



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

Departamento de Ingeniería Eléctrica
Sección De Electrónica Del Estado Sólido

“Sistema Integrado en Hardware para el Análisis y Procesamiento Básico de Imágenes Digitales”

T E S I S

Que presenta:

Ivan Shin Cao Chong Cervantes

Directores de tesis:

Dr. Mario Alfredo Reyes Barranca
Dr. José Antonio Moreno Cadenas

Ciudad de México

Junio, 2024



En memoria del
**Dr. José Antonio
Moreno Cadenas**

Contenido

- Objetivos
- Introducción
- Marco teórico
- Procesamiento de imágenes
- Aprendizaje Profundo
- Resultados
- Conclusiones
- Trabajo a futuro

Objetivo

Implementación de un sistema embebido en una Raspberry Pi 4 para procesar y analizar imágenes digitales empleando lógica difusa y redes neuronales.

Objetivos particulares

- ✓ Desarrollo de un sistema de inferencia difuso para el procesamiento de mejora de contraste en imágenes digitales.
- ✓ Extracción de características de imágenes de fondo de ojo para discernir entre imágenes con retinopatía diabética y sin retinopatía diabética.
- ✓ Implementación del sistema de inferencia difuso para la mejora de contraste de imágenes digitales en la Raspberry Pi 4.
- ✓ Implementación de la red neuronal para la extracción de características de imágenes de fondo de ojo en la Raspberry Pi 4.

Introducción

Motivación



Retinografía



Antecedentes



Recursos computacionales

Procesamiento de imágenes

Ha tomado un importante protagonismo diversas áreas.

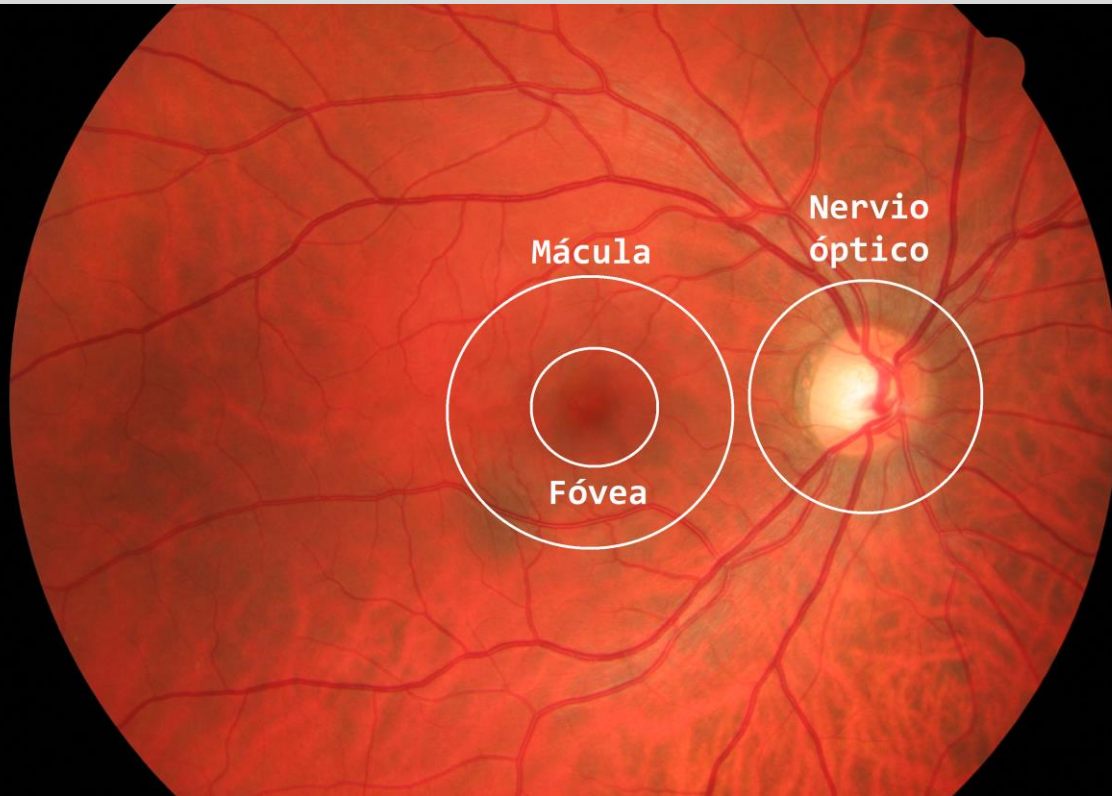
- Ciencia
- Industria
- Recreación
- Medicina

Solución de problemas

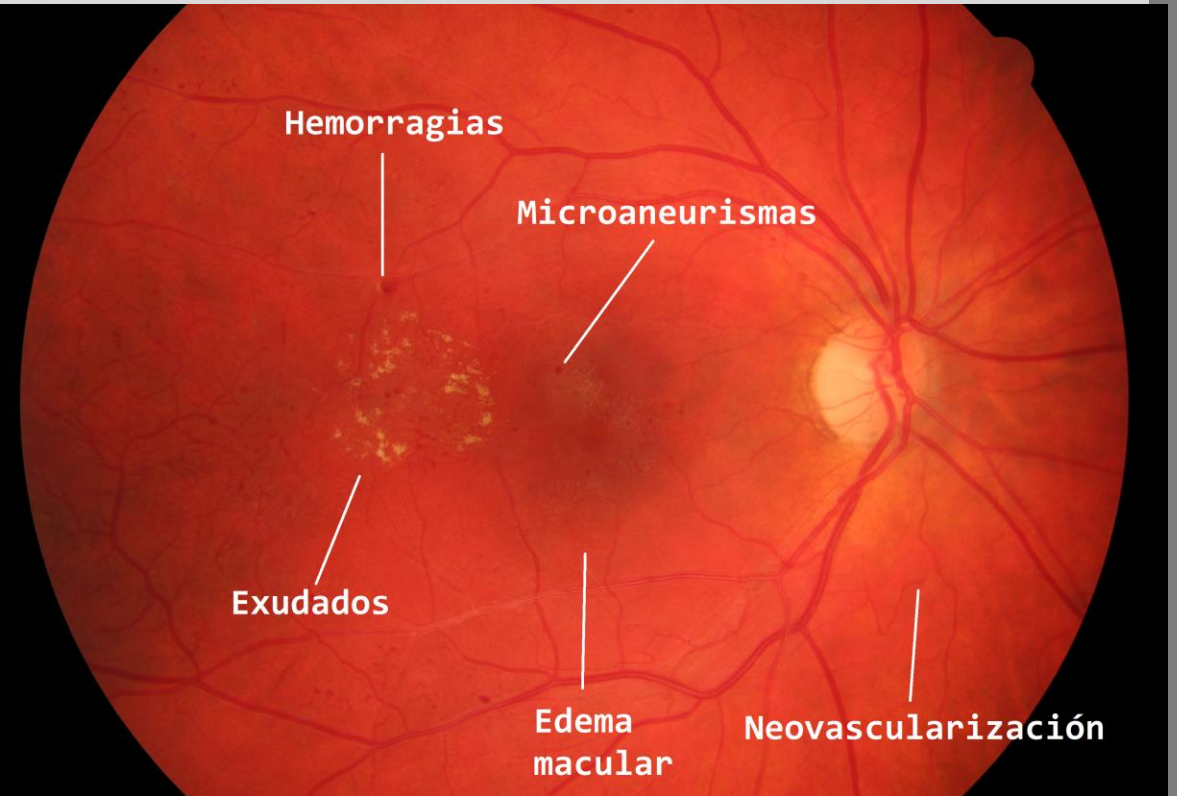
La retinopatía diabética es la principal causa de ceguera en México.

- Detección por medio de exámenes invasivos
- Retinografía

Imagen de fondo de ojo



Retinopatía diabética



- Adquisición de imágenes de fondo de ojo (1970)
- Mejora en la zona del nervio óptico (1989)
- Mejora de imagen retinal a color usando CLAHE (2013)
- Mejora de imagen retinal a color basado en ajuste de luminosidad y contraste (2019)
- Detección de retinopatía diabética y clasificación con lógica difusa (2019)
- Mejora de imagen retinal usando filtro pasa bajas y enraizamiento α (2020)

Recursos computacionales

Software y hardware

Computadora personal (PC)

- Procesador AMD Rizen 5 3500U

Raspberry Pi 4 modelo B

- Sistema operativo basado en kernel de Linux

Python

- Interfaz de desarrollo Spyder
- Google Colab (Computación en la nube)

Base de datos

- HRF-Universidad Friedrich-Alexander Universität Erlangen-Nürnberg, 45 imágenes
- MESSIDOR-ADCIS, 1200 imágenes
- Kaggle, 53586 imágenes

Marco teórico

Raspberry Pi 4 modelo B



Imágenes digitales



Lógica difusa



Redes neuronales

Raspberry Pi 4 modelo B

Computador de placa única de bajo costo y tamaño.

SoC (System on a Chip) BCM2711

- CPU: Quad-core Cortex-A72 (ARM v8) 64-bit @1.5 GHz
- GPU: Broadcom VideoCore VI 32-bit @ 500MHz
- RAM: LPDDR4-2400 SDRAM
- Caché: 32kB data + 48kB instruction L1 cache per core. 1MB L2 cache.



<pre> 00000000000000000000 J8 10000000000000000000 PoE Wi-Fi Pi Model 4B V1.4 D SoC RAM USB3 S I pwr hd hd C S USB2 m0 m1 I A V </pre>		<pre> Revision : d03114 SoC : BCM2711 RAM : 8GB Storage : MicroSD USB ports : 4 (of which 2 USB3) Ethernet ports : 1 (1000Mbps max. speed) Wi-fi : True Bluetooth : True Camera ports (CSI) : 1 Display ports (DSI) : 1 </pre>	
		<pre> J8: 3V3 (1) (2) 5V GPIO2 (3) (4) 5V GPIO3 (5) (6) GND GPIO4 (7) (8) GPIO14 GND (9) (10) GPIO15 GPIO17 (11) (12) GPIO18 GPIO27 (13) (14) GND GPIO22 (15) (16) GPIO23 3V3 (17) (18) GPIO24 GPIO10 (19) (20) GND GPIO9 (21) (22) GPIO25 GPIO11 (23) (24) GPIO8 GND (25) (26) GPIO7 GPIO0 (27) (28) GPIO1 GPIO5 (29) (30) GND GPIO6 (31) (32) GPIO12 GPIO13 (33) (34) GND GPIO19 (35) (36) GPIO16 GPIO26 (37) (38) GPIO20 GND (39) (40) GPIO21 </pre>	
		<pre> POE: TR01 (1) (2) TR00 TR03 (3) (4) TR02 </pre>	

Imagen digital

Definición

Representación bidimensional de una escena.

$$f(x, y) = \begin{bmatrix} f(0,0) & \dots & f(0, N-1) \\ \vdots & \ddots & \vdots \\ f(M-1,0) & \dots & f(M-1, N-1) \end{bmatrix}$$

Las matrices numéricas contienen y organizan la información de acuerdo con el espacio de color.

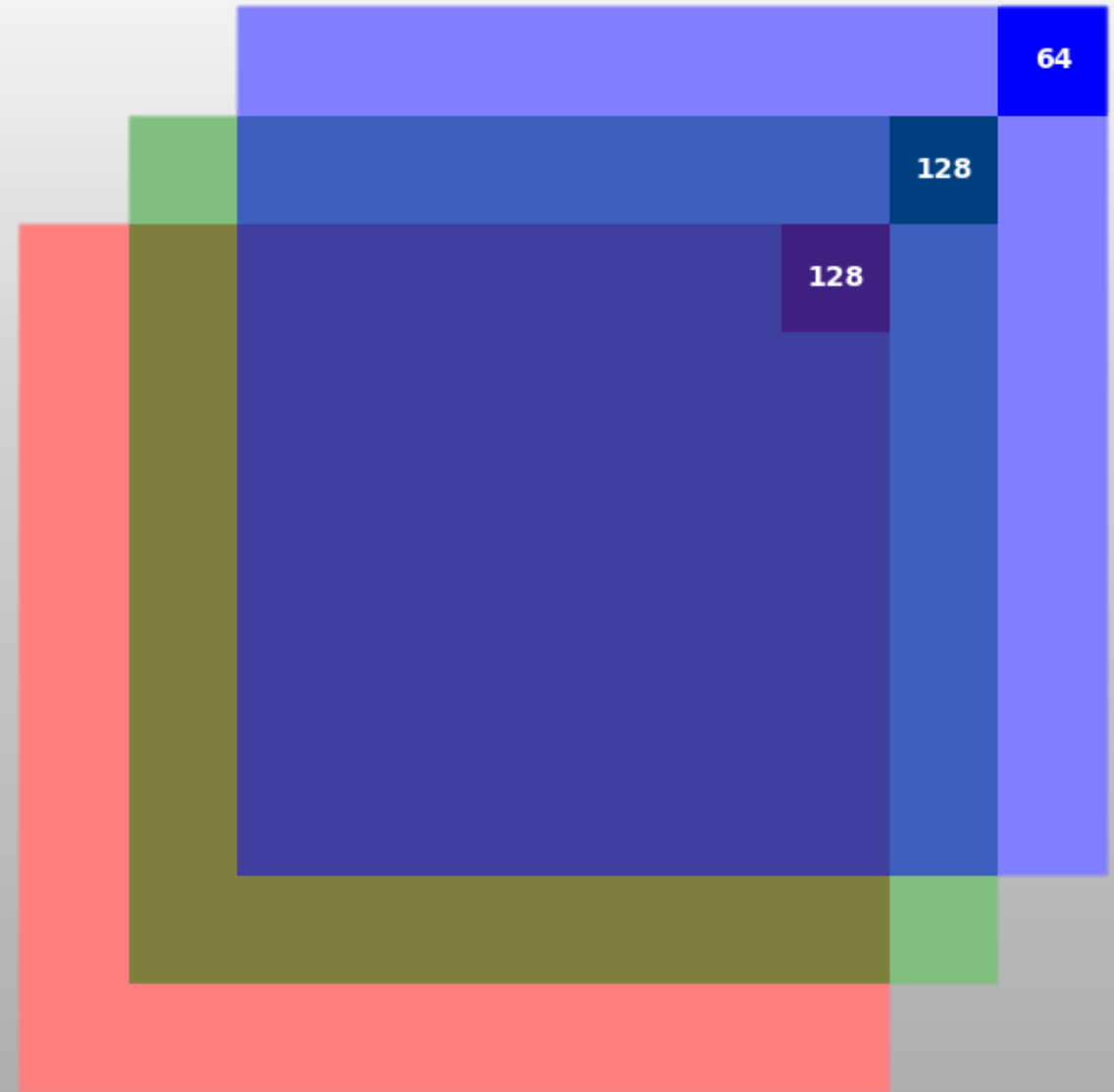


Imagen digital

Contraste

Variación de intensidades entre los píxeles de una imagen.

$$\sigma = SD = \sqrt{\frac{1}{MN} \sum_{j=0}^{N-1} \sum_{k=0}^{M-1} (i_{j,k} - \bar{i})^2}$$

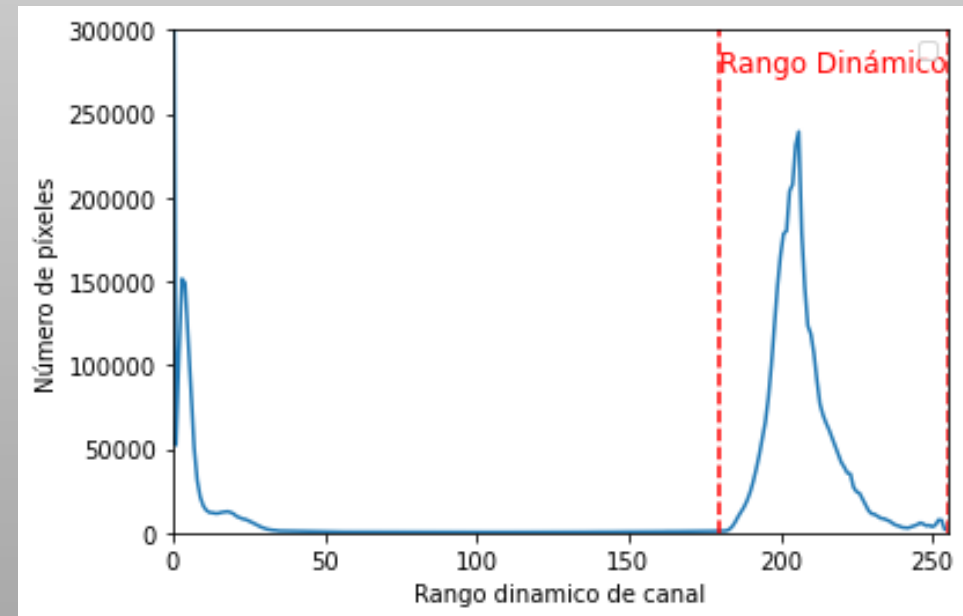
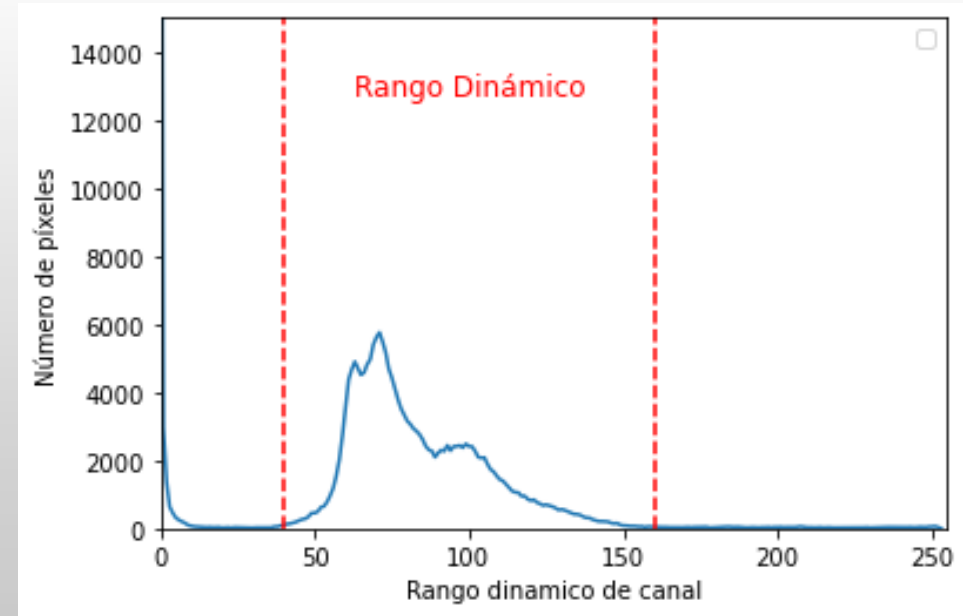
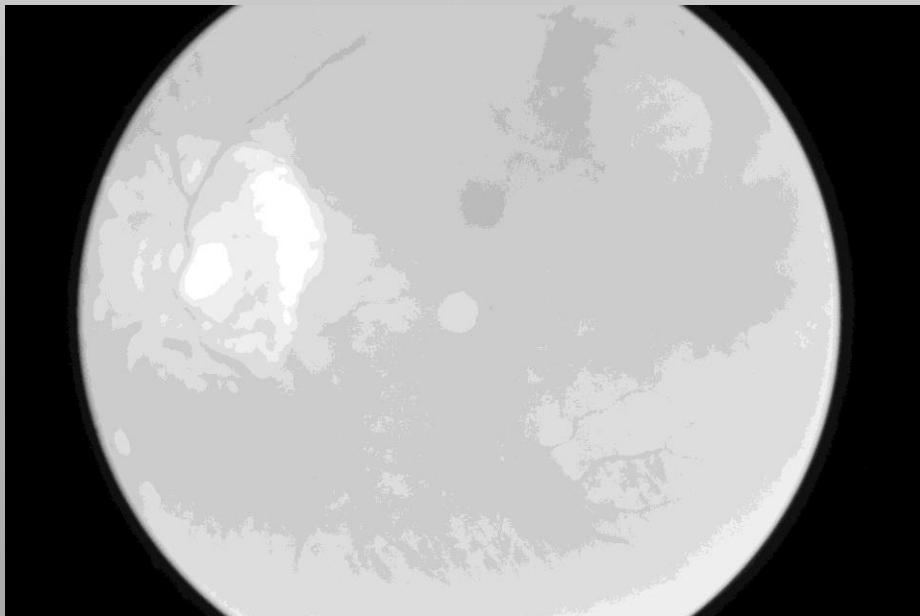
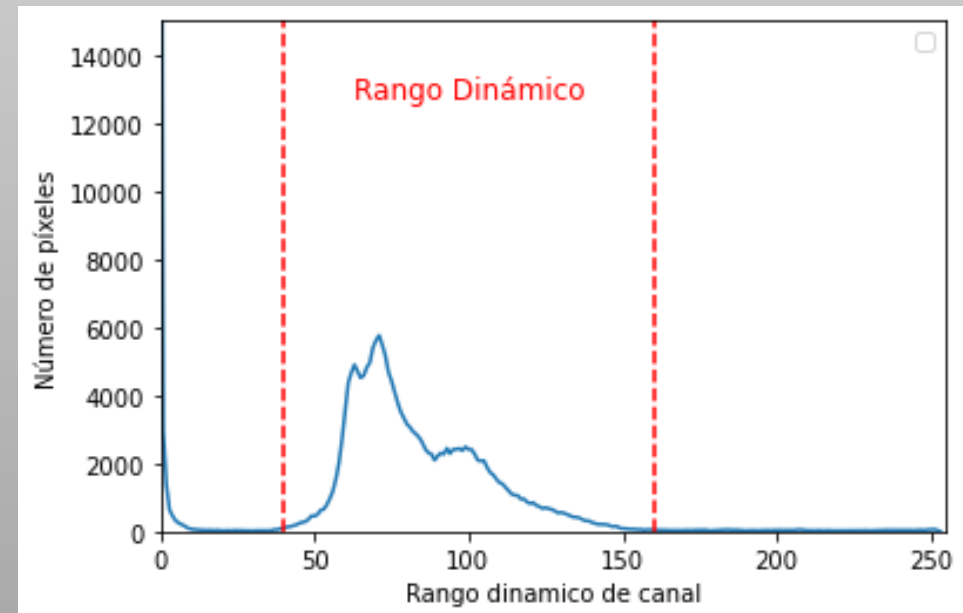
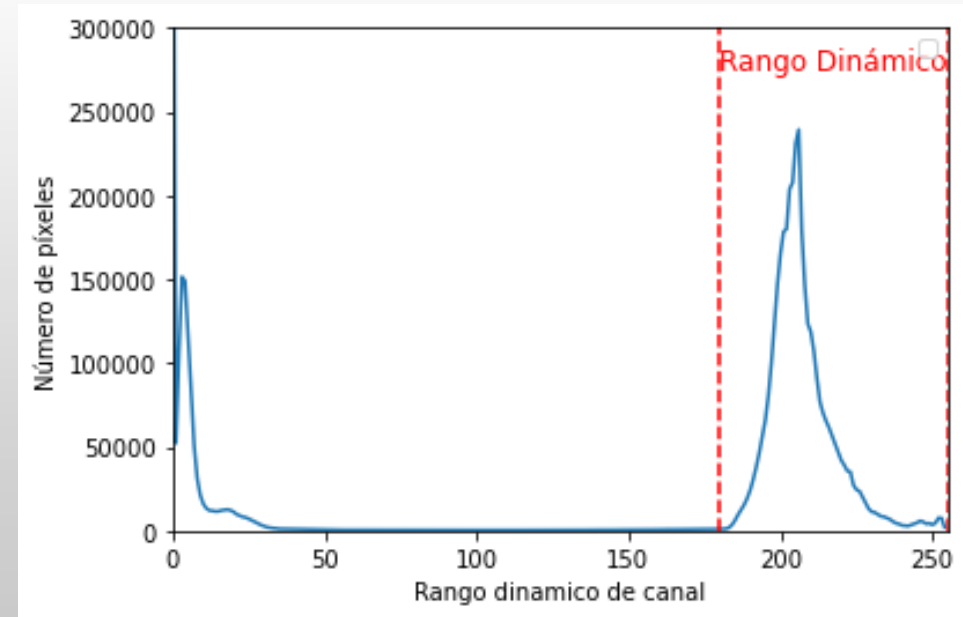
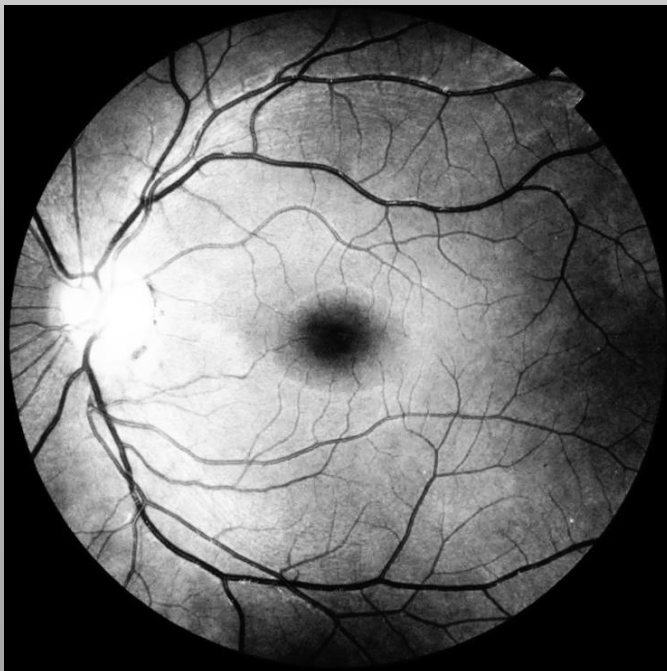


Imagen digital

Contraste

Variación de intensidades entre los píxeles de una imagen.

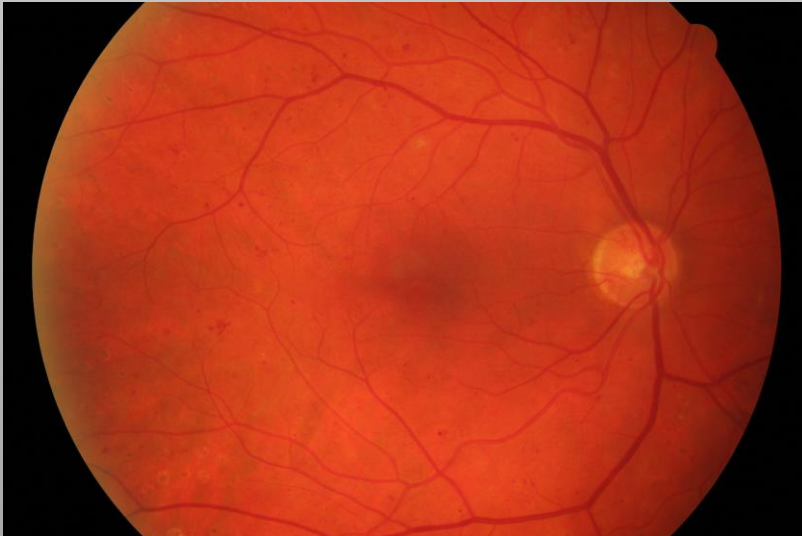
$$\sigma = SD = \sqrt{\frac{1}{MN} \sum_{j=0}^{N-1} \sum_{k=0}^{M-1} (i_{j,k} - \bar{i})^2}$$



Procesamiento de imágenes

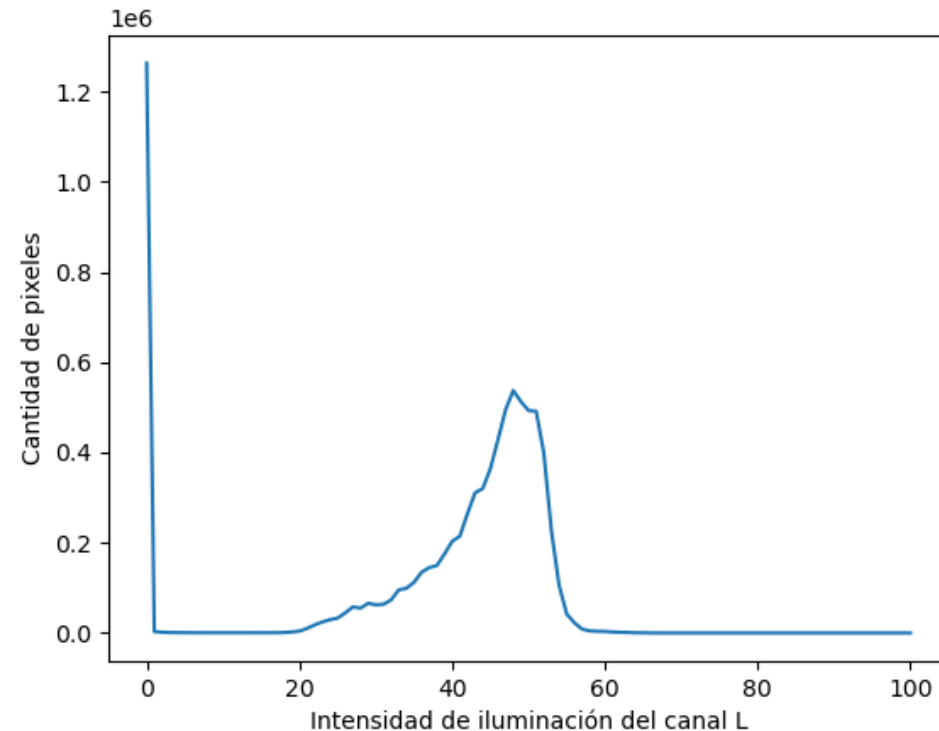
Técnicas y algoritmos para mejorar su calidad o facilitar su análisis e interpretación

- Transformación de intensidad
 - Mejora de iluminación
 - Mejora de contraste
- Filtros espaciales



Mejora de contraste

Procesamiento de imagen para incrementar el rango dinámico de una imagen.



Procesamiento de imágenes

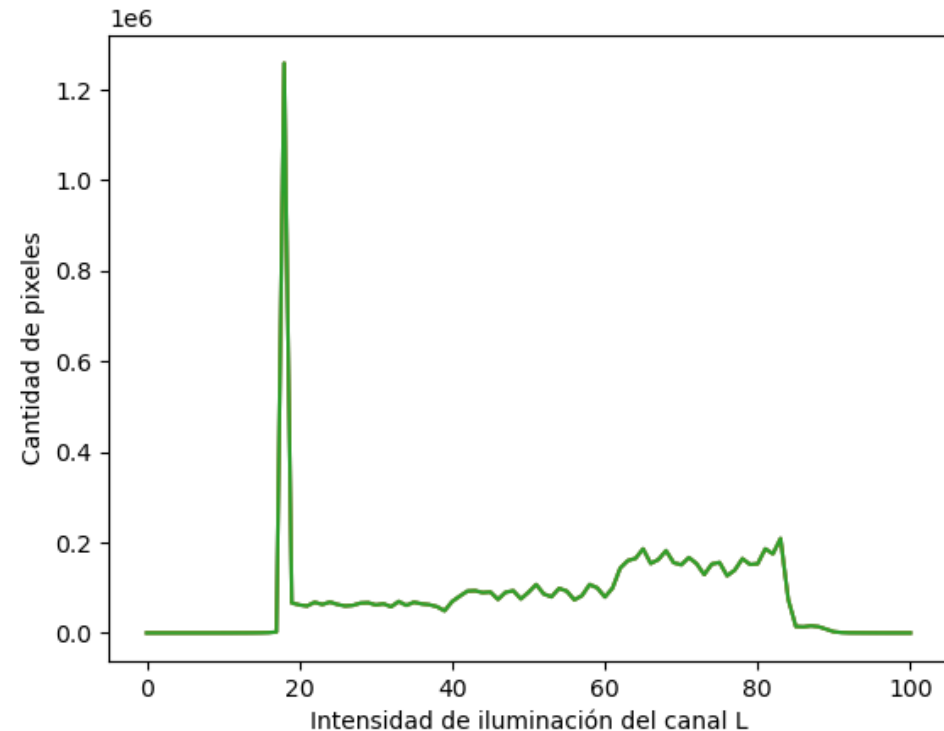
Técnicas y algoritmos para mejorar su calidad o facilitar su análisis e interpretación

- Transformación de intensidad
 - Mejora de iluminación
 - Mejora de contraste
- Filtros espaciales



Mejora de contraste

Procesamiento de imagen para incrementar el rango dinámico de una imagen.



Lógica difusa

Lógica
clásica

Valor
certero

25.2 °C

Lógica
difusa

Lenguaje
natural

Está
templado

Bases

- Funciones de membresía

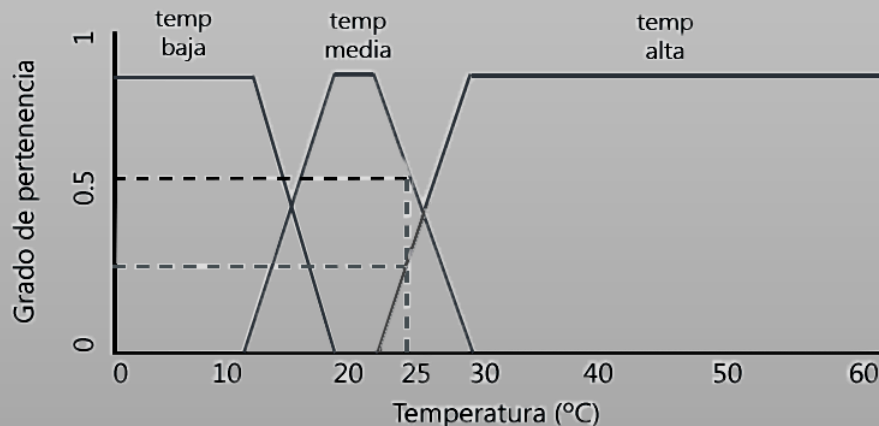
$$F = \{(x, \mu_F(x)) | x \in U, 0.0 \leq \mu_F \leq 1.0\}$$

- Triangular
- Trapezoidal
- Campana
- S (Sigmoidal)
- Z
- Gaussiana

- Variables lingüísticas

- Frio
- Templado
- Caliente

Las variables lingüísticas se definen con funciones de membresía



Sistemas de inferencia difusos (FIS)

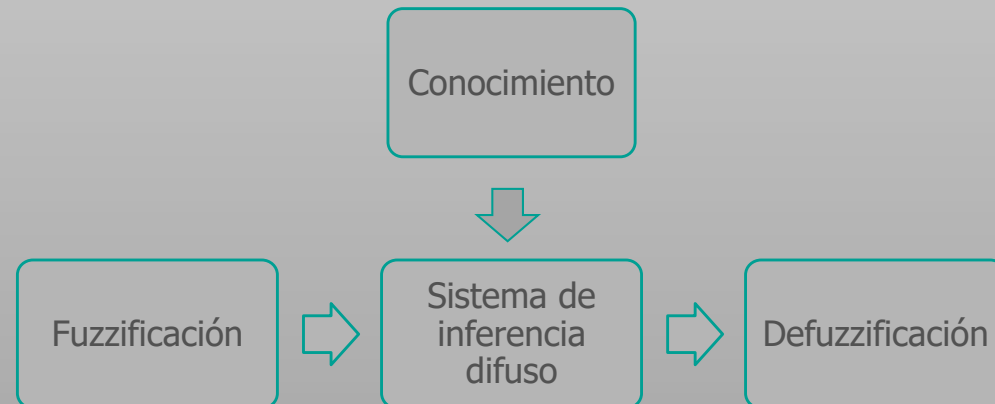
Son sistemas capaces de modelar la experiencia humana para mapear la salida

- Mamdani

SI x_1 es A_1 y x_2 es A_2 ENTONCES y_1 is B

- Sugeno

SI x es A y y es B ENTONCES z is $z = f(x, y)$



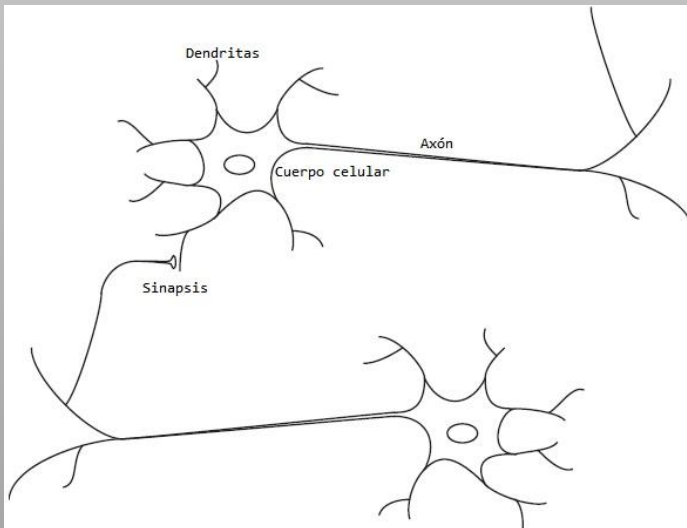
Redes neuronales artificiales

Inspiración biológica

El cerebro está conformado por 100 billones de neuronas conectadas con 10000 neuronas vecinas.

- Reconocimiento
- Asociación
- Clasificación
- Agrupación

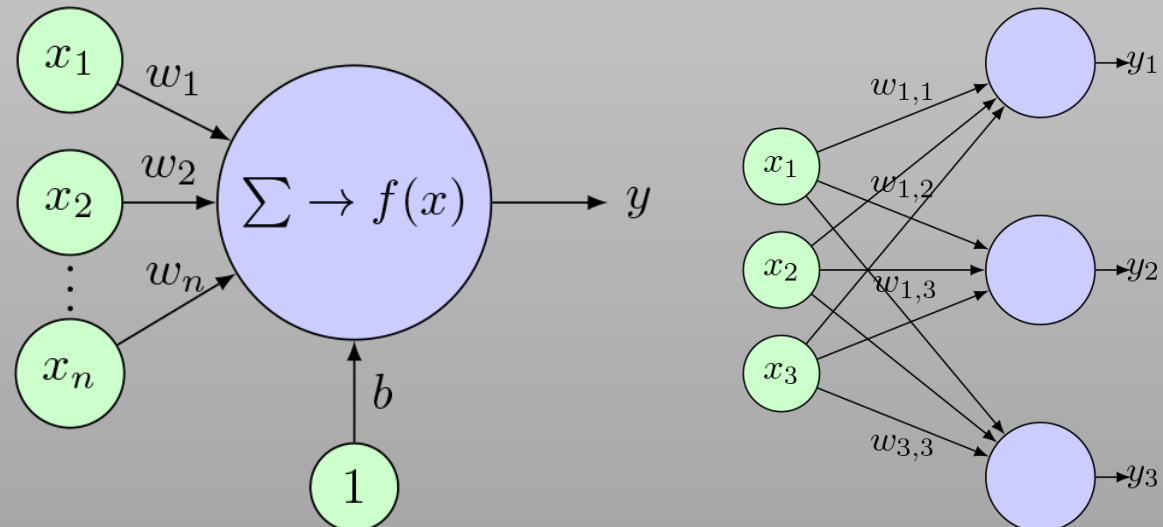
Recolectan y “resumen” las señales de estímulo de las neuronas vecinas



Neurona artificial

$$y = f \left(\sum_{i=1}^n (x_i w_i) + b \right)$$

- Pesos w_i
- Sesgo (Bias) b
- Función de activación

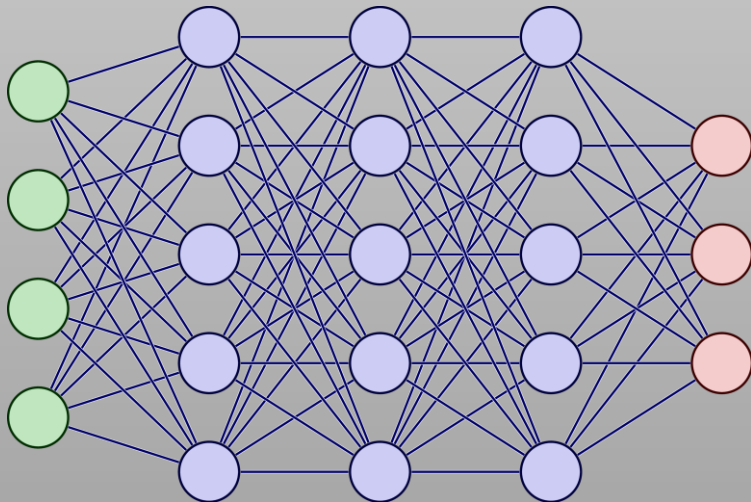


Redes neuronales artificiales

Redes neuronales artificiales (ANN)

Entrenamiento gracias al algoritmo de *backpropagation*

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje por refuerzo

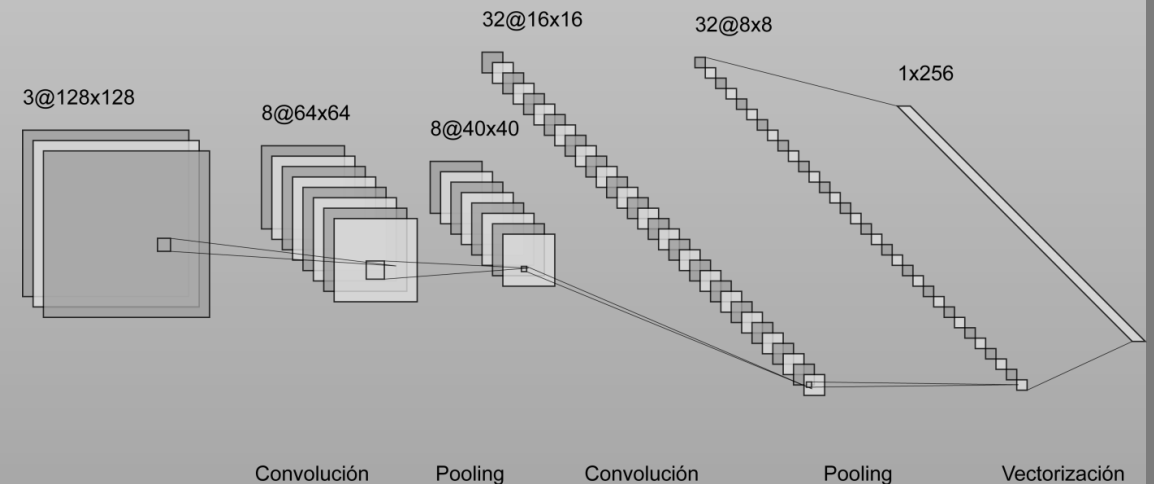


Redes neuronales convolucionales (CNN)

Están inspiradas en el sistema nervioso visual

1. Señal en la corteza visual
2. Campo receptivo local
3. Neuronas simples y complejas
4. Capas especializadas

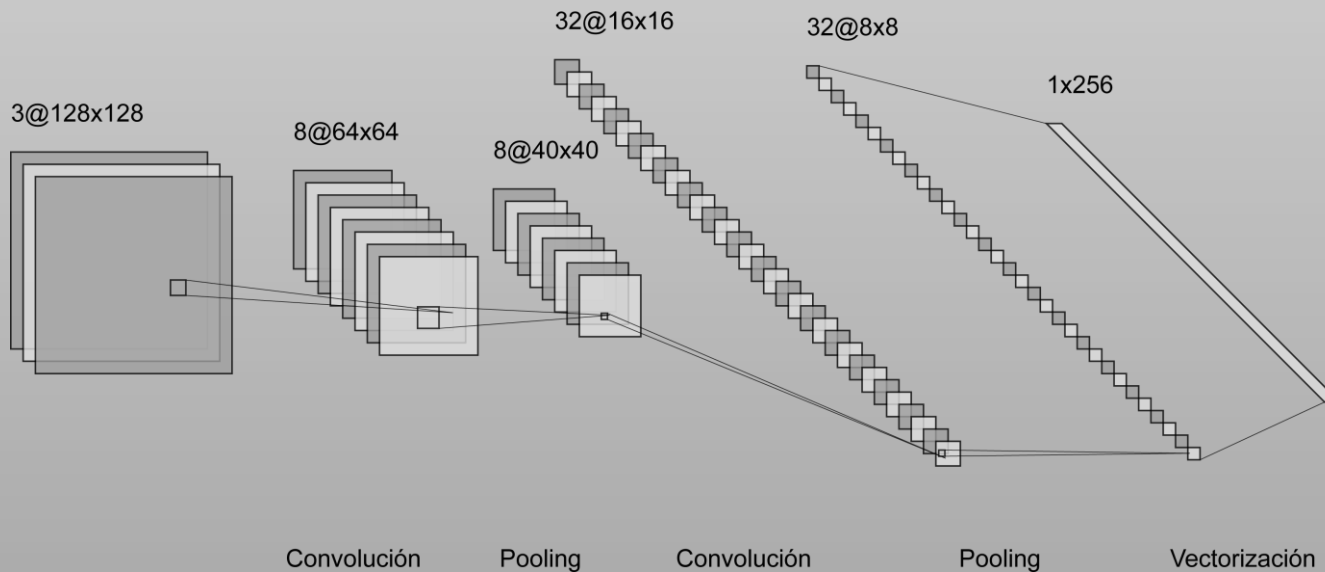
Se extraen características más complejas a medida que la información avanza en la red



Redes neuronales artificiales

CNN

- Convolución
- Submuestreo (Pooling)
- *Fully connected*



0	0	0	0	0	0	0
0	3	5	1	2	3	0
0	4	1	0	3	1	0
0	0	2	2	4	5	0
0	3	1	3	1	3	0
0	5	0	4	2	1	0
0	0	0	0	0	0	0

Imagen de 5×5

0	1	0
0	1	0
0	1	0

Kernel de 3×3

$*$

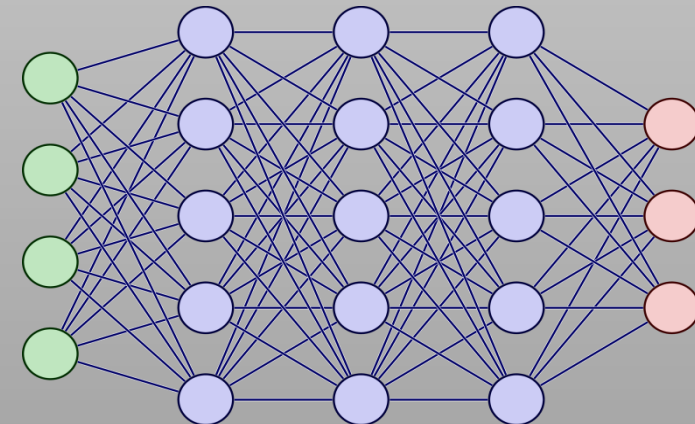
7	6	1	5	4
7	8	3	9	9
7	4	5	8	9
8	3	9	7	9
8	1	7	3	4

$=$ $I * K$

1	0	3	1
2	2	4	5
1	3	1	3
0	4	2	1

$\xrightarrow{\text{max pooling}}$

2	4	5
3	4	5
4	4	3



Procesamiento de imágenes

Metodología



FIS para índice de contraste



FIS para mejora de contraste

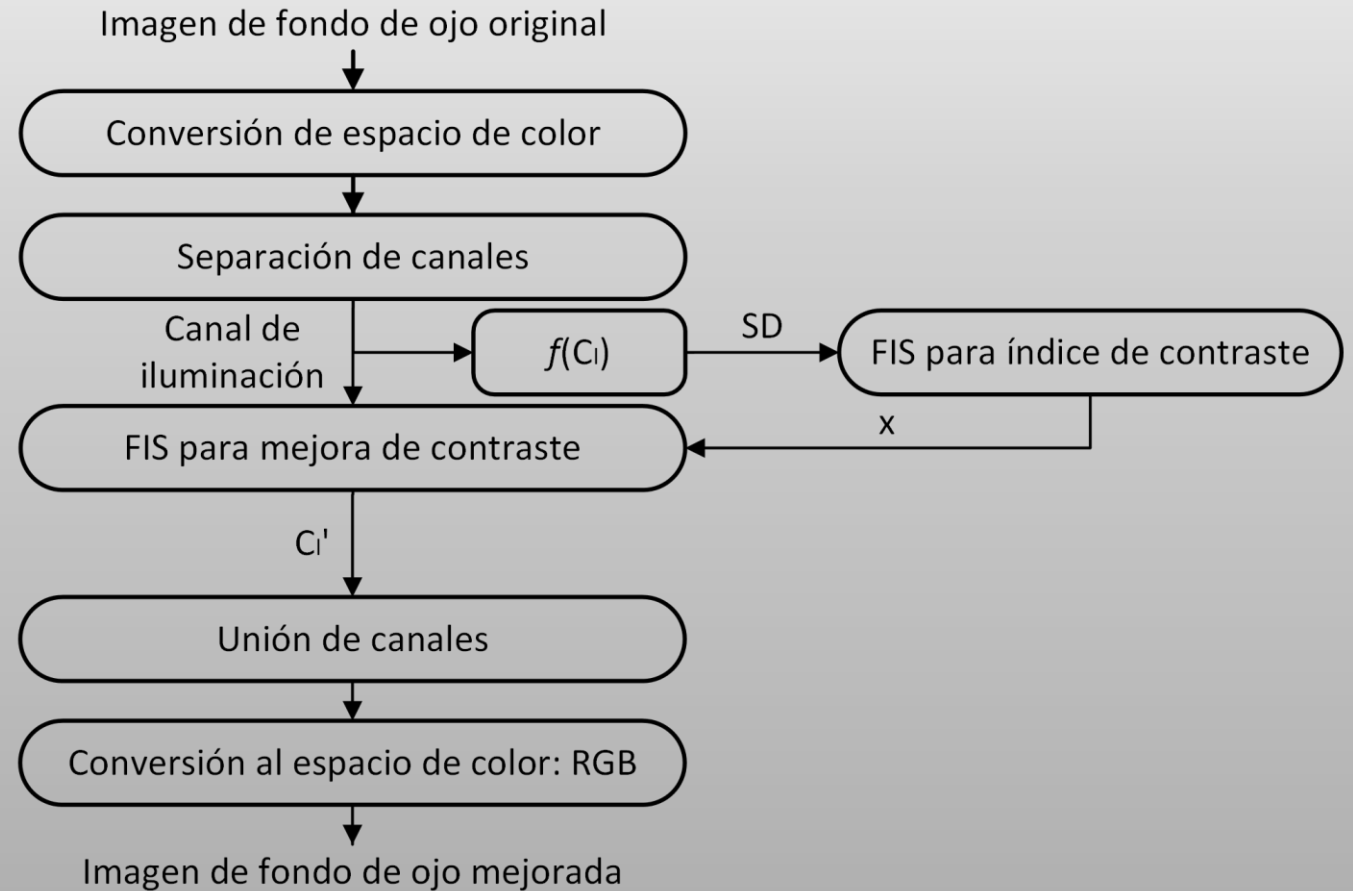


Resultados

Mejora de contraste

Consideraciones para mejorar el contraste en imágenes de fondo de ojo:

1. Mantener la tonalidad de los colores
2. Mantener la forma de las regiones de mácula
3. Mantener la forma de las regiones del nervio óptico



FIS para índice de contraste

Función Z

$$Z(\sigma) = \frac{1}{1 + e^{\alpha(\sigma - \beta)}}$$

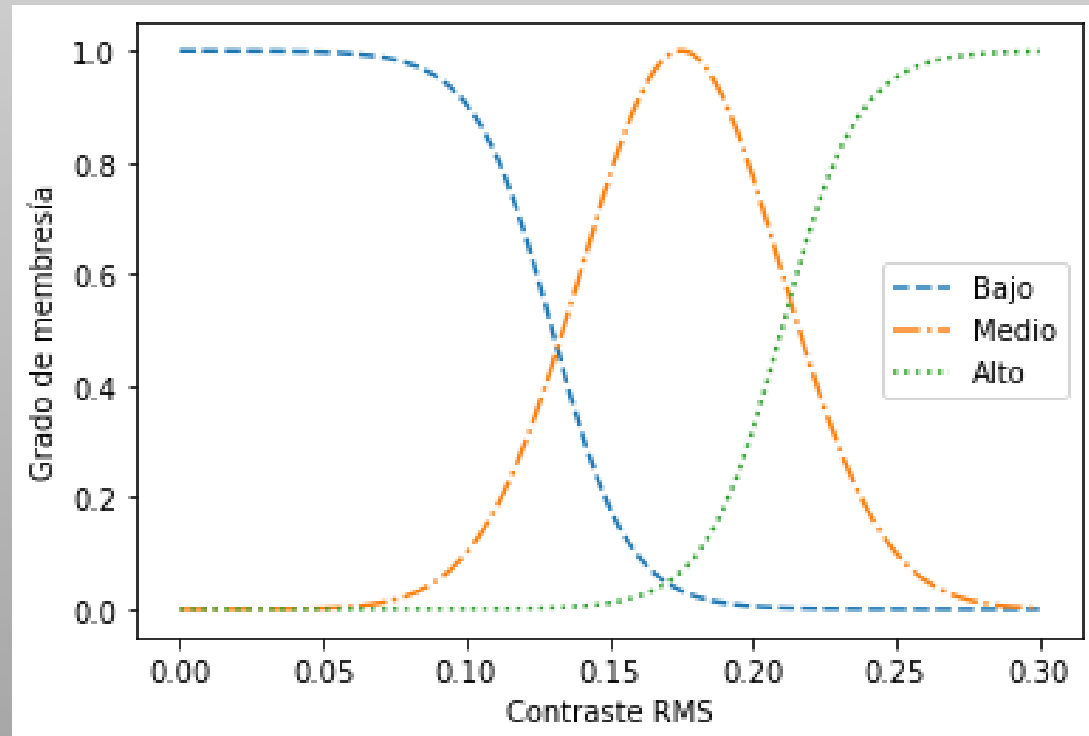
Función gaussiana

$$B(\sigma) = e^{-\frac{1}{2}\left(\frac{\sigma - \gamma}{\delta}\right)^2}$$

Función S

$$S(\sigma) = \frac{1}{1 + e^{-\alpha(\sigma - \varepsilon)}}$$

Función de membresía	Variable lingüística	Parámetro	Valor
Z	Bajo	α	0.750
		β	0.130
G	Medio	γ	0.175
		δ	0.035
S	Alto	α	0.750
		ε	0.210



FIS para índice de contraste

“Si tiene bajo contraste el ajuste deberá ser mayor, si tiene un contraste medio el ajuste será medio y si tiene un contraste alto el ajuste será poco”

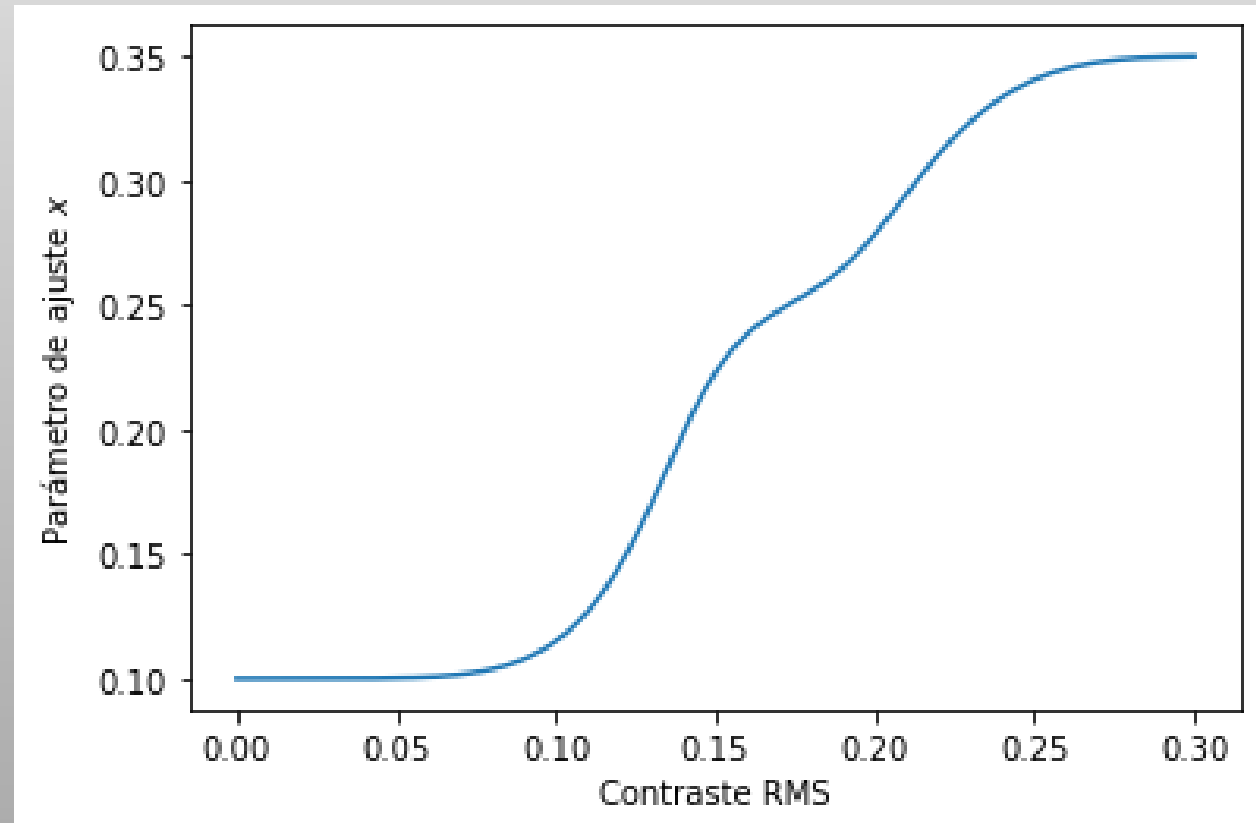
Reglas **SI-ENTONCES**

1. **SI** bajo **ENTONCES** f_{11} .
2. **SI** medio **ENTONCES** f_{12} .
3. **SI** alto **ENTONCES** f_{13} .

Defuzzificación (método WAM)

$$x = \frac{(\mu_{11} \cdot f_{11}) + (\mu_{12} \cdot f_{12}) + (\mu_{13} \cdot f_{13})}{\mu_{11} + \mu_{12} + \mu_{13}}$$

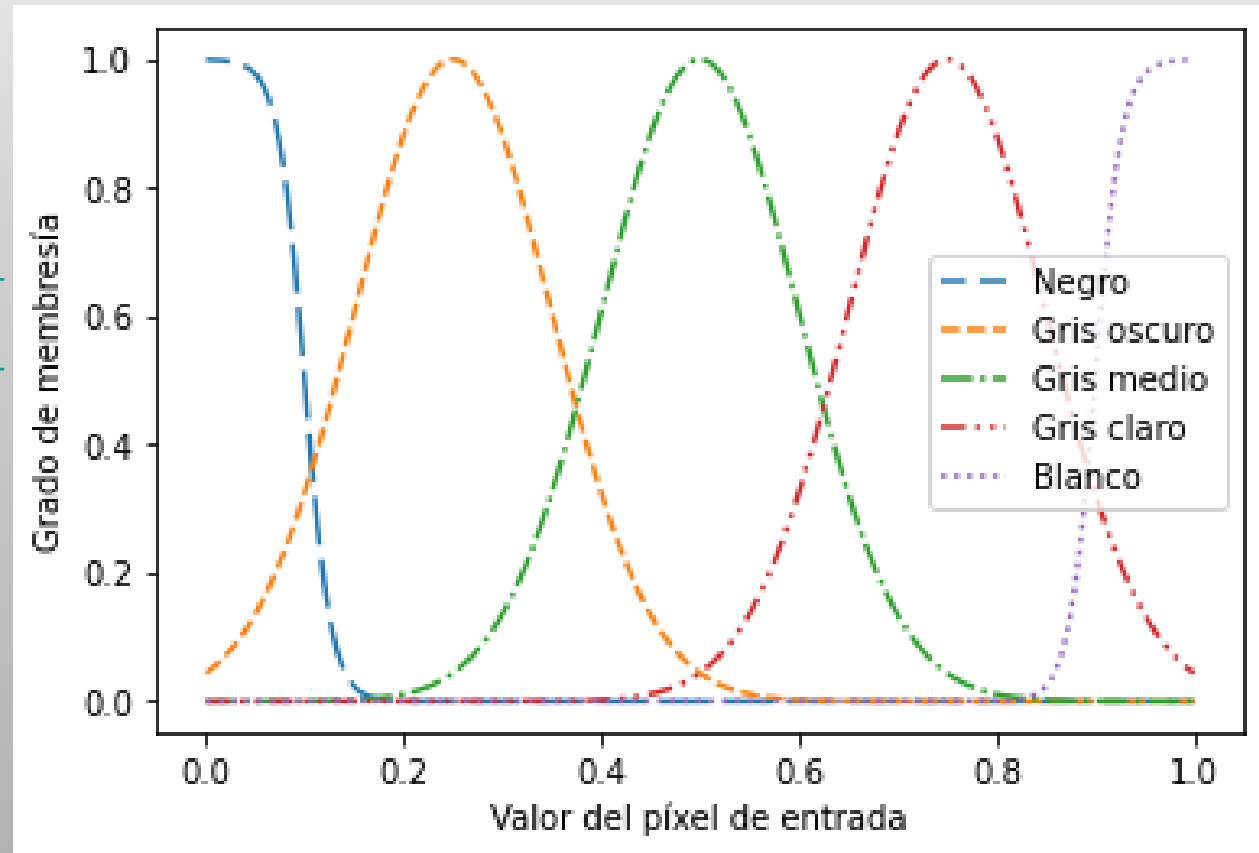
Singleton difuso	Valor numérico
f_{11}	0.10
f_{12}	0.25
f_{13}	0.35



FIS para mejora de contraste

El universo de discurso comprende todos los valores que puede tener un pixel.

Variable lingüística	Función de membresía	Parámetro	Valor
Negro	Z	α	0.75
		β	0.10
Gris oscuro	G_1	γ	0.25
		δ	0.10
Gris medio	G_2	γ	0.50
		δ	0.10
Gris claro	G_3	γ	0.75
		δ	0.10
Blanco	S	α	0.75
		ε	0.90



FIS para mejora de contraste

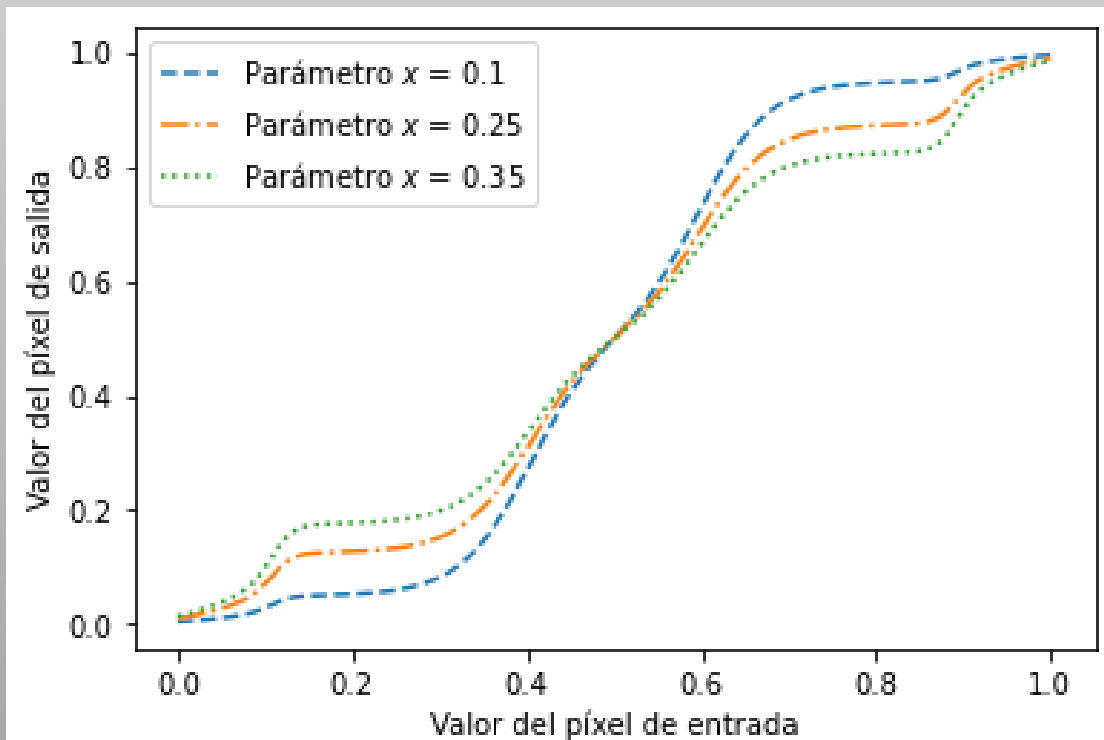
“Tonalidades negras se quedan negras, tonalidades oscuras deben ser más oscuras, tonalidades medias se quedan medias, tonalidades claras se aclararán más y tonalidades blancas se quedan blancas.”

Reglas **SI-ENTONCES**

1. **SI** negro **ENTONCES** negro.
2. **SI** gris oscuro **ENTONCES** gris más oscuro.
3. **SI** gris medio **ENTONCES** gris medio.
4. **SI** gris claro **ENTONCES** gris más claro.
5. **SI** blanco **ENTONCES** blanco.
6. **SI** gris oscuro **ENTONCES** negro.
7. **SI** gris claro **ENTONCES** blanco.

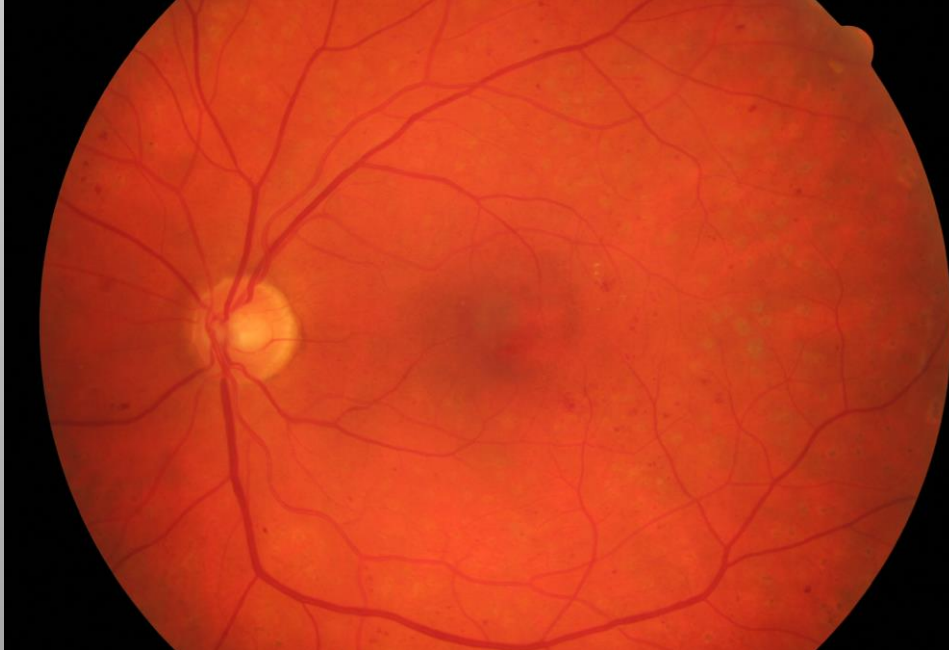
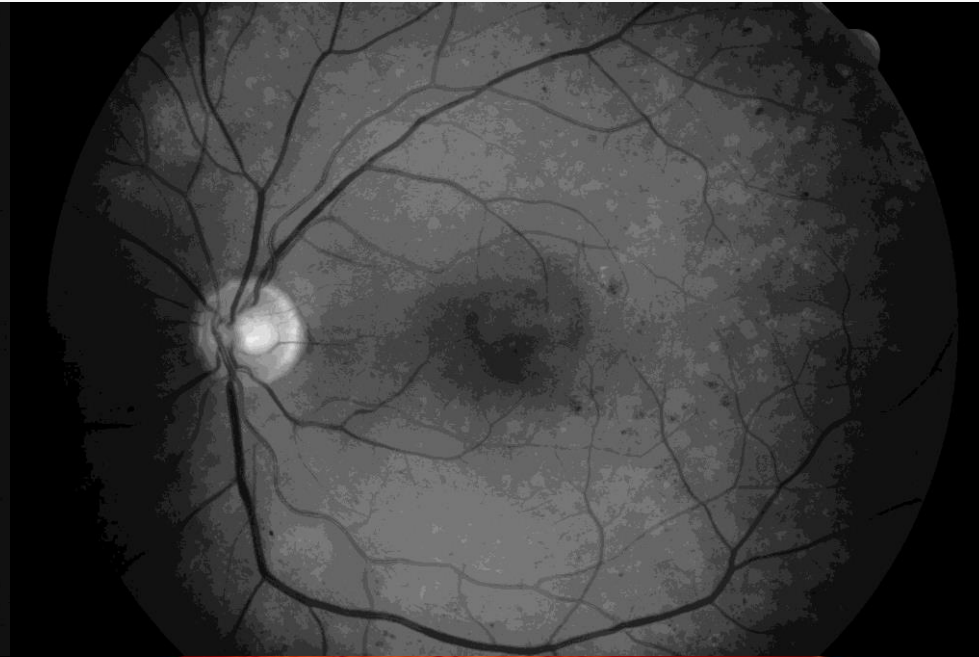
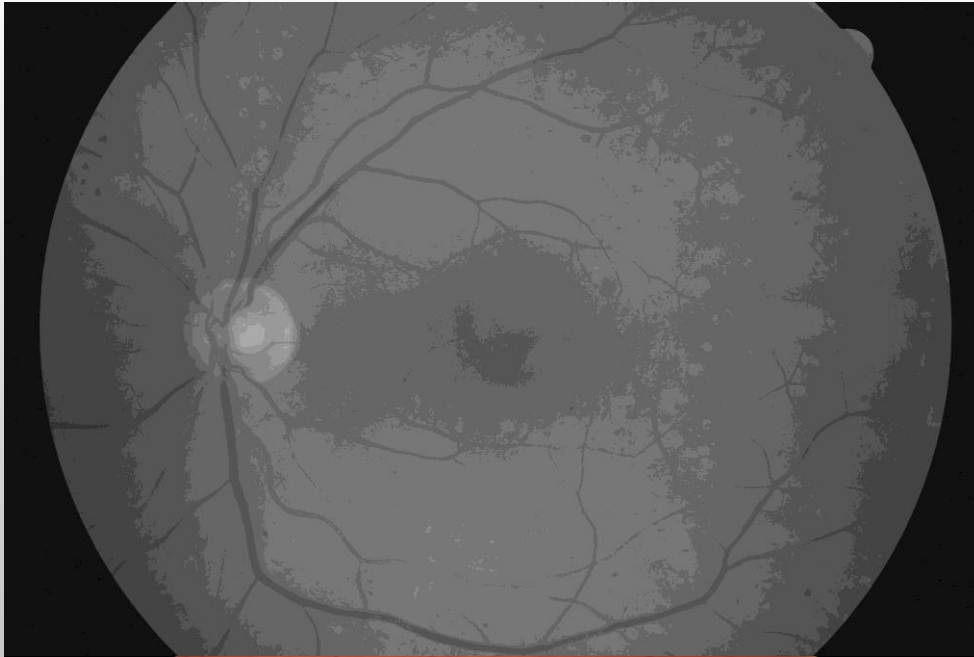
$$salida = \frac{\sum_{i=1}^n \mu(I)_{2i} \cdot f_{2i}}{\sum_{i=1}^n \mu(I)_{2i}}$$

Variable lingüística de salida	Singleton difuso	Valor numérico
Negro	f_{21}	0.0
Gris más oscuro	f_{22}	x
Gris medio	f_{23}	0.5
Gris más claro	f_{24}	$1 - x$
Blanco	f_{25}	1.0



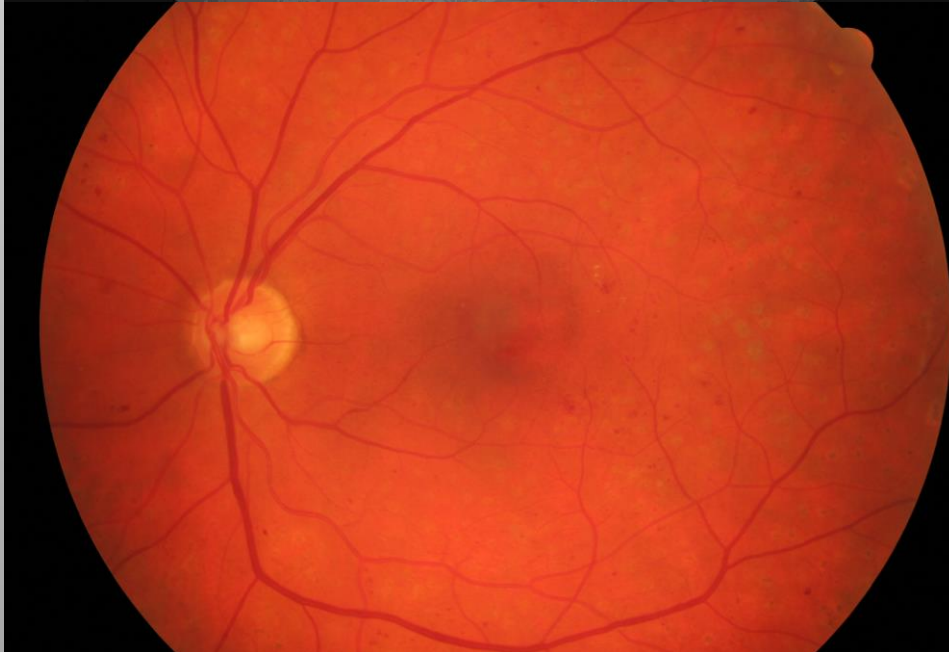
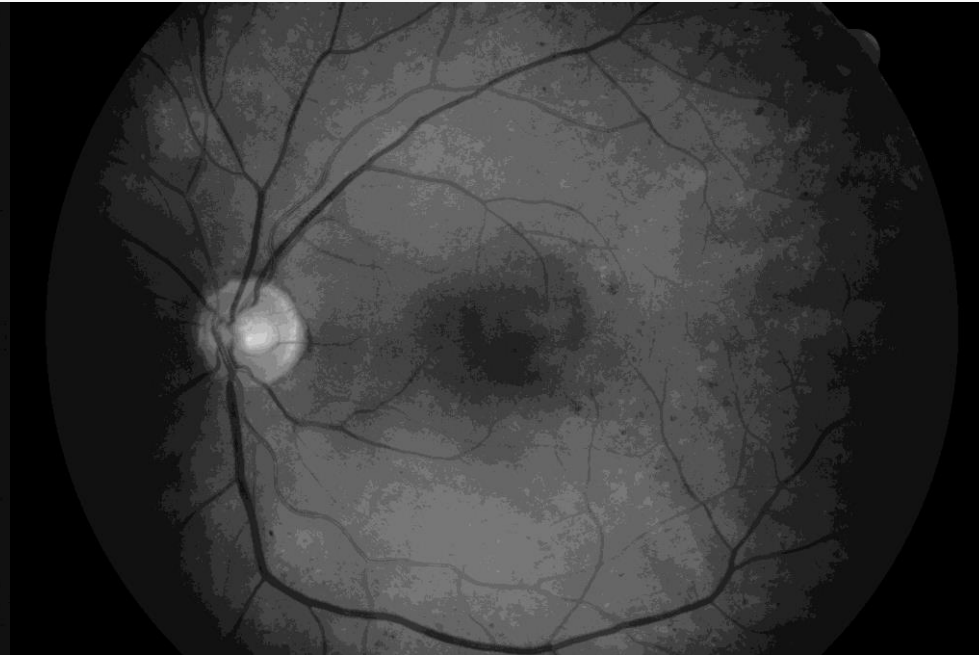
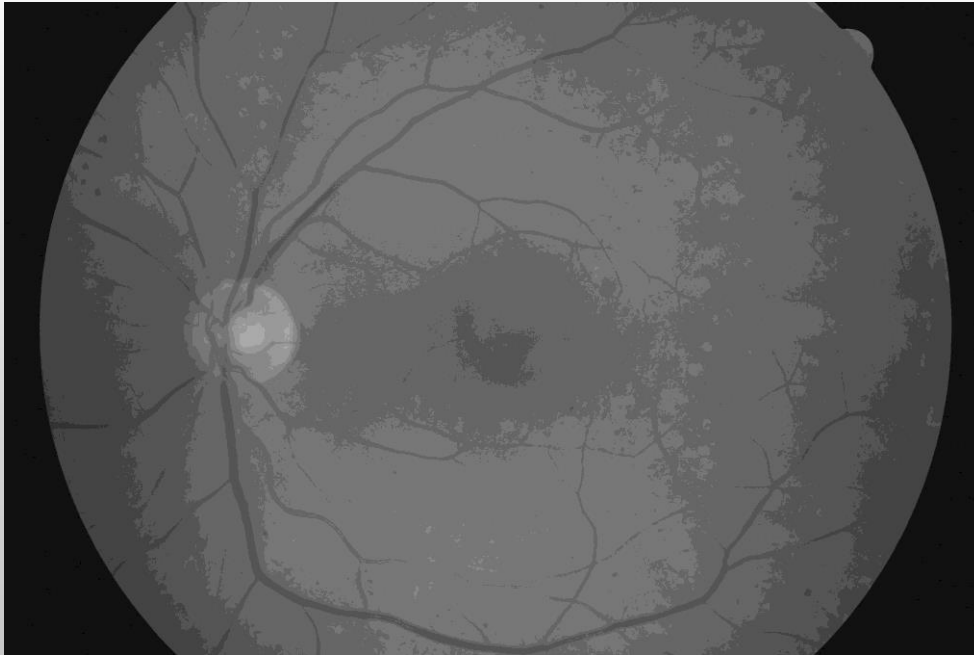
Original

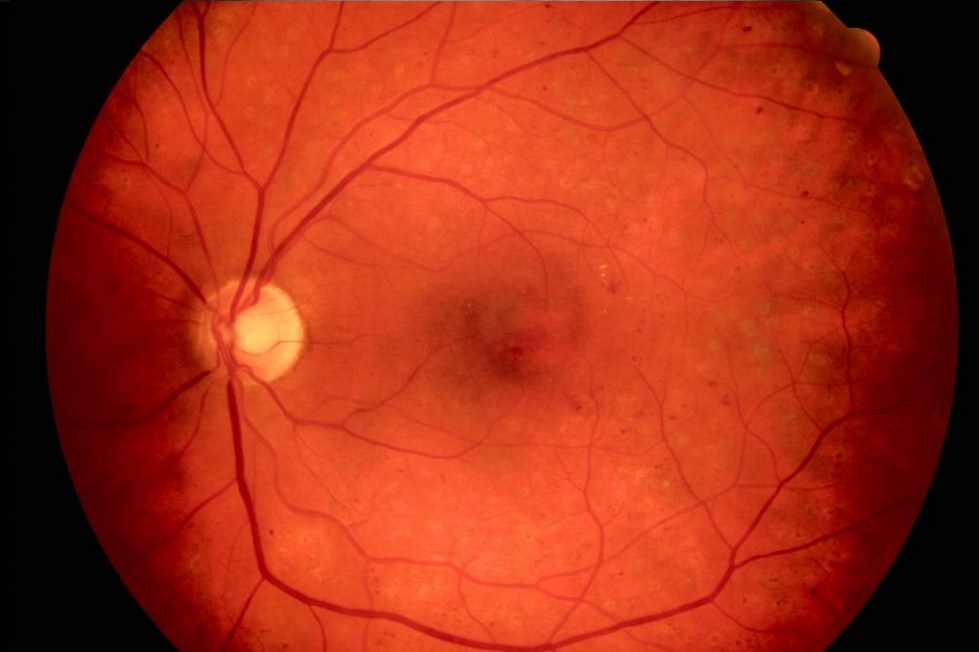
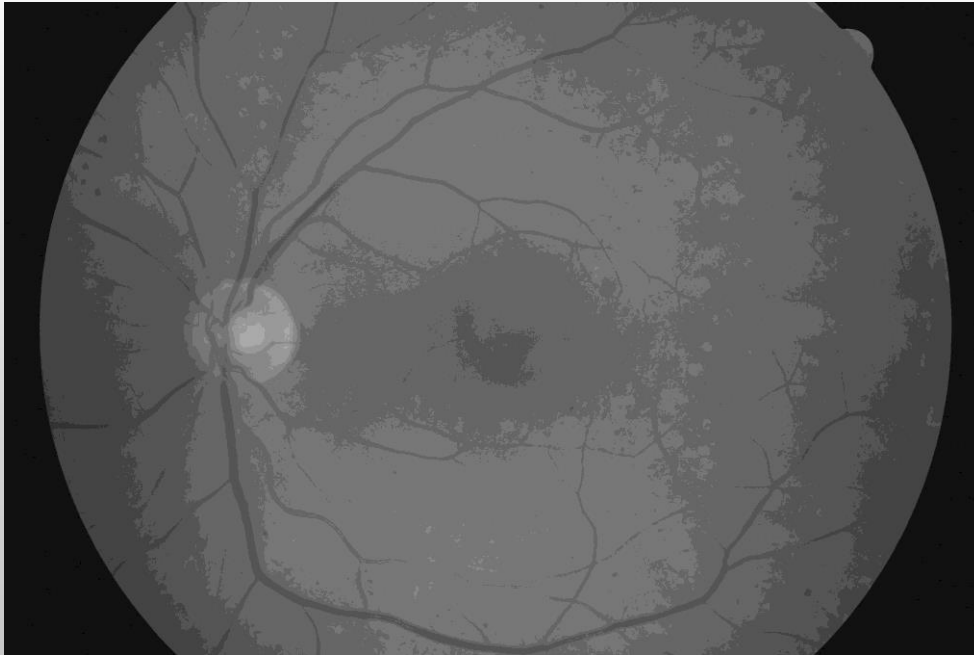
YCbCr

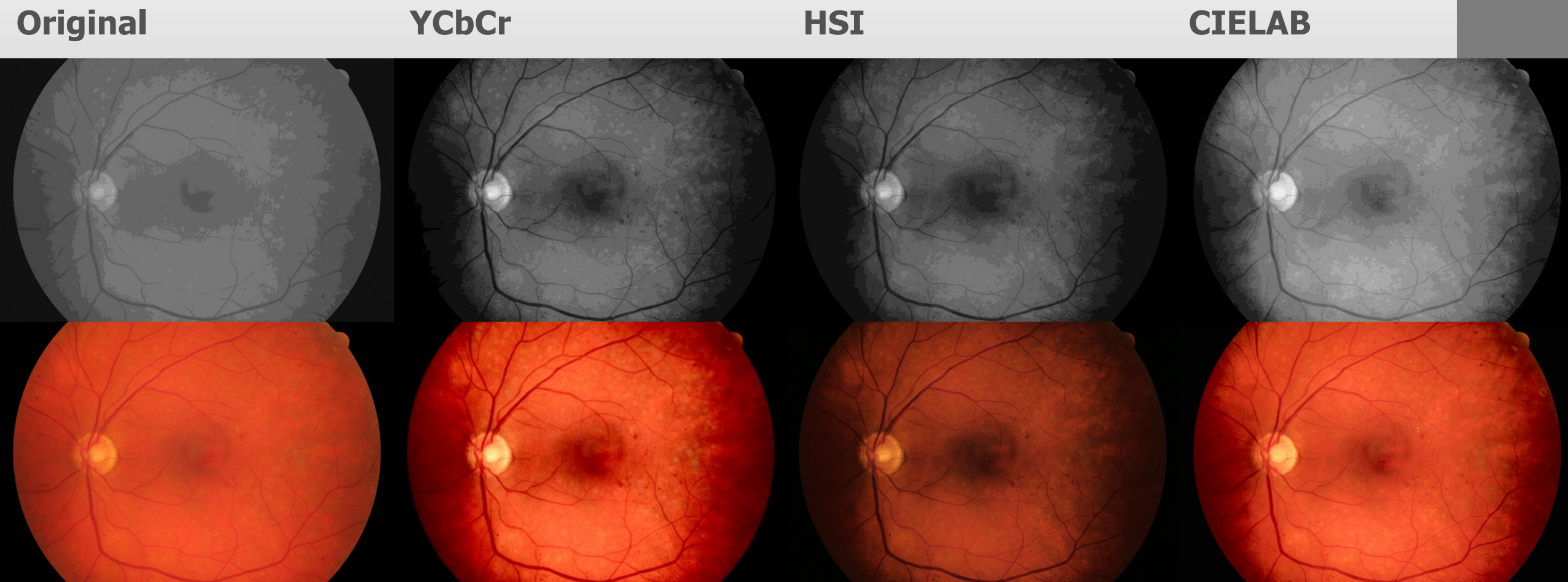


Original

HSI



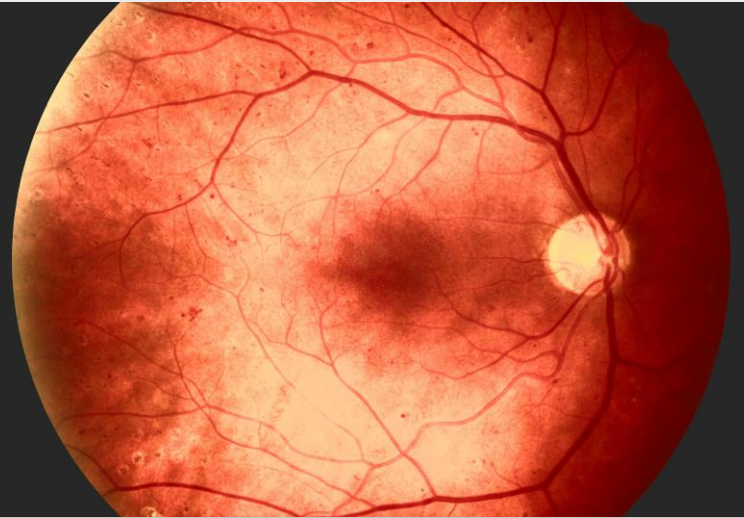




Resultados



Original



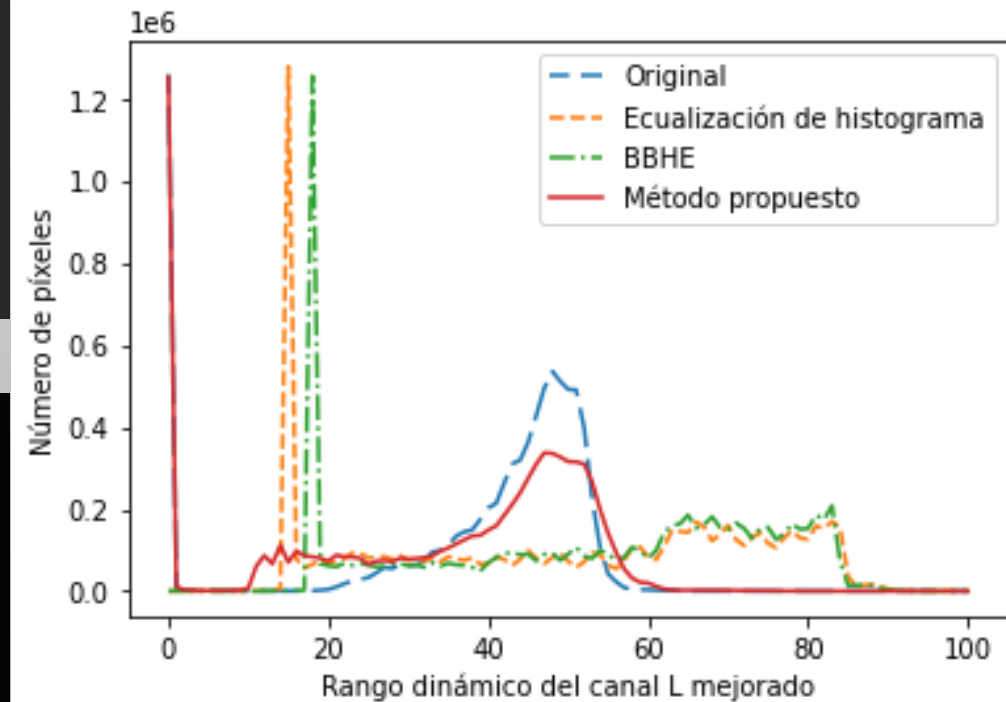
Ecualización de histograma



Ecualización de bi-histograma



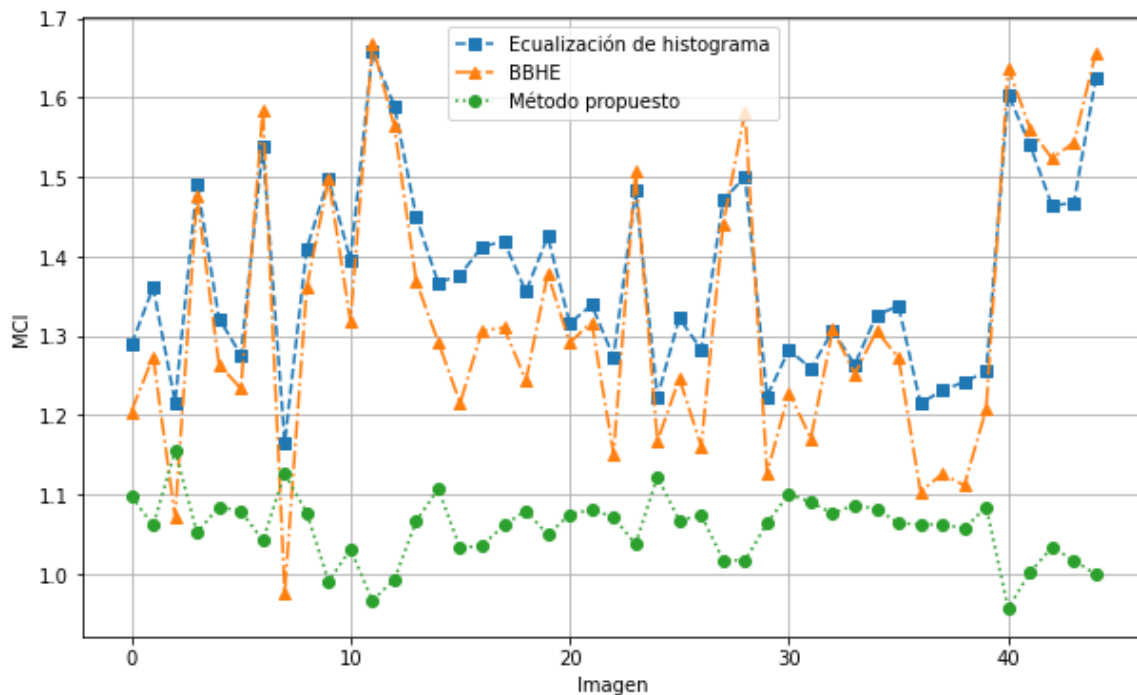
FIS



MCI

Indica si la imagen procesada tiene mayor o menor contraste.

$$MCI = \frac{\sigma_{in}}{\sigma_{out}}$$

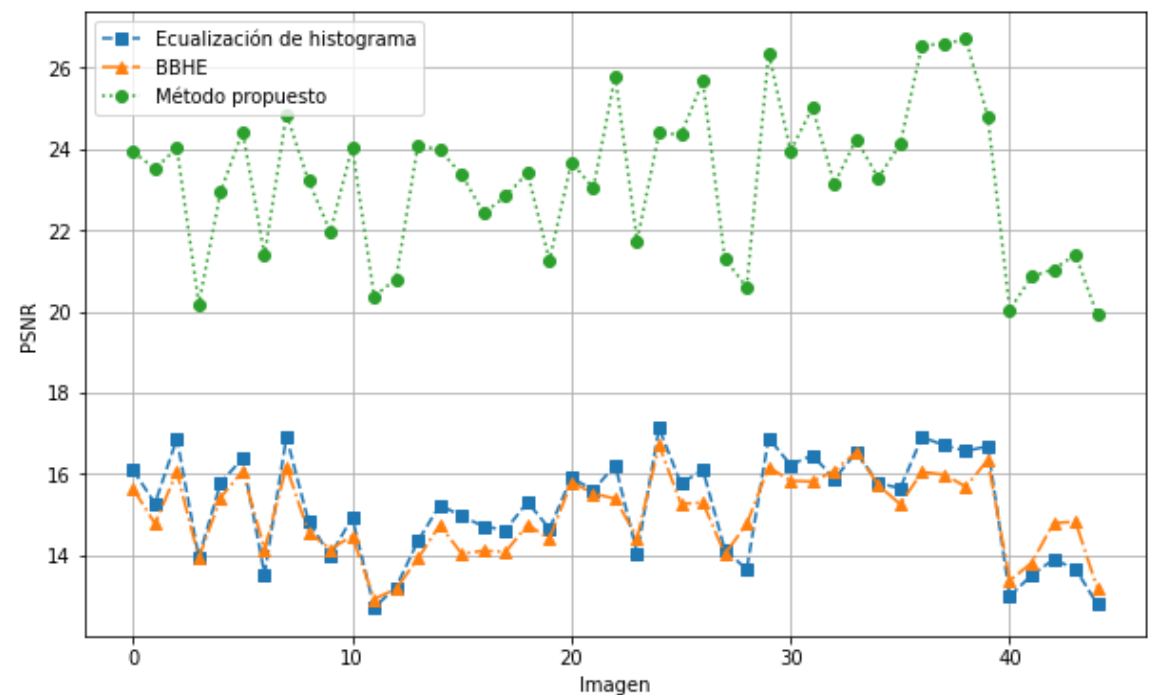


PSNR

Indicativo de la calidad del proceso

$$PSNR = 10 \log \left(\frac{MAX_I^2}{MSE} \right)$$

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i,j) - K(i,j)]^2$$

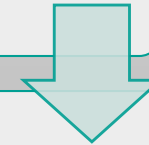


Resultados

Imagen	SD	MCI	PSNR	Opinión médica
Original	0.174	1.0	-	Buena
Ecualización de histograma	0.237	1.360	15.270	No adecuada
Ecualización de Bi- histograma	0.222	1.273	14.774	No adecuada
FIS	0.185	1.060	23.509	Muy buena

Aprendizaje profundo

YOLOv5



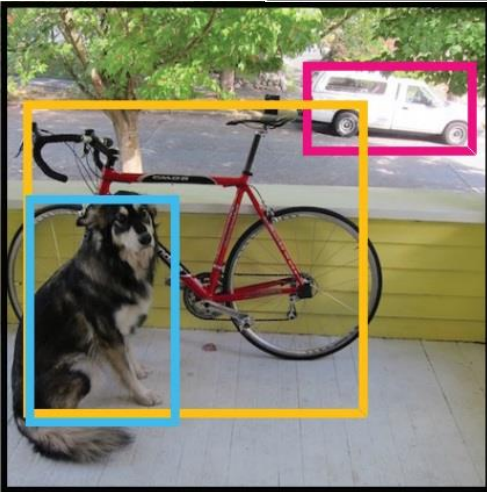
Entrenamiento de YOLOv5



Resultados del entrenamiento

YOLOv5 (You Only Look Once)

YOLOv1



YOLOv5

- Columna vertebral: Darknet-53
- Cuello: Capas de agrupamiento espacial
- Cabeza: Reducciones multiescala

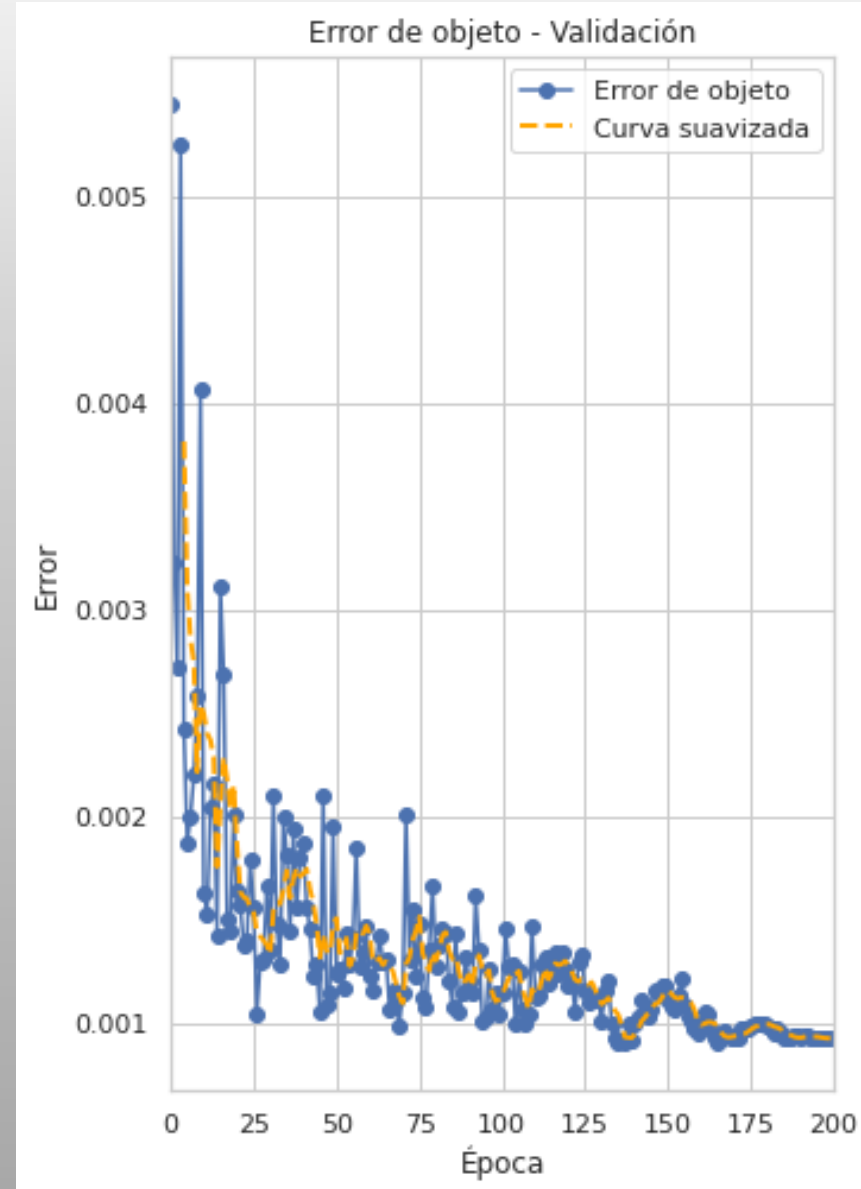
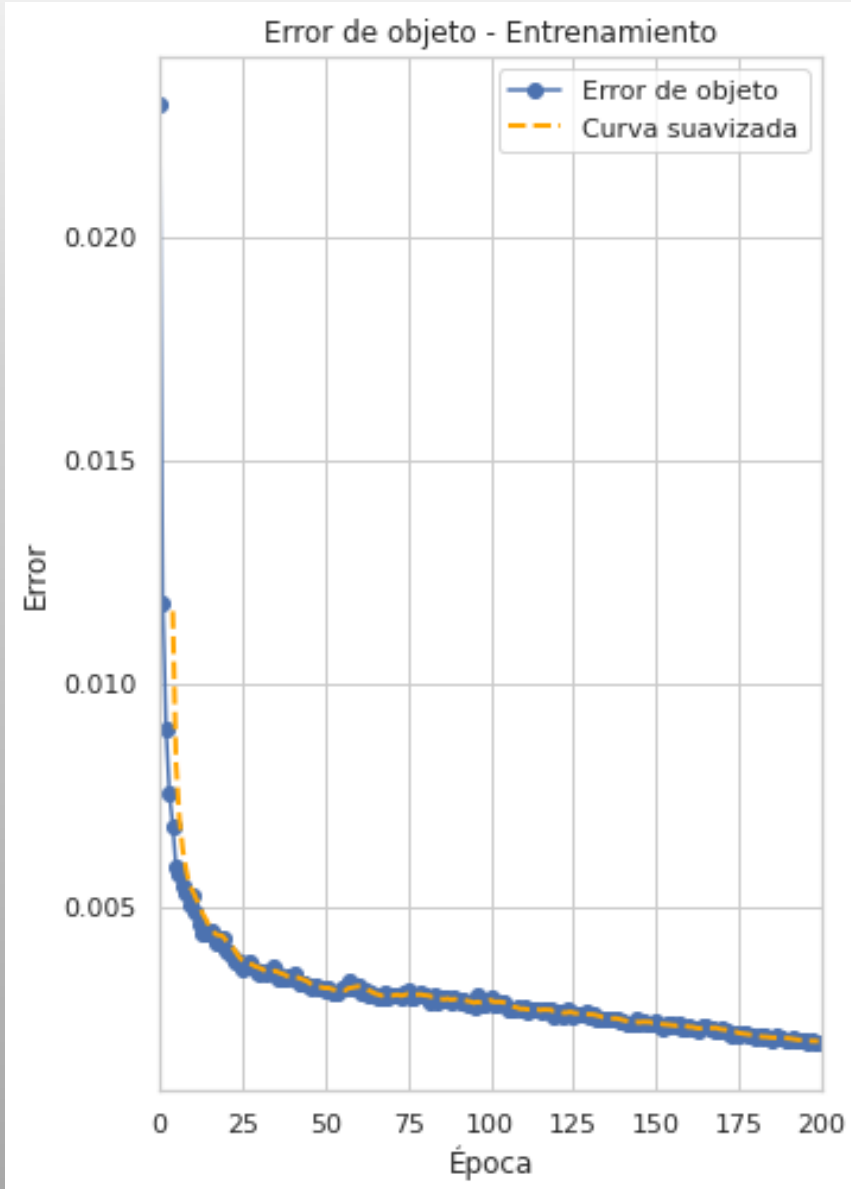
Función de pérdida:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{obj} + \lambda_3 \mathcal{L}_{loc}$$

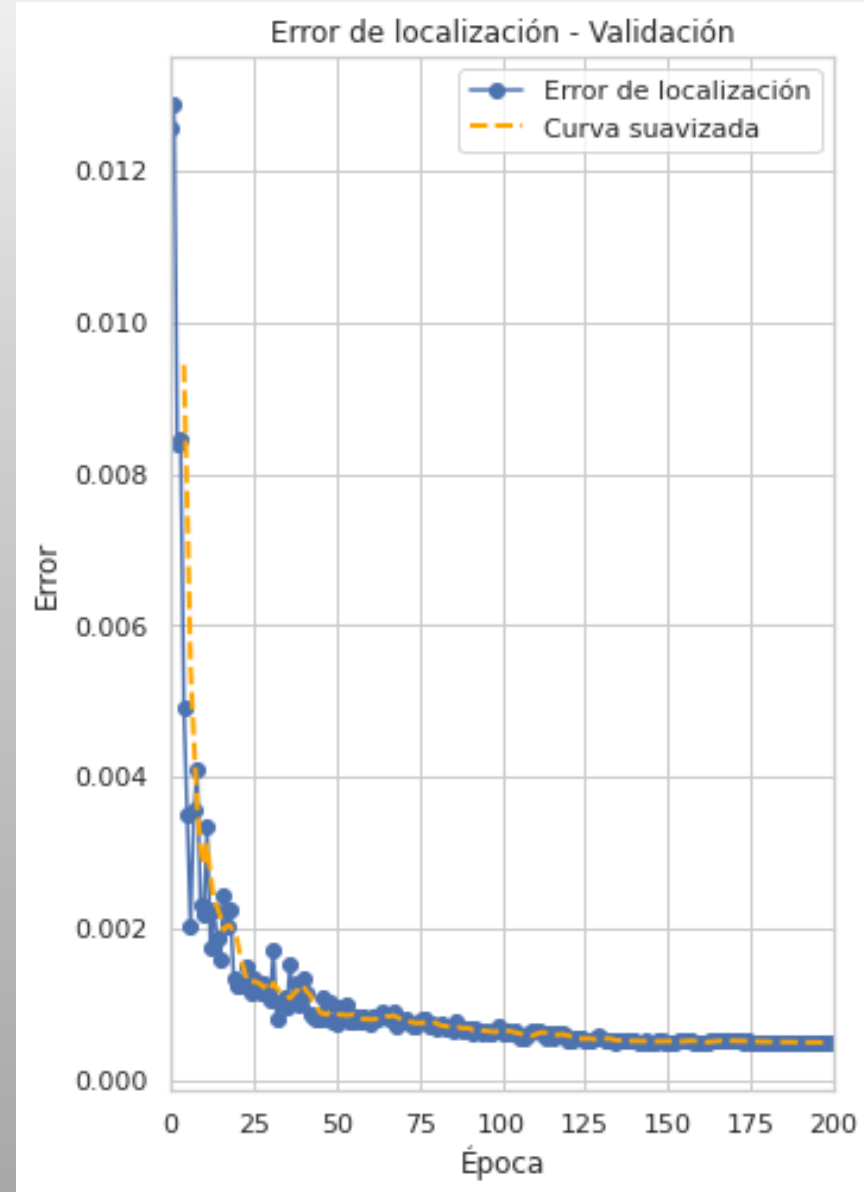
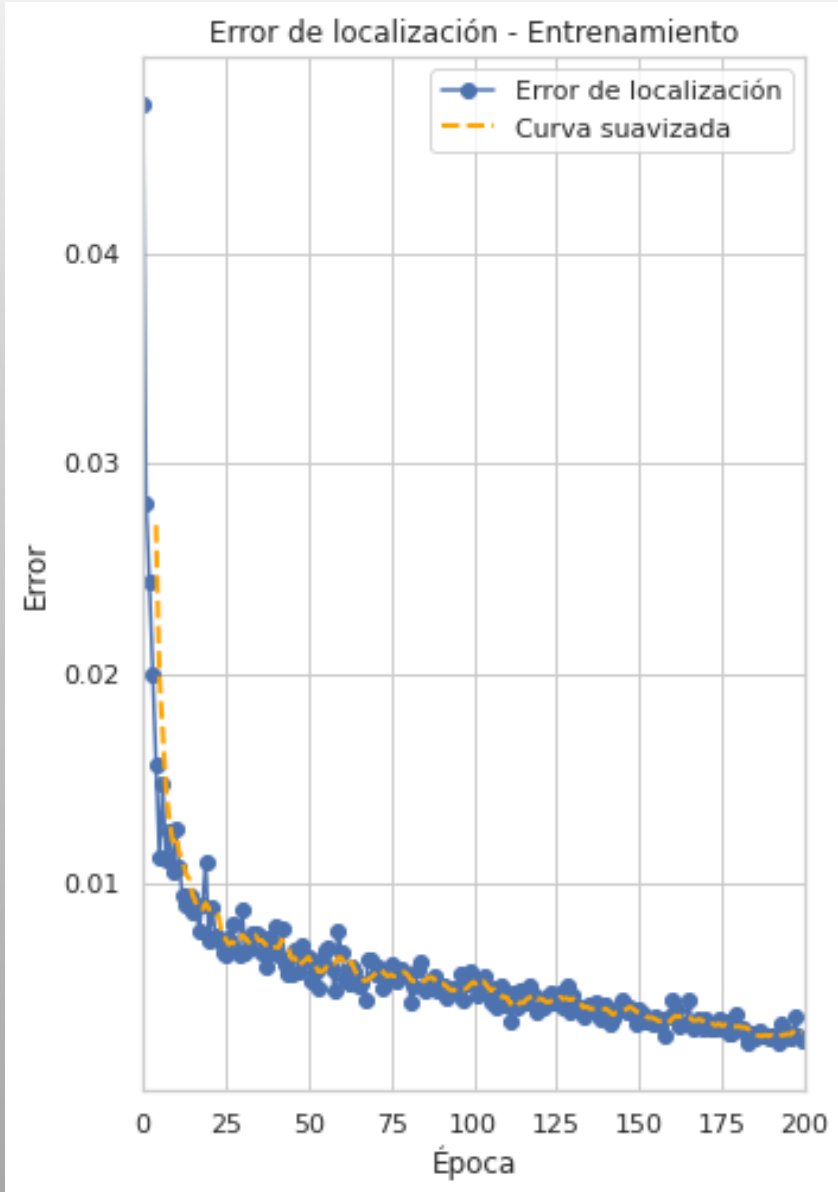
Entrenamiento de YOLOv5



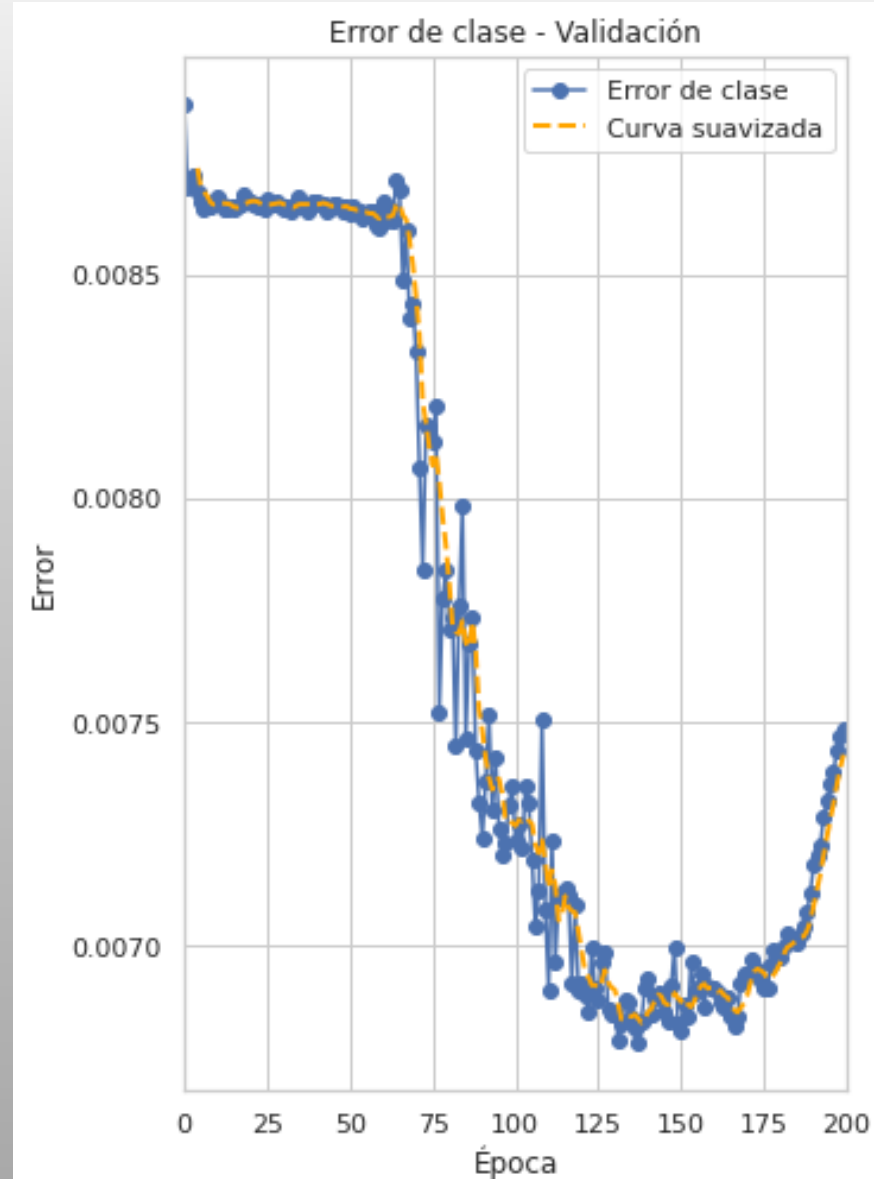
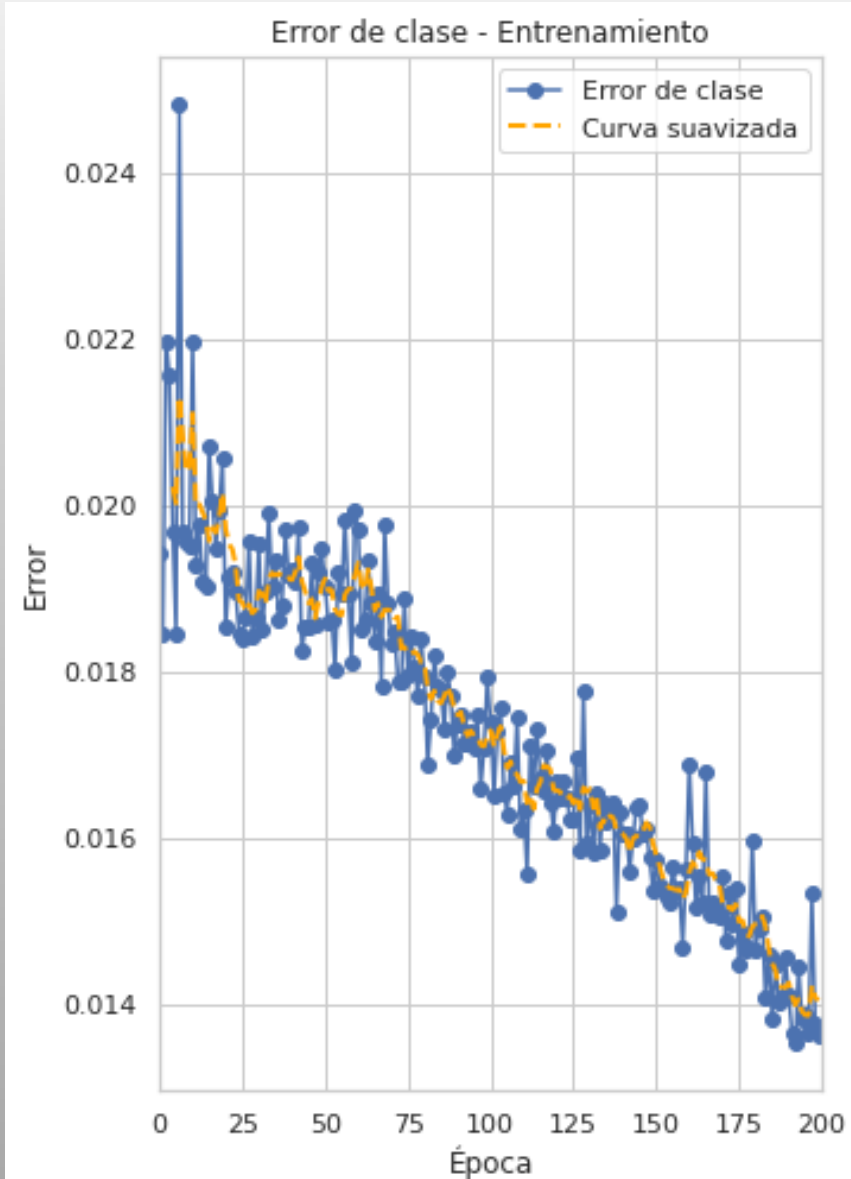
Resultados del entrenamiento



Resultados del entrenamiento



Resultados del entrenamiento

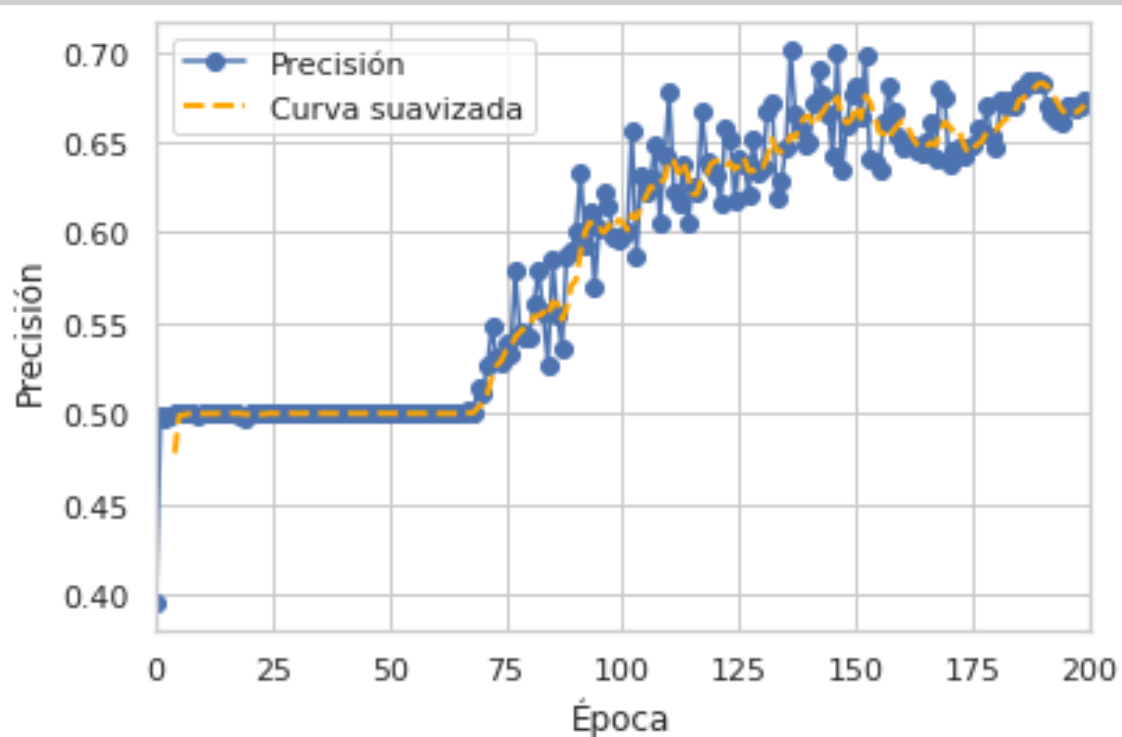


Resultados del entrenamiento

Precisión

Proporción de detecciones correctas

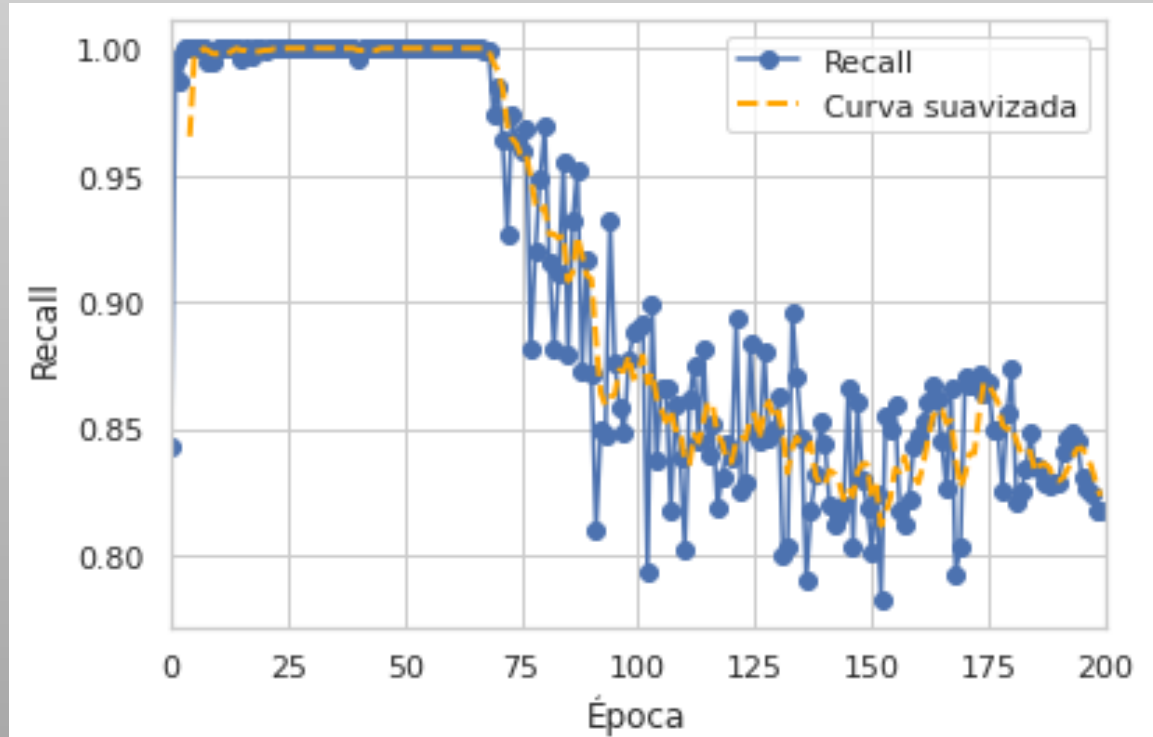
$$P = \frac{VP}{VP + FP}$$



Recall

Proporción de objetos reales

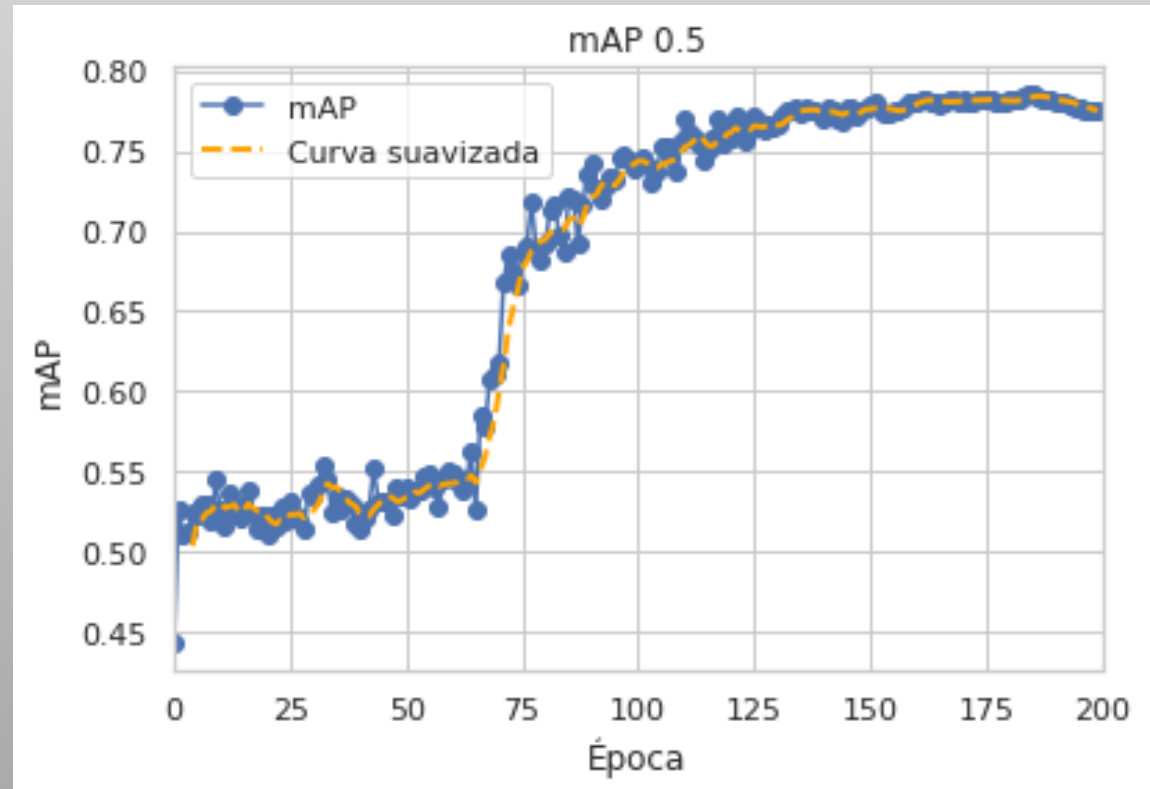
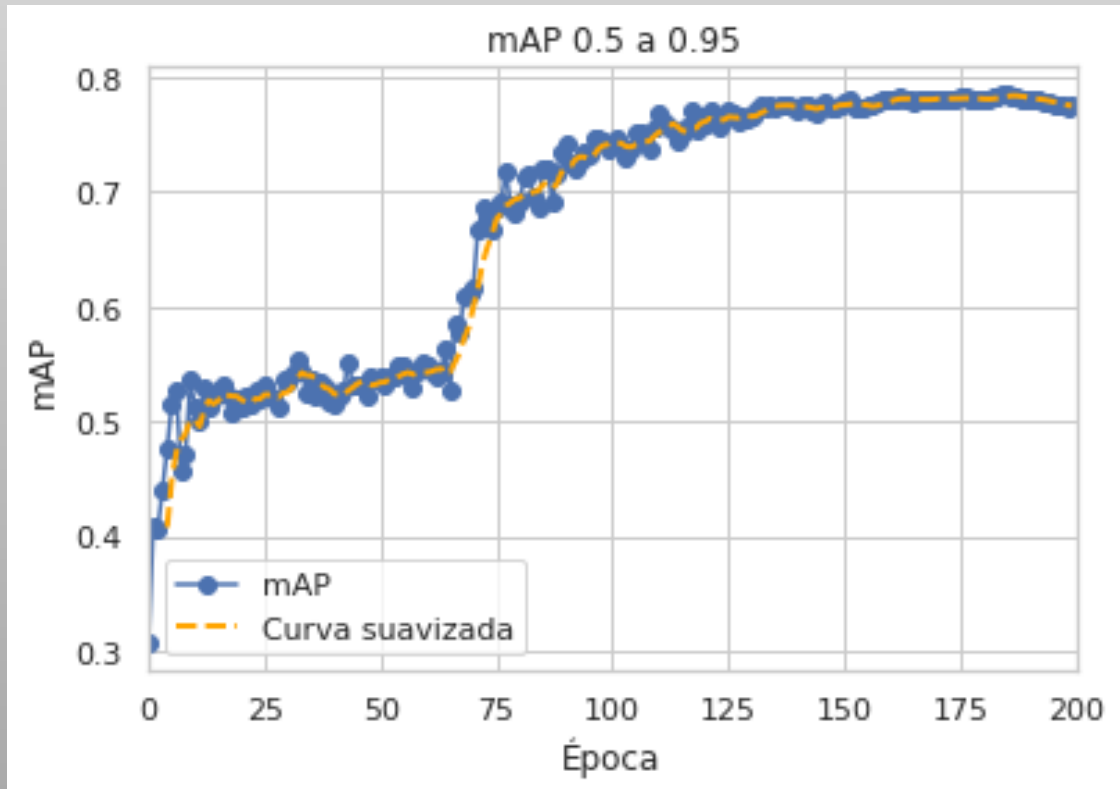
$$R = \frac{VP}{VP + FN}$$



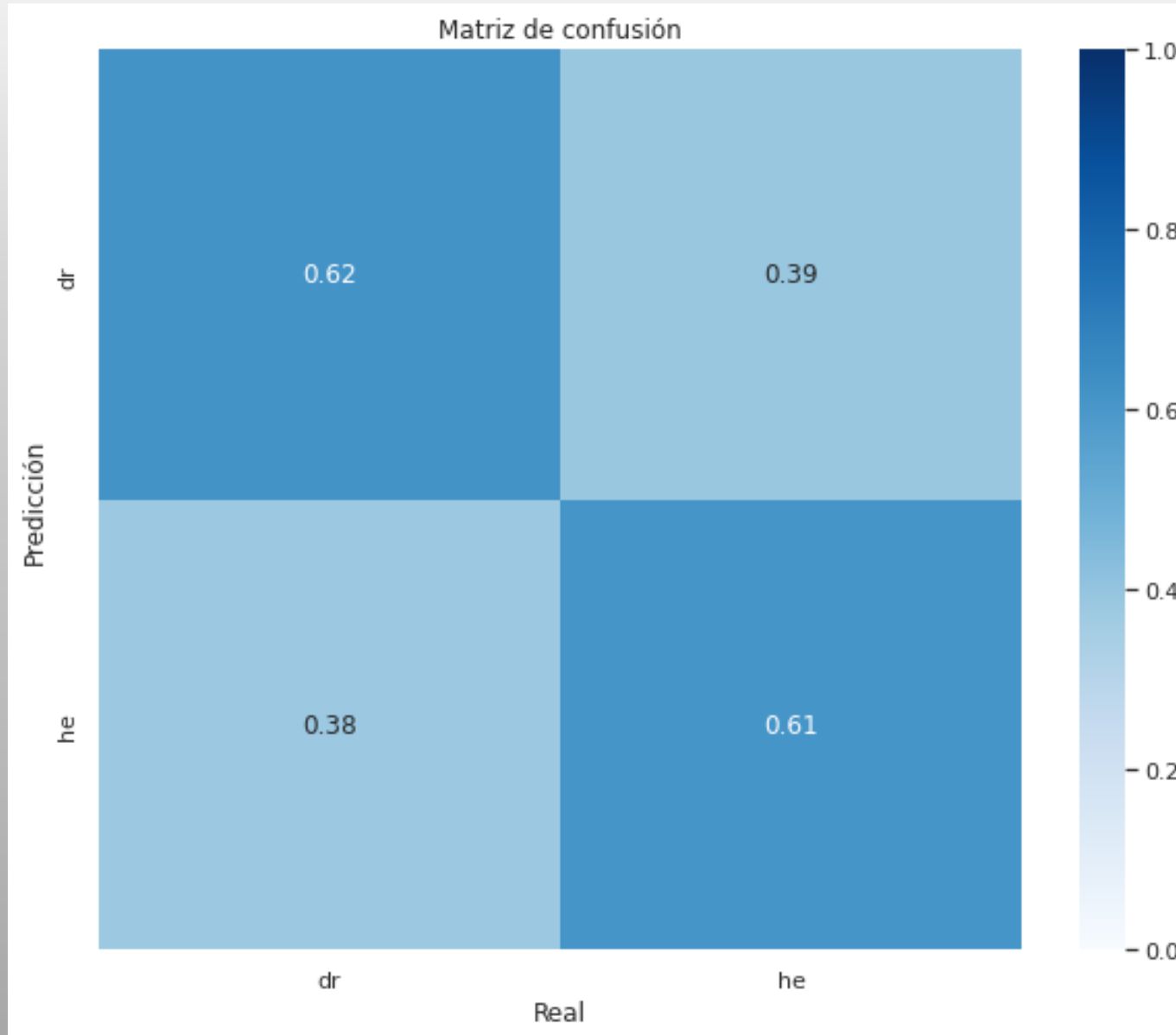
Resultados del entrenamiento

mAP (Mean Average Precision)

Cuantifica la precisión de un detector

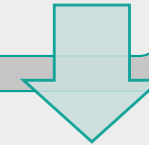


Resultados del entrenamiento



Resultados

YOLOv5s y YOLOv5m



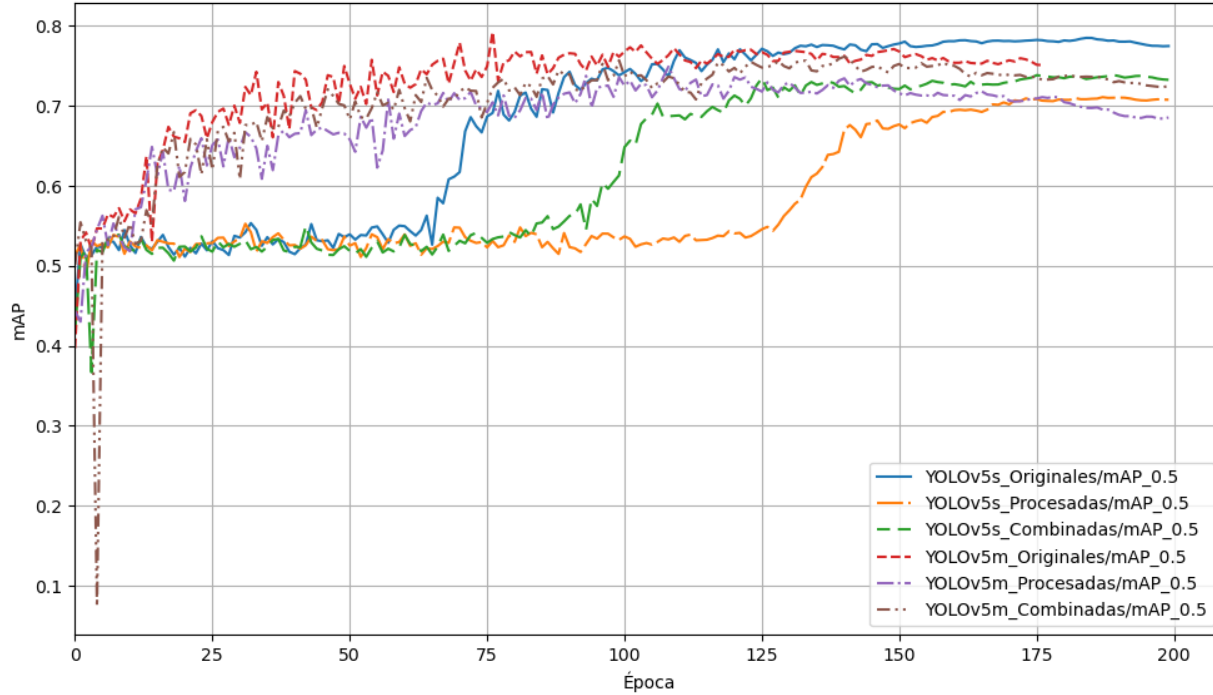
Implementación de YOLOv5



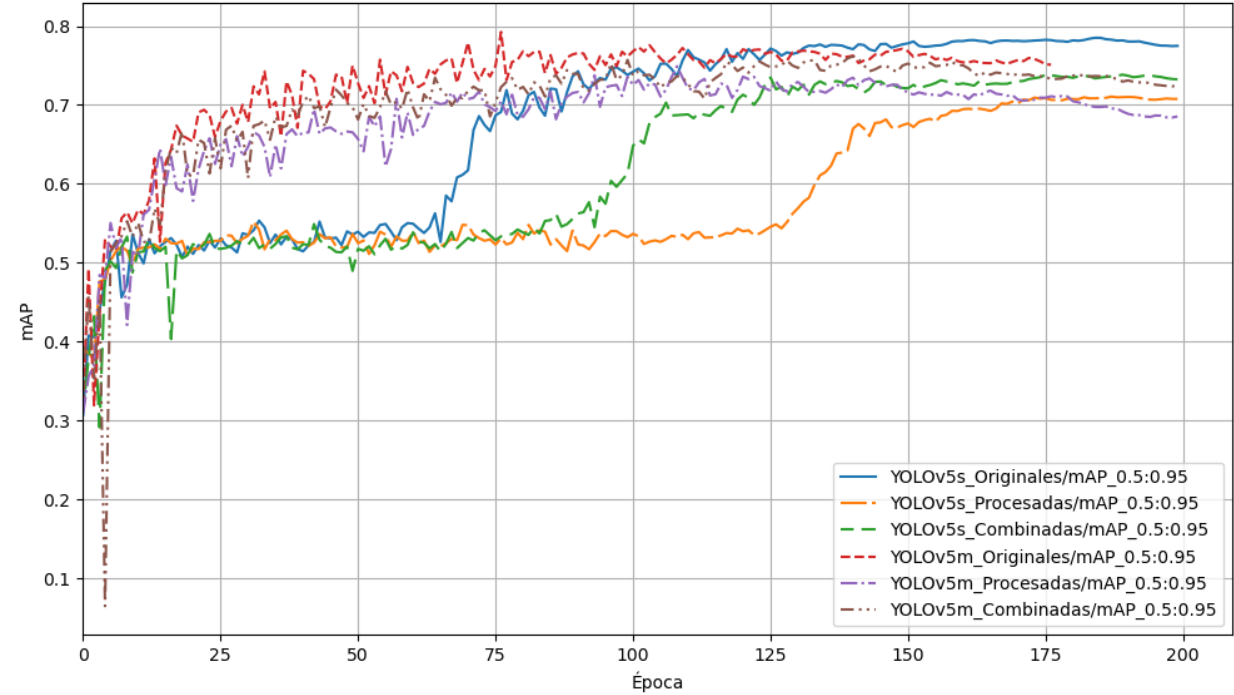
Sistema embebido

YOLOv5s y YOLOv5m

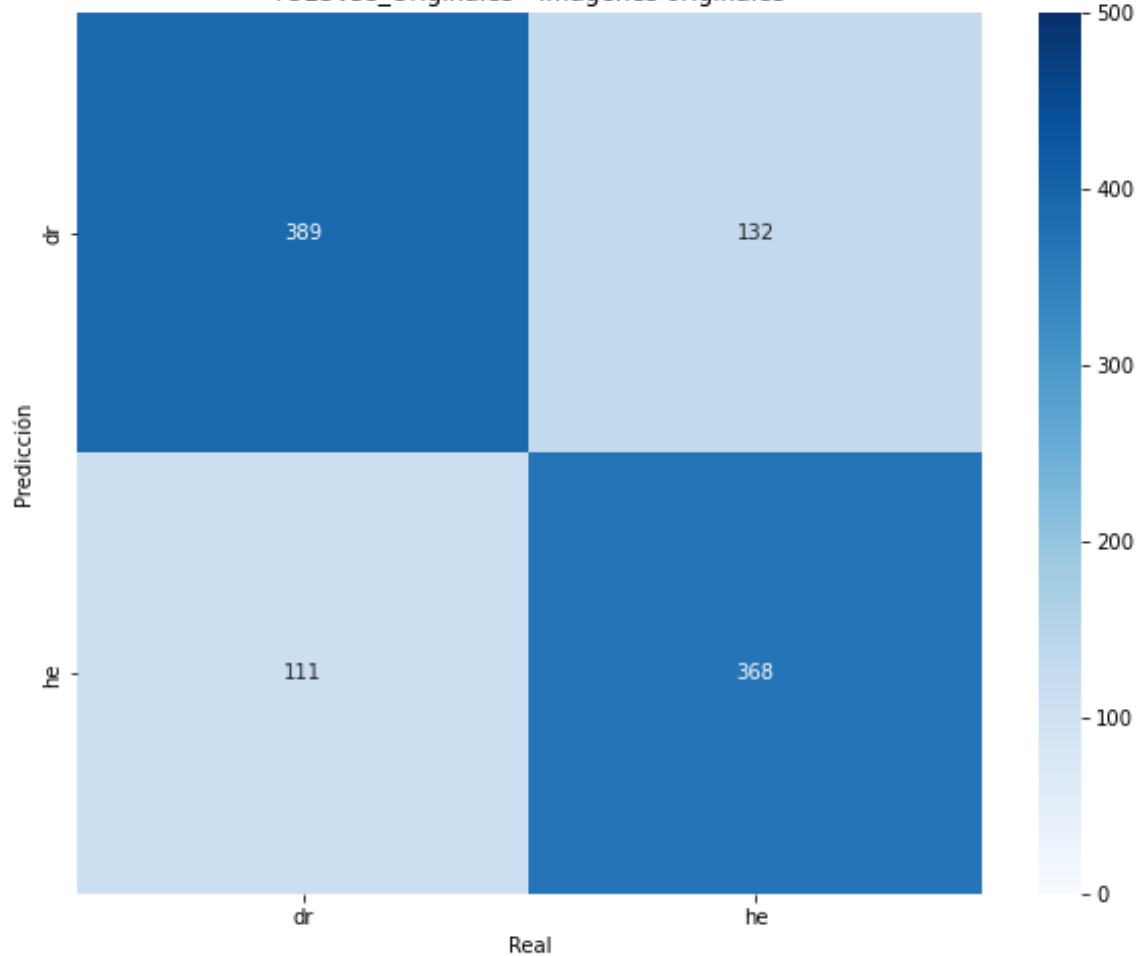
Evolución de mAP con un umbral de confianza de 0.5



Evolución de mAP con un umbral de confianza de 0.5 a 0.95

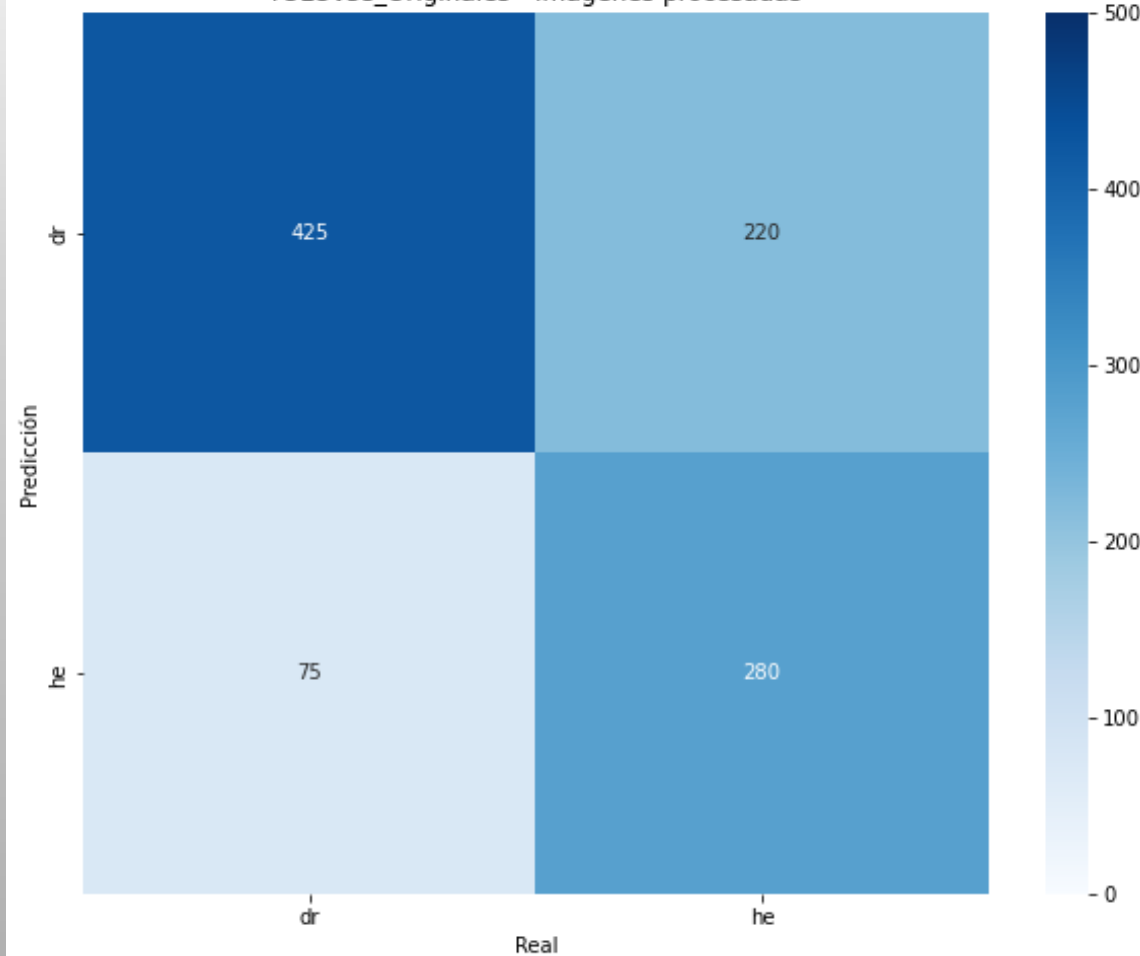


YOLOv5s_Originales - Imágenes originales



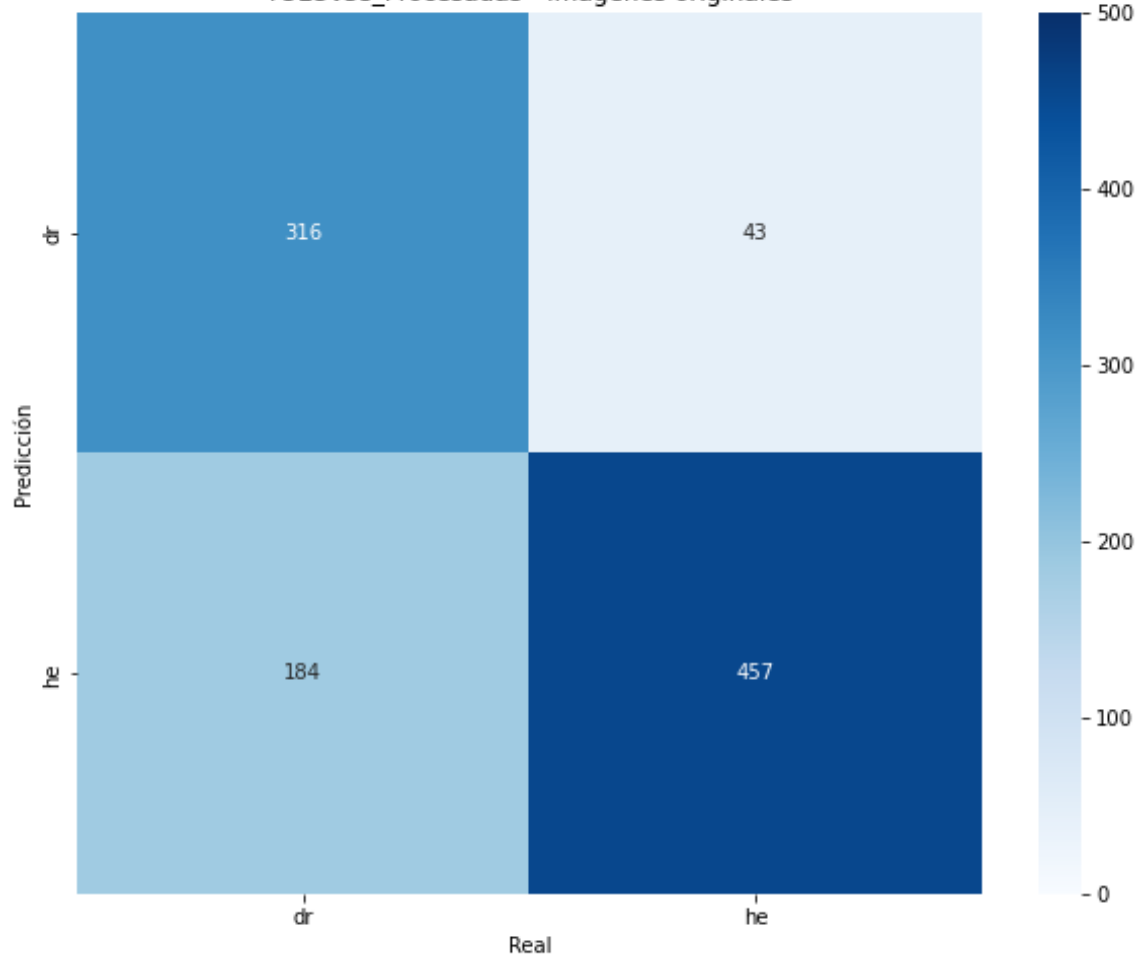
75.7%

YOLOv5s_Originales - Imágenes procesadas



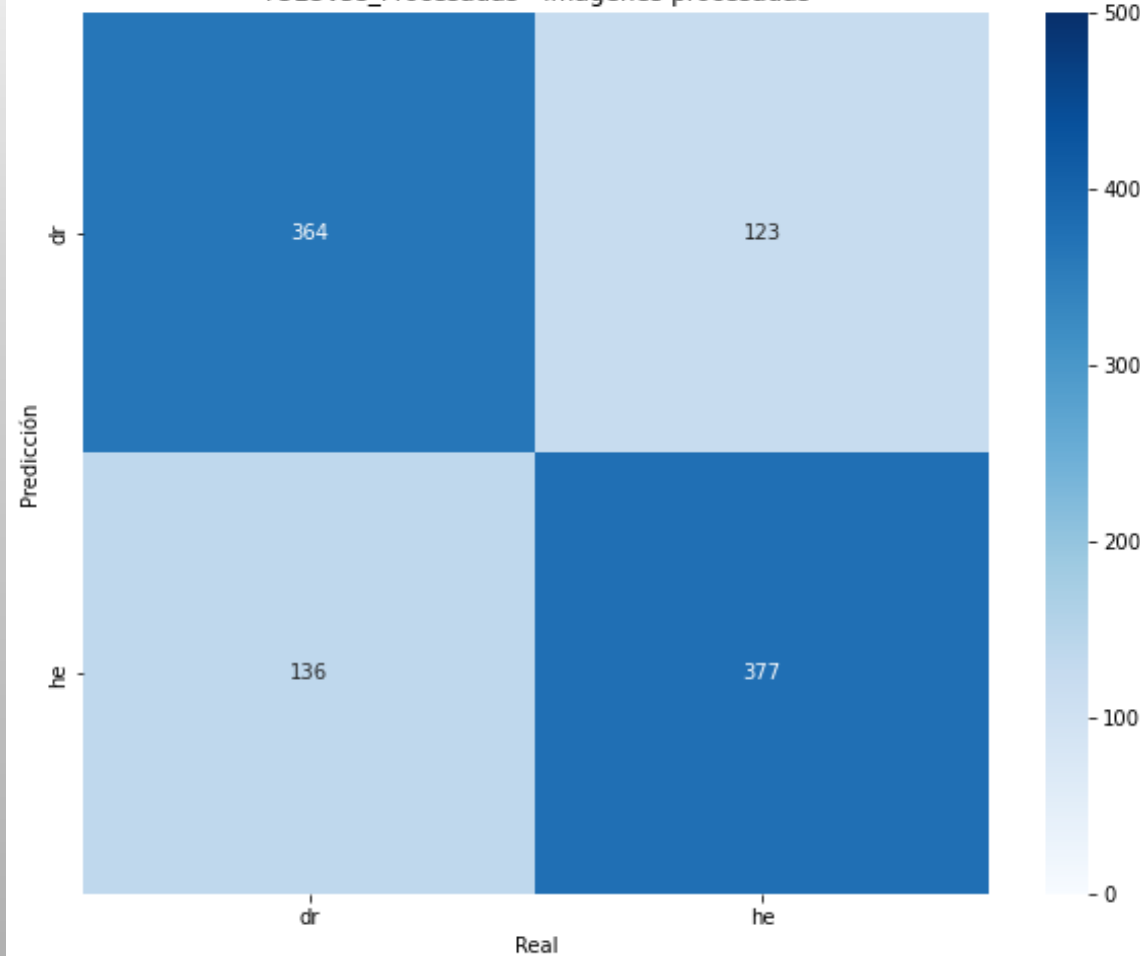
70.5%

YOLOv5s_Procesadas - Imágenes originales



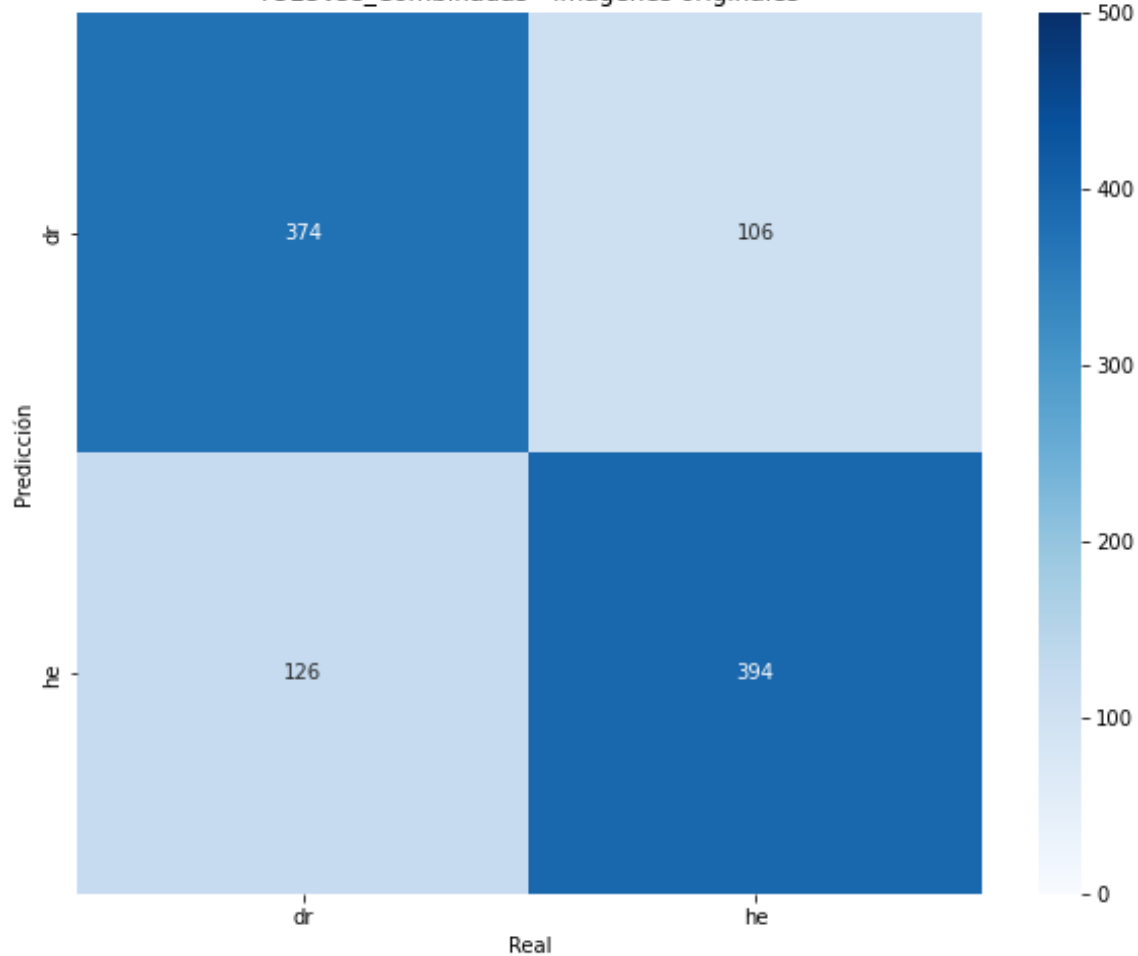
77.3%

YOLOv5s_Procesadas - Imágenes procesadas



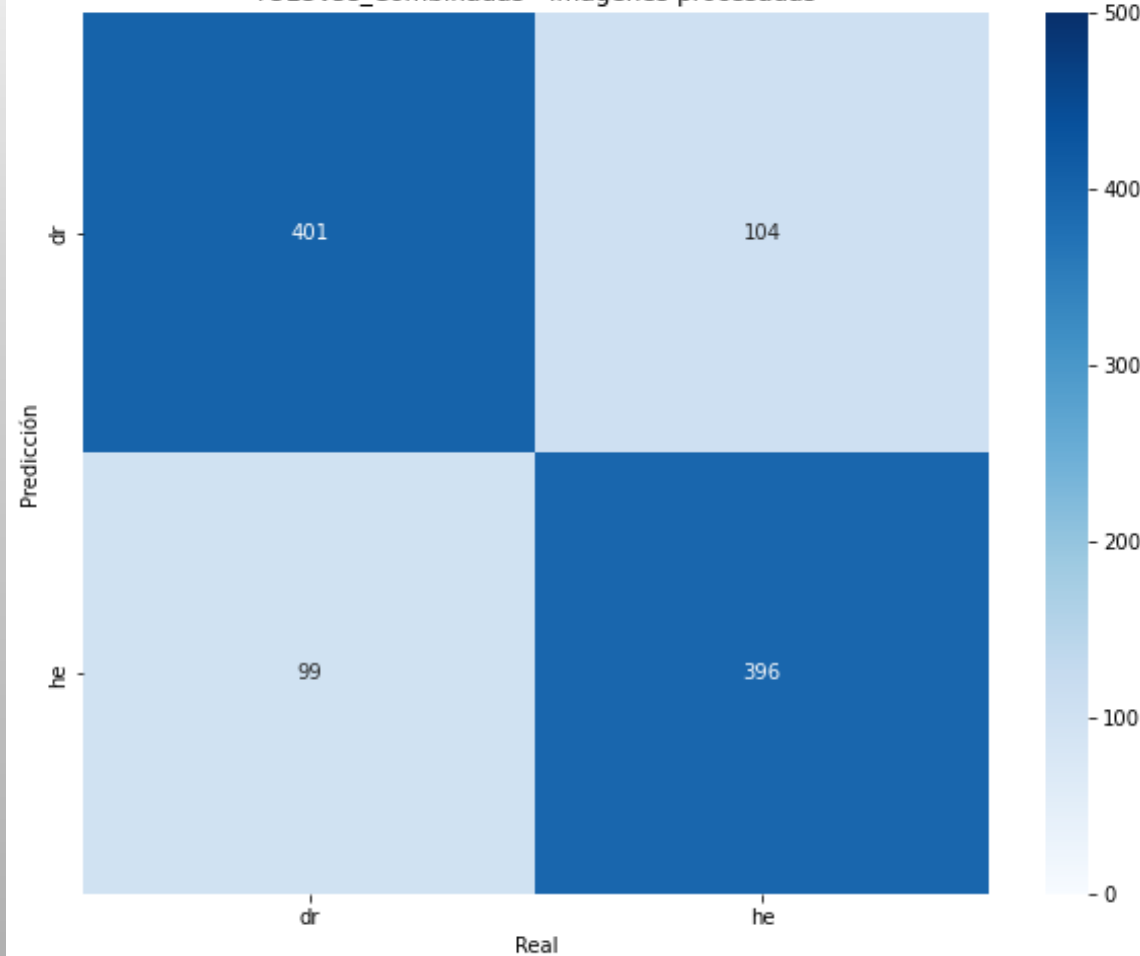
74.1%

YOLOv5s_Combinadas - Imágenes originales



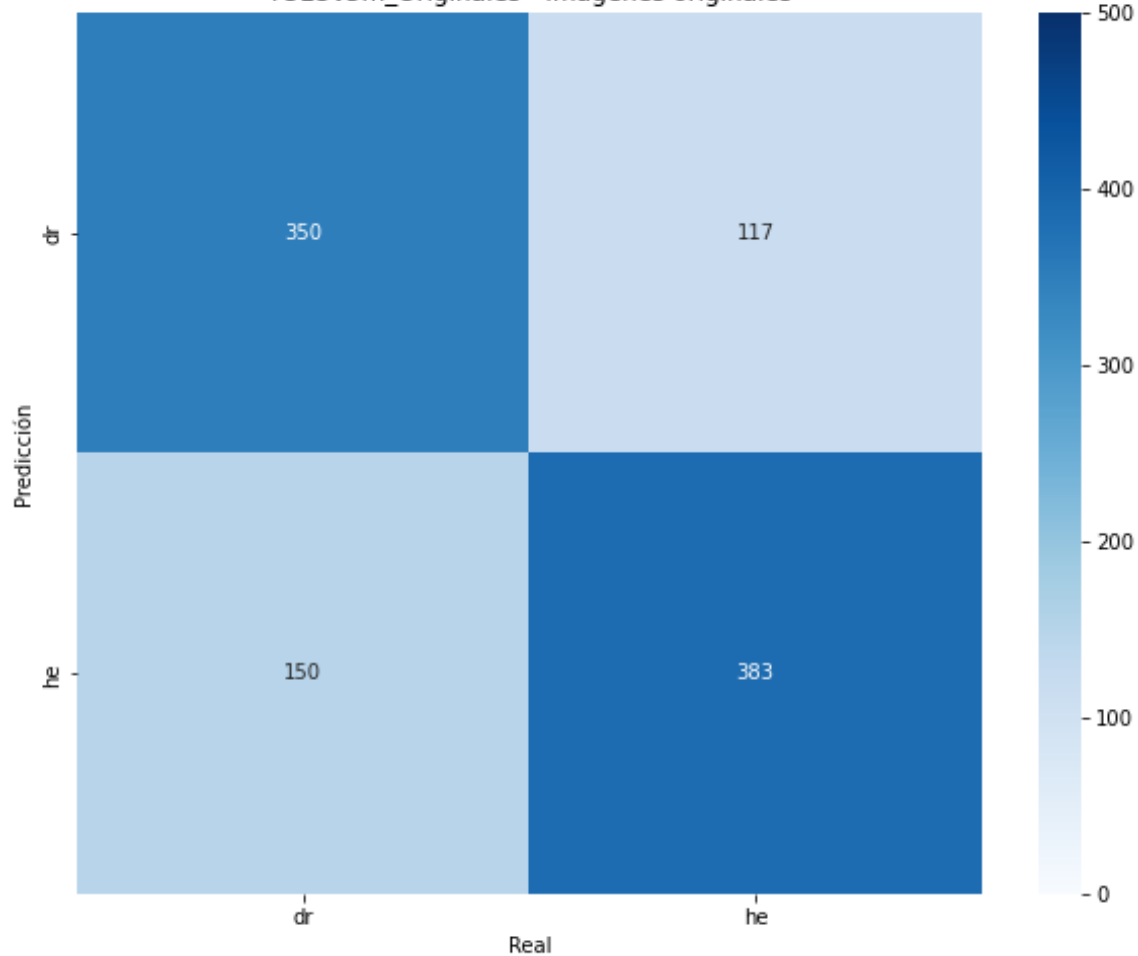
76.8%

YOLOv5s_Combinadas - Imágenes procesadas



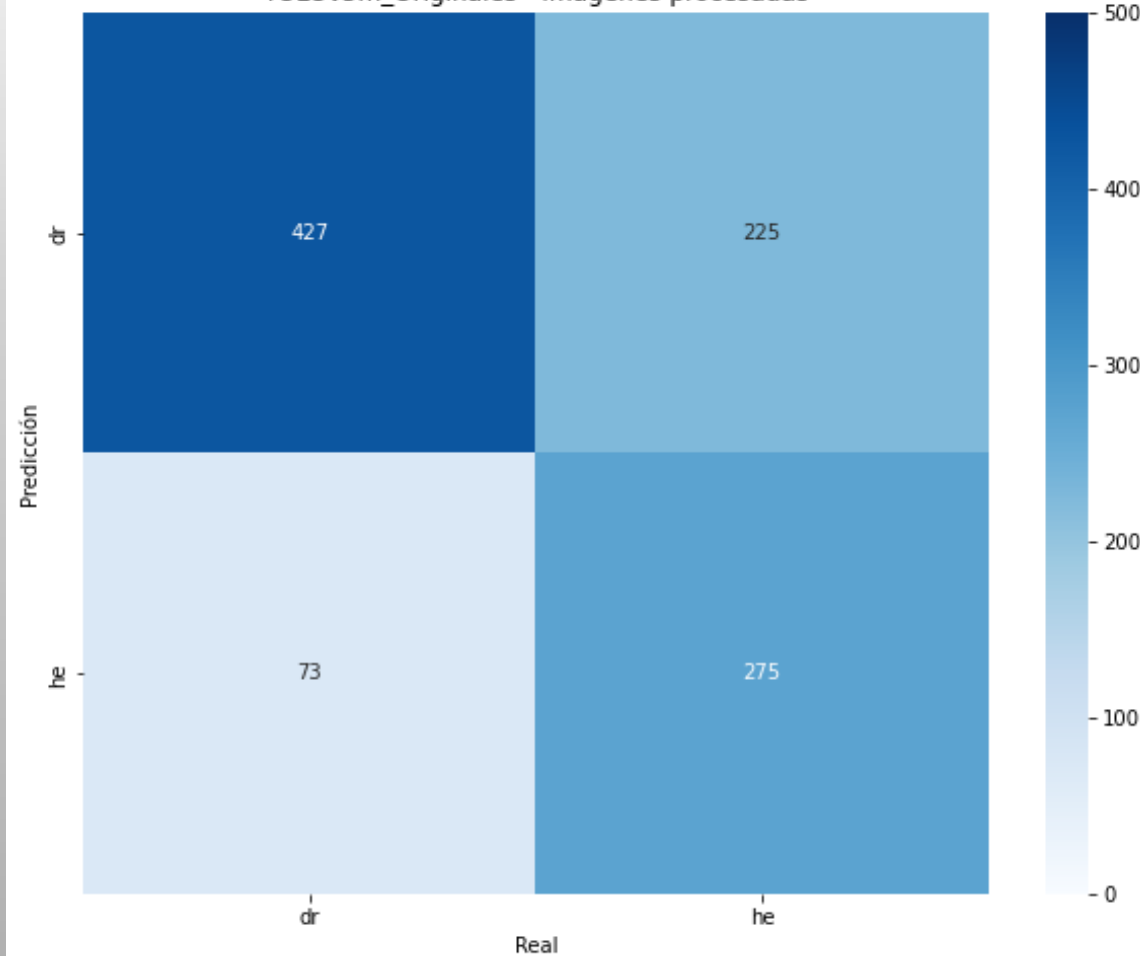
79.7%

YOLOv5m_Originales - Imágenes originales



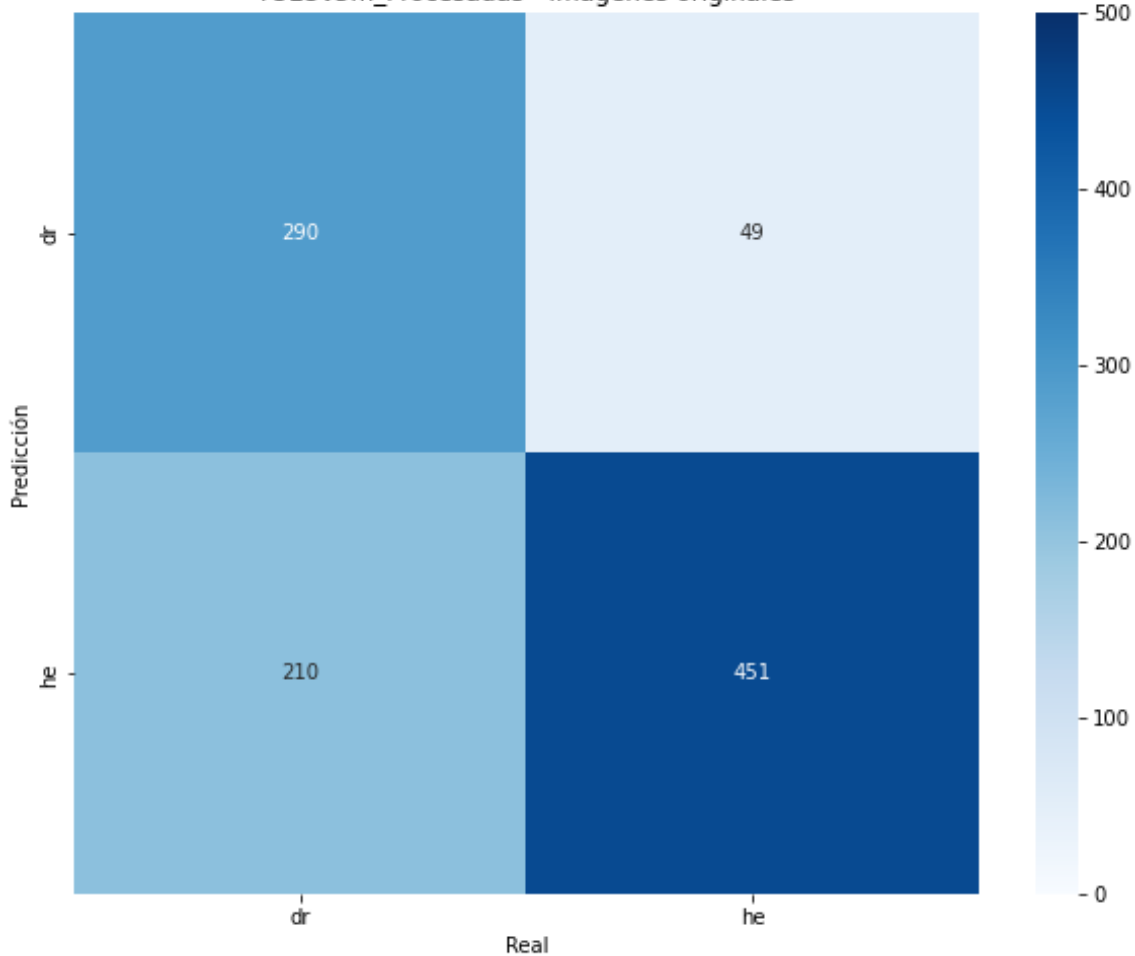
73.3%

YOLOv5m_Originales - Imágenes procesadas



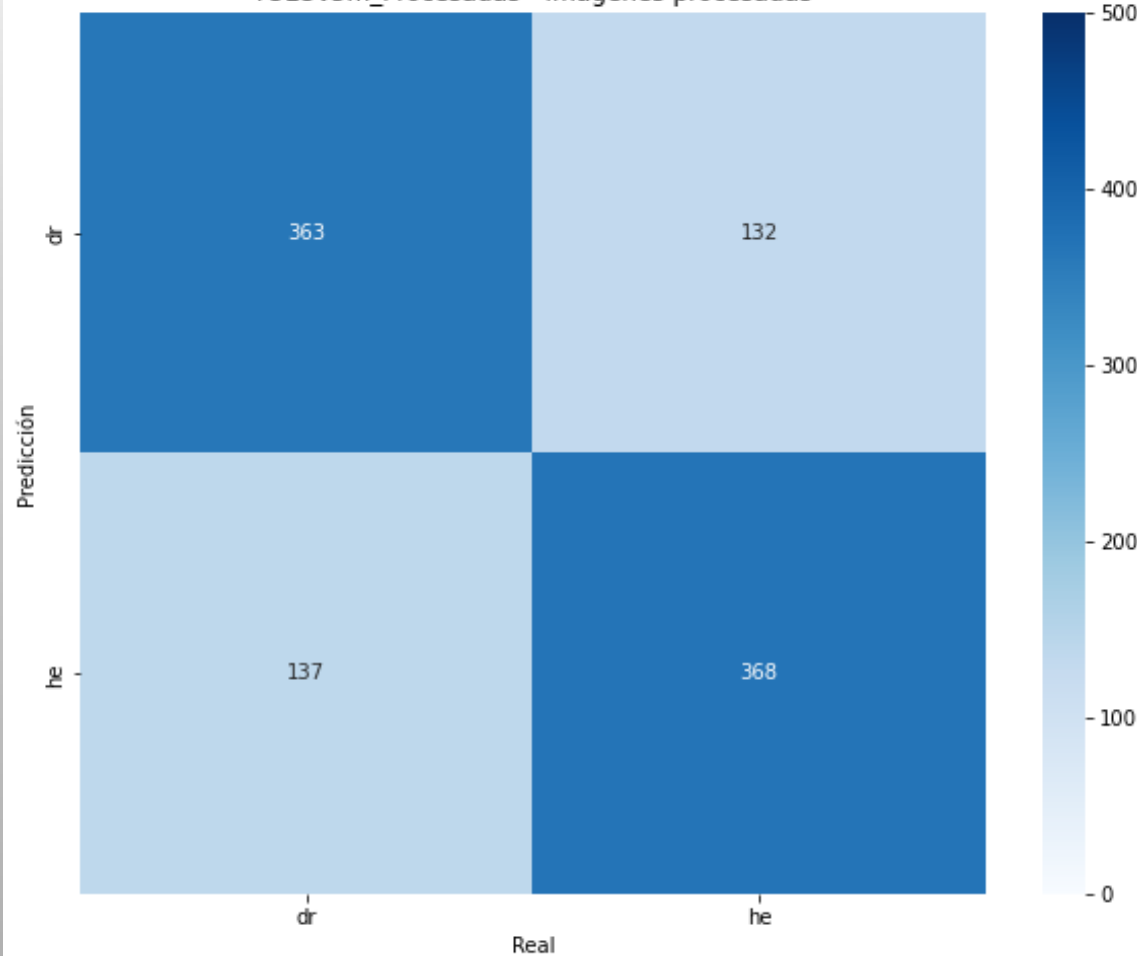
70.2%

YOLOv5m_Procesadas - Imágenes originales



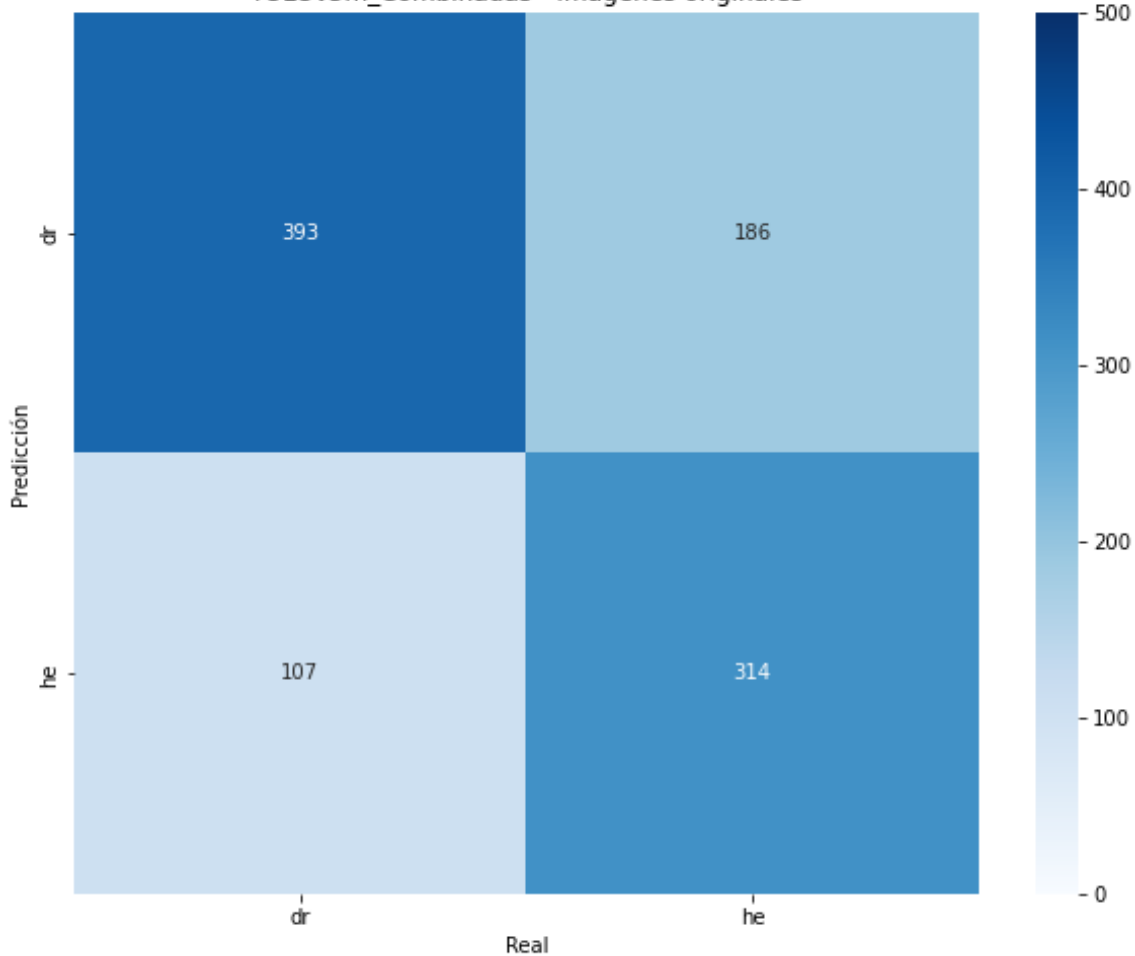
74.1%

YOLOv5m_Procesadas - Imágenes procesadas



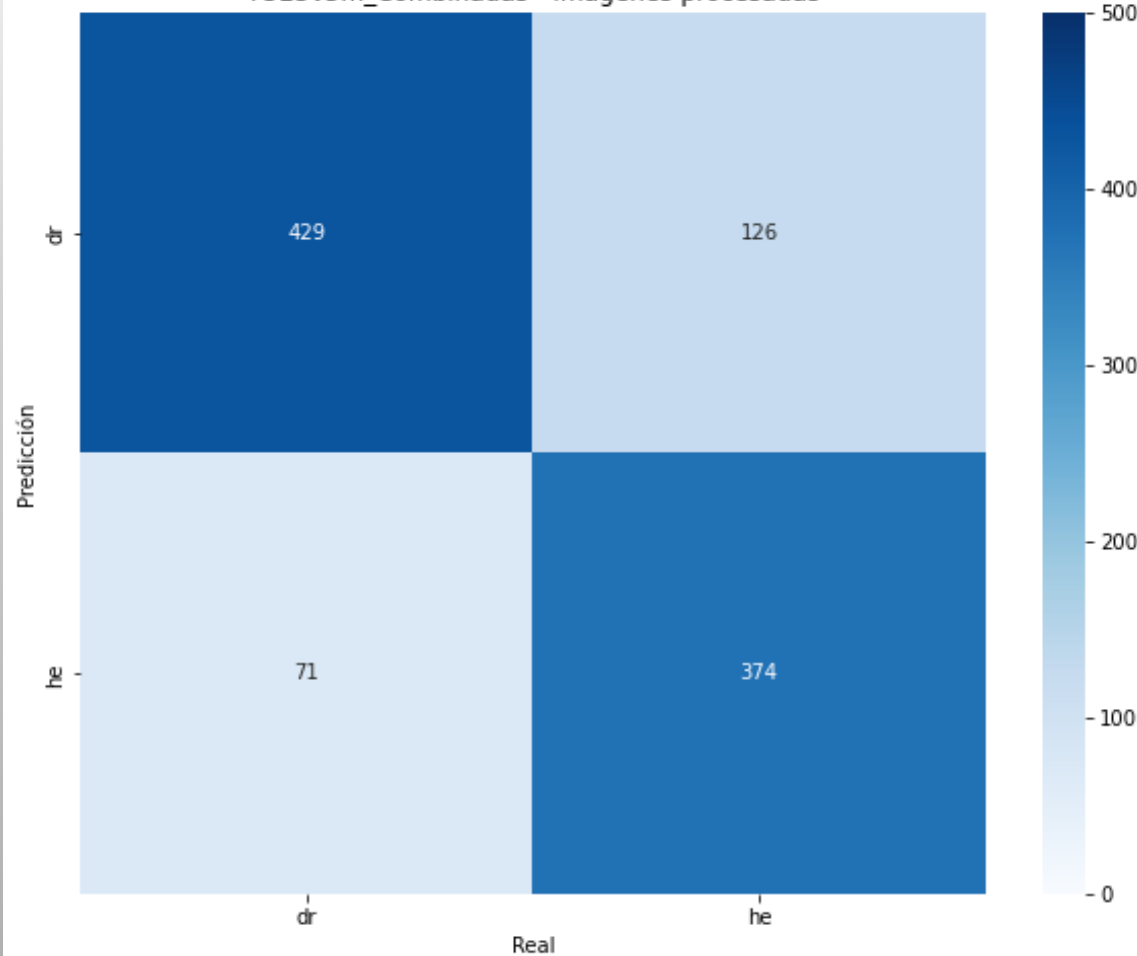
73.1%

YOLOv5m_Combinadas - Imágenes originales



70.7%

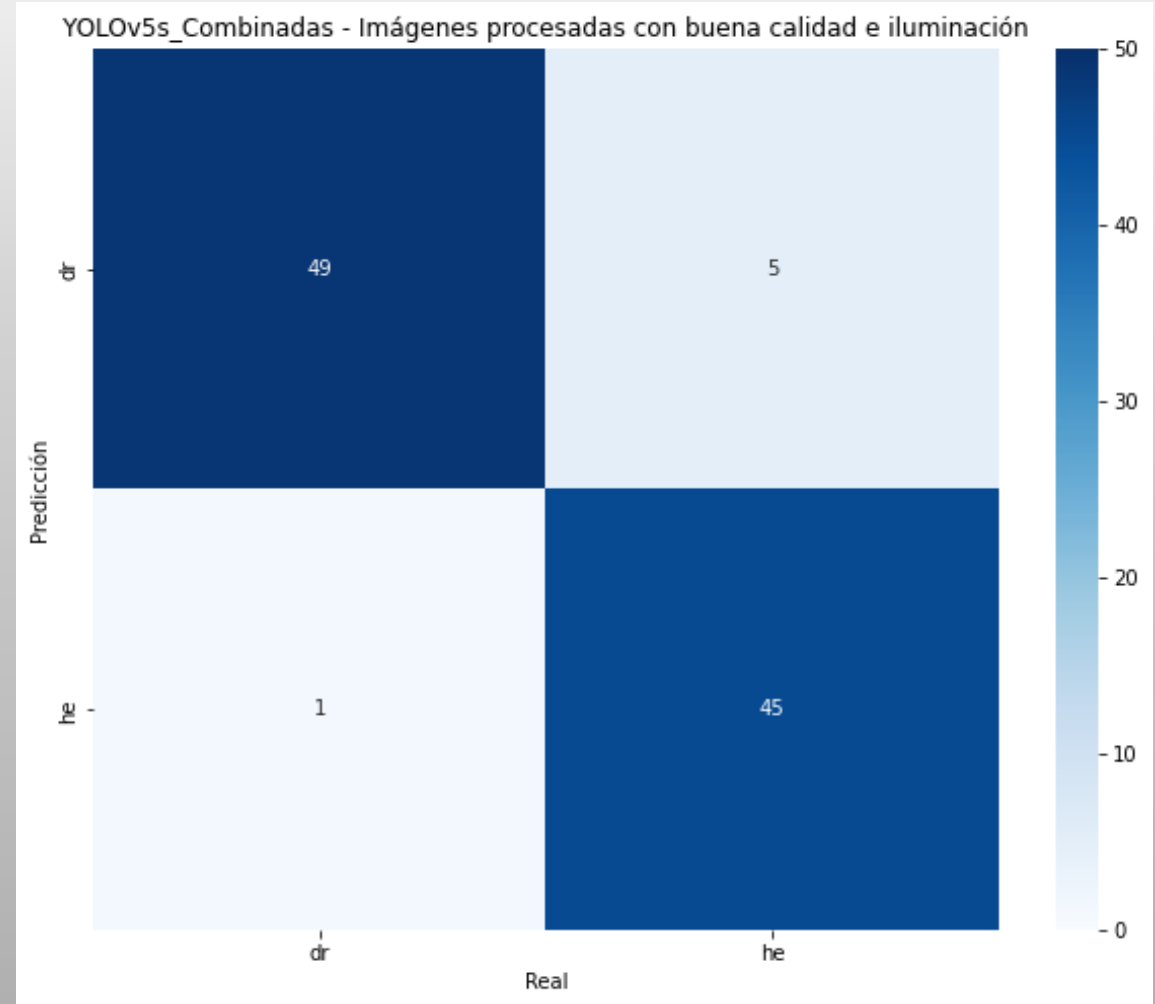
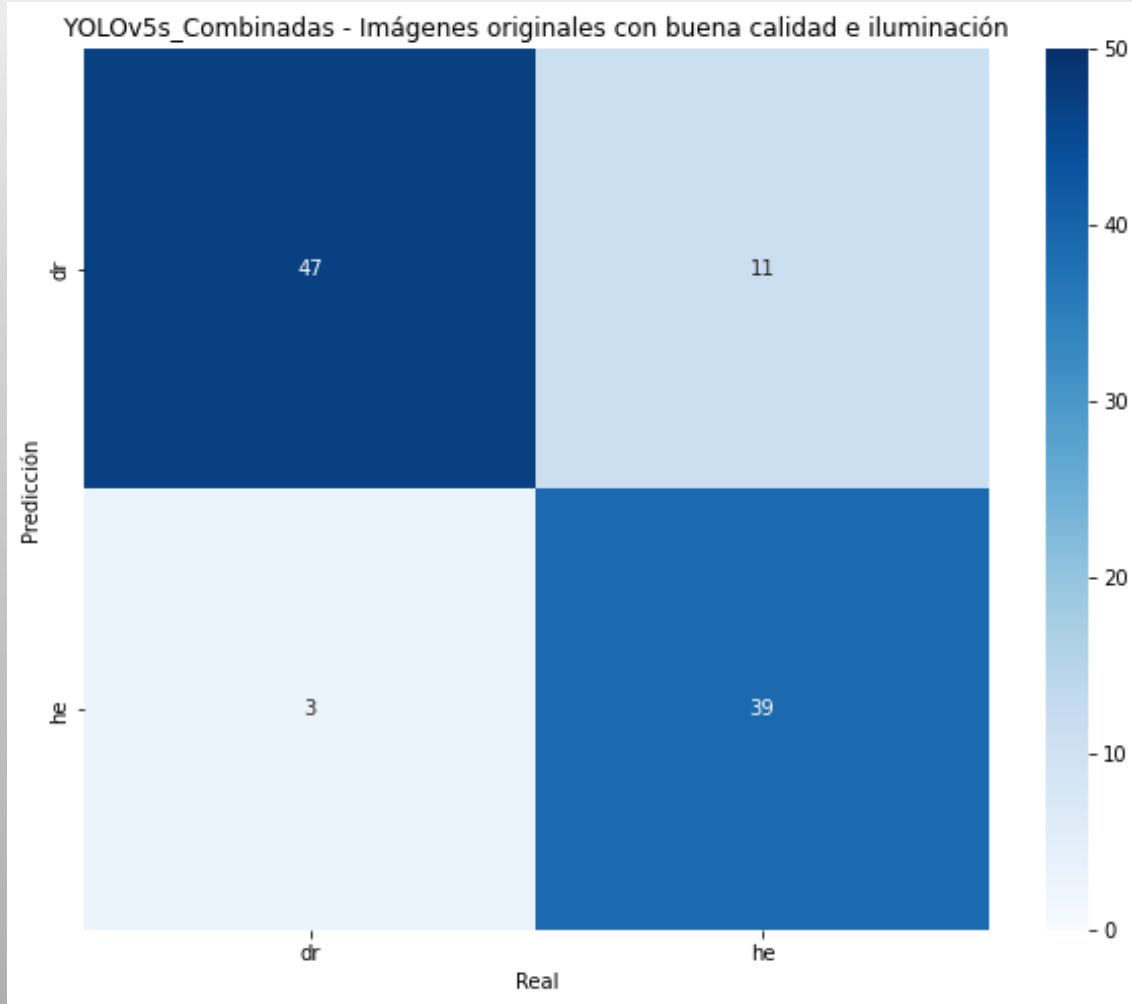
YOLOv5m_Combinadas - Imágenes procesadas

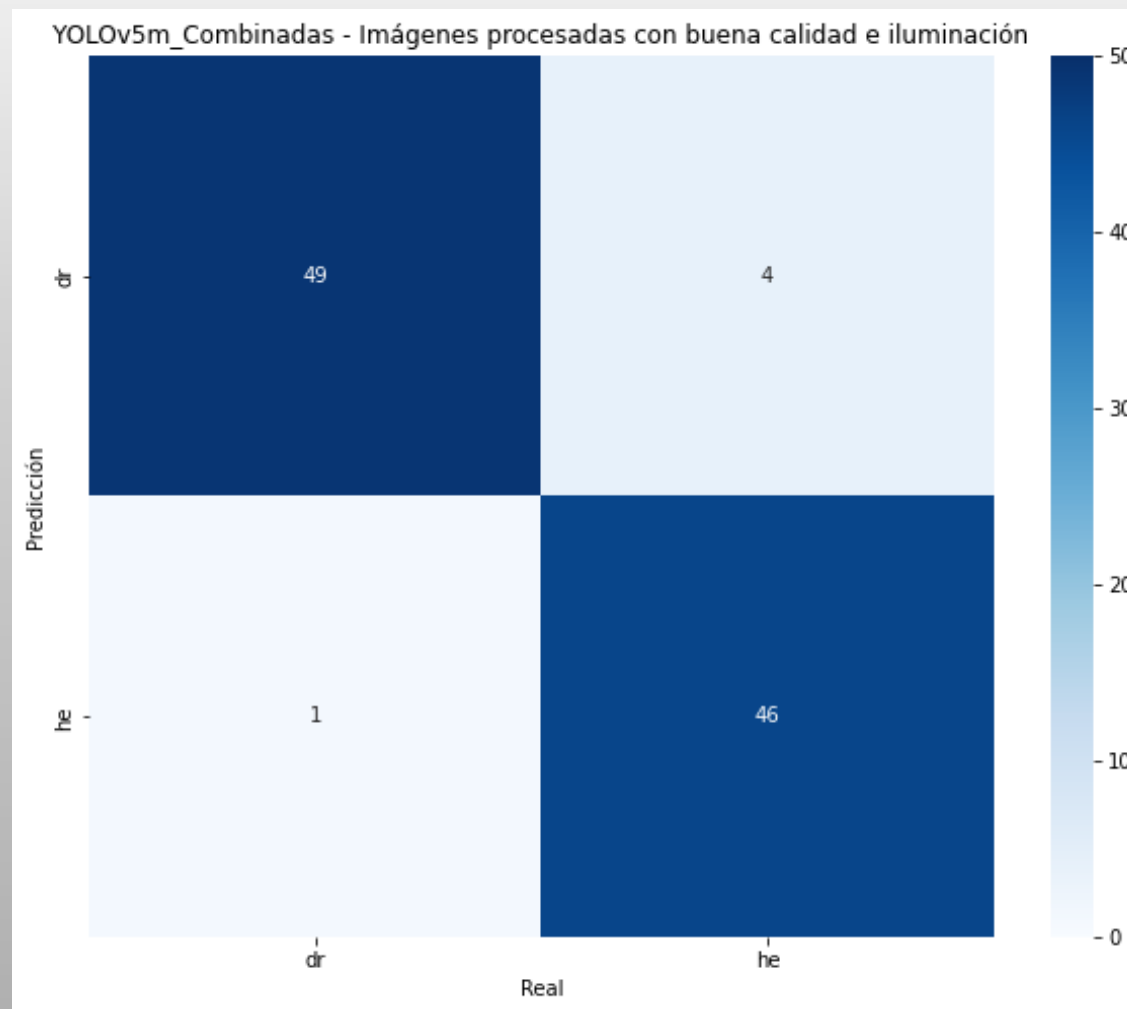
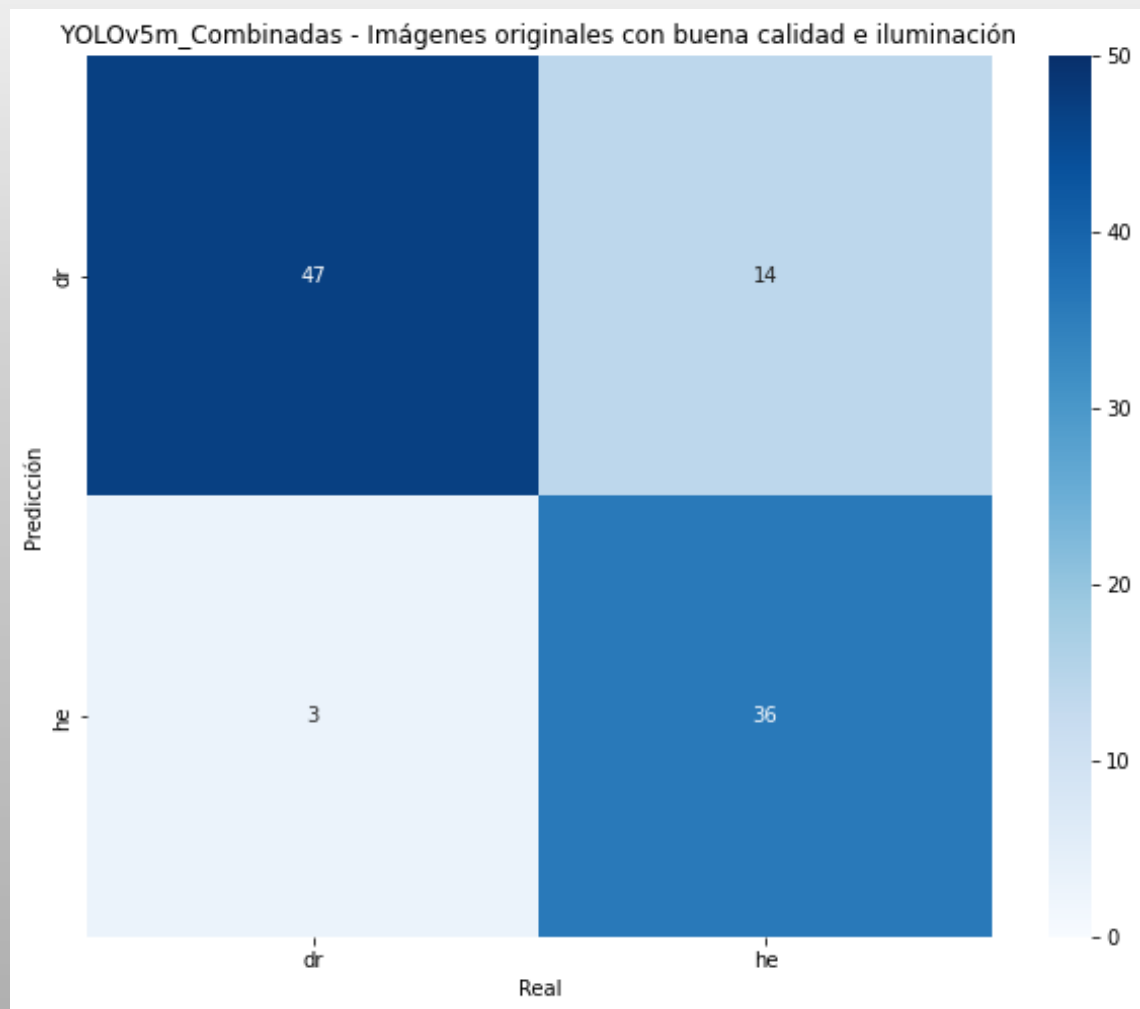


80.3%

YOLOv5s y YOLOv5m

Modelo	Imágenes de entrenamiento	Imágenes de inferencia	VP _{dr} (%)	VP _{he} (%)	VP _{total} (%)
YOLOv5s	Originales	Originales	77.8	73.6	75.7
		Procesadas	85.0	56.0	70.5
	Procesadas	Originales	63.2	91.4	77.3
		Procesadas	72.8	75.4	74.1
	Combinadas	Originales	74.8	78.8	76.8
		Procesadas	80.2	79.2	79.7
YOLOv5m	Originales	Originales	70.0	76.6	73.3
		Procesadas	85.4	55.0	70.2
	Procesadas	Originales	58.0	90.2	74.1
		Procesadas	72.6	73.6	73.1
	Combinadas	Originales	78.6	62.8	70.7
		Procesadas	85.8	74.8	80.3





YOLOv5s y YOLOv5m

Modelo	Imágenes de inferencia	VP _{dr} (%)	VP _{he} (%)	VP _{total} (%)
YOLOv5s	Originales	94.0	78.0	86.0
	Procesadas	98.0	90.0	94.0
YOLOv5m	Originales	94.0	72.0	83.0
	Procesadas	98.0	92.0	95.0

Implementación de YOLOv5

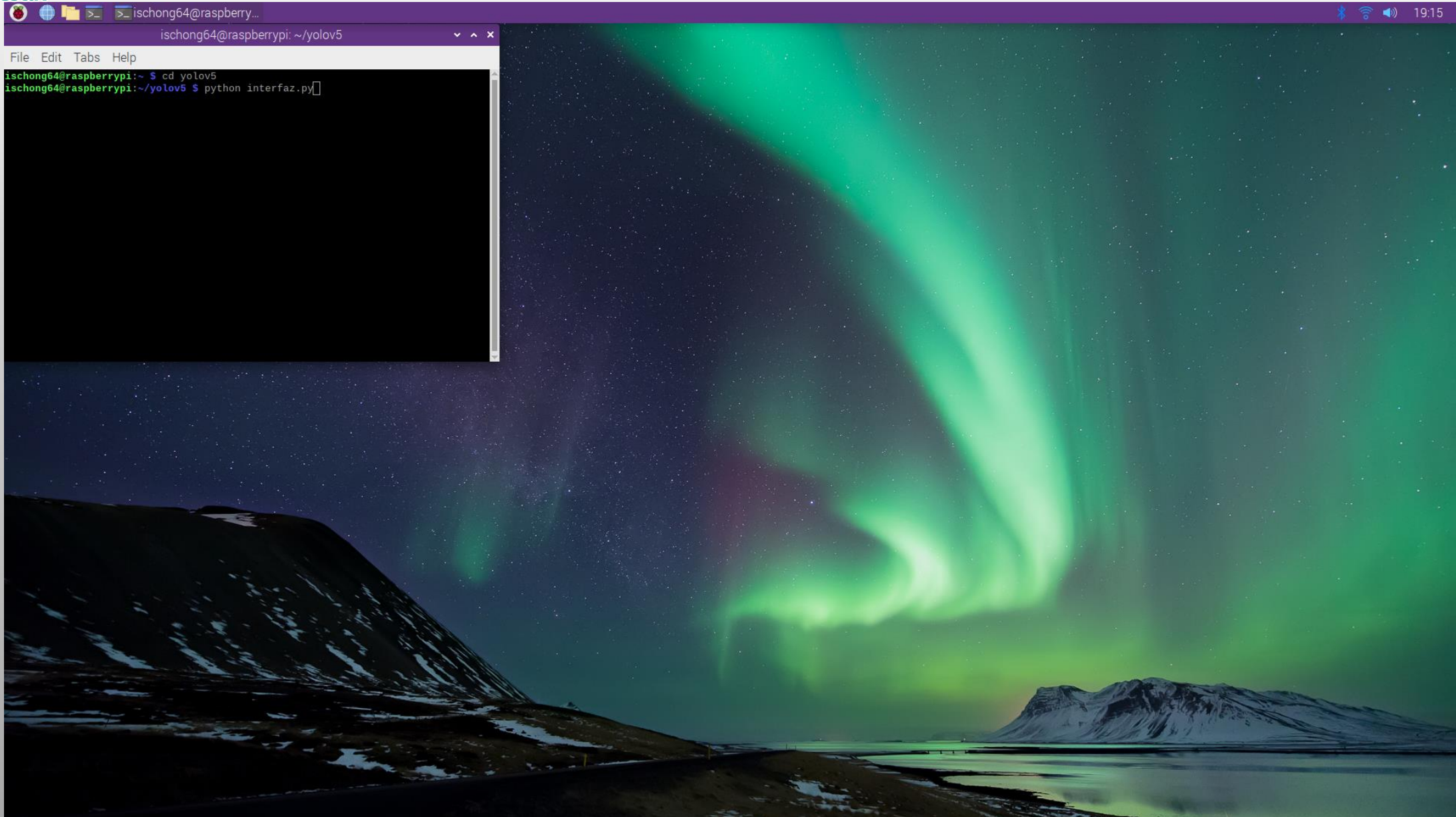
Procesador	BCM2711	AMD Ryzen 5
Arquitectura	ARMv8-A	ZEN+
Núcleos	4	4
Hilos	4	8
Frecuencia	1.4 GHz	2.1 a 3.7 GHz

Dispositivo	Proceso	YOLOv5s	YOLOv5m
PC	Preprocesamiento	0.6 ms	0.7 ms
	Inferencia	129.4 ms	326.9 ms
	NMS	0.6 ms	0.7 ms
Raspberry	Preprocesamiento	12.2 ms	13.6 ms
	Inferencia	1583.1 ms (1.6 s)	4090.7 ms (4 s)
	NMS	3.7 ms	4.1 ms

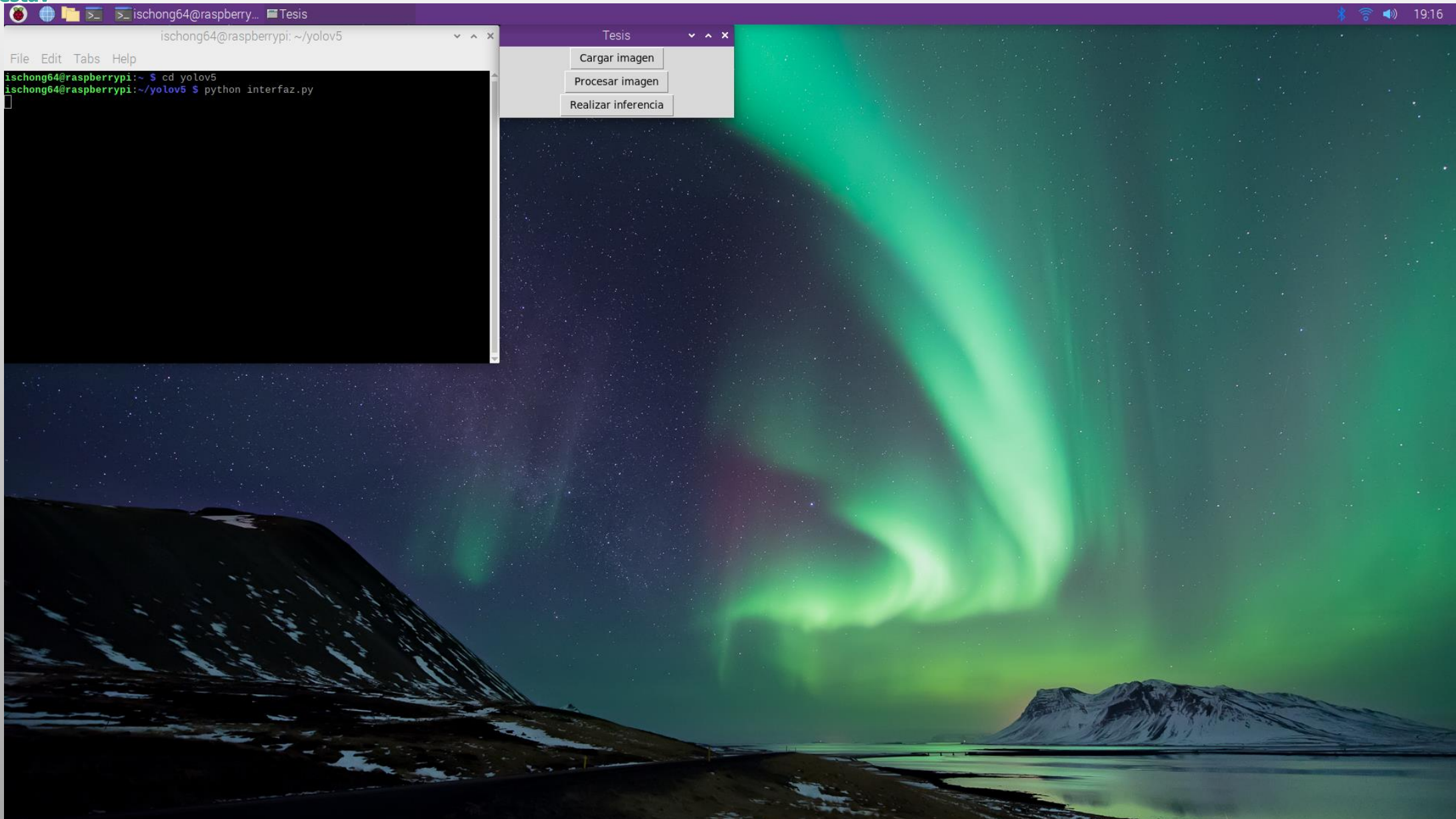
Implementación del sistema de mejora de contraste

Dispositivo	Tamaño (px)	Tiempo (s)	CPU (%)	RAM (%)
PC	640x426	2.21	7.6	1.9
	3504x2336	67.72 (1.12 min)	12.5	6.6
Raspberry	640x426	7.37	25	1.2
	3504x2336	225.42 (3.75 min)	25	5.5

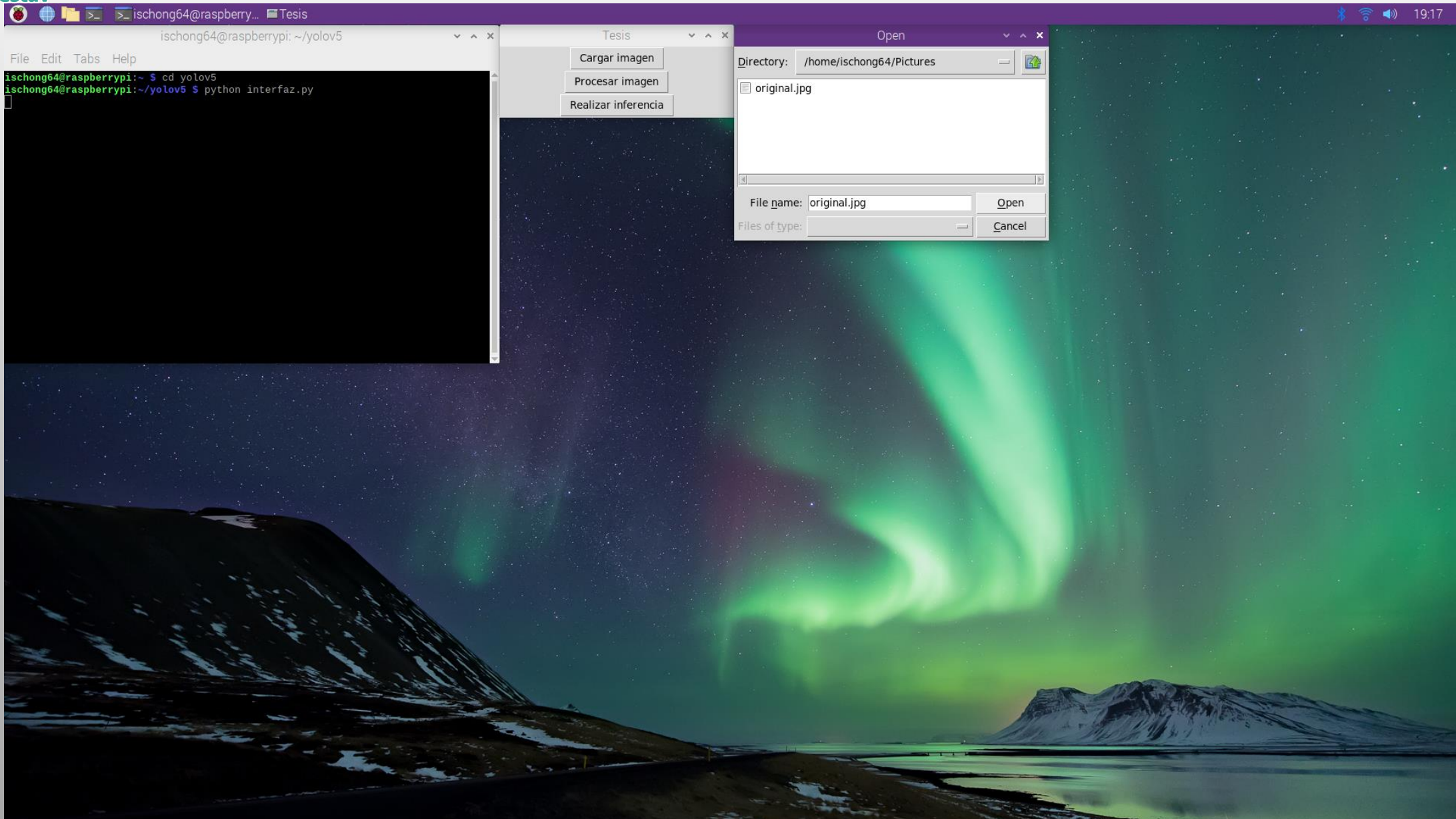
Implementación del sistema de procesamiento de imágenes de fondo de ojo



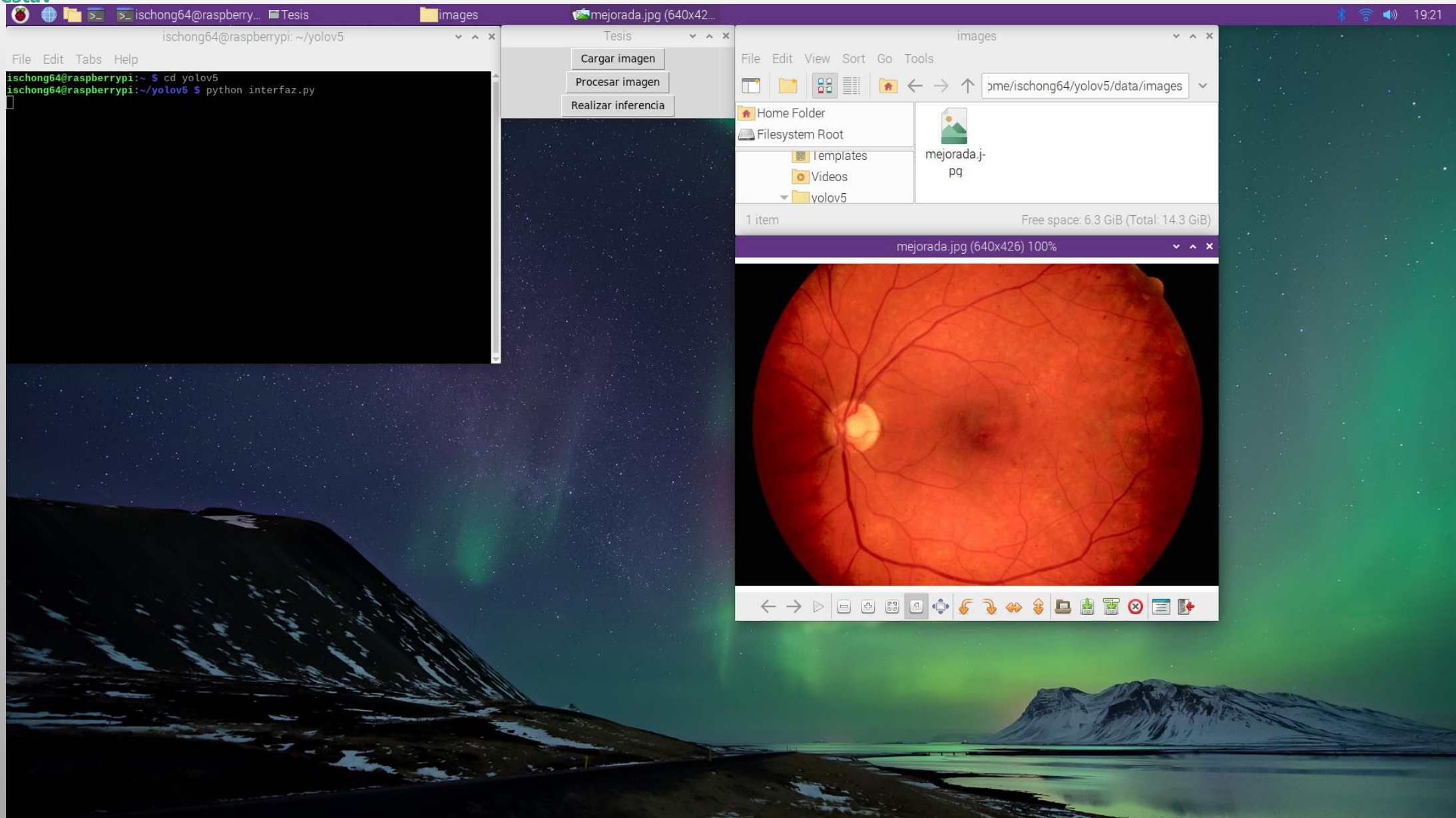
Implementación del sistema de procesamiento de imágenes de fondo de ojo



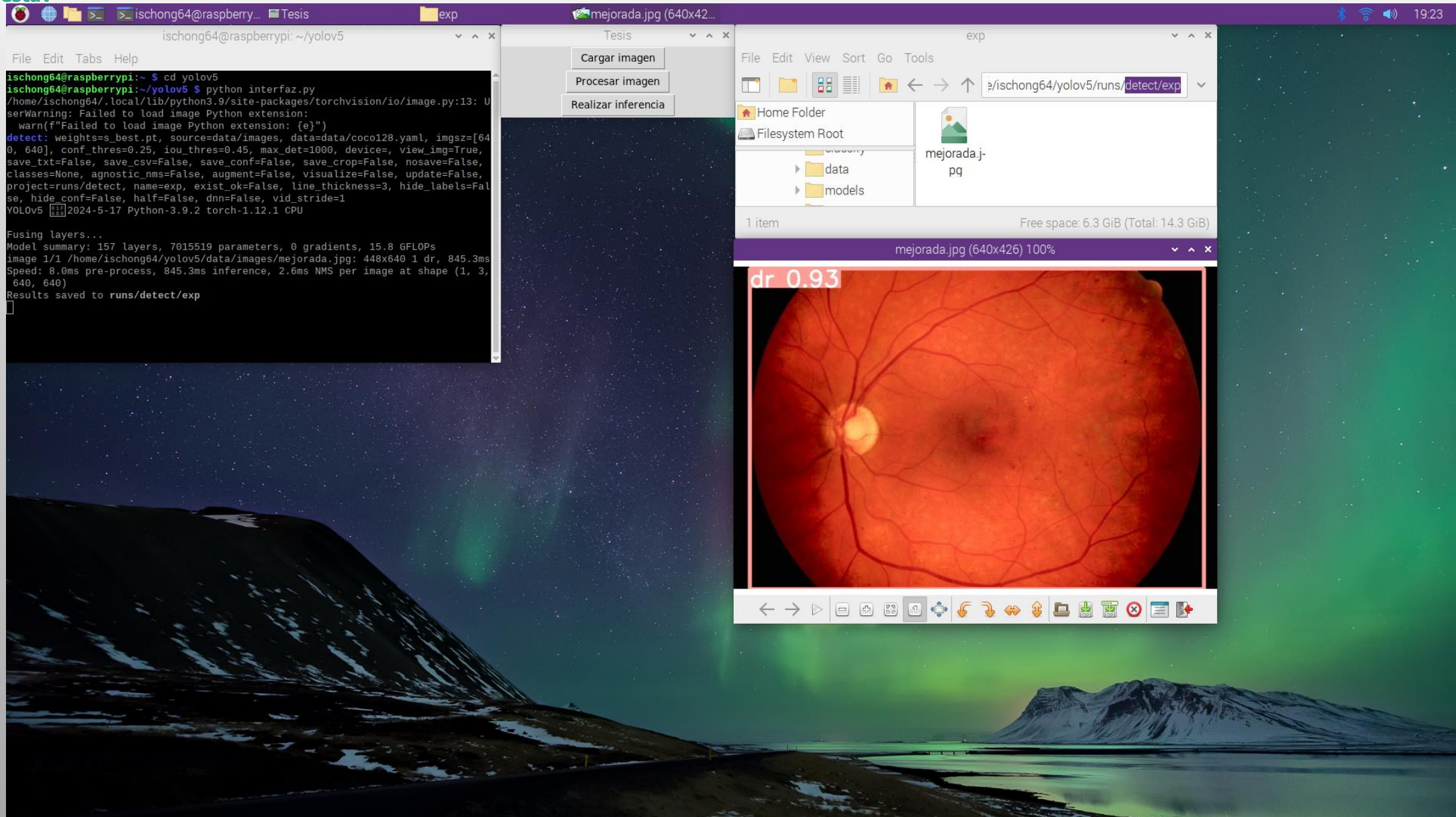
Implementación del sistema de procesamiento de imágenes de fondo de ojo



Implementación del sistema de procesamiento de imágenes de fondo de ojo



Implementación del sistema de procesamiento de imágenes de fondo de ojo



Conclusiones

- La correcta interpretación de este tipo de imágenes depende características particulares. Por ello, se tomó en cuenta el criterio de personal médico para diseñar el sistema de mejora de contraste.
- El uso de la lógica difusa permitió realizar una mejora de contraste adecuada para las imágenes de fondo de ojo utilizando el espacio de color CIELAB.
- La efectividad de la mejora de contraste utilizando la lógica difusa se puede observar en las gráficas del MCI y PSNR.

Conclusiones

- Las constantes fluctuaciones en el error de clase manifiestan un bajo rendimiento en la clasificación.
- Se obtuvo un mejor desempeño en los modelos entrenados con la combinación de imágenes originales y procesadas, realizando la inferencia en imágenes procesadas.
- Al probar los modelos en imágenes con una adecuada iluminación, el desempeño sube considerablemente.

Conclusiones

- Se demostró la viabilidad de implementación YOLOv5s y YOLOv5m y del sistema de mejora de contraste.
- El tiempo total de procesamiento en la Raspberry no representa un problema mayor, obteniendo el resultado final en 10 segundos con YOLOv5s y en 20 segundos con YOLOv5m.
- Se concluye la viabilidad de implementar sobre un sistema embebido, un sistema de procesamiento de imágenes utilizando la lógica difusa y redes neuronales convolucionales.

Trabajo a futuro

- Validar el sistema en escenarios reales.
- Optimizar el sistema mejora de contraste propuesto, ajustando la iluminación de las imágenes previo a la mejora de contraste.
- Explorar alternativas de hardware para superar las limitaciones actuales del entrenamiento.
- Adaptar el sistema para la detección de otras patologías.
- Explorar el entrenamiento con otras arquitecturas.
- Mejorar la interfaz de usuario.