Deep learning for image analysis in biology

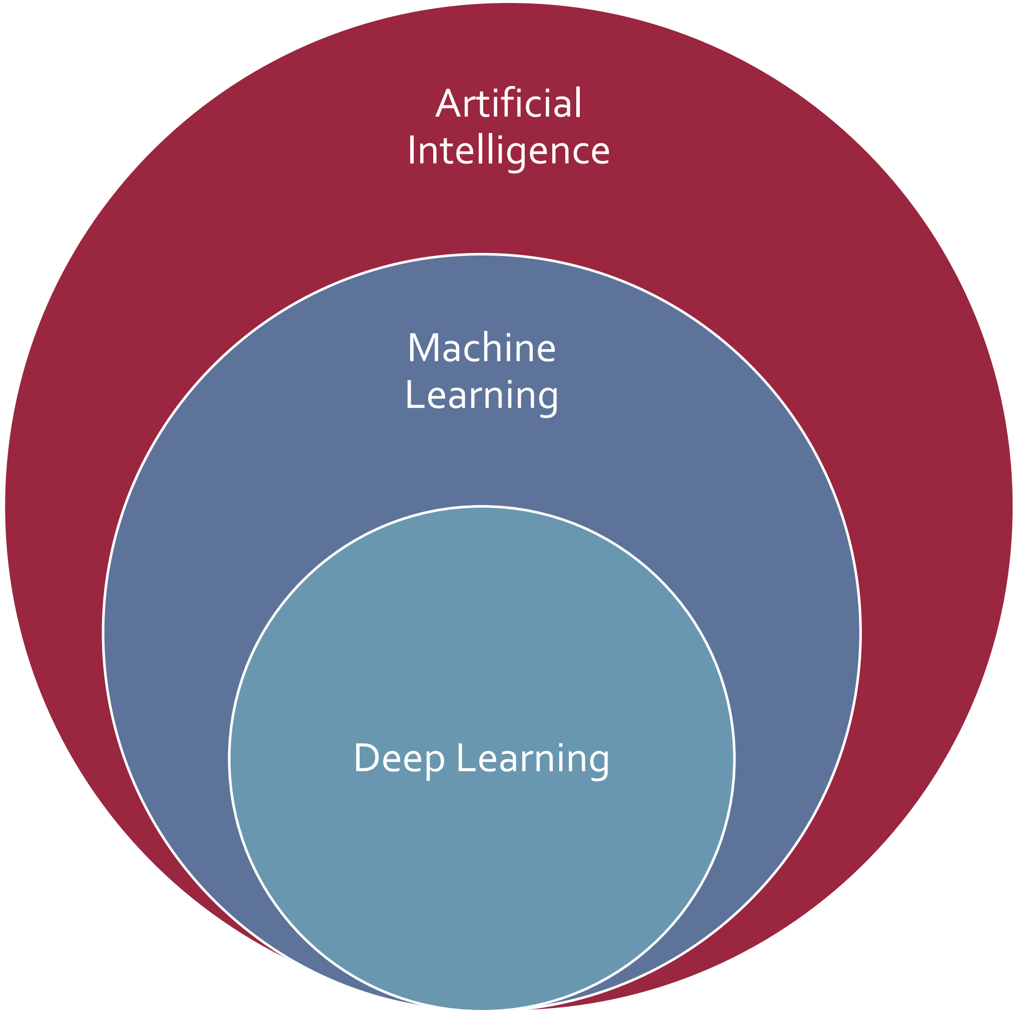
Maëlle Guillout - 02/03/2023

I. Artificial Intelligence, Deep Learning & Machine Learning

What is the difference?

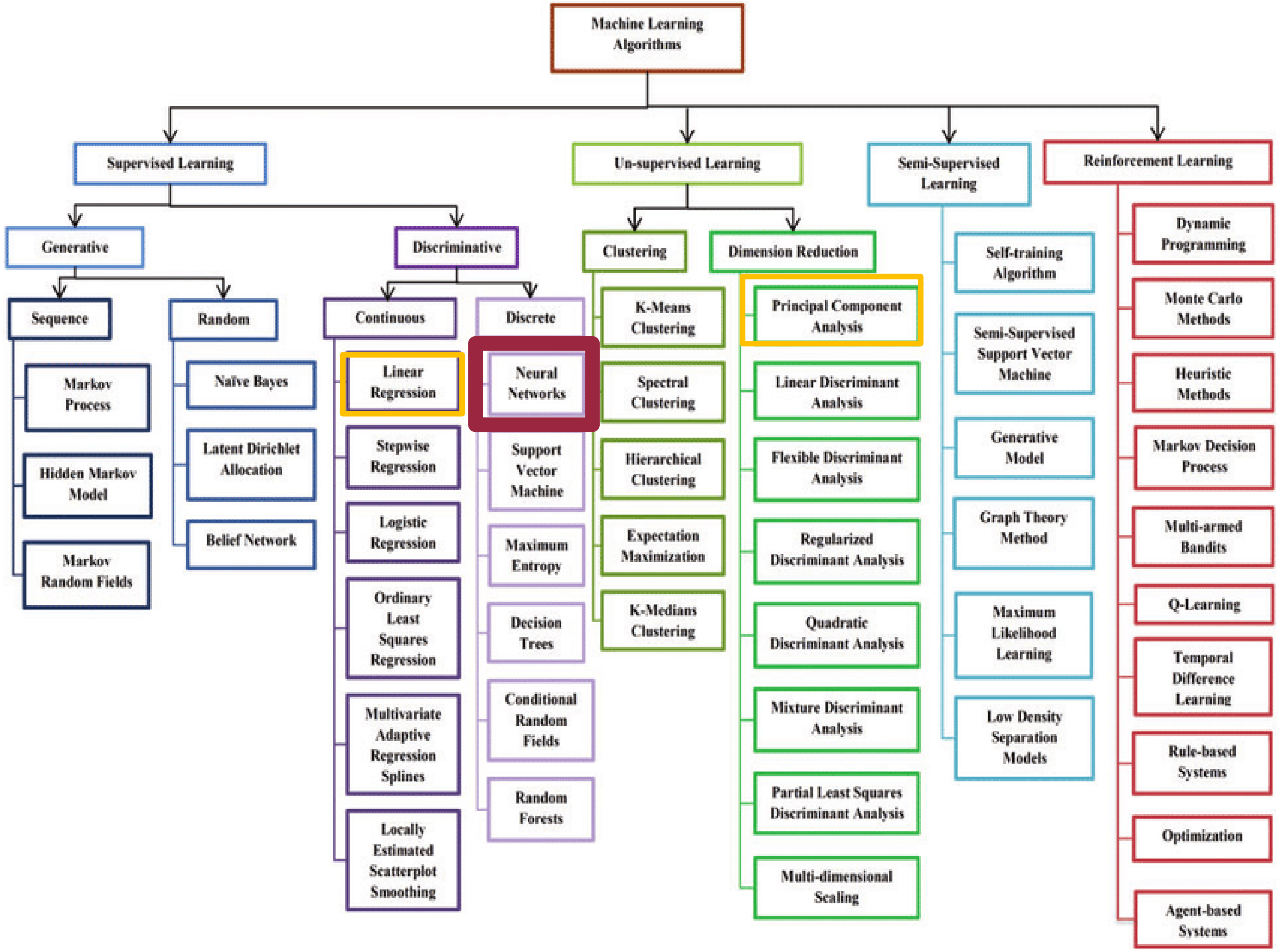
Definitions

- **Deep Learning is a sub-domain of Machine Learning which itself is a field of Artificial Intelligence**

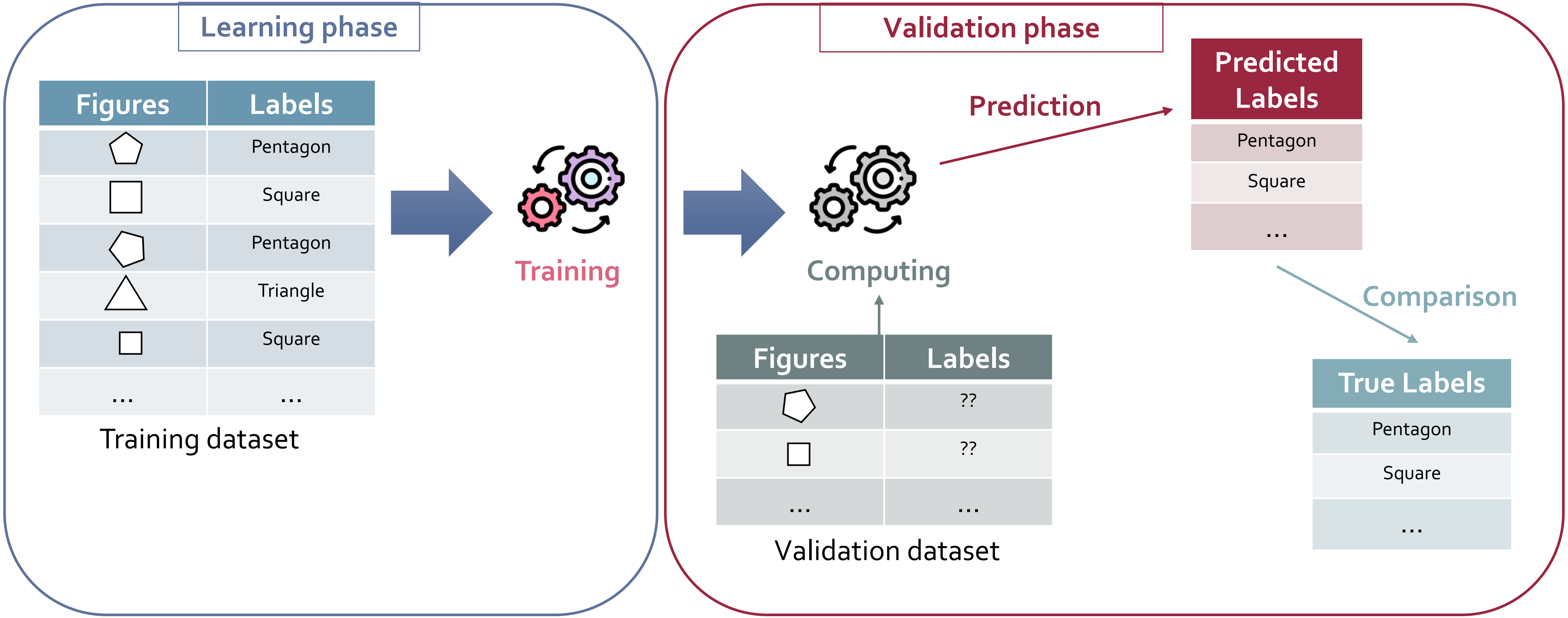


Definitions

- Deep Learning is a sub-domain of Machine Learning which itself is a field of Artificial Intelligence
- Machine Learning includes several statistical methods
- Deep learning methods are more complex and « autonomous »
- For image analysis➜ Neural Networks



Supervised Deep Learning



Supervised Learning: Main Issues

Amount of annotated data

Large dataset needed

Heterogeneity

Training dataset must represent the diversity of data

Annotation quality

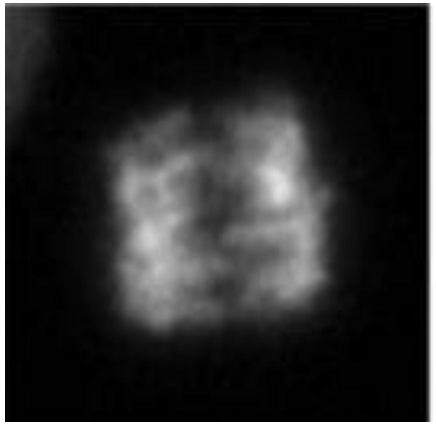
- Possible mistakes
- Possible bias

Deep learning for image analysis

Classification

Determines the class of each image

Is it a dog or a cat?

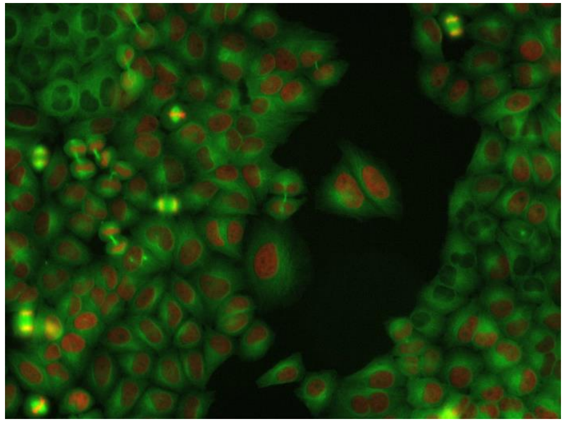


Which phase of the mitosis is it?

Segmentation

Determines the class **of each pixel** from each image

Which pixels belong to the class "dog" ?



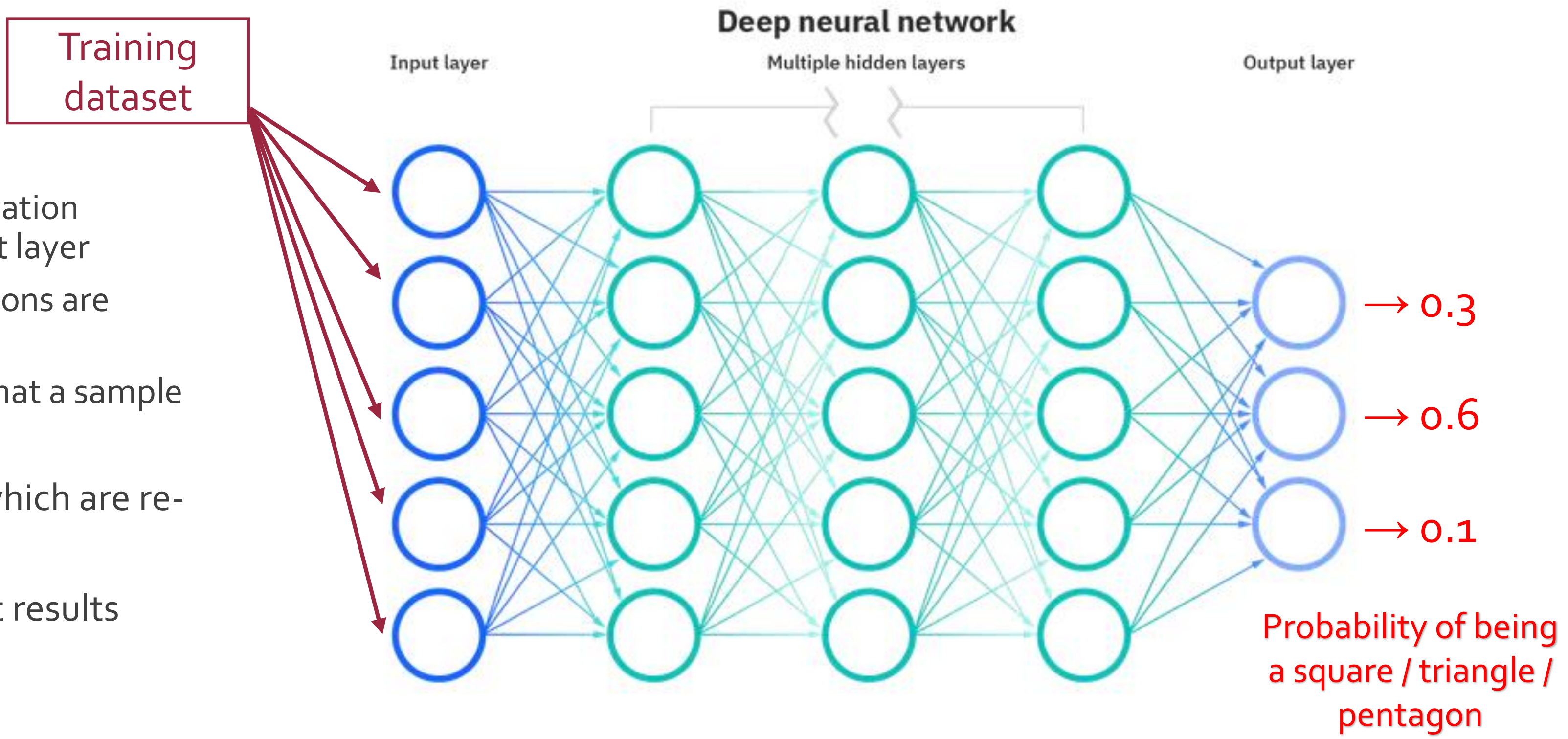
Where are the cells? How many are them?

Balluet et al., Journal of Microscopy, 2022
https://www.mocomakers.com/cats-versus-dogs/
Held et al., Nature Methods, 2010

Neural Networks

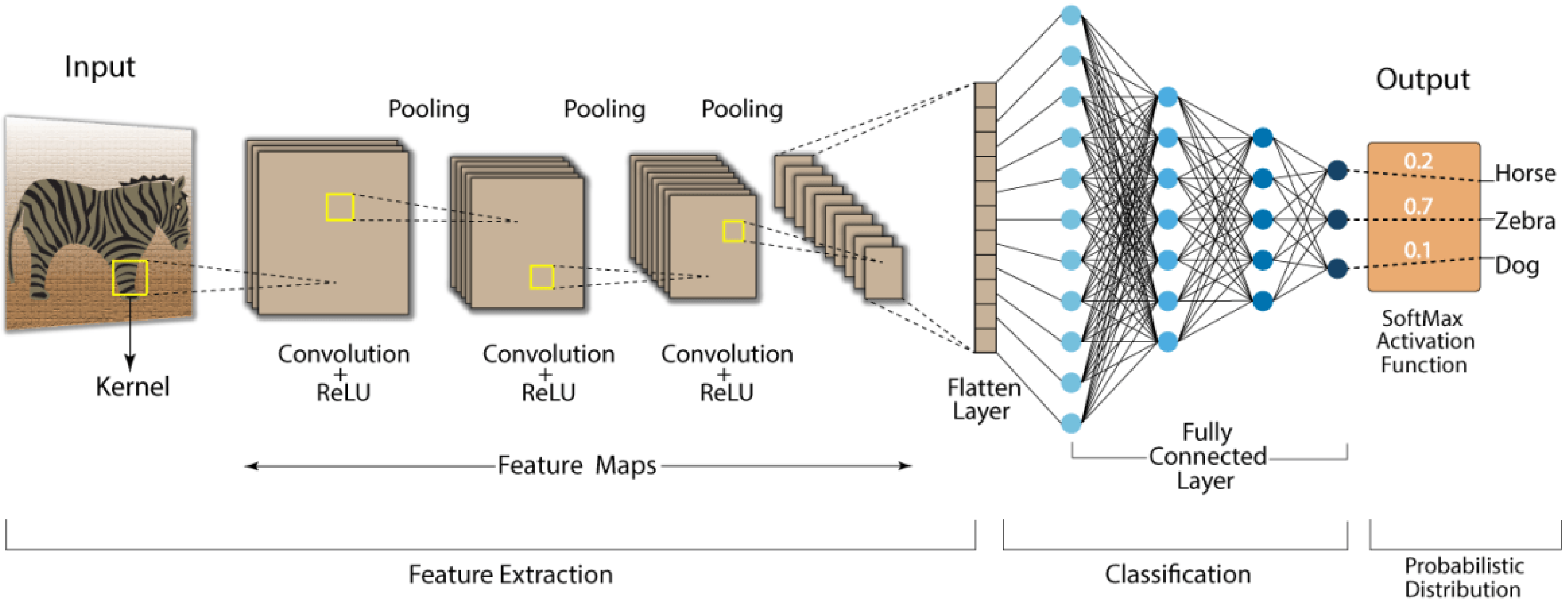
Neural networks are inspired from brain

- Neurons which treat information
- Organized in several layers
    - First: receive data and computes an activation function and transmits results to the next layer
    - According to previous results, some neurons are activated or not
    - Last: give the output (e.g. a probability that a sample belongs to each class)
- Layers are connected with different weights, which are re-calculated for each iteration
- Goal: minimize a cost function to have the best results

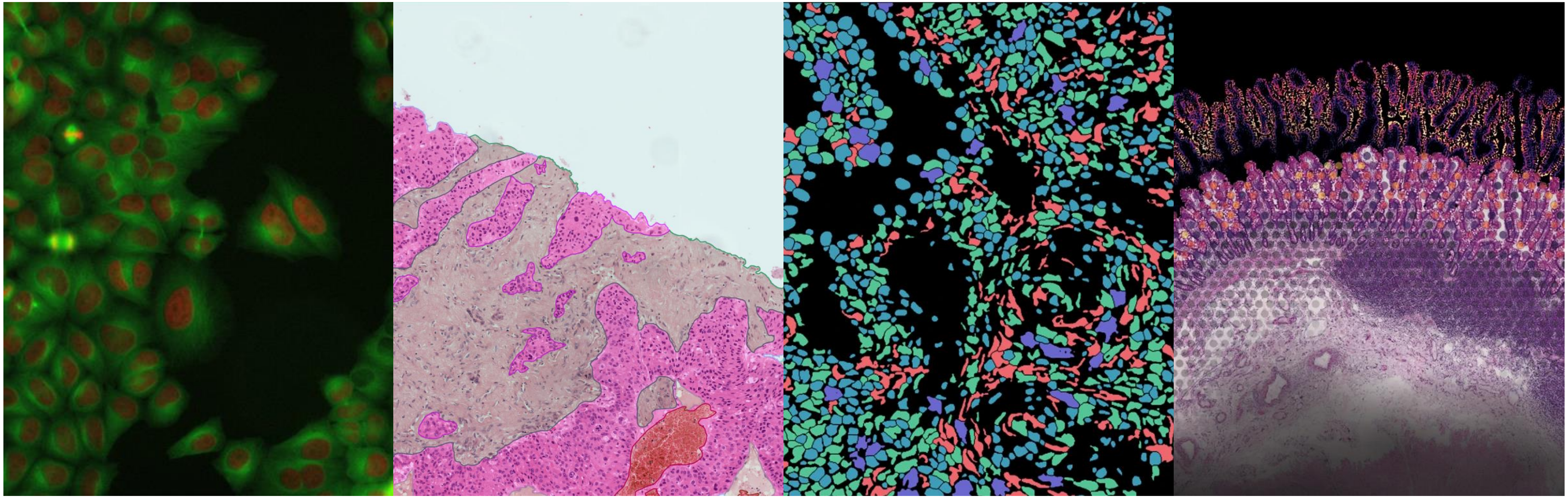


Convolutional Neural Networks (CNN)

A type of neural networks mainly used for image analysis



www.developersbreach.com

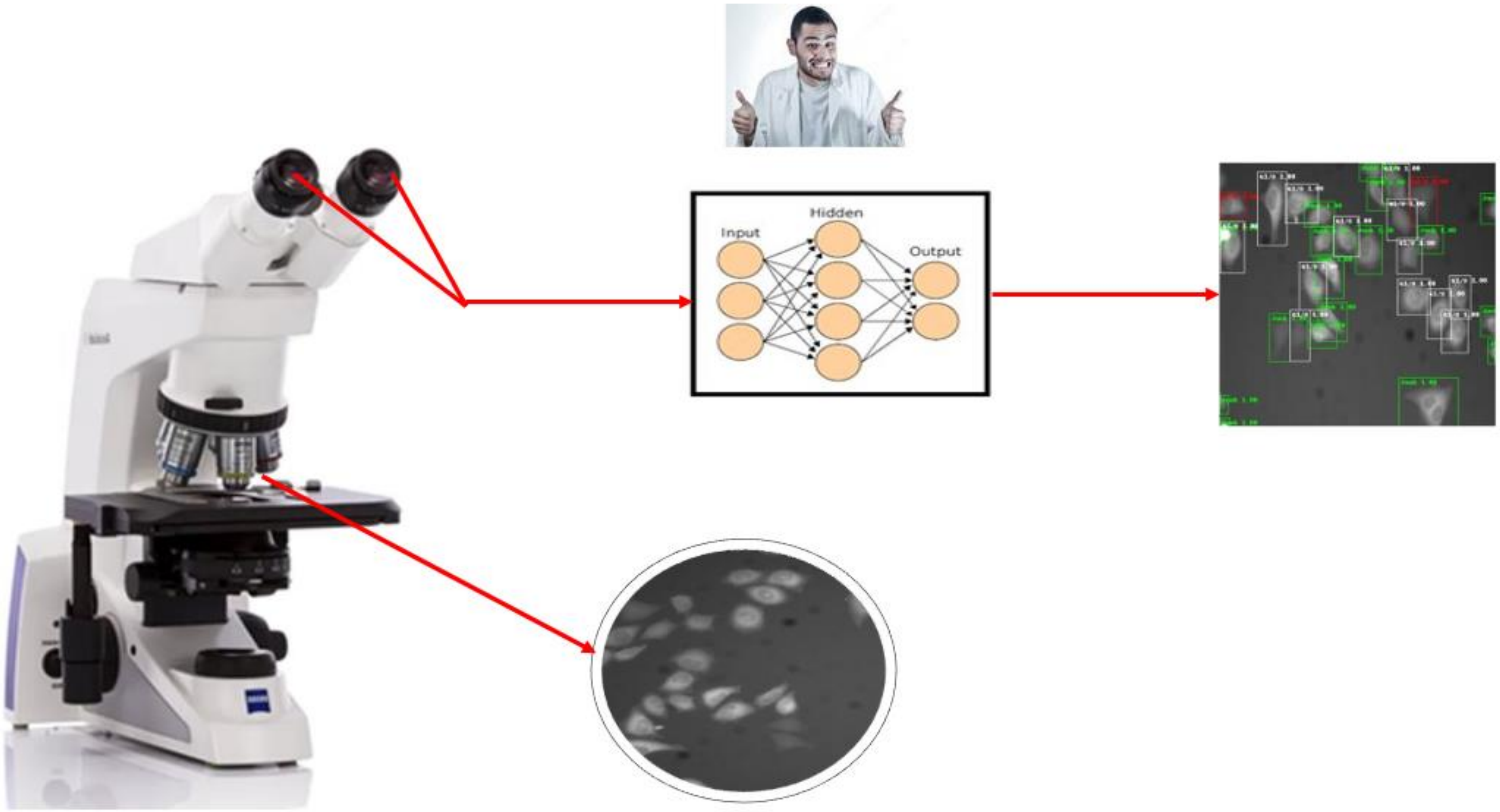


II. Applications of Deep Learning for image analysis in biology

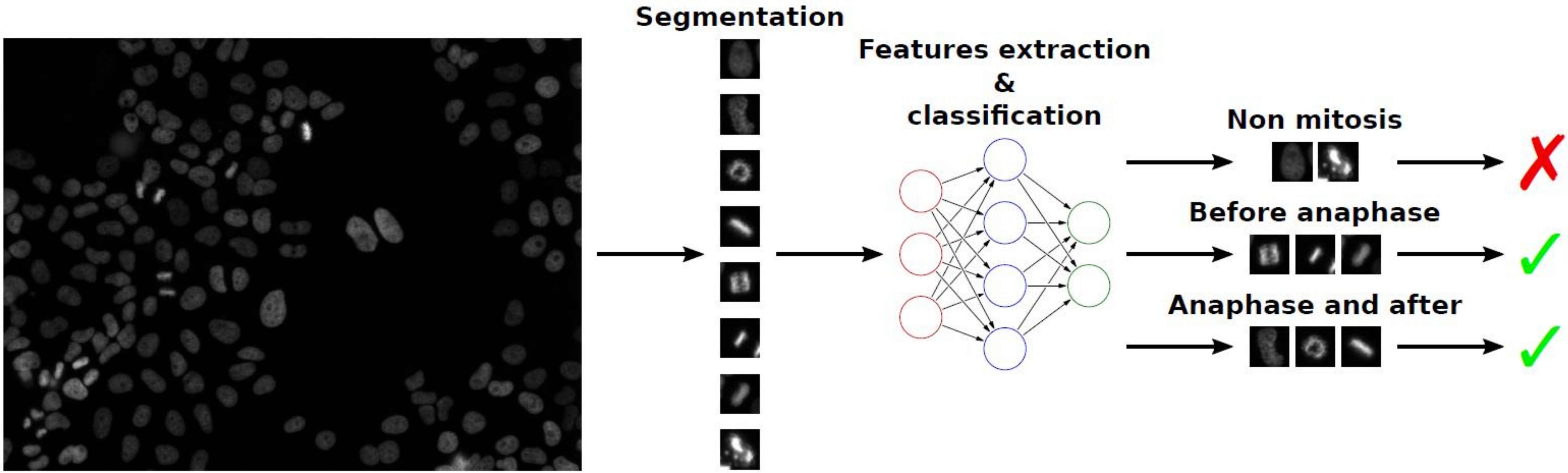
II.1 Deep Learning for smart microscopy

Smart Autonomous Microscopy

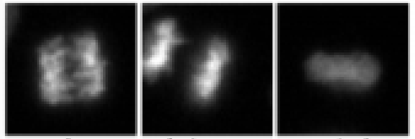
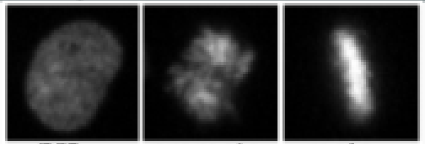
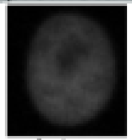
INSCOPER igdr



Detection of cells + Capture of rare events



Training Dataset for Classification: Cell Cognition

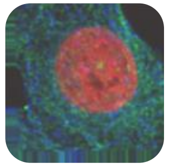
img	Class		Number of images	
 earlyana lateana telo	AA (After Anaphase)	earlyana	28 (Train: 20; Test: 8)	112 (Train: 80; Test: 32)
		lateana	42 (Train: 30; Test: 12)	
		telophase	42 (Train: 30; Test: 12)	
 pro prometa meta	BA (Before Anaphase)	prophase	38 (Train: 26; Test: 12)	112 (Train: 80; Test: 32)
		prometaphase	38 (Train: 26; Test: 12)	
		metaphase	36 (Train: 27; Test: 9)	
 inter	I (Interphase)		112 (Train: 80; Test: 32)	
 apo	J (Junk)		112 (Train: 80; Test: 32)	
	Total		448 (Train: 320; Test: 128)	

Held et al., Nature Methods, 2010

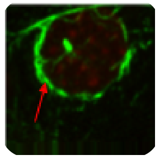
Generalization issue

Could the same classification model be used to classify:

Cells from various origins?



Cells with different markers?

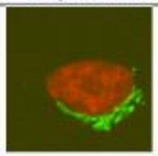
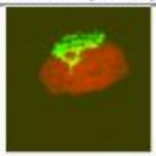
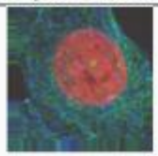
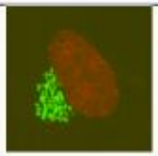
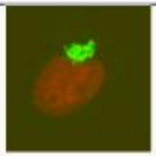
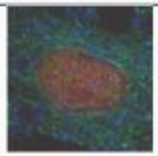


According to another biological question?

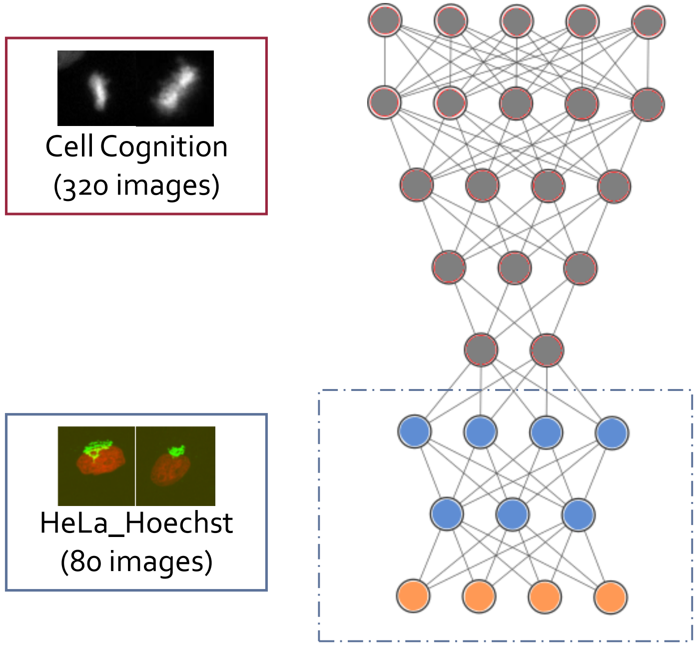
Transfer Learning

Retrain "lightly" the model → Small dataset with already labeled images required → How many images are required?

Table 1: Datasets from *Nagao et al* study. [1]

Dataset name	*RPE1_Hoechst_CENPF*	*HeLa_Hoechst_GM130*	HeLa_Hoechst_EB1
Cell line	Mouse retinal pigment epithelium	human cells (HeLa)	human cells (HeLa)
Markers	Hoechst (DNA) + CENP-F (centromeres)	Hoechst (DNA) + GM130 (Golgi)	Hoechst (DNA) + EB1 (microtubules)
Example of a cell in G2 phase			
Example of a cell not in G2 phase			

[1] Nagao, Y et al. Mol Biol Cell 31, 1346–1354 (2020)



II.2 Deep learning for histopathology

Why using deep learning for fibrosis detection?

To estimate proportion of fibrosis on a tissue

Faster analysis

+

Reduce bias

Reproducible research

25%

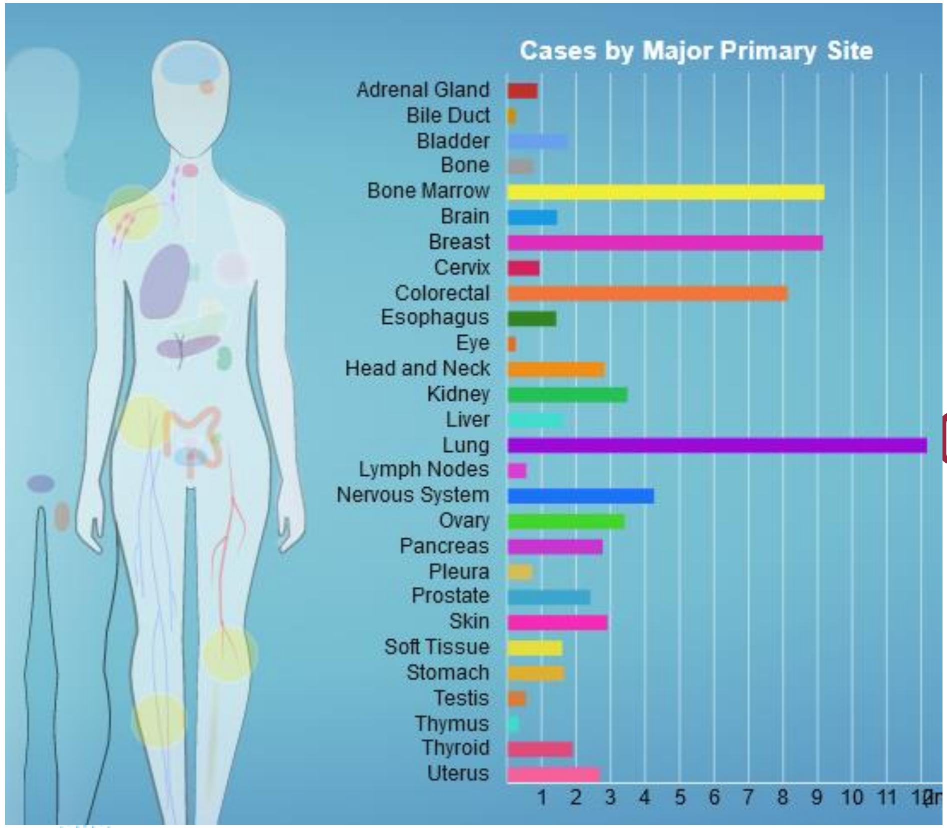
19.1%

Combining fibrosis analysis with omics analysis

Training Dataset

Images come from a databank build by the American National Cancer Institute

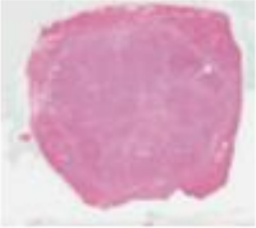
Cancer Genome Atlas (TCGA) contains data from 20,000 primary cancer (33 types)



Data Category	Cases (n=377)	Files (n=18 243)
Sequencing Reads	377	2 608
Transcriptome Profiling	376	1 698
Simple Nucleotide Variation	375	5 256
Copy Number Variation	377	3 074
DNA Methylation	377	1 290
Clinical	377	803
Biospecimen	377	1 634

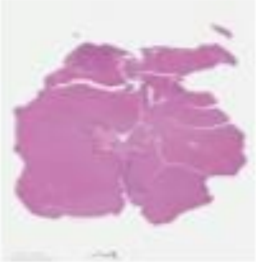
Downloaded histological data from 61 patients

- 135 svs slides
- Only 13 used for training
- Colored with H&E



TCGA-DD-A4NK-01Z-00-DX1.4FC242C7-5026-4400-8D8D-7BE954A6DE1E.svs - ScanScope image

TCGA-DD-A11D-01Z-00-DX1.3D607AE3-3910-406F-8EBA-3C1CDA0D34A6.svs - ScanScope image

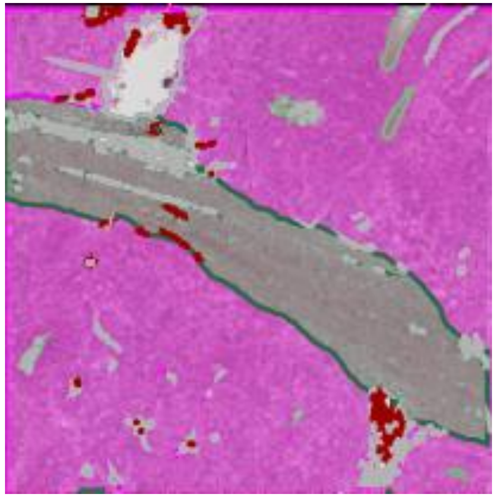


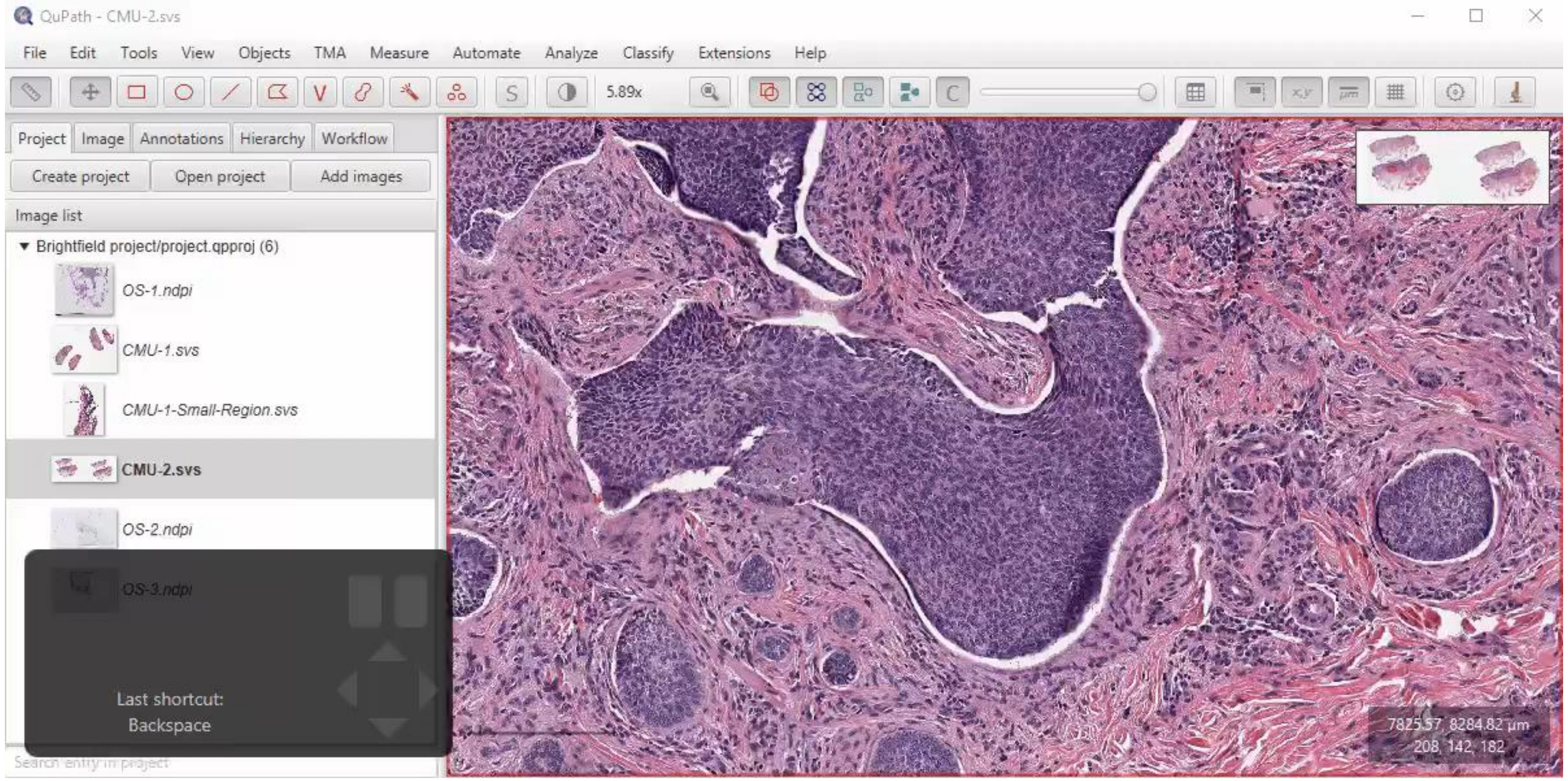
TCGA-DD-A73D-01Z-00-

QuPath

An free open source software for digital pathology image analysis

- Whole tissue visualisation
- TMA Analysis
- Stain estimation
- Automatic cell detection
- Automatic measurements and statistics
- Image annotation

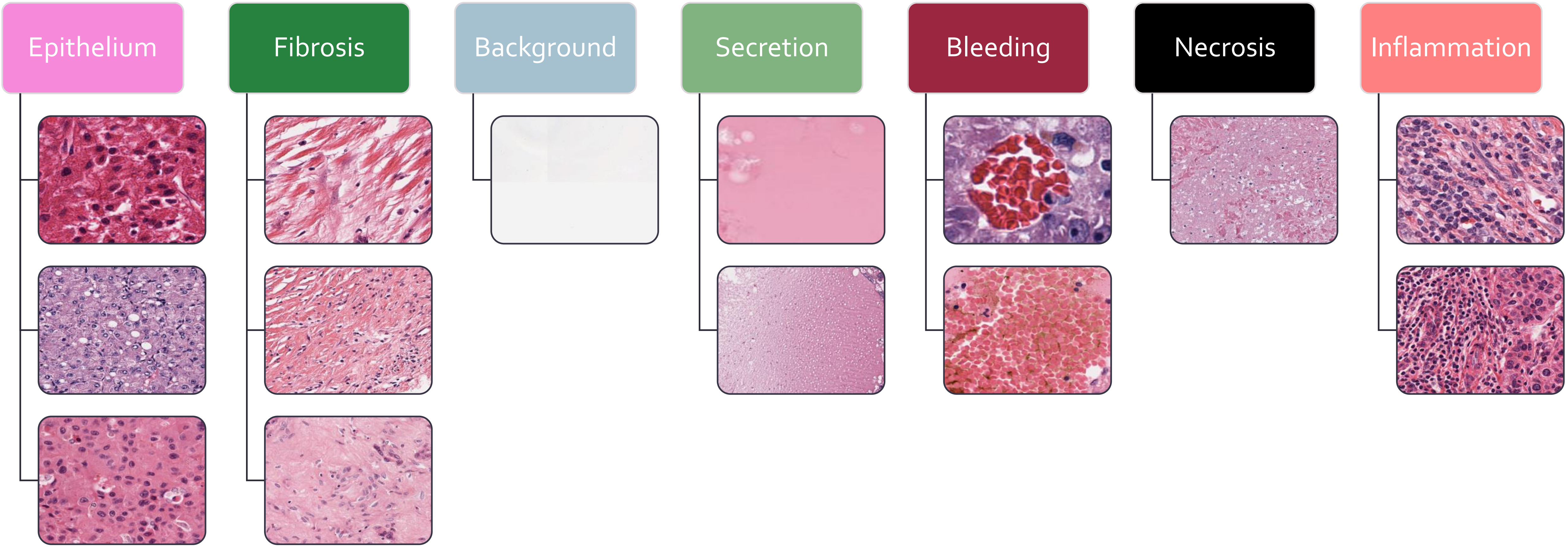




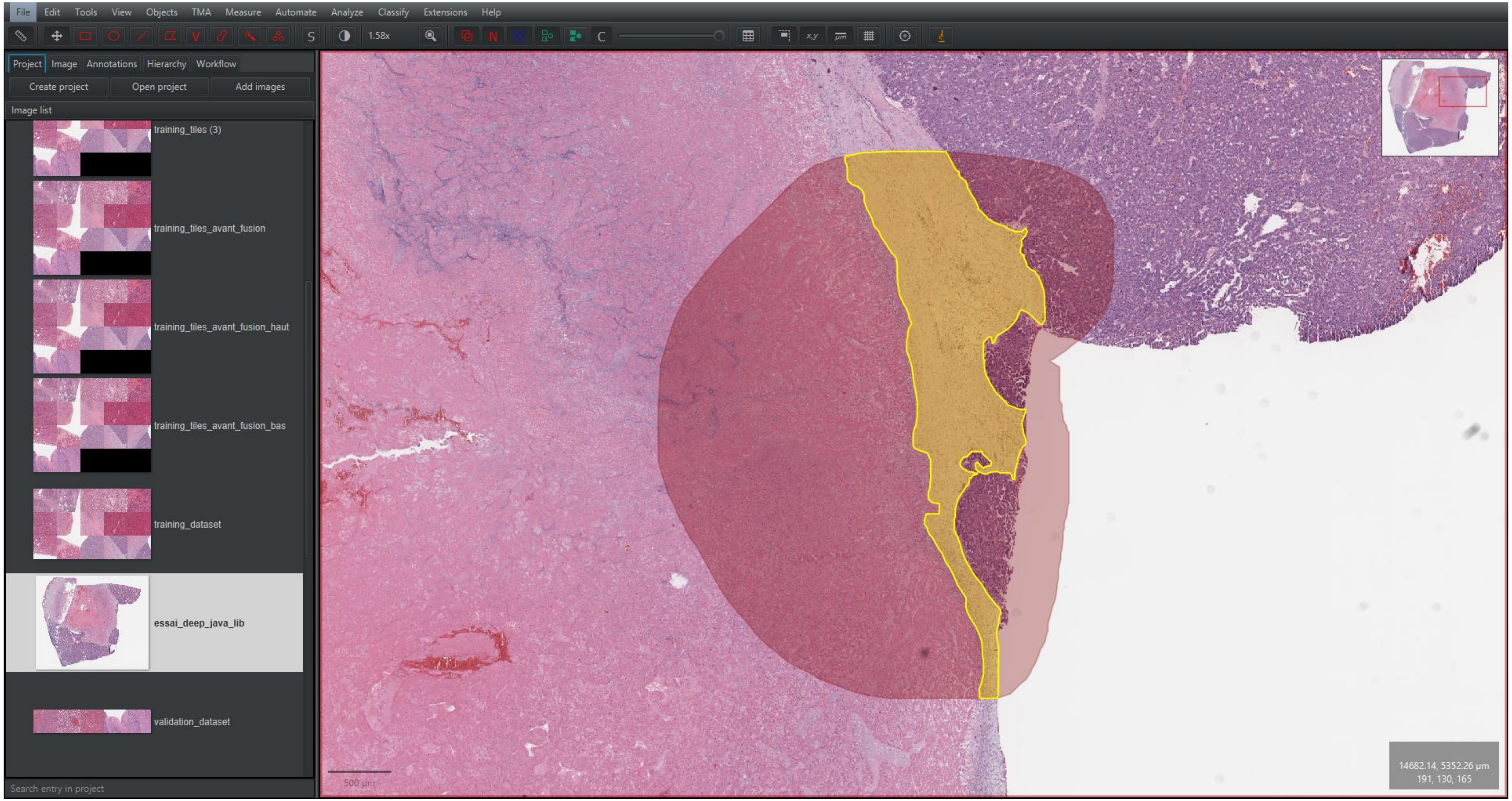
Bankhead et al., 2017

Image Annotation with Qupath

Classes used for segmentation



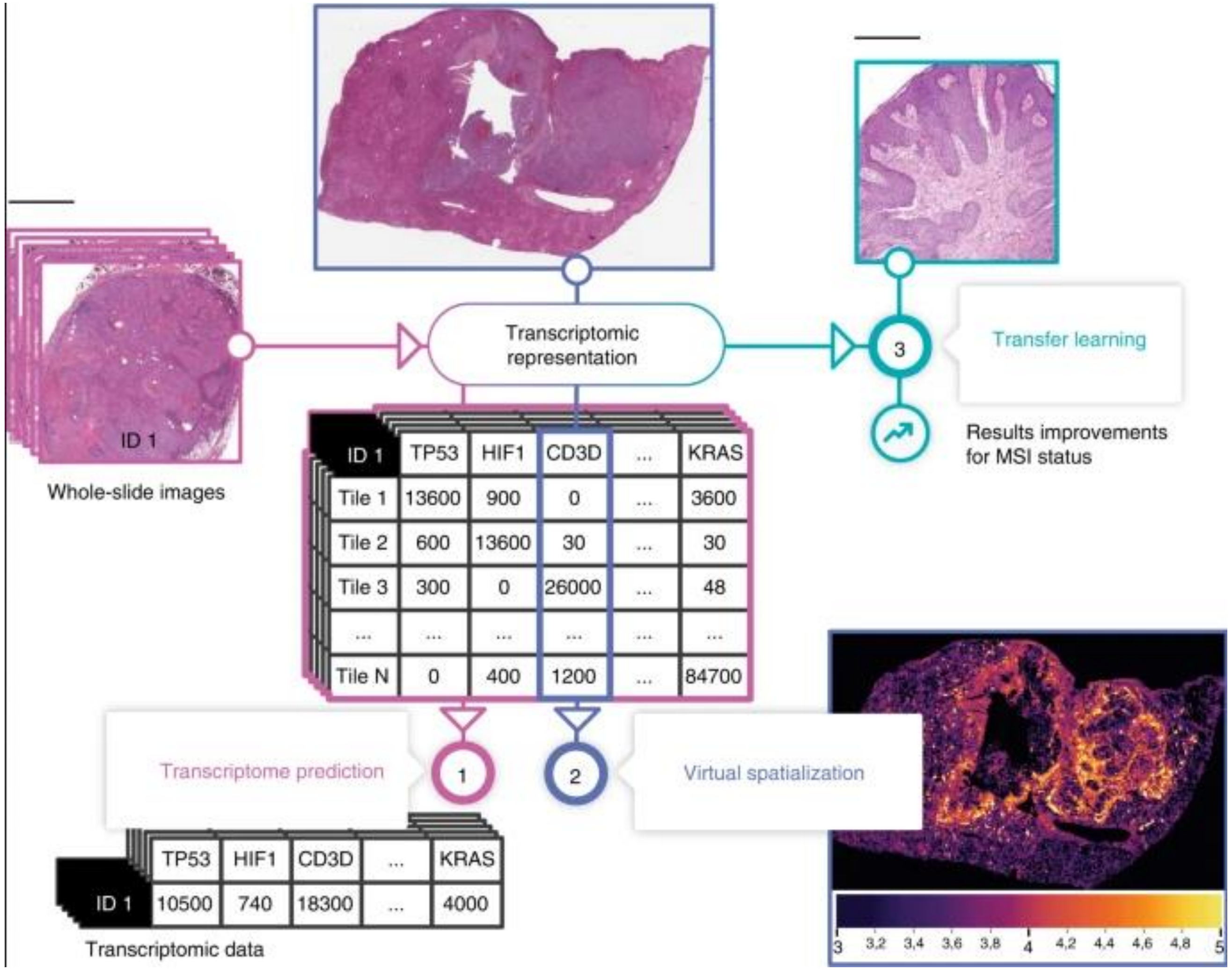
Model Integration in QuPath



Article | Open Access | Published: 03 August 2020

A deep learning model to predict RNA-Seq expression of tumours from whole slide images

Combination of Histology and Transcriptomics



Schmauch et al., 2020

Diagnostic

Quand l'intelligence artificielle permet d'identifier l'origine inconnue d'un cancer métastasé

Un jeune homme de 30 ans présentant un cancer métastasé d'origine inconnue a été le premier à tester un outil d'intelligence artificielle développé à l'Institut Curie. Le crible a permis d'identifier le rein comme l'organe présentant la tumeur d'origine et le traitement spécifique qui lui a permis de guérir.

Janvier 2020. Le cas d'un jeune homme de 30 ans présentant *« des métastases un peu partout »*, se souvient Sarah Watson, est confié au laboratoire de cette biologiste et oncologue de l'Institut Curie. *« Nous nous attendions à un diagnostic de sarcome* [cancer rare des tissus mous ou de l'os], *qui est ma spécialité. Mais la biopsie nous montre que c'est un carcinome* [cancer d'un tissu épithélial], explique-t-elle. *Nous ne savions pas quel était le primitif* [premier organe touché]. »

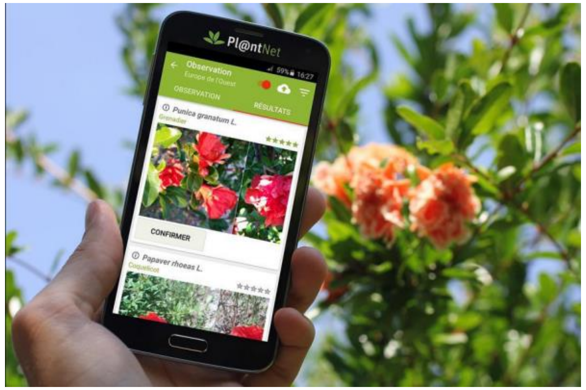
Article Le Monde, janvier 2023

Modèle entraîné sur la correlation données RNA-seq / histologiques → Arrive à prédire le profil transcriptomique à partir d'une image de lame en quelques minutes → Plus rapide que de sequencer l'ARN du patient et faire des analyses bioinformatiques

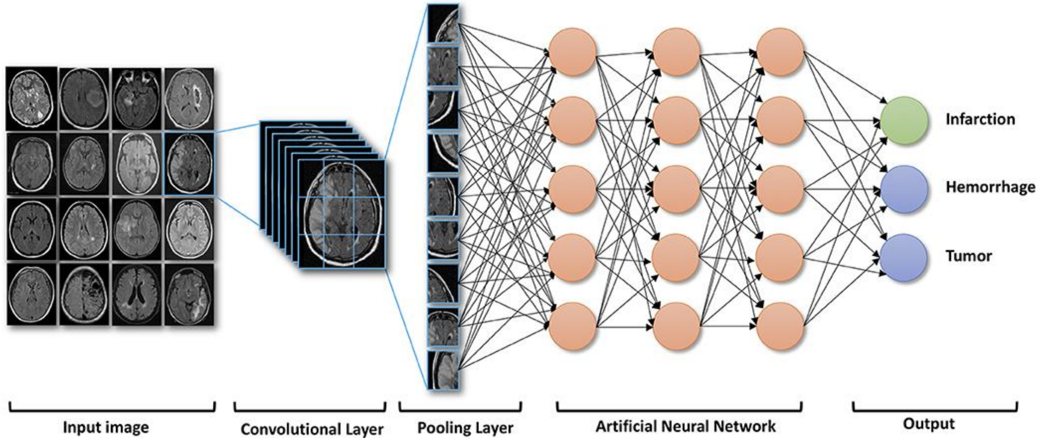
II.3 Other applications

Other applications in biology

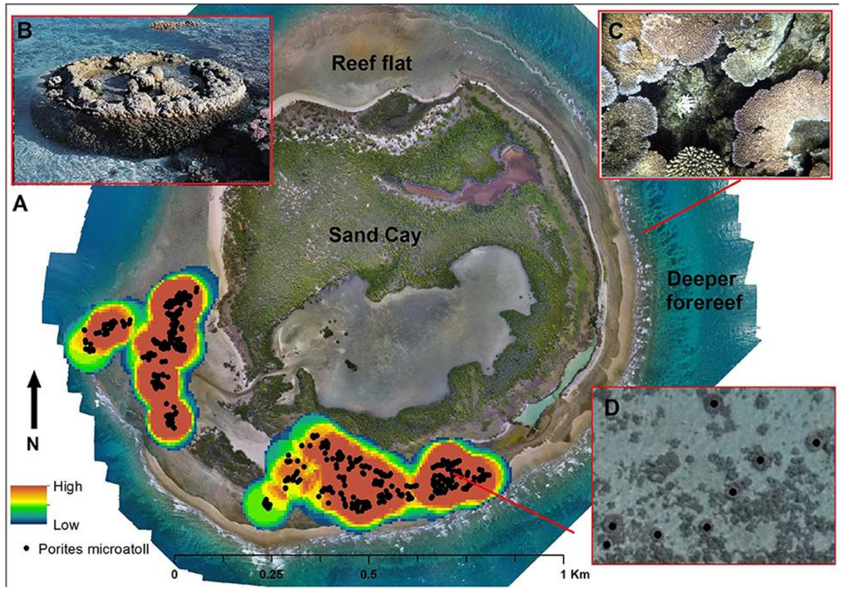
- Medical Imaging
- Animal Behaviour
- Plants recognition
- Ecology / Agronomy / environment



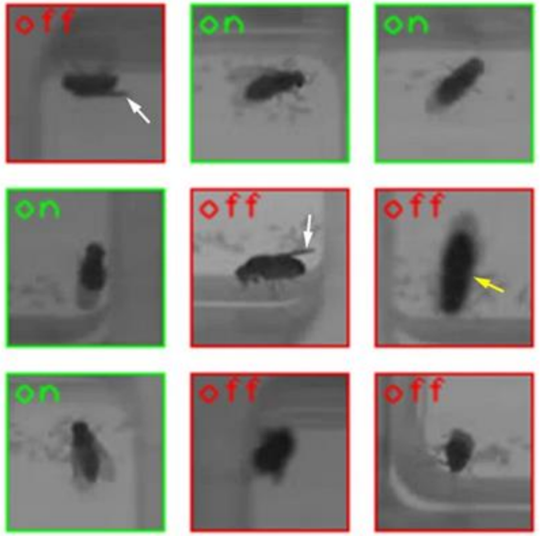
Pl@ntNet



G. Zaharchuk et al.



Hamylton et al.



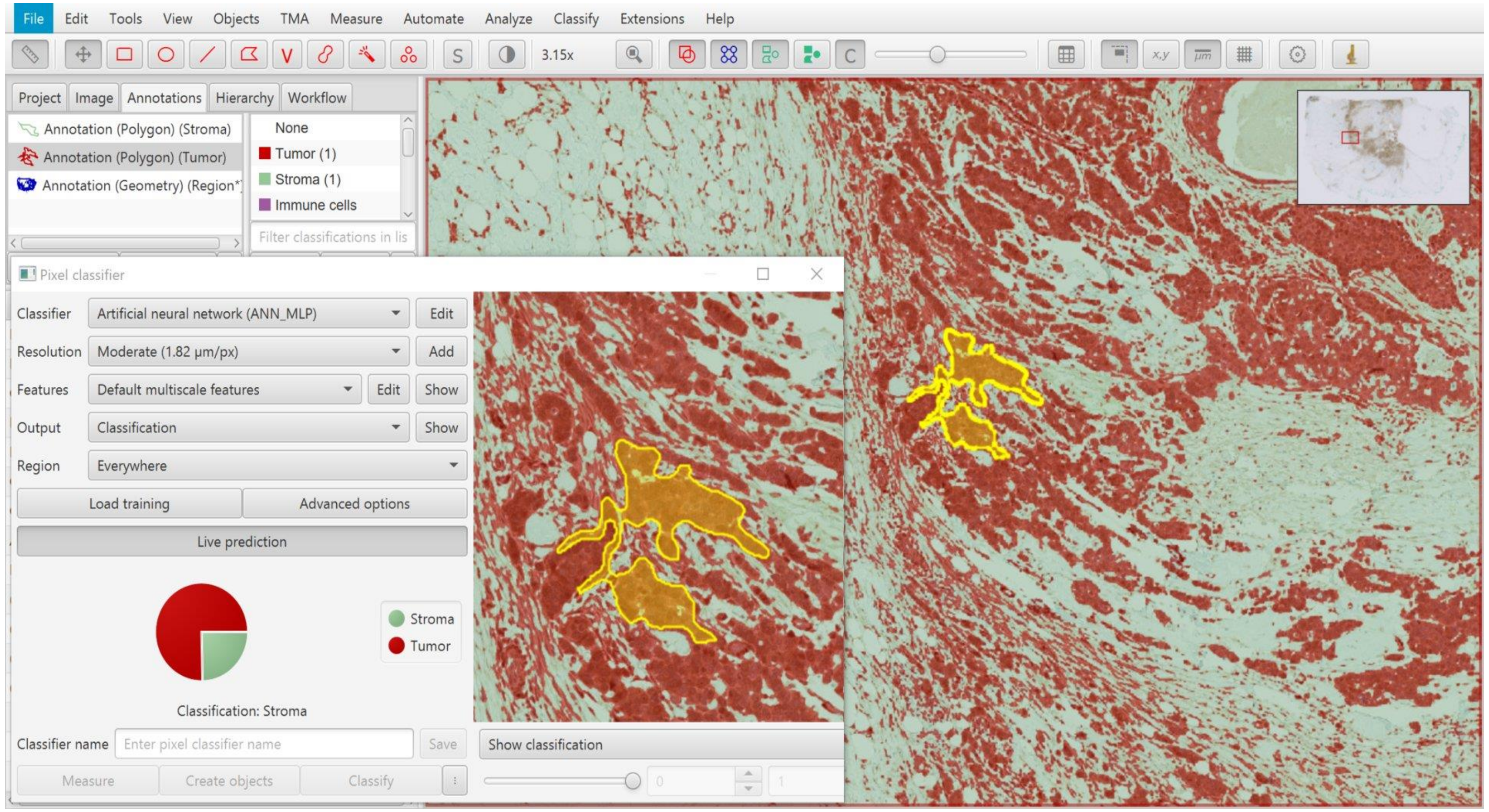
Stern et al.

Neural networks can be used to analyse other data than images: genomics (and other –omics), structure prediction for proteins, biomarkers identification …

III. How to use deep learning on images?

Without coding?

How to use deep learning models?



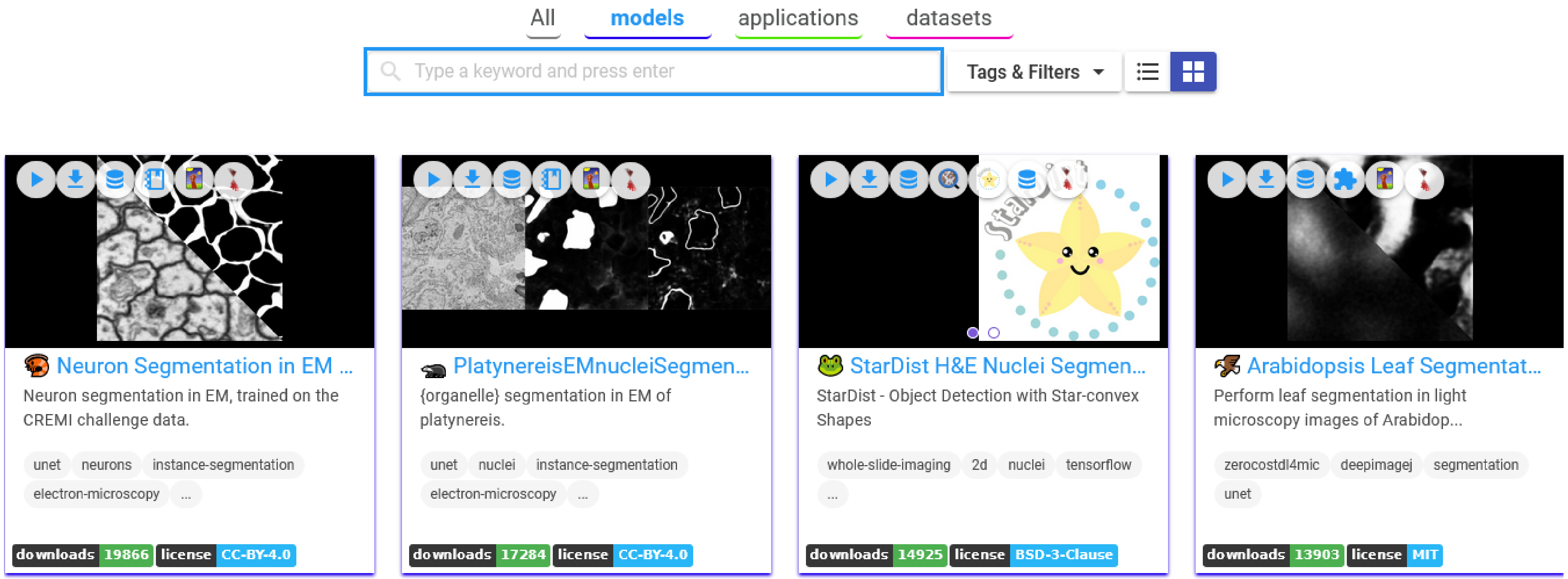
Maybe you already use some deep learning models without knowing

Stardist: Detection and segmentation of cell nuclei in fluorescence microscopy images

CellPose: Cell segmentation on stained tissues

Where to find deep learning models?

- DeepImageJ (for ImageJ/Fiji only)
- BioImage Model Zoo (for ImageJ, QuPath, ilastik ...)



If I want my to train my own model?

ZeroCostDL4Mic: easy-to-use notebooks

`pooling_steps`: Choosing a different number of pooling layers can affect the performance of the network. Each additional pooling step will also two additional convolutions. The network can learn more complex information but is also more likely to overfit. Achieving best performance may require testing different values here. **Default: 2**

`percentage_validation`: Input the percentage of your training dataset you want to use to validate the network during training. **Default value: 10**

`initial_learning_rate`: Input the initial value to be used as learning rate. **Default value: 0.0003**

`patch_width` **and** `patch_height`: The notebook crops the data in patches of fixed size prior to training. The dimensions of the patches can be defined here. When `Use_Default_Advanced_Parameters` is selected, the largest 2^n x 2^n patch that fits in the smallest dataset is chosen. Larger patches than 512x512 should **NOT** be selected for network stability.

`min_fraction`: Minimum fraction of pixels being foreground for a slected patch to be considered valid. It should be between 0 and 1.**Default value: 0.02** (2%)

Path to training images:

`Training_source:` " Insérer la valeur de type text ici

`Training_target:` " Insérer la valeur de type text ici

`model_name:` " Insérer la valeur de type text ici

`model_path:` " Insérer la valeur de type text ici

Training parameters:

Number of epochs

`number_of_epochs:` 100

Advanced parameters:

Python librairies: *tensorflow, pytorch*

Also some librairies for R: *imageseg, platypus*

How to learn?

Workshops at Biosit

Workshops

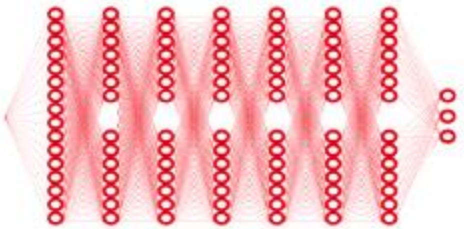
Bioimage Informatics with Fiji/ImageJ

Fall 2022

Whole-Slide Image Analysis and Quantitative Pathology with QuPath

Early 2023

Deep Learning for Biologists



Spring 2023

Ressources

- Médecine et Intelligence artificielle
  - Diagnostic du cancer du sein : l'intelligence artificielle d'Ibex bientôt réalité clinique à l'Institut Curie
  - https://www.ey.com/fr_fr/health/intelligence-artificielle-et-medecins-qui-va-gagner
  - https://www.inserm.fr/dossier/intelligence-artificielle-et-sante/
- Ethics and AI
  - Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward
  - Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?