

# Progresión Semana 1: 17/06 - 21/06

---

- Se ha descargado el conjunto de datos CADICA.
- Se ha realizado un análisis exploratorio del conjunto de datos, este análisis de datos exploratorio ha incluido inspeccionar:
  1. Cantidad de Videos por paciente
  2. Cantidad de Frames con Lesión por paciente
  3. Cantidad de Frames por tipo de Lesión por paciente (p0\_20, p20\_50, ... p100)
- Se ha programado una función para realizar un HoldOut al conjunto de datos. Se ha realizado la división por video, no por paciente, esto significa que un paciente puede tener sus videos en diferentes splits, pero los frames de un mismo video siempre terminarán en un mismo split.
- El HoldOut se ha realizado un esquema de train/val/test (60%, 20%, 20%).
- Se ha realizado una conversión del conjunto de datos al formato esperado por la detección de YOLOv8, esto incluye:
  1. Creación de dos carpetas, "images" y "labels".
  2. Cada carpeta tendrá los 3 splits (train, test, val).
  3. La carpeta labels tendrá un archivo .txt por cada frame con lesión, dejando los frames sin lesión sin ningún archivo .txt
  4. Se ha realizado una conversión del formato "[x,y,w,h], where (x,y) are the pixel coordinates of the top left corner, w is the width and h is the height of the bounding box." al formato esperado por YOLO, el cual consiste en encontrar el punto central de la bounding box, y especificarlo junto con la distancia a cada lado normalizada.
- Se ha creado una primera versión de la clase encargada del entrenamiento de la red YOLOv8 para la detección de aterosclerosis en las imágenes.
  1. Se han seleccionado solo las imágenes con lesiones de grados: p70\_90, p90\_98, p99, y p100
  2. El fichero `.yaml` de configuración de la red se ha configurado para que utilice todos los labels disponibles en el conjunto de datos, sin embargo, parece que no es buena idea hacerlo sin tener el resto de clases. Esto se debe a que el modelo parece administrar un orden de prioridad a las etiquetas al tener una codificación numérica nominal, haciendo que si duda, lo ponga en la clasificación más baja (0: p0\_20). Por lo tanto, para este primer entrenamiento, **solo se incluirán las etiquetas p70\_90 - p100** en el fichero yaml.

```
path: .Baseline_CADICA_Detection
train: images/train
val: images/val
test: images/test
```

```
names:
  0: p0_20
  1: p20_50
  2: p50_70
  3: p70_90
  4: p90_98
  5: p99
  6: p100
```





