estaban completamente libres de anomalías. Esta selección de las imágenes puede haber facilitado la tarea del modelo y, en consecuencia, haber favorecido a los buenos resultados obtenidos. Estudios previos han señalado que el rendimiento de los modelos de aprendizaje profundo tiende a disminuir cuando se aplican a imágenes con signos sutiles, fases incipientes o en condiciones dudosas. [24]

Para poner en contexto el rendimiento del modelo desarrollado en este trabajo, es útil compararlo con otras investigaciones previas. Por ejemplo, el modelo de Abramoff et al. (2019), basado en imágenes de retinopatía diabética (RD), obtuvo una sensibilidad del 96,8 % y una especificidad del 87 % en el conjunto Messidor-2. [24] Otros estudios, como el de de Gulshan et al. (2019) (Google), obtuvieron sensibilidades y especificidades de entre 90–97 % en diversos conjuntos de datos públicos de imágenes retinianas. [24] Aunque estos trabajos se enfocan en la detección específica de RD, permiten establecer un marco comparativo. En este sentido, el modelo desarrollado en este TFG ofrece una sensibilidad equiparable o superior, y una especificidad también elevada, aunque aplicada a una clasificación binaria general (ojo sano vs. patológico) y no a una patología concreta.

Además, estudios más recientes aplicados a otras enfermedades, como la degeneración macular asociada a la edad (DMAE) o el glaucoma, han demostrado también la eficacia del aprendizaje profundo en tareas de clasificación, seguimiento o predicción del riesgo. Por ejemplo, Borrelli et al. (2024) demostraron que los modelos de IA son capaces de analizar imágenes de OCT y detectar biomarcadores vinculados a un mayor riesgo de progresión de DMAE neovascular. [13] Por otro lado, Li et al. (2022) desarrollaron un sistema

de IA que predice la aparición y progresión del glaucoma con un área bajo la curva de hasta 0,91 en conjuntos de validación externos, [11] mientras que Lee et al. (2021) confirmaron que los cambios longitudinales en las predicciones de IA basadas en fotografías del fondo de ojo podían anticipar el desarrollo del glaucoma. [12]

Aunque el modelo del presente estudio no identifica patologías específicas, su función como herramienta de cribado general sigue siendo válida y útil, sobre todo si se considera su posible implementación en atención primaria. La presencia de falsos positivos no representa un riesgo clínico y, de hecho, puede considerarse una precaución aceptable si se garantiza que todos los casos de enfermedad real se van a remitir.

En el futuro, sería conveniente ampliar el entrenamiento y validación del modelo con conjuntos más grandes de datos más heterogéneos, que incluyan imágenes con alteraciones incipientes o dudosas, así como realizar una experiencia piloto en la práctica real que permita comparar el rendimiento del sistema con el de especialistas clínicos. De la misma manera, se podría probar usar redes neuronales más avanzadas y aprovechar modelos ya entrenados para mejorar la capacidad del sistema de funcionar bien en situaciones clínicas más difíciles. Además, esta herramienta podría automatizarse para que, cuando una retinografía sea clasificada como patológica, se envíe un aviso automático a los especialistas para su revisión.

## 5. CONCLUSIONES

En este trabajo se ha desarrollado un modelo de inteligencia artificial capaz de distinguir entre imágenes de fondo de ojo sanas y patológicas con una alta precisión. El modelo fue validado con imágenes no utilizadas en el entrenamiento, obteniendo una sensibilidad del 100 % y una especificidad del 93,4 %. Esto indica que el modelo fue capaz de detectar todas las imágenes patológicas, sin clasificar ningún ojo enfermo como sano, lo cual es especialmente relevante en el contexto clínico, donde lo prioritario es no pasar por alto ningún caso que requiera atención médica. El modelo cometió errores solo al clasificar algunos ojos sanos como patológicos, lo que supone falsos positivos, pero no implica riesgo para el paciente.

Es importante señalar que la mayoría de las imágenes patológicas empleadas en este estudio presentaban signos clínicos muy evidentes de enfermedad, lo que facilita la tarea de clasificación del modelo. En cambio, apenas se incluyeron casos con signos sutiles o dudosos, lo que podría explicar en parte el alto rendimiento observado. De cara a futuras aplicaciones, sería interesante evaluar el modelo con imágenes que presenten alteraciones leves o incipientes, ya que son estas las más complejas de clasificar incluso para profesionales.

Aunque el modelo no identifica la patología concreta, cumple satisfactoriamente su objetivo como herramienta de cribado inicial.

## 6. BIBLIOGRAFÍA

1. Pablo L, Garay-Aramburu G, García Layana A, Fernandez A, Vázquez I, Acebes X, et al. Assessing the economic burden of vision loss and irreversible legal blindness in Spain (2021–2030): a societal perspective. *Health Econ Rev.* 2024;14:70.
2. Organización Nacional de Ciegos Españoles (ONCE). Informe de datos estadísticos sobre afiliados y patologías visuales. Madrid: ONCE; 2023.
3. Martín Herranz R, Vecilla Antolínez G. Exploración del fondo de ojo. En: Martín Herranz R, editor. *Manual de Optometría.* Madrid: Elsevier; 2015:295.
4. i-Optics BV. EasyScan Retinal Imaging System. <https://medicolle.com/easyscan> (19 de mayo de 2025).
5. Lim JI, Rachitskaya AV, Hallak JA, Gholami S, Alam MN. Artificial intelligence for retinal diseases. *Asia Pac J Ophthalmol (Phila).* 2024;13:100096.
6. Islam MM, Yang HC, Poly TN, Jian WS, Jack Li YC. Aprendizaje profundo algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Comput Methods Programs Biomed.* 2020;191:105320.
7. Wu JH, Liu TYA, Hsu WT, Ho JH, Lee CC. Performance and limitation of aprendizaje automático algorithms for diabetic retinopathy screening: Meta-analysis. *J Med Internet Res.* 2021;23:e23863.
8. Dailidė G, Lagunavičius S, Balčiūnienė VJ. Efficacy of artificial intelligence in detecting diabetic retinopathy from retinal fundus images — a systematic review. *Ophthalmol J.* 2022;7:144–151.
9. Sánchez Cid P. Diagnóstico de la retinopatía diabética mediante técnicas de aprendizaje profundo [TFG]. Valladolid: Universidad de Valladolid; 2022.
10. Hernández Martínez P, Pérez Santamaría F. Uso en la práctica clínica de un método automático de detección de retinopatía diabética basado en inteligencia artificial. *Arch Soc Esp Oftalmol.* 2020;95:10:499–506.
11. Li F, Su Y, Lin F, et al. A deep-learning system predicts glaucoma incidence and progression using retinal photographs. *J Clin Invest.* 2022;132:e157968.
12. Lee T, Jammal AA, Mariottoni EB, Medeiros FA. Predicting glaucoma development with longitudinal aprendizaje profundo predictions from fundus photographs. *Am J Ophthalmol.* 2021;225:86–94.
13. Borrelli E, Serafino S, Ricardi F, et al. Aprendizaje profundo in neovascular age-related macular degeneration. *Medicina (Kaunas).* 2024;60:990
14. Mackay DD, Bruce BB. Non-mydríatic fundus photography: a practical review for the neurologist. *Pract Neurol.* 2016;16:343–351.
15. Marcen Grasa J, Antó Roca J. Óptica instrumental. En: *Servei Publicacions. Esquiascopia y oftalmoscopia.* Cap 24, apartado 3: Oftalmoscopia, p. 275; apartado 6: Oftalmoscopios tipo, subapartado 3: Cámara de fondo de ojo, p. 285. 1986.

16. Castillo Benítez VE, Castro Matto I, Mello Román JC, Vázquez Noguera JL, García-Torres M, Ayala J, et al. Dataset from fundus images for the study of diabetic retinopathy. *Data Brief*. 2021;36:107068.
17. Hoover A, Goldbaum M. The STARE Project: Structured Analysis of the Retina. <https://cecas.clemson.edu/~ahoover/stare/> (15 de abril de 2025).
18. STARE - Structured Analysis of the Retina. Diagnoses. <https://cecas.clemson.edu/~ahoover/stare/diagnoses/all-mq-codes.txt> (15 de abril de 2025).
19. Holm S, Russell G, Nourrit V, McLoughlin N. DR HAGIS-a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. *J Med Imaging (Bellingham)*. 2017;4:014503.
20. Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Fundus Image Database. <https://www5.cs.fau.de/research/data/fundus-images/> (15 de abril de 2025).
21. Calvo Herrero C, Román Díez R, Antuña Sánchez JC. Clasificación automática de imágenes de cielo mediante Inteligencia Artificial [TFG]. Valladolid: Universidad de Valladolid; 2023. Disponible en: <https://repositorio.uni-valladolid.es/xmlui/handle/123456789/xxx> (12 de mayo de 2025).
22. Malahina EA, Saitakela M, Bulan SJ, Lamabelawa MIJ, Belutowe YS. Teachable Machine: Optimization of herbal plant image classification based on epoch value, batch size and learning rate. *J Appl Data Sci*. 2024;5:1–10.
23. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56:45–50.
24. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and aprendizaje profundo in ophthalmology. *Br J Ophthalmol*. 2019;103:167–175.