

Diagnóstico asistido con algoritmos de clasificación: Caso de retinopatía diabética

Dalia Camacho García Formentí

Colaboradores

Dr. Alejandro Noriega Campero (MIT)

Dr. Abdullah Almaatouq (MIT)

Dra. Daniela Meizner (APEC)

Rami Manna (MIT)

Houssam Kherraz (MIT)

Contenido

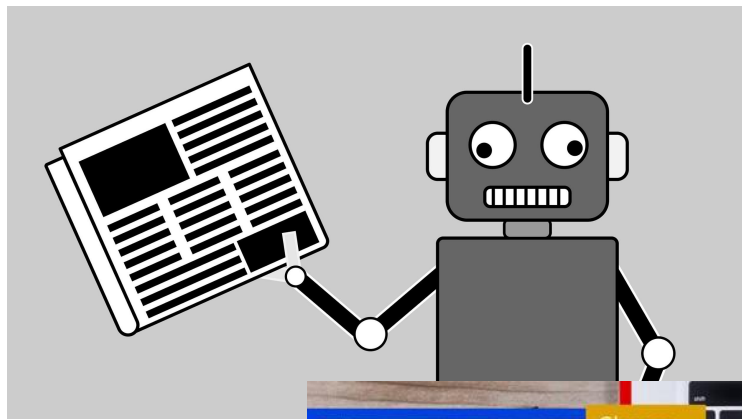
- **Introducción**
 - ¿Puedo confiar en los algoritmos?
 - Retinopatía diabética (RD).
- **Objetivo del proyecto**
- **Metodología**
 - Algoritmo de clasificación y métodos de explicación
 - Experimento
- **Resultados**
- **Conclusiones**
- **Referencias**



1. Introducción



Importancia de los algoritmos



¿Puedo confiar
en ellos?

¿Lobo o Husky?



Clasificador de nieve



(a) Husky classified as wolf



(b) Explanation

(Ribeiro 2016)

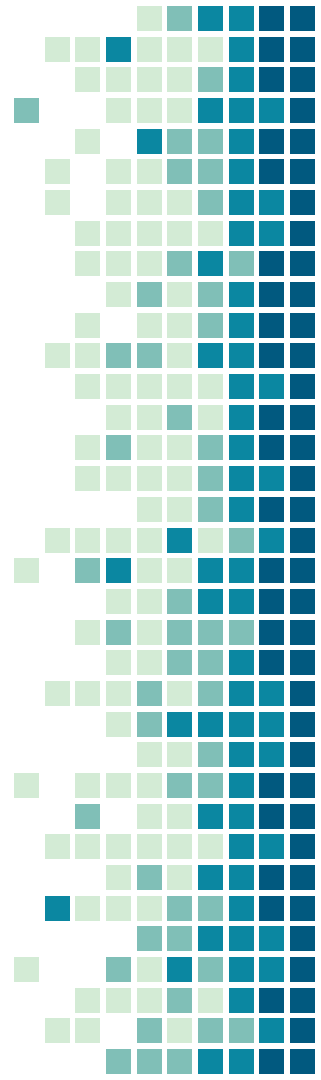
¿Y en diagnósticos médicos?

En 2015 Caruana buscó conocer la probabilidad de muerte de pacientes con neumonía

- Modelo 1: AUC=0.77
- Modelo 2: AUC=0.86



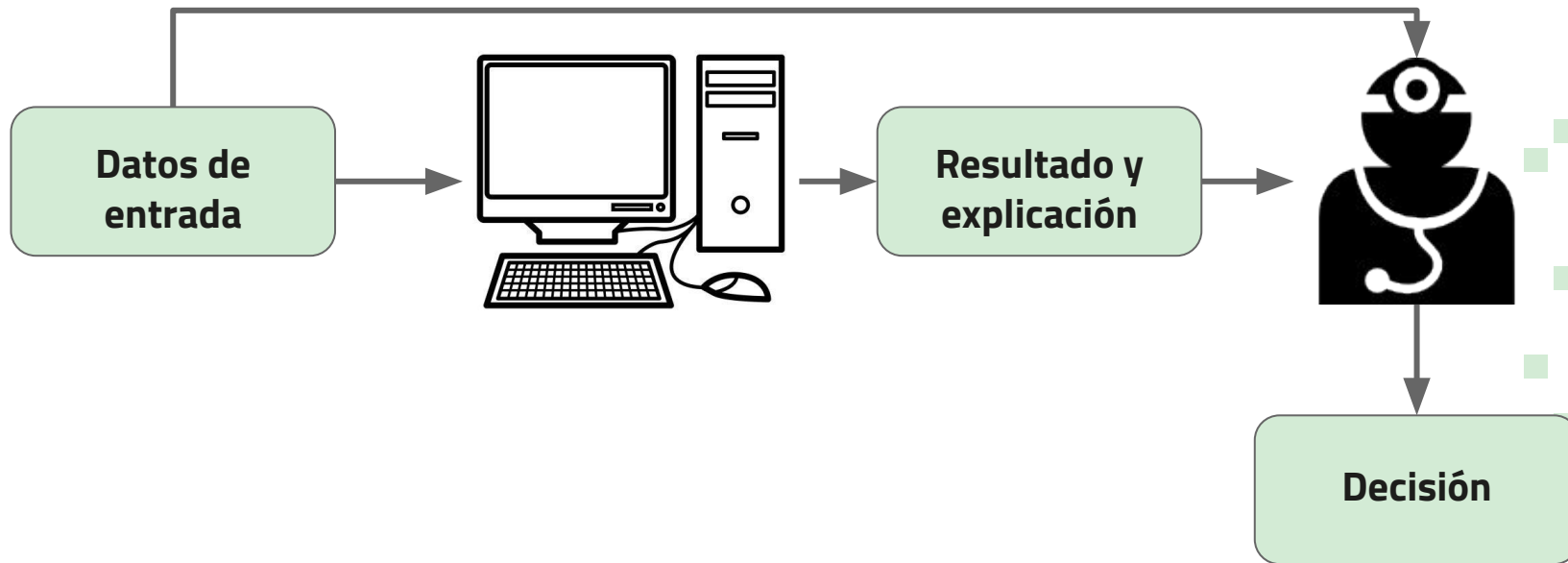
**Es necesario generar
explicaciones del
funcionamiento de los
algoritmos**



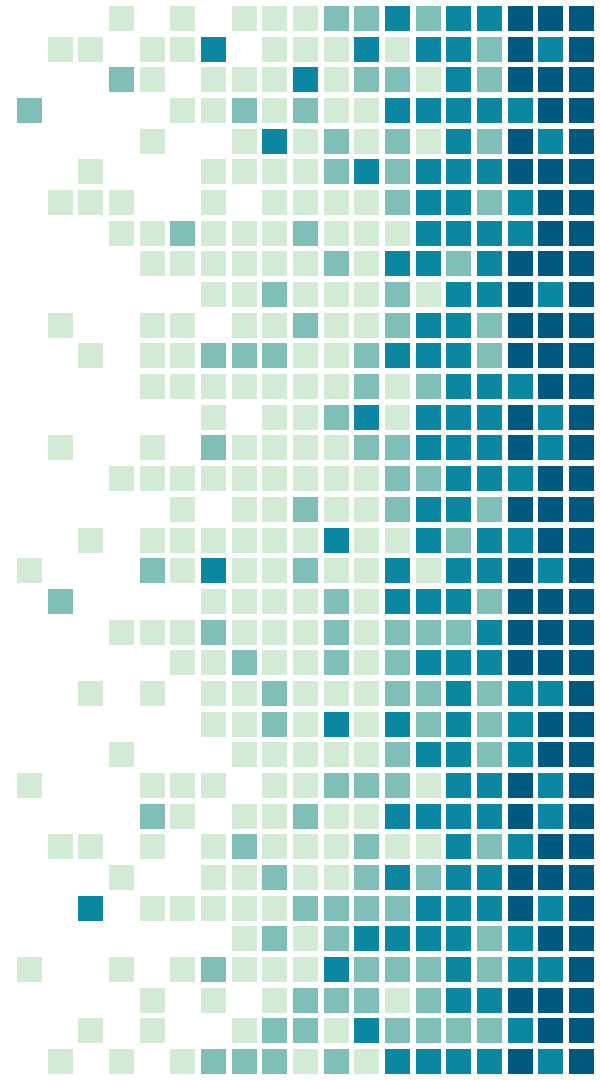
Las predicciones de un
algoritmo son correctas...

Pero no queremos que la
decisión dependa sólo del
algoritmo.

Esquema de decisión



1.2. Retinopatía Diabética



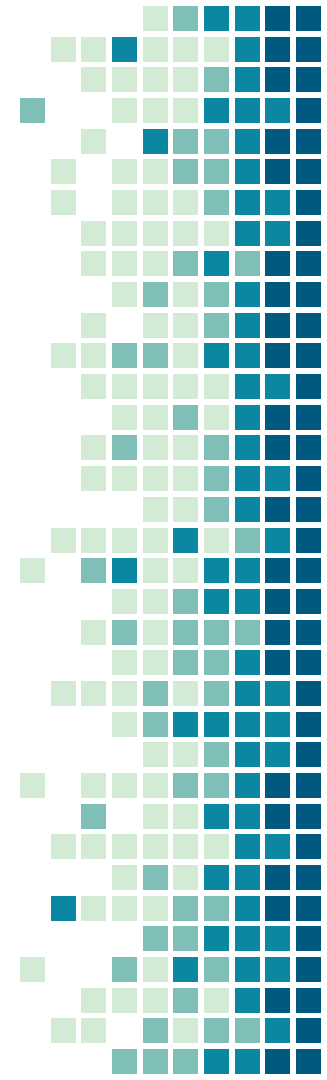
Diabetes



2014

8.5%

(OMS 2018)



Diabetes



2016

9.4% diagnosticados
(ENSANUT MC 2016)

2030

11.97%-18.29%
(Meza, 2015)



Retinopatía Diabética

2016

11% de individuos con diabetes diagnosticada

≈1% de la población

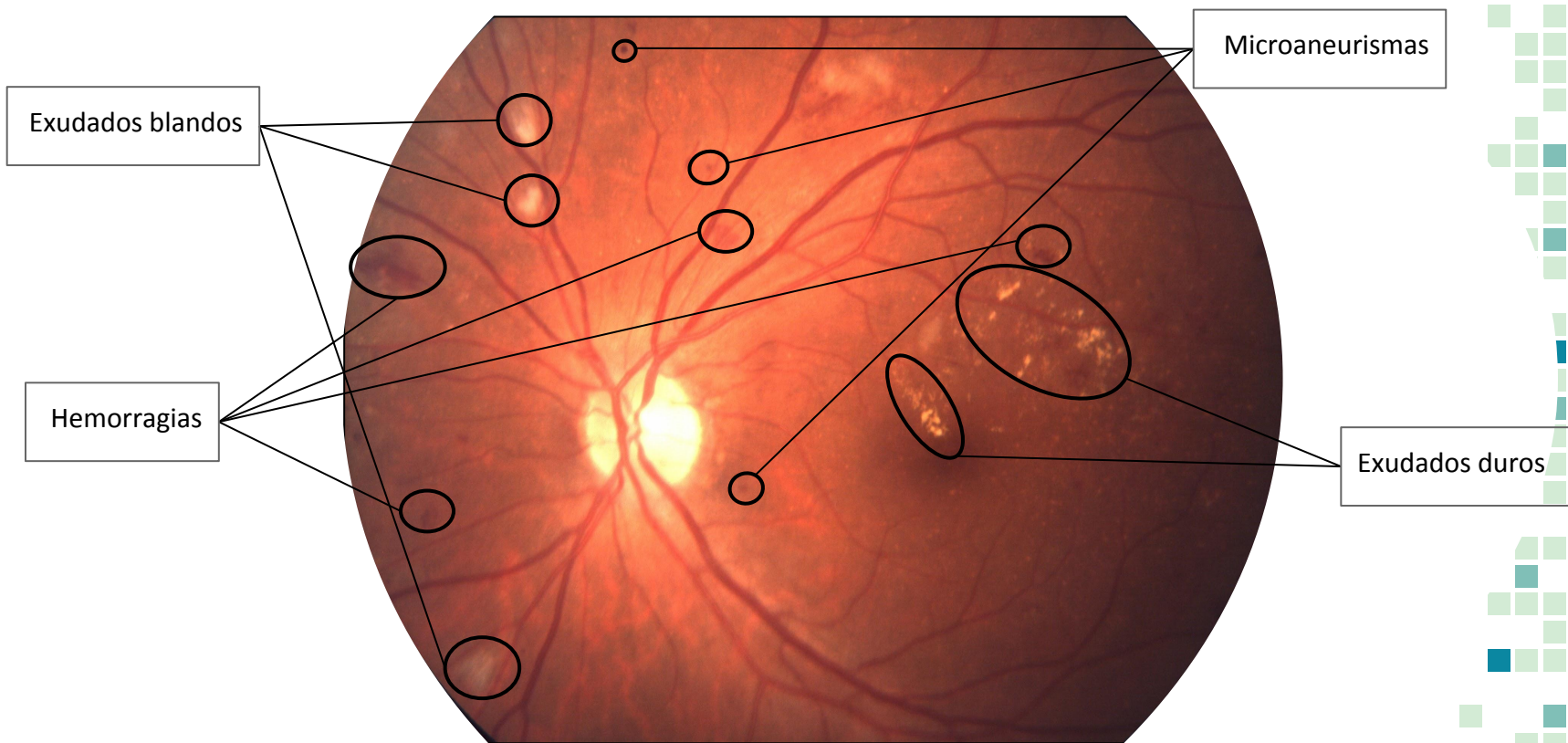
(ENSANUT MC 2016)

Segunda causa de ceguera y primera causa de deficiencia visual en mayores de 50 años

(López-Star 2018)



Lesiones

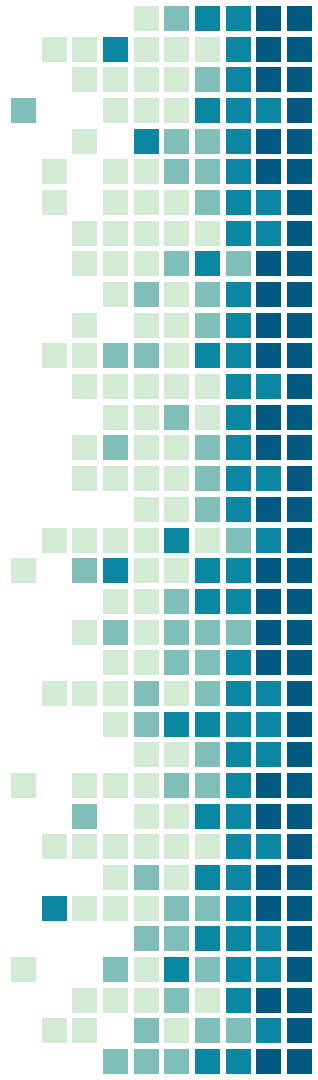


2. Objetivo del proyecto



Objetivo

Encontrar la mejor forma de presentar el resultado del algoritmo y su explicación para que los médicos realicen diagnósticos más acertados.

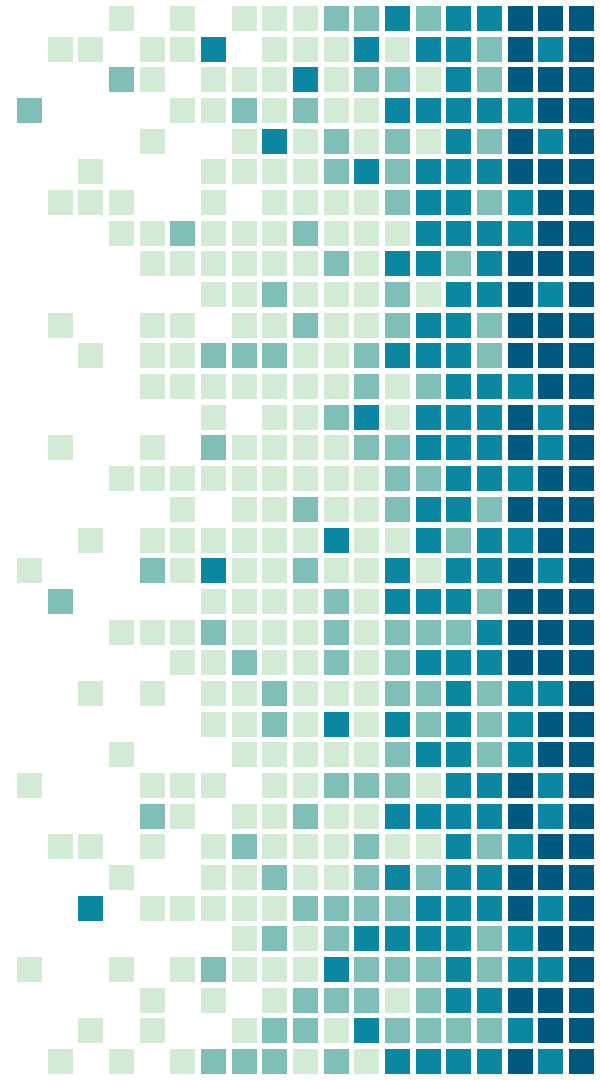


Preguntas intermedias

- **¿El desempeño de los médicos es mejor antes o después de conocer los resultados del algoritmo?**
- **¿Es mejor presentar la explicación o solamente el resultado?**
- **¿Cuál es la percepción que tienen sobre el algoritmo y su utilidad?**



3. Metodología



¿Con qué voy a
predecir?

Datos

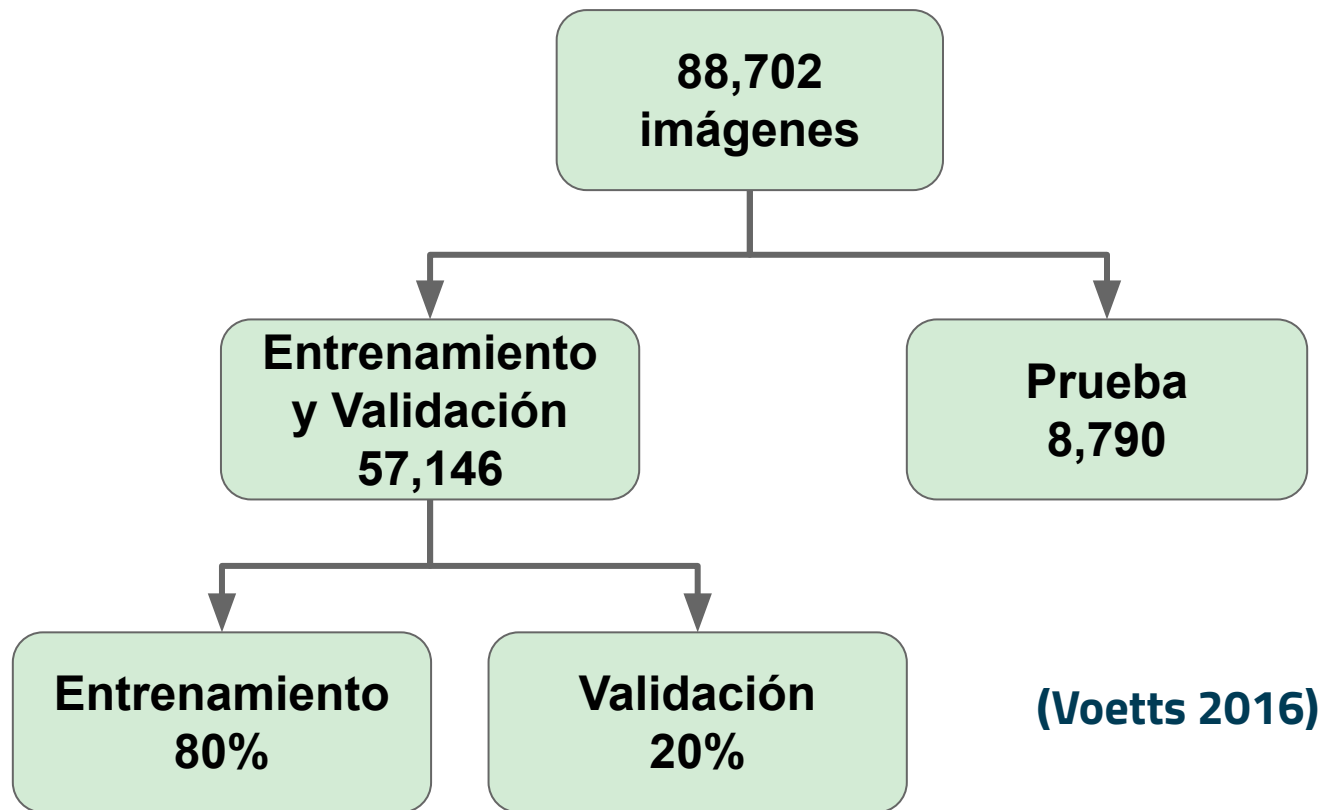
	EyePACS	APEC
RD no referible	71,584	112
RD referible	17,154	103
Total	88,702	215

Entrenamiento
Validación
Prueba

Prueba



Datos: EyePACS



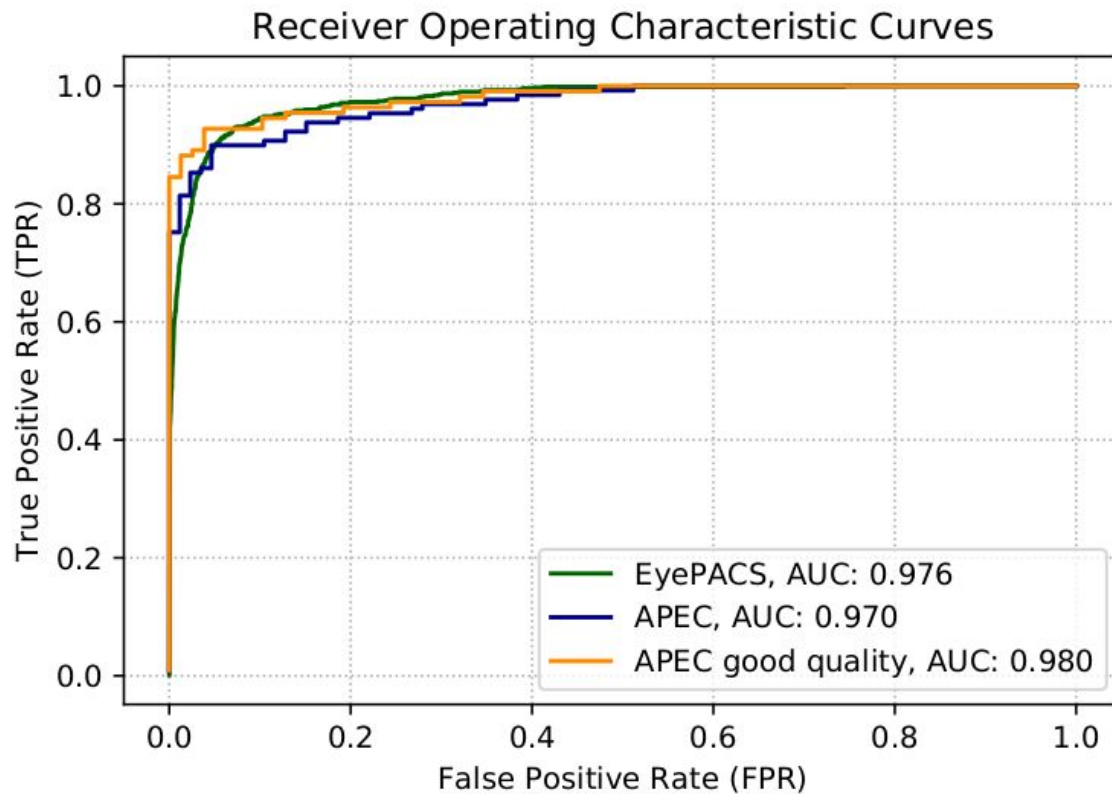
(Voetts 2016)

26

Modelo	ROC-AUC
1	0.974
2	0.960
3	0.969
Ensamble	0.977

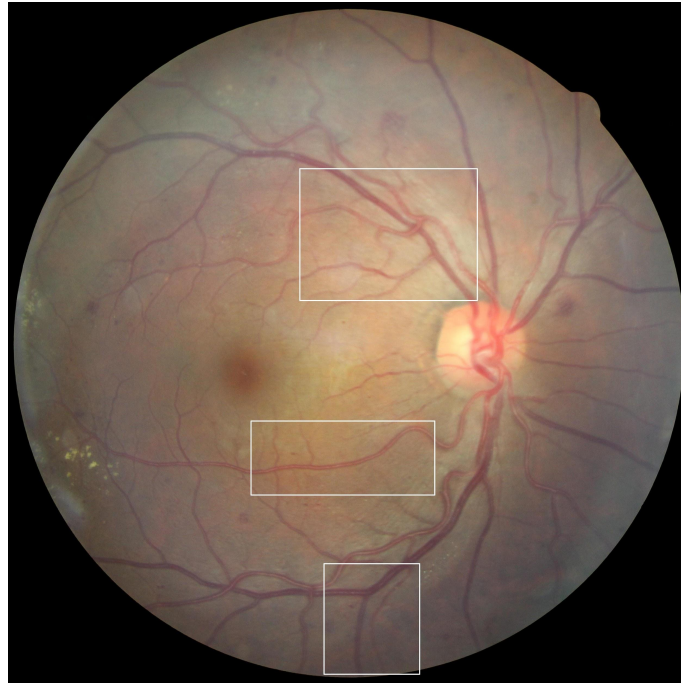
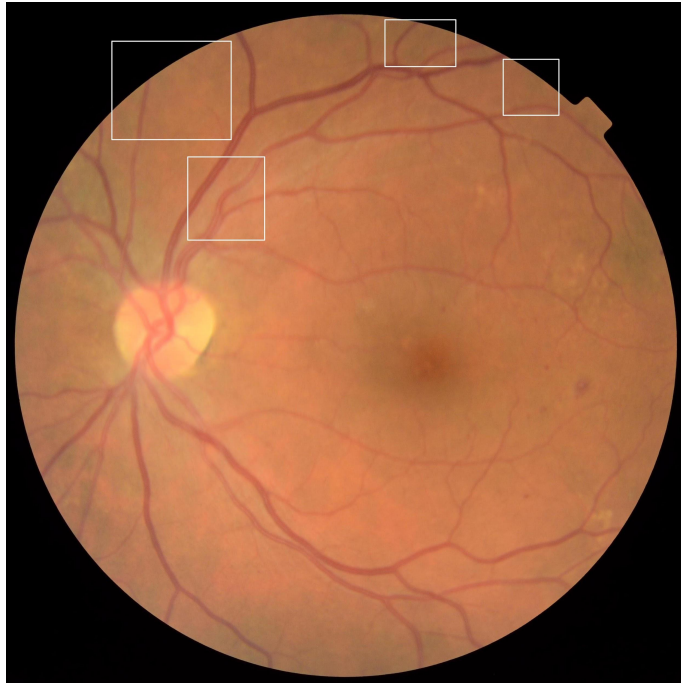


Resultados del modelo en APEC

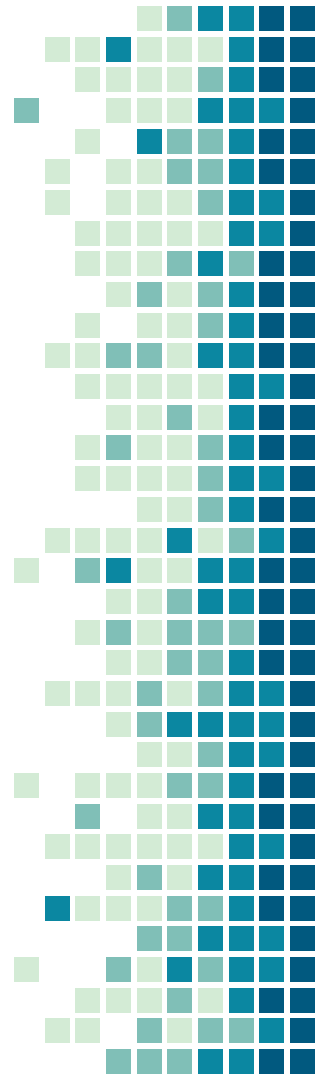
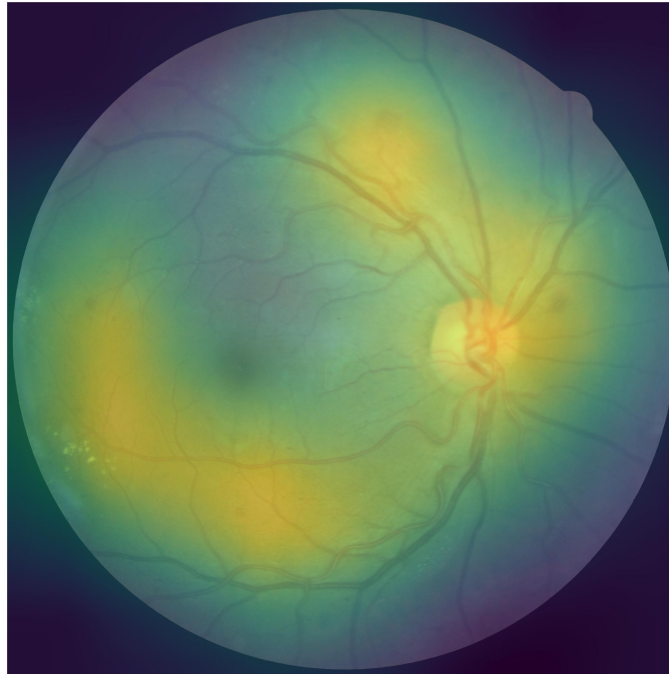
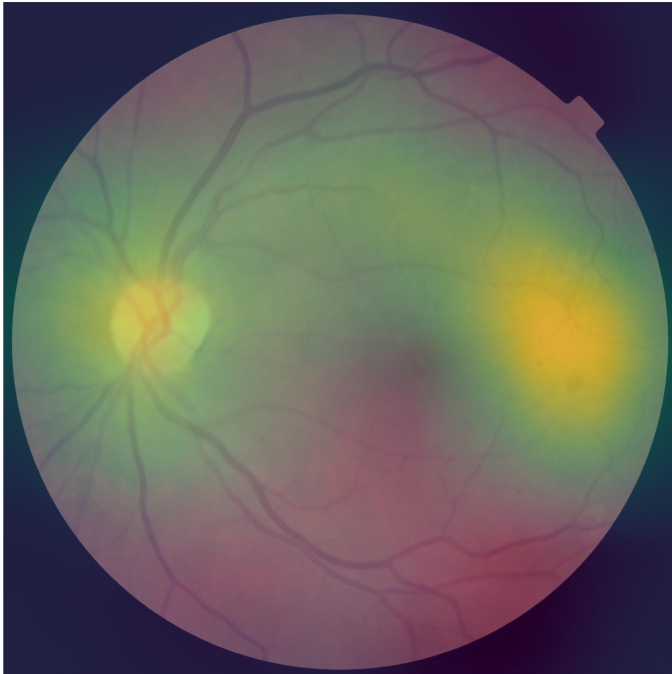


¿Y la
explicación?

Explicaciones interpretables localmente agnósticas a modelos: LIME



Ventanas deslizantes



Propagación de relevancia a través de las capas: LRP



Experimento

Participantes

- **17 estudiantes de la subespecialidad de retina en la APEC**

Año de especialidad	Participantes
2	3
3	12
4	2

Esquemas de tratamiento

- **Solitario**
- **Respuesta binaria**
- **Mapa de calor y respuesta numérica**

4. Resultados



Desempeño general

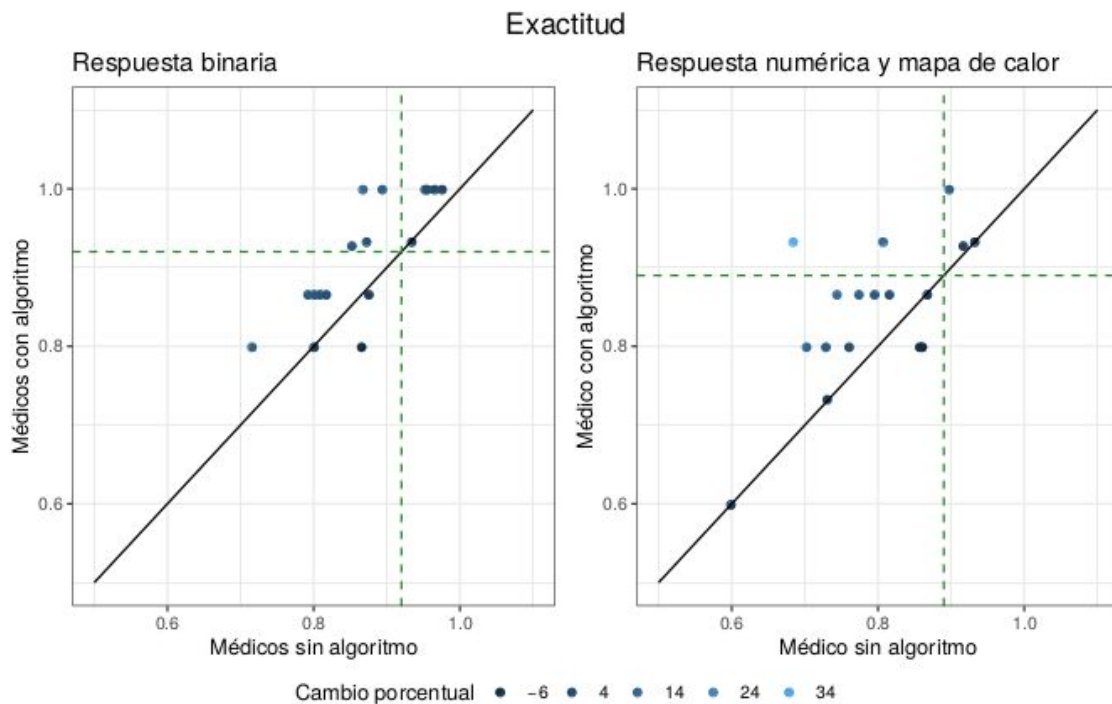
Modalidad	Algoritmo	Médicos	Médicos con algoritmo	Mejora porcentual
Todas	0.92 (0.90, 0.925)	0.82 (0.80, 0.85)	-	-
Solitario	0.93 (0.91, 0.96)	0.87 (0.83, 0.91)	-	-
Respuesta binaria	0.92 (0.89, 0.95)	0.87 (0.83, 0.90)	0.91 (0.88, 0.95)	5.53 (2.91, 8.02)
Respuesta numérica y mapa de calor	0.89 (0.85, 0.92)	0.79 (0.75, 0.83)	0.85 (0.8, 0.89)	7.39 (2.91, 12.46)

Apreciación de los médicos

Modalidad	No	A veces	Sí	NA
Respuesta binaria	1	4	8	4
Respuesta numérica y mapas de calor	3	1	7	6



Desempeño individual



“ *I think that it could be a useful screening tool for technicians for a quick referral.* ”



“ *It doesn't recognize the
vascular proliferations near
the nerve.* ”



“ For me, the most useful element would be the attention maps. The confidence percentages are ambiguous



5. Conclusiones



Preguntas intermedias

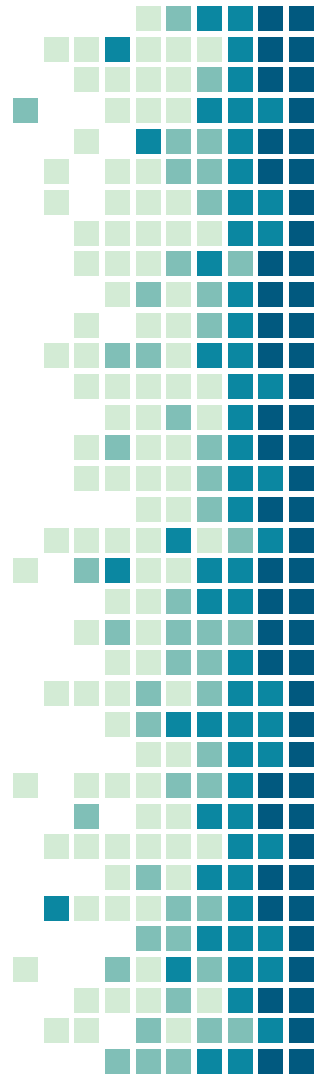
- **¿El desempeño de los médicos es mejor antes o después de conocer los resultados del algoritmo? El desempeño es mejor después de conocer los resultados.**
- **¿Es mejor presentar la explicación o solamente el resultado? Parece ser que el presentar el mapa de calor da mejores resultados, pero no es estadísticamente mejor que el resultado binario.**
- **¿Cuál es la percepción que tienen sobre el algoritmo y su utilidad? La mayoría creen que les ayudó en al menos algunos casos y que es una buena herramienta, aunque no es perfecta.**



Objetivo

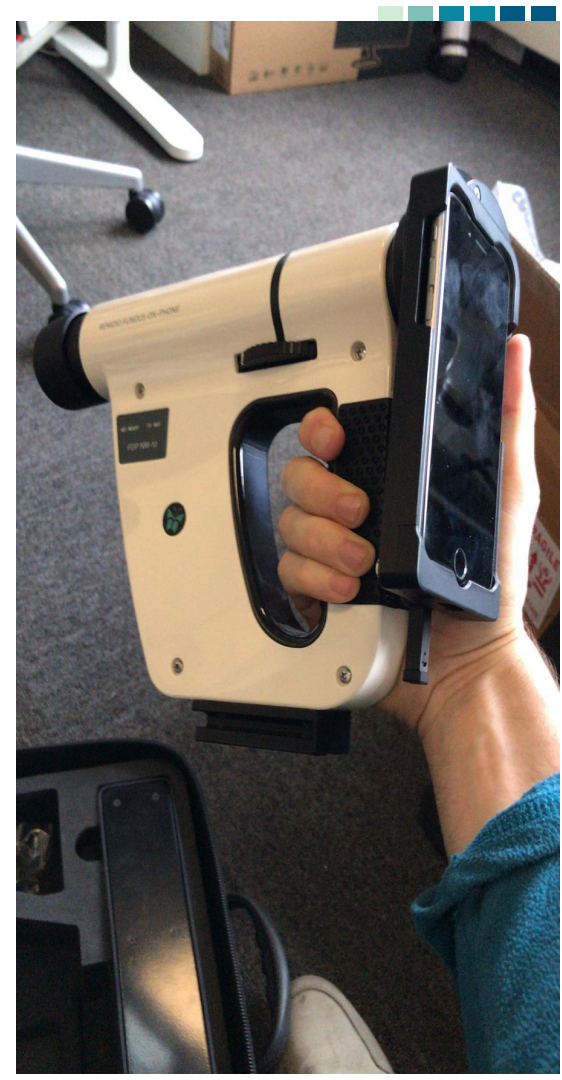
Encontrar la mejor forma de presentar el resultado del algoritmo y su explicación para que los médicos realicen diagnósticos más acertados.

- **Es mejor presentar los resultados del algoritmo que no hacerlo.**
- **El mapa de calor es útil y les da mayor información sobre cómo funciona el algoritmo.**
- **Los niveles de confianza son confusos.**
- **Combinar mapas de calor con respuesta binaria (escenario no evaluado)**



Trabajo futuro

- Definir un algoritmo que distinga imágenes de buena calidad y de mala calidad.
- Extender el algoritmo de clasificación de RD a otras enfermedades de la retina haciendo uso de *transfer learning*.
- Llevar el diagnóstico a zonas donde no hay médicos especialistas en retina.
- Llevar esta herramienta a los médicos para que cuenten con más información al momento de hacer el diagnóstico.



6. Referencias



Referencias

- Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. ArXiv, 2016.
<http://arxiv.org/abs/1602.04938>.
- Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, pages 1721–1730, New York, NY, USA, 2015. ACM.
<http://doi.acm.org/10.1145/2783258.2788613>.
- World Health Organization. Diabetes, Oct 2018.
<https://www.who.int/es/news-room/fact-sheets/detail/diabetes>.
- Rosalba Rojas-Martínez, Ana Basto-Abreu, Carlos A Aguilar-Salinas, Emiliano Zárate-Rojas, Salvador Villalpando, and Tonatiuh Barrientos-Gutiérrez. Prevalencia de diabetes por diagnóstico médico previo en México. 2018. Salud Pública de México, 60:224 – 232, 06
http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342018000300003&nrm=iso.

Referencias

- R. Meza, T. Barrientos-Gutierrez, R. Rojas-Martinez, N. Reynoso-Noveron, L. S. Palacio-Mejia, E. Lazcano-Ponce, and M. Hernandez-Avila. Burden of type 2 diabetes in Mexico: past, current and future prevalence and incidence rates. *Prev Med*, 81:445–450, Dec 2015.
- EM López-Star, K Allison-Eckert, H Limburg, I Brea-Rodríguez, and VC Lansingh. Evaluación rápida de la ceguera evitable, incluida la retinopatía diabética, en Querétaro, México. *Rev Mex Oftalmol*, 92(2):8493, marzo-abril 2018
<https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=85759>.
- Kaggle. Diabetic retinopathy detection.
<https://www.kaggle.com/c/diabetic-retinopathy-detection/data>.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.
- Mike Voets, Kajsa Mollersen, and Lars Ailo Bongo. Replication study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *CoRR*, abs/1803.04337, 2018. <http://arxiv.org/abs/1803.04337>.

Referencias

- **Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, 2017.**

