

Evaluación retrospectiva de un primer modelo de inteligencia artificial argentino para tamizaje automático de retinopatía diabética referible a partir de fotografías de fondo de ojo

Dr. Ing. José Ignacio Orlando, Ing. Tomás Castilla, Dr. Alejandro Koch, Dr. Ignacio Larrabide, Dra. Marcela Martínez y Dra. Mercedes Leguía



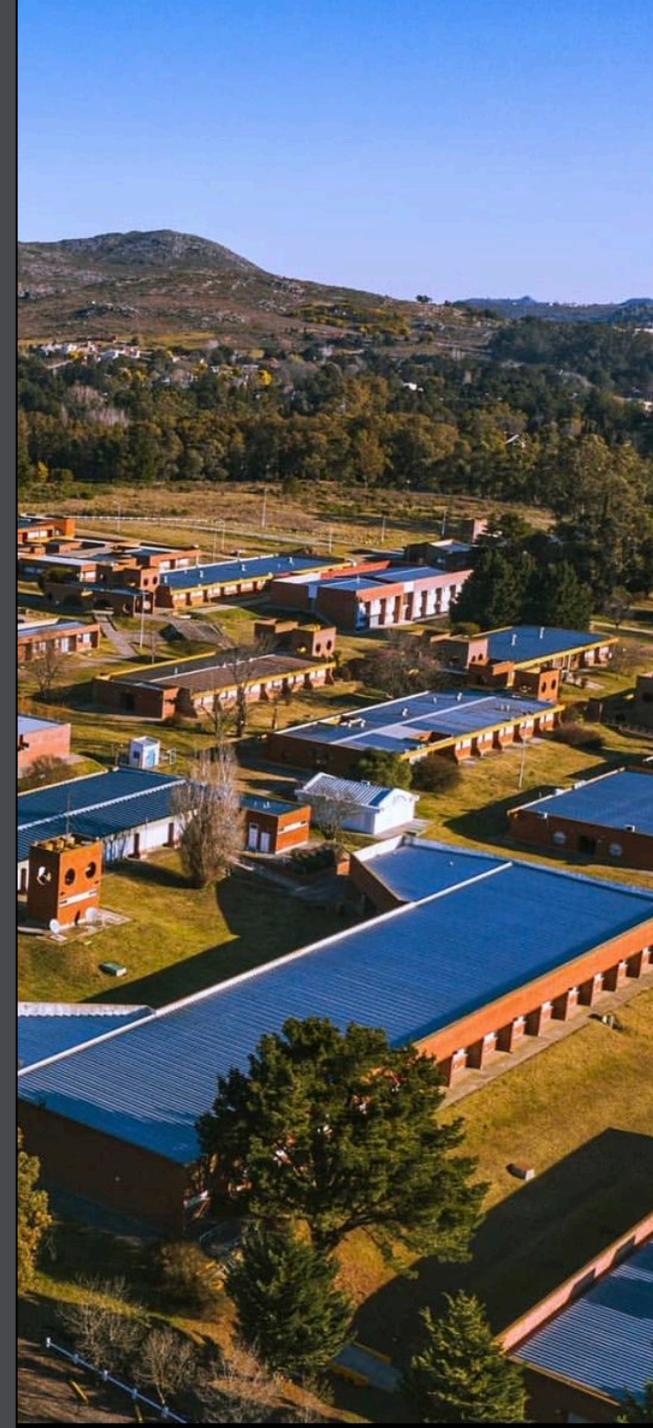
UNICEN
Universidad Nacional del Centro
de la Provincia de Buenos Aires



martínez
CENTRO DE OFTALMOLOGÍA



ALTA COMPLEJIDAD EN RED
Hospital El Cruce
Dr. Néstor Carlos Kirchner



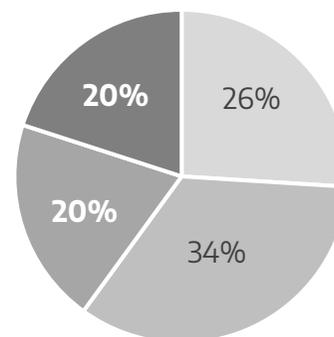
La retinopatía diabética en Argentina

Un **16% de los casos** de ceguera en Argentina se deben a la retinopatía diabética no tratada ¹



El **77%** de las personas diabéticas **no cumple con el control oftalmológico anual** ¹

Más del **20%** de las personas que acuden a la consulta por síntomas ingresan **en estado de ceguera legal** ²



Agudeza visual (AV) en la primera consulta ²

- AV normal
- AV aceptable
- AV baja/moderada
- Ceguera legal



¹ Encuesta nacional de ceguera y deficiencia visual evitable en Argentina, 2013

Barrenechea et al. Pan American Journal of Public Health. 2013

² Cuidado ocular del paciente con diabetes: opiniones y recomendaciones de la SARyV.

Carrasco et al. Revista OCE (CAO). 2013. https://oftalmologos.org.ar/oce_anteriores/items/show/68

³ Retinopatía diabética y agudeza visual en la primera consulta.

Vouilloz et al. 11º Jornadas Científicas y de Gestión del Hospital El Cruce. 2018.

<https://repositorio.hospitalelcruce.org/xmlui/handle/123456789/742?show=full>

La telemedicina como herramienta de diagnóstico

Técnicos capacitados obtienen retinografías en nodos de captura

- No es necesario un oftalmólogo en el lugar para obtener el estudio
- Alivia la carga de trabajo del profesional (se terceriza la captura de la imagen)

La imagen se analiza remotamente en un centro de informes

- Los oftalmólogos reciben los estudios, los analizan y generan el diagnóstico correspondiente.
- No es necesario que el oftalmólogo se desplace hacia la población bajo análisis

↑ Más facilidad de acceso al control oftalmológico anual

↑ Enfoque costo-efectivo para prevenir ceguera por diabetes

↓ Dificultades de escalabilidad
Cuando la red crece, puede saturarse el centro de informes.

Clinical science



Telemedicine for detecting diabetic retinopathy: a systematic review and meta-analysis

Lili Shi,^{1,2} Huiqun Wu,¹ Jiancheng Dong,¹ Kui Jiang,¹ Xiting Lu,³ Jian Shi⁴

Original Investigation | Journal Club

February 2015

Diabetes Eye Screening in Urban Settings Serving Minority Populations

Detection of Diabetic Retinopathy and Other Ocular Findings Using Telemedicine

Cynthia Owsley, PhD¹; Gerald McGwin Jr, PhD^{1,2}; David J. Lee, PhD³; et al

> Author Affiliations | Article Information

JAMA Ophthalmol. 2015;133(2):174-181. doi:10.1001/jamaophthalmol.2014.4652

Evaluación de un programa de teleoftalmología para prevención de la ceguera por diabetes en una zona rural de la Argentina

Assessment of a teleophthalmology program for the prevention of diabetes blindness in a rural area of Argentina

Avaliação de um programa de teleoftalmologia para a prevenção da cegueira por diabetes em uma área rural da Argentina

Tomás Ortiz-Basso¹, Paula Verónica Gomez², Analia Boffelli³, Ariel Paladini⁴.

La diabetes es la principal causa de ceguera en personas laboralmente activas. A través de una foto del ojo, los médicos pueden detectar a tiempo esta enfermedad. Esta forma de atención se llama teleoftalmología y permite un rápido examen de ojos, evitando la consulta presencial. El objetivo fue evaluar el funcionamiento de un programa teleoftalmológico en una zona rural de la Argentina. Se observó que la realización del examen de ojos luego de comenzar el programa mejoró en un 39%. Desarrollar este tipo de programas es fundamental para prevenir la ceguera por diabetes.



Ophthalmology
Volume 123, Issue 12, December 2016, Pages 2571-2580



Original article

Cost-effectiveness of a National Telemedicine Diabetic Retinopathy Screening Program in Singapore

Hai V. Nguyen PhD¹, Gavin Siew Wei Tan MMed(Ophth), FAMS², Robyn Jennifer Tapp PhD^{3,4}, Shweta Mital MSc⁵, Daniel Shu Wei Ting MD, PhD², Hon Tym Wong FRCSed(Ophth), MMed(Ophth)⁶, Colin S. Tan FRCSed(Ophth), MMed(Ophth)⁶, Augustinus Laude FRCSed(Ophth), FAMS(Ophth)⁷, E. Shyong Tai FRCP, PhD⁸, Ngiap Chuan Tan MMed(FM), MCI⁹, Eric A. Finkelstein PhD, MHA⁵, Tien Yin Wong MMed(Ophth), PhD^{2,10}, * Ecosse L. Lamoureux MSc, PhD^{2,10}, * 83

Inteligencia artificial para tamizaje de casos de riesgo

La inteligencia artificial es una tecnología madura!

- Resultados comparables con el humano en problemas de visión artificial.
- **Ya hay soluciones comerciales seguras para detectar retinopatía diabética**

Detección automática de casos de riesgo utilizando IA

- Reduce la carga de trabajo para informes (triage de casos de riesgo o inciertos, tamizaje de casos detectados como positivos)
- Una opinión adicional y de bajo costo para colaborar en el diagnóstico médico.

↑ Más eficiencia de las redes de teleoftalmología

↑ Menores costos y mayor efectividad en el diagnóstico

↓ Soluciones comerciales muy costosas, en moneda extranjera

↓ Requieren cámaras específicas

npj | Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices

Michael D. Abràmoff^{1,2,3,4}, Philip T. Lavin⁵, Michele Birch⁶, Nilay Shah⁷ and James C. Folk^{1,2,3}

Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems

Diabetes Care 2021;44:1168–1175 | <https://doi.org/10.2337/dc20-1877>

Aaron Y. Lee,^{1,2,3} Ryan T. Yanagihara,¹ Cecilia S. Lee,^{1,2} Marian Blazes,¹ Hoon C. Jung,^{1,2} Yewlin E. Chee,¹ Michael D. Gencarella,¹ Harry Gee,⁴ April Y. Maa,^{5,6} Glenn C. Cockerham,^{7,8} Mary Lynch,^{5,9} and Edward J. Boyko,^{10,11}

Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients **BMJ**

Peter Heydon¹, Catherine Egan,^{1,2} Louis Bolter,³ Ryan Chambers,³ John Anderson,³ Steve Aldington,⁴ Irene M Stratton,⁴ Peter Henry Scanlon⁴, Laura Webster,⁵ Samantha Mann,⁵ Alain du Chemin,⁵ Christopher G Owen⁶, Adnan Tufail,^{1,2} Alicja Regina Rudnicka⁶

Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study

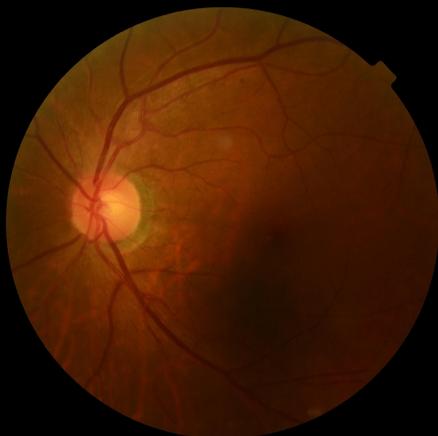
Yuchen Xie*, Quang D Nguyen*, Haslina Hamzah*, Gilbert Lim, Valentina Bellemo, Dinesh V Gunasekeran, Michelle Y T Yip, Xin Qi Lee, Wynne Hsu, Mong Li Lee, Colin S Tan, Hon Tym Wong, Ecosse L Lamoureux, Gavin SW Tan¹, Tien Y Wong¹, Eric A Finkelstein¹, Daniel SW Ting¹

Hacia una IA nacional para tamizaje de RD

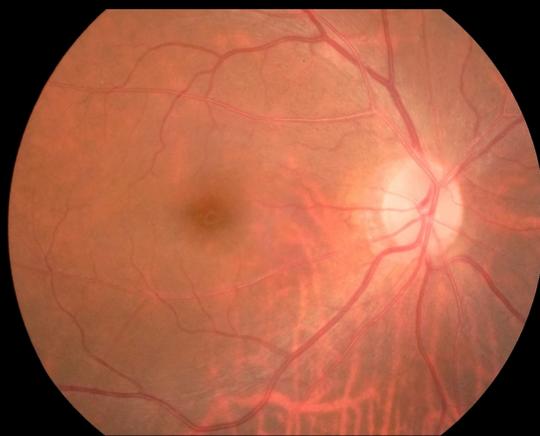
Un modelo de inteligencia artificial para diferenciar casos referibles de no referibles

- Los casos referibles requieren ser revisados por un profesional para determinar si requiere tratamiento.
- Los no referibles pueden esperar 1 año para realizarse un nuevo control.

Sin signos de RD
(Grado 0)



RDNP leve
(Grado 1)



RDNP moderada
(Grado 2)



RDNP severa
(Grado 3)



RDP
(Grado 4)



Casos no referibles

Casos referibles

¹ Imágenes y etiquetas extraídas del banco de datos públicos DDR (<https://github.com/nkicsl/DDR-dataset>)

² Escalas de severidad de retinopatía diabética según ICDR. Wilkinson et al. 2003, *Ophthalmology*.

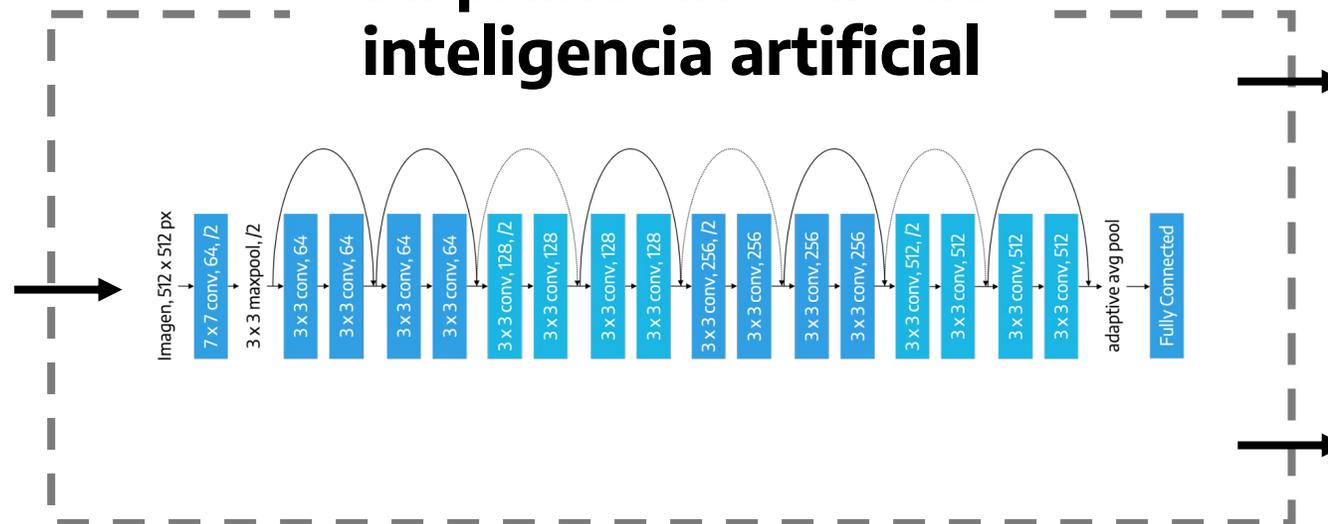
Hacia una IA nacional para tamizaje de RD

Entrada



Fotografía de fondo de ojo

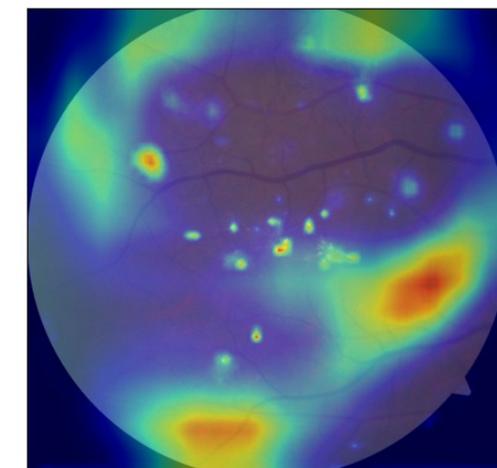
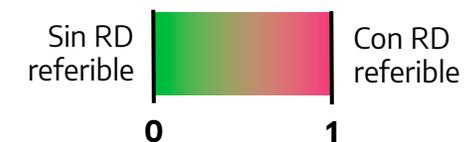
Un primer modelo de inteligencia artificial



Red neuronal convolucional de clasificación
Mejora de contraste
Datos de entrenamiento de orígenes diversos
Estrategia de aumentación de imágenes calibrada
XGradCAM para feedback

Salidas

1) Probabilidad de ser un caso de RD referible

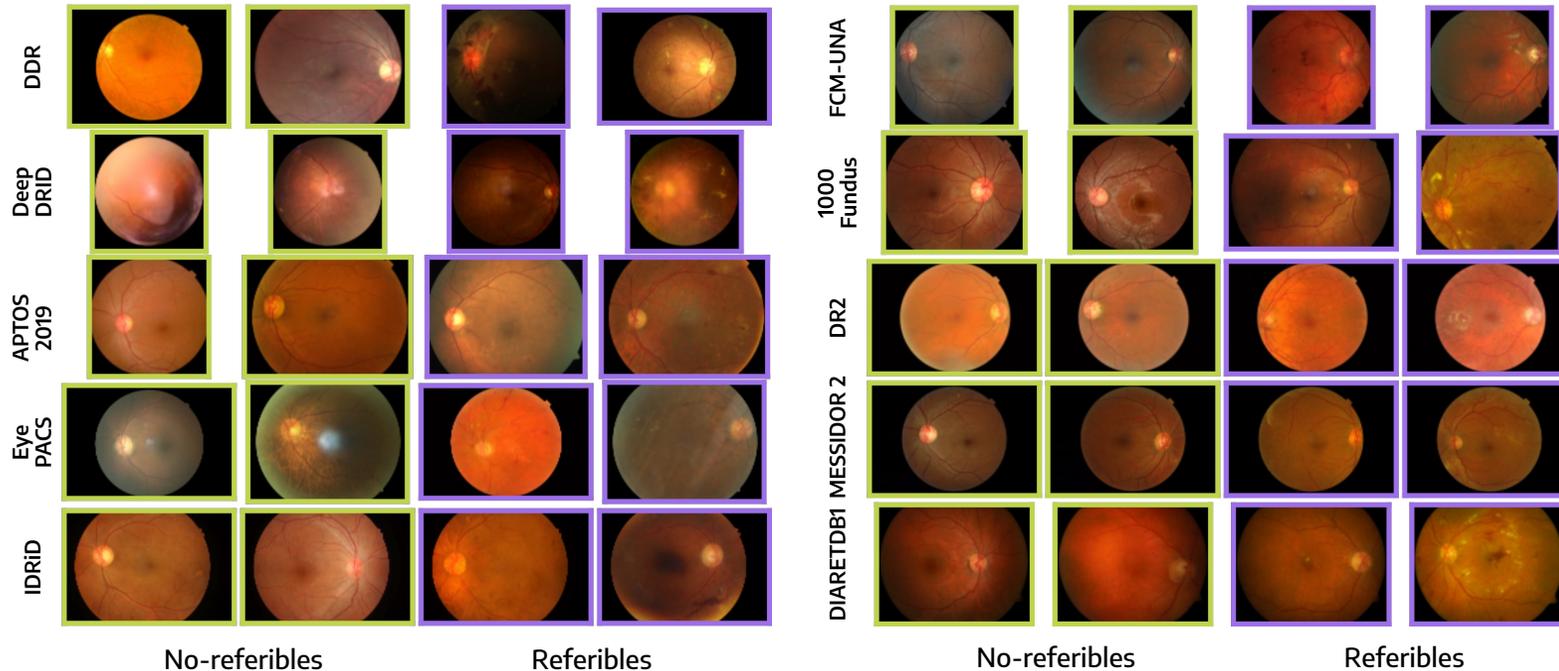


2) Regiones tenidas en cuenta por el modelo

Materiales

117.168 imágenes en total, recolectadas a partir de **bancos de datos públicos y privados**

- Sin RD referible: 89.433 imágenes.
- Con RD referible: 27.735 imágenes.



Materiales

117.168 imágenes en total, recolectadas a partir de **bancos de datos públicos y privados**

- Sin RD referible: 89.433 imágenes.
- Con RD referible: 27.735 imágenes.

Separadas en 3 conjuntos disjuntos:

Entrenamiento	Validación	Test
39.592	9.569	61.526
Usadas para entrenar el modelo predictivo	Usadas para ajustar el modelo predictivo (selección de puntos operativos)	Usadas para evaluar el modelo (no vistas por el algoritmo)
		Estudio observacional retrospectivo

Materiales: evaluación

Estudio observacional retrospectivo sobre las 61.525 imágenes de test

- ➔ **Valores altos de sensibilidad y especificidad**
Más sensibilidad, más casos de riesgo detectados, menos riesgo de ceguera.
Más especificidad, menor cantidad de falsos positivos, menos trabajo para informe.
- ➔ **Robustez a variaciones en cámaras, comorbilidades, resoluciones, étnias, ...**
Mayor aplicabilidad del sistema en la infraestructura ya existente, menores costos.
- ➔ **Feedback para el/la profesional**
Facilidad para confirmar/descartar el diagnóstico automático, más eficiencia en el informe, menor efecto “caja negra”.

Materiales: evaluación

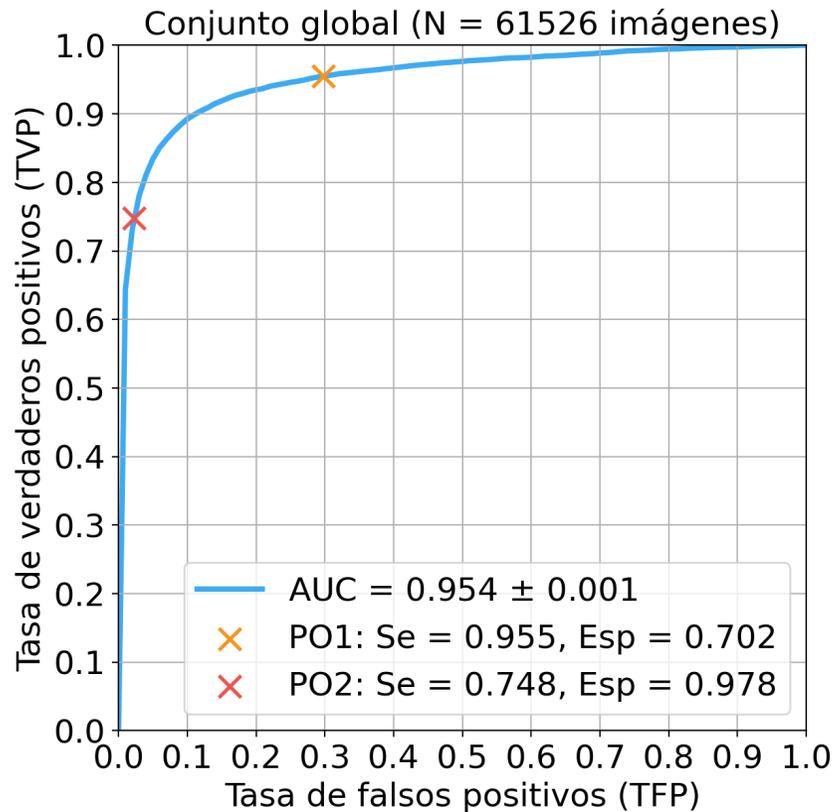
Estudio observacional retrospectivo sobre las 61.525 imágenes de test

Conjunto de datos	Número de imágenes	País de origen de los datos	Cámaras de fondo de ojo		
			Modelo	Midriática?	FOV
DDR	3.759	China	41 diferentes, no reportadas *		45°
IDRiD	103	India	Kowa VC-10 α	SI	50°
MESSIDOR 2	1.744	Francia	Topcon TRC NW6	SI y NO	45°
EyePACS	53.576	EE.UU.	No reportado		
DR2	435	Brasil	Topcon TRC NW8	NO	45°
1000Fundus	144	China	Zeiss FF450 Plus IR, Topcon TRC-50DX	SI	35-50°
DeepDRID	400	China	Topcon TRC NW400	No reportado	45°
DIARETDB1	89	Finlandia	Zeiss FF450plus	No reportado	50°
FCM-UNA	757	Paraguay	Zeiss Visucam 500	SI	45°
Martínez	484	Argentina	Cristal Vue NFC-700	NO	45°
HEC	35	Argentina	Topcon TRC NW8	NO	45°

* Principalmente modelos Topcon D7000, Topcon TRC NW48, Nikon D5200 y Canon CR 2.



Resultados cuantitativos



Dos puntos operativos (PO)

Seleccionados a partir de datos de validación, para controlar sensibilidad y especificidad ¹

PO1: probabilidad de referibilidad > 0.02

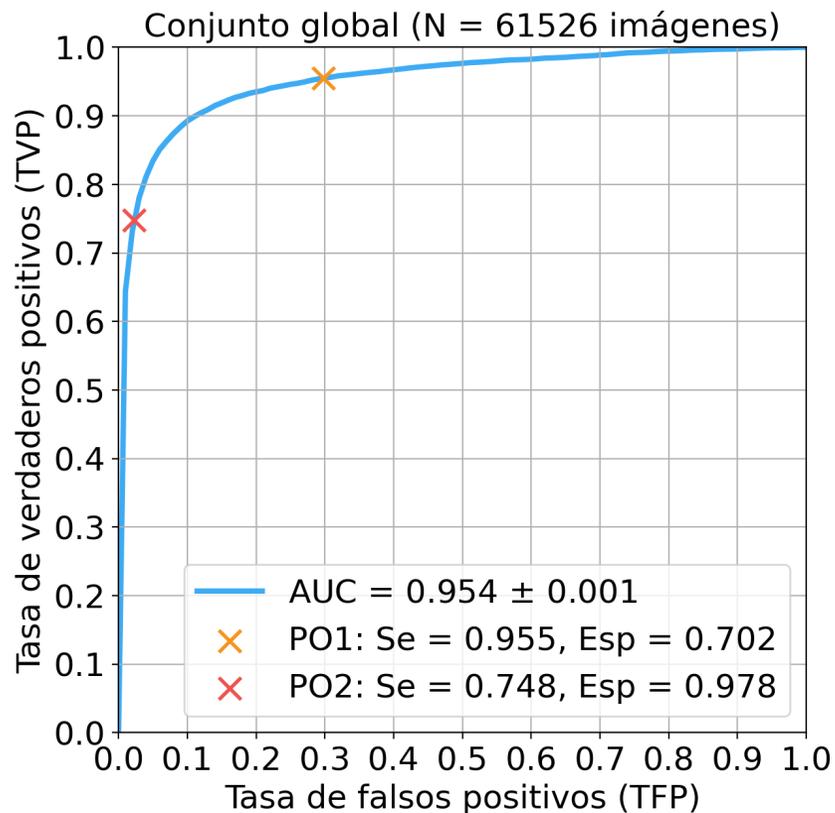
95.5% de casos positivos detectados,
29.8% de falsos positivos

PO2: probabilidad de referibilidad > 0.51

74.8% de casos positivos detectados,
2.2% de falsos positivos

¹ Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, Gulshan et al. JAMA. 2016

Resultados cuantitativos



Conjunto de datos	Número de imágenes	Métricas		
		AUC	Sensibilidad	Especificidad
DDR	3.759	0.965	75%	98%
IDRiD	103	0.949	83%	90%
MESSIDOR 2	1.744	0.973	89%	94%
EyePACS	53.576	0.951	73%	98%
DR2	435	0.974	84%	96%
1000Fundus	144	1.000	100%	96%
DeepDRID	400	0.959	88%	87%
DIARETDB1	89	0.981	96%	93%
FCM-UNA	757	0.986	88%	99%
Martínez	484	0.955	80%	94%
HEC	35	0.961	100%	86%

Rendimiento por conjunto con sensibilidades y especificidades superiores al 73% y el 86%

Resultados cuantitativos

Conjunto de datos	Modelo	Métricas		
		AUC	Sensibilidad	Especificidad
DDR	Zago et al. 2020	0.833	-	-
	Este modelo	0.965	75%	98%
IDRiD	Zago et al. 2020	0.796	-	-
	Hervella et al. 2022	0.944	-	-
	Este modelo	0.949	83%	90%
MESSIDOR 2	Gargeya et al. 2017	0.940	93%	87%
	Voets et al. 2019	0.853	82%	69%
	Zago et al. 2020	0.944	90%	87%
	Este modelo	0.973	89%	94%
EyePACS	Pires et al. 2019	0.946	-	-
	Este modelo	0.951	73%	98%
DR2	Pires et al. 2019	0.963	-	-
	Este modelo	0.974	84%	96%

Rendimiento similar o superior al de otros modelos internacionales evaluados sobre los mismos conjuntos

¹ Zago GT, Andreão RV, Dorizzi B, Salles EOT. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Comput Biol Med*, 2020; 116: e103537.

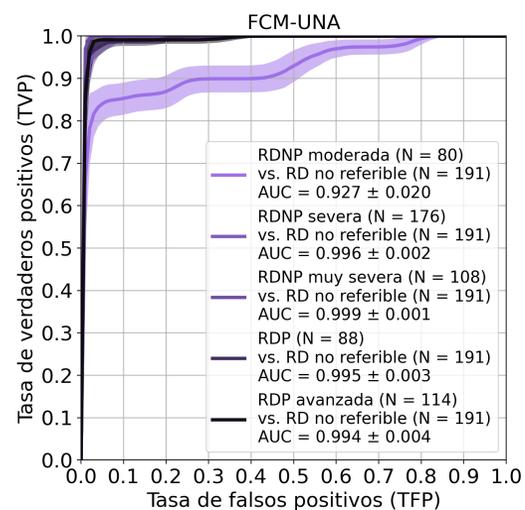
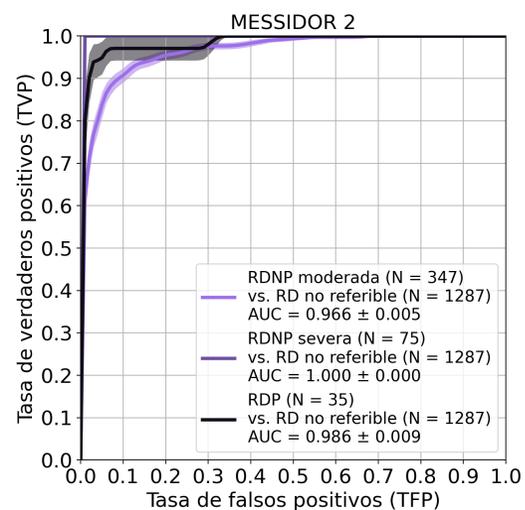
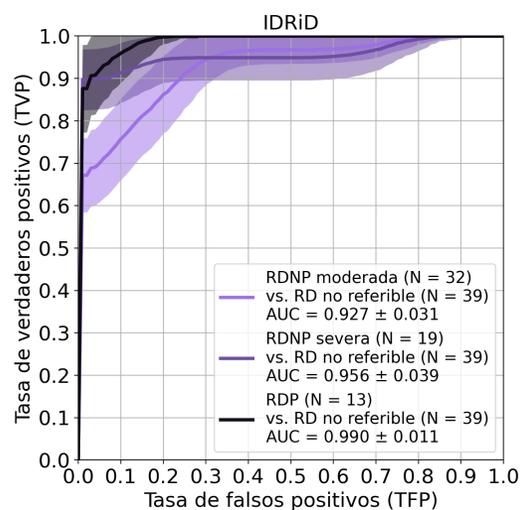
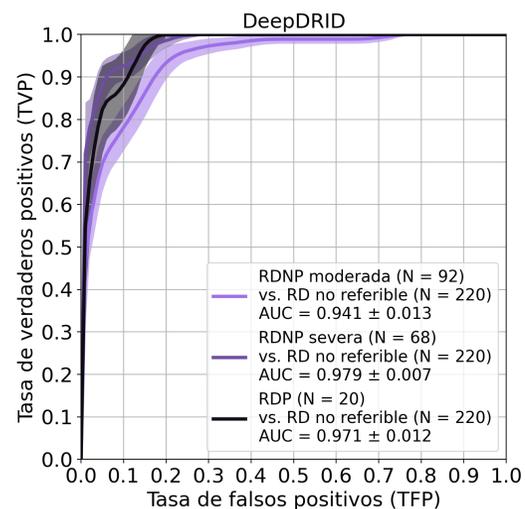
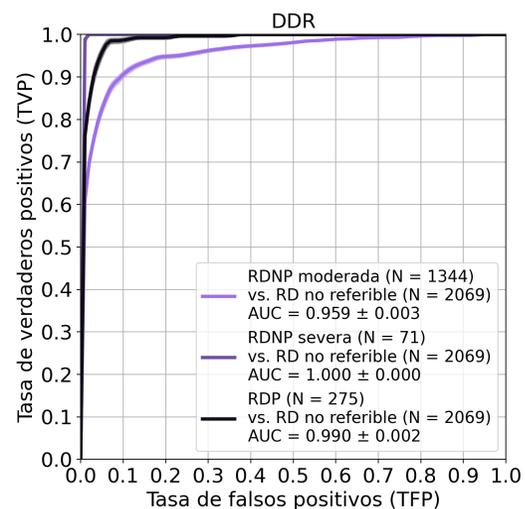
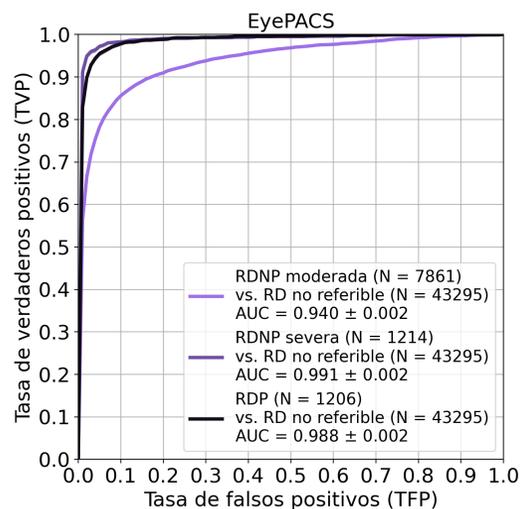
² Hervella A, Rouco J, Novo J, Ortega M. Multimodal image encoding pre-training for diabetic retinopathy grading. *Comput Biol Med*, 2022; 143: e105302.

³ Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 2017; 124(7): 962-969.

⁴ Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PloS one*, 2019; 14(6): e0217541.

⁵ Pires R, Avila S, Wainer J, Valle E, Abràmoff, MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. *Artif Intell Med*, 2019; 96: 93-106.

Resultados cuantitativos



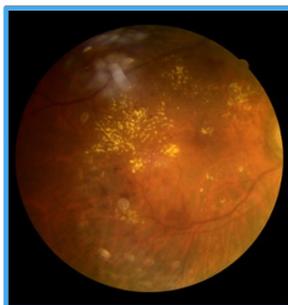
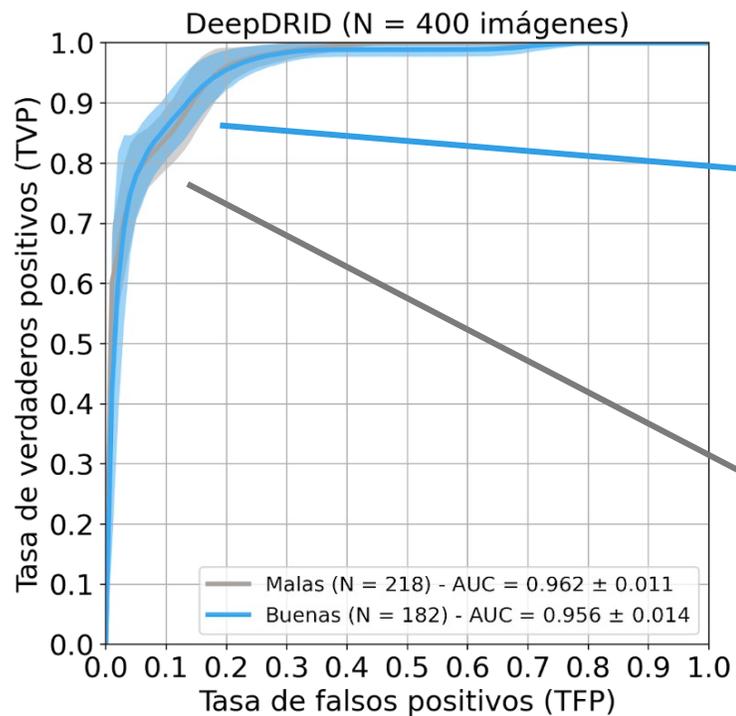
Evaluación según el grado de la retinopatía diabética

Los **grados más severos** de la enfermedad se detectan con **mayor efectividad** que los casos de **RDNP moderada**

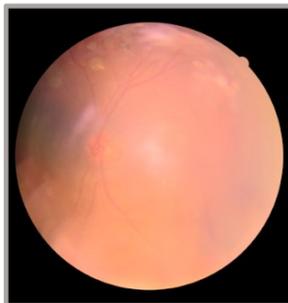
RDNP: retinopatía diabética no proliferativa
RDP: retinopatía diabética proliferativa

Resultados cuantitativos

Evaluación según la calidad de las imágenes



Buenas

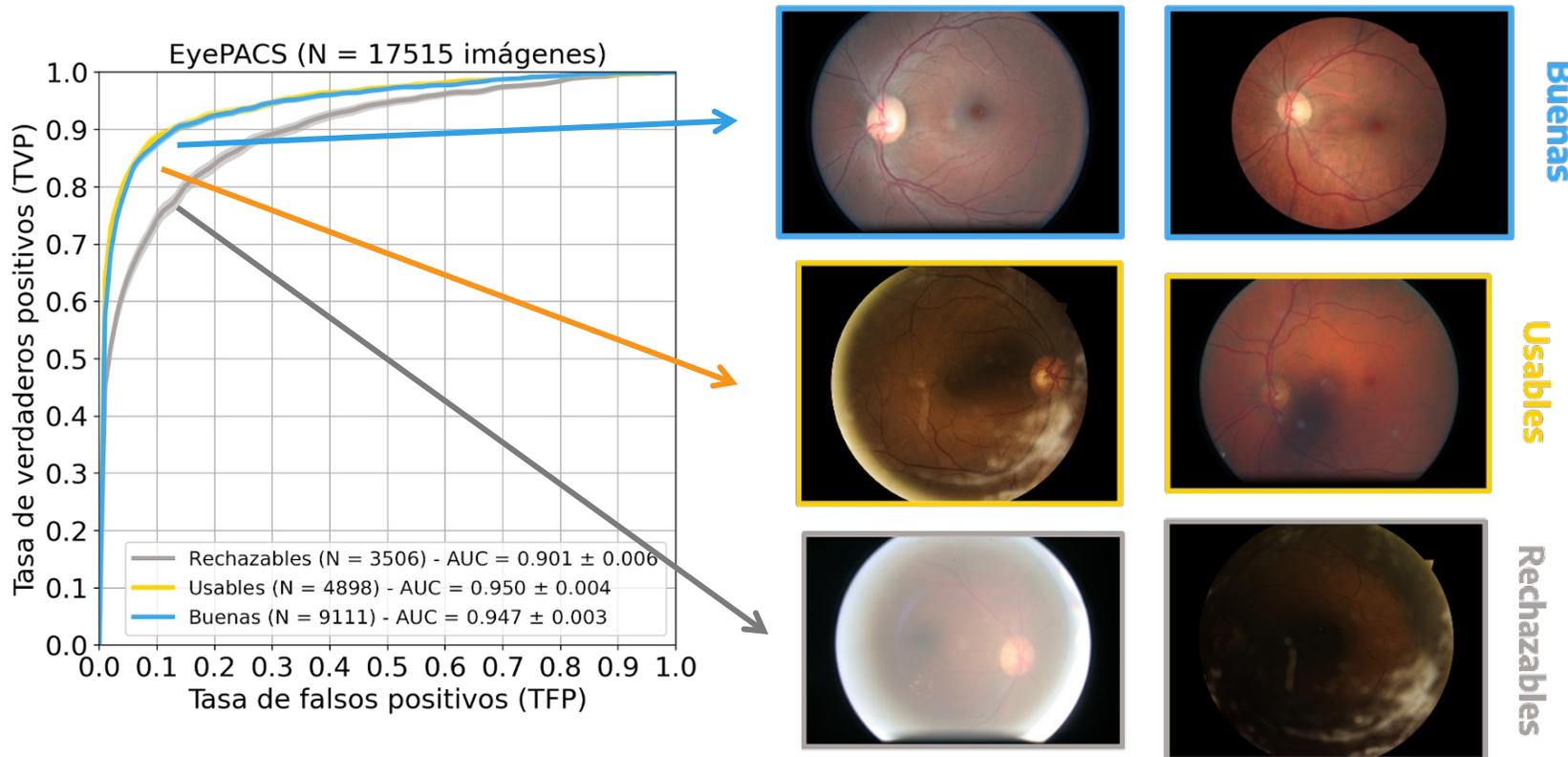


Malas

En imágenes con artefactos de captura pero en las que es posible observar el disco, la copa óptica y las arcadas vasculares, el modelo no presenta pérdidas significativas de performance.

Resultados cuantitativos

Evaluación según la calidad de las imágenes



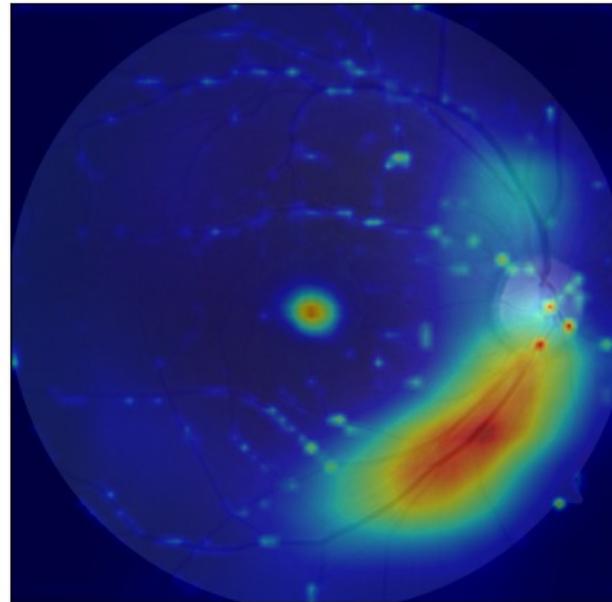
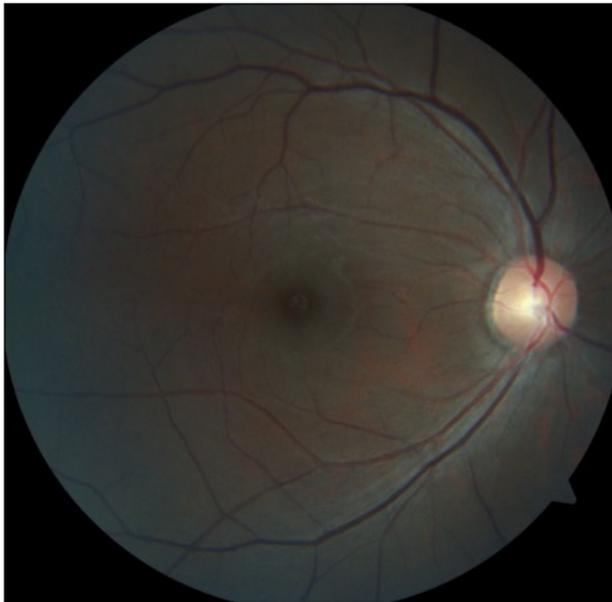
En imágenes con artefactos de captura pero en las que es posible observar el disco, la copa óptica y las arcadas vasculares, el modelo no presenta pérdidas significativas de performance.

Se observan pérdidas significativas en la capacidad de detección cuando las imágenes son de muy mala calidad

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos **no referibles**, el modelo estudia las regiones de las arcadas vasculares y de la mácula, en donde no halla lesiones.



Etiqueta real:
no referible
(sin signos de RD)

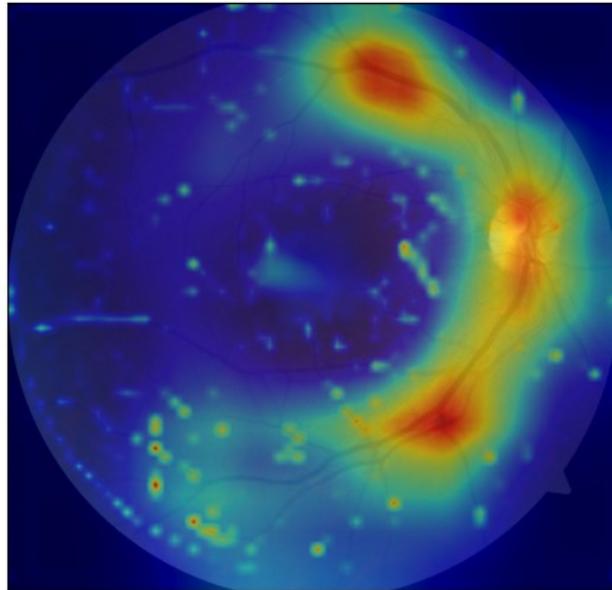
Predicción:
99.9% de probabilidad
de ser no referible

← Feedback para el
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos **no referibles**, el modelo estudia las regiones de las arcadas vasculares y de la mácula, en donde no halla lesiones.



Etiqueta real:
no referible
(RDNP leve)

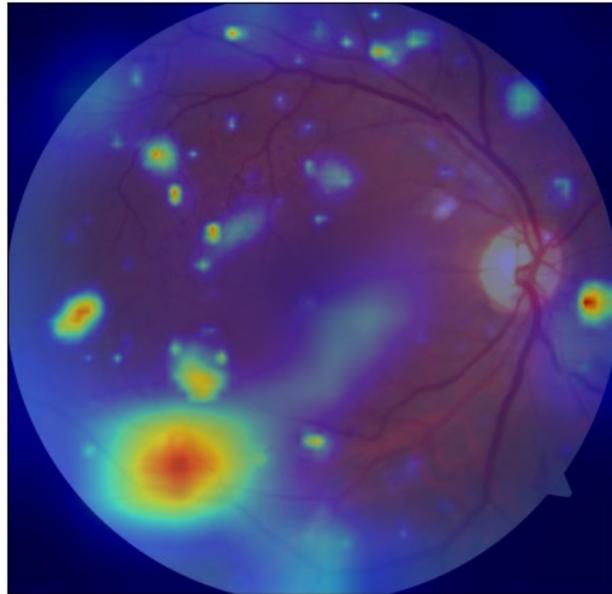
Predicción:
99.9% de probabilidad
de ser no referible

← **Feedback para el profesional**

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos que nuestra IA reconoce **referibles**, el modelo identifica lesiones típicas de la RD (en el ejemplo, **hemorragias**).



Etiqueta real:
referible
(RDNP muy severa)

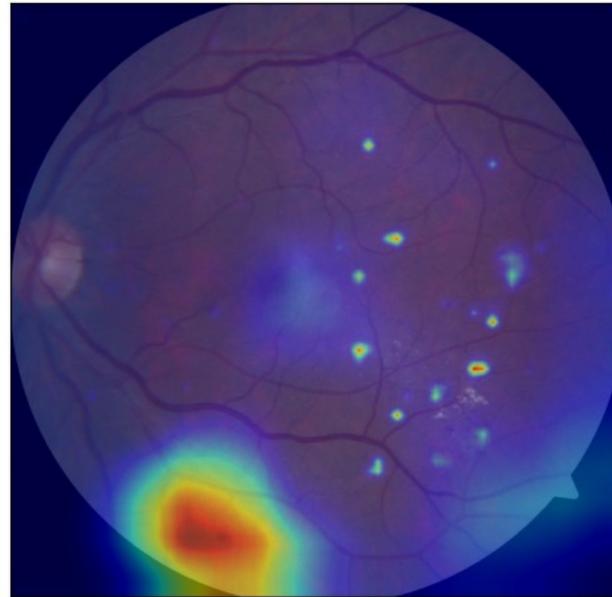
Predicción:
99.9% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos que nuestra IA reconoce **referibles**, el modelo identifica lesiones típicas de la RD (en el ejemplo, **hemorragias**).



Etiqueta real:

referible
(RDNP moderada)

Predicción:

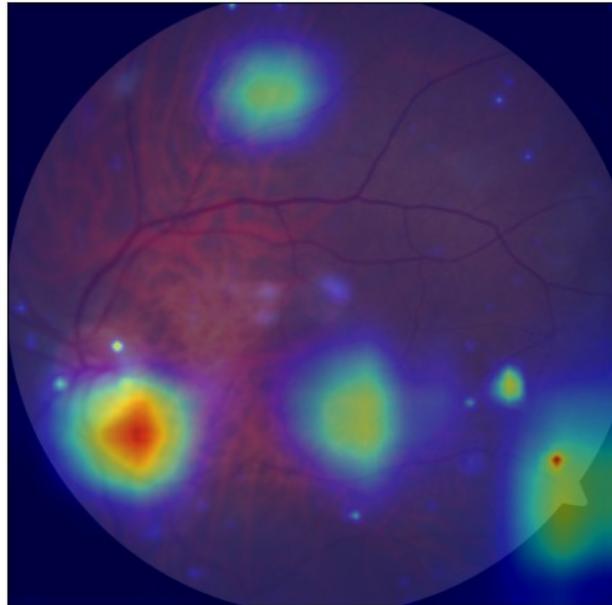
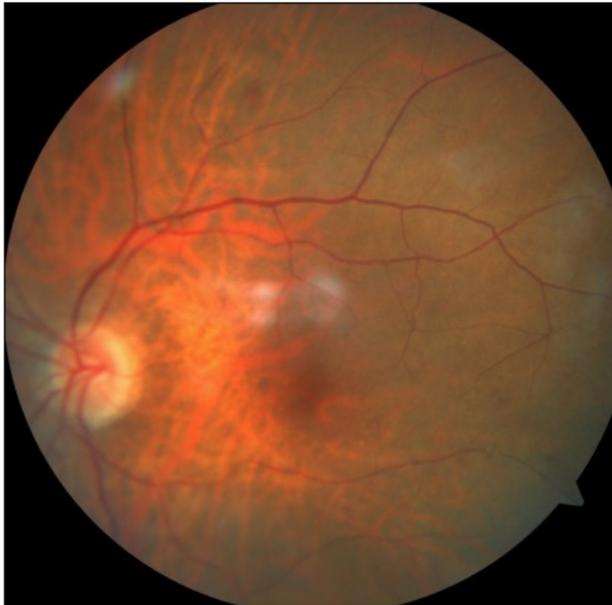
99.6% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos que nuestra IA reconoce **referibles**, el modelo identifica lesiones típicas de la RD (en el ejemplo, **hemorragias**).



Etiqueta real:

referible
(RDNP moderada)

Predicción:

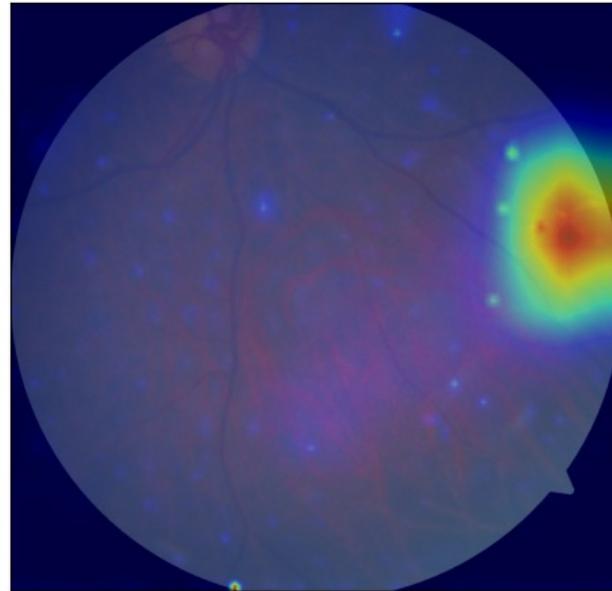
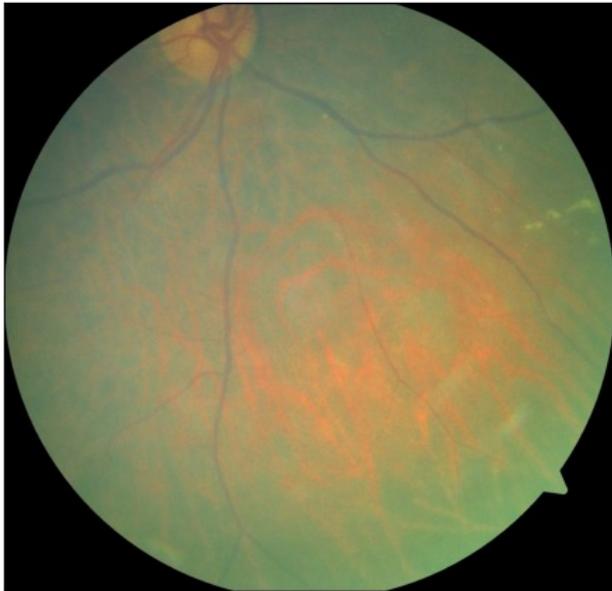
85.7% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos que nuestra IA reconoce **referibles**, el modelo identifica lesiones típicas de la RD (en el ejemplo, **exudados**).



Etiqueta real:
referible
(RDNP severa)

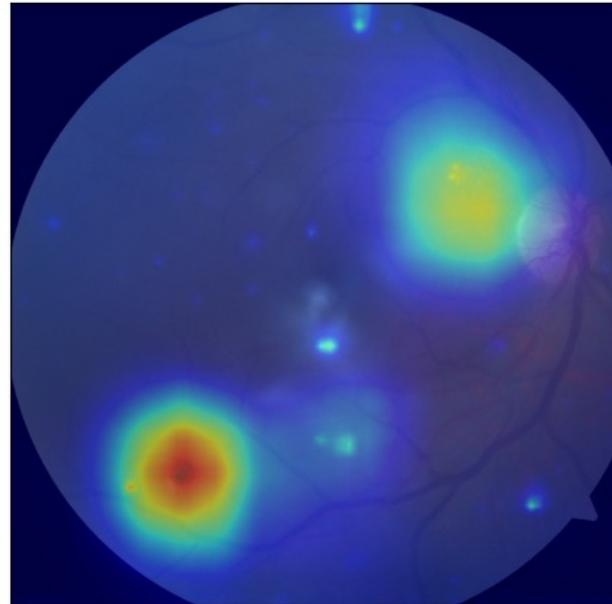
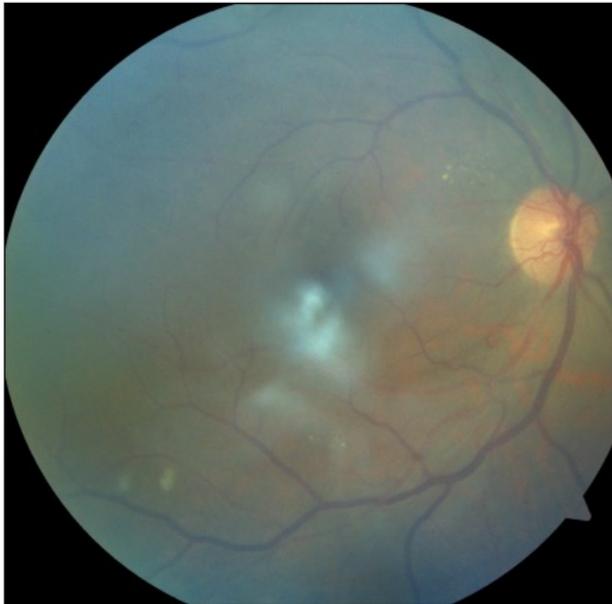
Predicción:
96.1% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En casos que nuestra IA reconoce **referibles**, el modelo identifica lesiones típicas de la RD (en el ejemplo, **exudados**).



Etiqueta real:
referible
(RDNP severa)

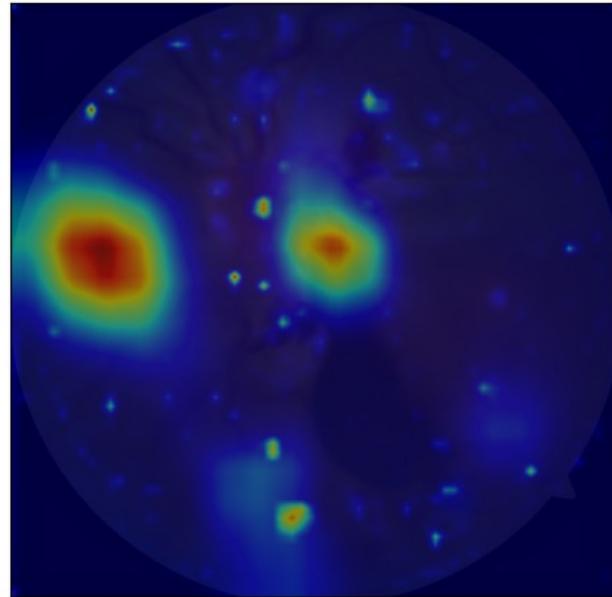
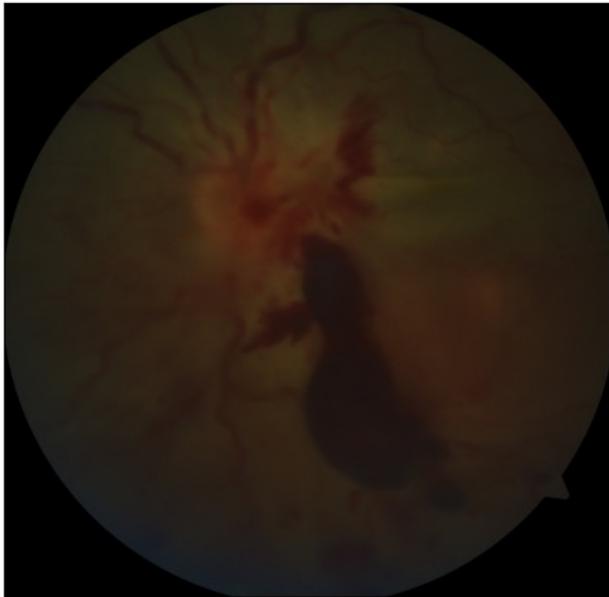
Predicción:
96.1% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Resultados cualitativos

Análisis de las regiones consideradas por el algoritmo para justificar su respuesta

En los casos **proliferativos**, el modelo **no detecta neovascularizaciones** o **hemorragias muy grandes**



Etiqueta real:
referible
(RDP)

Predicción:
99.7% de probabilidad
de ser RD referible

← **Feedback para el**
profesional

Conclusiones

El modelo propuesto demostró:

- **Valores altos de sensibilidad y especificidad para detectar casos referibles**
Impacto positivo en el tamizado de los casos de riesgo
- **Robustez a la presencia de artefactos de captura**
Mayor aplicabilidad del modelo en un contexto real de telemedicina
- **Robustez a cambios en el dispositivo de adquisición**
Mayor aplicabilidad del modelo utilizando la capacidad instalada del sistema médico actual
- **Respuestas cualitativas compatibles con las observaciones clínicas**
Mayor seguridad de la herramienta.

Limitaciones del estudio:

- **Evaluación en casos con otras patologías o enfermedades concomitantes**
¿Detecta enfermedades con lesiones similares a las de la RD? ¿Qué ocurre con las que no son parecidas?
- **Malos resultados en imágenes de muy mala calidad**
Es necesario detectarlas previamente o pueden dar respuestas equívocas.
- **Análisis de incerteza**
¿Qué ocurre en los casos en los que el algoritmo presenta probabilidades de salida cercanas al 50%?

Trabajos futuros

En los próximos meses trabajaremos incorporando al modelo:

- Detección de casos de edema macular diabético.
- Detección de RDP como una clase adicional, para mejorar la identificación de neovascularizaciones.
- Entradas auxiliares (otras mejoras de contraste, segmentaciones vasculares, localizaciones del disco y la fovea)

Estudio nacional a mayor escala

- Utilizar imágenes capturadas en nuestro país para entrenar el modelo y mejorar aún más los resultados.
- Realizar evaluaciones con datos de otras instituciones clínicas y profesionales argentinos.

¿Les interesa colaborar con nosotrxs?

¡Necesitamos más imágenes y más colaboradores clínicxs!



jjorlando@pladema.exa.unicen.edu.ar



[@ignaciorlando](https://twitter.com/ignaciorlando)



<https://ignaciorlando.github.io>

retinar

¡Muchas gracias por su atención!
¿Preguntas?



Resumen extendido
y slides aquí:



¿Les interesa colaborar con nosotrxs?
¡Necesitamos más imágenes y más colaboradores clínicxs!



jjorlando@pladema.exa.unicen.edu.ar



[@ignaciorlando](https://twitter.com/ignaciorlando)



<https://ignaciorlando.github.io>

Evaluación retrospectiva de un primer modelo de inteligencia artificial argentino para tamizaje automático de retinopatía diabética referible a partir de fotografías de fondo de ojo

Dr. Ing. José Ignacio Orlando, Ing. Tomás Castilla, Dr. Alejandro Koch, Dr. Ignacio Larrabide, Dra. Marcela Martínez y Dra. Mercedes Leguía



UNICEN
Universidad Nacional del Centro
de la Provincia de Buenos Aires



martínez
CENTRO DE OFTALMOLOGÍA



ALTA COMPLEJIDAD EN RED
Hospital El Cruce
Dr. Néstor Carlos Kirchner

