

Deep learning sin código

Code free deep learning-AutoML

M. Olivera, C. Méndez Mangana, G. Santolària Rossell[†], Antón Barraquer Kargacin

Resumen

Mediante las herramientas de aprendizaje profundo sin código (CFDL, *code-free deep learning*), es posible desarrollar algoritmos con desempeño similar a los desarrollados de forma manual mediante la programación de redes neuronales, pero en este caso, sin la necesidad de programar código. Los sistemas basados en la nube, accesibles desde cualquier ordenador, a través de una simple interfaz web, hacen posible el desarrollo de estos algoritmos.

Grupos internacionales han analizado múltiples aplicaciones de esta técnica, especialmente en retinografías, pero nuestro grupo se encuentra probando los límites de esta técnica aplicados a la angiografía por tomografía de coherencia óptica (OCTA) y fotografías del segmento anterior para la detección del fracaso del injerto con interesantes resultados. Sin duda, el desarrollo de estas nuevas tecnologías hace más accesible el mundo de la inteligencia artificial (IA) a grupos de investigación más diversos, y son un importante paso en dirección de la democratización de la atención sanitaria.

Palabras clave: AutoML. Libre de código. Aprendizaje profundo automatizado. OCTA. Fracaso de injerto.

Resum

Mitjançant eines de *code-free deep learning* (CFDL) és possible desenvolupar algorismes amb rendiments similars als desenvolupats de forma clàssica mitjançant la programació manual de xarxes neuronals, però en aquest cas sense la necessitat de programar codi. Per mitjà de sistemes basats en el núvol, accessibles des de qualsevol ordinador, amb una simple interfície web és possible el desenvolupament d'aquests.

Grups internacionals han analitzat diverses aplicacions d'aquesta tècnica, especialment en retinografies, però el nostre grup està experimentant amb els límits d'aquesta tècnica aplicats a l'OCTA i fotografies del segment anterior per a la detecció del fracàs de l'implant amb resultats interessants. Sens dubte, el desenvolupament d'aquestes noves tecnologies fa de fàcil accés el món de la IA a grups de recerca més diversos i són un pas important en direcció a la democratització de l'atenció sanitària.

Paraules clau: AutoML. Lliure de codi. Aprenentatge profund automatitzat. OCTA. Fracàs de l'empelt.

Abstract

Using code-free deep learning (CFDL) tools, it is possible to develop algorithms with performances similar to those developed classically through manual programming of neural networks, but in this case without the need to program code, using cloud-based systems, accessible from any computer, through a simple web interface.

International groups have analyzed several applications of this technique, especially in retinography, but our group is experimenting with the limits of this technique applied to OCTA and photographs of the anterior segment for the detection of implant failure with interesting results. Undoubtedly, the development of these new technologies makes the world of AI easily accessible to more diverse research groups and are an important step in the direction of the democratization of healthcare.

Key words: AutoML. Code-free. Automated deep learning. OCTA. Graft failure.

6.1. Deep learning sin código

Code free deep learning-AutoML

M. Olivera¹, C. Méndez Mangana², G. Santolària Rossell†, Antón Barraquer Kargacin³

¹Complejo Hospitalario Universitario Insular Materno Infantil-Instituto Canario de la Retina. Las Palmas de Gran Canaria.

²Centro de ojos de La Coruña. ³Centro de Oftalmología Barraquer. Barcelona.

Correspondencia:

Maximiliano Olivera

E-mail: mxolivera@gmail.com



En memoria del Ing. Gil Santolària Rossell (1993-2023), especialista en Inteligencia Artificial. Coautor de capítulos de la presente ponencia, publicaciones científicas y congresos internacionales del grupo de investigación en IA del Centro de Oftalmología Barraquer. Además de sus logros profesionales, y siempre dispuesto a ayudar a los demás, Gil era una persona maravillosa con quien tendremos imborrables recuerdos y viajes compartidos. Tu legado perdurará en cada uno de nosotros. Descansa en paz amigo.

Introducción

Desde el comienzo de la actual época dorada de la IA aplicada a modelos de regresión, clasificación o detección, grandes equipos de ingenieros y expertos de área (por ejemplo: médicos) han colaborado mano a mano para el desarrollo de los algoritmos. A diferencia de un programa (*software*) convencional, que requiere instrucciones paso a paso para su funcionamiento, un algoritmo de IA necesita una serie de instrucciones que sirvan de guía sobre cómo se debe extraer información a partir de los datos a analizar (Figura 1).

A medida que el desarrollo de soluciones basadas en IA ha ido avanzando, se han desarrollado librerías de código abierto de *software* como Tensorflow¹ (por Google Brain Team) o Pytorch² (por Facebook AI Research Lab), basadas en lenguajes de programación como Python³ y C++⁴, que permiten el desarrollo de algoritmos (como redes neuronales) mediante instrucciones de código relativamente simples, con una sintaxis amigable. Estas librerías permiten además reutilizar arquitecturas de redes previamente entrenadas por los equipos de desarrollo, fácilmente adaptables (reentrenables) a los nuevos datos, en un proceso denominado transferencia de aprendizaje o *transfer learning*.

Para el desarrollo de todo el proceso descrito, es necesario que los usuarios o equipos de trabajo posean un profundo conocimiento

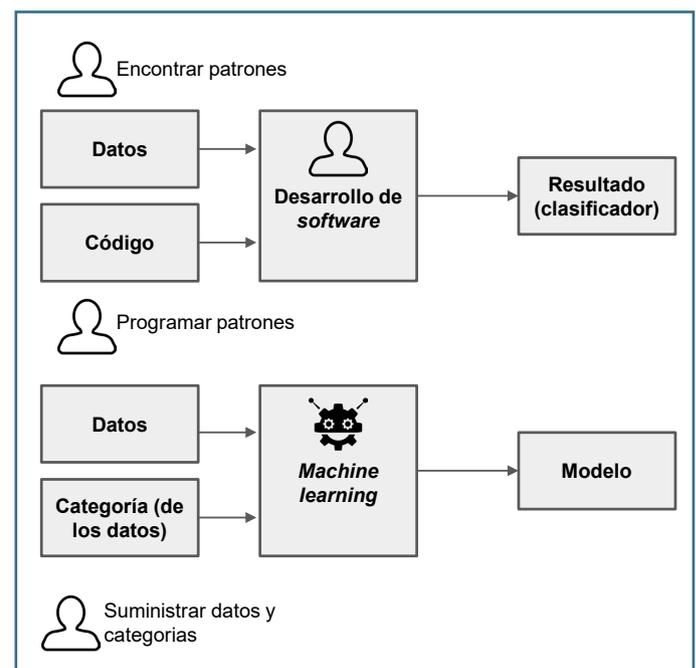


Figura 1. Esquema comparativo entre desarrollo de *software* tradicional frente a *machine learning*.

en programación en los lenguajes necesarios, experiencia previa y conocimiento del funcionamiento de las redes neuronales.

Siguiendo la misma lógica de simplificar el proceso del desarrollo de herramientas de IA, las mismas empresas tecnológicas ya mencionadas y otras como Apple, Microsoft o Amazon, entre otras, están trabajando en optimizar aún más este desarrollo. De esta forma, nace el concepto de “aprendizaje profundo sin código”, el cual podemos definir como un conjunto de herramientas, habitualmente basadas en la nube, que permiten a los usuarios crear, entrenar y desplegar modelos de aprendizaje profundo, sin la necesidad de dar forma a la arquitectura o programar las instrucciones para el desarrollo del modelo.

Técnicamente, si se posee un conjunto de datos, imágenes, por ejemplo, con sus respectivas etiquetas (categorías), mediante una simple interfaz gráfica web, es posible entrenar un modelo de clasificación, analizar las métricas, ponerlo a prueba e implementarlo tanto para su uso *online* como mediante un archivo descargable para integrar en aplicaciones móviles.

Cada una de dichas empresas tecnológicas ofrecen plataformas web con interfaz gráfica sencilla, haciéndolas más accesibles a públicos menos técnicos.

Aprendizaje profundo sin código frente a *deep learning* “clásico”

Mediante este proceso, teóricamente no es necesario el conocimiento de ningún tipo en programación, aunque en nuestra experiencia, el conocimiento del mismo puede facilitar y agilizar el proceso fundamentalmente a la hora de la ingestión y procesamiento de los datos (Figura 2).

Otro aspecto importante a la hora de considerar el uso del CFDL es el de los recursos de *hardware* (medios materiales como ordenadores). Los procesos computacionales involucrados en el desarrollo de algoritmos de IA son intensivos desde el punto

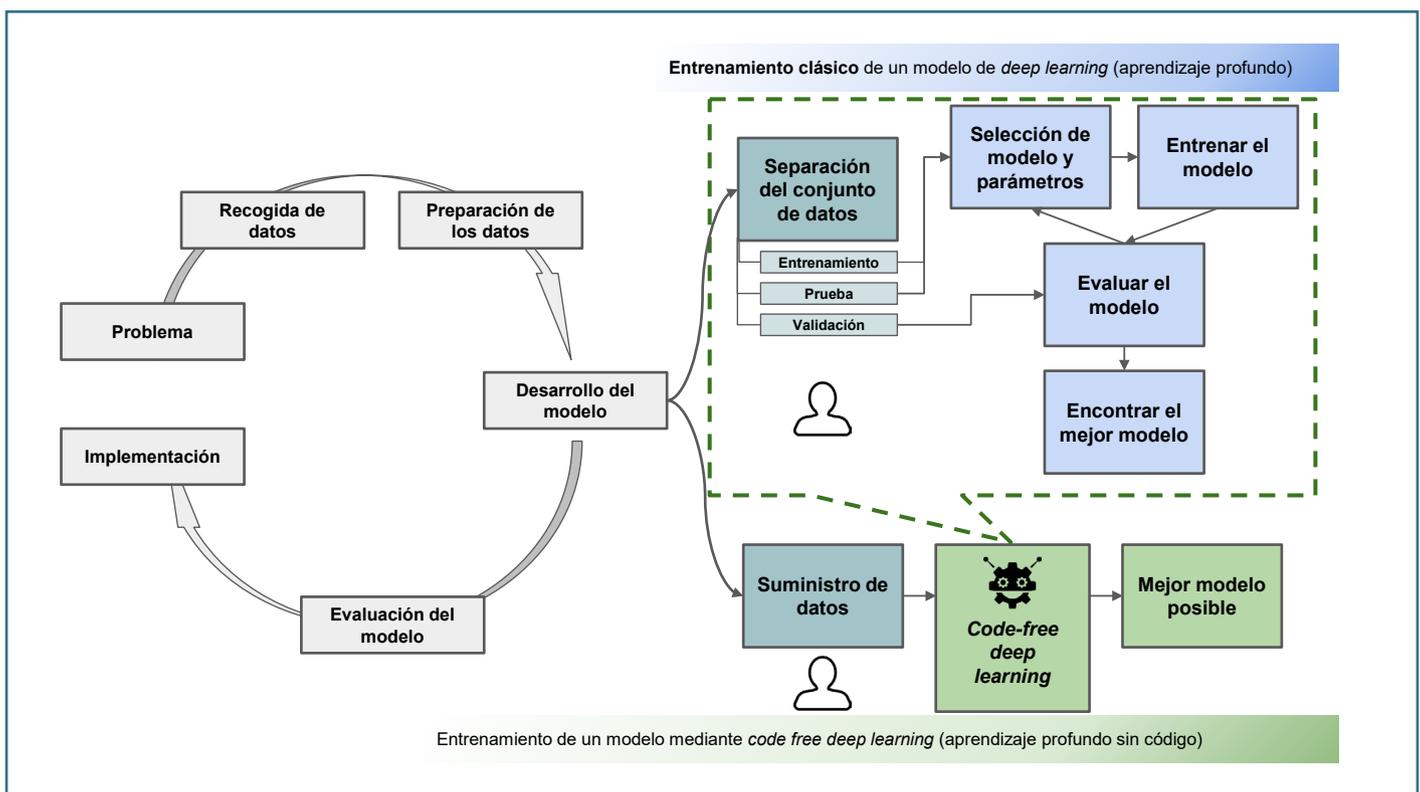


Figura 2. Ciclo de vida de un proyecto de *machine learning* y detalle comparativo de la fase de desarrollo del modelo con la forma clásica frente a aprendizaje profundo sin código (CFDL, *code free deep learning*). Para cada paso del proceso de desarrollo clásico es necesario contar con personal entrenado y con conocimientos de programación en los lenguajes y librerías correspondientes. En el esquema, se pueden ver todos los procesos que se realizan de forma automatizada por el CFDL (línea verde discontinua).

de vista del procesamiento de datos, millones de operaciones matemáticas son ejecutadas en cada fase de entrenamiento de los modelos.

Hoy en día, las gráficas diseñadas para la ejecución de entornos gráficos en tres dimensiones (videojuegos o diseño) están disponibles en cualquier tienda informática, son realmente útiles para el desarrollo de modelos simples de clasificación (lo más habitual en nuestro medio) y, con un coste contenido, es posible adquirir todos los componentes necesarios para su desarrollo.

Esta solución es la empleada por la mayoría de los grupos de investigación trabajando en el área de la IA, pero desde el punto de vista de demanda energética no es la más eficiente, y existen tarjetas gráficas diseñadas específicamente para procesos de aprendizaje automático o *machine learning*, que con una alta eficiencia (y muy elevado coste) pueden optimizar este proceso.

Sin embargo, las plataformas de CFDL, al estar basadas en la nube, no requieren la adquisición de ningún tipo de *hardware* especializado (pero no están exentas de coste), y cualquier ordenador conectado a internet, con capacidad de transferencia de datos, puede ser suficiente.

De esta forma, los procesos de entrenamiento realizados en remoto mediante esta tecnología no solo no requieren la adquisición de *hardware*, sino que también ofrecen la ventaja de pagar solo lo que se usa (alquiler), mediante una facturación basada en los recursos y el tiempo utilizados. Teniendo en cuenta que la mayor parte del tiempo de un proceso de *machine learning*⁵ se gasta en la gestión de los datos y el diseño del experimento, y que el uso del poder de cálculo se hace solo durante el entrenamiento de los modelos, el gasto en equipamiento puede no estar del todo

justificado si no se realizan intensas tareas de entrenamiento de forma frecuente, siendo las soluciones basadas en la nube, como el CFDL, una gran opción.

Publicaciones y trabajos destacados en el campo

Desde los inicios de esta tecnología, se ha hipotetizado sobre su uso en medicina, especialmente en oftalmología. En el año 2019⁶, el grupo del profesor P. Keane, de UCL-Moorfields, demostró que los modelos desarrollados sobre conjuntos de datos de alta calidad de retinografías, OCT, lesiones cutáneas y radiografías torácicas, mediante CFDL, utilizando la plataforma AutoML de Google, obtienen un desempeño comparable a los modelos de *deep learning* o aprendizaje profundo clásicos, considerados como el estándar de la industria, en problemas de clasificación simples.

Las limitaciones sobre la interpretabilidad de los resultados, la falta de acceso completo a la arquitectura empleada y la carencia de conjuntos de datos robustos de libre disponibilidad, son analizados en el artículo citado, y siguen siendo algunas de las limitaciones, que los detractores de esta tecnología siguen indicando a día de hoy.

El mismo grupo ha comparado diferentes plataformas de CFDL en su desempeño ante conjuntos de datos o *datasets* de dominio público en imagen multimodal de retina⁷ (Messidor-2, EyePACS, Waterloo, Kermany y IDRID) (Tabla 1), con buenos desempeños en general, cada uno con sus virtudes y defectos (lectura recomendada), con potenciales ámbitos de desempeño ideales para cada uno.

Puntaje F1	Waterloo (OCT)	Kermany (OCT)	EyePACS (retinografía)	Messidor (retinografía)	Media
Amazon	0,98	0,99	0,90	0,89	0,94
Apple	0,79	0,52	0,82	0,75	0,72
CarifAI	0,79	N/A	N/A	0,69	0,74
Google	0,94	0,98	0,92	0,85	0,92
MedicMind	0,97	N/A	N/A	0,84	0,91
Microsoft	0,95	0,91	0,84	0,85	0,89
Media	0,95	0,94	0,88	0,84	

OCT: tomografía de coherencia óptica. En negrita, se resalta el mejor resultado.
Adaptado de Korot *et al.*⁷.

Tabla 1. Desempeño mediante puntaje F1 ($2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$) de cada uno de los sistemas analizados de aprendizaje profundo sin código o *code-free deep learning* en cada uno de los conjuntos de datos.

Por último, el mismo grupo de Moorfields Eye Hospital ha publicado recientemente un artículo donde describen el proceso de desarrollo y validación internacional de un modelo desarrollado mediante CFDL y un modelo clásico para la detección de enfermedad plus en retinopatía del prematuro⁸.

Los modelos desarrollados a partir de 7.414 imágenes retinianas mediante ambas técnicas tienen métricas de desempeño comparables (área bajo la curva de 0,986 para el modelo clásico y 0,989 para el modelo CFDL), pero sin embargo, el modelo desarrollado a partir de CFDL pierde cierta capacidad de generalización con imágenes adquiridas con otros dispositivos durante la validación externa, lo que según los autores debe ser considerado a la hora de la implementación de estos modelos en la práctica habitual, para evitar sesgos que puedan ser negativos para los pacientes.

Nuestra experiencia

OCTA y retinopatía diabética

Siguiendo las recomendaciones generales, en nuestro grupo, hemos utilizado esta herramienta para el desarrollo de dos líneas de investigación. Por un lado, la detección y clasificación de la retinopatía diabética mediante OCTA⁹ y, por otro, la detección

del fracaso del injerto postqueratoplastia mediante fotografía del segmento anterior.

En el caso de la OCTA, el conjunto de datos proviene de un ensayo clínico¹⁰ orientado al seguimiento longitudinal, tanto de variables oculares como sistémicas, de pacientes diabéticos tipo 1 y controles. El uso de esta herramienta permite explorar cada modalidad de imagen (retinografía, OCT-b, OCTA), cada plexo y área de escaneo (OCTA) por separado, con la capacidad de esta herramienta de entrenar modelos de clasificación para diferentes tareas, como la presencia o no de retinopatía diabética y el grado de severidad de la misma. Los hallazgos preliminares nos permiten identificar los modos de imagen con mayor capacidad de discriminación, y a partir de allí, avanzar en tareas más complejas hacia el campo de la oculómica (Figura 3).

Queratoplastias y fracaso del injerto

Otra línea de investigación en desarrollo por nuestro grupo se focaliza en la detección de fracaso del injerto en pacientes postqueratoplastia¹¹. En este caso, hemos seleccionado imágenes del segmento anterior de pacientes a los que se les había realizado una queratoplastia penetrante y han tenido un seguimiento a lo largo de, al menos, un año. Para el desarrollo de este proyecto,

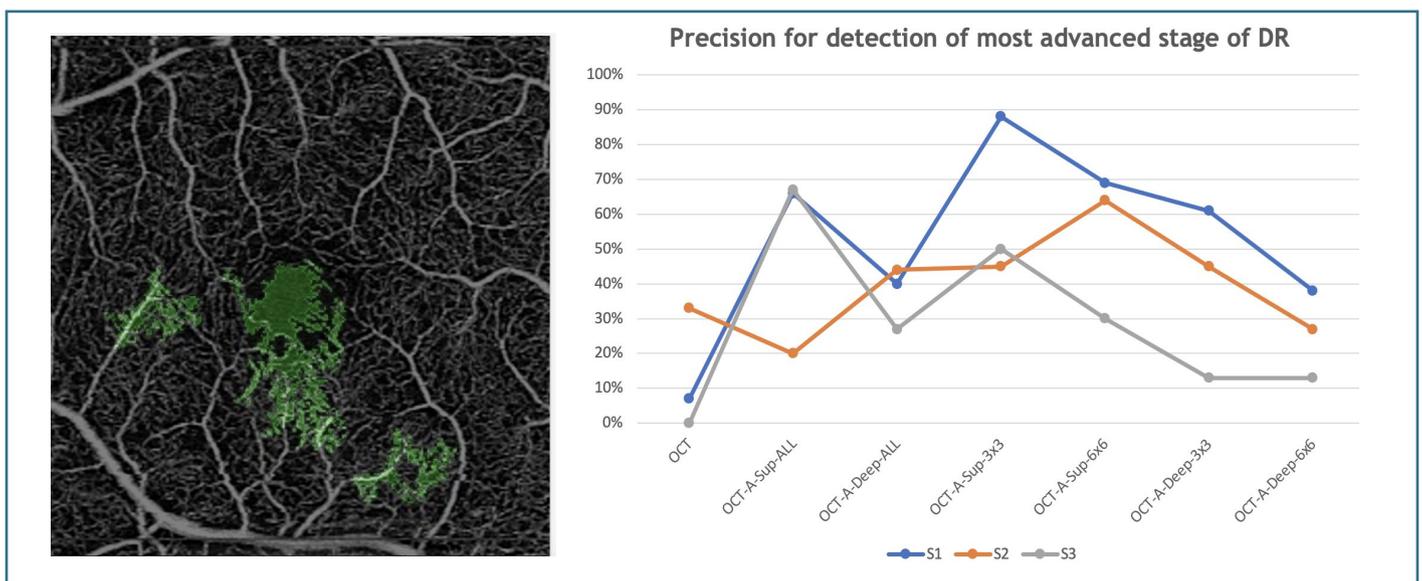


Figura 3. Ejemplo de un caso de angiografía por tomografía de coherencia óptica (OCTA) en el plexo superficial para la clasificación de la retinopatía diabética. El modelo de clasificación utiliza información de la zona avascular foveal y de algunas tramas vasculares para la toma de su decisión. Performance de los modelos en cada tipo de captura de OCTA para la detección de la retinopatía diabética según diferentes estrategias⁹.

hemos optado por el uso del CFDL en dos fases, que se explican a continuación.

Prueba de concepto

Después de las fases iniciales de observación y detección del problema, el planteamiento técnico, la recogida de datos (imágenes) y la verificación de la calidad del etiquetado de las mismas, lo siguiente era saber si era posible entrenar un modelo de visión por computador para clasificar a las imágenes entre fracaso y no fracaso del injerto, ante la ausencia de bibliografía sobre problemas similares y ciertas limitaciones de recursos, problemas sobre qué preprocesado de imágenes realizar, qué arquitecturas utilizar y el *hardware* necesario.

Mediante CFDL, hemos podido entrenar un modelo simple de clasificación binaria: fracaso/sano, en clases balanceadas (50% de representatividad de cada una) y a resolución completa (nativa). El resultado de esta prueba de concepto superó ampliamente nuestras expectativas, mostrando una sensibilidad y especificidad por encima del 90% y un puntaje F1 del 95,99%

Preprocesado de datos, enmascaramiento de imágenes y clases desbalanceadas

En todo modelo de *deep learning* que involucre imágenes, es importante contar con una muestra normalizada de imágenes, idealmente a baja resolución, debido a que el proceso de convoluciones (análisis píxel a píxel por el cual las redes neuronales convolucionales analizan las imágenes suministradas) involucrado en el entrenamiento de la red neural analiza píxel a píxel cada imagen, requiriendo el análisis de alrededor de 3 millones de puntos de información para una imagen de 1 megapíxel (MP), requiriendo un gran poder de cálculo informático. Además, durante la prueba

de concepto, hemos detectado mediante mapas de activación que, en algunas imágenes, el modelo parecía “fijarse” en áreas que no eran clínicamente relevantes, como las pestañas.

Para tratar estas observaciones, mediante el preprocesado de las imágenes, hemos entrenado modelos con imágenes¹² a diversas resoluciones, desde 1,3 MP hasta 0,07 MP, sin máscaras y con máscaras agresivas que eliminan las estructuras anexas más allá de la conjuntiva (Figura 4).

Para aumentar el desafío, a diferencia de la prueba de concepto inicial, estos modelos fueron entrenados con las clases desbalanceadas, con una incidencia del 20% de fracasos, que es la real de nuestro conjunto de datos y está en línea con la bibliografía.

Con este tratamiento del conjunto de datos, hemos visto que, en las imágenes sin máscara, el desempeño de los modelos baja drásticamente, al mismo tiempo que se reduce la resolución de las mismas, pero, sin embargo, al aplicar máscaras automatizadas (a partir de un *script*-programa en Python-proprietario), los modelos vuelven a obtener muy buen desempeño, detectando hasta el 85% de los fracasos en las resoluciones más bajas.

Consideraciones finales y conclusiones

Como algunos de los máximos referentes del área explican¹³, el CFDL puede representar la próxima fase de la atención e investigación sanitaria asistida por IA. Esta tecnología aporta la novedad de simplificar el proceso necesario para el desarrollo de los modelos, ya que, a pesar de comenzar con la misma estructura de planteamiento del problema, recogida y etiquetado de datos,

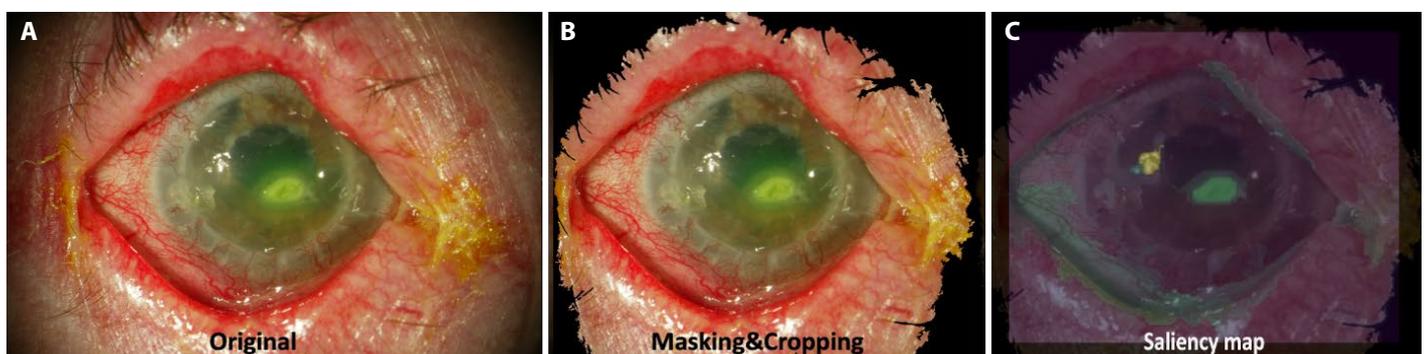


Figura 4. Ejemplo de procesamiento de datos. (A) Imagen original a resolución nativa de 1,3 MP. (B) Imagen con máscara y resolución acotada a 0,07 MP (imagen escalada para la representación en este medio). (C) Se ha aplicado sobre la imagen tratada un mapa de activación correspondiente al modelo de detección de fallo del injerto.

el proceso de entrenamiento de los modelos se hace de forma automatizada. No es necesario programar la red neural, tampoco es necesario procesar las imágenes; simplemente transfiriendo los archivos a los servidores, dentro de carpetas que indiquen su "etiqueta", es posible, a través de una interfaz gráfica, llegar a entrenar el modelo (Figura 5).

Sin embargo, como hemos intentado remarcar a lo largo del presente capítulo, esto no implica la automatización completa del desarrollo del modelo, el conocimiento (*domain-expertise*) de un oftalmólogo que asegure el planteamiento de un problema lógico, la supervisión por dichos expertos en momentos importantes del desarrollo del modelo y la calidad de los datos, son fundamentales.

Hoy en día, también es importante conocer y entender conceptos básicos de estadística y *machine learning* para poder analizar y entender de forma adecuada los resultados de los procesos de entrenamiento. Otro aspecto fundamental es tener en cuenta que, en el caso de que busquemos desarrollar modelos con imágenes propias, estas serán transferidas a servidores externos (Google, Amazon, Apple...), por lo que es fundamental el proceso de anonimización de las mismas, que no permitan la identificación de los pacientes, quienes además deben haber prestado su consentimiento por escrito para el uso en investigación de sus imágenes clínicas.

Una buena práctica en este sentido es, además del cambio de nombre de los archivos, borrar sus metadatos, que son datos

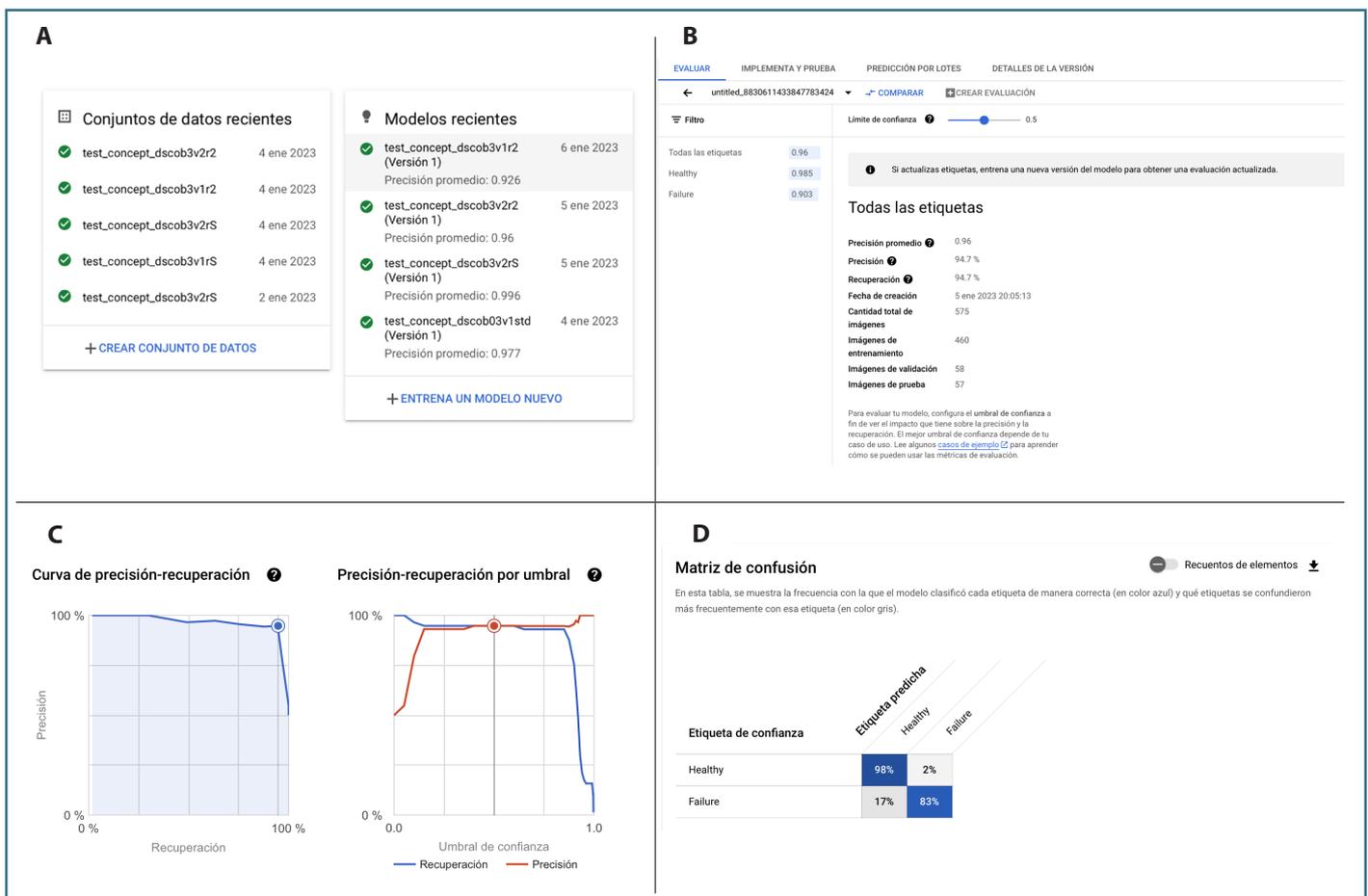


Figura 5. Elementos de la interfaz gráfica de usuario de la plataforma Vertex-AI (AutoML de Google). **(A)** Resumen de los conjuntos de datos y modelos más recientes. **(B)** Resumen de métricas fundamentales de un modelo. **(C)** Curvas de precisión-sensibilidad y precisión según el punto de corte de confianza. **(D)** Matriz de confusión correspondiente a un modelo.

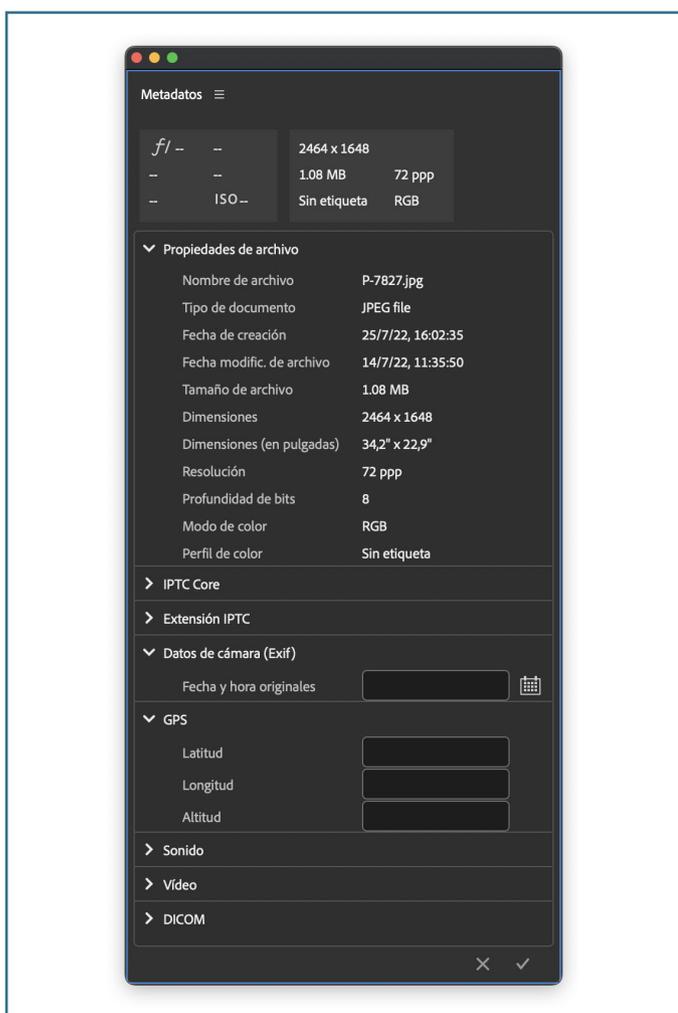


Figura 6. Ejemplo de imagen anonimizada, donde el nombre del archivo no contiene ningún dato trazable, y los metadatos se encuentran borrados o modificados con datos genéricos que evitan la trazabilidad del archivo original.

incluidos no accesibles a “simple vista” por los usuarios, pero sí fácilmente desde el apartado “propiedades” con cualquier explorador de archivos de nuestros ordenadores (Figura 6).

En conclusión, el CFDL es una gran herramienta que puede permitirle a grupos pequeños de investigadores, sin la necesidad de contar con un experto en *machine learning* y con limitaciones de recursos locales, el desarrollo de modelos o, al menos, pruebas de concepto de modelos de *deep learning* con posibilidades de desempeño comparable al abordaje clásico.

Es crucial, sin embargo, que miembros del equipo tengan conocimientos de nivel experto en el área de estudio que aseguren la calidad de los datos, y un nivel mínimo de comprensión de

machine learning que permita analizar de forma correcta el desempeño de los modelos y alcanzar conclusiones adecuadas.

Bibliografía

1. TensorFlow. [Internet]. [Citado 16 Jun 2021]. Disponible en: <https://www.tensorflow.org/>
2. PyTorch. [Internet]. The Linux Foundation. [Citado 16 Jun 2021]. Disponible en: <https://pytorch.org>
3. Python. [Internet]. Python Software Foundation. [Citado 15 May 2023]. Disponible en: <https://www.python.org/>
4. Standard C++. [Internet]. C++ Foundation. [Citado 15 May 2023]. Disponible en: <https://isocpp.org/>
5. Paleyes A, Urma RG, Lawrence ND. Challenges in Deploying Machine Learning: a Survey of Case Studies. [Internet]. En: arXiv.org. 18 Nov 2020. [Actualizado 19 May 2022]. Disponible en: <http://arxiv.org/abs/2011.09926>
6. Faes L, Wagner SK, Fu DJ, Liu X, Korot E, Ledsam JR, *et al.* Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health.* 2019;1:e232-42.
7. Korot E, Guan Z, Ferraz D, Wagner SK, Zhang G, Liu X, *et al.* Code-free deep learning for multi-modality medical image classification. *Nat Mach Intell.* 2021;3(4):288-98.
8. Wagner SK, Liefers B, Radia M, Zhang G, Struyven R, Faes L, *et al.* Development and international validation of custom-engineered and code-free deep-learning models for detection of plus disease in retinopathy of prematurity: a retrospective study. *Lancet Digit Health.* 2023;5(6):e340-9.
9. Olivera M, Bernal-Morales C, Alé-Chilet A, Barraso M, Marin S, Feu S, *et al.* Automated Machine learning models applied to Optical Coherence Tomography Angiography for detection and classification of Diabetic Retinopathy in Diabetes Mellitus type 1. *Invest Ophthalmol Vis Sci.* 2022;63(7):2326.
10. Zarranz-Ventura J, Barraso M, Alé-Chilet A, Hernández T, Oliva C, Gascón J, *et al.* Evaluation of microvascular changes in the perifoveal vascular network using optical coherence tomography angiography (OCTA) in type I diabetes mellitus: a large scale prospective trial. *BMC Med Imaging.* 2019;19(1):91.
11. Mangana CM, Kargacin AB, Fernández-Engroba J, Taña P, Santolaria G, Olivera M, *et al.* Detection of graft failure in post-keratoplasty patients by Automated Deep Learning. *Invest Ophthalmol Vis Sci.* 2022;63(7):2330.
12. Mangana CM, Kargacin AB, Engroba JF, Ferragut-Alegre Á, Tañá P, Zarranz-Ventura J, *et al.* Post-keratoplasty graft failure detection with Code Free Deep Learning: Image processing and real-life incidence influence on training performance. *Invest Ophthalmol Vis Sci.* 2023;64(8):1103.
13. O'Byrne C, Keane PA. Code-free deep learning: The next phase of AI-enabled healthcare? [Internet]. En: Ophthalmologytimes.com. 28 Oct 2021. [Citado 2 May 2023]. Disponible en: <https://europe.ophthalmologytimes.com/view/code-free-deep-learning-the-next-phase-of-ai-enabled-healthcare->