**MEDICAL DIAGNOSIS: HELICOBACTER**

For the detection is the examination of histological samples with immunohistochemical staining.

Visual examination is highly time consuming due to the size of images (120000x16000 pixels) and the different densities of the bacteria.

**Objectives:**

* Detection of Helicobacter in immunohistochemical histopathological images
* Identification of main problems in small sample size unbalanced data sets
* Detection of Anomalous staining using self-supervised learning
* Attention Mechanisms
* Assessment of Reproducibility

**Challenges for ML-DL methods:**

* Small sample size of WSI (245 cases) of huge size
* Highly unbalance in the set of patches
* Different densities of Hpylori

Consideration: It is not normally on the centers but in the borders the helicobacter

**SYSTEM PIPELINE**

**1.DATA PREPARATION:** extract patches among sample border



With the train do data augmentation and random. Then you have **dataset1**: Patients with annotated patches for cross-validation of models. And then **Dataset2**: independent set of non-annotated patches for assessment of reproducibility (test diagnosis).

2. **PATCH CLASSIFICATION**: identify Hpylori presence in patches

- **classification approach**: Data augmentation, shawoll architecture (conbolutional unit). Flatten.

- **anomaly detector**: self-supervised learning of visual representations. Reconstruction error detects anomalous staining.

**3. PATIENT DIAGNOSIS**: Aggregate patch outputs for diagnosis

**SPECIFIC OBJECTIVES:**

* **IMPLEMENT TWO SYSTEMS:**
  + **Anomaly detector with adaptive thresholding**
  + **Classification approach with attention mechanism**
* **COMPARE PERFOMANCE:**
  + **Cross-Validation: patch and patient level**
  + **Reproducibility: patient level**

- **Adaptive thresholding**: percentage of positive patches with adaptative thresholding

**- Attention Mechanism**: sample representation space using attention for patient classification

CLASSIFICATION APPROACH:

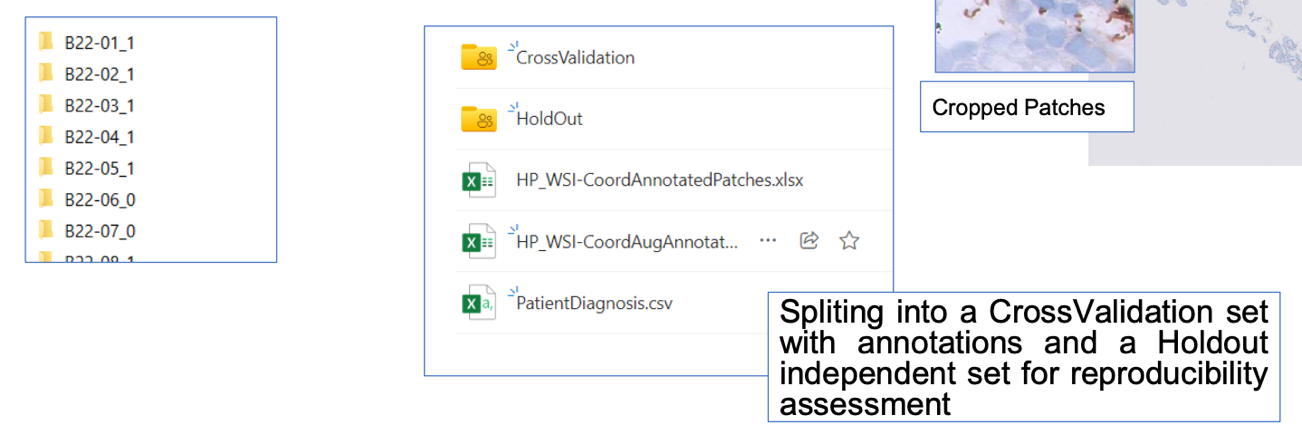
**DATASET: QUIRON SALUT**

Tiff images with CODI I DENSITAT. 3 sections en densitat according Hpilory density: NEGATIVA, BAIXA, ALTA DENSITAT. Això en “PatientDIagnosis.csv”.

El codi és B22-01, B22-02, …B22-n.

Patches of size 256x256 were cropped along the border of the tissue samples of one of the sections.

Cropped images (.png) are inside folders identified by the PatID and the tissue section as PatID\_Section#



**CrossValidationSet:**

Subset of 123 cases (77 positive ones), an expert annotated 1211 patches (161 being positive ones).

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Coordinates (I,j,w,h) of the annotated patches. Presence: -1 (no helico), 1(helico), 0(not clear). Are in HP\_WSI-CoordAnnotatePatches.xlsx.

This file only contains annotated patches. The remaining ones of the set are negative patches.

Tabla

Descripción generada automáticamente

**HoldOut Set:**

The remaining cases (116) are an independent HoldOut set for verification of the reproducibility of Patient Diagnosis methods.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente