

IA en biologie et santé: de la donnée à la décision clinique

Transformer des données complexes en décisions médicales grâce à l'IA

Marwa AFNOUCH

Post-doc CNRS

 UMR 1283-8199
METAB-OMICS
Multi-omique et physiopathologie des maladies métaboliques

 Université
de Lille

 Institut
Pasteur
de Lille

 Funded by the
European Union

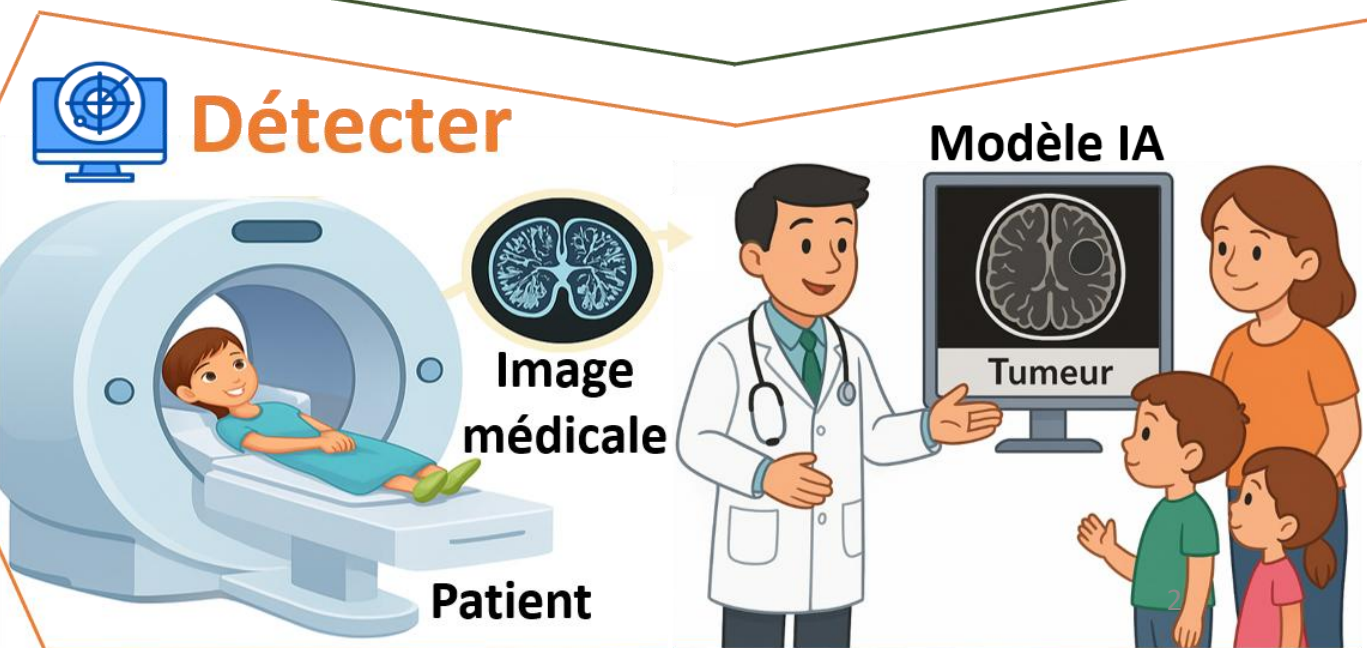
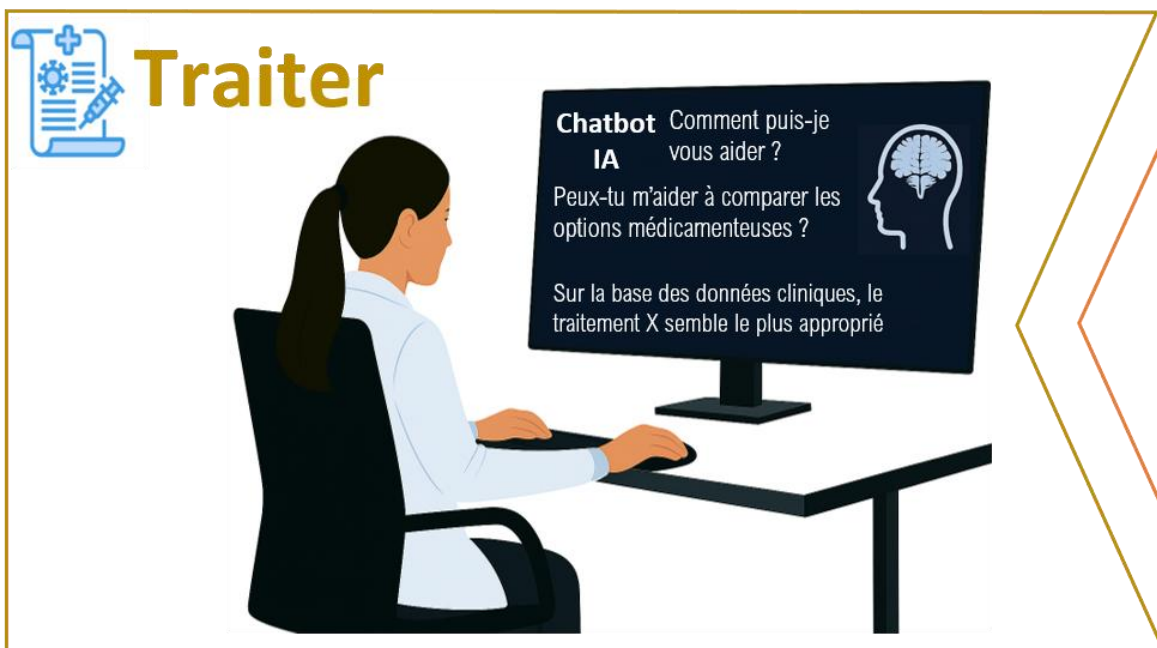
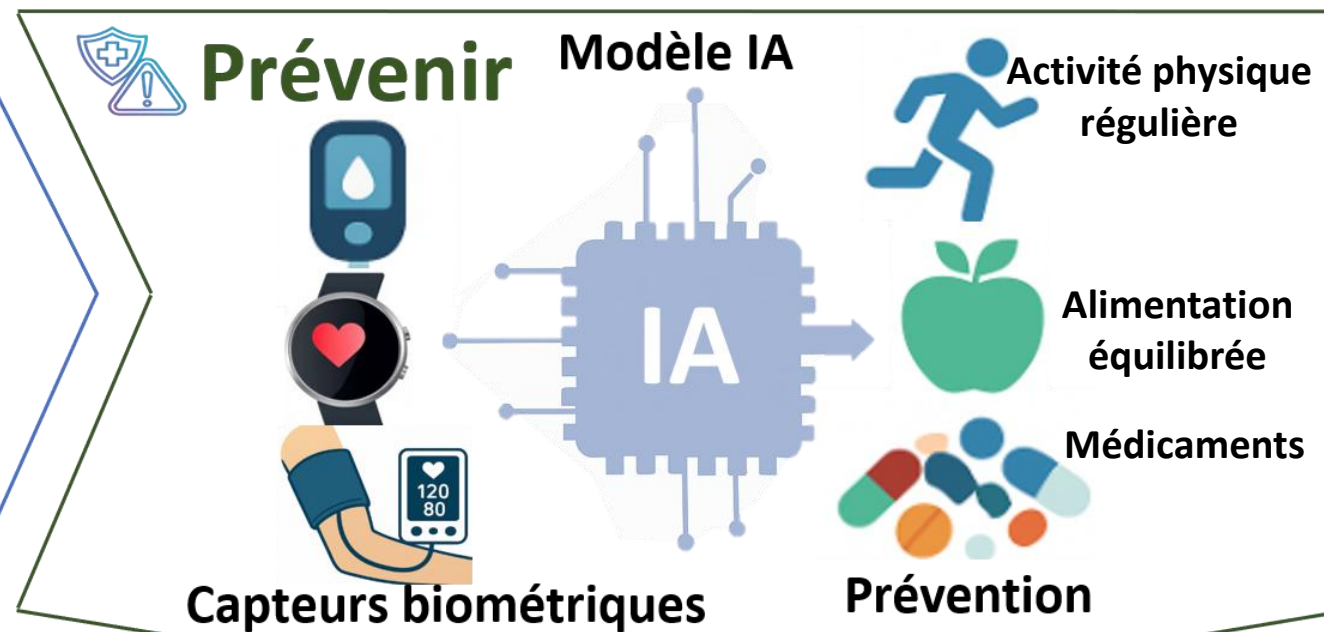
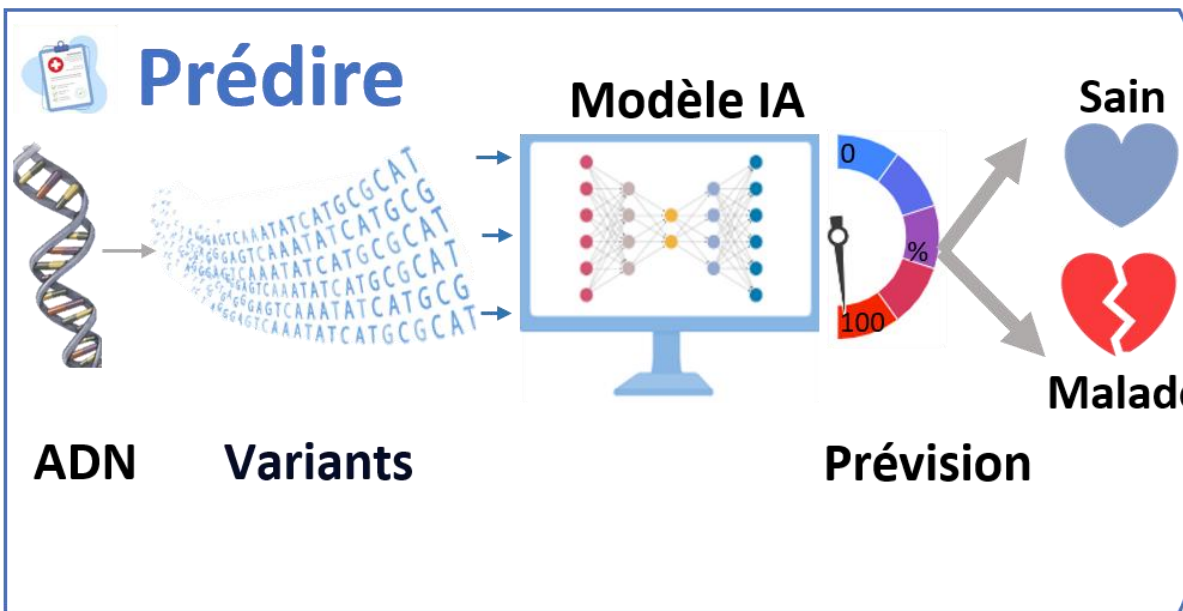
 CHU
LILLE

 Inserm

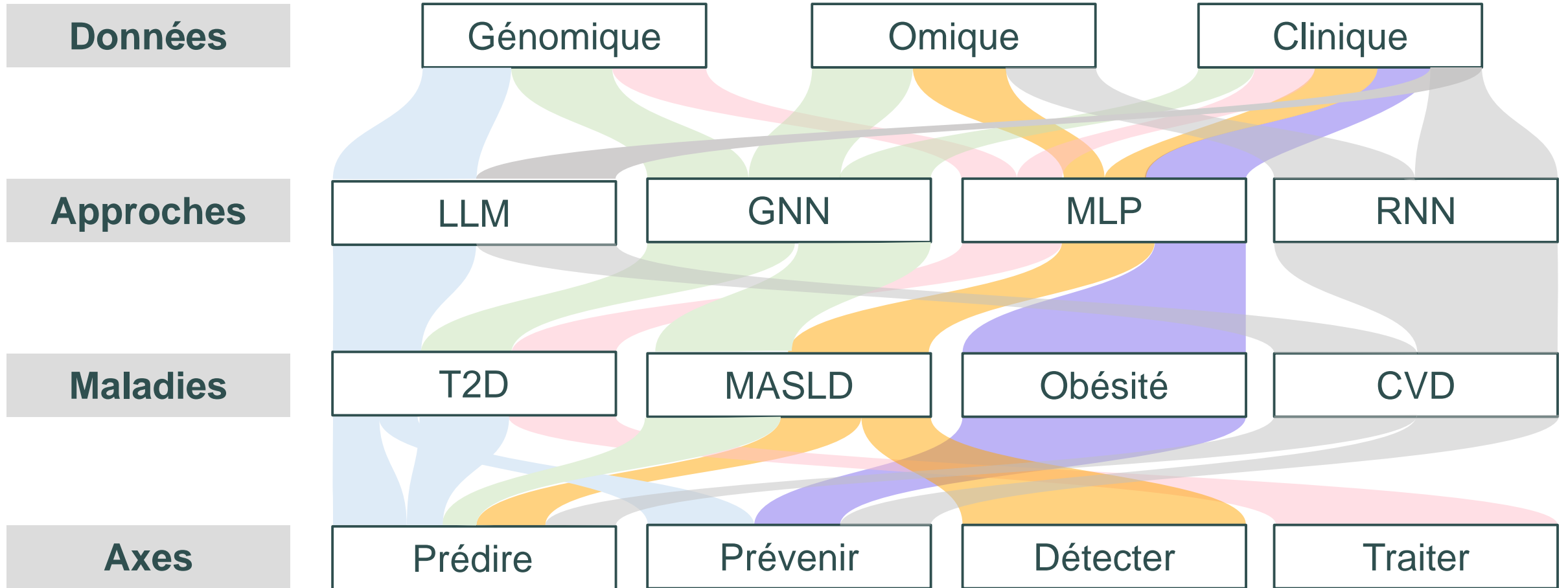
 cnrs

 PreciDIAB
National Center for
Precision Diabetic Medicine

L'IA en santé et biologie



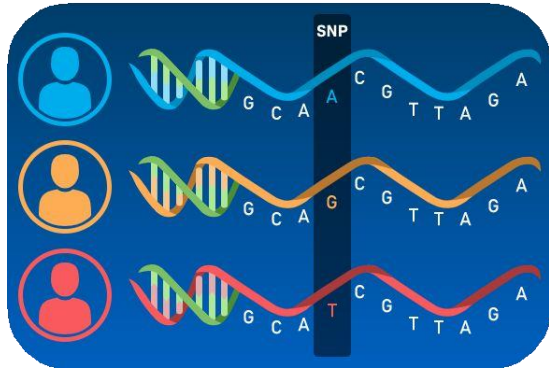
L'IA en santé et biologie



Prédire

Prédiction du risque de diabète de type 2 (T2D)

Input

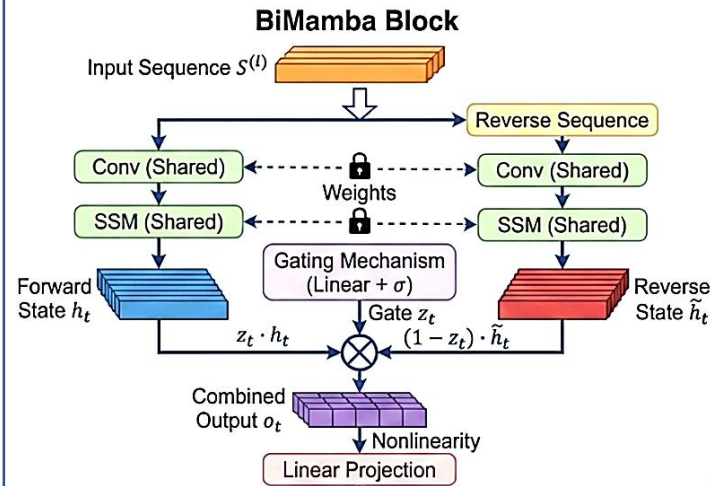


Données génomiques

...AGGCTAGGATATCGATAAGCTGACTGAT...

SNPs ($\sim 10^6$)
Cohorte WTCCC
 $p \gg n$
16 Go VRAM, 6 GPU

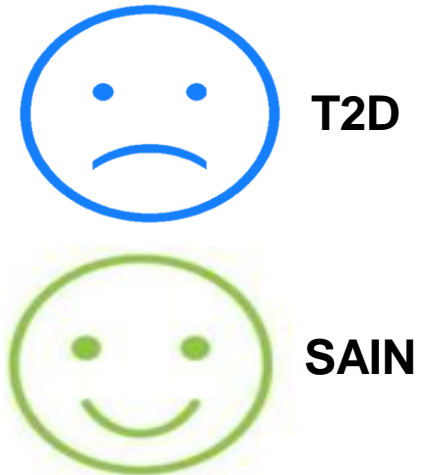
Model



Transformer $O(n^2)$ ✗
BiMamba $O(n)$ ✓

$\times 10^8$ paramètres

Output

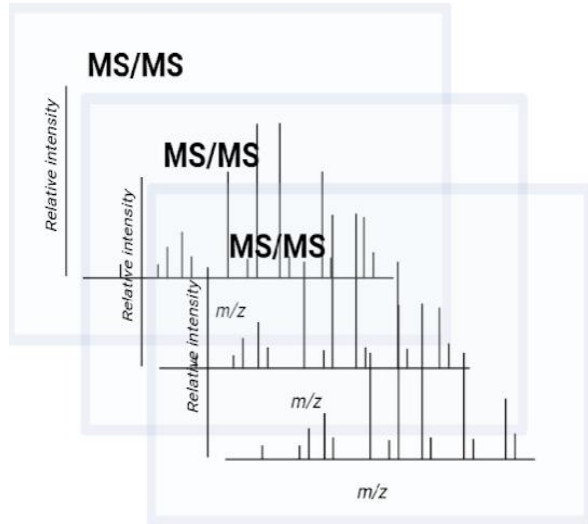


AUC
0.70

🔍 SHAP · top SNPs explicatifs

Prédire des structures moléculaires à partir de spectres

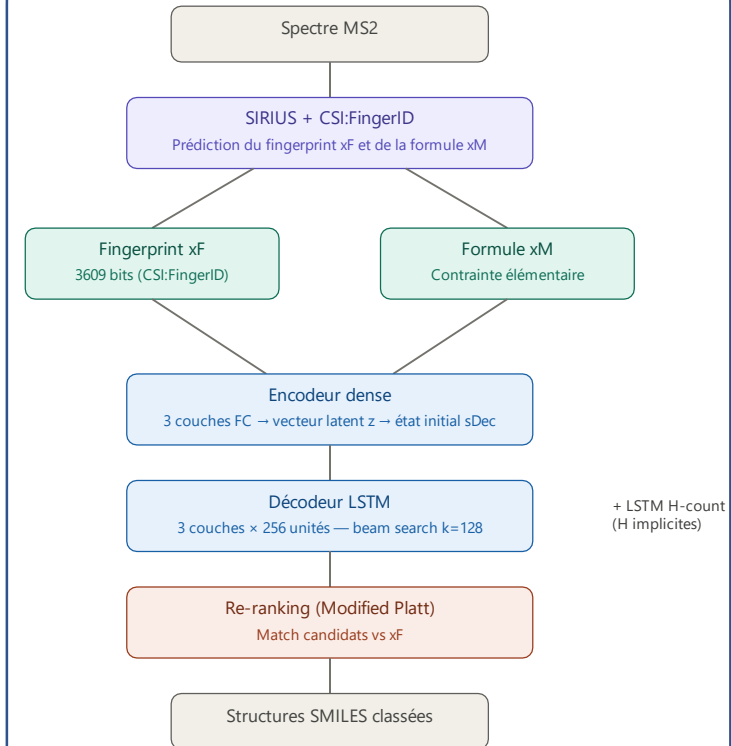
Input



Spectres de
masse(plasma/sang)
~20 000 spectres

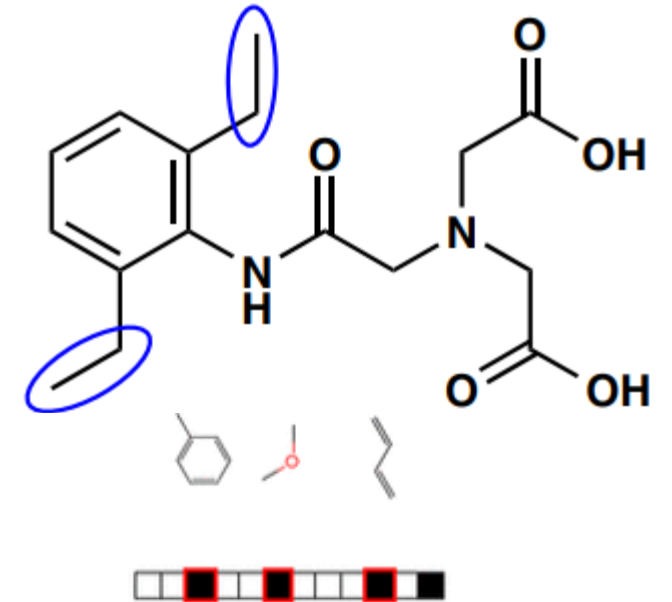
Data split :80% train, 20% test

Model



Modèle de génération guidée par
fingerprint moléculaire
(MSNovelist)

Output



Structures moléculaires
candidates

Prévenir

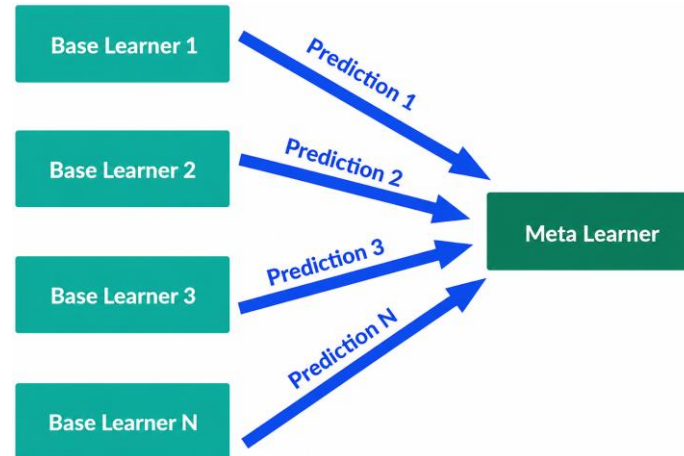
Prévention du surpoids et de l'obésité de l'enfant

Input



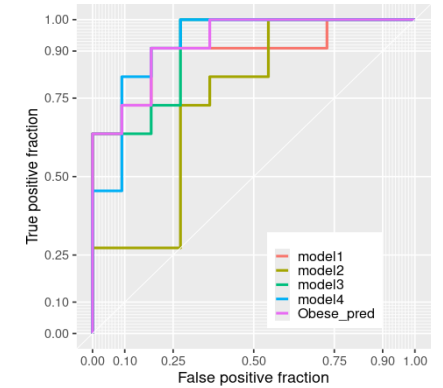
- Mode de vie
- Facteurs comportementaux
- Données cliniques

Model

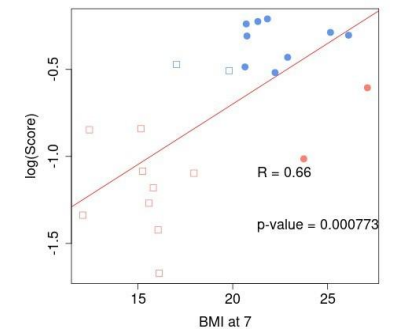


Ensemble de 4 MLP
Réseau de neurones simple

Output

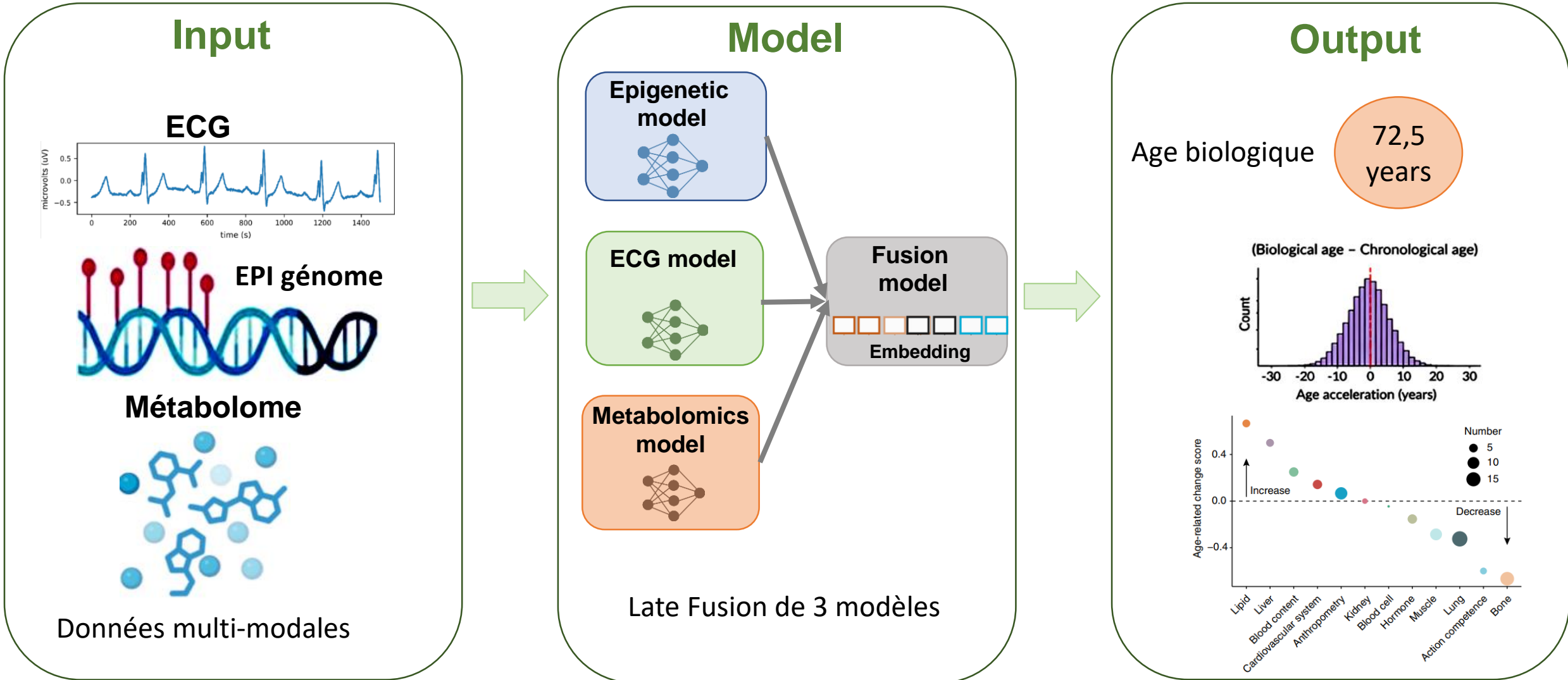


Score de risque



Trajectoire de IMC

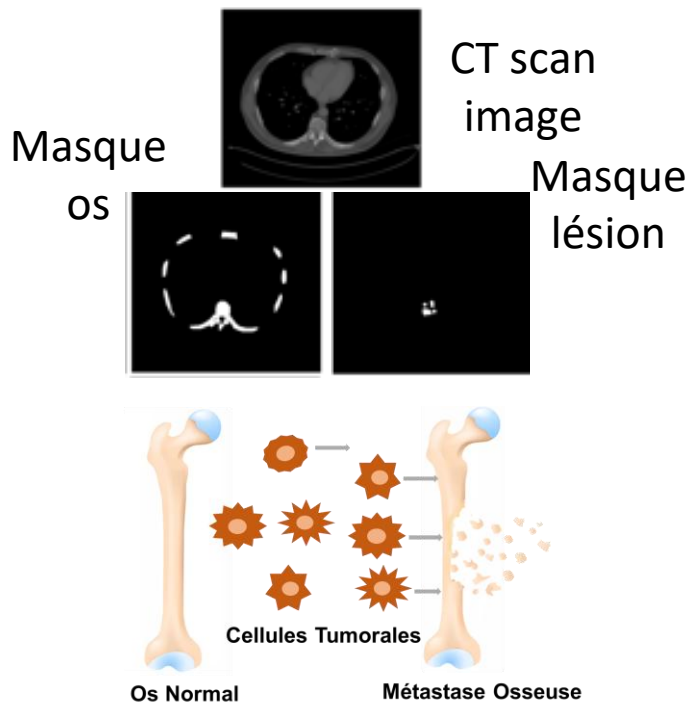
Estimation multimodale de l'âge biologique pour la prévention



Détecter

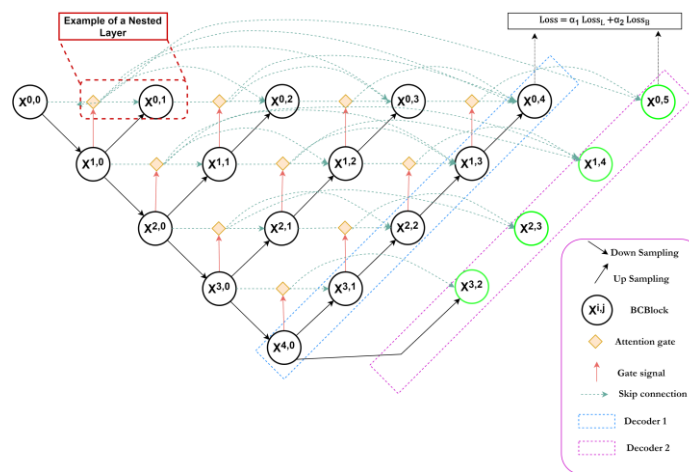
Détection des métastases osseuses

Input



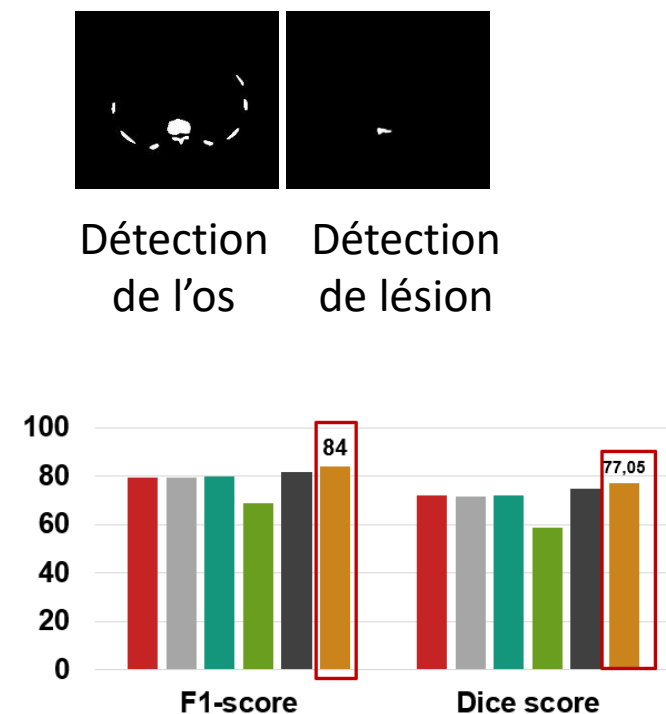
BM-seg Dataset avec 100K images

Model



Hybrid AttUnet++
CNN avec double décodeurs

Output



Traiter

Prédiction de traitements antidiabétiques

Input



Données cliniques



Données génétiques

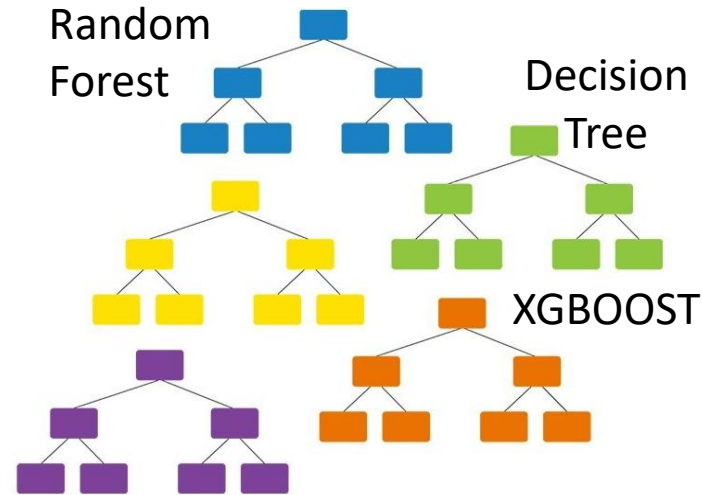
3 cohortes:

Precidiag: 3130 patients

Corbeil : 1997 patients

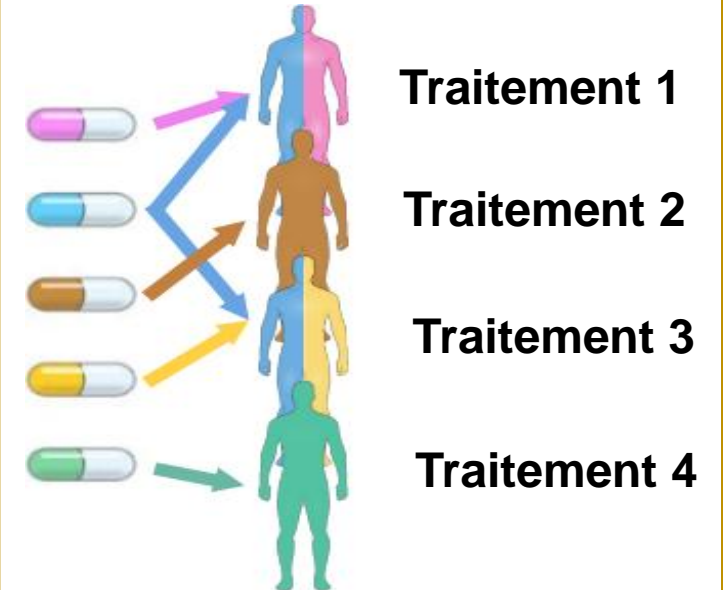
UKBB: 9312 patients

Model



Comparaison des modèles de
Machine Learning

Output



Médecine de précision

Les défis de l'IA en santé

La réalité des données médicales

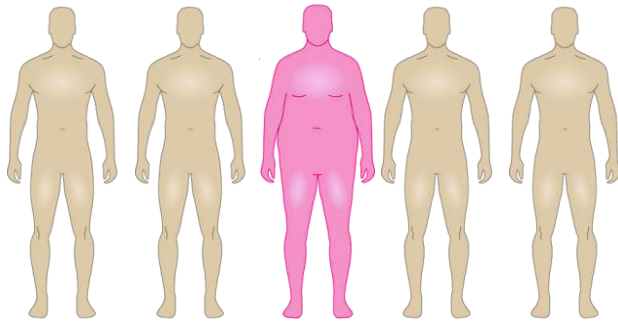
Formats hétérogènes

Formats hétérogènes non interopérables



Biais dans les données

Les jeux d'apprentissage sont déséquilibrés



Qualité des données

Champs libres, abréviations, fautes de frappe...



QUALITÉ DES DONNÉES FAIBLE
Bruit - Incohérences - Erreurs
Manquants - Redondances

1254	01/02/23	abc	--
1254A	--	3000	999/999
...

Biais de cohorte

Un modèle entraîné sur l'hôpital A échoue sur l'hôpital B

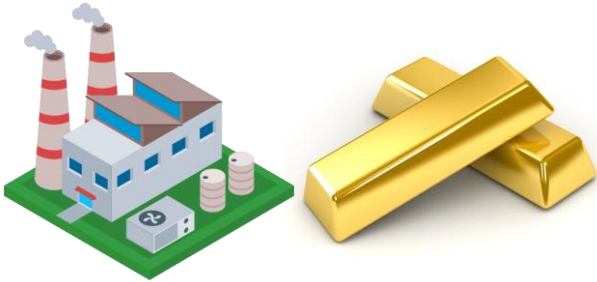


Contrainte technique et accès aux données

Contraintes matérielles

Utilisation coûteuse

- GPU / HPC nécessaires
- Mémoire élevée
- Temps de calcul long



Barrière technique

Compétences interdisciplinaires

- Environnements complexes
- Formation nécessaire
- Outils multiples



Accès sécurisé aux données

Données sensibles

- Accès restreint (RGPD)
- Serveurs distants (HPC)



Reproductibilité

Dépendance aux environnements

L'application des modèles et l'interprétation des résultats peut dépendre du contexte



Équipe & collaborations



Marc-Emmanuel
DUMAS
DR CNRS



Amna ABDELGADER
-KHAMIS
DR Univ Lille



Nicolas
GAMBARDELLA
DR CNRS



Zahra
ELHAMRAOUI
IR Univ Lille



Smaïn FETTEM
IR Univ Lille



Francesc PUIG
CASTELLVI
Post-Doc INSERM



Salla AKERBLOM
Doctorante CNRS



Omaira BINAN
Doctorante Univ Lille