## Symposium SFP : Intelligence artificielle et anatomie pathologique

Carrefour Pathologie 2019 – 6 novembre 2019 – Palais des congrès, Paris

# Fouille de données, connaissances structurées et apprentissage peu supervisé

## La clé de l'aide au diagnostic en anatomie pathologie

**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

**François-Xavier Frenois, PhD**
fx.frenois@canceropole-gso.org

**Plateforme Imag'IN de l'IUC :** https://www.imagin.univ-tlse3.fr/ImagIn
**GitHub :** https://github.com/ouatataz

***AUCUN LIEN D'INTÉRÊT***

# L'intelligence artificielle est née dans les années 1950

Dartmouth Conference, 1956


Au cours des décennies qui ont suivi, l'IA a été présentée alternativement comme une grande opportunité de développement de nos civilisations, et en même temps très souvent décriée (manque de puissance informatique)



**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Deux courants de recherche se sont constitués…*

- **IA dite « forte » : *General AI***
  - *Concevoir des machines possédants tous nos sens et capables de raisonner comme les humains*
- **IA dite « faible » : *Narrow AI***
  - ***Concevoir des machines capables d'aider les humains dans leurs tâches***

**MÉDECINE**

| | | |
|---|---|---|
| **AIDE À LA DÉCISION**<br>Diagnostic et thérapeutique | **MÉDECINE DE PRÉCISION**<br>Recommandations de traitements personnalisés | **MÉDECINE PRÉDICTIVE**<br>Prédiction d'une maladie et de son évolution |
| **CHIRURGIE ASSISTÉE PAR ORDINATEURS**<br>Robots | **PRÉVENTION**<br>Anticipation d'une épidémie<br>Pharmacovigilance | **IOT - OBJETS CONNECTÉS**<br>Hospitalisation à domicile<br>Personnes âgées |

  - ***Mobilise de nombreuses disciplines*** *(mathématiques, informatique, sciences cognitives, philosophie) + connaissances spécialisées des domaines auxquels on souhaite l'appliquer (médecine, transports, etc…)*
  - ***Systèmes limités dans leurs capacité d'adaptation :*** *doivent être adaptés pour accomplir des tâches pour lesquelles ils n'ont pas été initialement conçus*

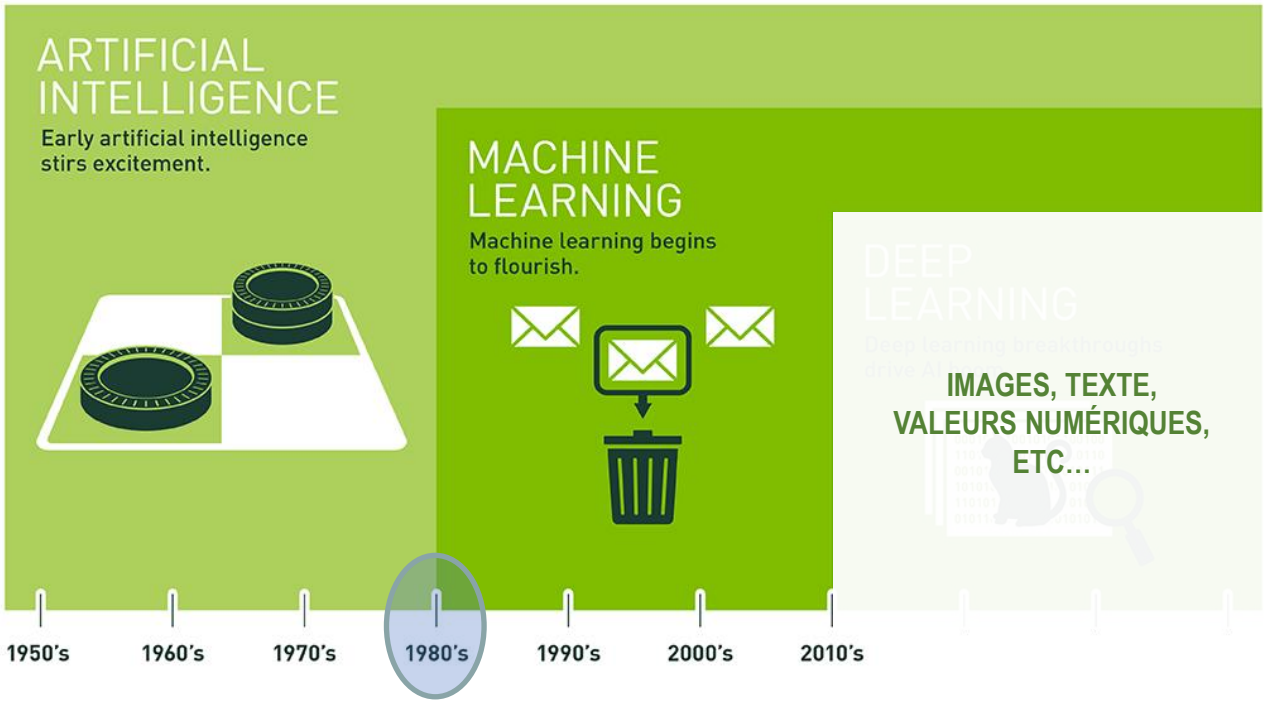

**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *L'apprentissage automatique*

*À la base des technologies mises en œuvre dans l'IA faible*

### Algorithmes qui raisonnent sur des données : Big Data

- ***Machine Learning*** *depuis les années 1980 (apprentissage machine)*





ACCUMULATION DE DONNÉES
DIMINUTION DES COÛTS DE STOCKAGE

Ces algorithmes se fondent sur des approches statistiques pour donner aux ordinateurs la capacité d'apprendre à partir de données puis de faire des prédictions sur de nouvelles données

AMÉLIORATION DES PERFORMANCES POUR RÉSOUDRE DES TÂCHES SANS ÊTRE EXPLICITEMENT PROGRAMMÉS POUR CHACUNE



**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Machine learning : modèle statistique ou probabiliste*

*Différentes approches algorithmiques souvent utilisées en combinaison pour aboutir à une classification*

- **Apprentissage supervisé et non supervisé :** selon les informations disponibles durant la phase d'apprentissage…
  - **Supervisé : les données d'entrée sont fournies au système avec leur label** (i.e. image H&E + Diagnostic associé)
  - **Non supervisé : pas de label**, on cherche à déterminer la **structure sous jacentes des données**
  - **Autres :** semi-supervisé, par renforcement, transfert learning…
- **Données possibles :** vecteurs de caractéristiques (*n* dimensions), graphes, arbres, courbes…

**Image RVB 8 bits** (256 niveaux de gris / canal)

DOMINANTE VERTE

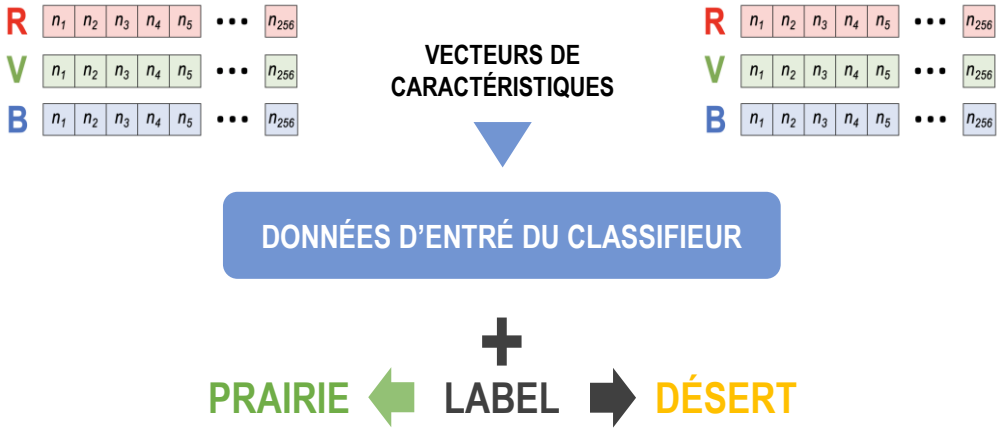



**Image RVB 8 bits** (256 niveaux de gris / canal)

DOMINANTE JAUNE

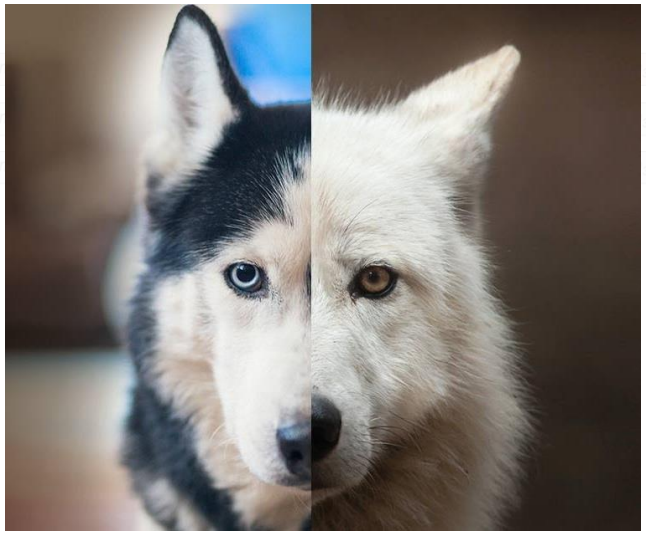


**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Machine learning : modèle statistique ou probabiliste*

*Différentes approches algorithmiques souvent utilisées en combinaison pour aboutir à une classification*

- **Apprentissage supervisé et non supervisé :** selon les informations disponibles durant la phase d'apprentissage…
  - **Supervisé : les données d'entrée sont fournies au système avec leur label** (i.e. image H&E + Diagnostic associé)
  - **Non supervisé : pas de label**, on cherche à déterminer la **structure sous jacentes des données**
  - **Autres :** semi-supervisé, par renforcement, transfert learning…
- **Données possibles :** vecteurs de caractéristiques (*n* dimensions), graphes, arbres, courbes…





**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Machine learning : modèle statistique ou probabiliste*

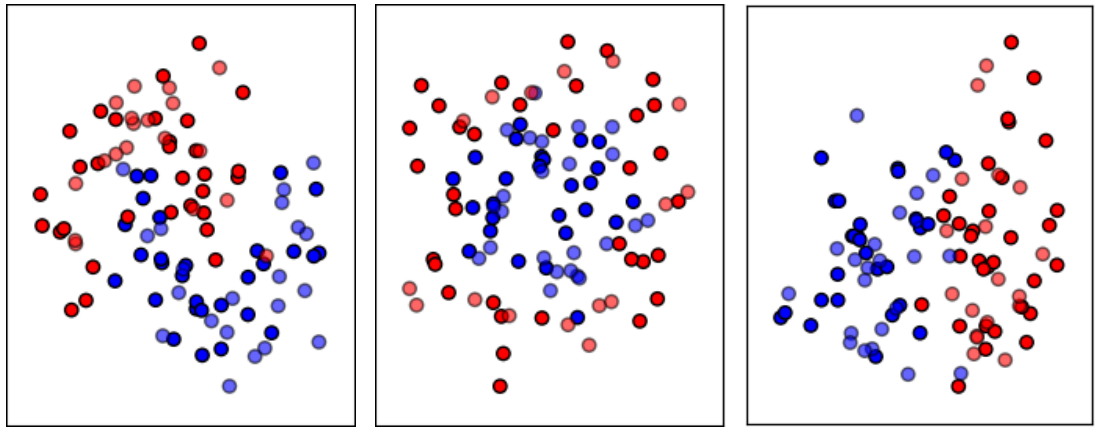
*Principaux algorithmes et code de test*

## Principaux algorithmes :

- **Supervisé :** SVM, ARBRES DE DÉCISION (RANDOM FOREST), KNN, PERCEPTRON (RÉSEAU DE NEURONES)…
- **Non supervisé :** K-MEANS…

## Python : librairie « scikit-learn »

*https://scikit-learn.org/stable/index.html*



**DATASET #1**
`make_moons`

**DATASET #2**
`make_circles`

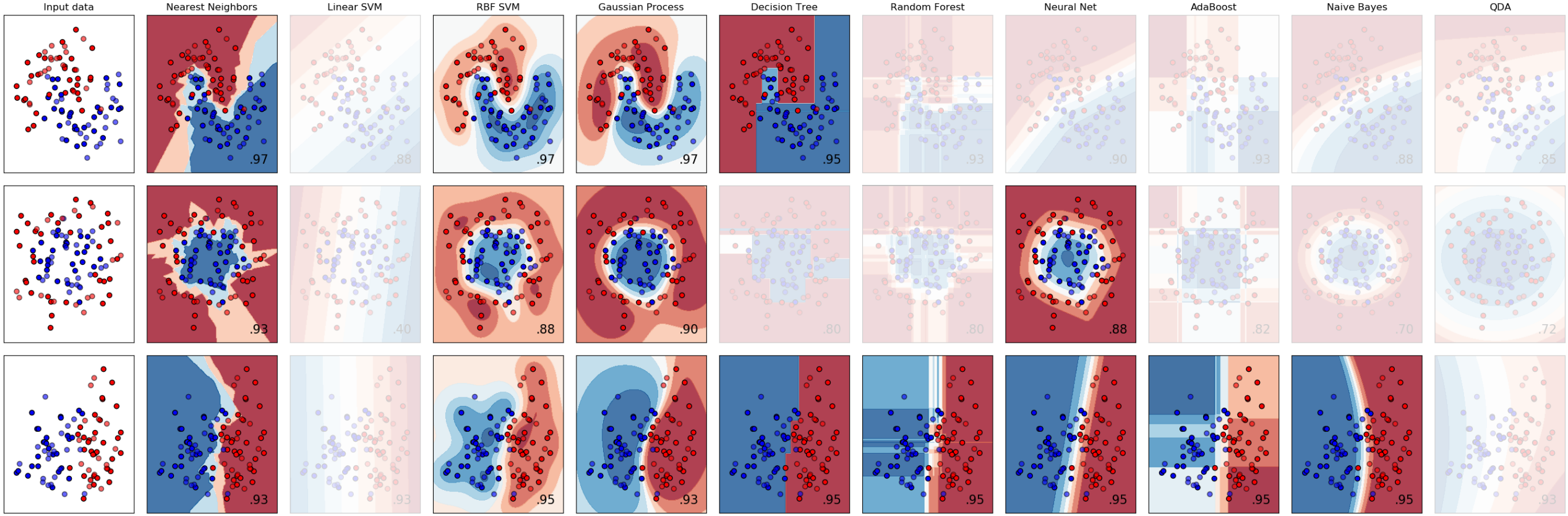
**DATASET #3**
`linearly_separable`

```python
# Initialize classifiers
    names = [
        "Nearest Neighbors",
        "Linear SVM",
        "RBF SVM",
        "Gaussian Process",
        "Decision Tree",
        "Random Forest",
        "Neural Net",
        "AdaBoost",
        "Naive Bayes",
        "QDA"
    ]
    classifiers = [
        KNeighborsClassifier(3),
        SVC(kernel="linear", C=0.025),
        SVC(gamma=2, C=1),
        GaussianProcessClassifier(1.0 * RBF(1.0)),
        DecisionTreeClassifier(max_depth=5),
        RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1),
        MLPClassifier(alpha=1, max_iter=1000),
        AdaBoostClassifier(),
        GaussianNB(),
        QuadraticDiscriminantAnalysis()
    ]

# Generate a random n-class classification problem
    X, y = make_classification(
        n_features=2,
        n_redundant=0,
        n_informative=2,
        random_state=1,
        n_clusters_per_class=1
    )

# Initialize datasets
    datasets = [
        make_moons(noise=0.3, random_state=0),
        make_circles(noise=0.2, factor=0.5, random_state=1),
        linearly_separable
    ]

# Iterate over datasets and pre-process
# Split datasets into training and test part
    for ds_cnt, ds in enumerate(datasets):
        X, y = ds
        X = StandardScaler().fit_transform(X)
        X_train, X_test, y_train, y_test = \ train_test_split(X, y, test_size=.4, random_state=42)

        # iterate over classifiers
        for name, clf in zip(names, classifiers):
            clf.fit(X_train, y_train)
            score = clf.score(X_test, y_test)
```

Arnaud Abreu & François-Xavier FRENOIS, PhD
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
Plateforme Imag'IN de l'IUC - *https://www.imagin.univ-tlse3.fr*

# *Machine learning : modèle statistique ou probabiliste*

*Principaux algorithmes et code de test*



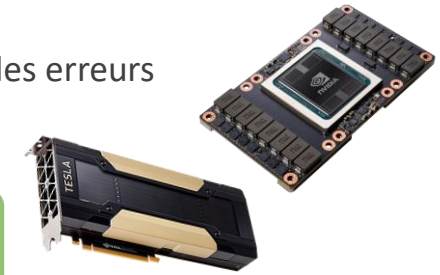

**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Réseaux de neurones : algorithmes inspirés du fonctionnement du cerveau biologique*

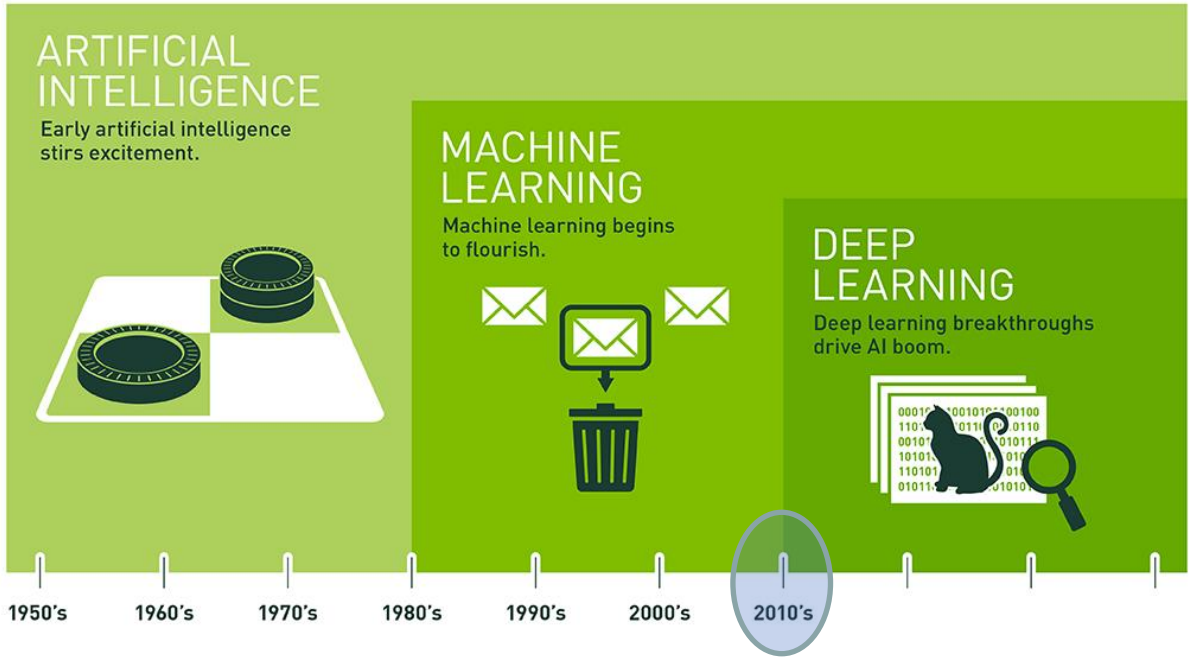
*Simulent un réseau de neurones organisés en différentes couches, échangeant les uns avec les autres*

## **Réseaux de neurones profonds convolutionnels (CNN) :** classification d'images

- **Apprentissage par essais et erreurs :** renforcement des poids synaptiques entre les couches et rétro-propagation des erreurs
- **Très utilisés depuis 2012 :** augmentation de la puissance de calcul parallélisable offerte par les cartes GPU…



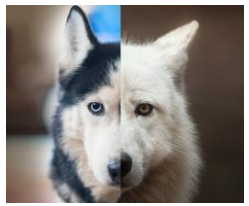
CARTES GPU
NVIDIA TESLA V100




**Particularité des réseaux profonds :** le nombre de couches est important (niveau d'abstraction très élevé).

CLASSIQUE

PROFOND

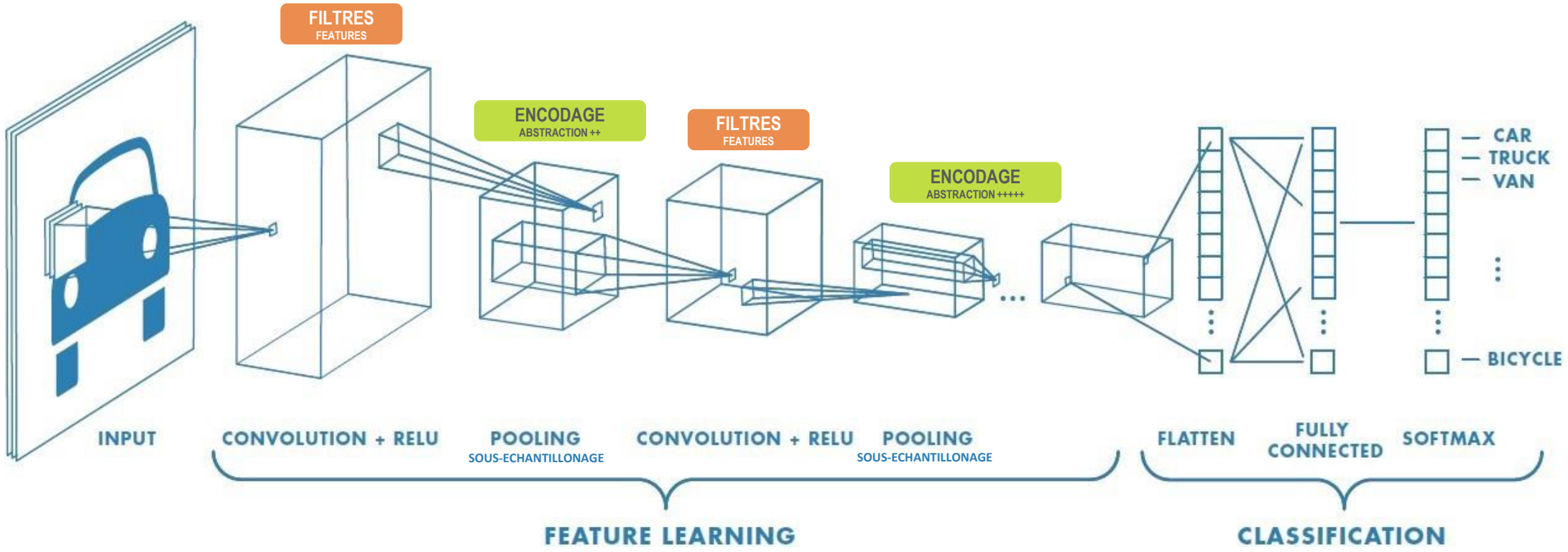


**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

http://scs.ryerson.ca/~aharley/vis/conv/

## *Design d'un CNN profond*

*Convolution + MaxPooling + Perceptron*



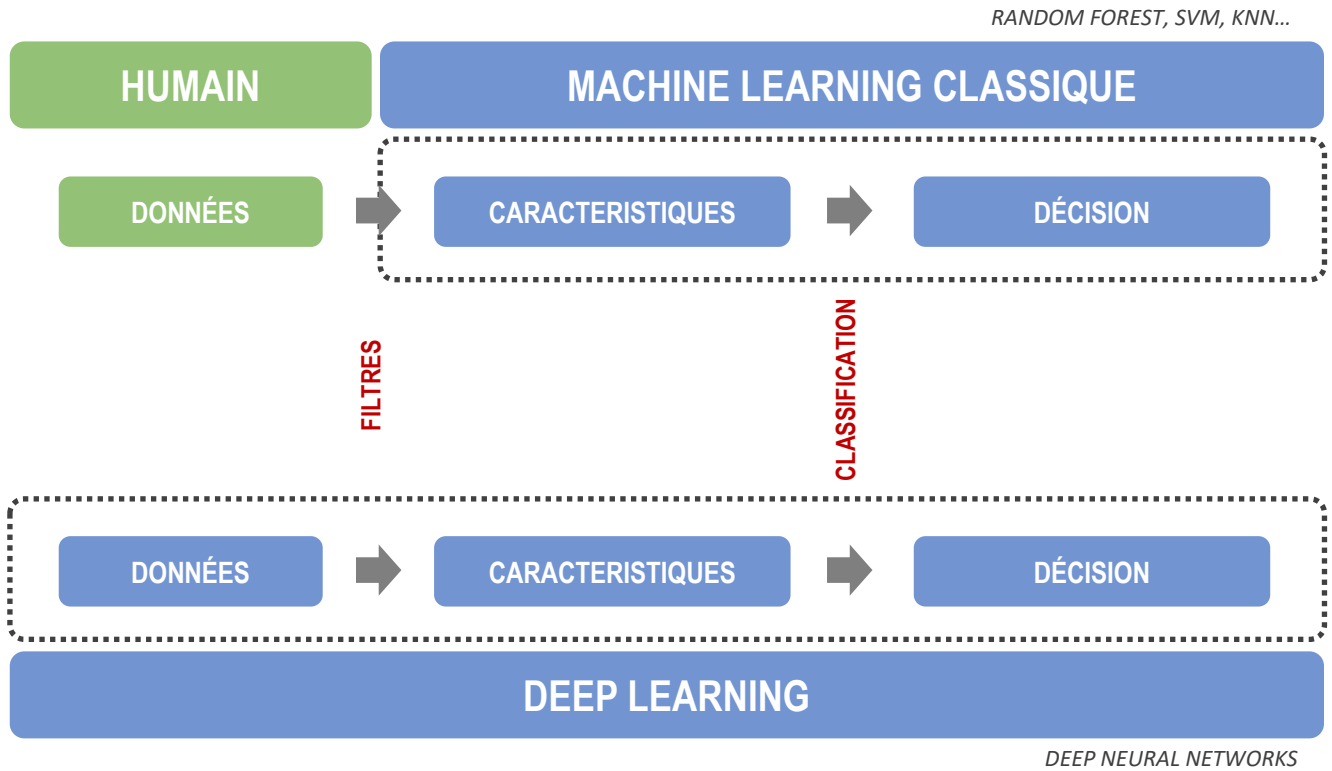

**Déterminer automatiquement un espace de caractéristique tout en encodant l'image**



**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Machine learning versus deep learning*

*L'espace des caractéristiques est connu et défini par l'homme dans le machine learning, alors qu'un réseau de neurones va lui-même déterminer les caractéristiques qui expliquent les données pour permettre la prise de décision*




**Mise en oeuvre** (infrastructures et algorithmes éprouvés)
**Caractéristiques connues :** hiérarchie, corrélations…

**Intervention humaine :** objectivité et complexité diminuées
**Temps perdu** en design de caractéristiques et métriques

**Pas d'élaboration de caractéristiques, décisions très complexes**
**Variété de données :** images, langage, etc…

**Lourd à mettre en oeuvre**
**Quantité de données** pour l'apprentissage : $10^5$ ~ $10^6$
**Boîte noire** (caractéristiques inconnues), **explicabilité**



**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

**Arnaud Abreu**
aabreu@unistra.fr
arnaud.abreu@roche.com

**GitHub :** https://github.com/ArnaudAbreu

*Employé de l'Institut Roche*

## Deep-Learning, principales limitations et mauvaises tendances :

- **Mise en place : Rassembler, séparer, pré-traiter, labéliser, entraîner et paramétrer, post-traiter** (Compétences + Temps)
- **Déploiement : Méthode statistique**…
- **Applicatif : Souvent atomique, très limité et décevant**
- **Explicabilité : Boîte noire, pixels -> décision sans intermédiaire**


**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

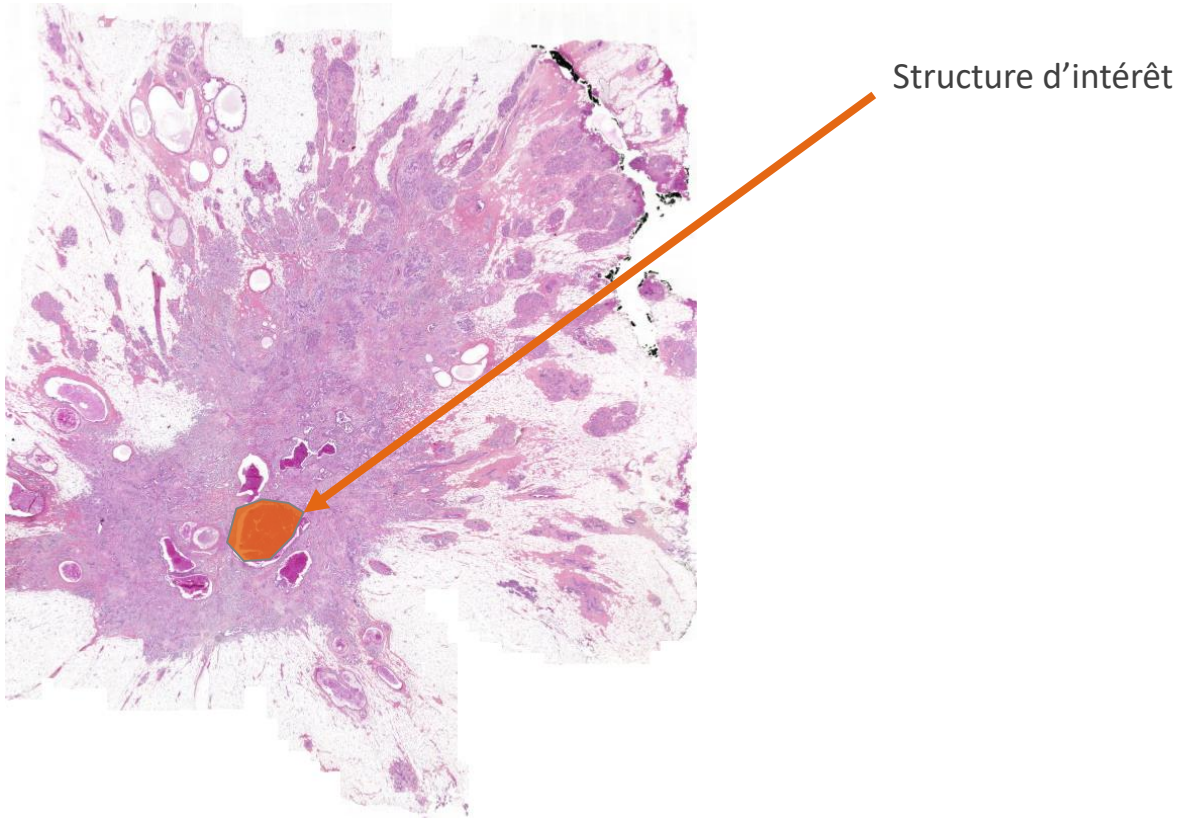
## Deep-Learning, principales limitations et mauvaises tendances :

- **Mise en place :** Rassembler, séparer, pré-traiter, **Labéliser**, entraîner et paramétrer, post-traiter (Compétences + Temps)
- **Déploiement :** Méthode statistique…
- **Applicatif :** Souvent atomique, très limité et décevant → **Exhaustivité**
- **Explicabilité :** Boîte noire, pixels -> décision sans intermédiaire


**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Exhaustivité et explicabilité de l'analyse : le langage du pathologiste*

*Formalisme rigoureux de la structuration de la connaissance*

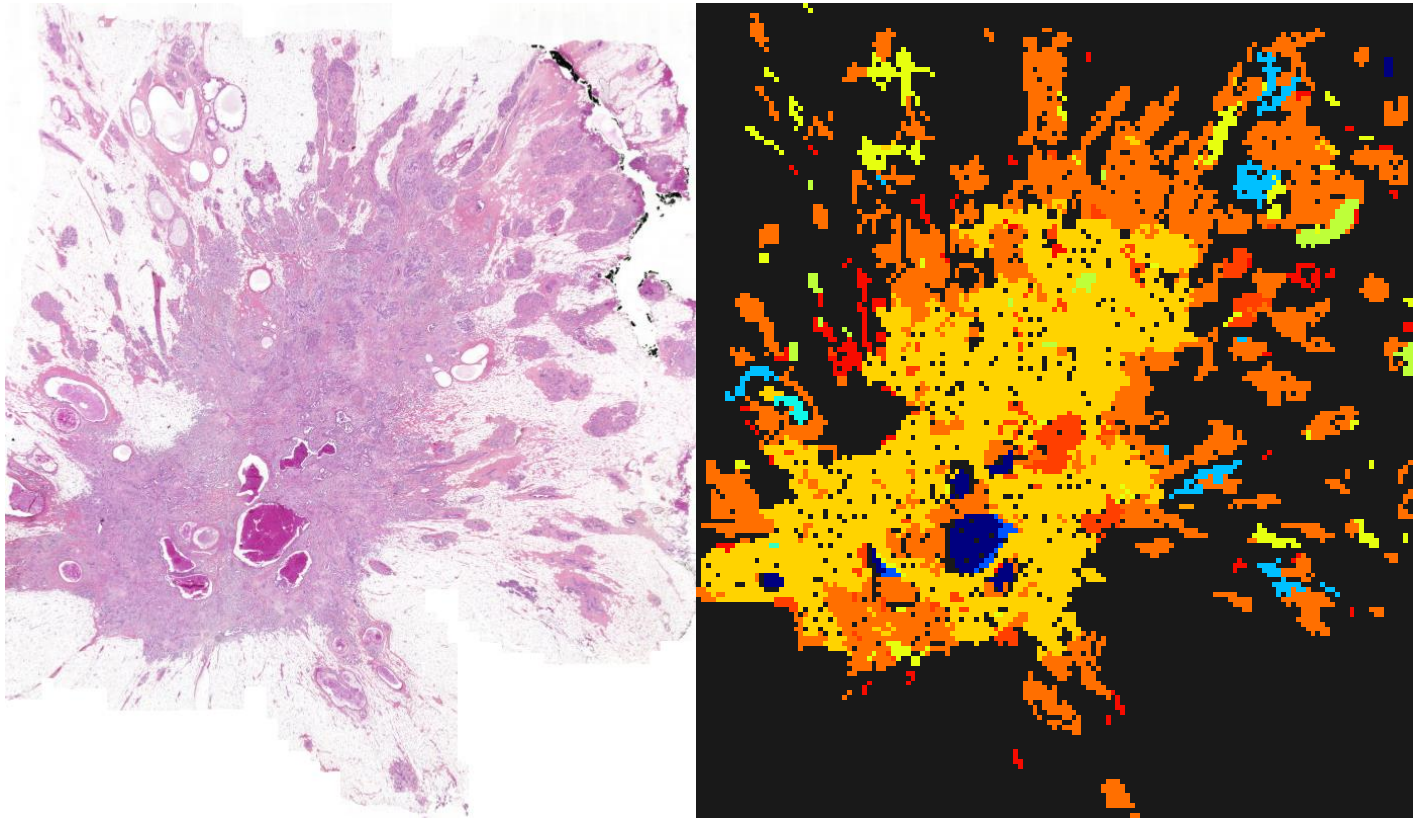





**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Exhaustivité et explicabilité de l'analyse : le langage du pathologiste*

*Formalisme rigoureux de la structuration de la connaissance*

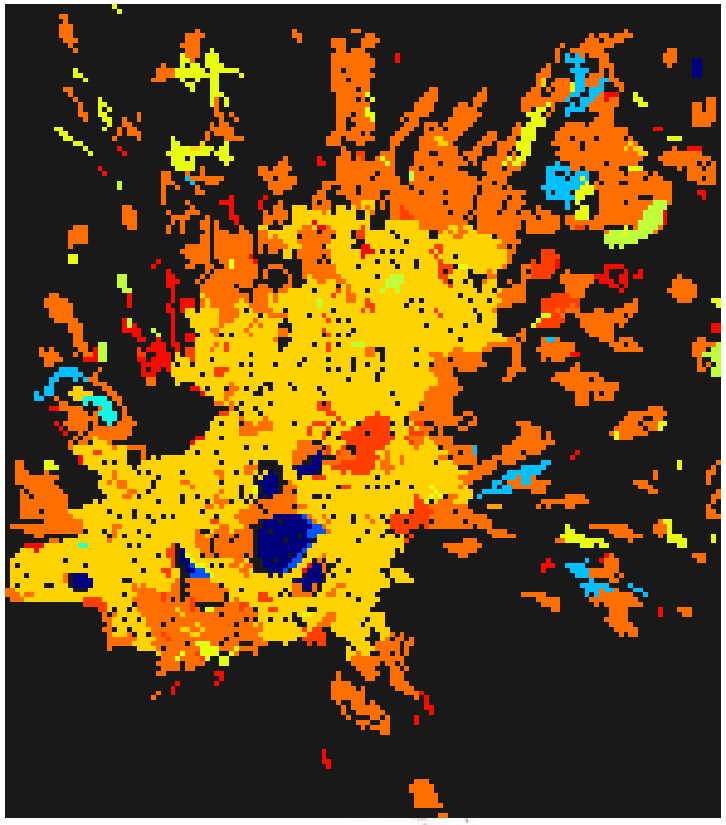




**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Exhaustivité et explicabilité de l'analyse : le langage du pathologiste*

*Formalisme rigoureux de la structuration de la connaissance*

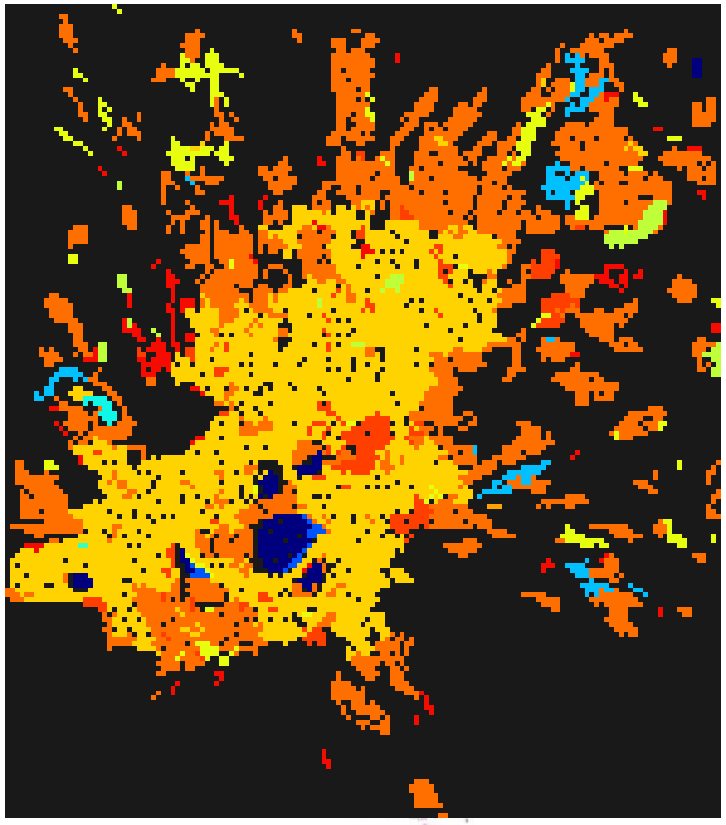


**Identification**
Segmentation,
classification


**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

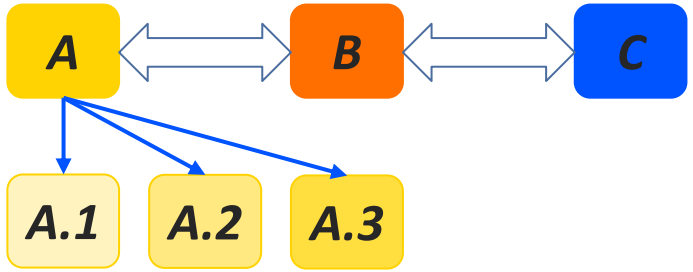
# *Exhaustivité et explicabilité de l'analyse : le langage du pathologiste*

*Formalisme rigoureux de la structuration de la connaissance*



**Identification**
Segmentation, classification

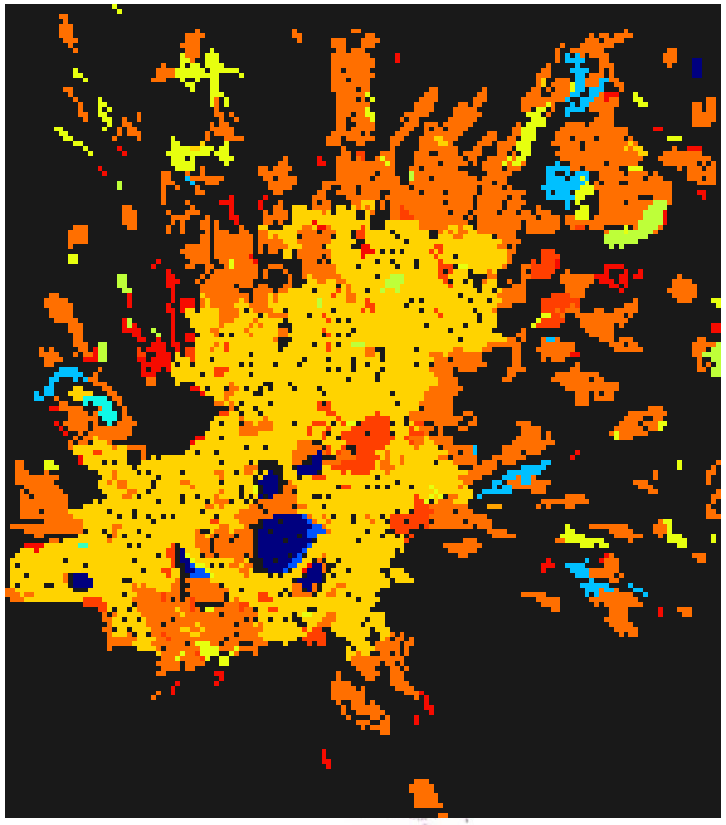
**Hiérarchie conceptuelle**
Sous-type des structures




**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Exhaustivité et explicabilité de l'analyse : le langage du pathologiste*

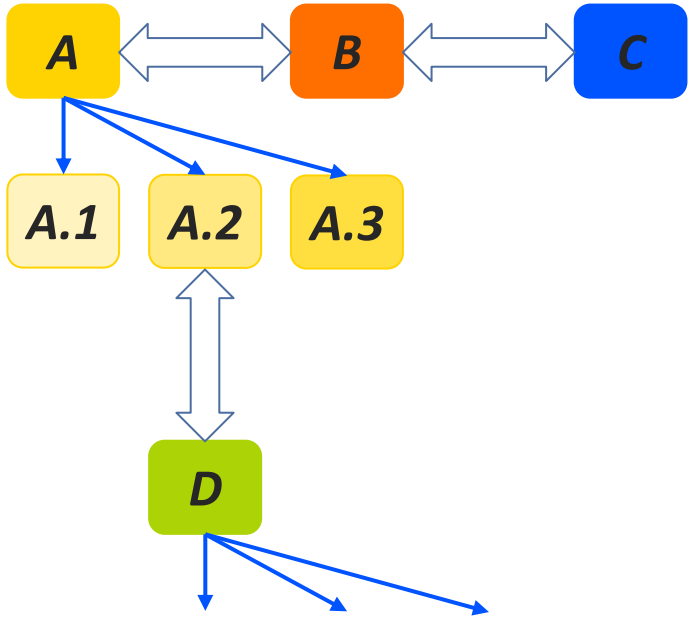
*Formalisme rigoureux de la structuration de la connaissance*



**Identification**
Segmentation, classification

**Hiérarchie conceptuelle**
Sous-type des structures

**Hiérarchie spatiale**
Concepts visibles à plus forte résolution

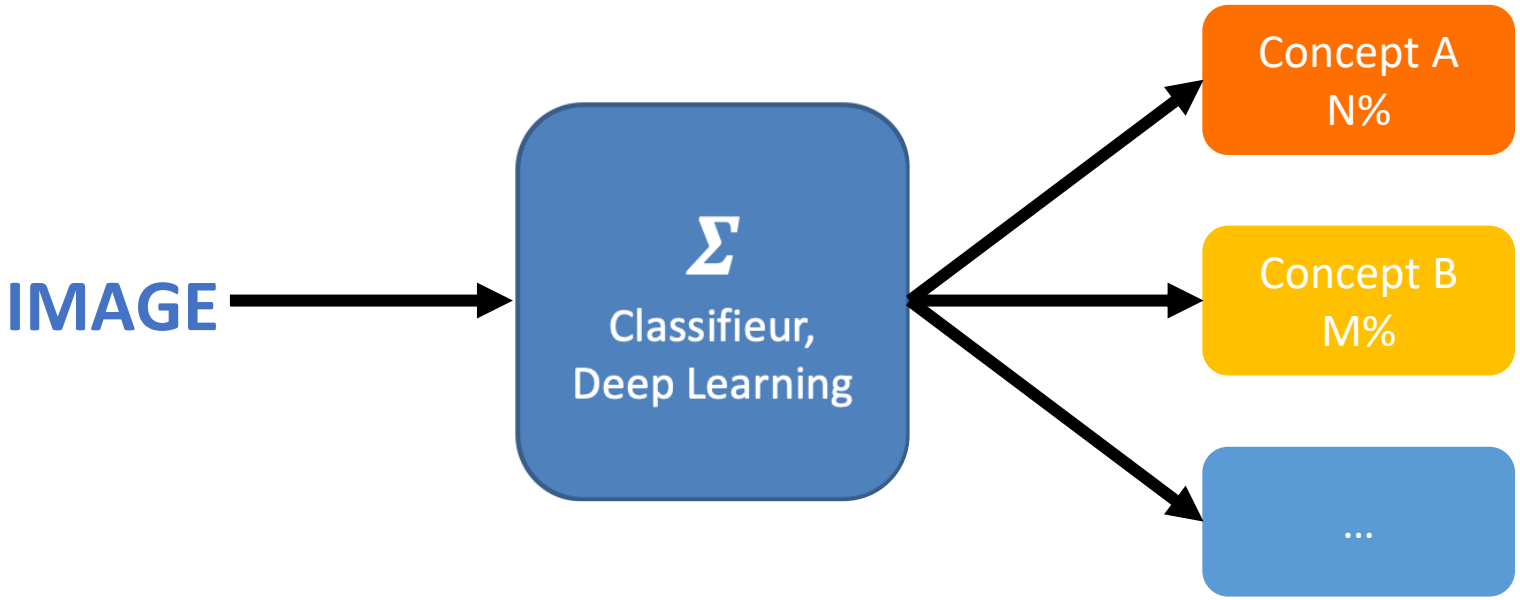



***Décision obtenue par déduction***

**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
**Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Approche machine-learning, le problème de l'annotation*

*Un grand nombre de classes, Une grande variabilité, des experts qui ont autre chose à faire…*

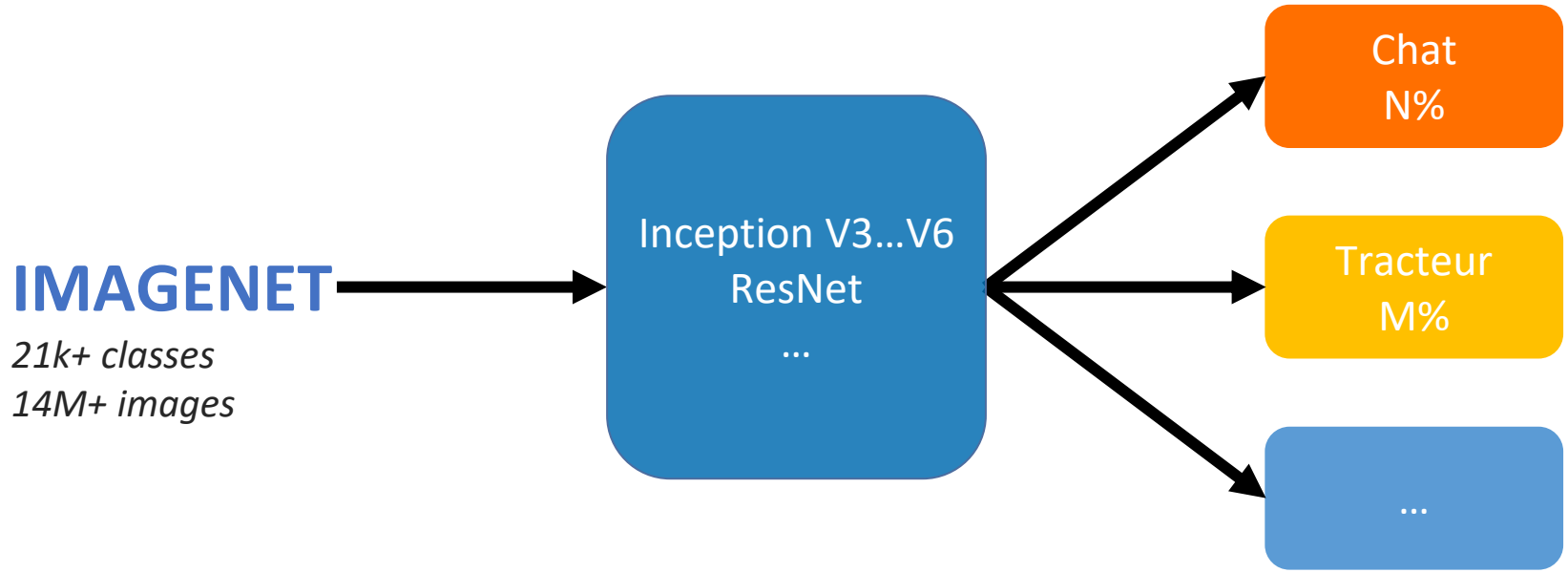





**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Approche machine-learning, le problème de l'annotation*

*Transfert de connaissances, utilisation de réseaux entraînés sur des tâches très générales*

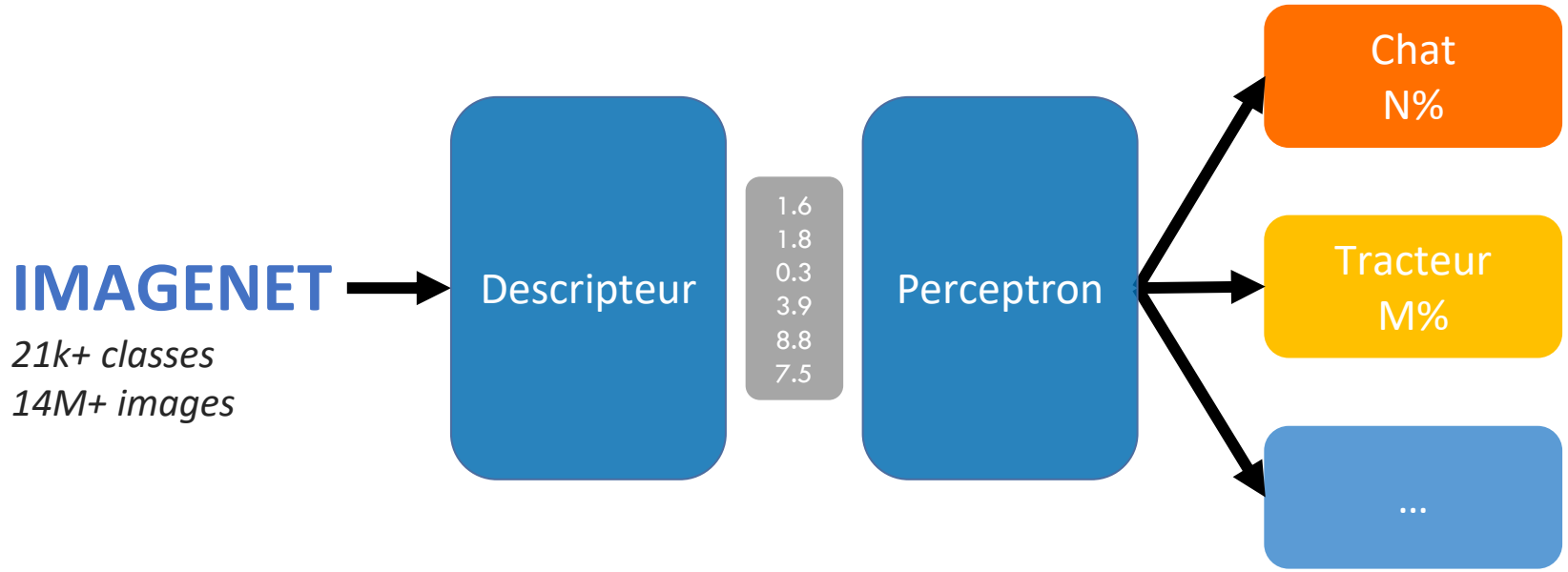





**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Approche machine-learning, le problème de l'annotation*

*Transfert de connaissances, utilisation de réseaux entraînés sur des tâches très générales*





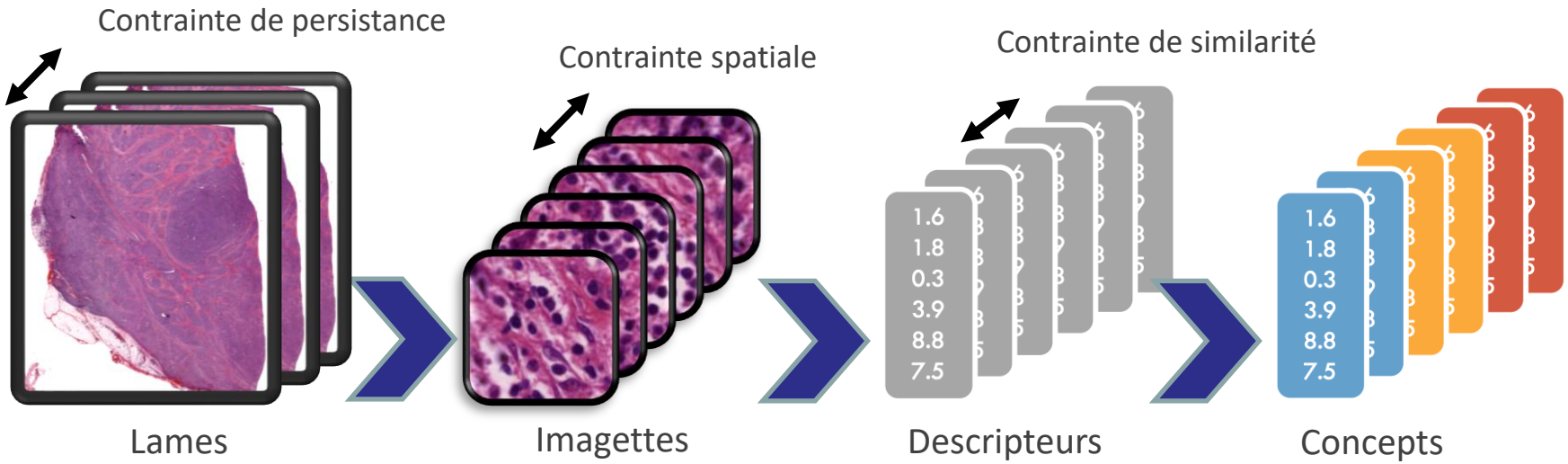
**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# *Approche machine-learning, le problème de l'annotation*

*Transfert de connaissances, clustering dans l'espace des caractéristiques*

## Clustering sous certaines contraintes

- **Contrainte spatiale de transition lente :** Le concept d'un pixel est sûrement le même que celui de son voisin
- **Contrainte de persistance du motif :** Un concept est observable chez plusieurs patients
- **Contrainte de similarité :** Deux segments d'un même concept doivent avoir une représentation similaire



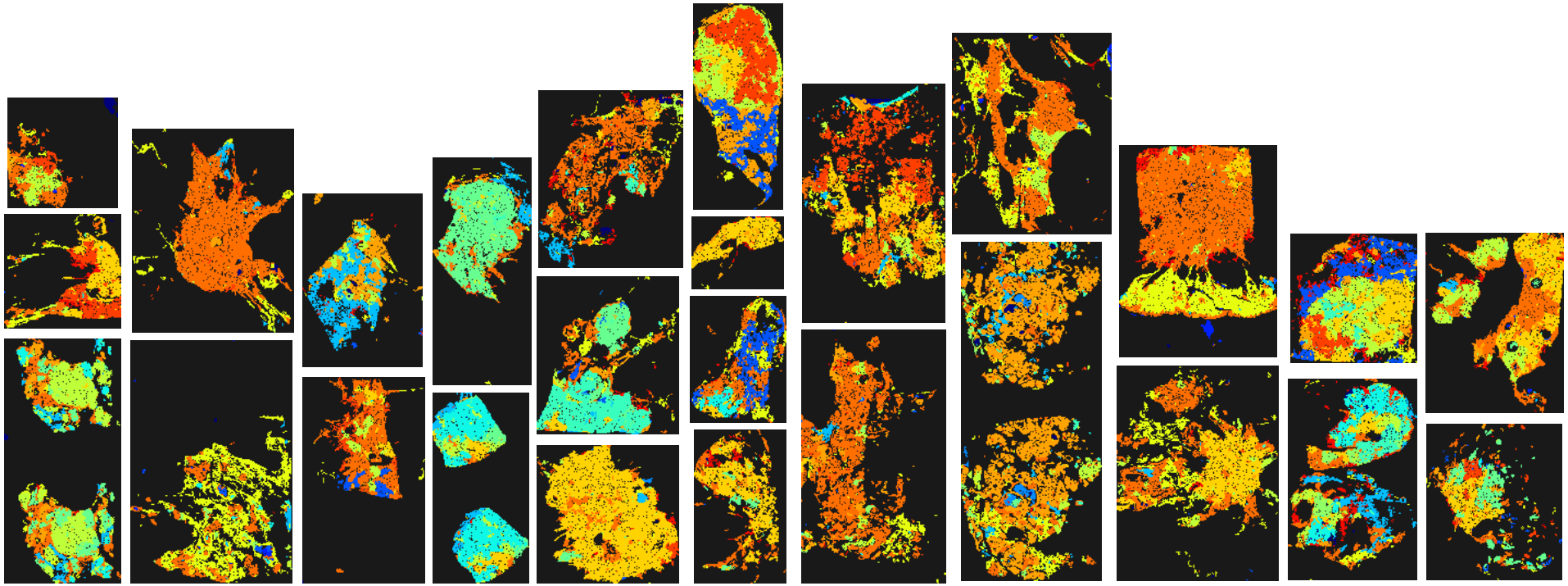



**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

## *Applications, perspectives*

*Usage sur de très grands datasets*

- **Sélection de zones d'intérêt :** exploration de lame optimisée
- **Établissement de sous-concepts pertinents :** Émergence de nouveaux biomarqueurs
- **Requête de bases d'images par mots-clefs :** Études cliniques et regroupement de patients facilités





**Arnaud Abreu & François-Xavier FRENOIS, PhD**
Laboratoire d'Anatomie et Cytologie Pathologiques – IUC Toulouse Oncopôle
IBiSA **Plateforme Imag'IN de l'IUC** - *https://www.imagin.univ-tlse3.fr*

# MERCI POUR VOTRE ATTENTION

**Département de Pathologie (CHU Toulouse)**
- **Pr Pierre Brousset**
- Dr François-Xavier Frenois
- Mr Arnaud Abreu
- Dr Camille Franchet
- Dr Camille Laurent
- Me Myriam Marty
- Mr Gaël Gascoin
- … et l'ensemble du personnel

**Plateforme Imag'IN de l'IUC**
- Dr François-Xavier Frenois
- Dr Nathalie Van Acker
- Me Eveline Caussat
- Me Stéphanie Grenard
- Me Gabrielle Perez

**Laboratoire iCube (Strasbourg)**
- Dr Cédric Wemmert
- Dr Benoit Naegel