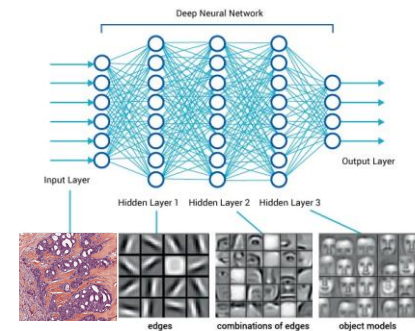
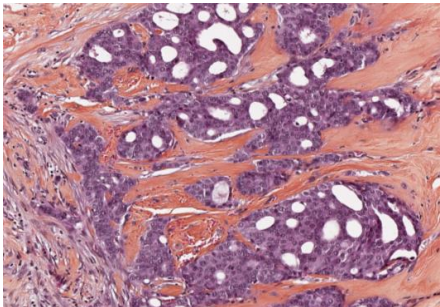


La pathologie à l'heure de l'intelligence artificielle

Quelles questions biologiques ?

Exemple de la pathologie mammaire



Dr Guillaume Bataillon

Service d'anatomie cytologie pathologiques
Pôle de médecine diagnostique et thérapeutique
Institut Curie

Plan

1. Quelles questions biologiques poser ?

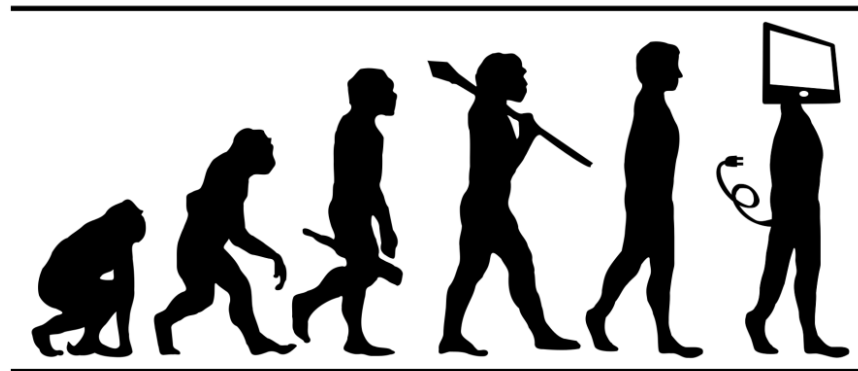
Le projet répond-il à un besoin ?

2. Quels sont les prérequis ?

Existe-t-il des données de qualité en quantité suffisante ?

3. Quelle approche algorithmique utiliser ?

4. Exemples



1. Questions biologiques : 2 types d'approches

Fonctionnelle, reproductive programmatiste

Reproduire une tâche spécifique selon l'état de l'art.

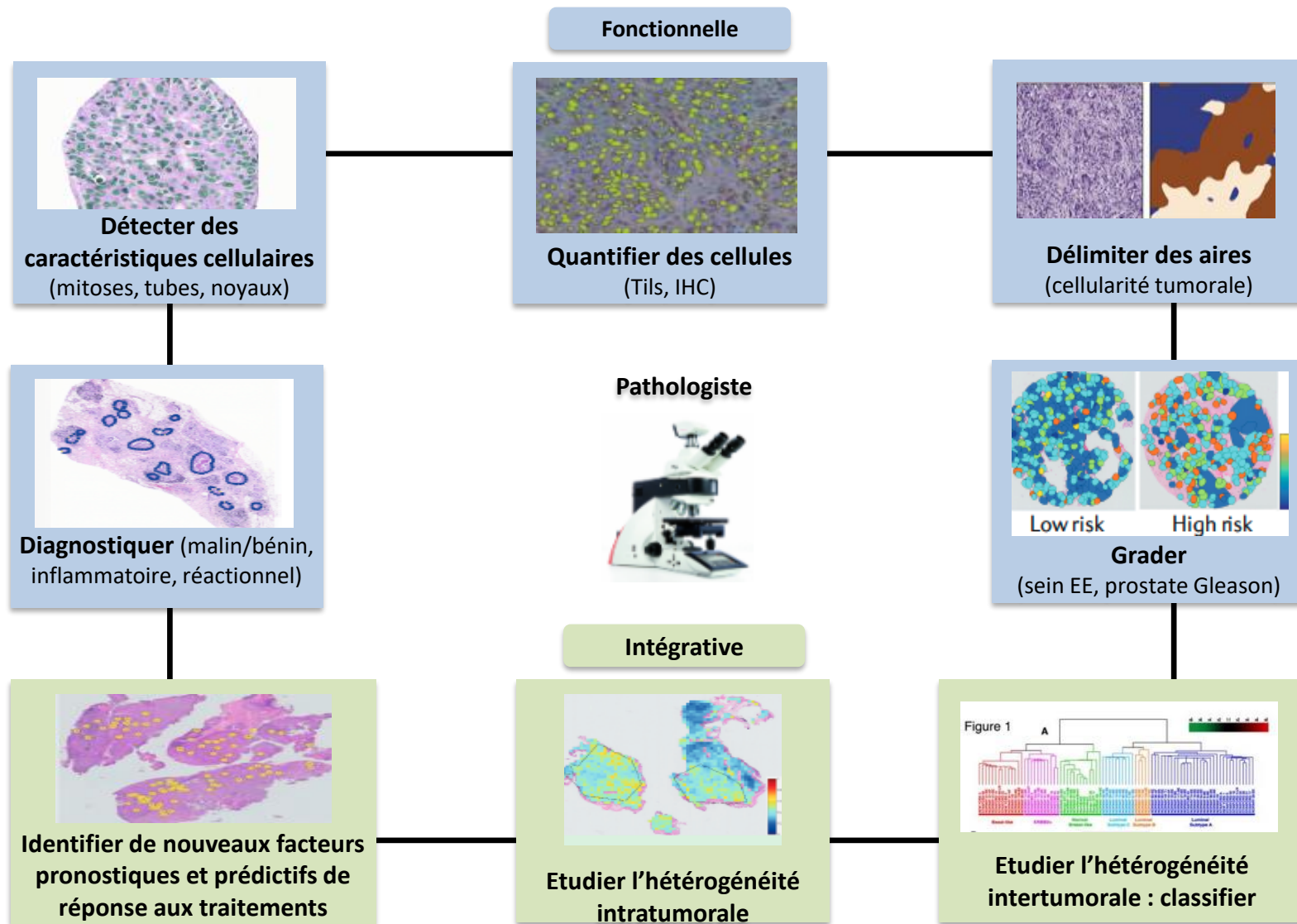
Automatisation des tâches répétitives et chronophages : aide au diagnostic, contrôle qualité, tri

Productive, intégrative généraliste

Prédire, à partir de données complexes, un paramètre non accessible à l'observation classique.

Prédire les caractéristiques clinico-moléculaires des tumeurs

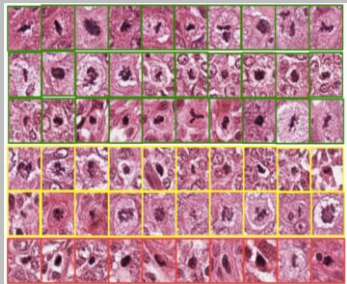
1. Questions biologiques en pathologie



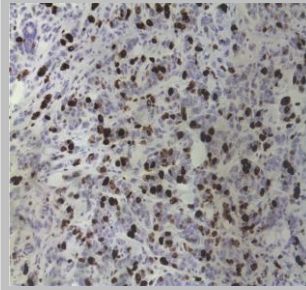
1. Questions biologiques en pathologie mammaire

Wang et al. J Medical Imag 2014

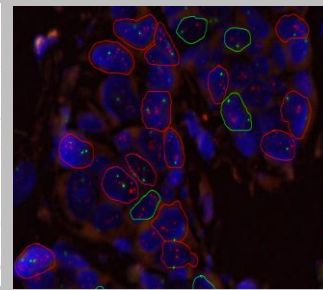
Isola Tuominen Immunoratio 2010



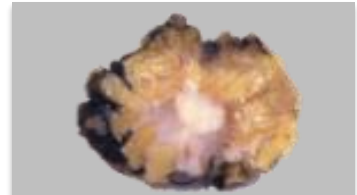
Mitoses



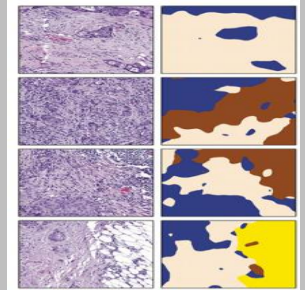
Ki67, RO-RP



FISH HER2



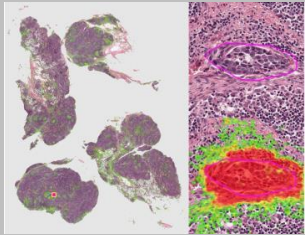
Taille tumorale et limites chirurgicales



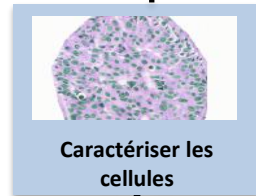
Cellularité tumorale
TILs, Nécrose, Embols

Lundin et al, J Pathol Inform 2015

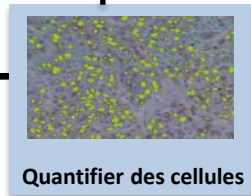
Camelyon 2016



Métastases ganglionnaires



Caractériser les cellules



Quantifier des cellules



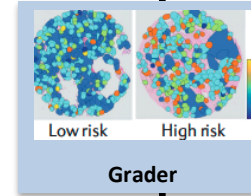
Délimiter des aires



Diagnostiquer



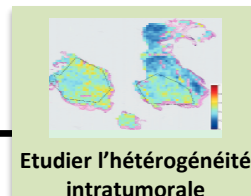
Pathologiste



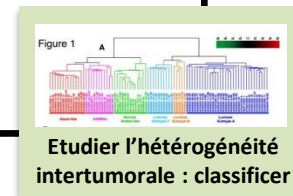
Grader



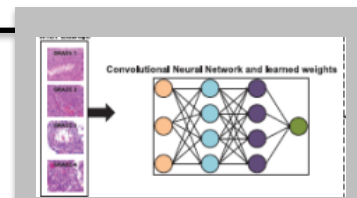
Identifier des facteurs pronostiques et prédictifs



Etudier l'hétérogénéité intratumorale



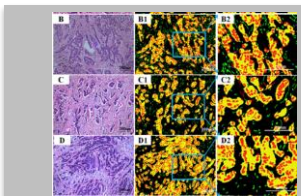
Etudier l'hétérogénéité intertumorale : classifier



Grade

Srivastava et al Biomedic Info 2018

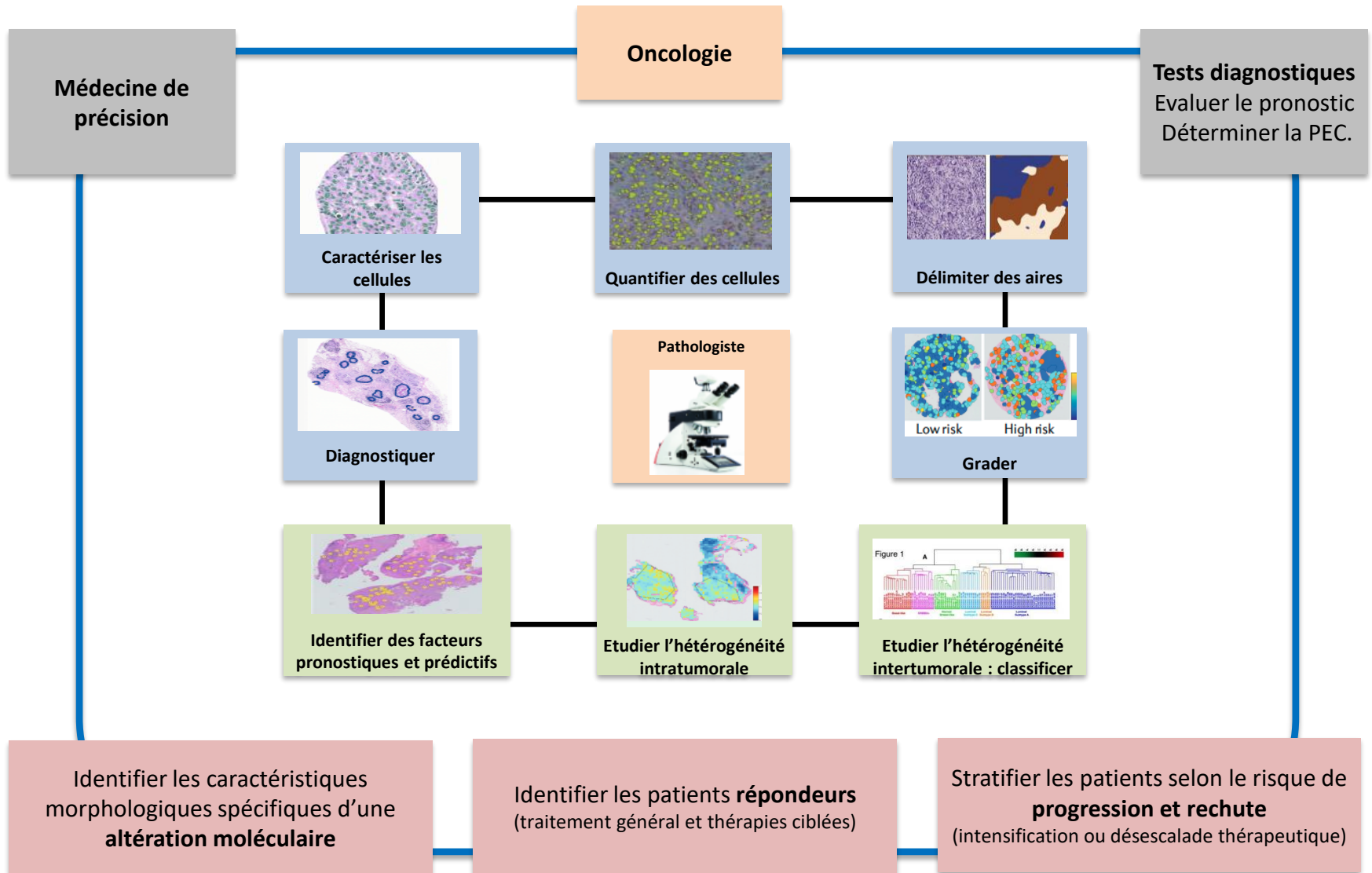
Steiner et al, Am J Surg Pathol 2018



Facteurs pronostiques :
cellularité tumorale, %
stroma

Chen et al Sci rep 2015

1. Questions biologiques en cancérologie



Plan

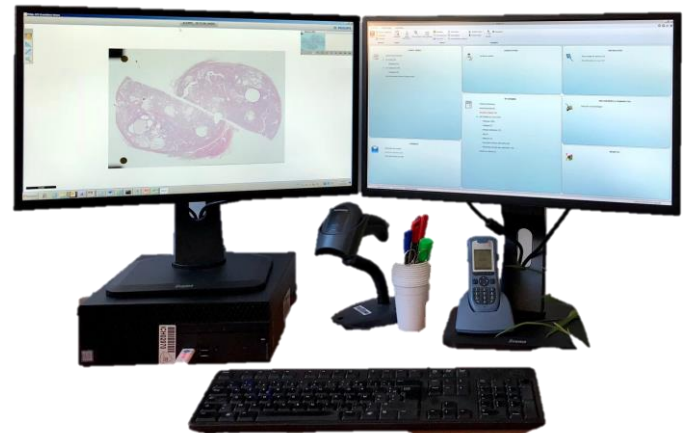
1. Quelles questions biologiques poser ?

Le projet répond-il à un besoin ?

2. Quels sont les prérequis ?

Quelle base de données utiliser ?

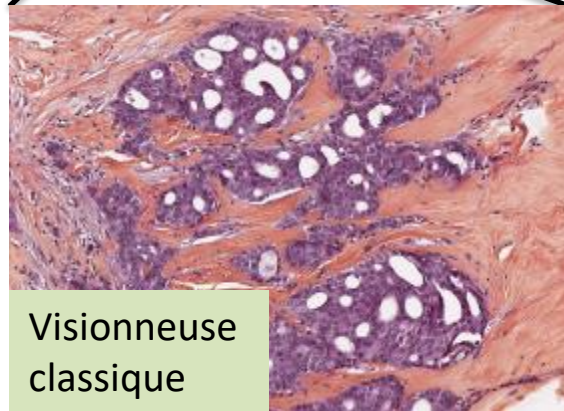
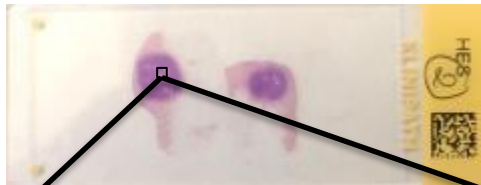
Existe-t-il des données de qualité en quantité suffisante ?



2.1. Prérequis : numérisation des lames



Appareil photo / Microscope



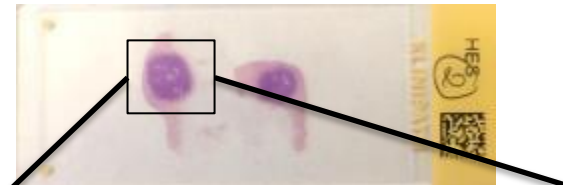
Visionneuse classique

Jpeg 1 Mo

50Mo/jour (10Go/an)



Scanner



Visionneuse dédiée

Format propriétaire 300 Mo – 4 Go

Standardisation **en cours** : DICOM/Opensource

150 000 lames/an

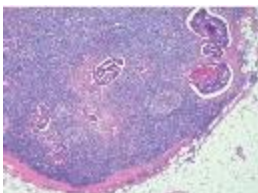
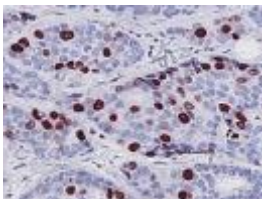
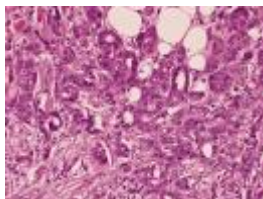
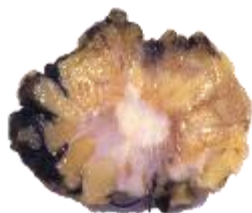
1Po/an

2.1. Prérequis : numérisation du laboratoire (comparatif)

	Pathologie	Génétique	Radiologie
Coût matériel	Scanner de lame Digipath 80lames/J (25k€) Aperio AT2 700lames/J (200k€) 0,76-0,86€/GB	Sequencurs (pool possible) MiSeq 8GB/24h (99k€) NovaSeq 6TB/44h (990k€) 0,9-34€/GB	IRM (500-1500k€) Scanner (500-1000k€)
Coût consommable	Pas de surcoût	MiSeq 1000€ NovaSeq 9000€	Pas de surcoût
Taille des fichiers	0,5-4GB/WSI	2GB/WES	0,08GB/Radio 0,1GB IRM 0,5GB/Scanner
Coût stockage	1-10To/J 0,01€/GB/mois Plus la base est importante plus elle a de la valeur		1-50GB/J
Coût analyse	Visualisation des WSI Annotation des régions d'intérêt	Contrôle de qualité Alignement sur génome de référence	Visualisation Annotation
Algorithmes - Collaboration - Opensource - Ergonomique	Détection des noyaux, mitoses, Degré de pléomorphisme Infiltrat lymphocytaire	Indentification des variants : SNV, Indel Annotation fonctionnelle des variants Etude CNV Analyse des microsatellites	Hierarchiser les examens Radiomique

2.2. Prérequis : structuration des données en pathologie

MASTECTOMIE PARTIELLE DROITE (QSE) ET GANGLION SENTINELLE AXILLAIRE



EXAMEN MACROSCOPIQUE

Taille du prélèvement : 56 mm X 40 mm X 30 mm
Orientation du prélèvement : **Repères inférieur, externe, superficiel**
Lésion(s) macroscopique(s) : **unique de 13 mm X 8 mm**

EXAMEN MICROSCOPIQUE

Carcinome

Type histologique (OMS 2012) : Carcinome Mucineux

Taille du carcinome infiltrant : **13 mm**

Grade Elston & Ellis : **1**

Différenciation : **2** Noyaux : **2** Mitoses : **1**

Nombre de mitoses : **0,64/mm²**

Emboles vasculaires : **Non vus**

CCIS associé : **Non**

TIL du stroma : **20 %**

Immunohistochimie

- Récepteur à l'Estrogène : 100 % cellules positives, Intensité : 3,
Statut RE GEPFICS 1999 (seuil à 10%) : positif
- Récepteur à la Progestérone : 100 % cellules positives, Intensité : 3,
Statut RP GEPFICS 1999 (seuil à 10%) : positif
- HER2 : 0 % cellules positives, Intensité : 0, **Score HER2 GEPFICS 2014 : 0**
- **Index Ki-67 : 5 %**

HER France

Berges d'exérèse sur pièce : **non atteintes** Carcinome infiltrant à 6 mm de la berge externe

Ganglions sentinelles

- Ganglion sentinelle n° GS1 : Statut définitif : **pas de cellules cancéreuses**

CONCLUSION

Carcinome Mucineux

Grade E&E1, RE+, RP+, HER2 0, Ki67 5 %

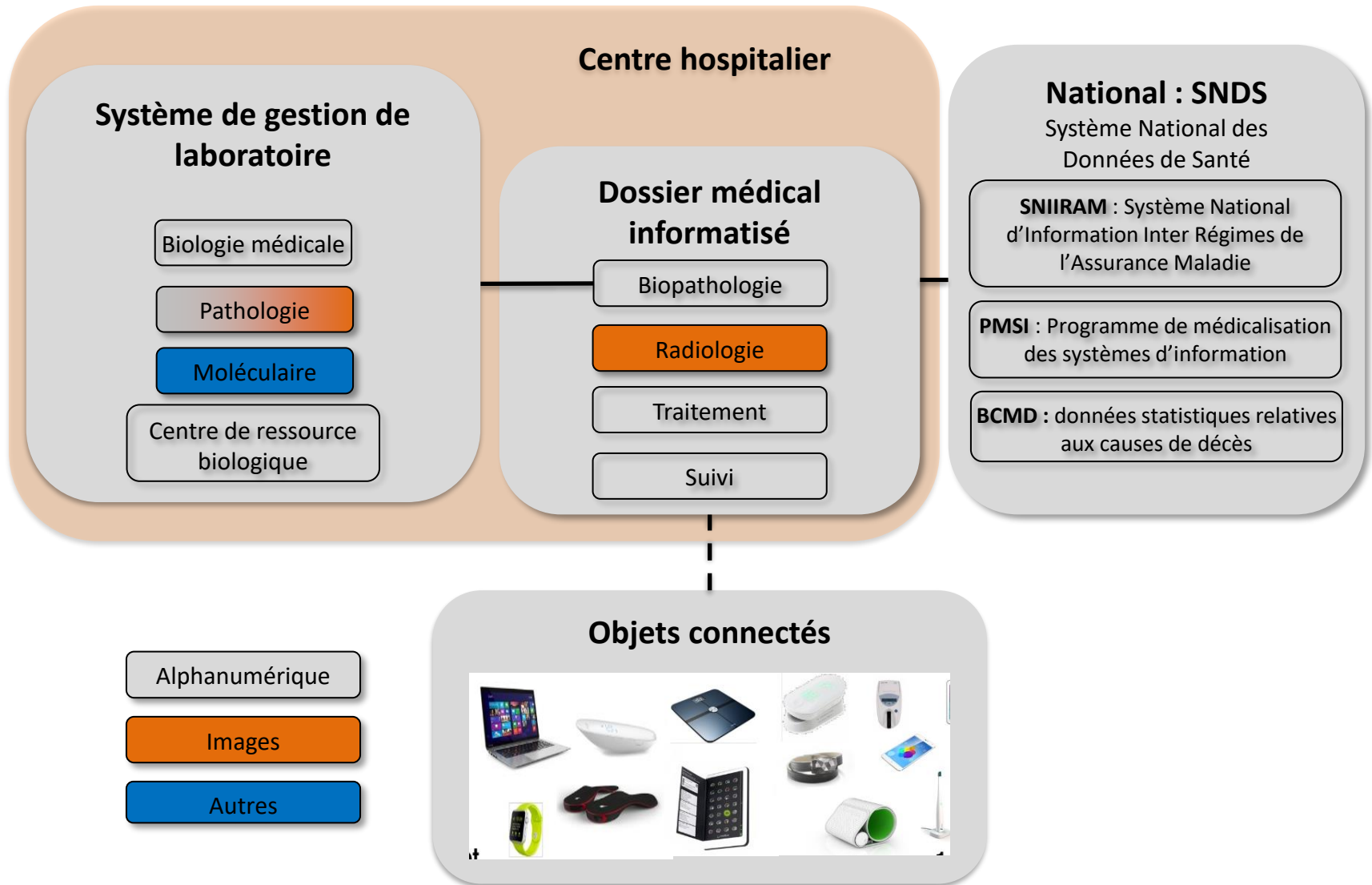
Qualité de l'exérèse chirurgicale : **R0**

Stade UICC/TNM version 7 : **pT1cN0**

CODE CIM : C50.9.8480/3

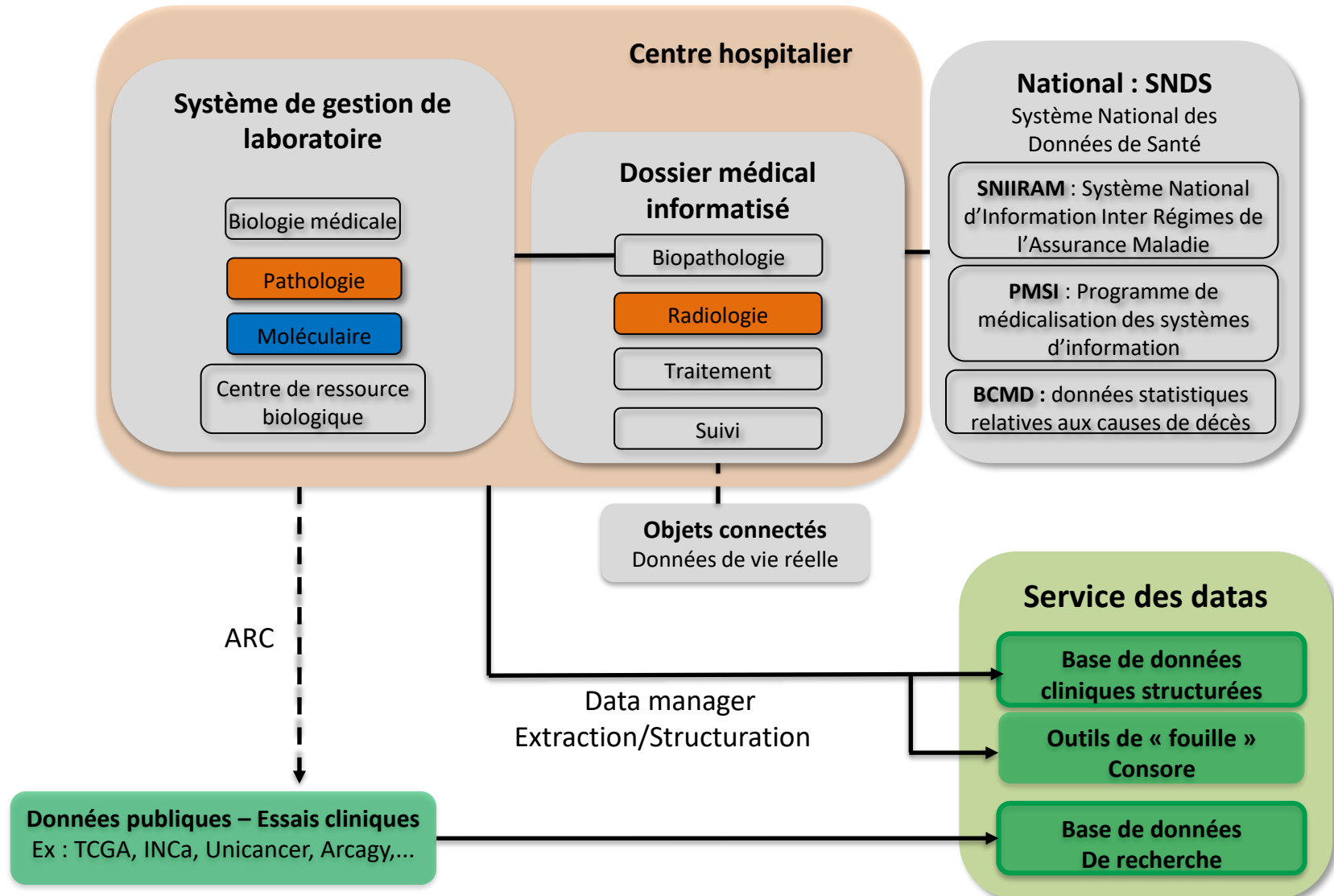
2.2. Prérequis : structuration des données cliniques

Les données en santé



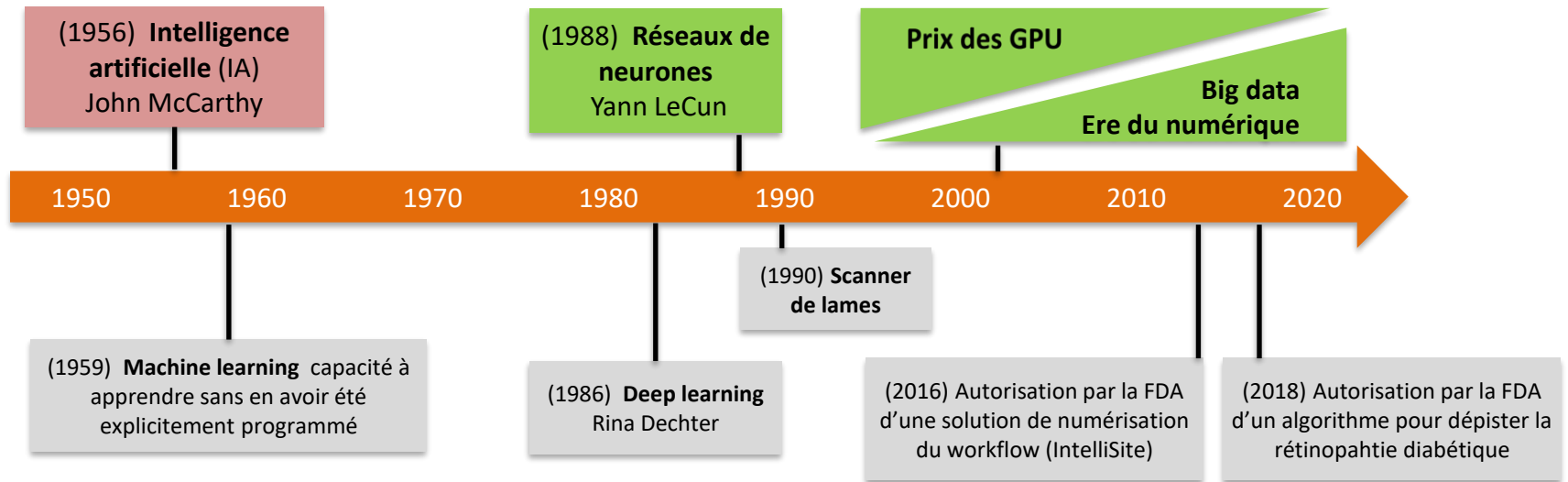
2.2. Prérequis : structuration des données cliniques

Les bases de données



Plan

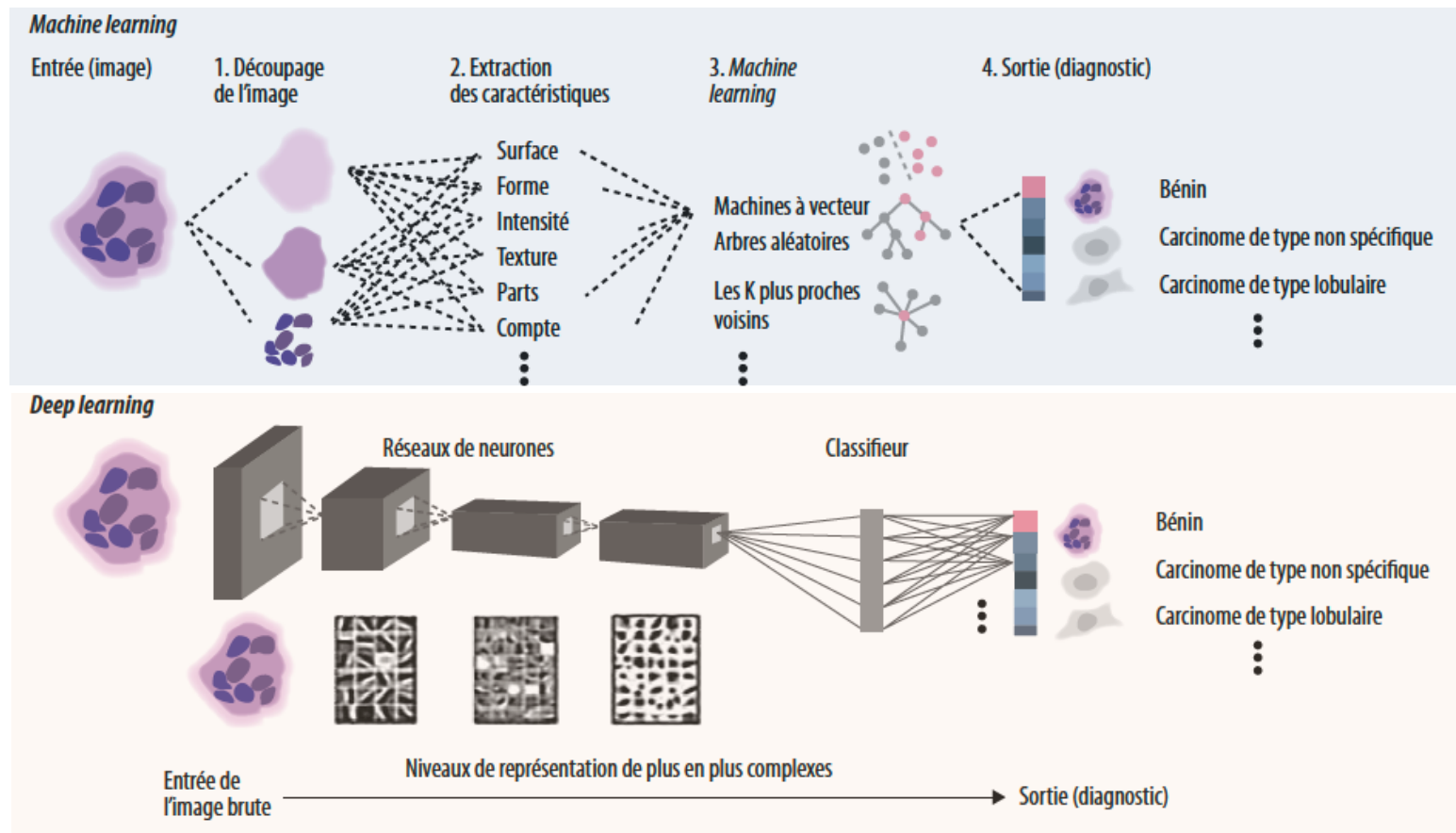
1. Quelles questions biologiques poser ?
2. Quels sont les prérequis ?
3. Quelle approche algorithmique utiliser ?



3. L'intelligence artificielle en médecine et en pathologie : les concepts

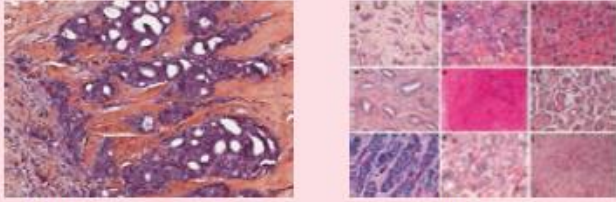
Les algorithmes

Intelligences artificielles : science de la programmation permettant à un algorithme, de réaliser des tâches en tenant compte de nombreux paramètres, imitant l'intelligence humaine de manière efficace, adaptative et évolutive




3. L'intelligence artificielle en médecine et en pathologie : les concepts Méthodes d'apprentissage

Supervisée



Type non spécifique Autres

Quel type de tumeur est-ce ?

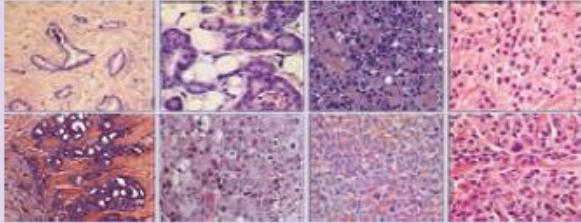


C'est un type non spécifique


Tâches bien définies

- Grader, scorer
- Prévisions

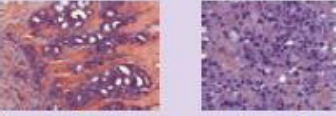
Non supervisée



Combien de types de tumeurs ?



Deux types de tumeurs

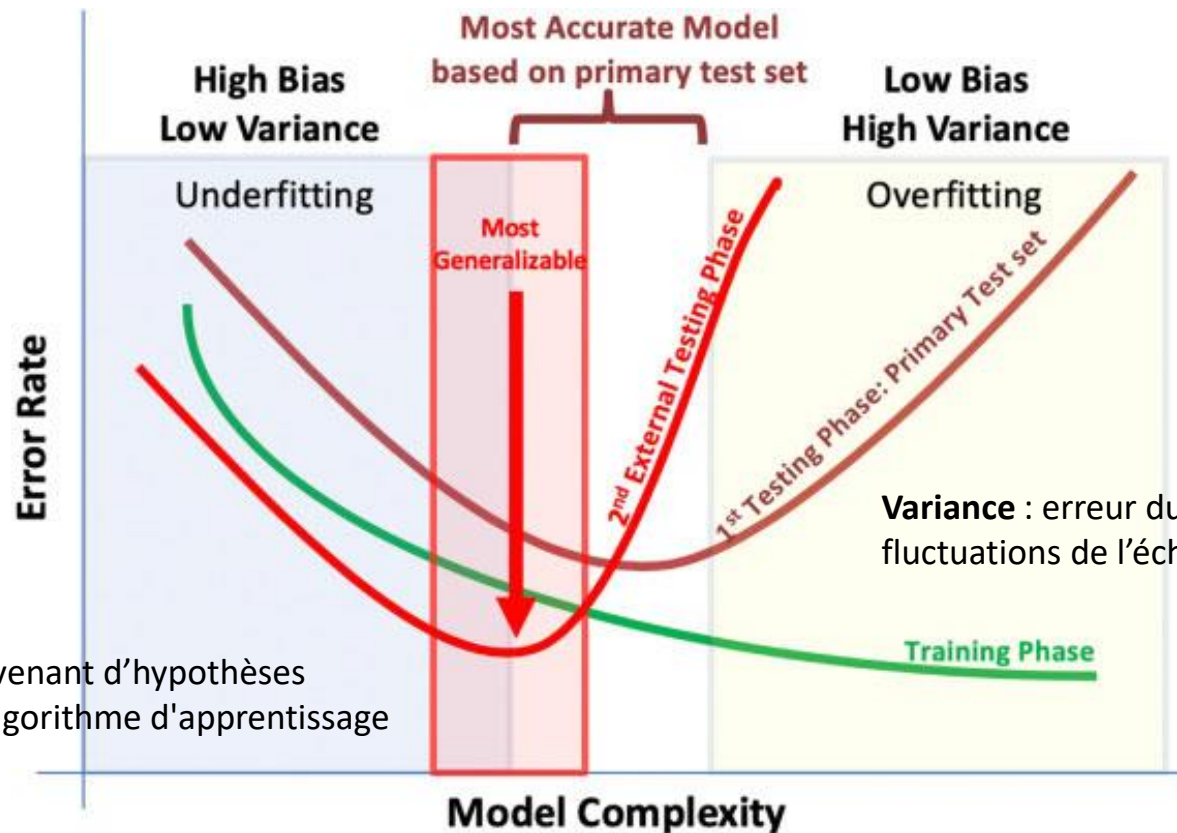


Non spécifique Lobulaire

Données en vrac

- Classifier, identifier des groupes
- Associations
- Réduction de dimensions

3. L'intelligence artificielle en médecine et en pathologie : les concepts Le choix du modèle supervisé : dilemme biais et variance



Biais : erreur provenant d'hypothèses erronées dans l'algorithme d'apprentissage

Variance : erreur due à la sensibilité aux fluctuations de l'échantillon d'apprentissage



4. Exemples en pathologie mammaire



Classer des biopsies mammaires

Thèse de mathématique : Nicolas Jouvin

Analyse non supervisée

Algorithmes : réseaux de neurones

Base de données : comptes rendus et images Jpeg



4.1 Classifier des biopsies mammaires

1^{ère} partie : Analyse de texte non supervisée – clustering non supervisé

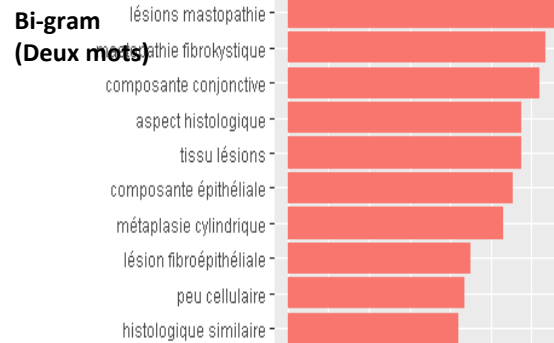
Corps de texte des comptes rendus de biopsies mammaires (2004-2017) : 17 000

```

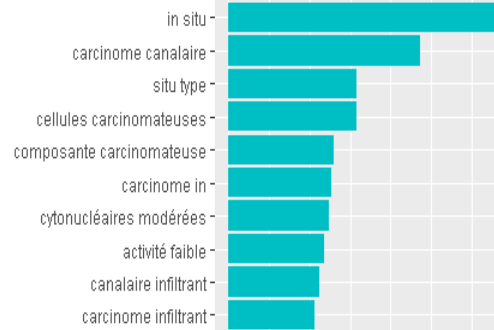
Data: X Compte rendus
Result: Clustering Y Nombre de classes
Input: Q, K, any initializations for Y and β

1  $\mathcal{L} \leftarrow \text{VEM}(X, Y)$ 
2 for  $l \leftarrow 1$  to  $T$  do
3    $Y^{(old)} \leftarrow Y$ 
4   for  $i \leftarrow 1$  to  $N$  do
5     Find  $l$  such that  $Y_{il} = 1$ 
6     for  $q \leftarrow 1$  to  $Q$  do
7       if  $q \neq l$  then
8         Set  $Y_{iq}^{(tmp)} = 1$  and
9          $\mathcal{L}[q] \leftarrow \text{VEM}(X, Y^{(tmp)})$ 
10        Compute:  $\Delta_i(l, q) \leftarrow \mathcal{L}[q] - \mathcal{L}$ 
11      else
12         $\Delta_i(l, q) \leftarrow 0$ 
13      end
14    end
15     $q^* \leftarrow \arg \max_q \Delta_i(l, q)$ 
16    if  $q^* \neq l$  and  $C_i > 1$  then
17      Set  $Y_{iq^*} = 1$ , and  $\mathcal{L} \leftarrow \mathcal{L}[q^*]$ 
18    end
19    if  $Y == Y^{(old)}$  then Break;
20  end
  
```

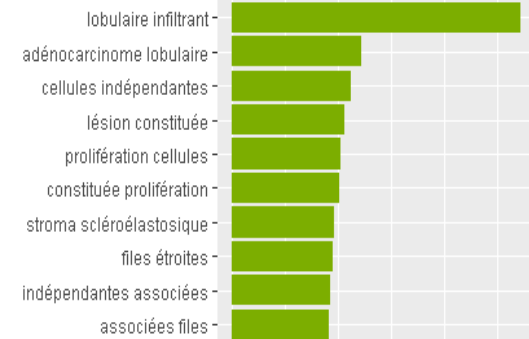
1. Bénin (7000)



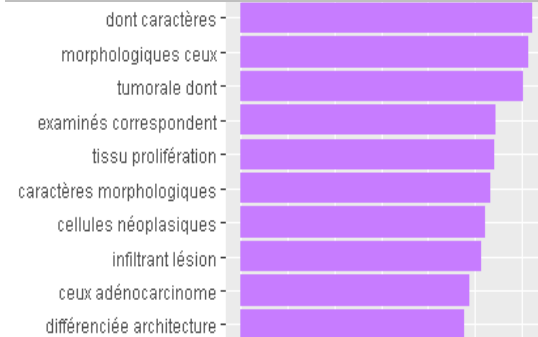
2. In situ (2000)



3. Infiltrant Lobulaire (1000)



4. Infiltrant non spécifique (7000)

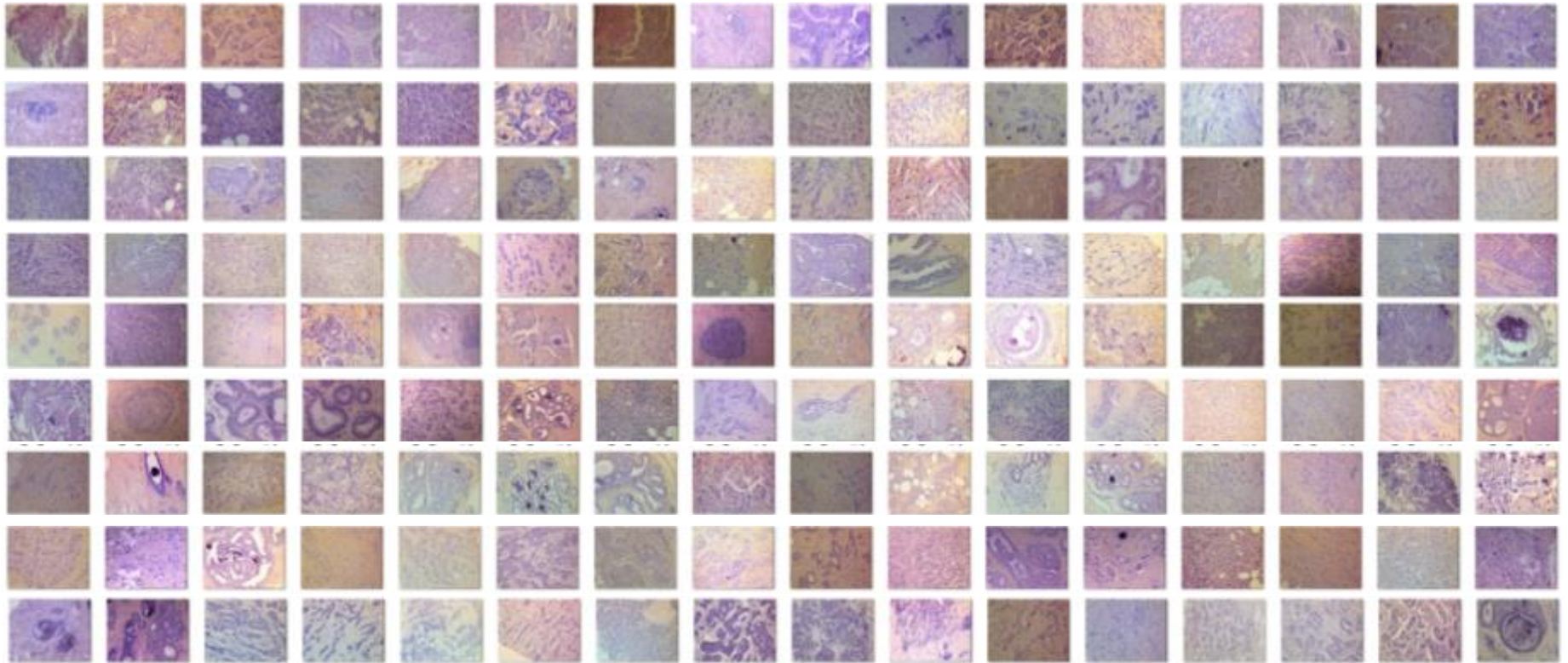


Etude des bases de données non labellisées

4.1 Classer des biopsies mammaires

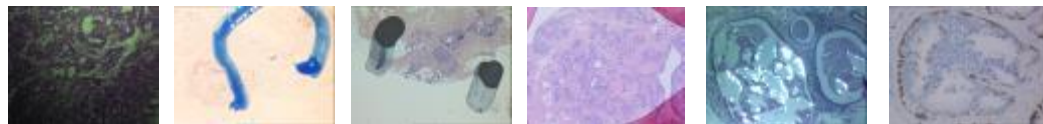
2^{ème} partie (en cours) : Analyse de texte/Images non supervisée – clustering

Images Jpeg de biopsies mammaires (2004-2017) : 50 000 (17 000 cas)



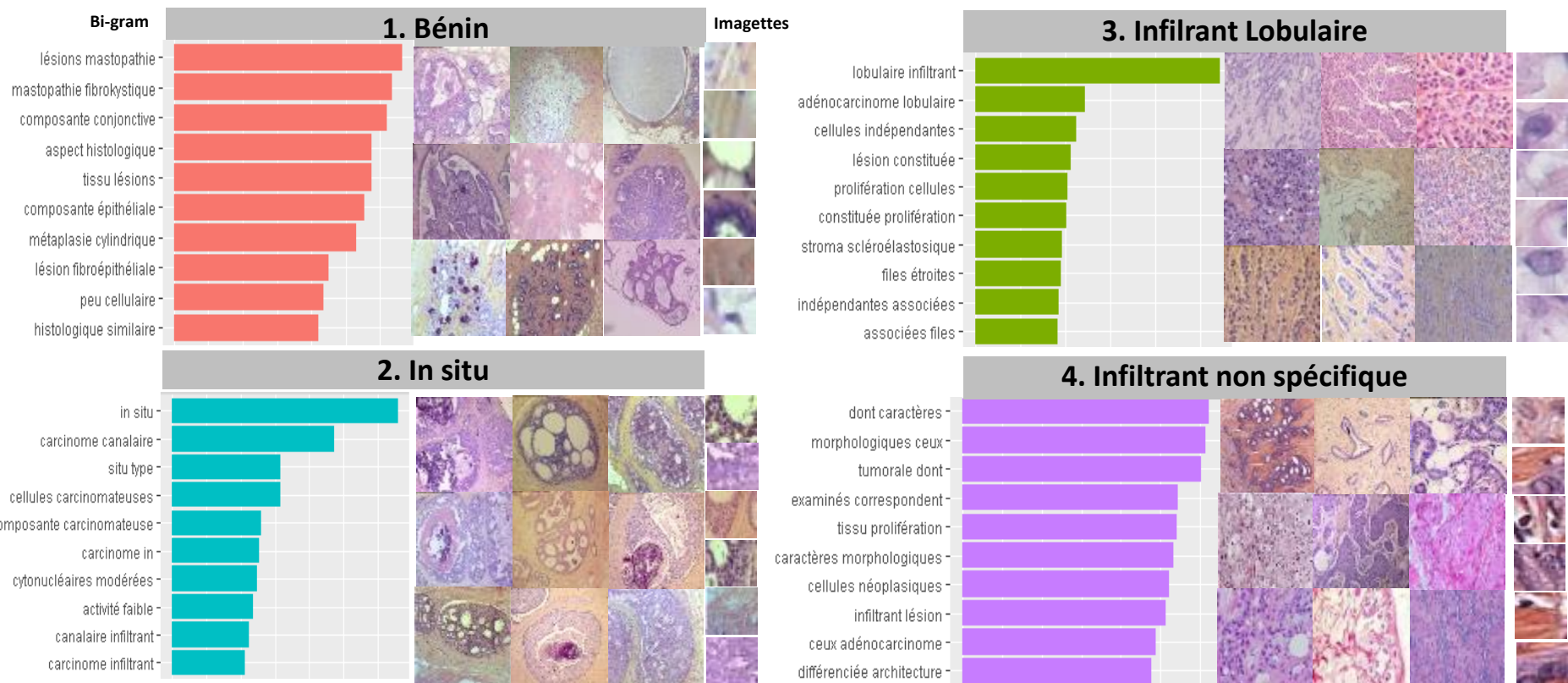
Limites : annotations (type d'images, grandissement), qualité hétérogène, zone d'interêt

Travail de « curation »



4.1 Classifier des biopsies mammaires

2^{ème} partie (en cours) : Analyse de texte/Images non supervisée – clustering



Aide au diagnostic

Classification, facteurs pronostiques et prédictifs de réponse aux traitements

4. Exemples en pathologie mammaire



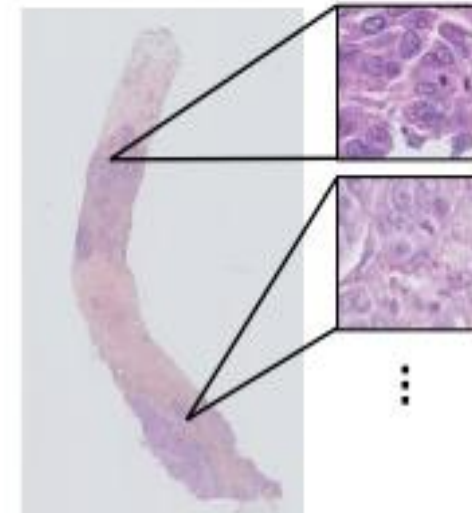
Classer des biopsies mammaires

Thèse de mathématique : Nicolas Jouvin

Analyse non supervisée

Algorithmes : réseaux de neurones

Base de données : comptes rendus et images Jpeg



Prédire la réponse à la chimiothérapie néoadjuvante dans le cancer du sein triple négatif

Thèse au CBIO : Peter Naylor

Analyse supervisée (RCB)

Algorithmes : réseaux de neurones

Base de données : images WSI



4.2. Prédiction de la réponse aux traitements néoadjuvants

Le challenge

Biopsie initiale



Diagnostic
Facteurs pronostiques et prédictifs
De réponse aux traitements

Chimiothérapie Néoadjuvante (NACT)



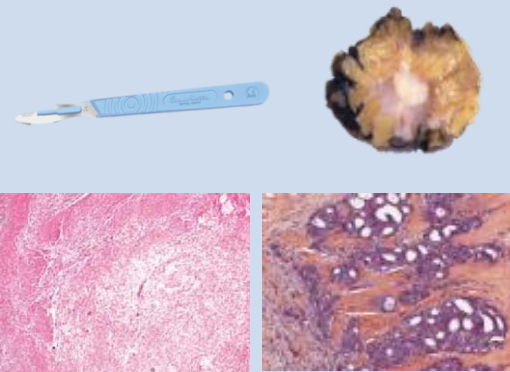
Indication

- Localement avancé (N+, >5cm), inopérable
- HER2, TNBC, grade, IM, Ki67, TILs élevés

Bénéfices

- Diminution de la taille tumorale
- Pronostic (récidive, survie)

Chirurgie



Complete

Absence

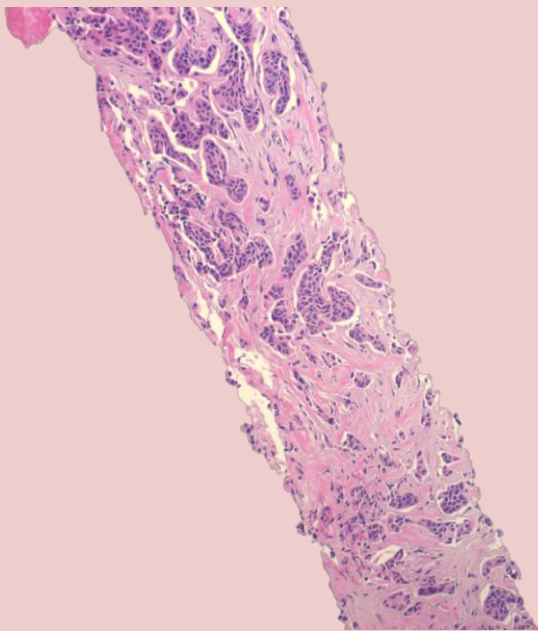
Mesure de la réponse
Residual Cancer Burden (RCB) :
presence/absence de tumeur invasive dans le
lit tumoral/ganglions

Défi : prédire si la tumeur va répondre à la chimiothérapie

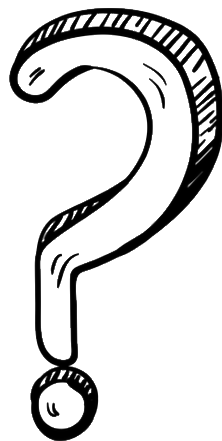
4.2. Prédiction de la réponse aux traitements néoadjuvants

Une approche algorithmique

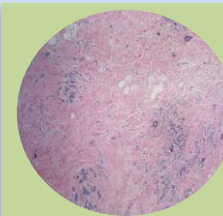
Biopsie mammaire initiale



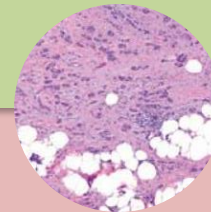
Lames HES



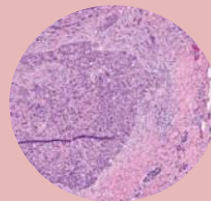
Chirurgie après traitement néoadjuvant



RCB-0 = pCR
complete response



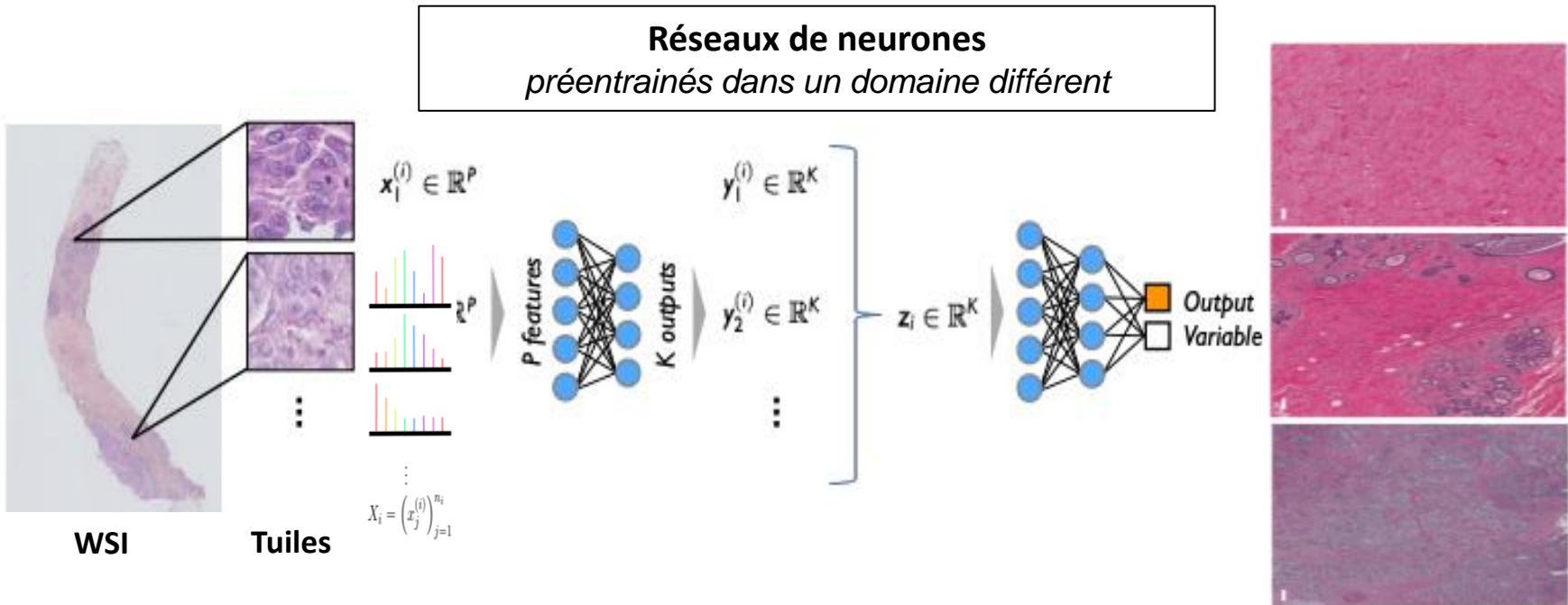
RCB-I
RCB-II
intermediate outcomes



RCB-III
extensive residual disease

4.2. Prédiction de la réponse aux traitements néoadjuvants

Matériels et méthodes



X Entrée : lames HES biopsies mammaires

x_i : vecteurs descripteurs des tuiles

y_i : labels intermédiaires des tuiles

z_i : vecteur descripteur du patient

Y Sortie : label du patient RCB (binarisé) 0-I VS II-III

4.2. Prédiction de la réponse aux traitements néoadjuvants

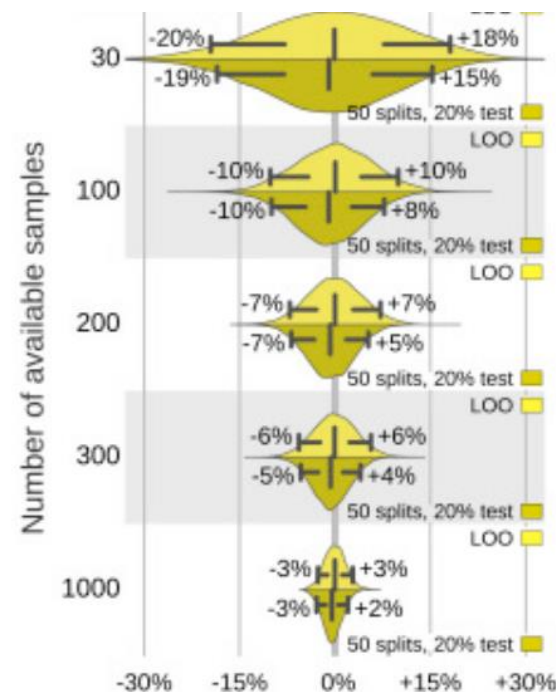
Résultats

Modèle	Manuel*	Réseaux de neurones
120 lames	59,8%	60,6%
180 lames	66,4%	68,0%

Nombre de cas correctement classifiés / nombre de cas total

*Modèle classique « random forest »

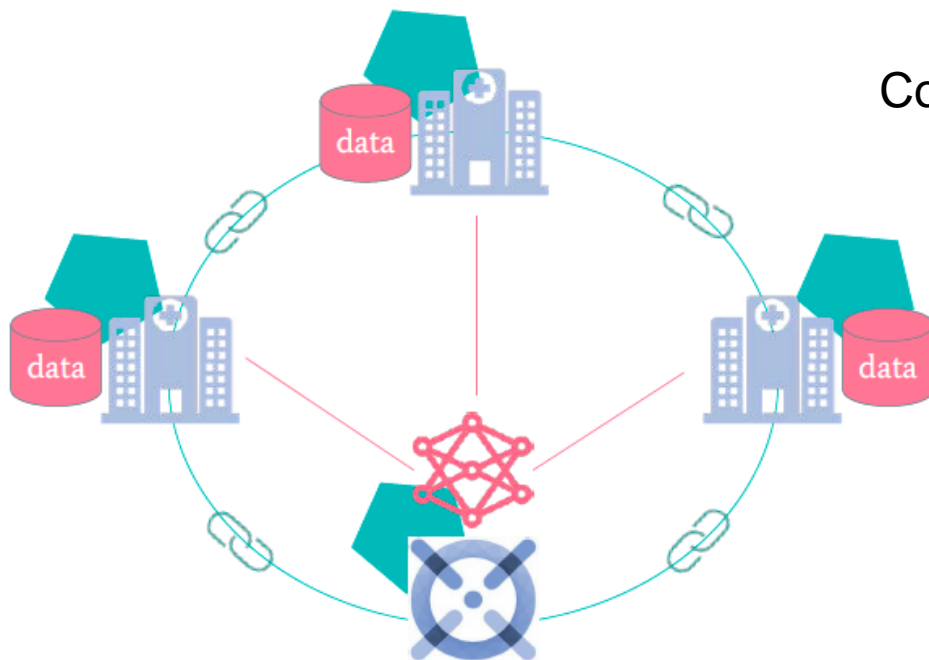
- Index mitotique
- Grade (EE)
- Cellularité tumorale
- % de stroma
- % de lymphocytes



Erreur de prédiction en fonction du nombre d'échantillons

4.2. Prédiction de la réponse aux traitements néoadjuvants

Amélioration : augmenter la base de données et la reproductibilité



Concept : faire voyager les algorithmes

Les données personnelles restent dans les hôpitaux

Les opérations sont enregistrées dans un registre commun

Projet multicentrique informatique et scientifique :

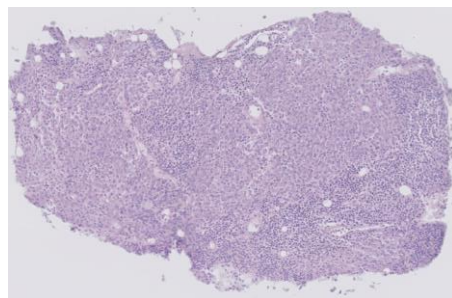
- Partenariat : Curie, CLB, Owkin
- Améliorer la reproductibilité inter-centres
- Augmenter le nombre de cas : 1400 cas

620 cas
120 CLB
500 Curie Paris

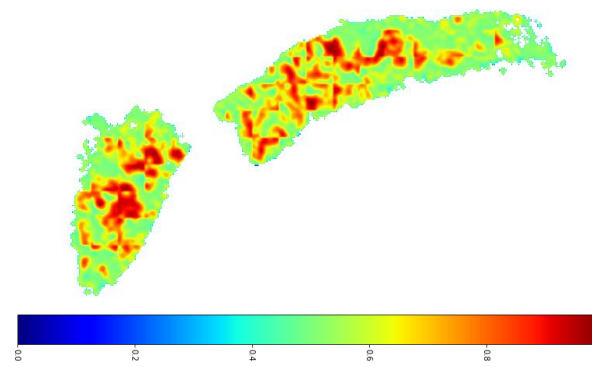
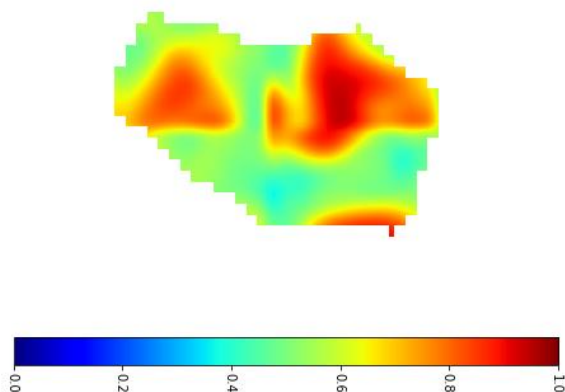
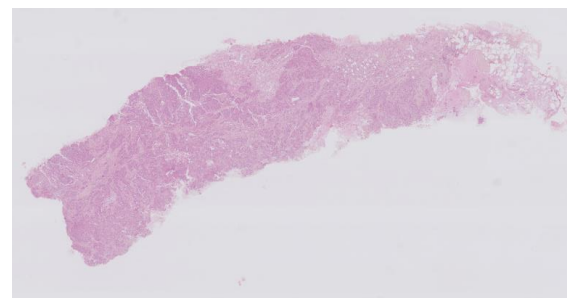
780 cas Curie Paris
Extraction du RCB ?

4.2. Prédiction de la réponse aux traitements néoadjuvants Amélioration : visualisation des résultats (Heatmaps)

Ganglion
métastatique



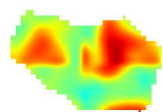
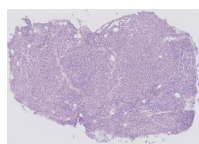
Biopsie mammaire
Carcinome NST



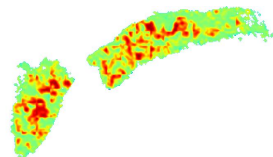
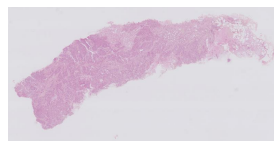
Visualisation de données connues
Identification de cellules tumorales

4.2. Prédiction de la réponse aux traitements néoadjuvants Amélioration : visualisation des résultats (Heatmaps)

Ganglion
métastatique

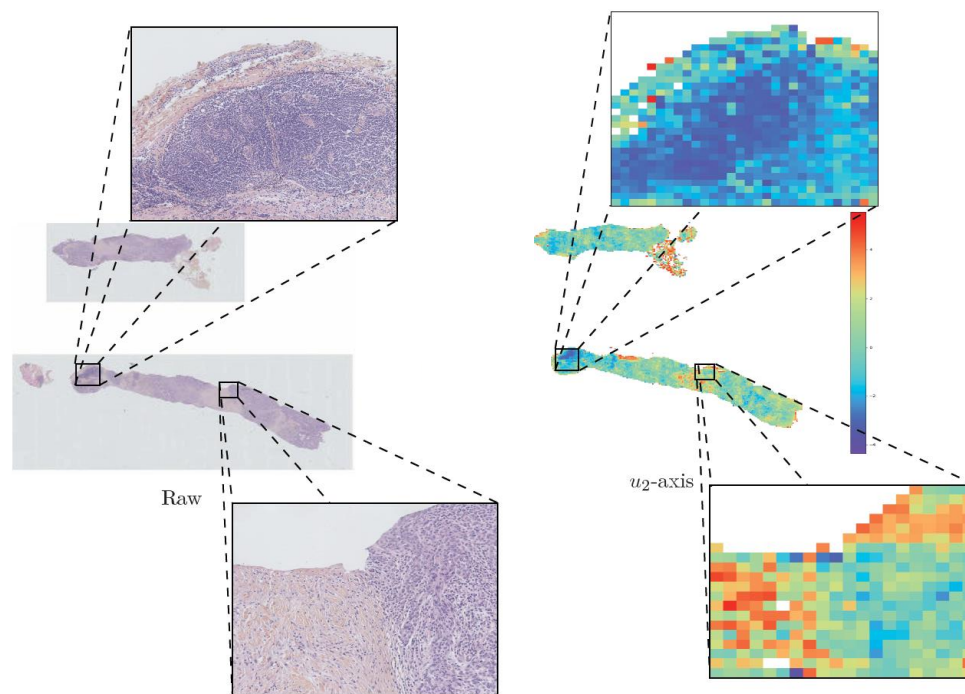


Biopsie mammaire
Carcinome NST



Visualisation de données connues
Identification de cellules tumorales

Biopsie mammaire initiale



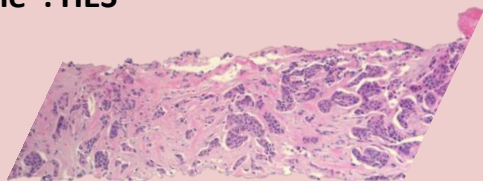
Visualisation de données complexes
Prédiction de la réponse au traitement

4.2. Prédiction de la réponse aux traitements néoadjuvants

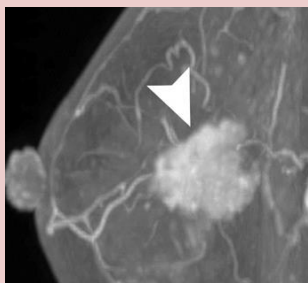
Amélioration : approche multidisciplinaire radiologie et moléculaire

Tumeur initiale

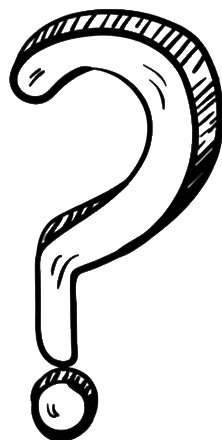
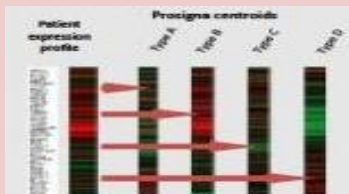
Biopsie : HES



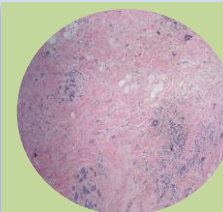
Imagerie initial : TNM radiologique



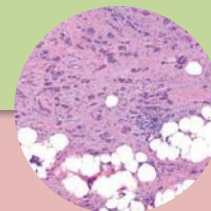
Données moléculaires : signature transcriptomique



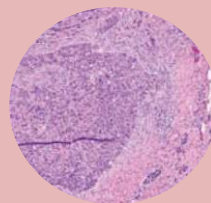
Chirurgie après traitement néoadjuvant



RCB-0 = pCR
complete response



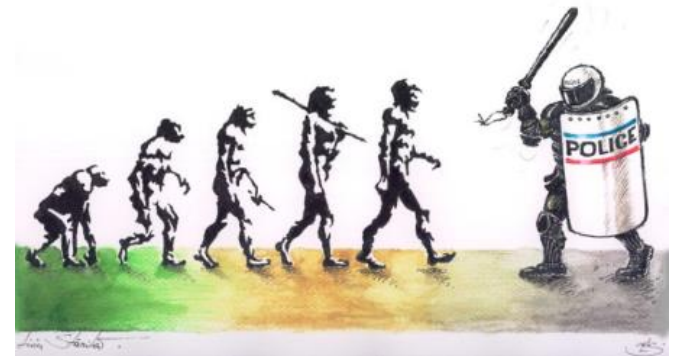
RCB-I
RCB-II
intermediate outcomes



RCB-III
extensive residual disease

Limites

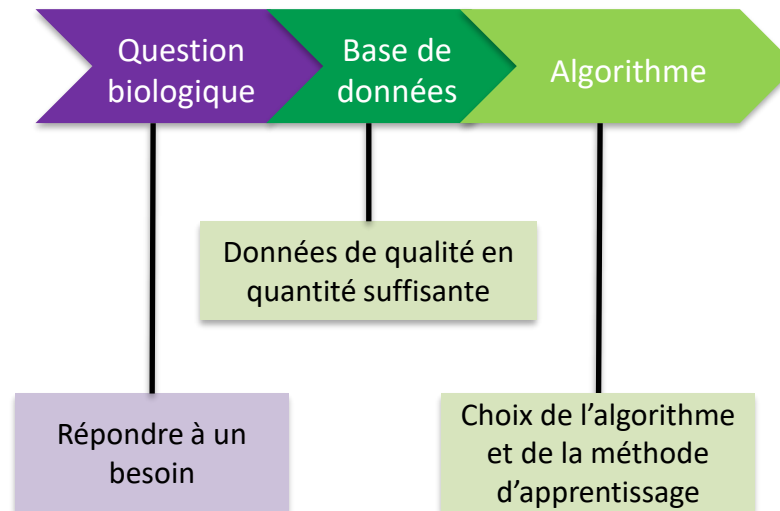
- Technique : standardisation
 - Fixation : formol
 - Coloration : automate (HES VS HE)
 - Numérisation : coût de l'équipement, du stockage (WSI ~ Go)
- Bases de données : qualité et quantité
 - Discordance inter-observateur
 - Annotation : chronophage, fastidieuse, choix de l'échelle
- Algorithmes : effet boîte noire des résultats – visualisation
- Juridique
 - Confidentialité des données
 - Certification



Conclusion : comment répondre à une question biologique ?

Equipe pluridisciplinaire

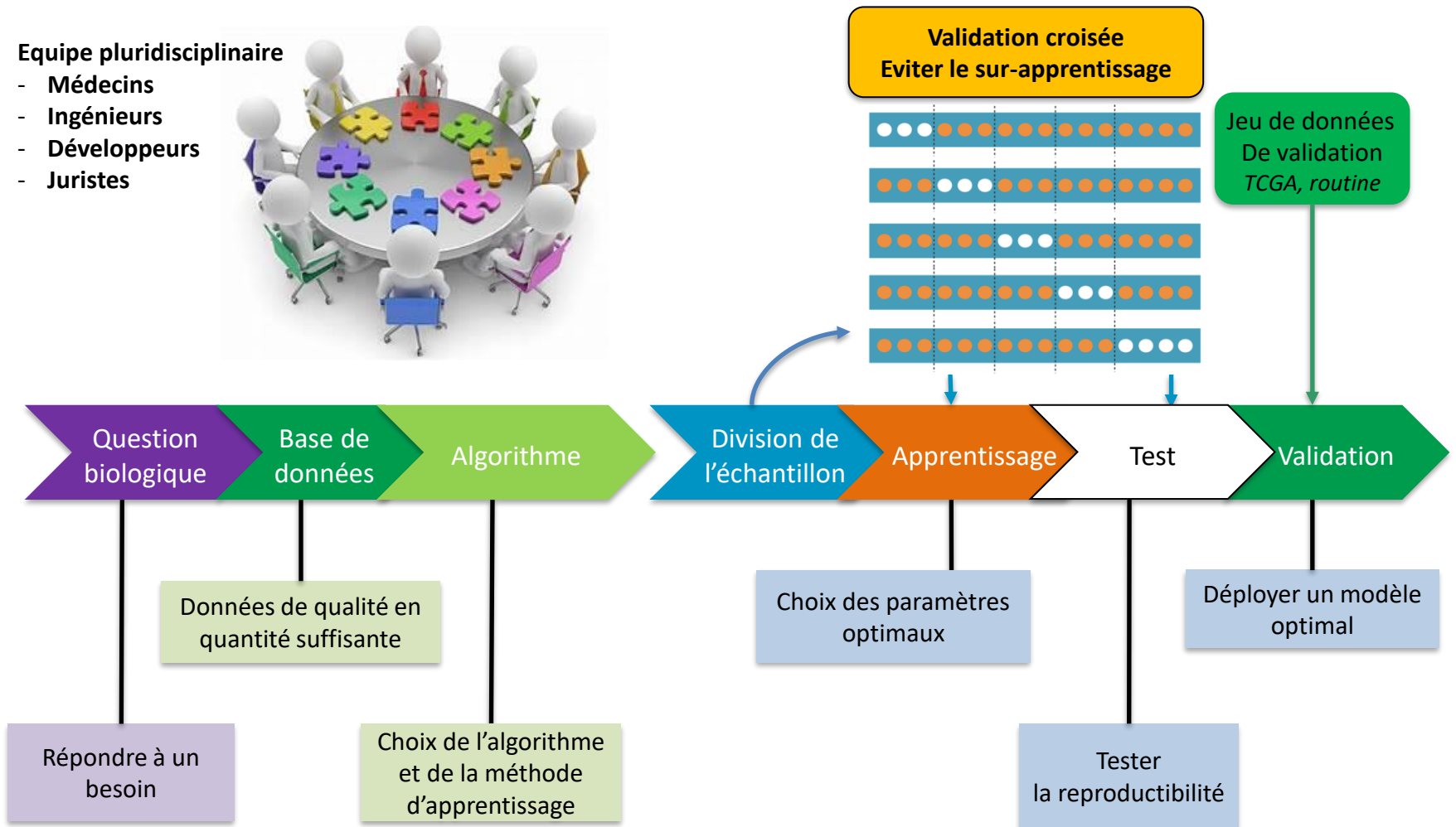
- Médecins
- Ingénieurs
- Développeurs
- Juristes



Conclusion : comment répondre à une question biologique ?

Equipe pluridisciplinaire

- Médecins
- Ingénieurs
- Développeurs
- Juristes



Merci



Qiu Zhijie, Map of Technological Ethics 2017