ML in Applications

Dipartimento di Automatica e Informatica
Politecnico di Torino, Torino, ITALY

Politecnico
di Torino

Lab 3

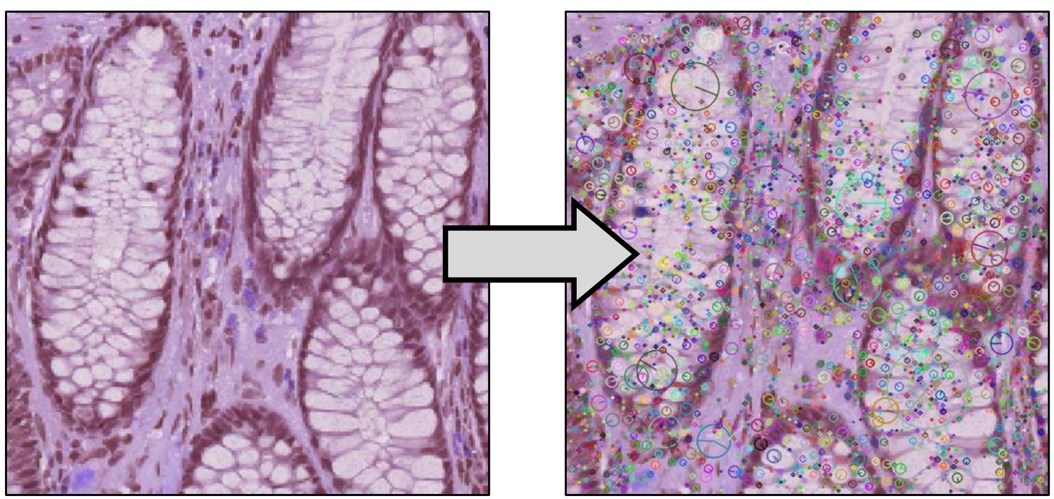
Image classification via bag of (visual) features

Bag of Visual Words

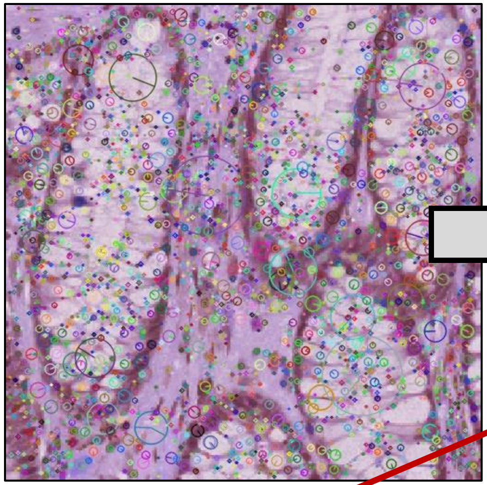
- Bag of Visual Words (or features) is an extension to the NLP algorithm Bag of Words
- Can be considered the state-of-the-art algorithm NOT deep learning based
- Essentially, it follows 3 steps
  - **Features** extraction
  - Construction of visual vocabulary by **clustering**, followed by frequency analysis
  - **Classification** of images, based on the generated vocabulary

Features extraction

- We will leverage the SIFT (Scale Invariant Feature Transform) algorithm developed by David Lowe, now available on OPENCV (patent expired)
  - First detects *key-points* in an images (the number changes among images)
  - Second calculates the decriptors (always 128 for SIFT)



Features extraction



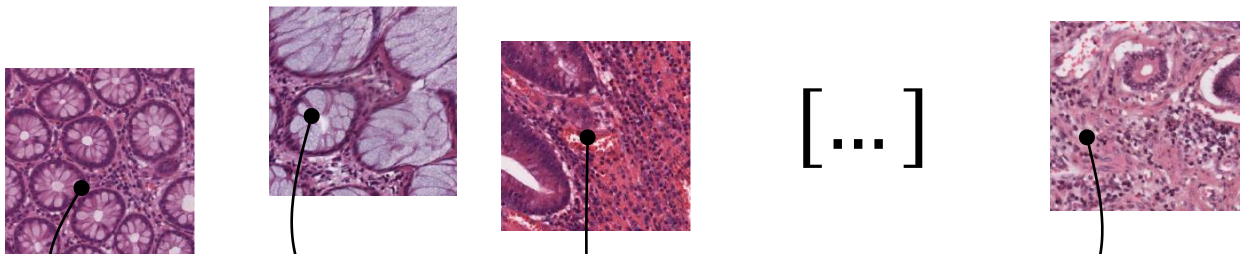
```
[36]  1 helper = ImageHelpers()
      2 kp, des = helper.features(image)

[37]  1 print(len(kp))
      2 print(len(des))
      3 print(des.shape)

3629
3629
(3629, 128)
```

The number of key-points (and hence descriptors) varies among images

- Given a single image of our training set, we get several descriptors, each one of shape (*128,)*

From features extraction to clustering

- Over the whole dataset of ***N*** images, we get a list of ***N*** descriptors blocks, one for each image:



$$\boldsymbol{L} = [\square_1, \square_2, \square_3, \square_4, \dots, \square_N]$$

Descriptors associated to image 1, with shape (*k1*, *128*)

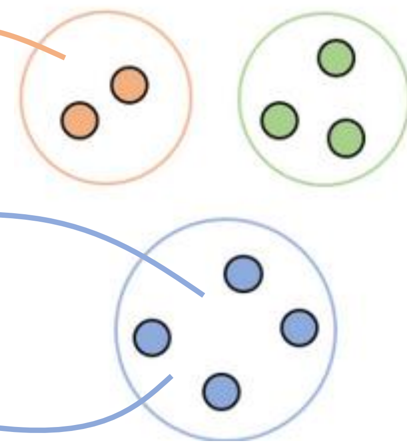
- We need to convert ***L*** to an array ***D*** of shape *(k1+k2+k3+k4+…+kN, 128)* to be fed to clustering through *sklearn*

Construction of visual vocabulary by clustering

$$\boldsymbol{D} \rightarrow ndarry\ with\ shape = (k1 + k2 + \cdots + kN, 128) = (K, 128)$$

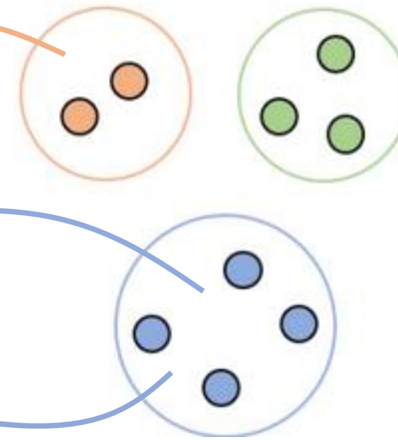
$$\boldsymbol{D} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,128} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,128} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{K,1} & x_{K,2} & x_{K,3} & \cdots & x_{K,128} \end{bmatrix} \xrightarrow{\text{clustering}} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}$$

$$\boldsymbol{D} = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,128} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,128} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{K,1} & x_{K,2} & x_{K,3} & \cdots & x_{K,128} \end{bmatrix}$$



Construction of visual vocabulary by clustering

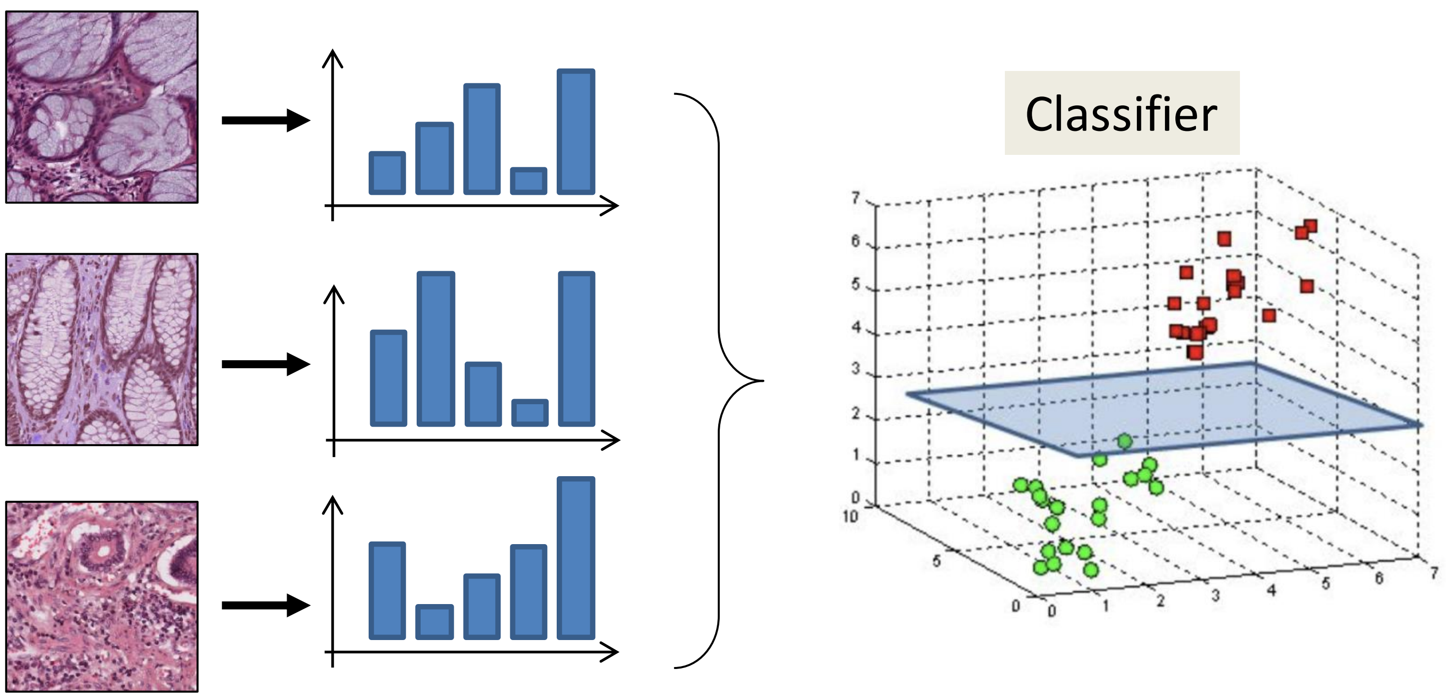
$$D = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots & x_{1,128} \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots & x_{2,128} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{K,1} & x_{K,2} & x_{K,3} & \cdots & x_{K,128} \end{bmatrix}$$



- Basically, **each descriptor** of the training set is **associated** with a **cluster**
- Each **cluster** is a **visual word**
- Descriptors belonging to the same cluster are **«synonyms»**

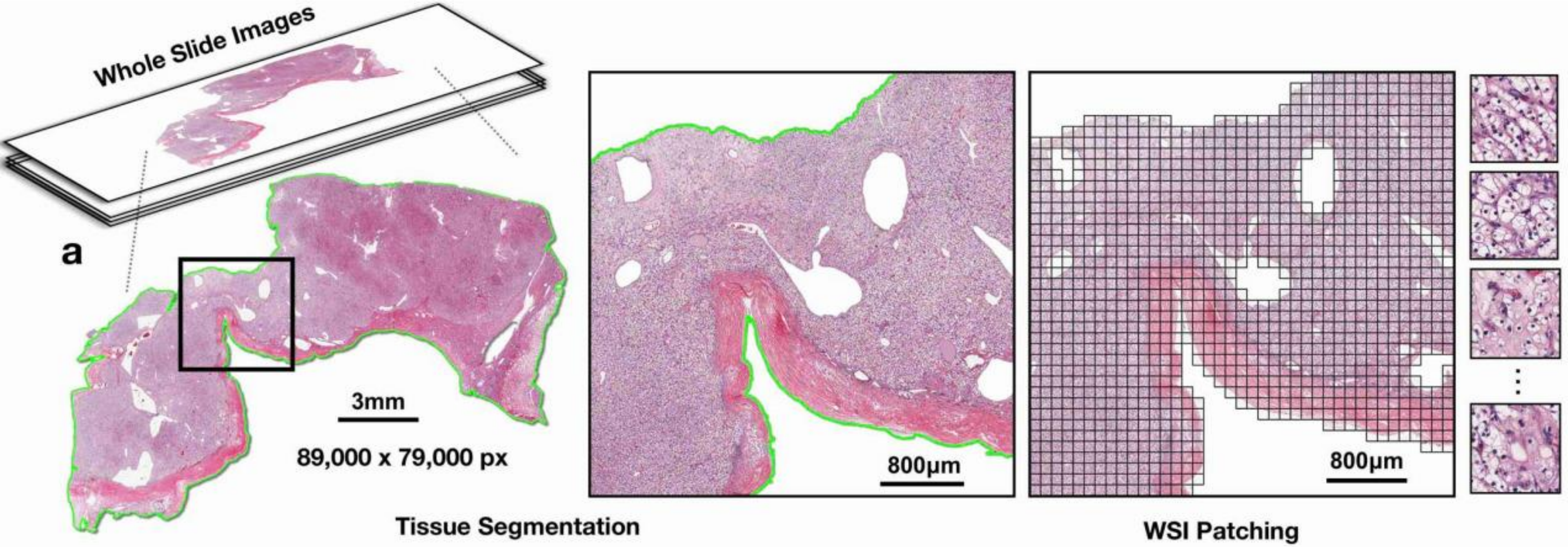
Classification

- Finally, the **image** is represented by a **histogram of the visual words** by counting the number of features that belong to each visual word (i.e., cluster)



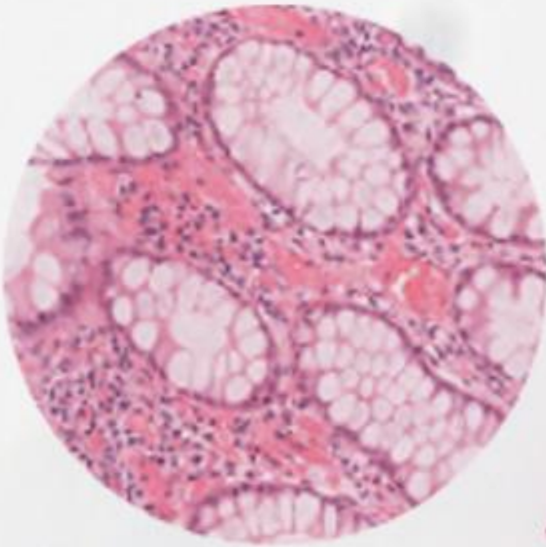
Case study

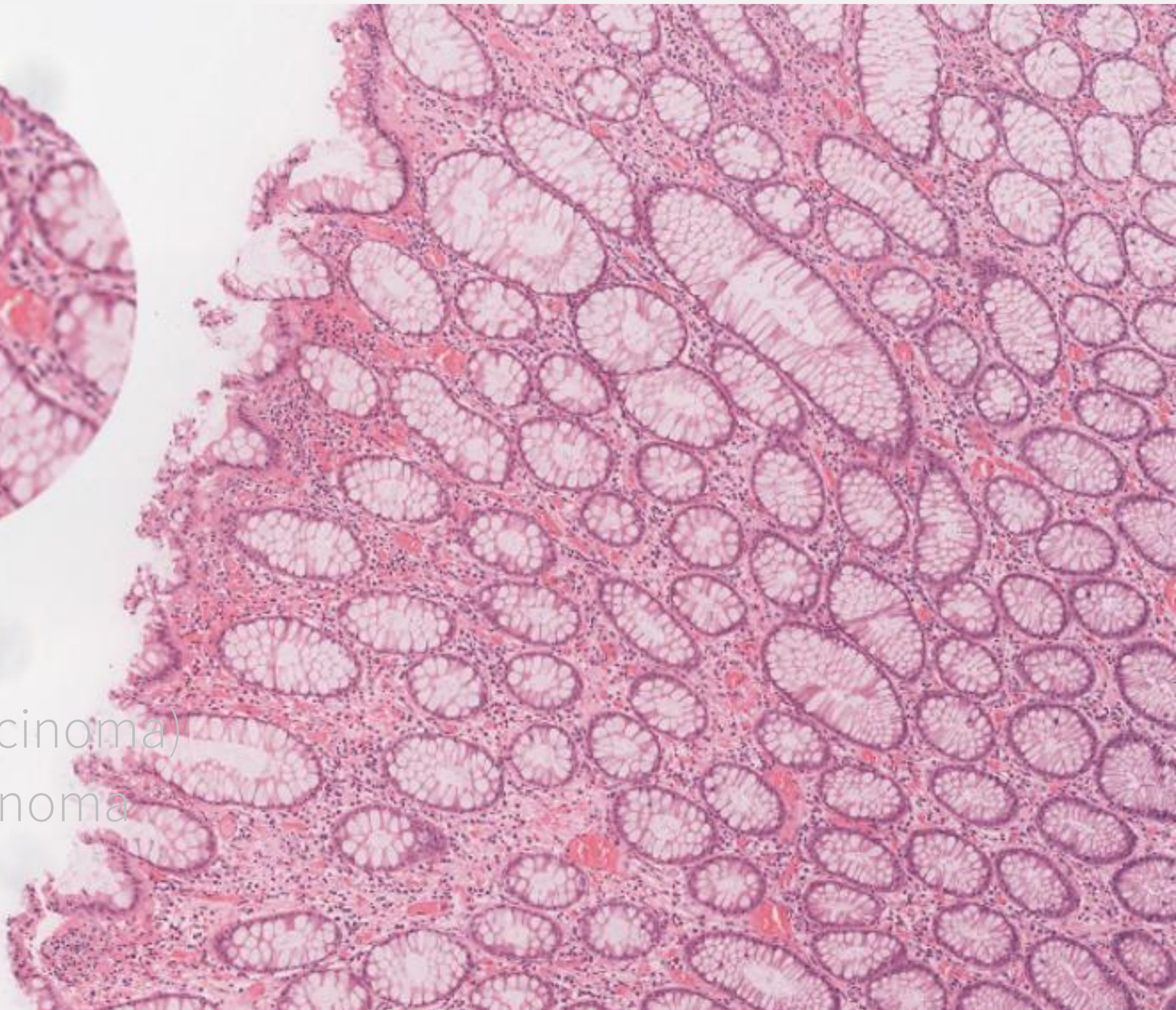
- Colorectal classification via Whole Slide Images (WSIs)



Source: Lu, M. Y., Williamson, D. F., Chen, T. Y., Chen, R. J., Barbieri, M., & Mahmood, F. (2021). Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, *5*(6), 555-570.

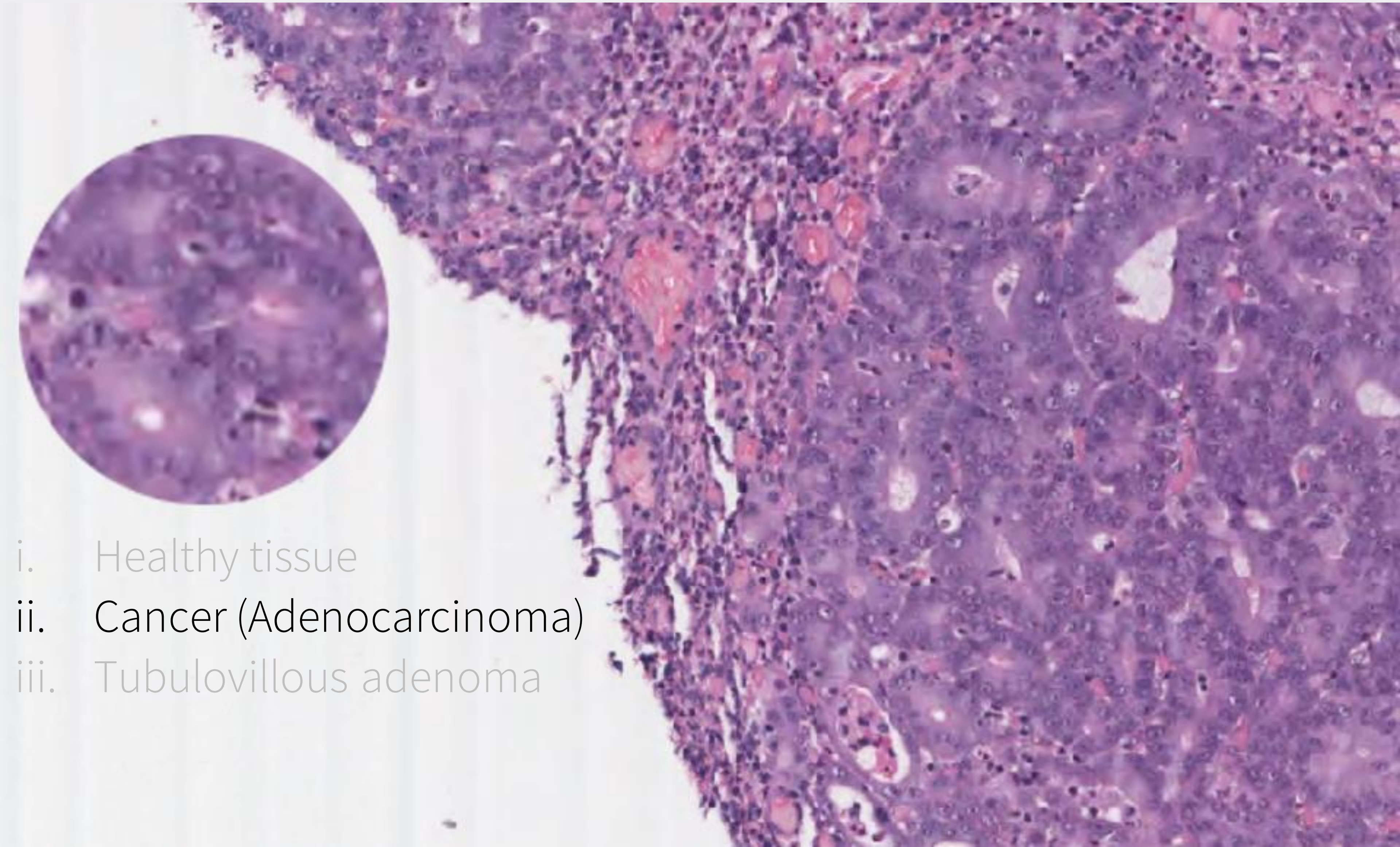
Dataset





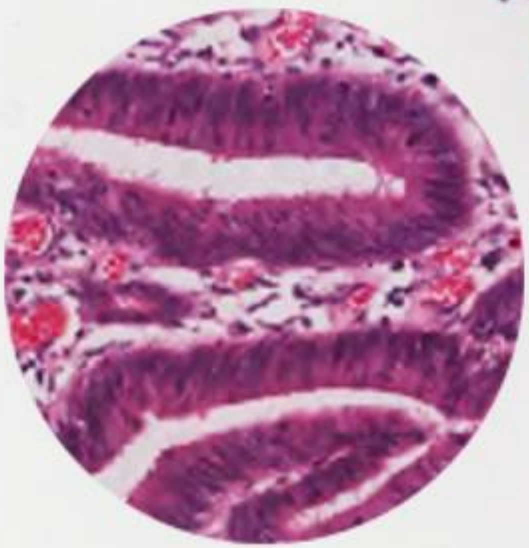
i. Healthy tissue
ii. Cancer (Adenocarcinoma)
iii. Tubulovillous adenoma

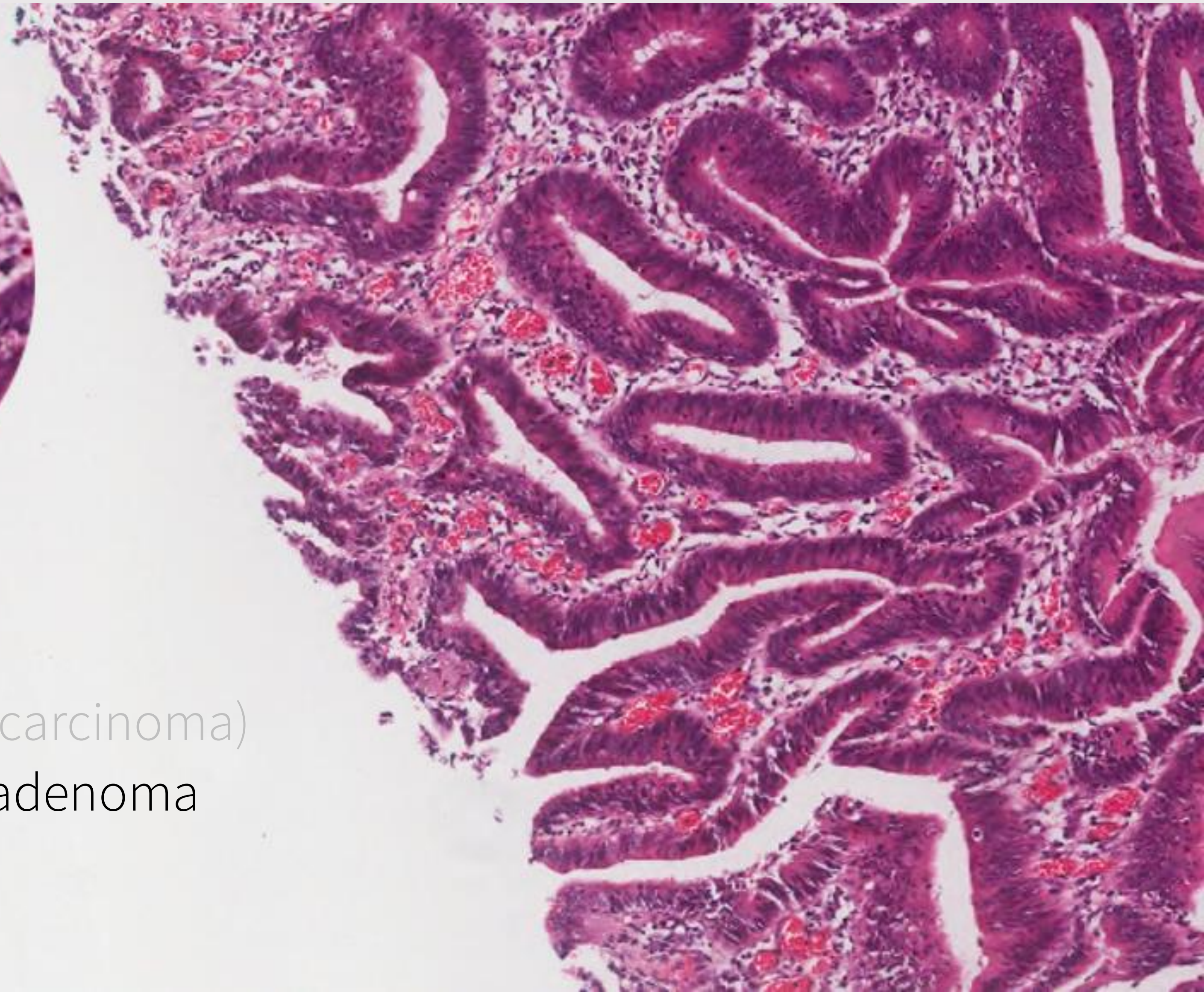
Dataset



i. Healthy tissue
ii. Cancer (Adenocarcinoma)
iii. Tubulovillous adenoma

Dataset





i. Healthy tissue
ii. Cancer (Adenocarcinoma)
iii. Tubulovillous adenoma