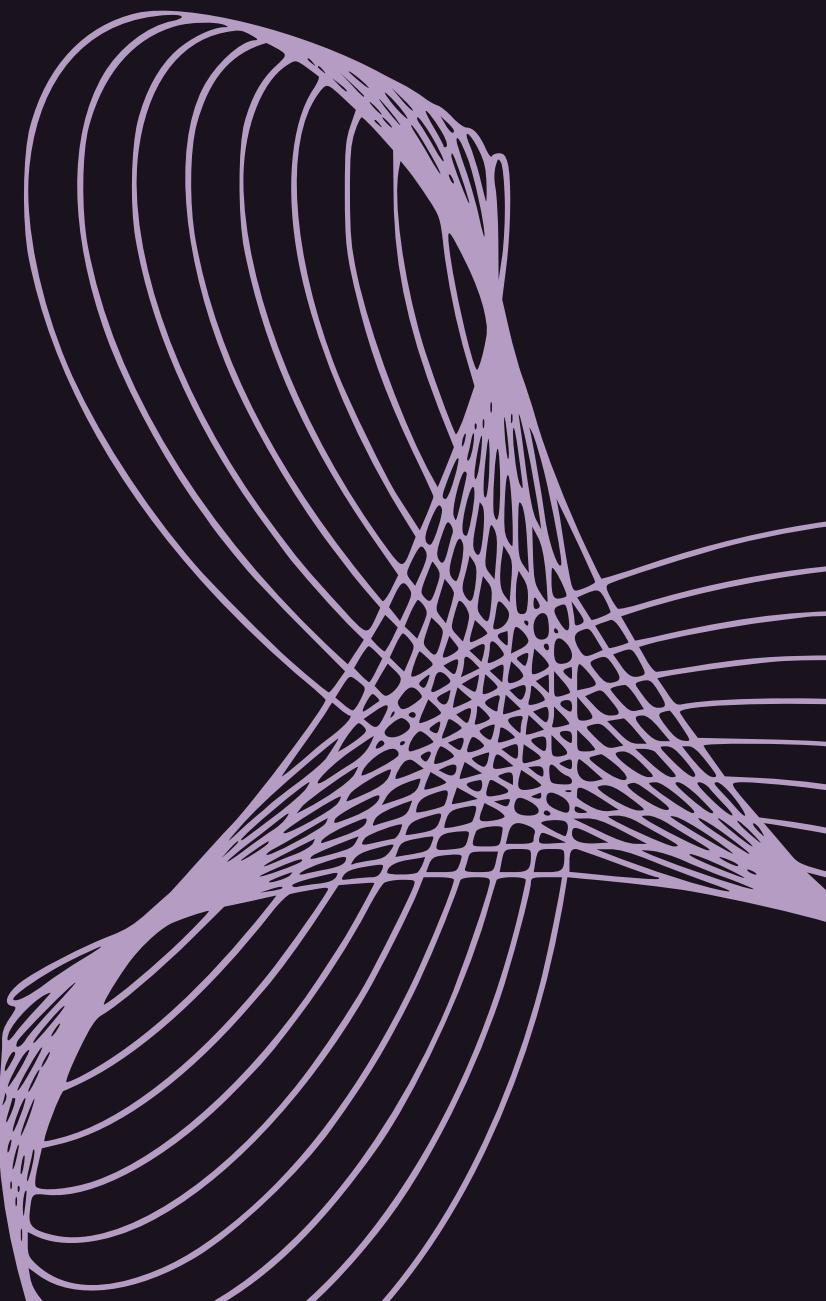
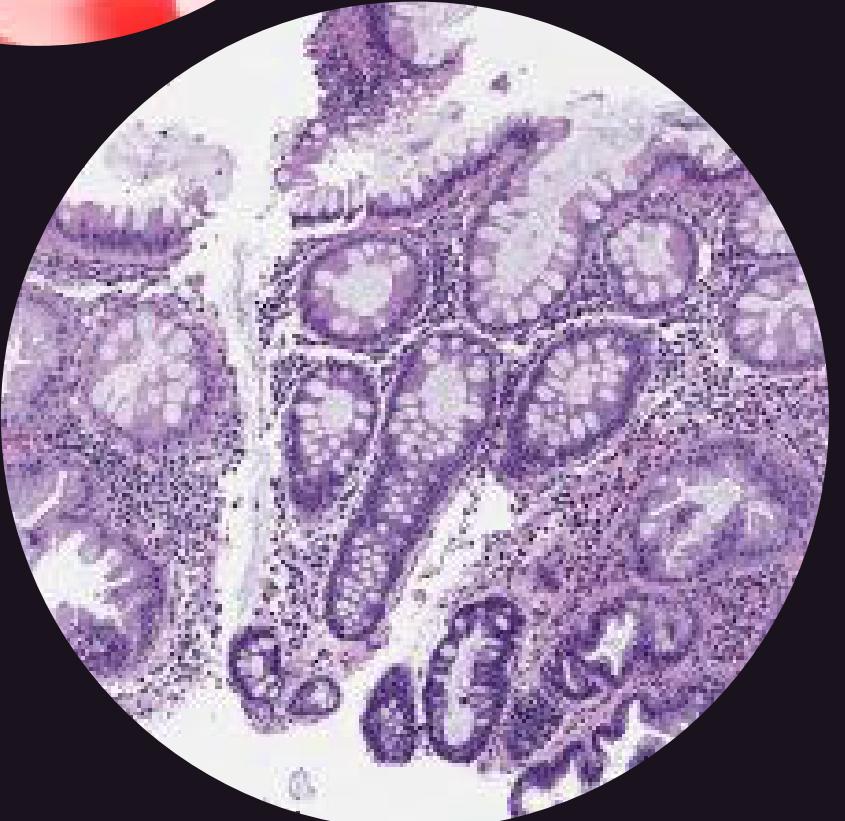


Clasificación de Pólipsos Colorrectales en Imágenes Histológicas con Redes Neuronales Convolucionales y Transfer Learning

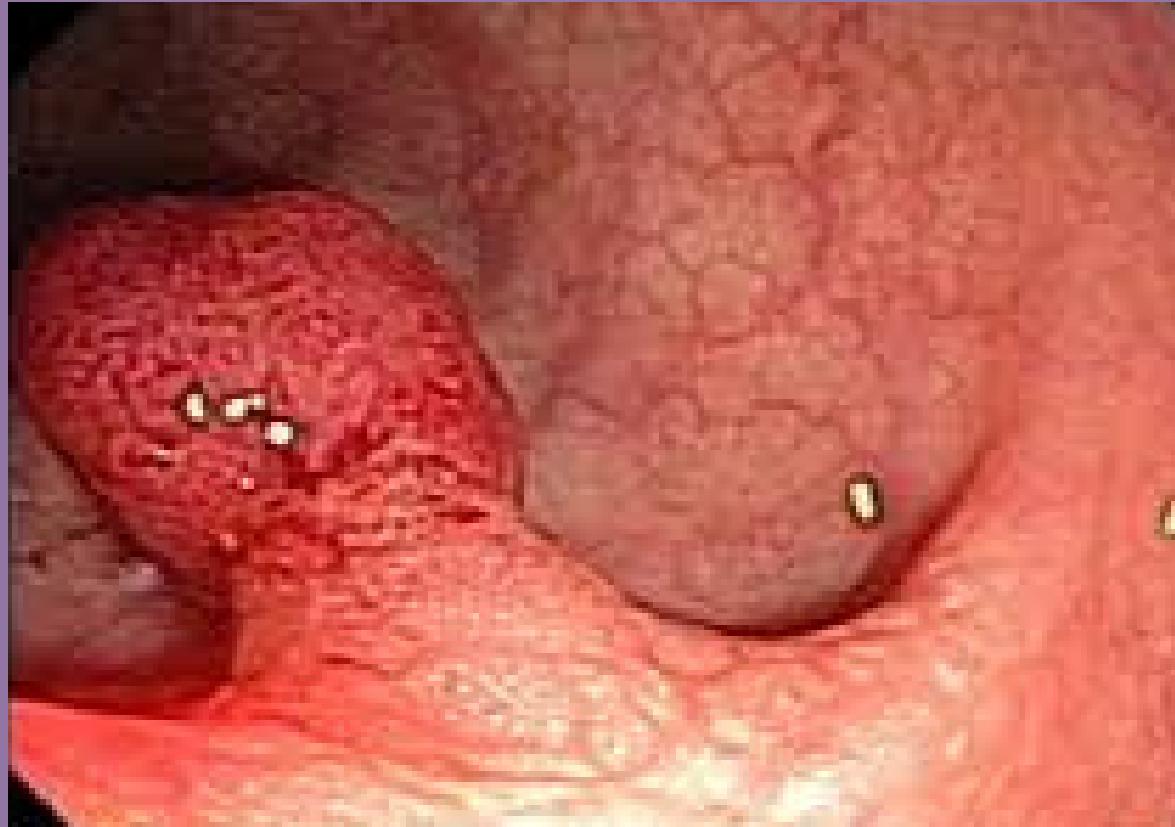
Alejandra Ortiz Medina



Polipos colorrectales



Un pólipo es un crecimiento formado por células que sobresale del tejido de la pared de un espacio hueco, como el intestino.



Adenoma serrado sésil SSA

Están formados por células ganglionares que recubren el interior del intestino grueso y tienen una alta probabilidad de resultar en cancer.



Pólipo hiperplásico HP

Este tipo de pólipos se caracteriza por un crecimiento excesivo de células en las glándulas intestinales. Tienen baja probabilidad de ser cancerígenas.



Conjunto de datos MHIST

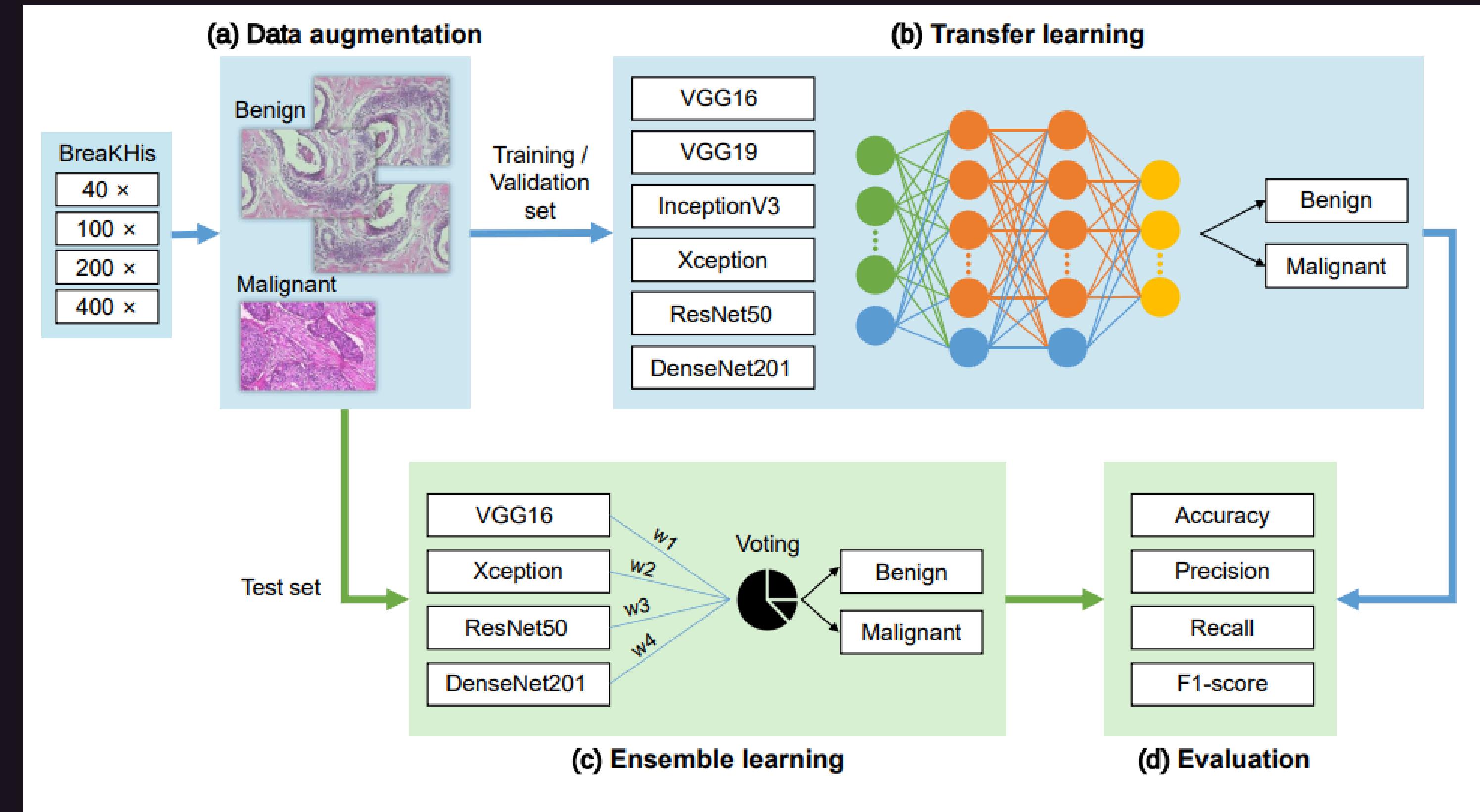


- Datos disponibles en:
<https://bmirds.github.io/MHIST/>
- Conjunto de datos compuesto por una carpeta con **3151** imágenes de muestras histopatológicas.
- Un archivo de anotaciones .csv que contiene las etiquetas **HP** y **SSA**, el nombre de la imagen, el tipo de partición (**train o test**) y el número de anotaciones por cada muestra

Objetivo

Clasificar imágenes de muestras histopatológicas para determinar si presentan pólipos de tipo hiperplásico o adenocarcinoma



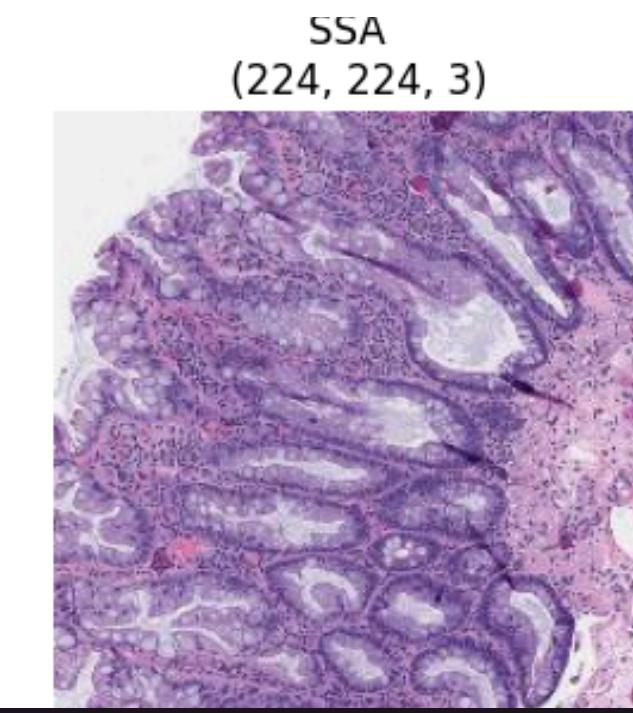
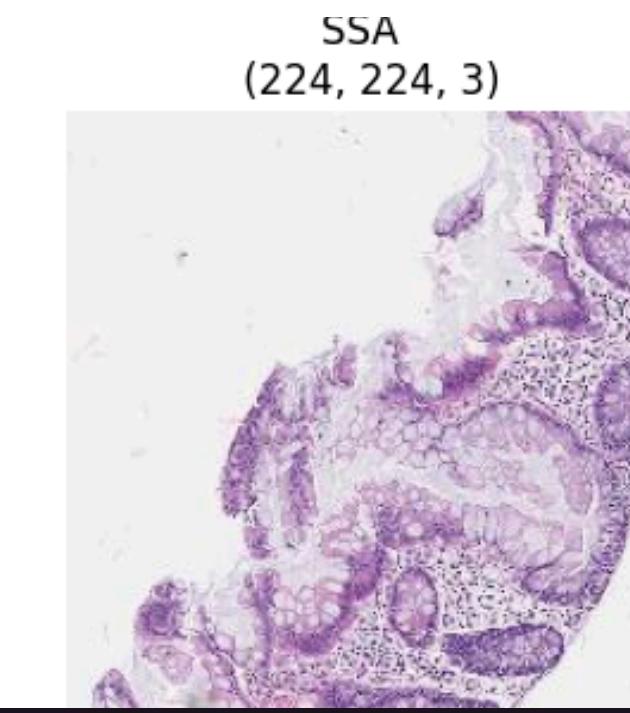
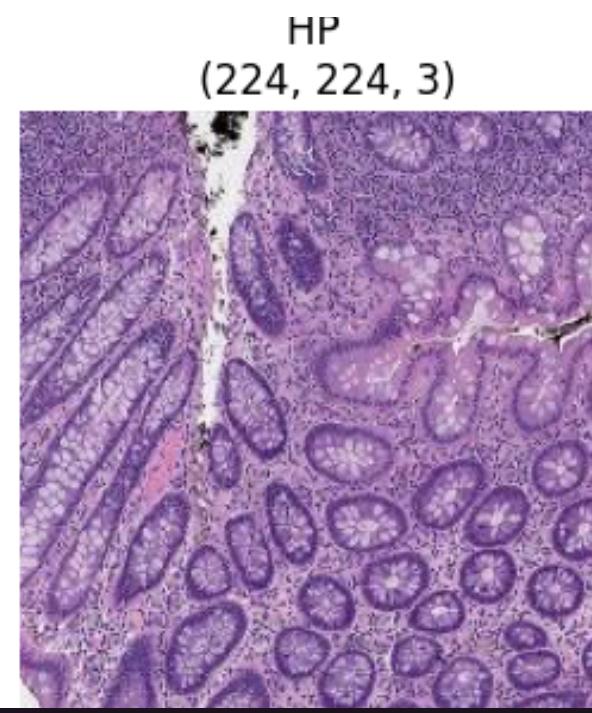
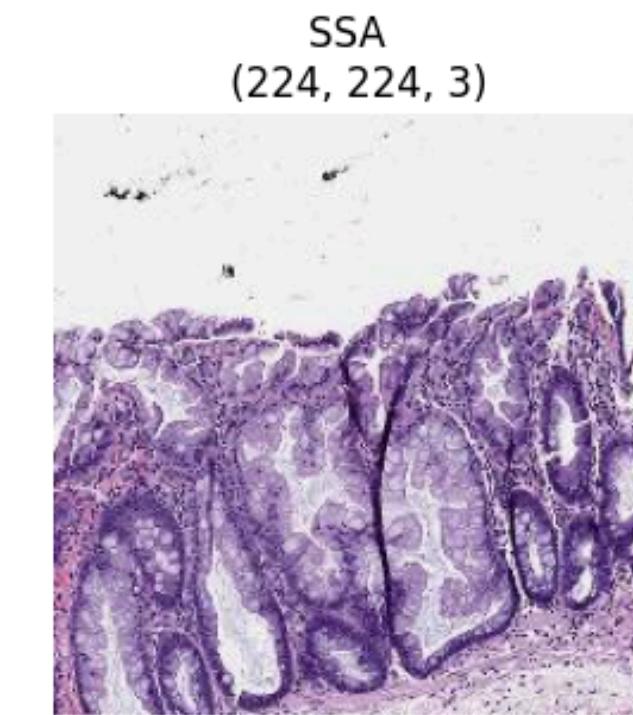
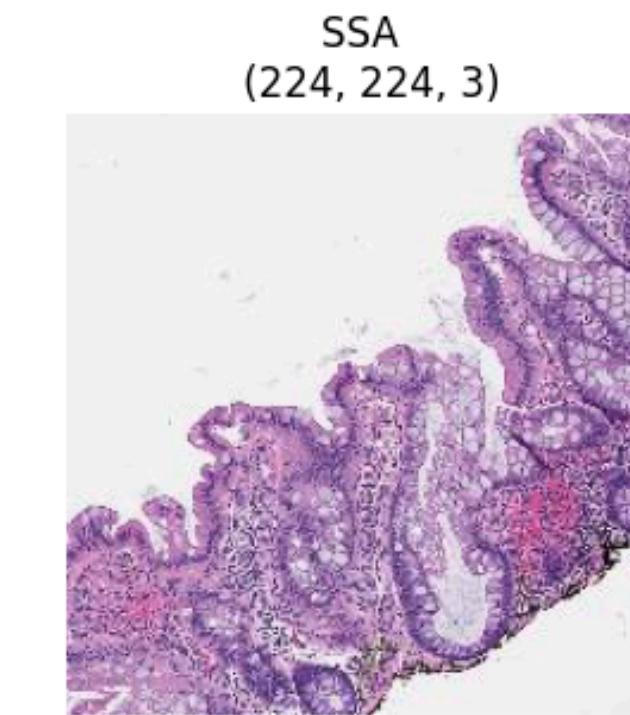
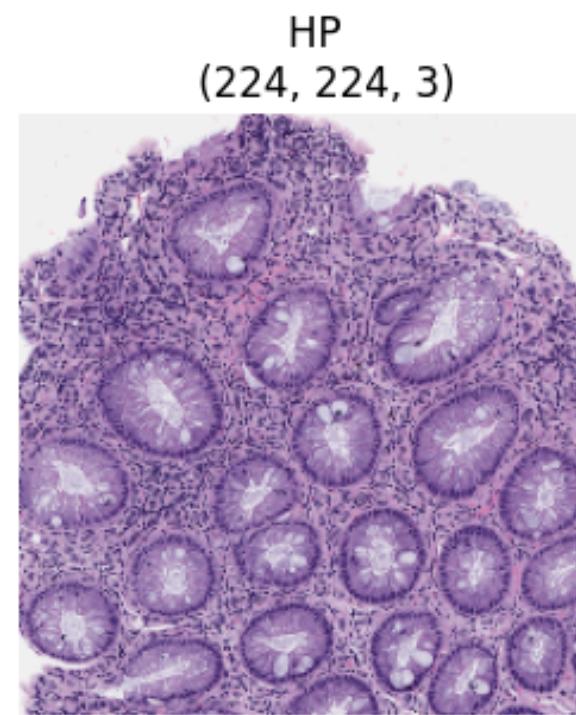
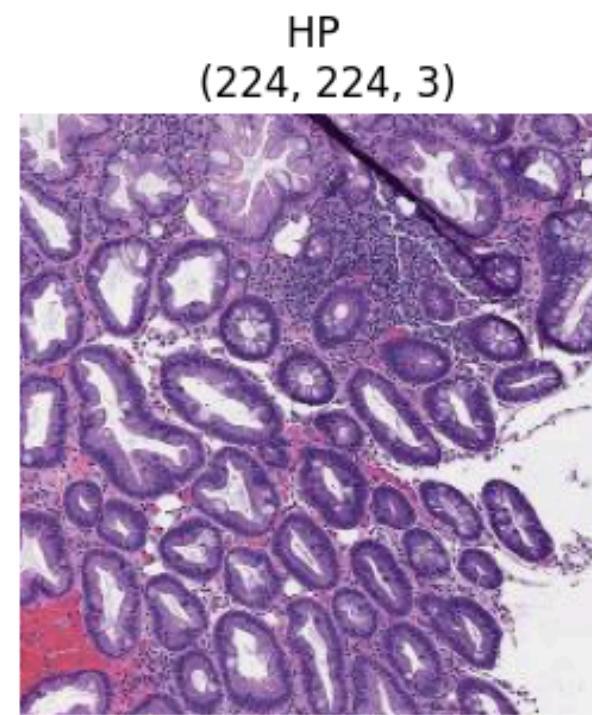




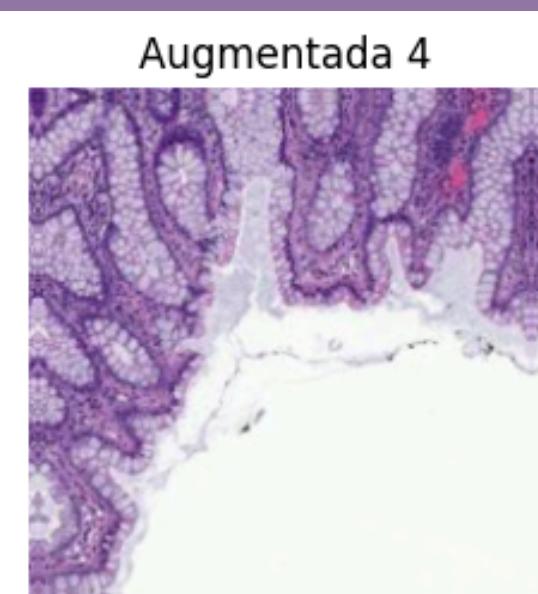
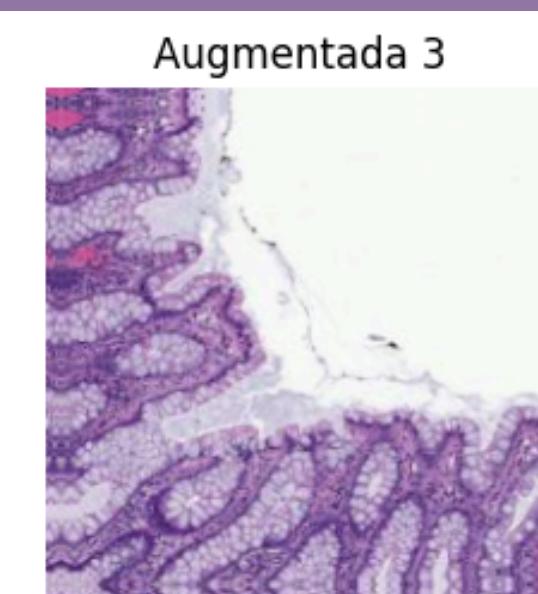
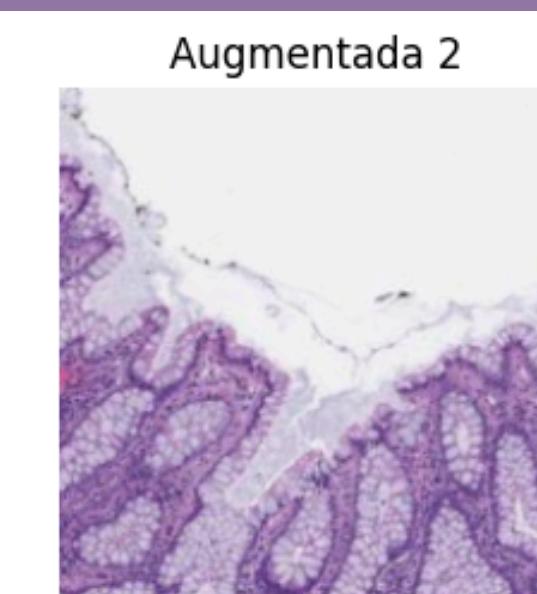
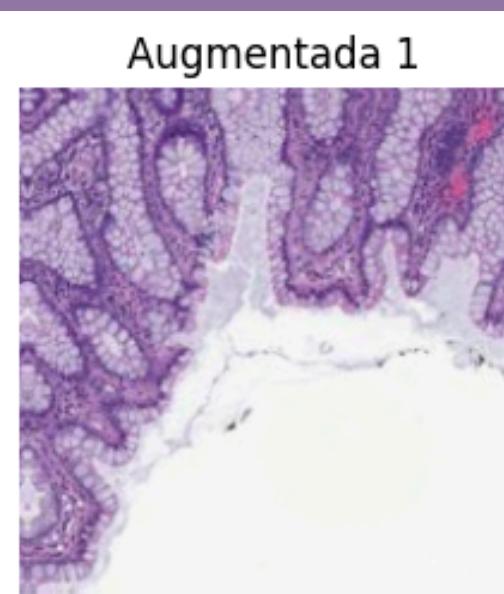
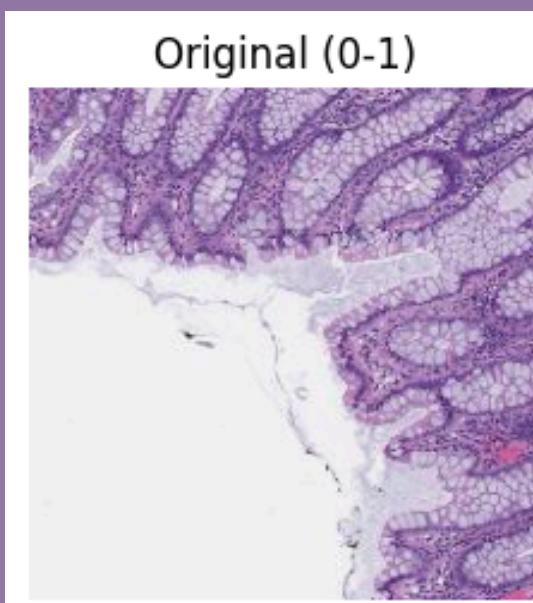
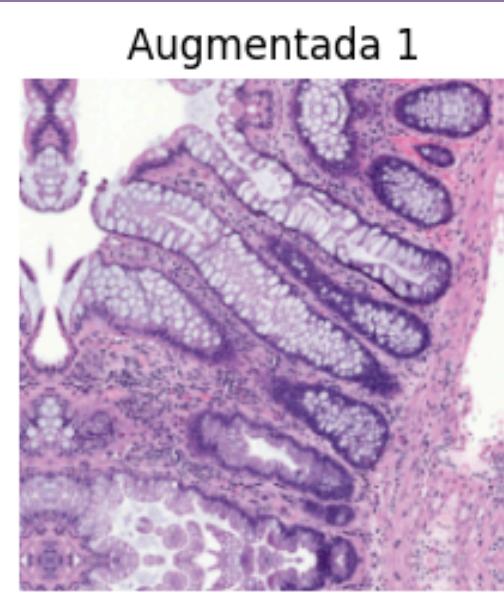
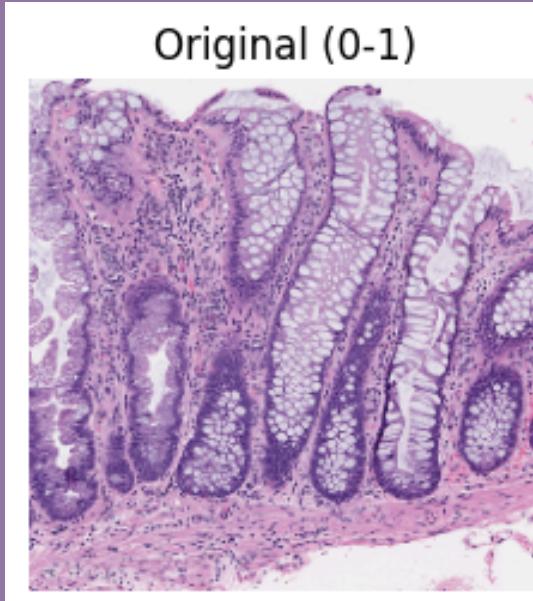
Entrenamiento



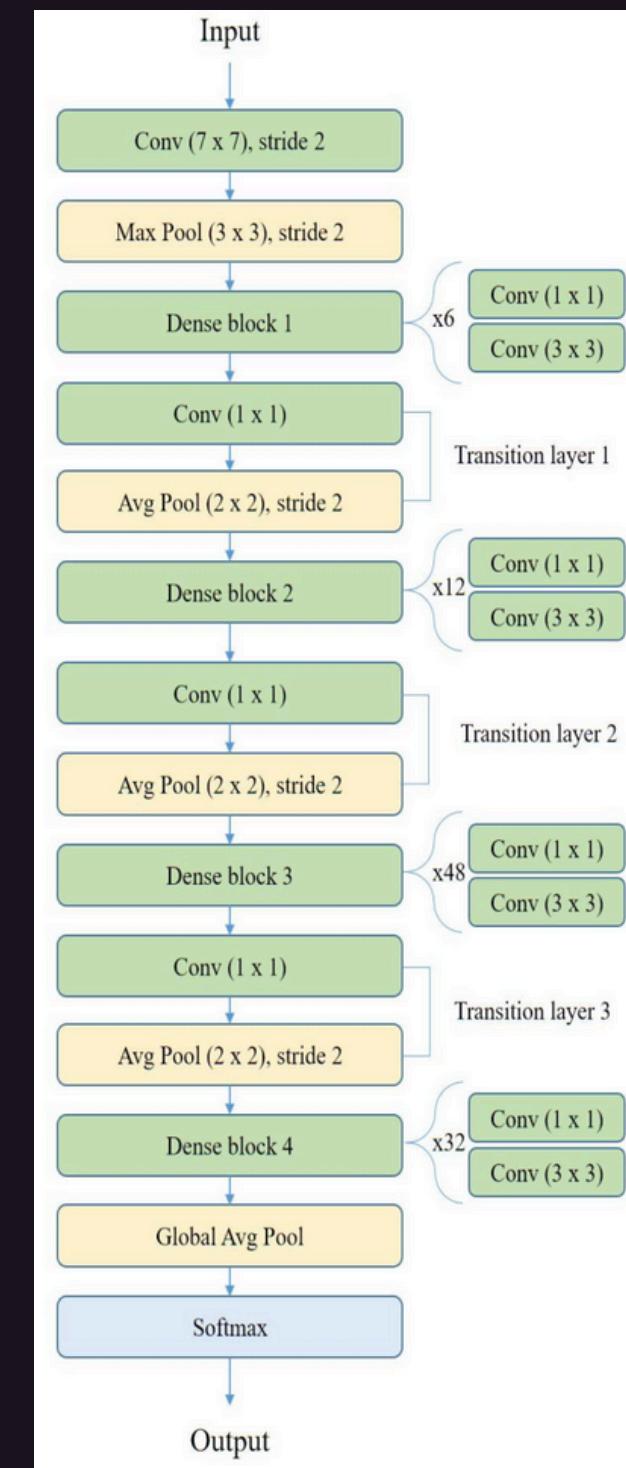
Algunos ejemplos



Data Augmentation



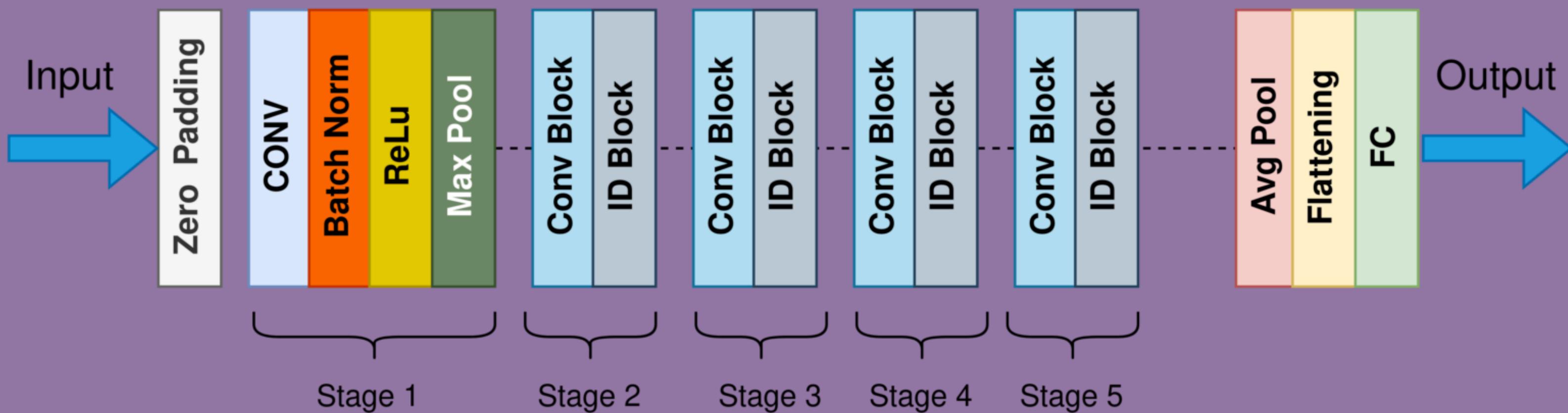
DenseNet201



Tomado de https://www.researchgate.net/figure/DenseNet201-architecture_fig2_363776890

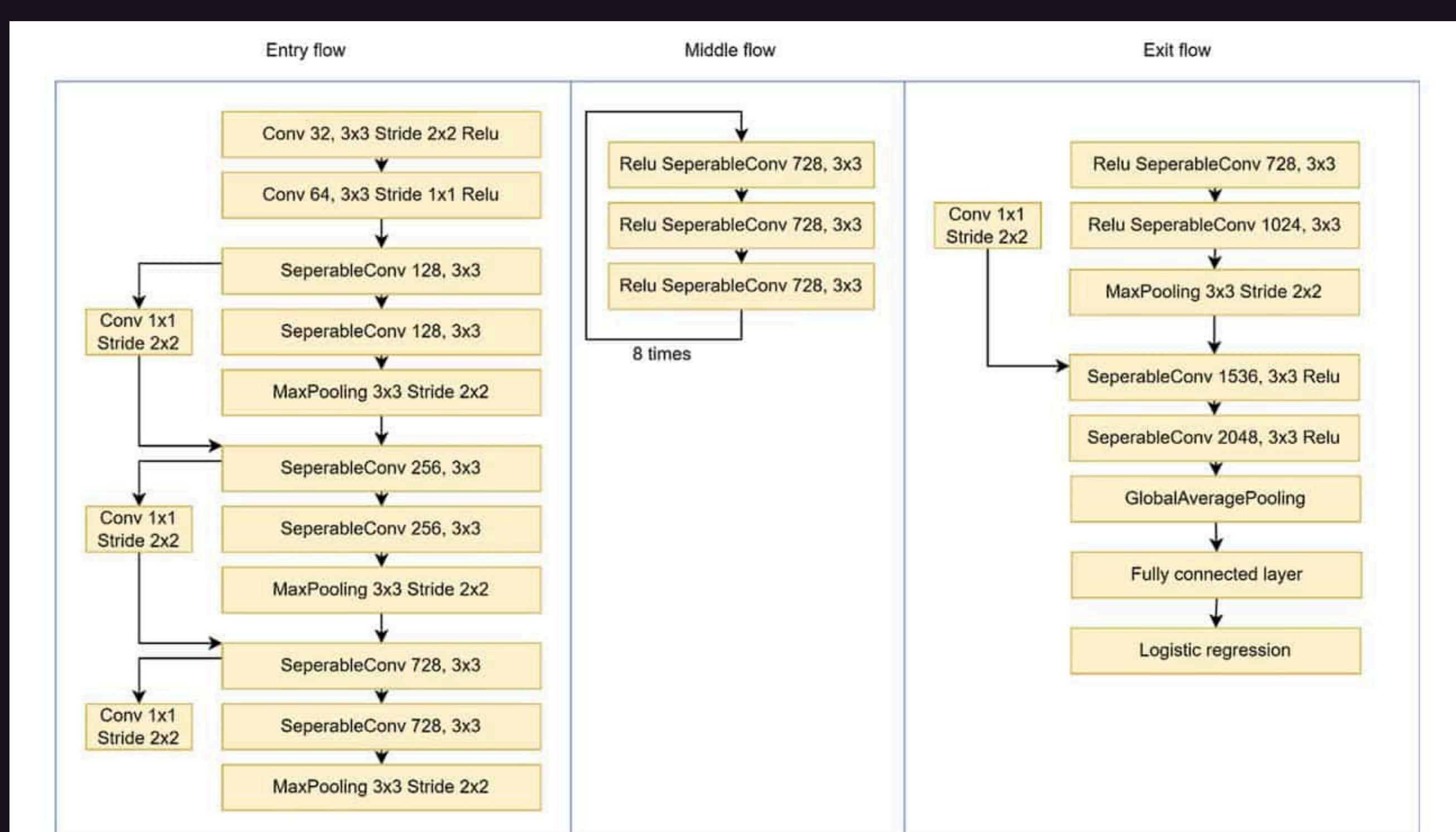
Resnet50

ResNet50 Model Architecture



Tomado de <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758/>

Xception



Tomado de <https://medium.com/@kdk199604/xception-deep-learnings-leap-beyond-inception-05a708c205f9>

Transfer Learning

- **Pesos** - ImageNet
- **Cabezal** - GlobalAveragePooling + Dropout(0.5) + Dense(1, Sigmoid)
- **Warm-up** - 6 épocas con base congelada (LR e-3).
- **Optimizador** - Adam.
- **Callbacks** - EarlyStopping (Patience=8) y ReduceLROnPlateau.

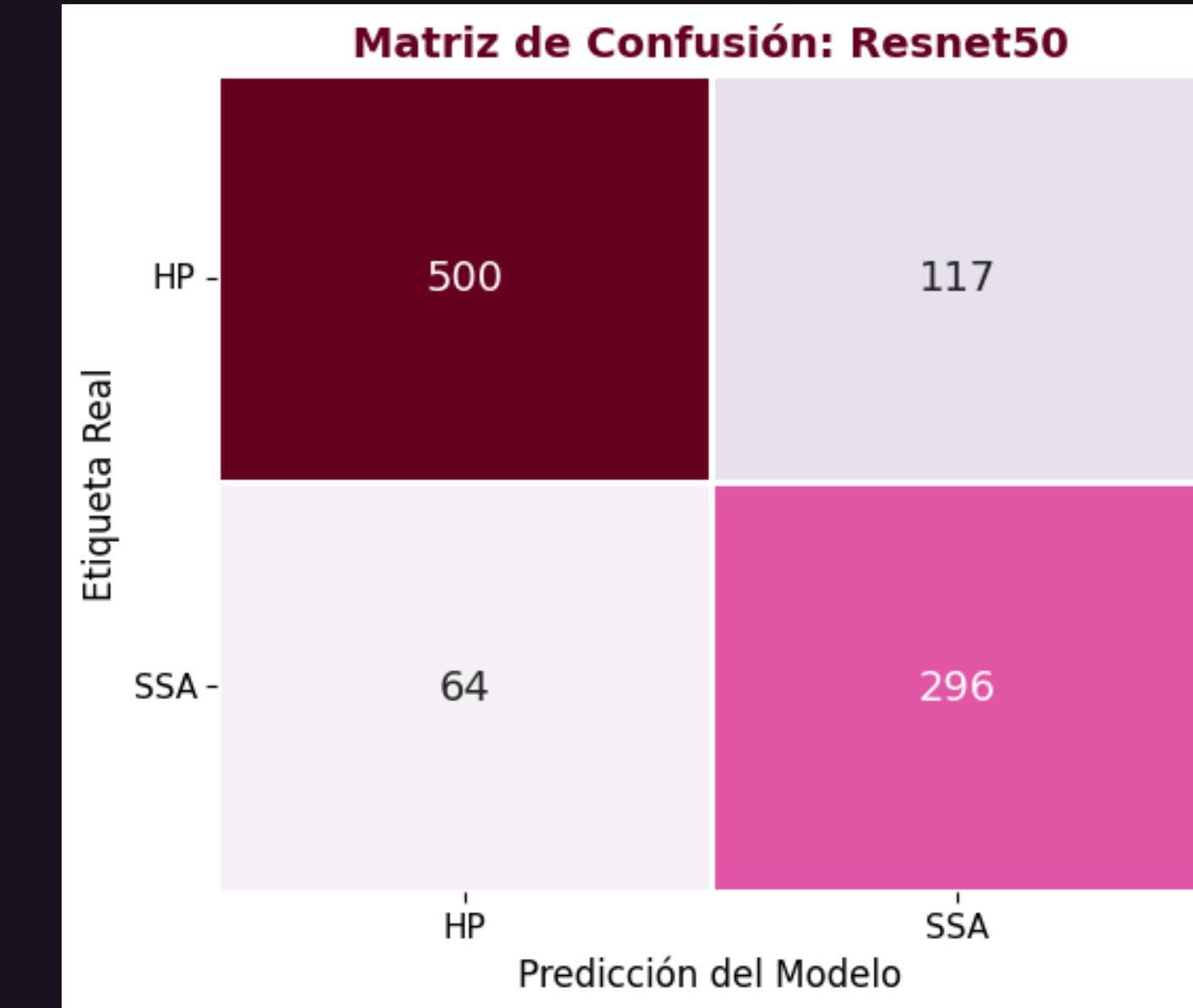
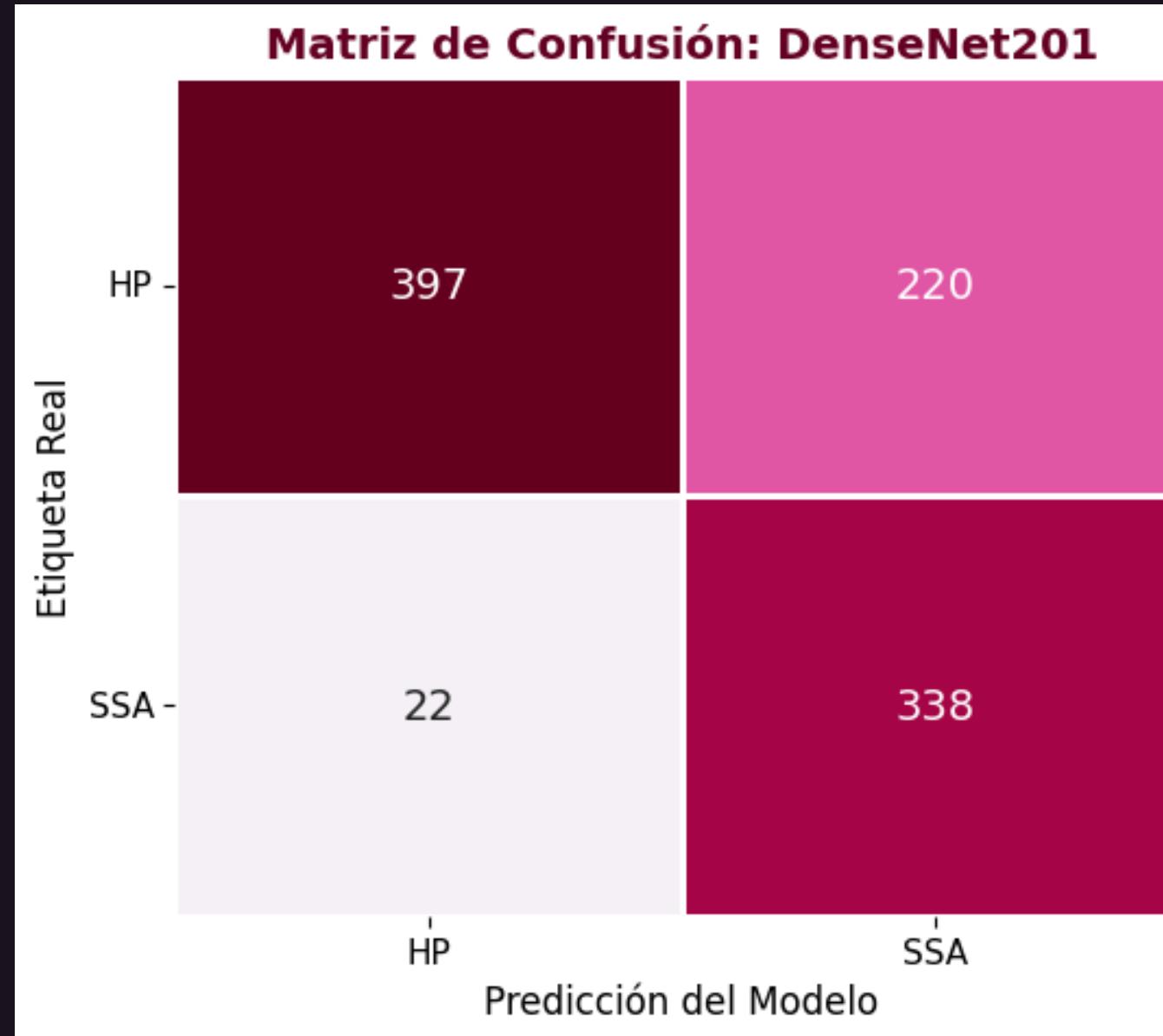
Resultados



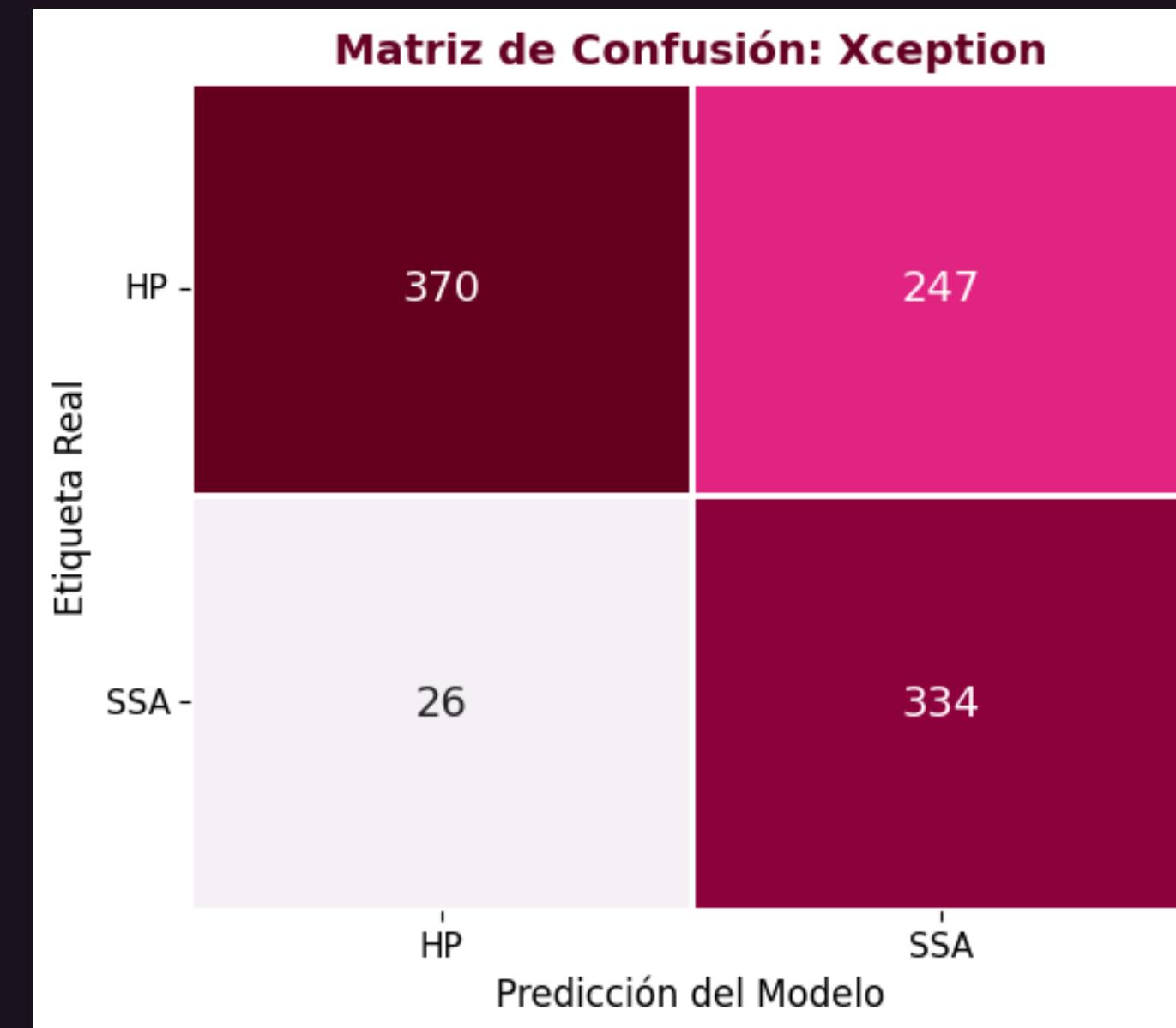
AUC



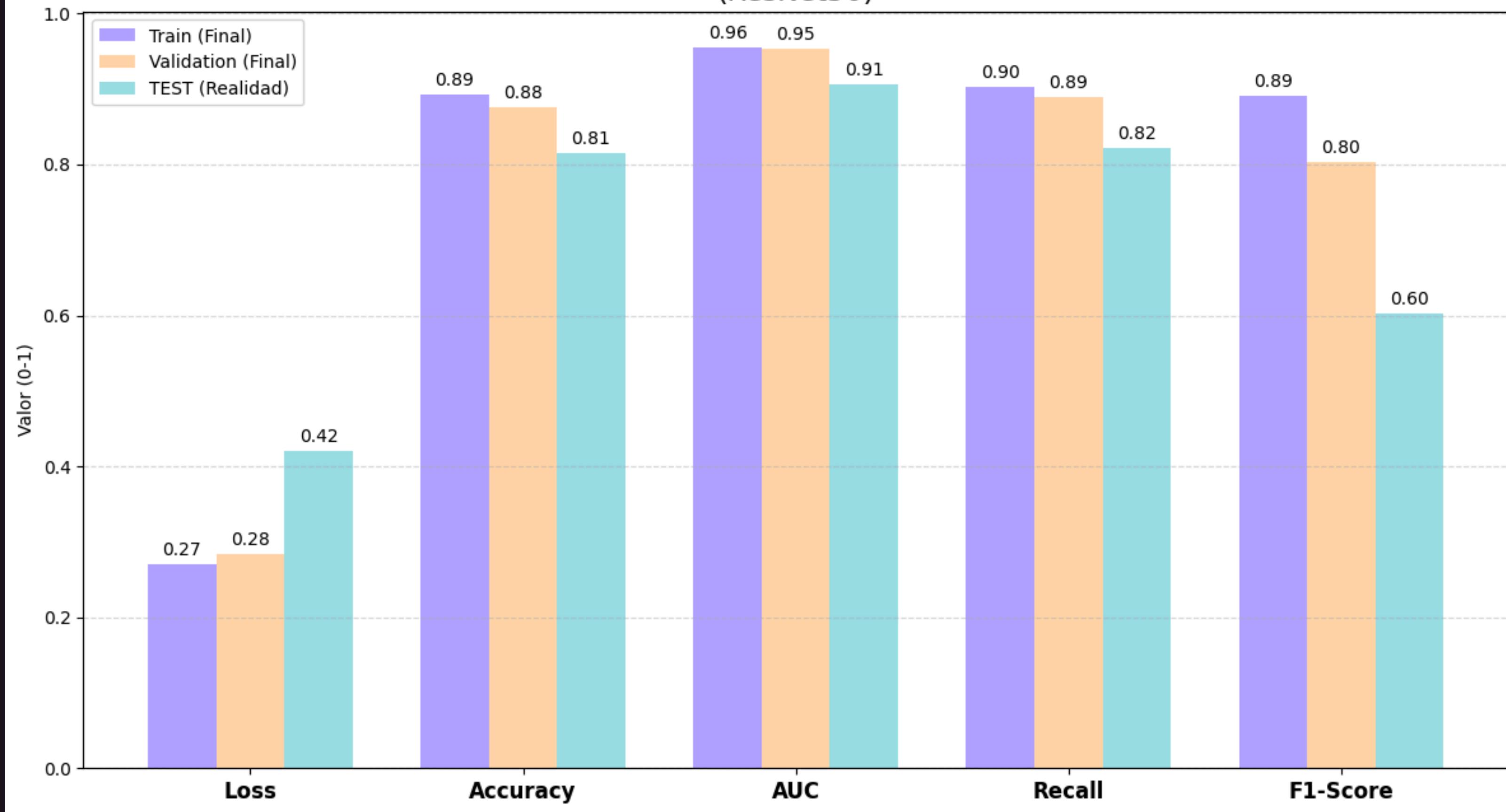
Matriz de confusión



Matriz de confusión



Entrenamiento vs Validación vs Test (ResNet50)

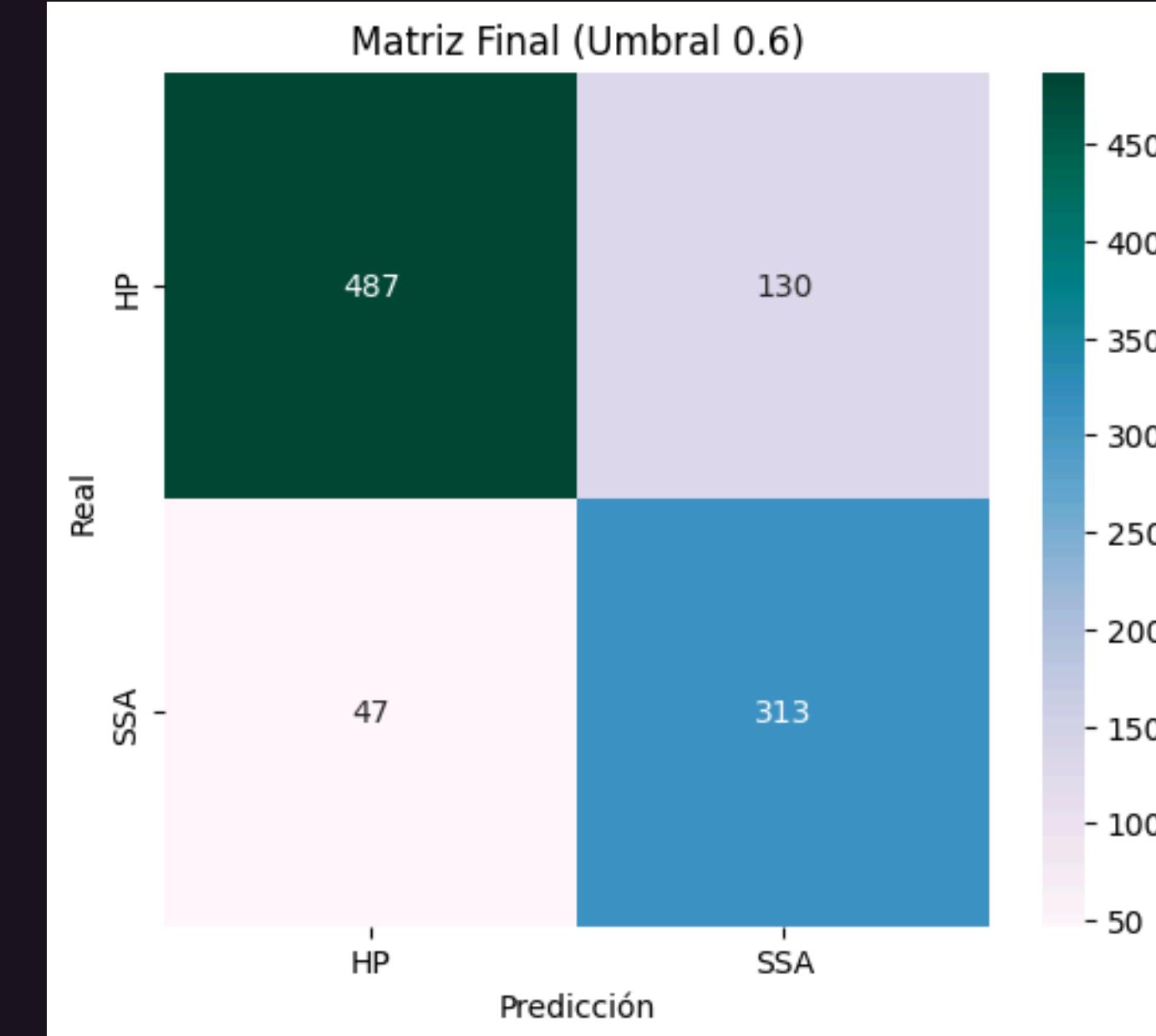
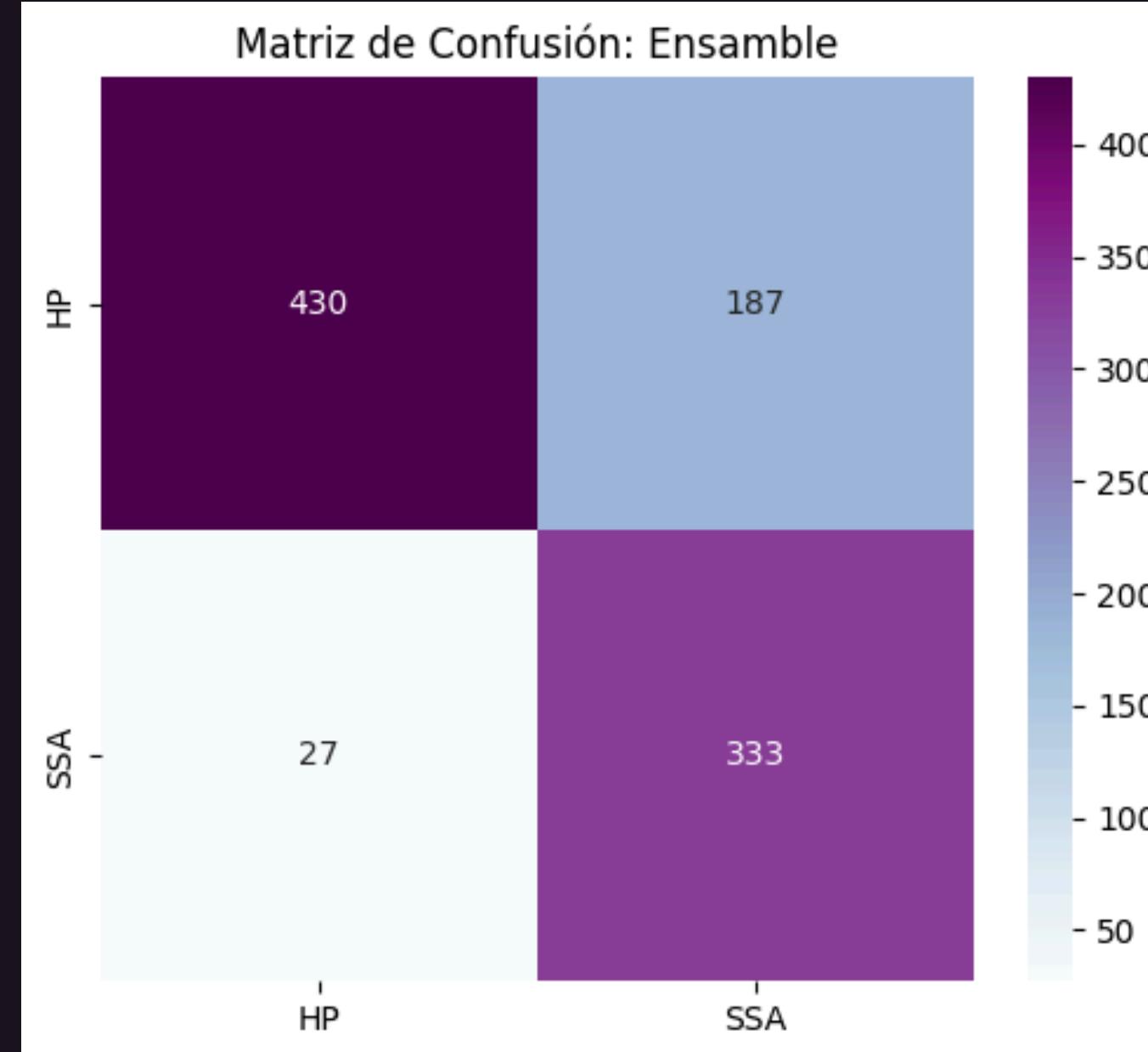


Ensemble learning - Soft voting

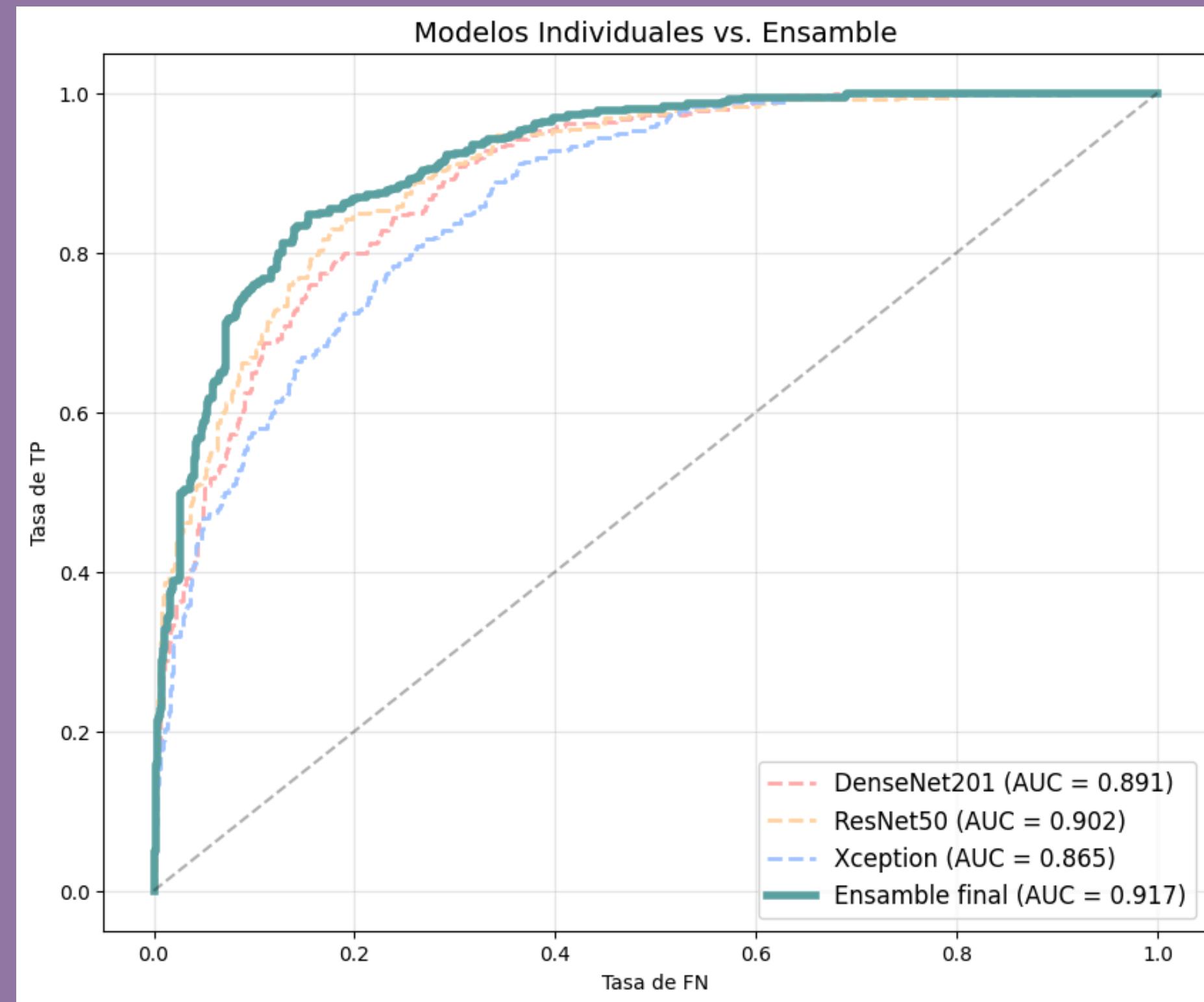
$$P_{final}(\mathbf{x}) = \frac{\sum_{i=1}^N w_i \cdot P_i(\mathbf{x})}{\sum_{i=1}^N w_i}$$

$$W_i = \frac{AUC_i}{\sum AUC_{total}}$$

Matriz de confusión ensamble



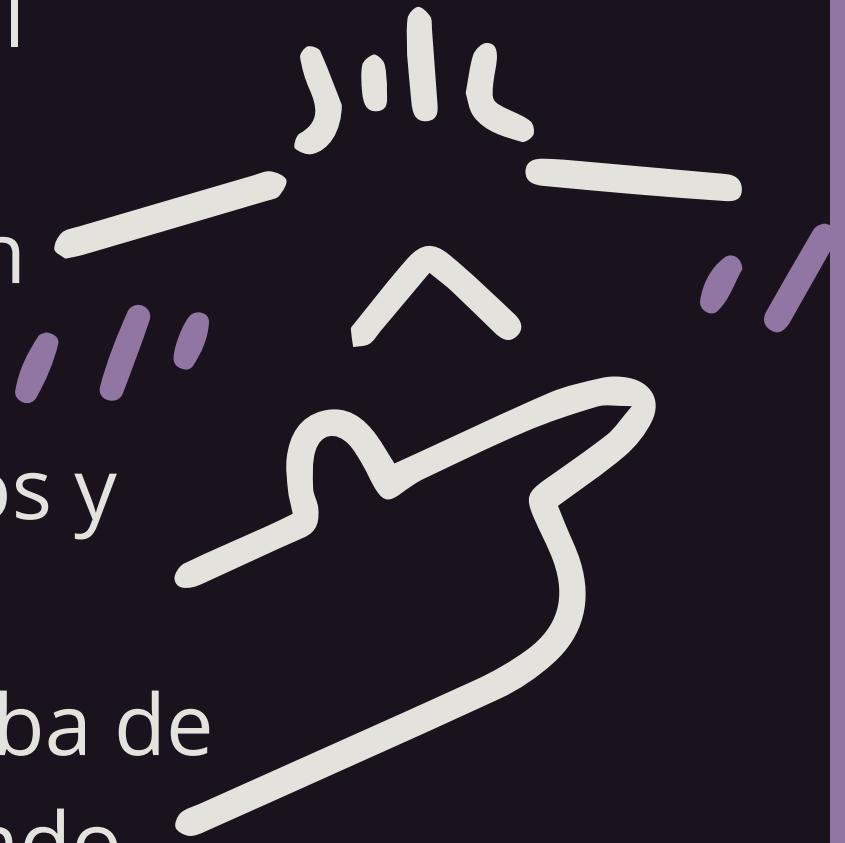
Resultados ponderados para AUC de ensamble



	precision	recall	f1-score	support
HP	0.91	0.79	0.85	617
SSA	0.71	0.87	0.78	360
accuracy			0.82	977
macro avg	0.81	0.83	0.81	977
weighted avg	0.84	0.82	0.82	977

Conclusiones

- La combinación de modelos mediante Soft Voting superó el rendimiento de cualquier red individual.
- El modelo ensanle demostró un equilibrio ideal, se logró un Recall del 87% con una Precisión aceptable del 71%
- Al especializar las últimas capas de modelos pre-entrenados y congelar **BatchNormalization**.
- Los modelos demostraron alta fiabilidad al superar la prueba de sanidad y al estabilizar sus curvas de aprendizaje, asegurando que el 82% de Accuracy final no se debió a overfitting



Enlaces

- https://colab.research.google.com/drive/11k0DLI8WfnEN_3GgLipBkse5sGnvYxs8?usp=sharing
- [ClearML](#)

