Histopathologie numérique et intelligence artificielle

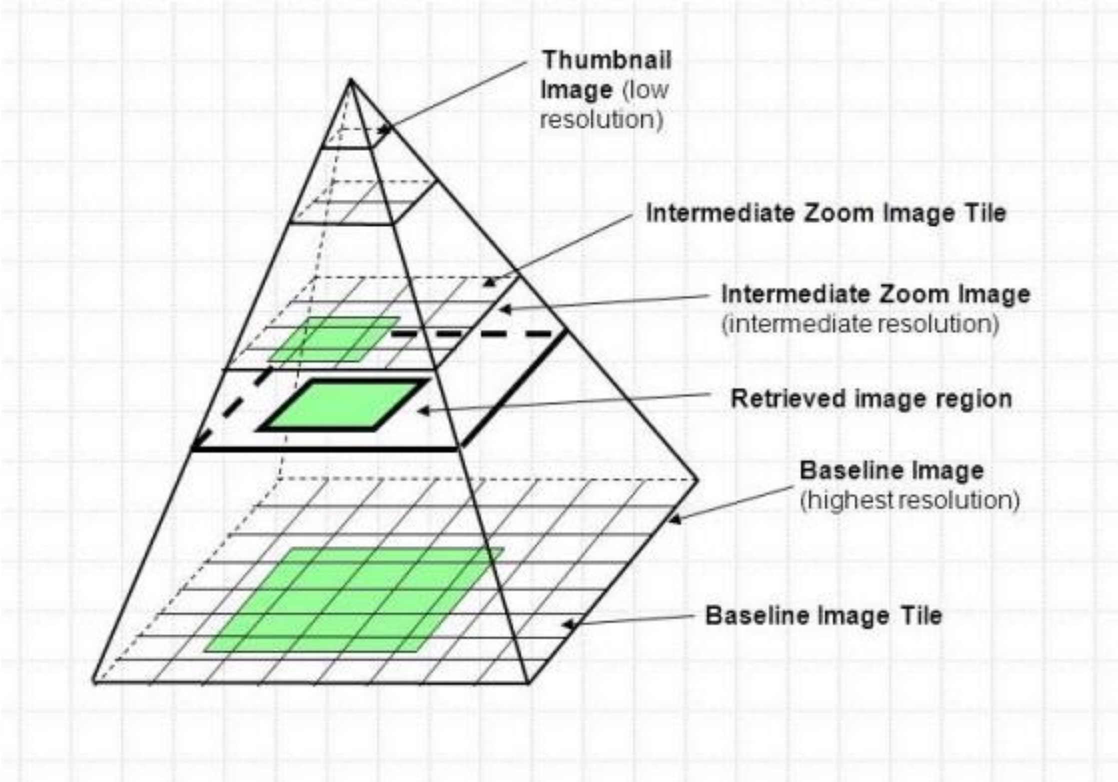
Xavier DESCOMBES - INRIA

EQUIPE
MORPHEME
Sophia Antipolis Mediterranee

2022

Histopathology data

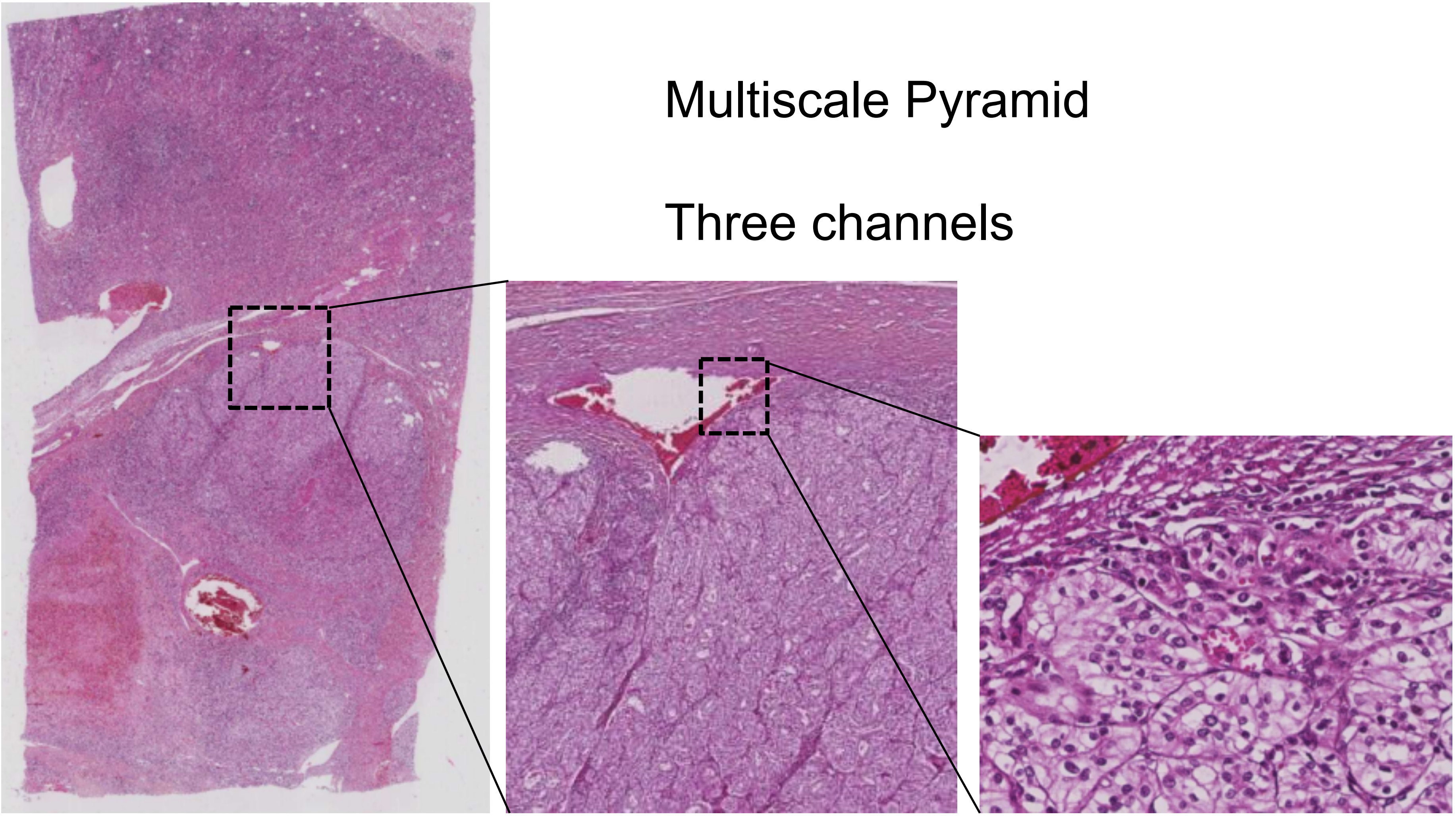
Pyramidal structure of data (Whole_Slide Imaging format)



Histopathological data

Multiscale Pyramid

Three channels



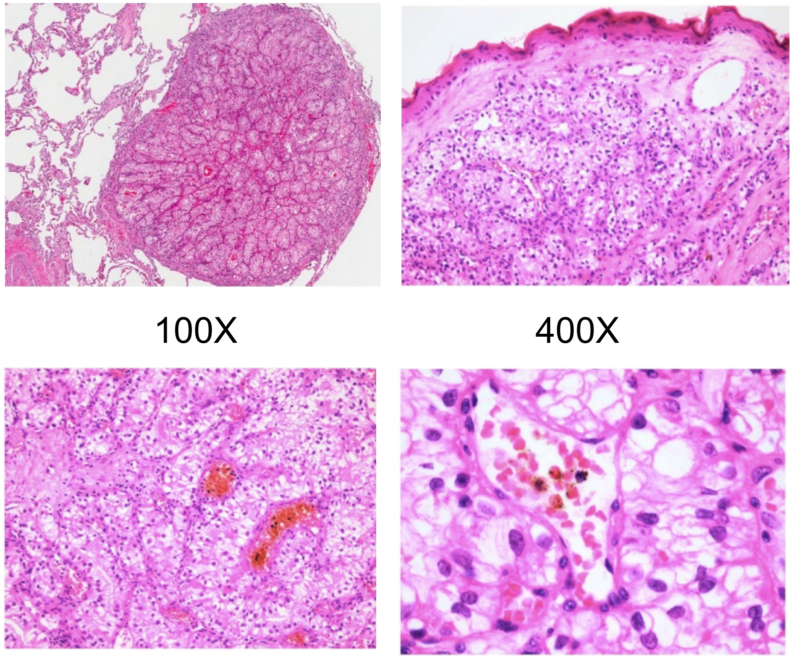
(822 x 1.365) (13.152 x 21.840) (52.608 x 87.360)

Background

Pathological is the gold standard for RCC diagnosis

4.7X 40X

100X 400X



Kassam, Karim , et al. ***Case Reports in Dermatological Medicine,*** 2013 (2013): 1-3.

Challenges

- The size of the pathological image is large.
- The diagnosis is time consuming and laborious.
- Inconsistent diagnosis, poor repeatability.
- Lack of pathologists.

Background

Computer Aided Diagnosis (CAD)

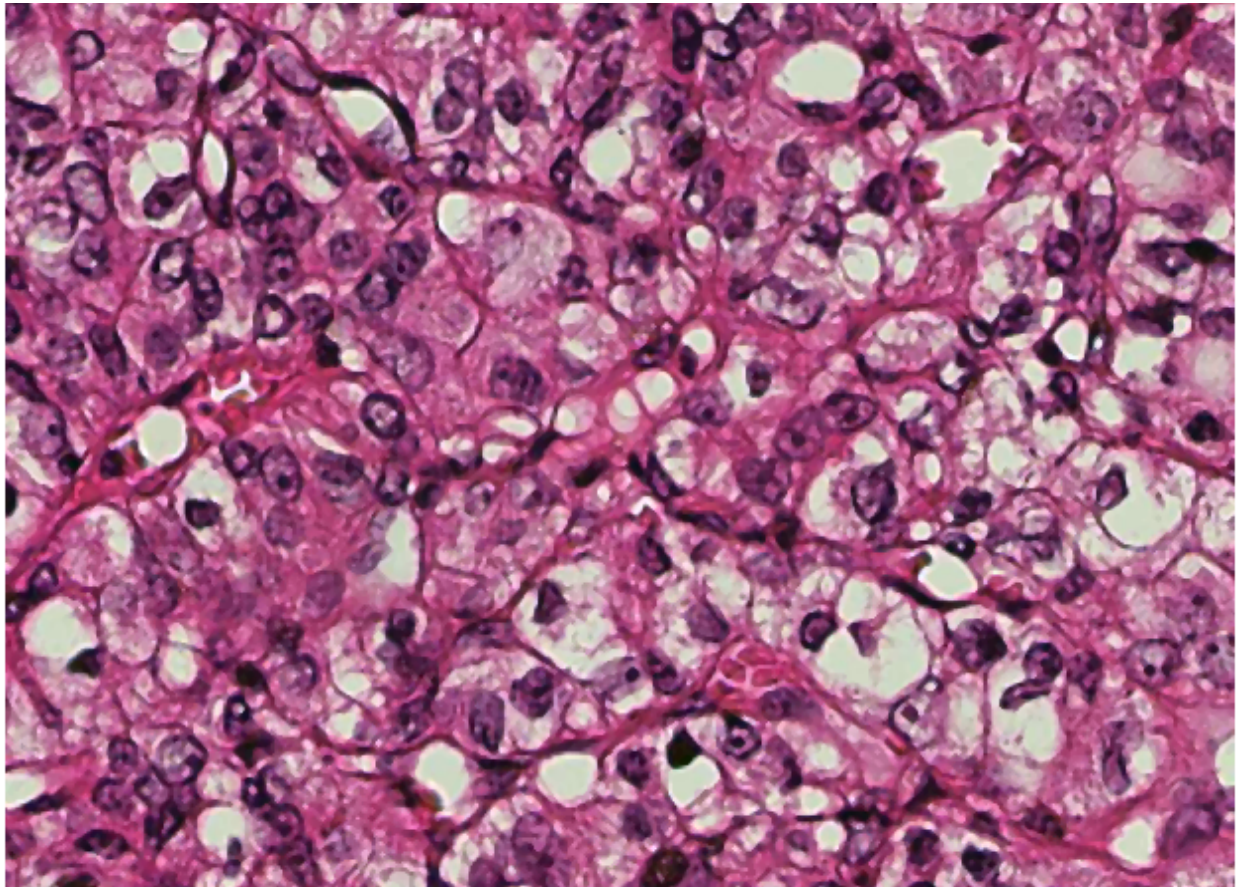
- CAD is an interdisciplinary technology combining elements of artificial intelligence and computer vision with radiological and pathology image processing.

- CAD also has potential future applications in digital pathology with the advent of whole-slide imaging and machine learning algorithms. And it's being investigated for the standard H&E stain.

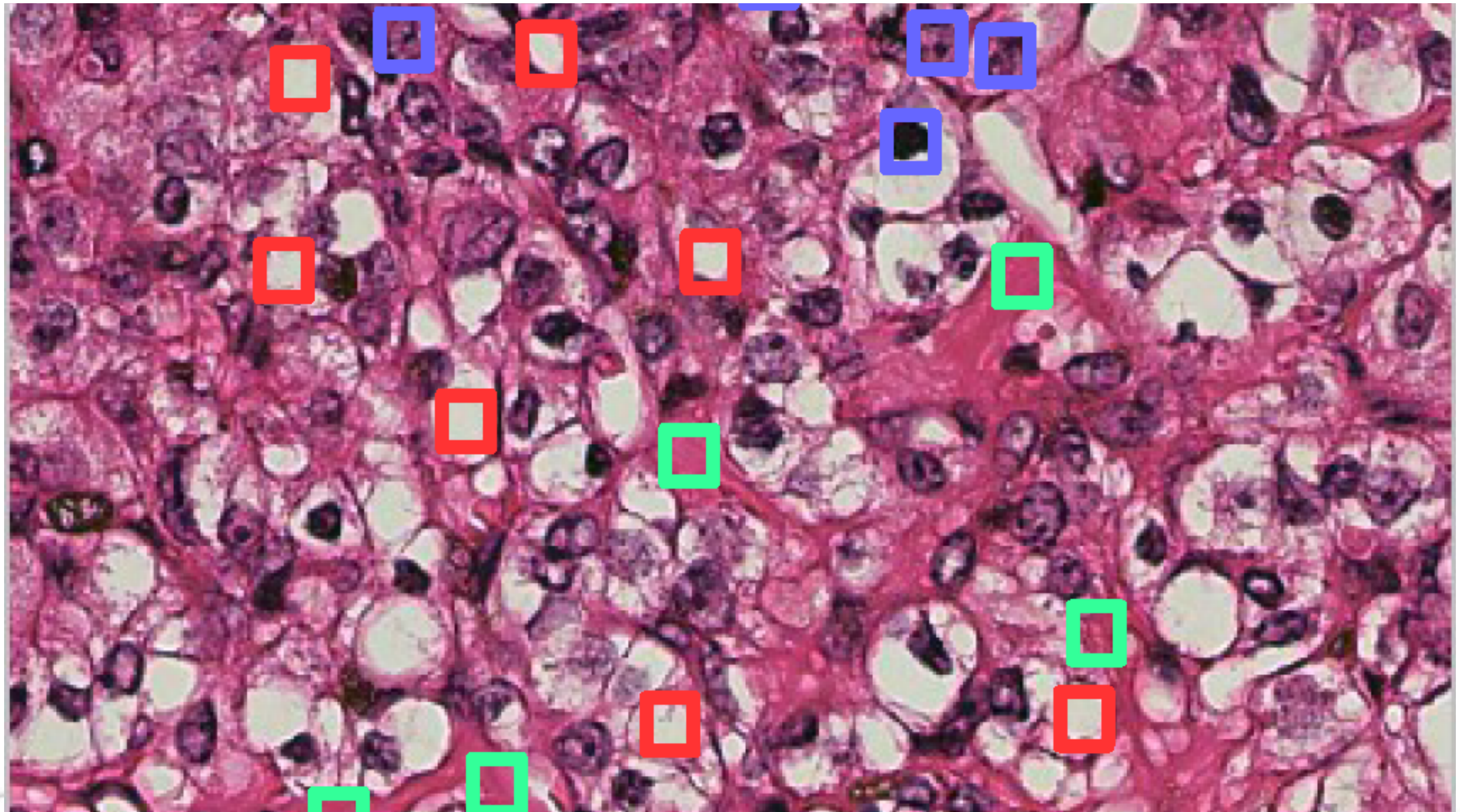
- The application of machine learning in clinical diagnosis can save manpower, time and even reducing the risk of misdiagnosis.

Shiraishi, Junji, et al. "Computer-aided diagnosis and artificial intelligence in clinical imaging." ***Seminars in nuclear medicine***. Vol. 41. No. 6. WB Saunders, 2011.

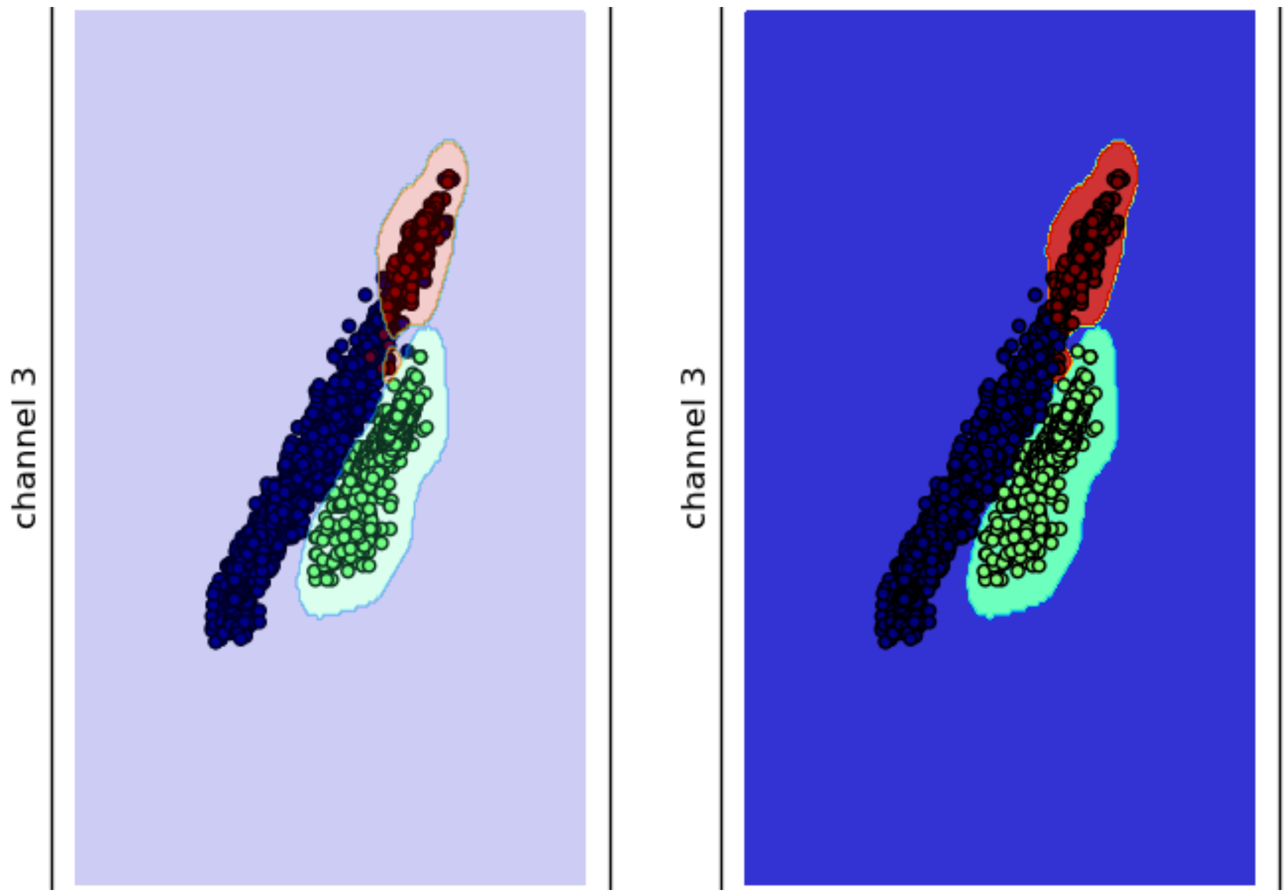
First example: Nuclei segmentation



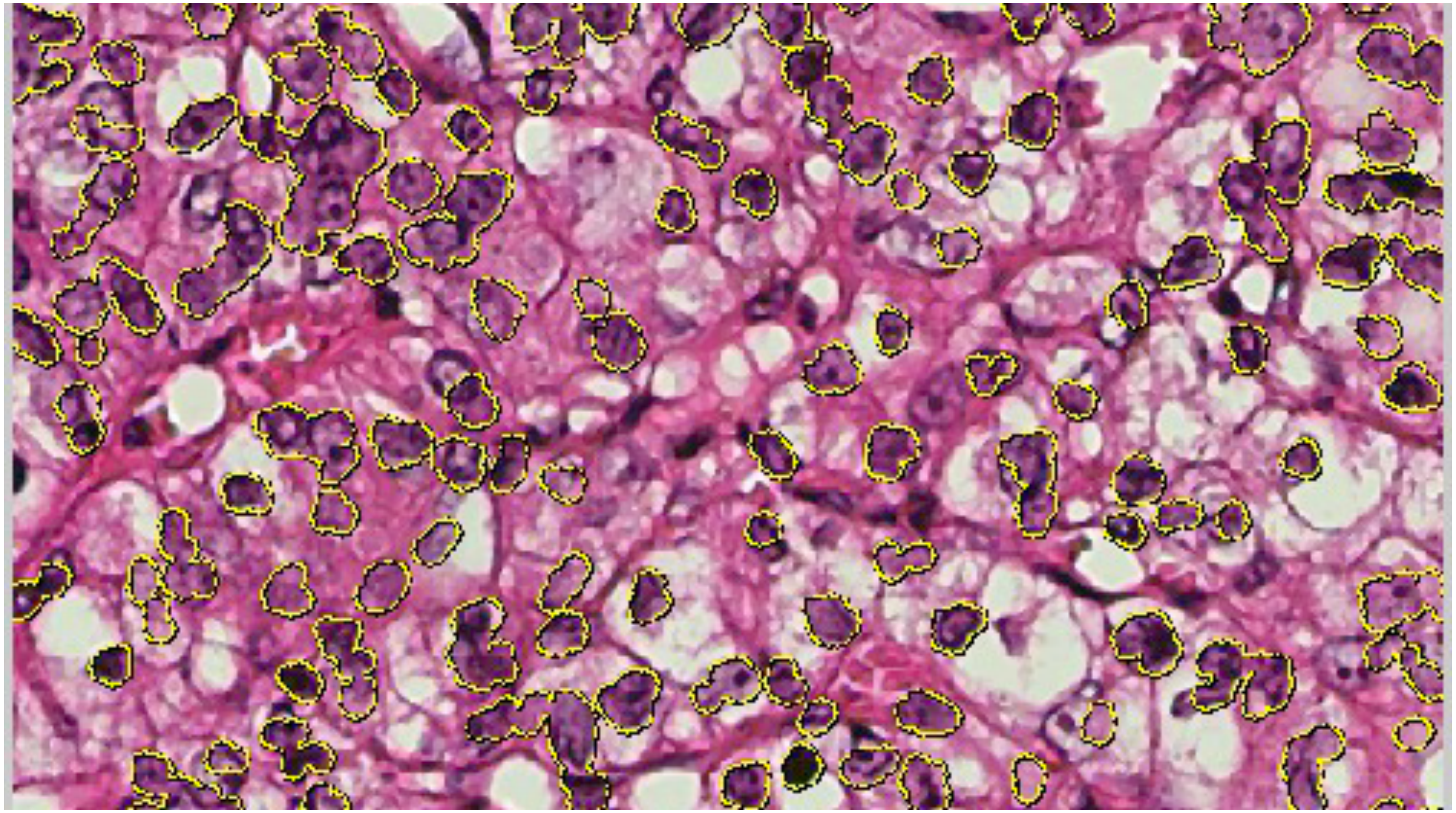
Learning set



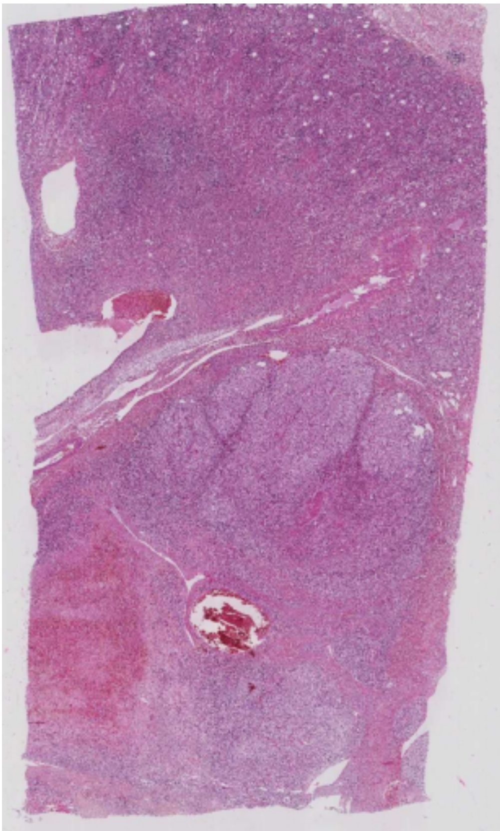
SVM classification

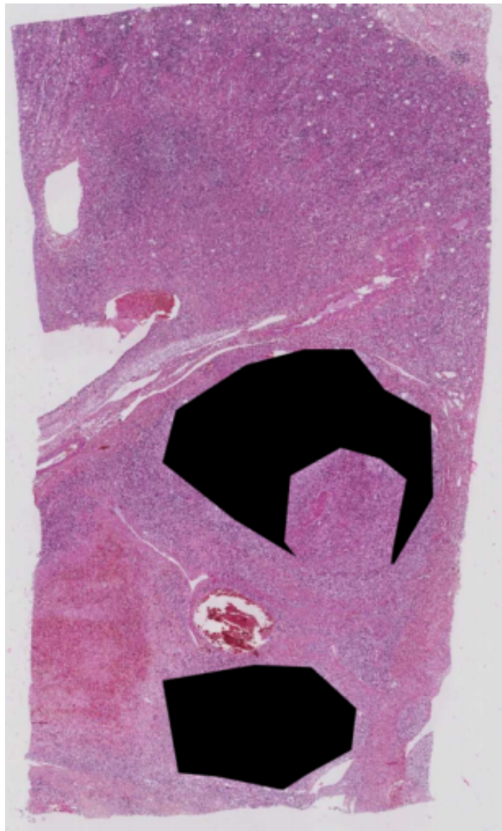


SVM Classification (plus « cleaning »)



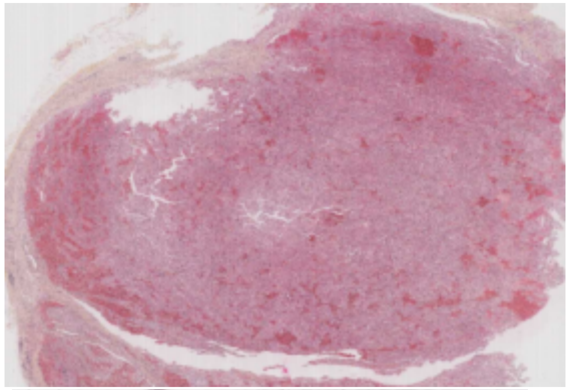
Classification of tumor areas (ROI)

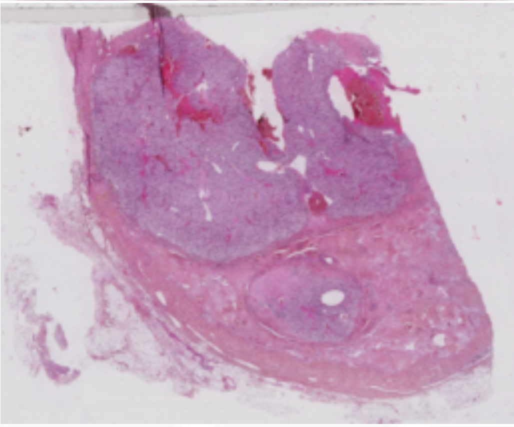


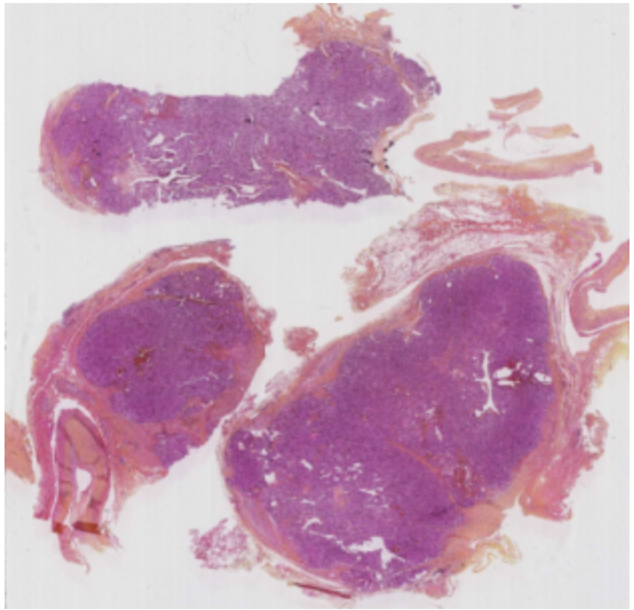


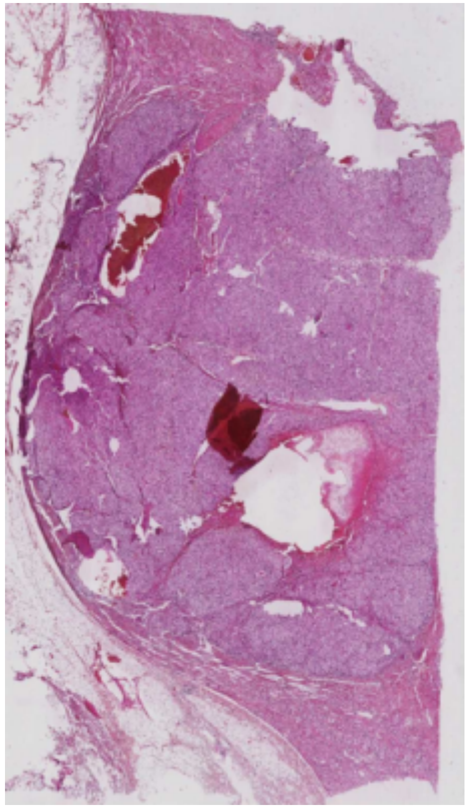
Challenge

- Variability between and within images
- Non informative areas (fat, blood…)
- Huge datasets (12 Go ≅ 100 000 pixels per axis)

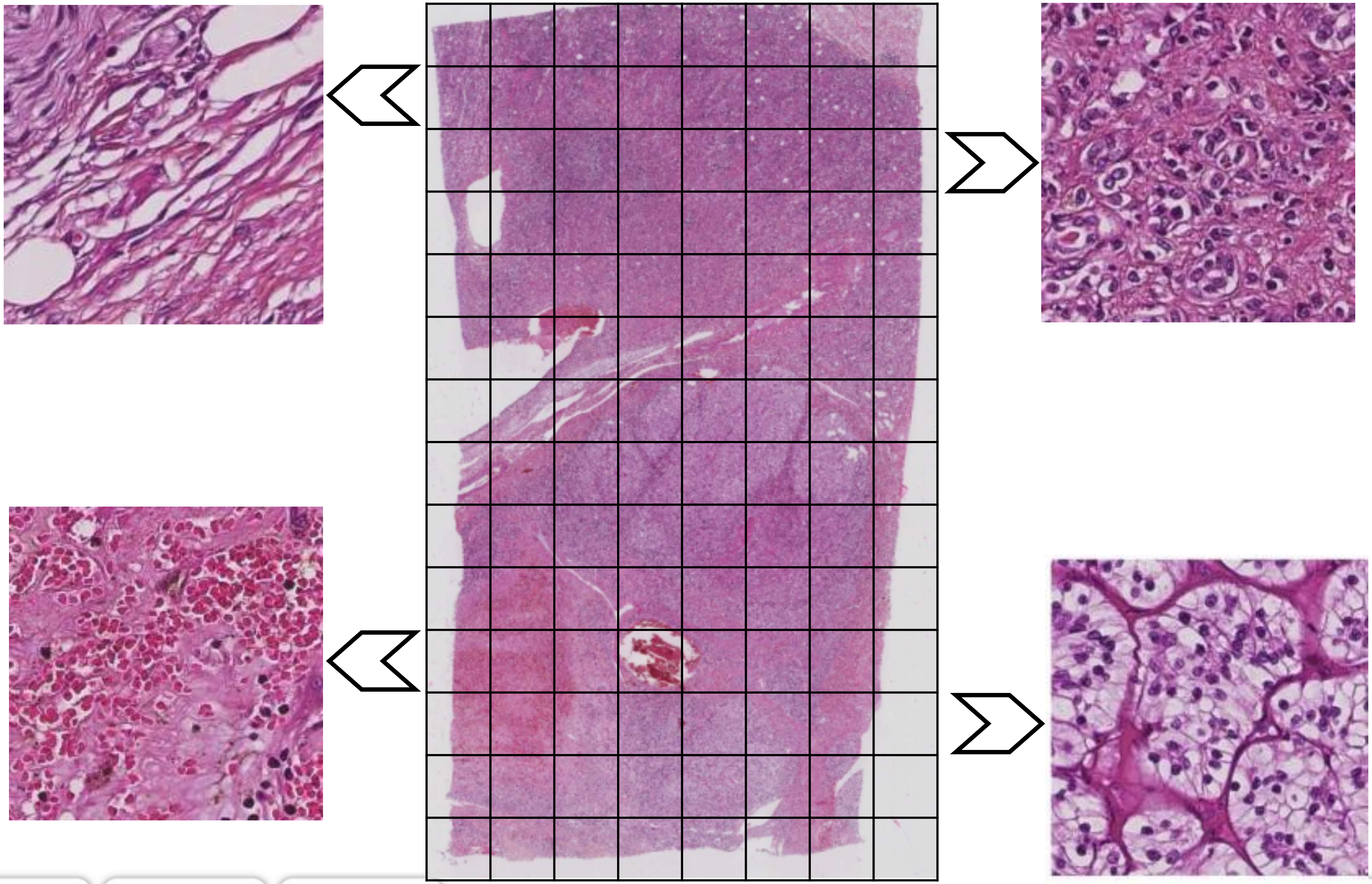








Patches classification



Tumor

Pre-Processing : color deconvolution

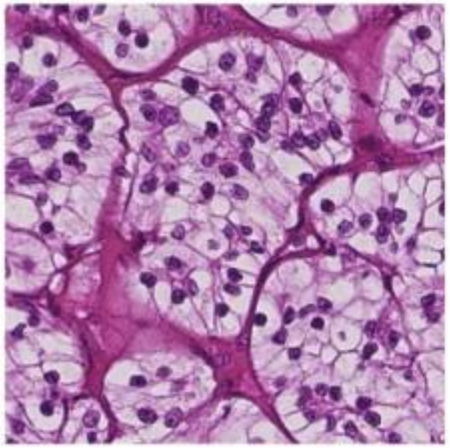
$$\forall c \in \mathbb{N}^2, H(c) = \frac{Rouge(c)}{C3(c)},$$

$$\forall c \in \mathbb{N}^2, E(c) = \frac{Vert(c)}{Rouge(c)}.$$

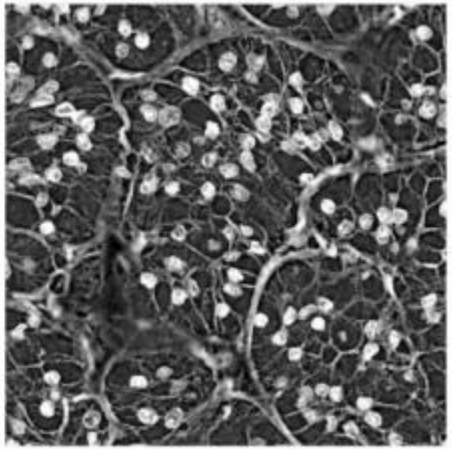
$$\forall c \in \mathbb{N}^2, C1(c) = \arctan\left[\frac{Rouge(c)}{max\left(Vert(c), Bleu(c)\right)}\right]$$

$$\forall c \in \mathbb{N}^2, C2(c) = \arctan\left[\frac{Vert(c)}{max\left(Rouge(c), Bleu(c)\right)}\right]$$

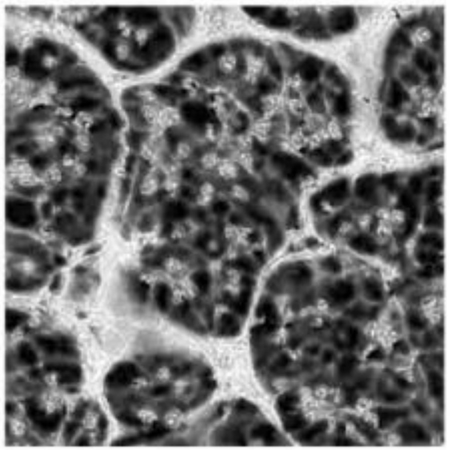
$$\forall c \in \mathbb{N}^2, C3(c) = \arctan\left[\frac{Bleu(c)}{max\left(Rouge(c), Vert(c)\right)}\right]$$



RGB image



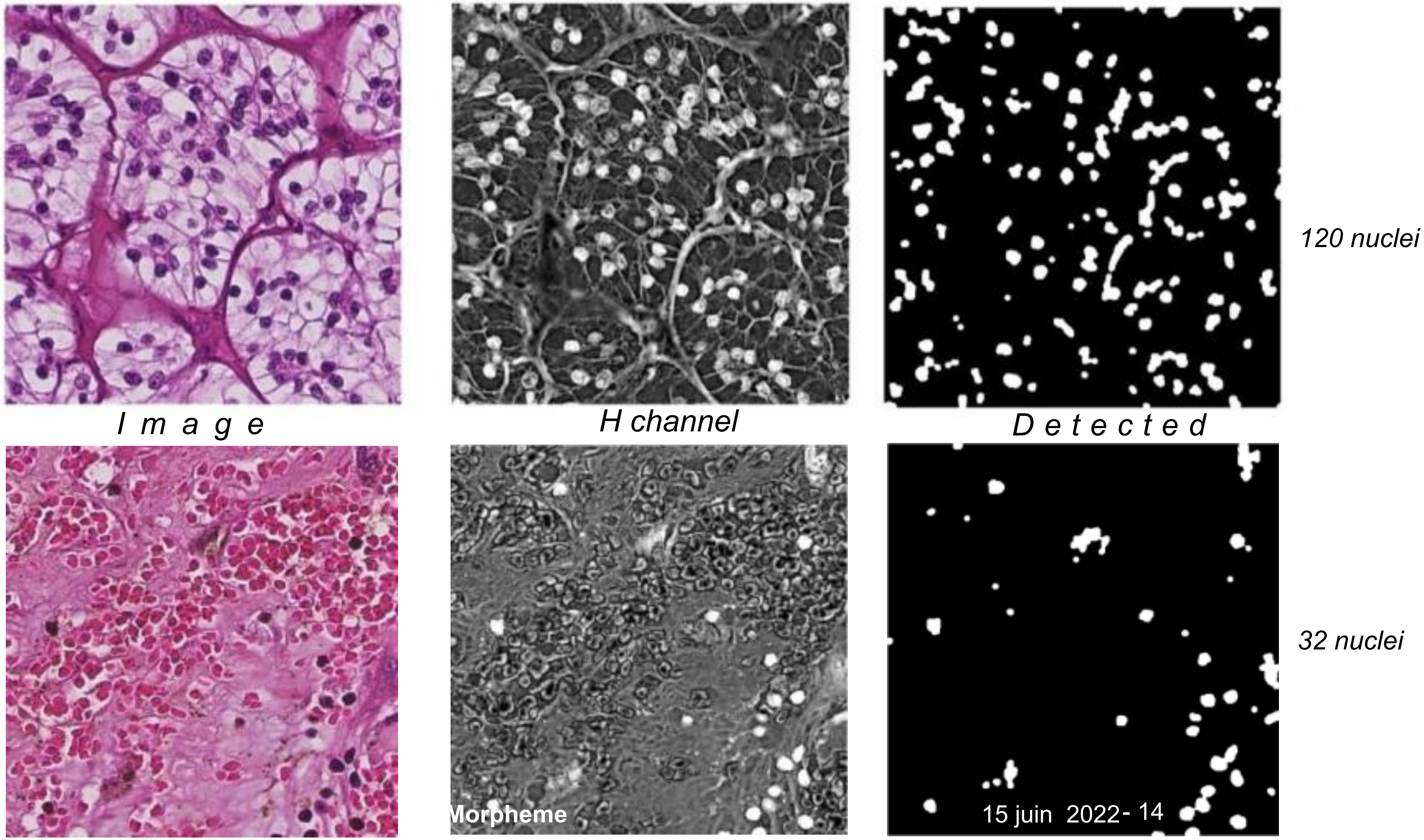
H channel



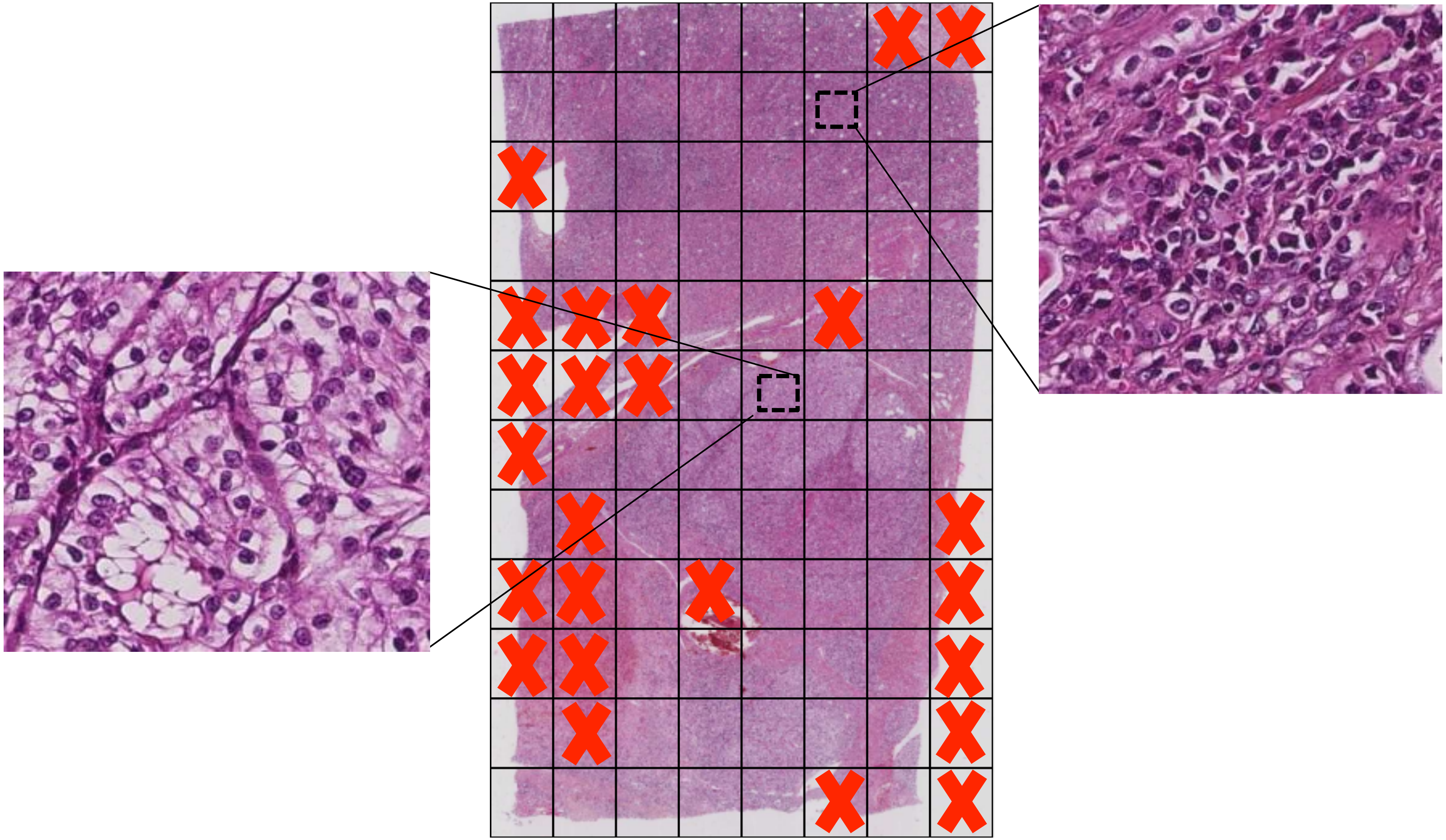
E channel

Reduce dataset

Remove patches with few cells



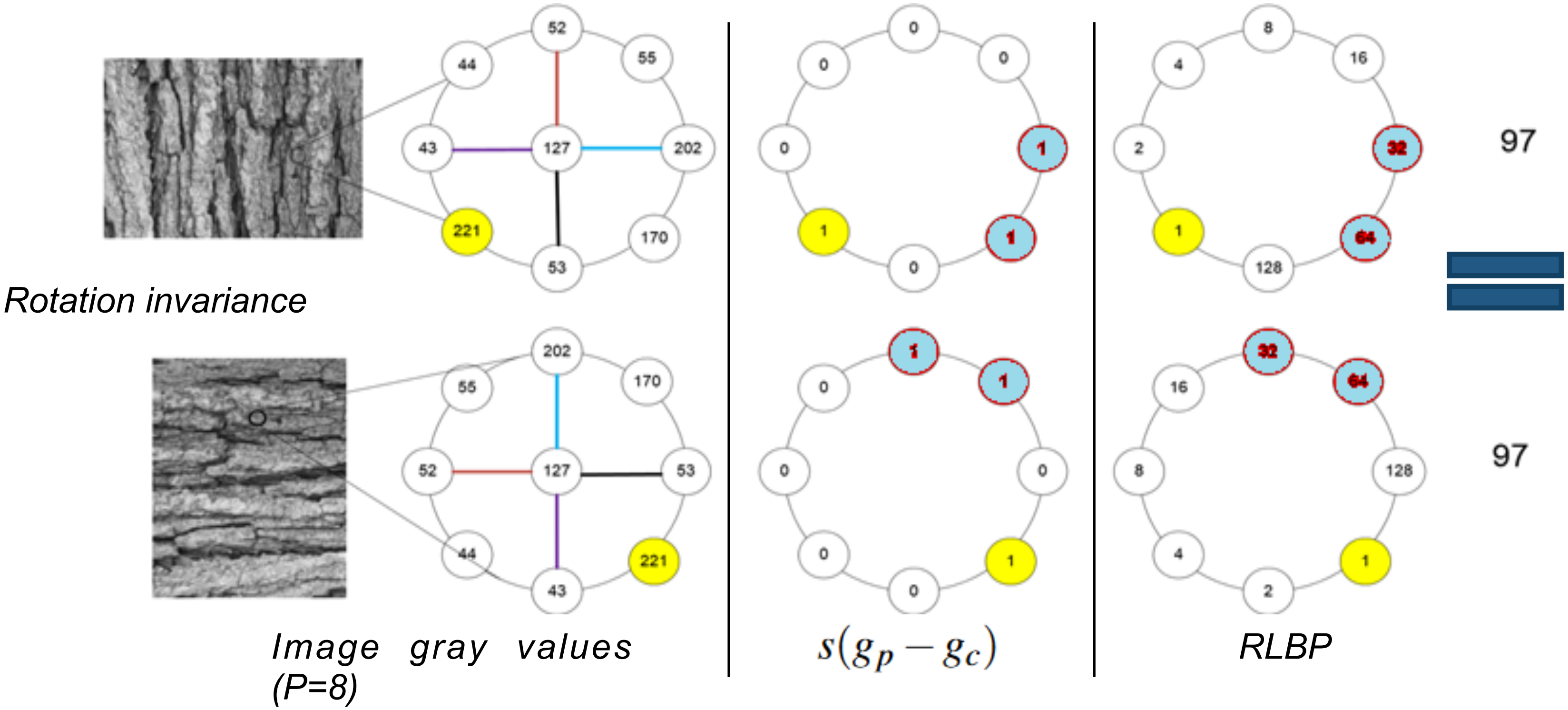
Reduce dataset



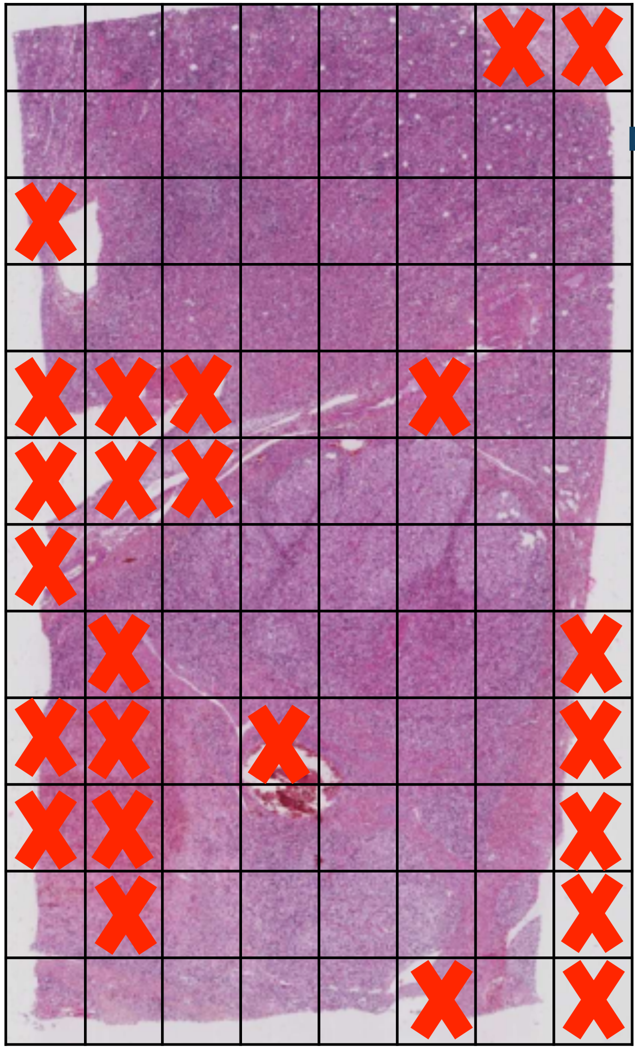
Features extraction : local binary patterns

$$D = \underset{p \in (0,1 \ldots P-1)}{\arg\max} \; |g_p - g_c|$$

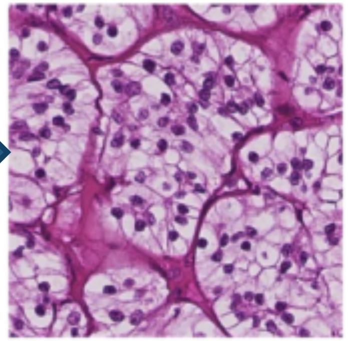
$$RLBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^{mod(p-D,P)}$$



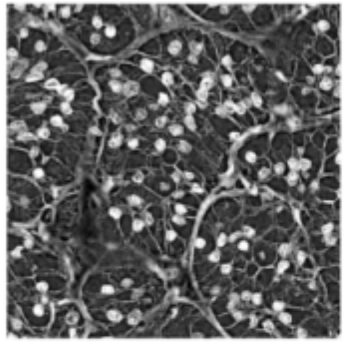
Classification : k-means



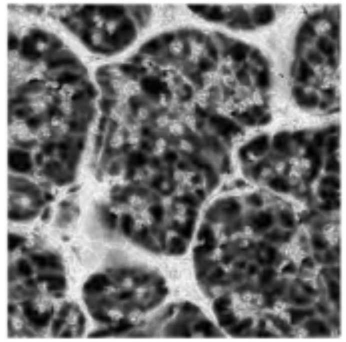
Result after nuclei detection



RGB image



H channel



E channel

RLBP- H	14	440	8700	…	…	…	…	…	…	745
RLBP- E	945	560	163	…	…	…	…	…	…	12

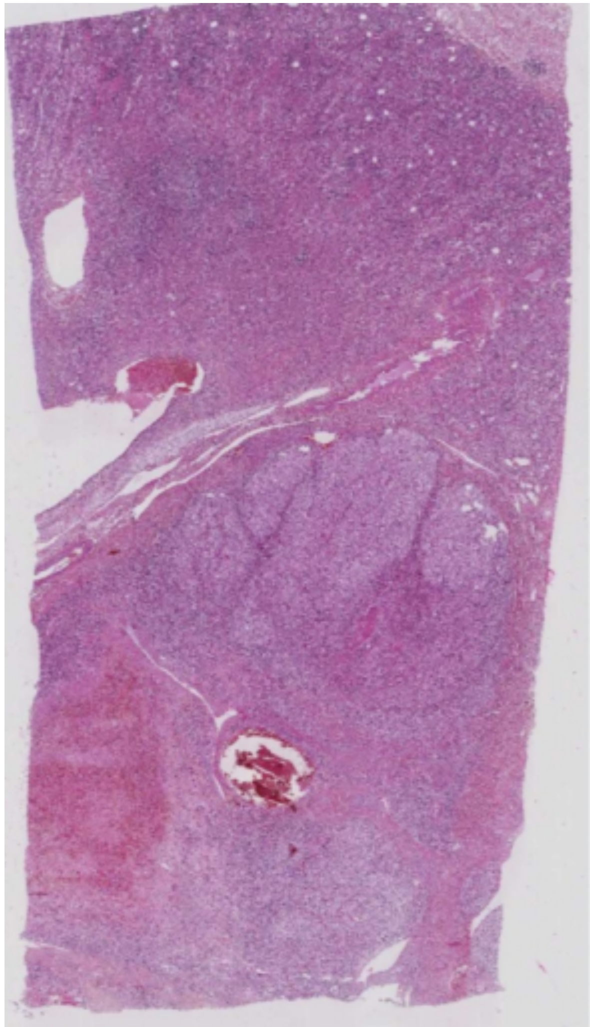
P=16, R=3 (65.536 x 2 patterns)

Find the Most Frequent patterns from the training images

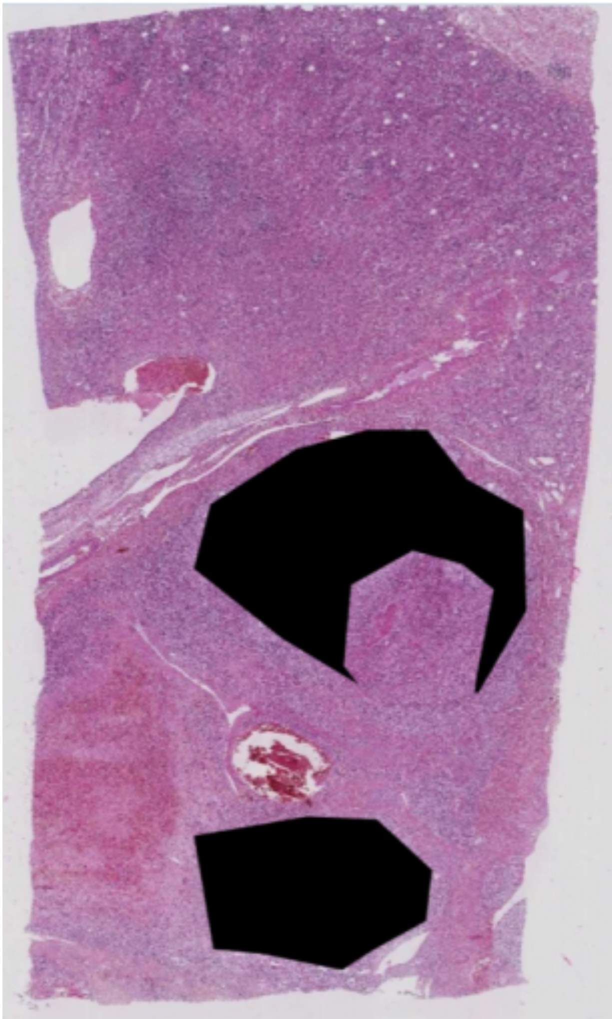
Classification K-NN

- *Classes number: 02 (Tumor / Not-tumor)*
- *Learning : 10 slides ➔ 850 patches*
- *Test : 05 slides*

Result on learning set



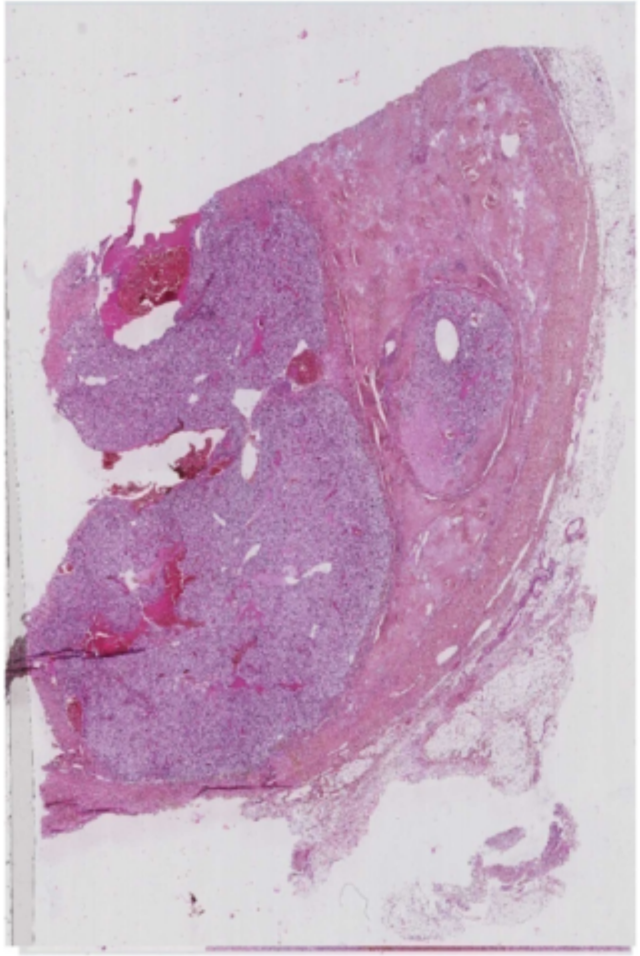
Input image

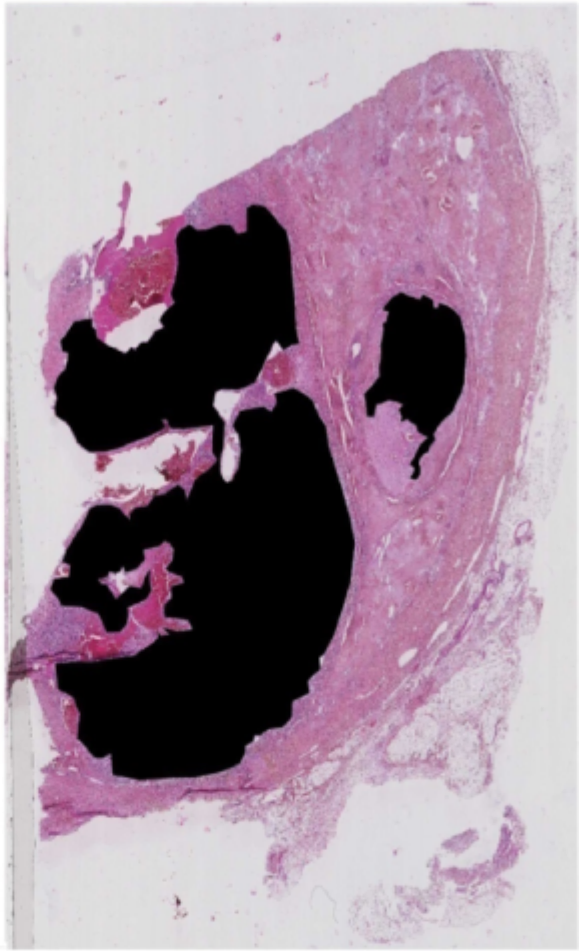


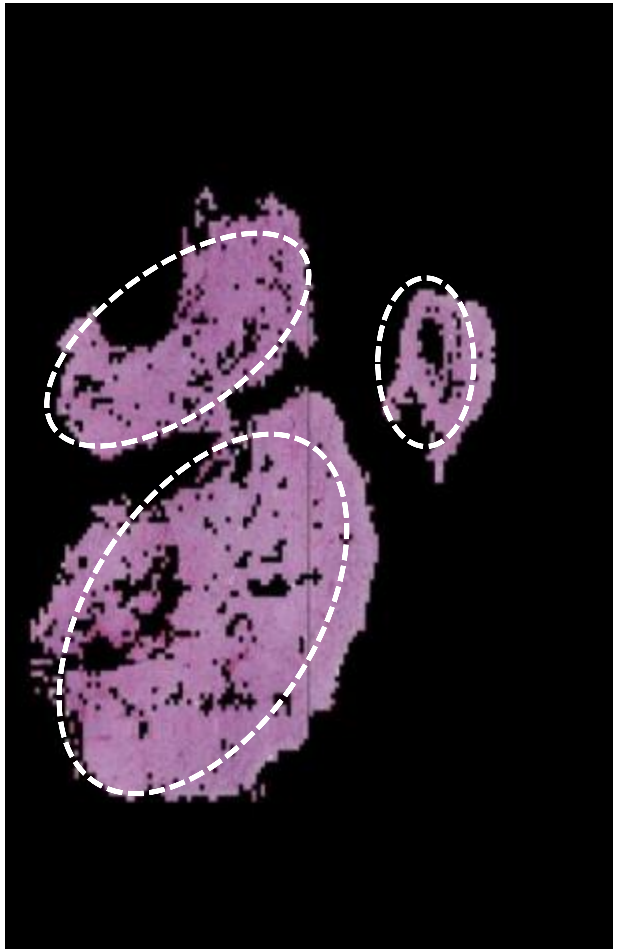
Ground Truth

Result

Result on test set

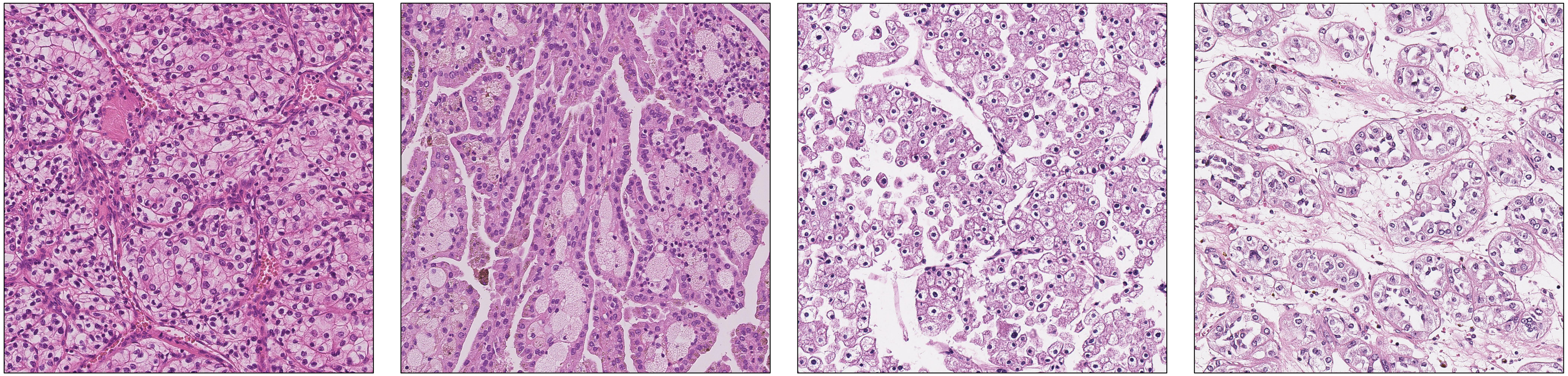






Input image *Ground Truth* *Result*

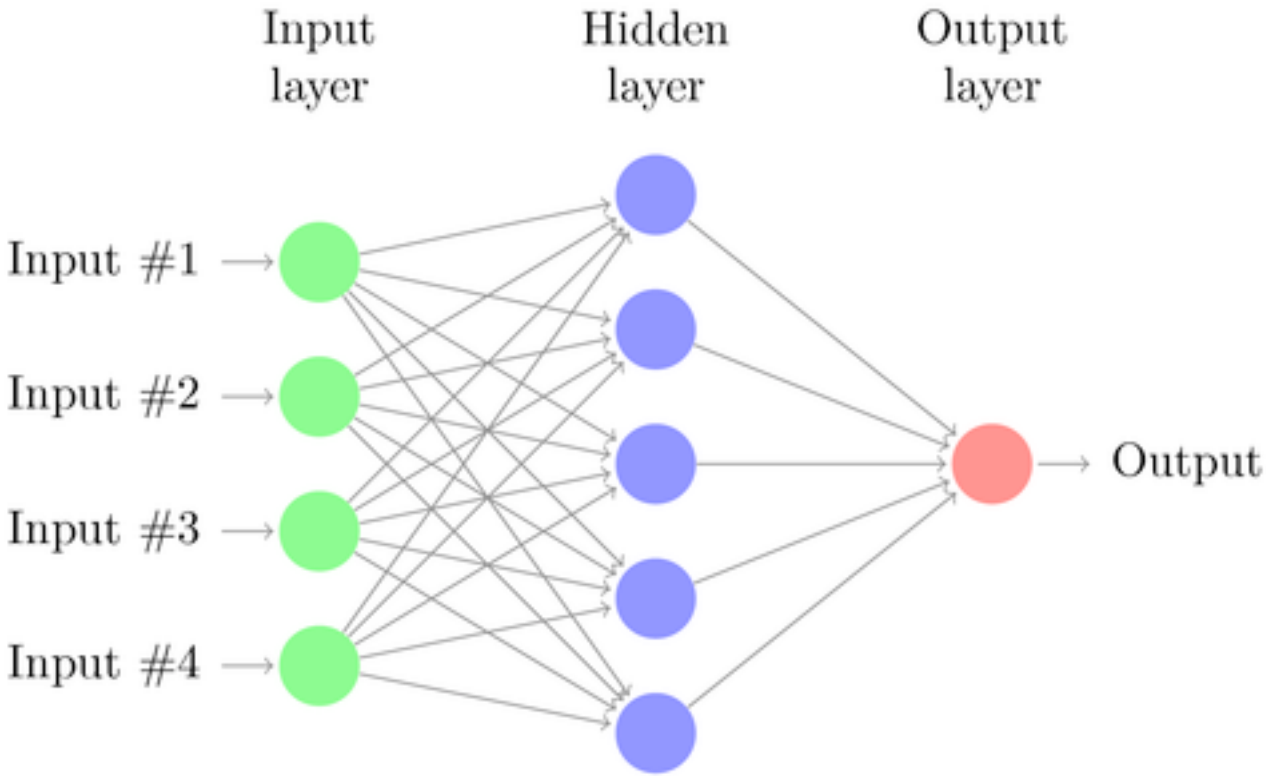
Kidney cancer classification



(a) (b) (c) (d)

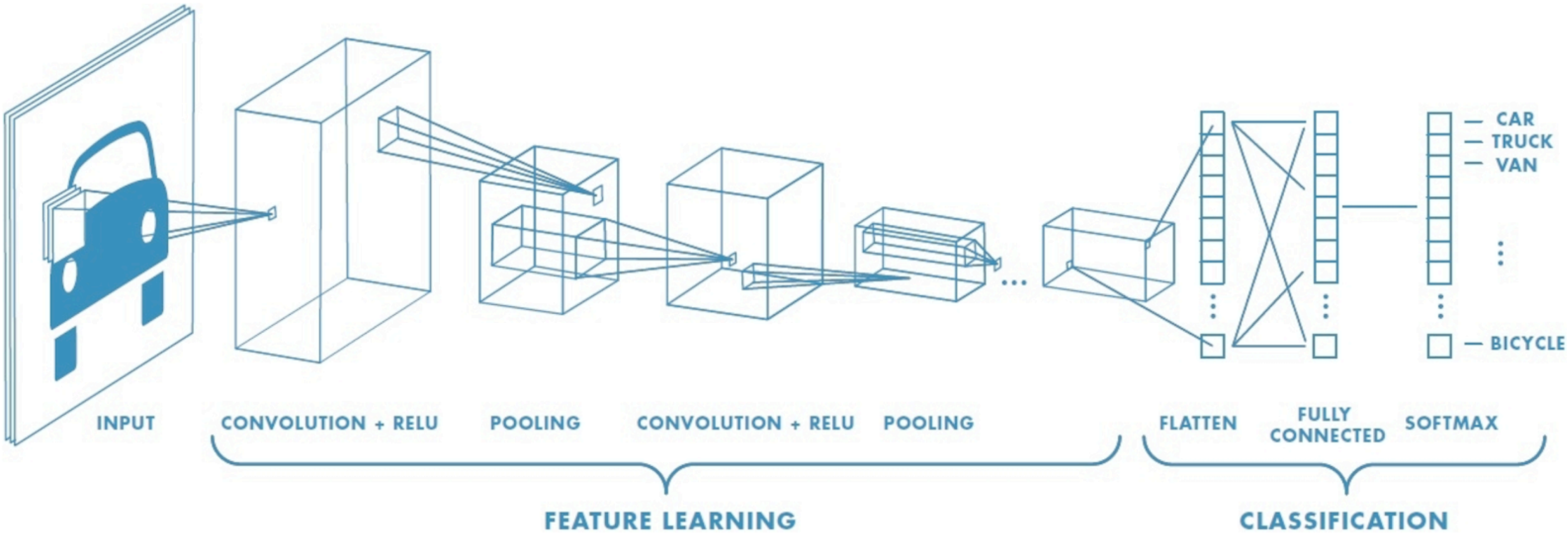
Neural Networks : The multi-layers perceptron

Perform classification by learning



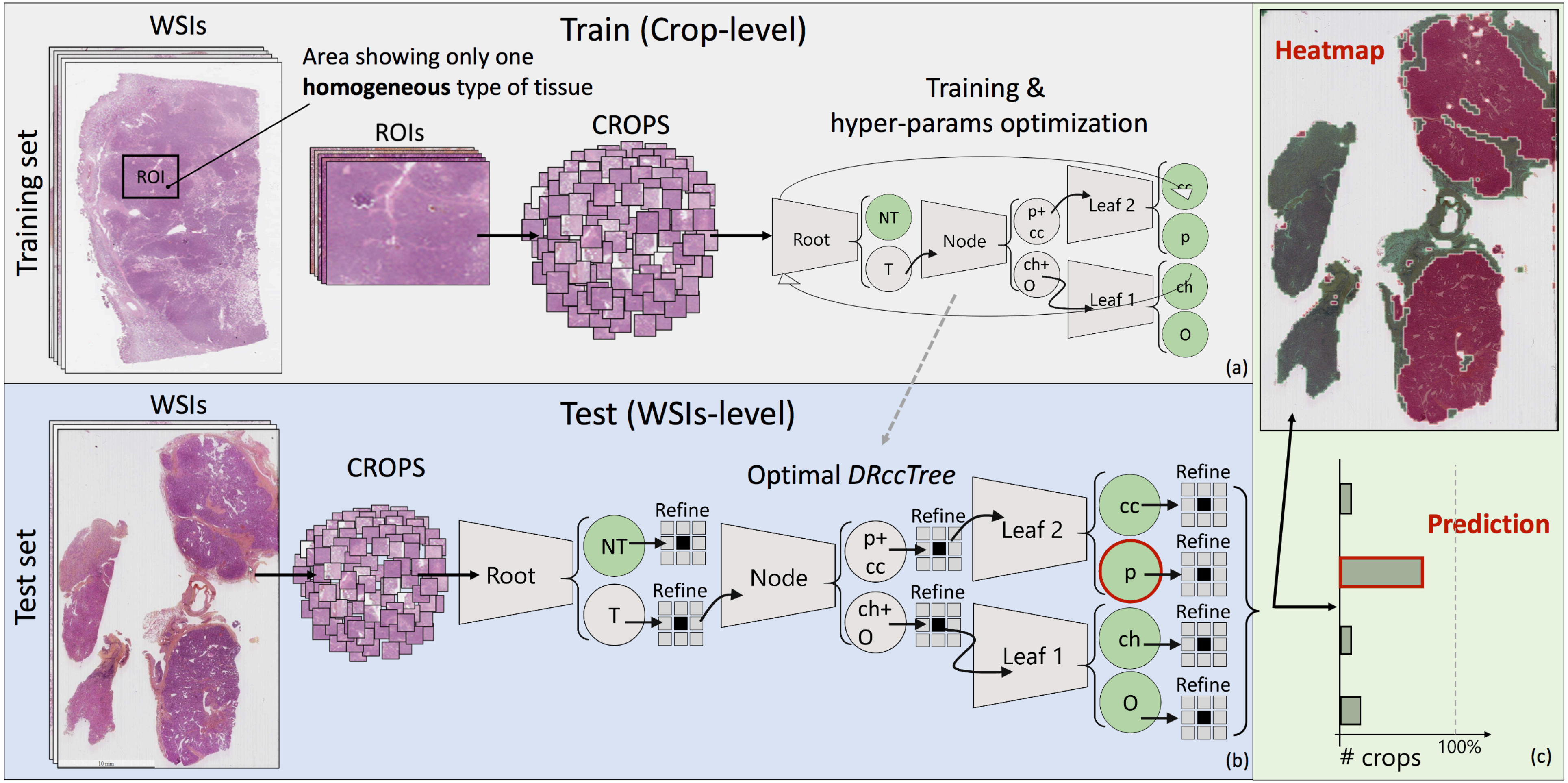
-> weights are learned from examples

Convolutional neural network : CNN

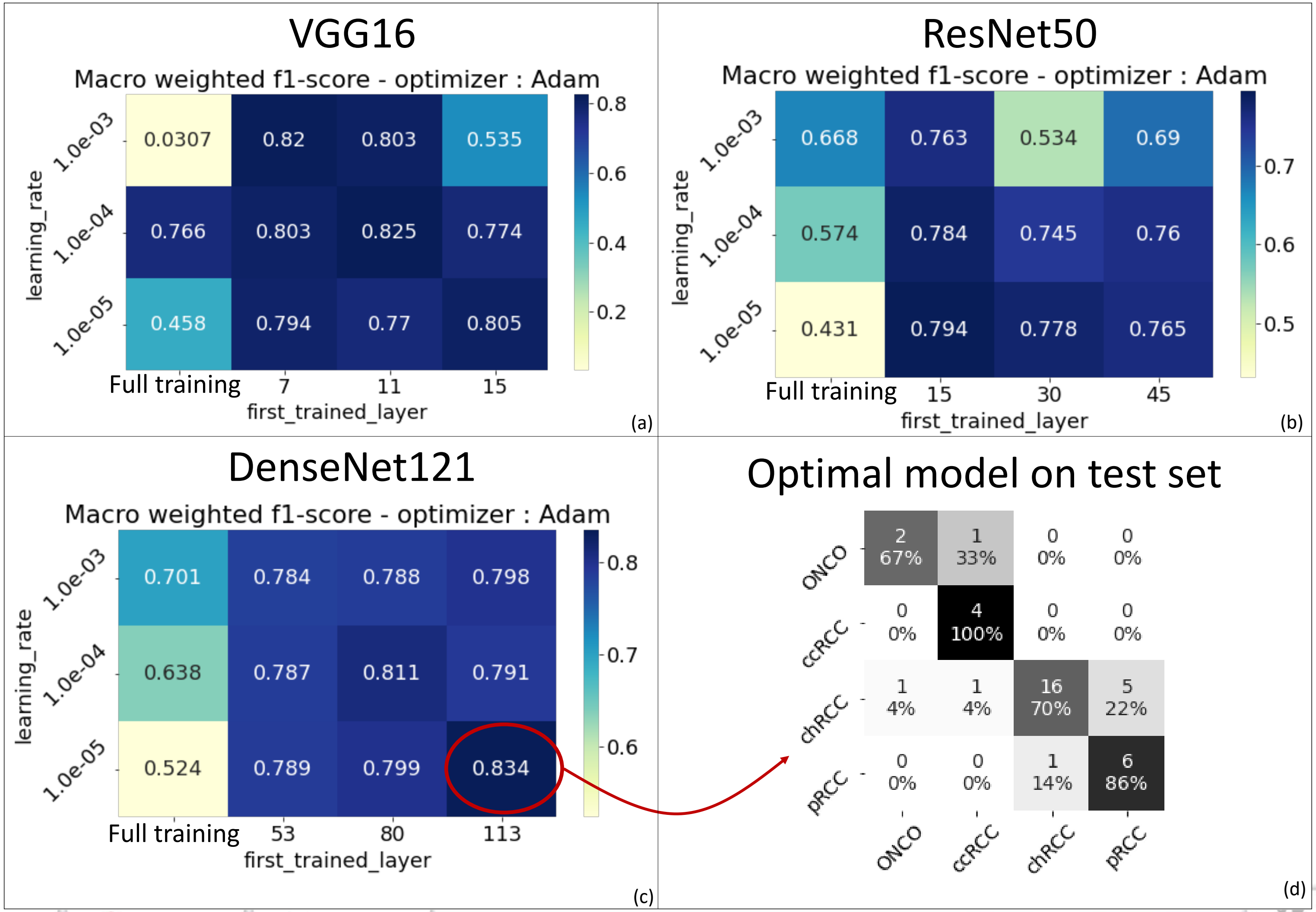


https://medium.com/@himadrisankarchatterjee/a-basic-introduction-to-convolutional-neural-network-8e39019b27c4

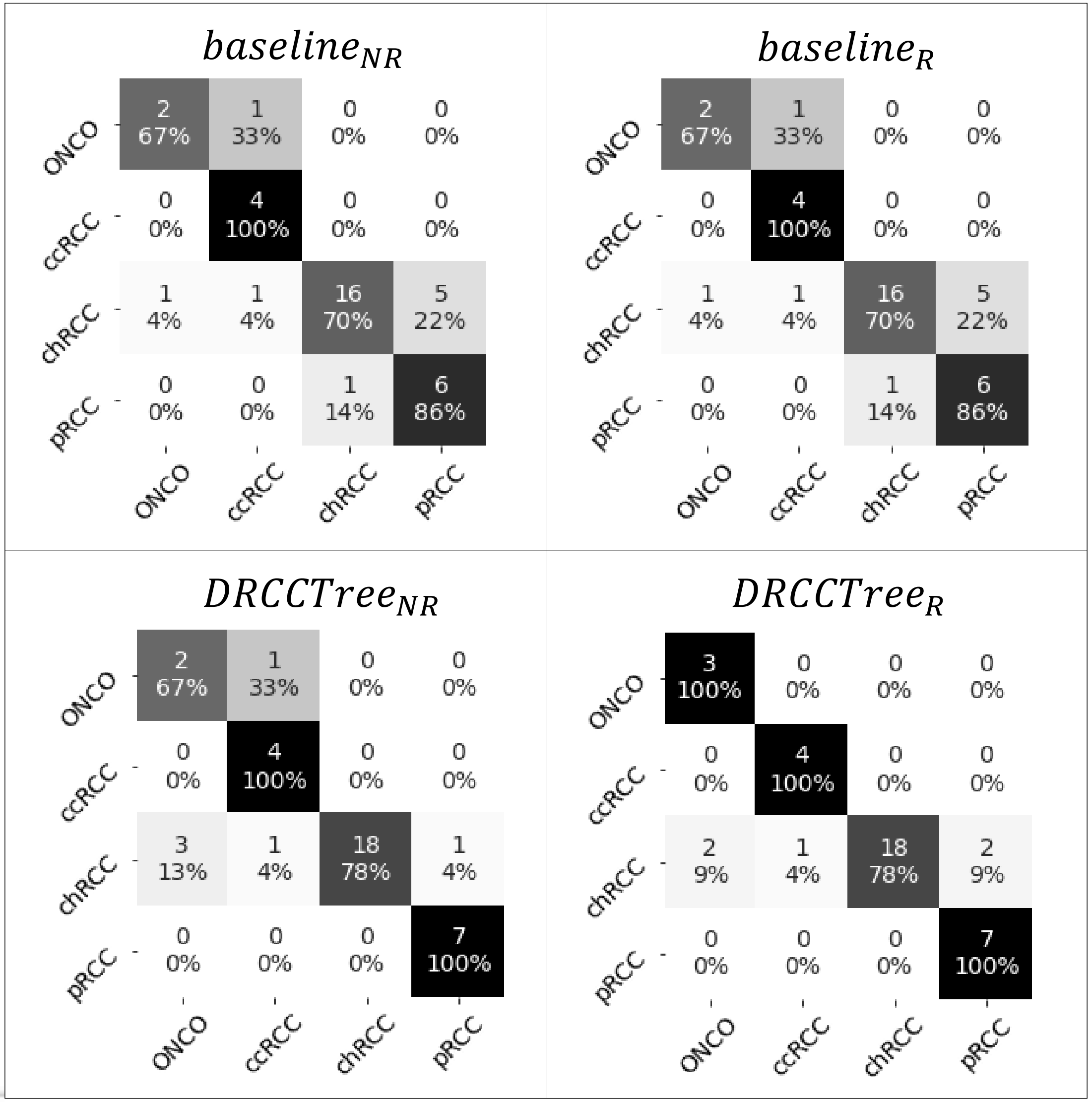
Kidney cancer classification



Kidney cancer classification

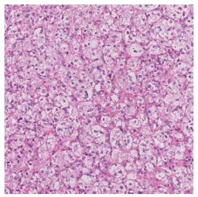


Kidney cancer classification

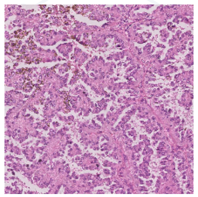


Classification from vascular network

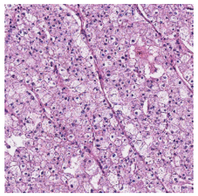
Motivations



ccRCC



pRCC



Chromophobe

1. Most of the current classification research is focused on cell characteristics. We consider **hand-crafted features of the vascular network for classifying RCCs and study resulting performances**.

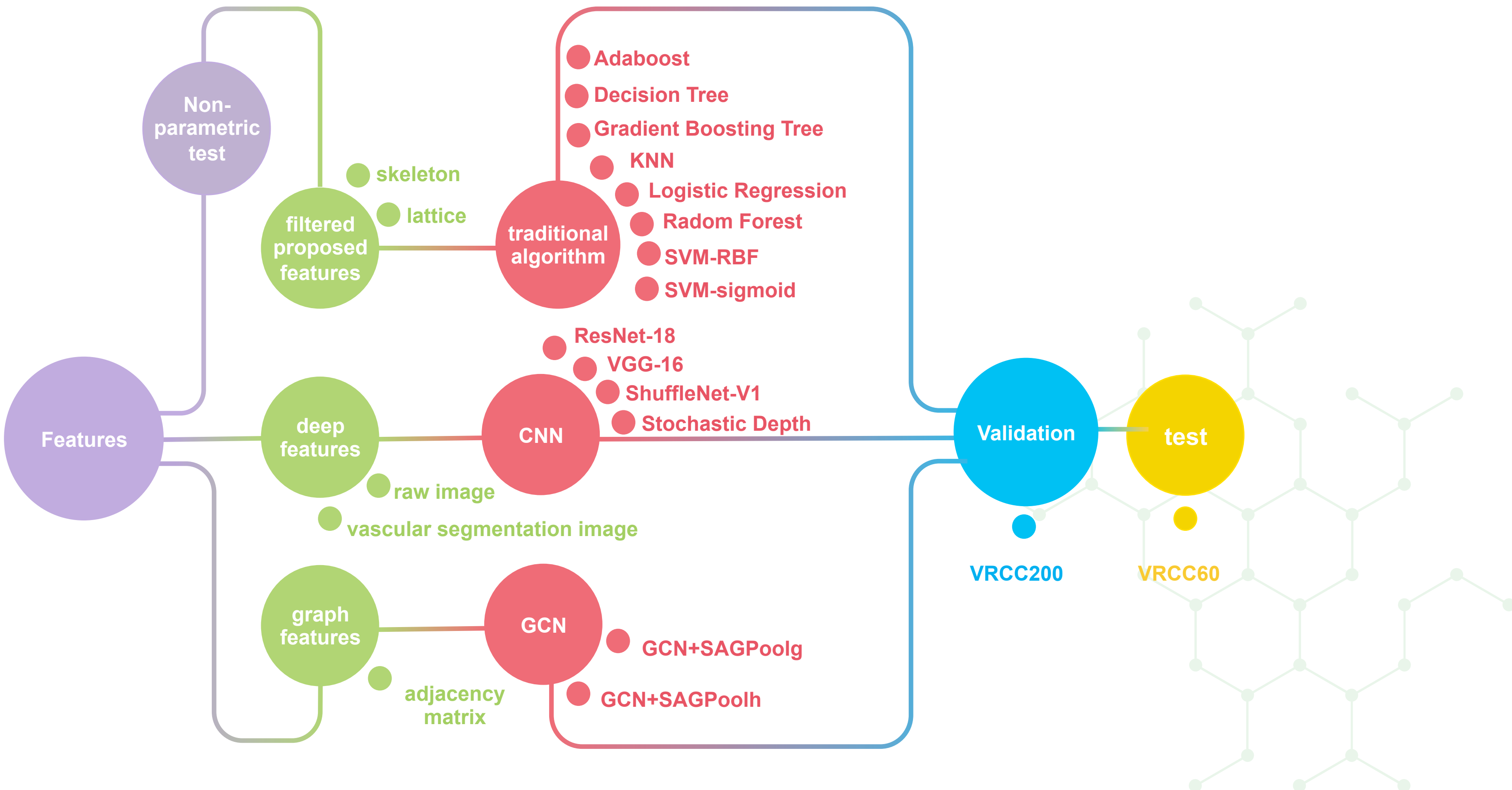
2. whether the **vascular structure of RCCs is sufficient to fully characterize the RCCs subtype** remains questionable. For example, **ccRCC is characterized by a "fishnet"** structure, the **pRCC has a tree-like structure and Chromophobe has linear structure**.

3. There are **no public** vascular annotated histological image **datasets** and **benchmark** for RCCs classification based on the vascular network currently.

4. If the **vascular network** can be used **alone** for **RCCs classification** is an open issue**.**

Experiments

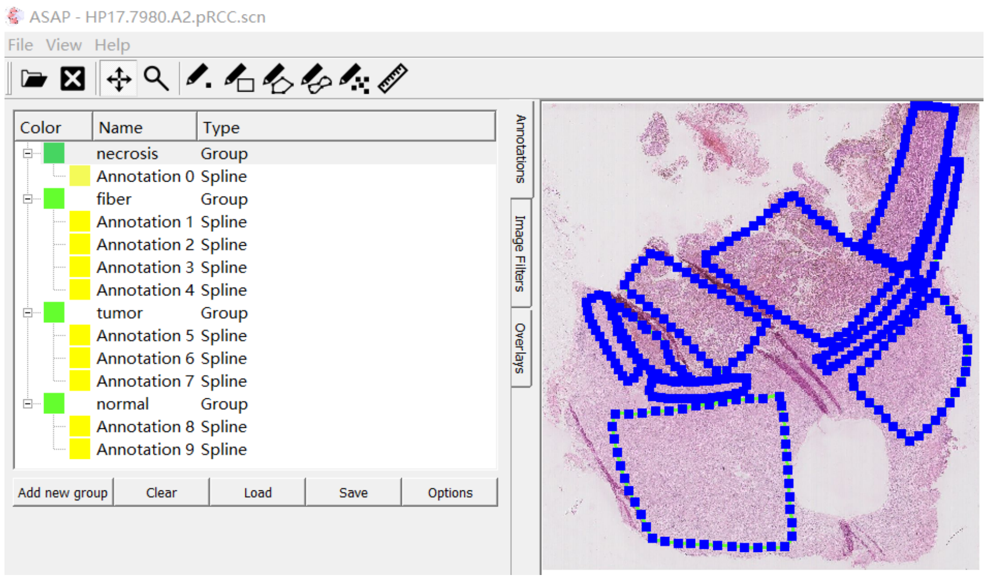
Overview of our experiments

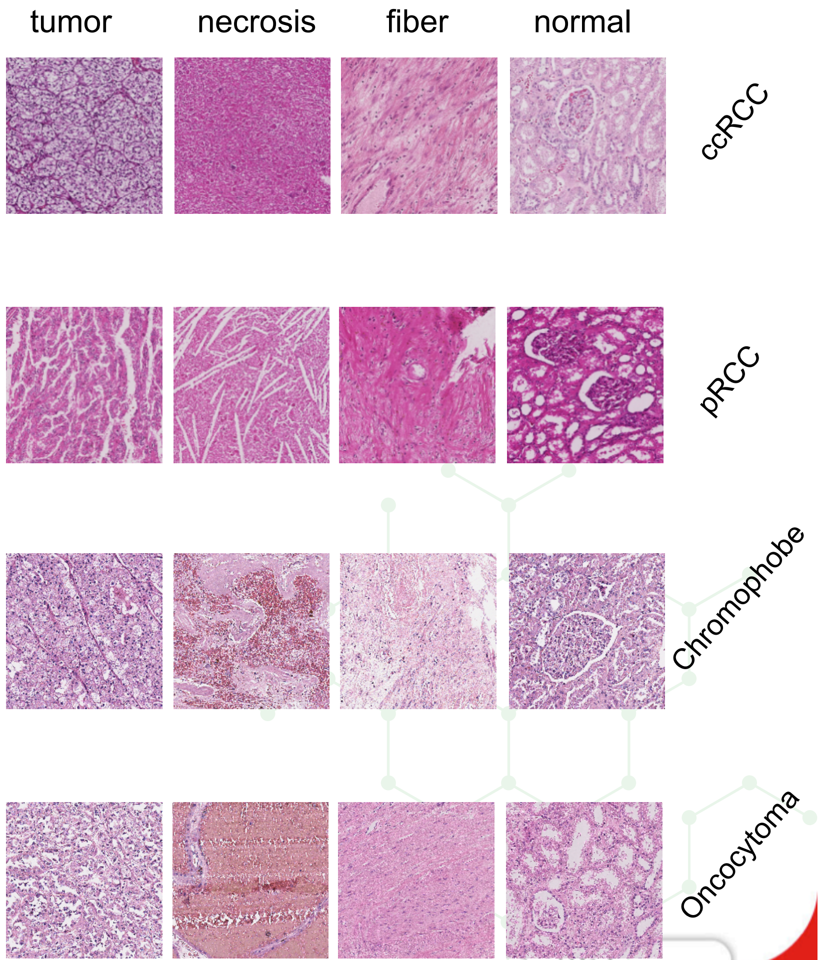


Dataset

BigRCC dataset building

Use ASAP software to get image patches (2000*2000 pixels) about tumor and non-tumor (necrosis、fiber、normal)

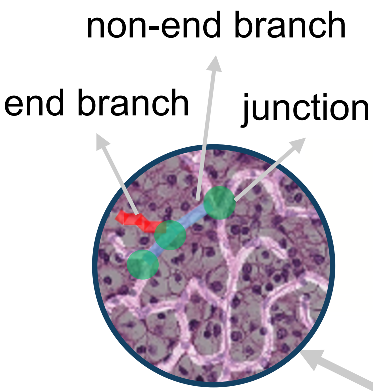




	Non-tumor			Tumor	Total
	Necrosis	Fiber	Normal		
ccRCC	3324	1941	7459	27287	39986
pRCC	1602	920	2105	13637	18254
Chromophobe	79	170	1037	3134	4420
Oncocytoma	36	48	438	248	770

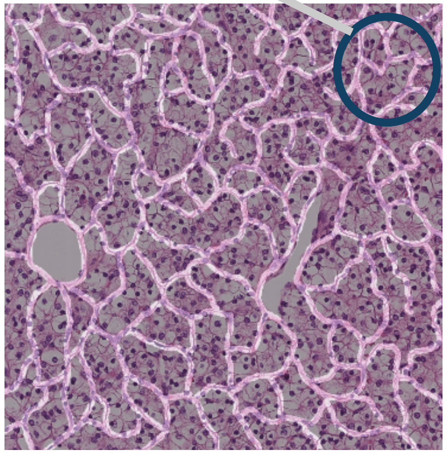
Vascular annotate of RCC subtype

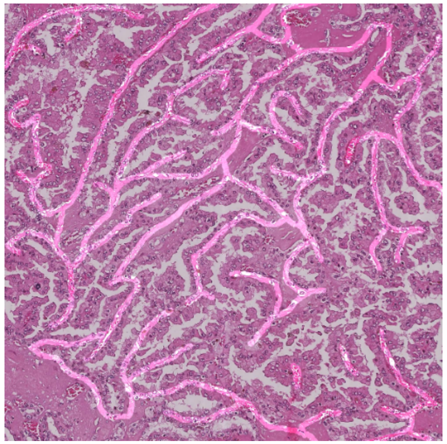
Vascular annotated dataset building

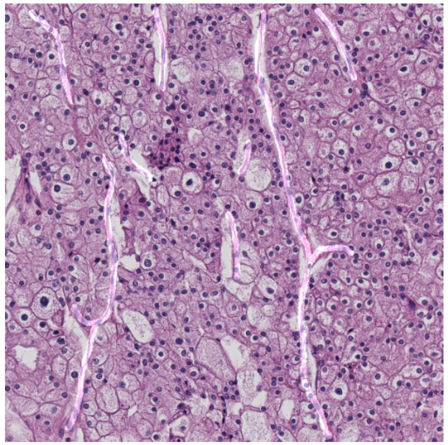


ImageJ

	image patches	patients
ccRCC	130	13
pRCC	130	14
chromophobe	166	4







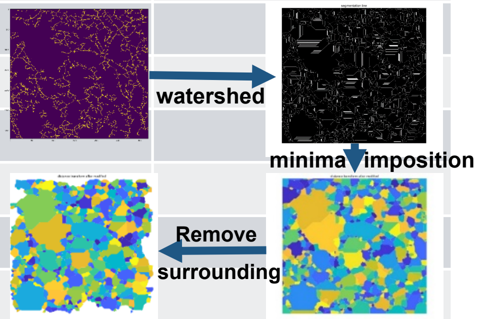
ccRCC pRCC Chromophobe

Features

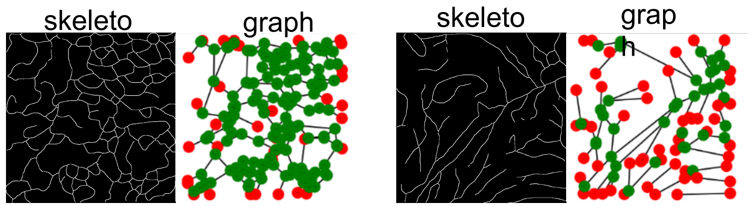
Vascular network features extraction

1. Hand-crafted Features (our proposed)

Skleton Features	Note	Lattice Features	Note
NE	the **N**umber of **E**nd branches	**mean area**	mean of all lattice areas
LE	average **L**ength of the **E**nd branches	**median area**	median of all lattice areas
small NE	**NE** that LE **less than** nuclear size × 10	**mean perimeter**	mean of all lattice perimeters
long NE	**NE** that LE **more than** nuclear size × 10	**median perimeter**	median of all lattice perimeters
NJ	**N**umber of **J**unctions	**mean eccentricity**	mean of all lattice eccentricities
LJ	average **L**ength of the non-**E**nd branches	**median eccentricity**	mean of all lattice eccentricities
density	sum of skeleton pixels		
NE/NJ	NE/NJ Ratio		
LE/LJ	LE/LJ Ratio		
NE/(LJ+LE)	NE/(LJ+LE) Ratio		
NJ/(LJ+LE)	NJ/(LJ+LE) Ratio		
LJ/(LJ+LE)	LJ/(LJ+LE) Ratio		



2. Graph Features

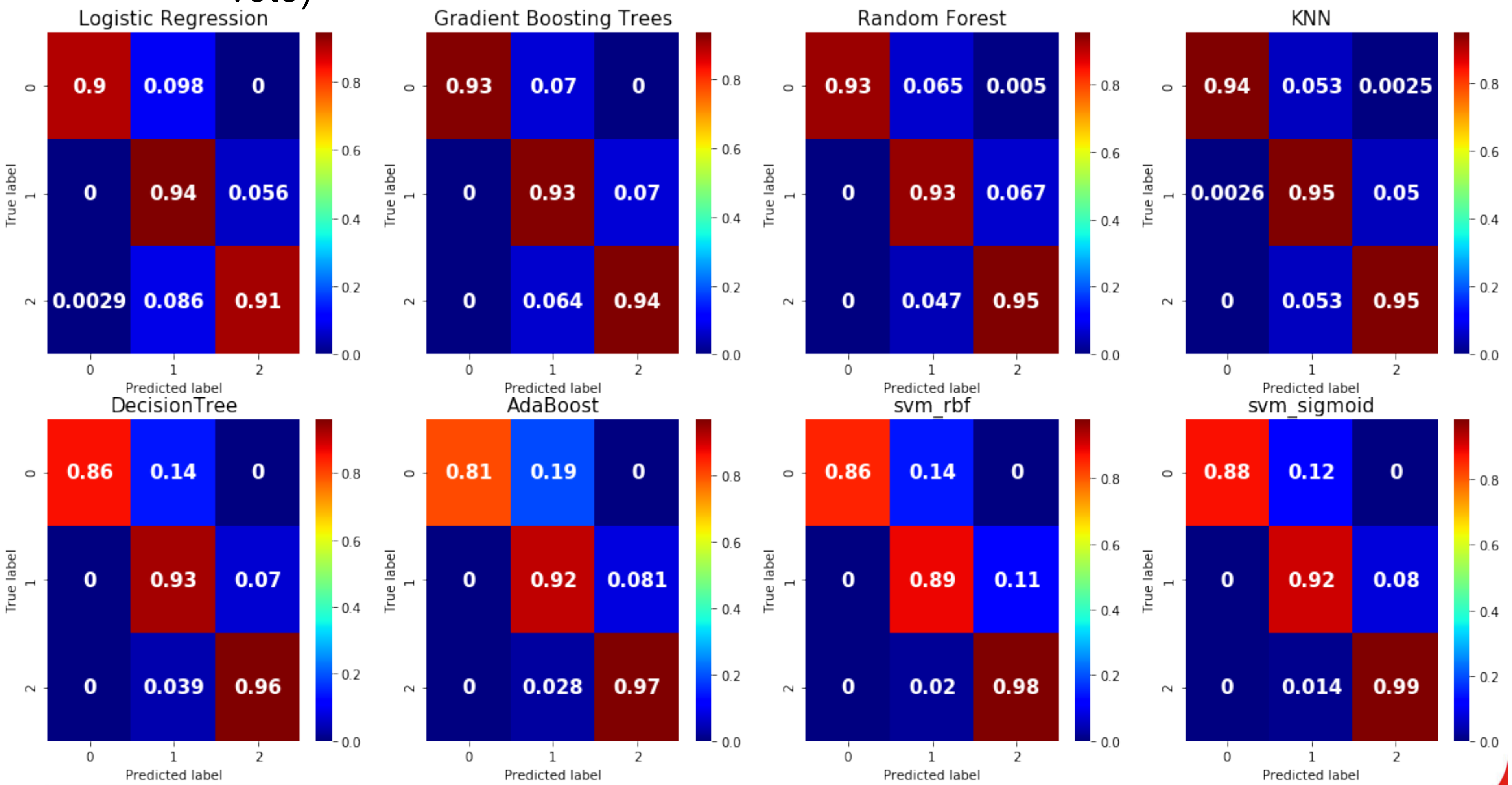


ccRCC **pRCC**

- Red point represent **end branch** of the vascular network, green point represent **junction** of the vascular network.
- This feature **only contain the topological information** of vascular network.
- Then we take the **adjacency matrix** of the graph as **graph features** and put them into GCN (Graph Convolutional Network).

Result 1-2 – Performance on traditional algorithms

2. over_sampling & Leave one out at patient level (majority vote)



Achievements

1. We present the **first work** to investigate the importance of **geometric and topological properties of the vascular network** for **RCCs classification**.

2. We proposed two sets of **hand-crafted features**, **skeleton and lattice features** to represent the vascular network, which is extracted from the vascular network segmentation images.

3. We build new vascular annotated dataset for RCCs histopathological image classification.

4. We build **benchmark results** on the vascular annotated dataset. We compare 3 types of inputs: **proposed features, graph features, and deep features**, showing that **our proposed** features based traditional classifier can provide **best results at small datasets**, and **can classify ccRCC and pRCC robustly**.

Problem & Motivation

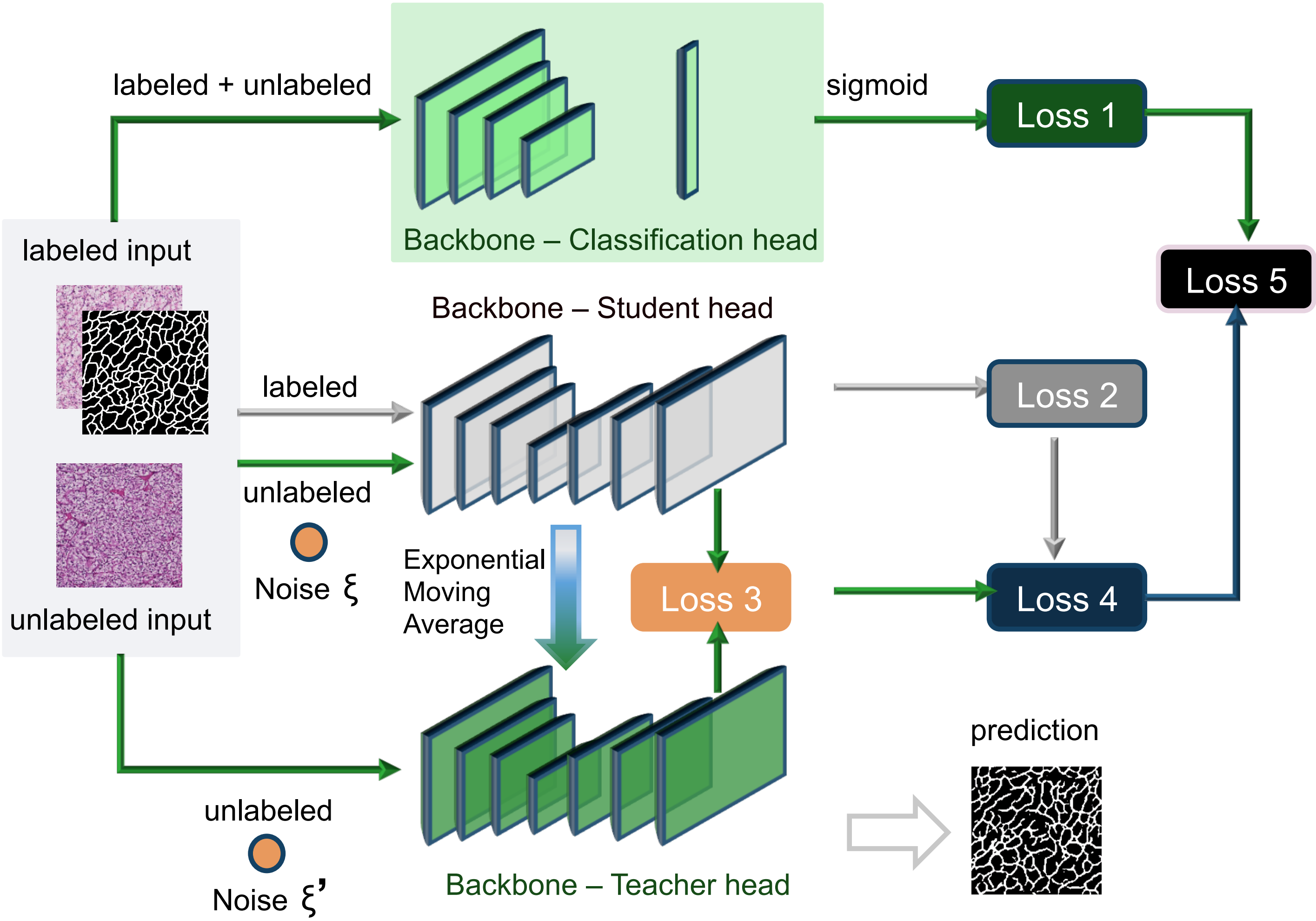
- Our **former work** has clarified the importance of the vascular network in discriminating RCC subtypes, but their application has been **limited to some extent due to manually vascular segmentation**.

- **Supervised learning relies on labels**, which is a challenge for vascular segmentation task from histopathological images due to **lacking ground truth.**

- This encouraged us to find out if it was possible to improve the vascular network segmentation performance **using unlabeled datasets**. This is indeed the paradigm of semi-supervised learning (**SSL**) models.

- Compared with the **difficulty** of obtaining manually a vascular network **mask for the segmentation task**, the **labeling for the classification** task is **easy** to obtain. We conjectured that **joint supervised classification and SSL for vascular network segmentation**, both embedded in a multi-task learning (**MTL**) model, may improve the performance of vascular network segmentation in RCC histopathological images.

MTL-SSL model structure



Loss1
Classification loss:
CrossEntropyLoss

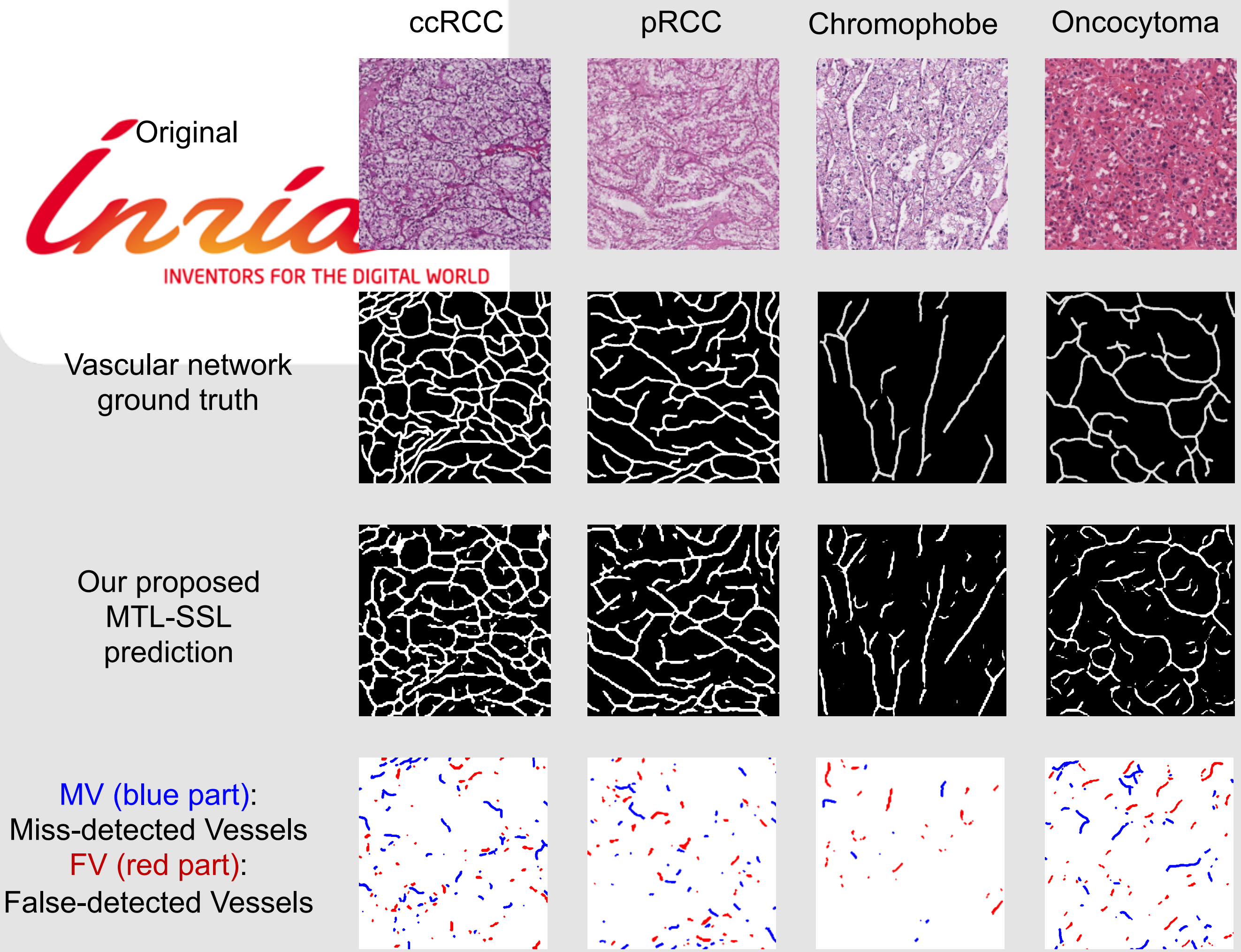
Loss2
Supervised loss:
0.5 * (DiceLoss + CrossEntropyLoss)

Loss3
ConsistencyLoss

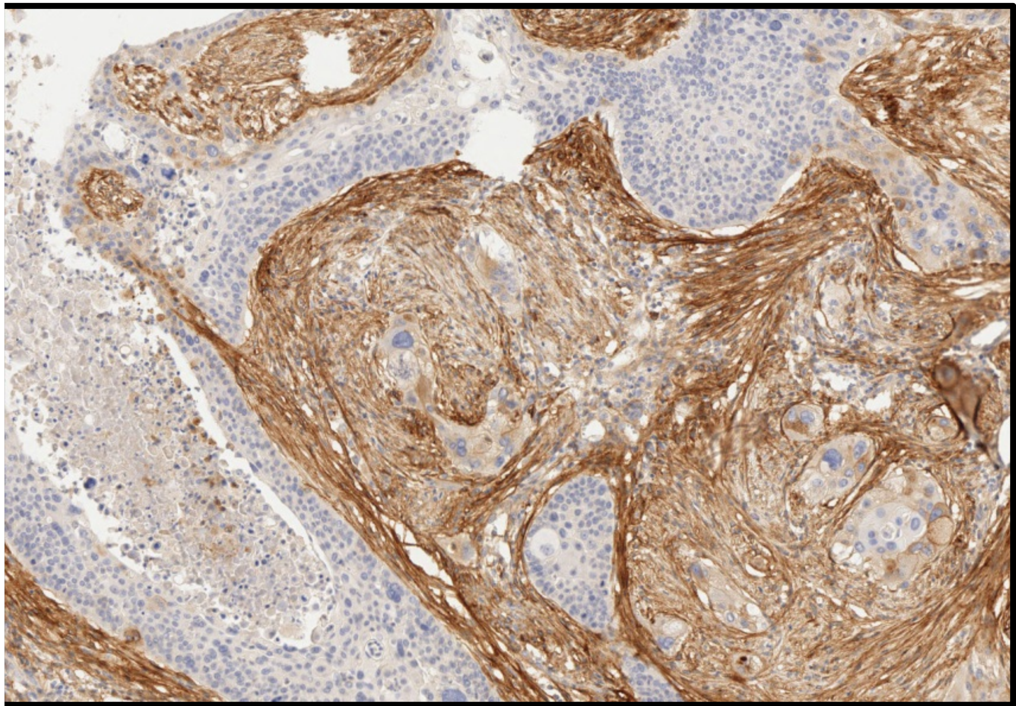
Loss4
SSL loss:
Supervised loss + consistency_weight * ConsistencyLoss

Loss5
Total loss:
Classification loss * classification weight + SSL loss * SSL weight

Result 1 – Segmentation results of our model



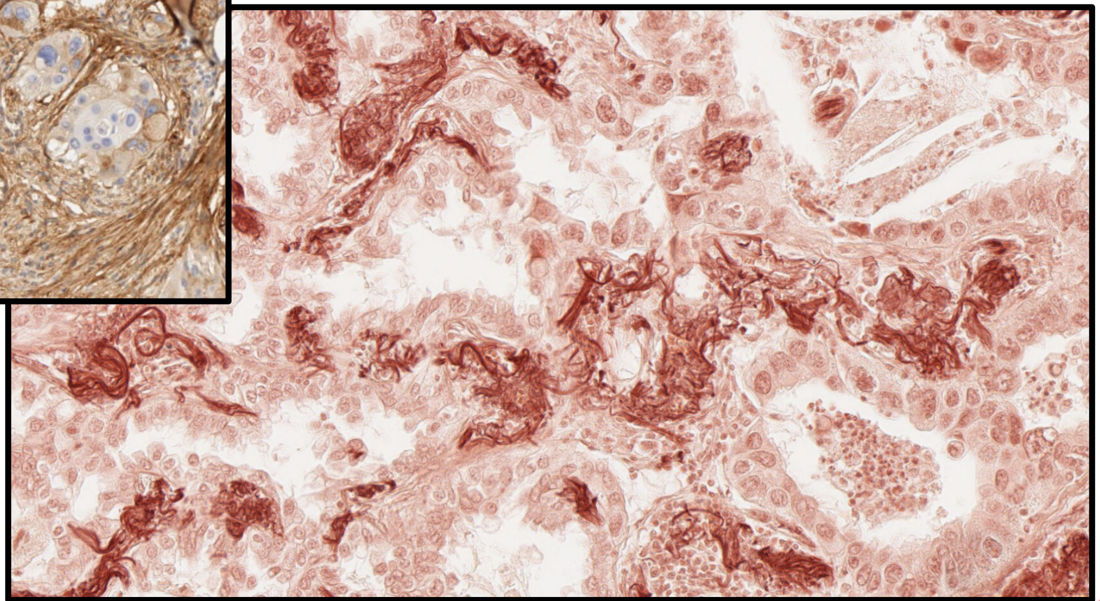
Next steps : other colorations



Immunostaining

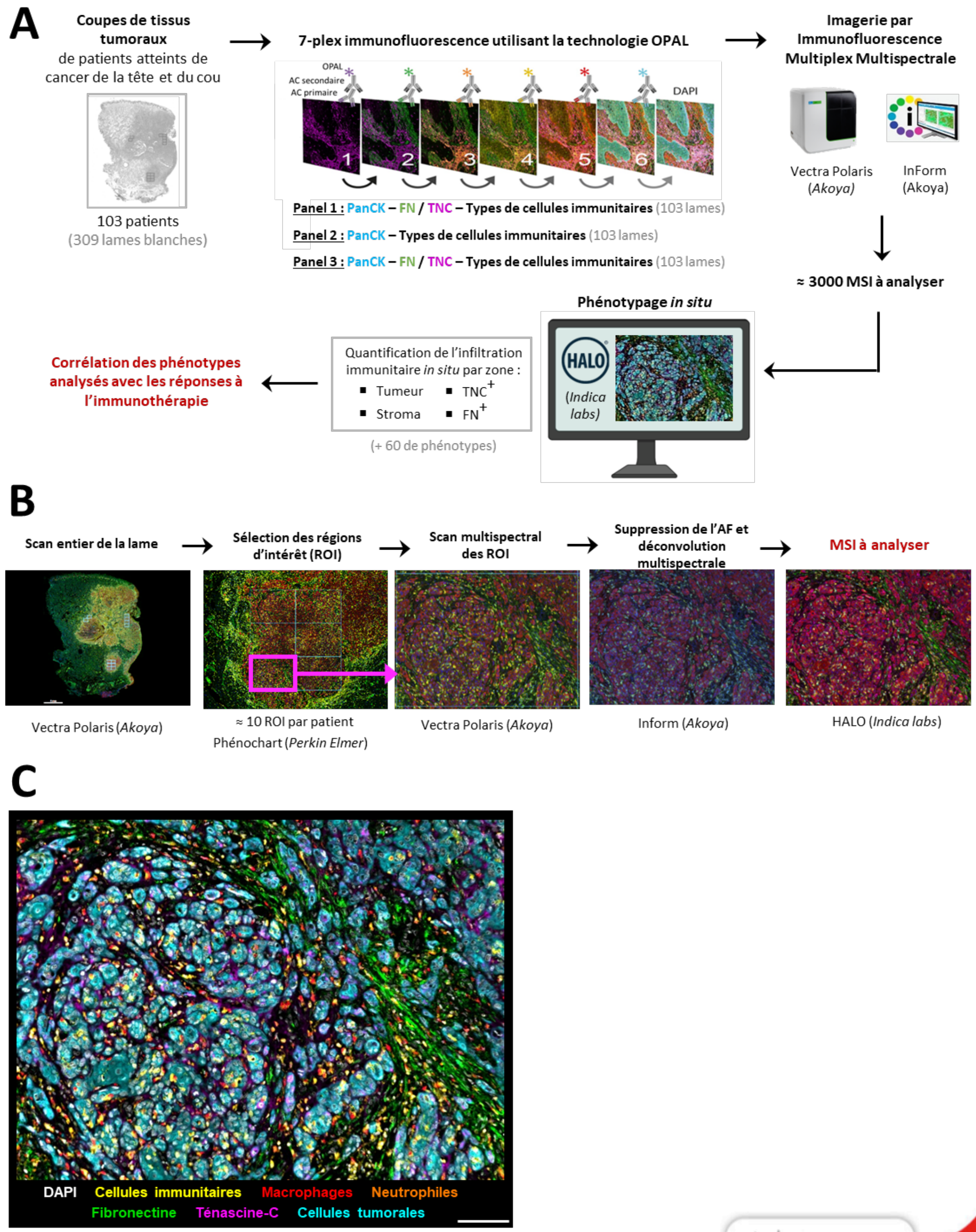
Orceine

…

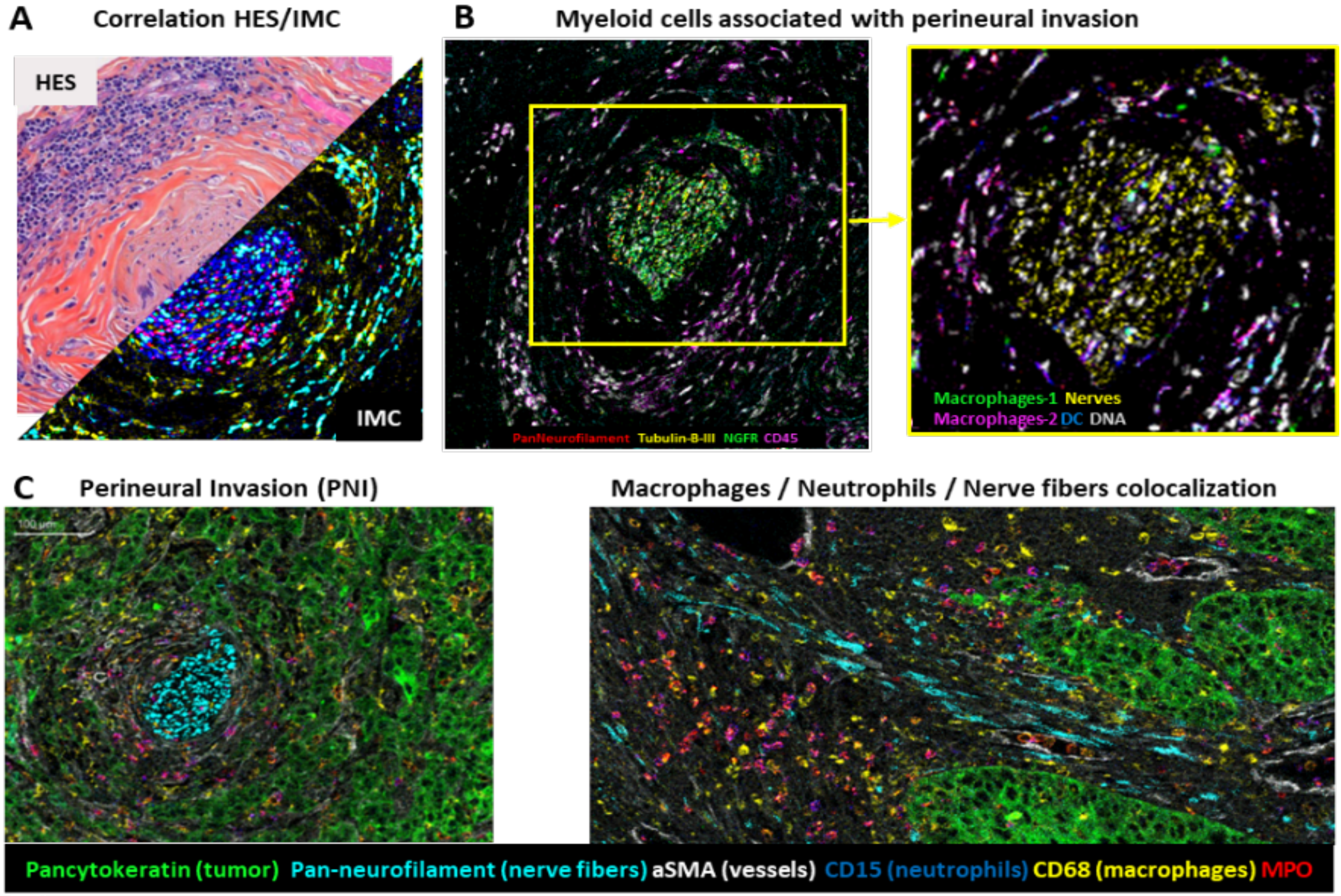


Next steps :

multispectral scanner



Next steps : mass cytometry



Conclusion : computation histopathology

- Can help for diagnoses by
  - providing quantitative parameters
  - by confirming or not first diagnose
  - by defining ROI
- Can provide a medical research tool
  - by exibiting features of interests
  - by detecting specific local/global patterns