

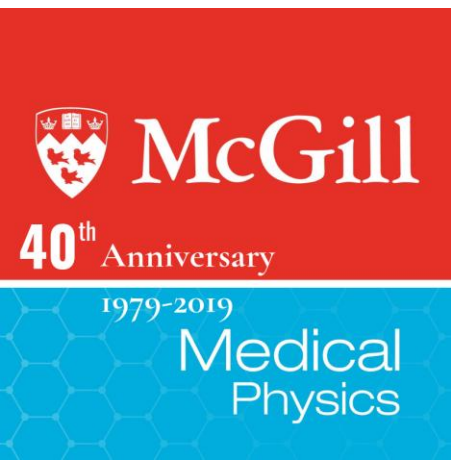
IA multimodale pour améliorer la stratification des risques et la qualité des soins contre le cancer au Québec

Présentatrice: Yujing Zou^{1,2,3}

Directrice: Shirin Abbasinejad Enger^{1,2,3}

Collaborateurs: Juan Duran^{1,2,3}, Harry Glickman^{1,2,3}, Farhood Farahnak^{1,2,3}

MSSS -Congrès PQC, Nov 20th, 2025



¹ Unité de physique médicale, Département d'oncologie, Université McGill, Montréal, Canada

² Institut québécois d'intelligence artificielle – Mila (Montreal Institute for Learning Algorithms), Montréal, Canada

³ Institut Lady Davis de recherches médicales, Hôpital général juif, Université McGill, Montréal (Québec), Canada

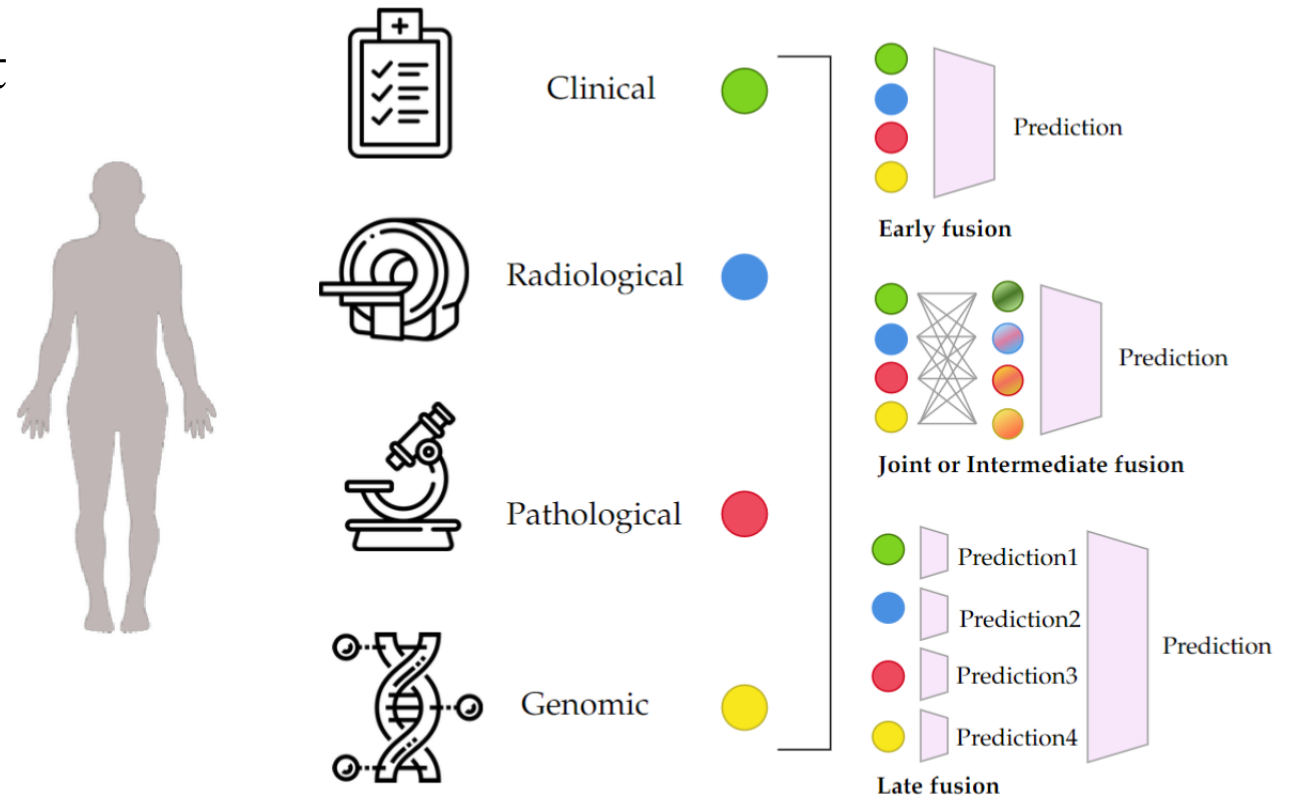


Motivation: Modèle de prédiction multimodal des résultats du traitement

Développement d'un modèle de prédiction des résultats de traitement spécifique au patient, basé sur l'apprentissage profond, capable de prédire avec précision les combinaisons de traitements optimales pour les patients atteints de cancer à partir de données multimodales.

Pourquoi la prédiction pré-traitement est essentielle

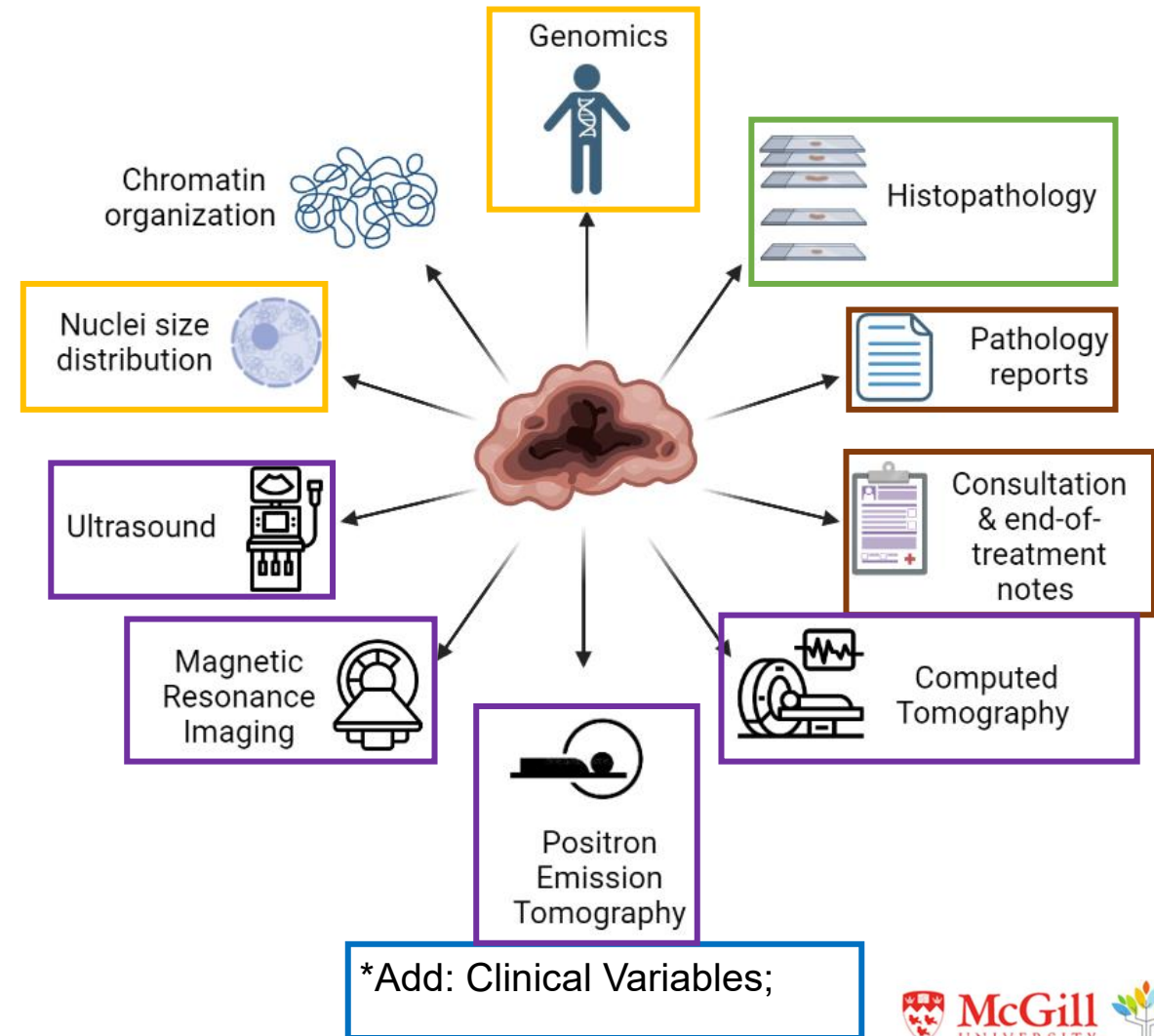
- Patients similaires → évolutions différentes
- Préparation manuelle lourde
- Variabilité entre cliniciens/établissements
- Confidentialité stricte : aucun cloud
- Prédictions fiables nécessaires avant traitement



Problème: Fragmentation et silos

Le défi québécois : données dispersées, duplication d'efforts

- Données éparpillées (clinique, imagerie, pathologie, rapports)
- Duplication des efforts de curation
- Modalités manquantes fréquentes
- Besoin d'un schéma commun et d'outils automatisés



Objectifs du projet: Construire un pipeline prédictif complet, local et répliquable

Importance de la prédiction pré-traitement et de la curation de données à grande échelle

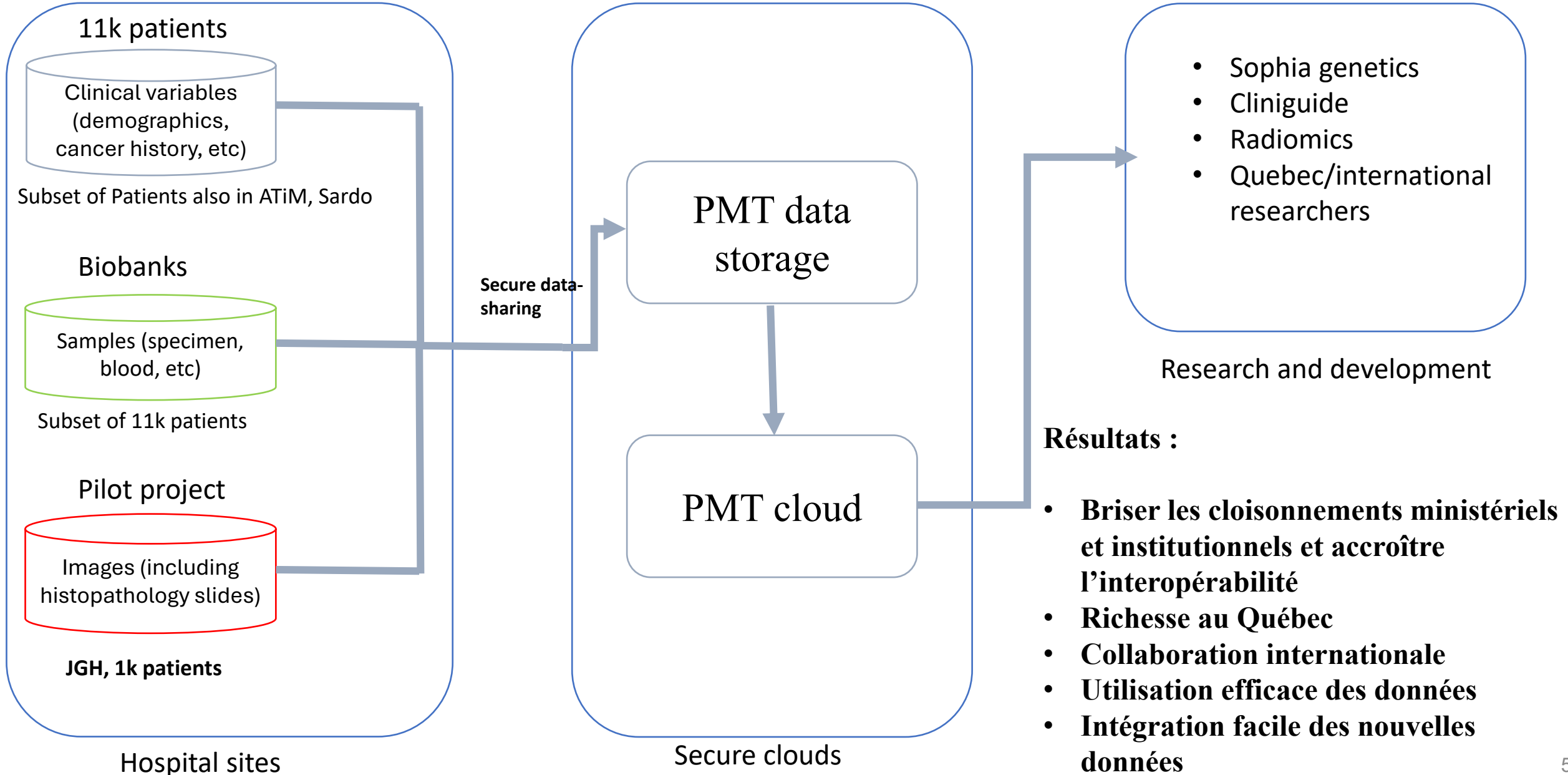
Structurer automatiquement les notes oncologiques via un LLM local

Combiner clinique, texte, imagerie TEP-TDM et pathologie numérique

Intégrer les scores dans les comités tumoraux

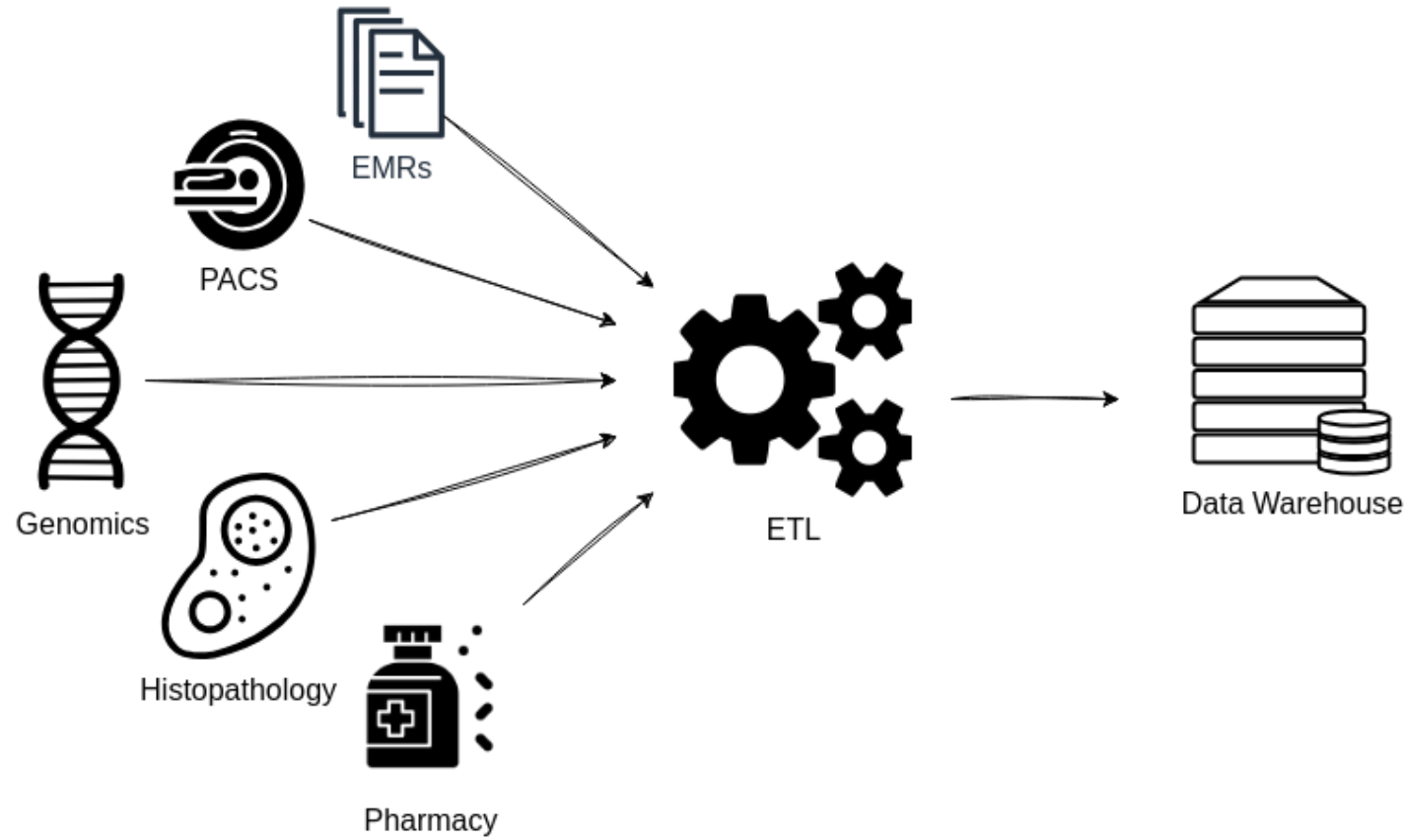
Offrir un cadre reproductible pour d'autres établissements du Québec

Projet pilote de l'Hôpital général juif: curation de données multimodales en oncologie



Automatisation de l'ingestion sécurisée des données (ETL hospitalier)

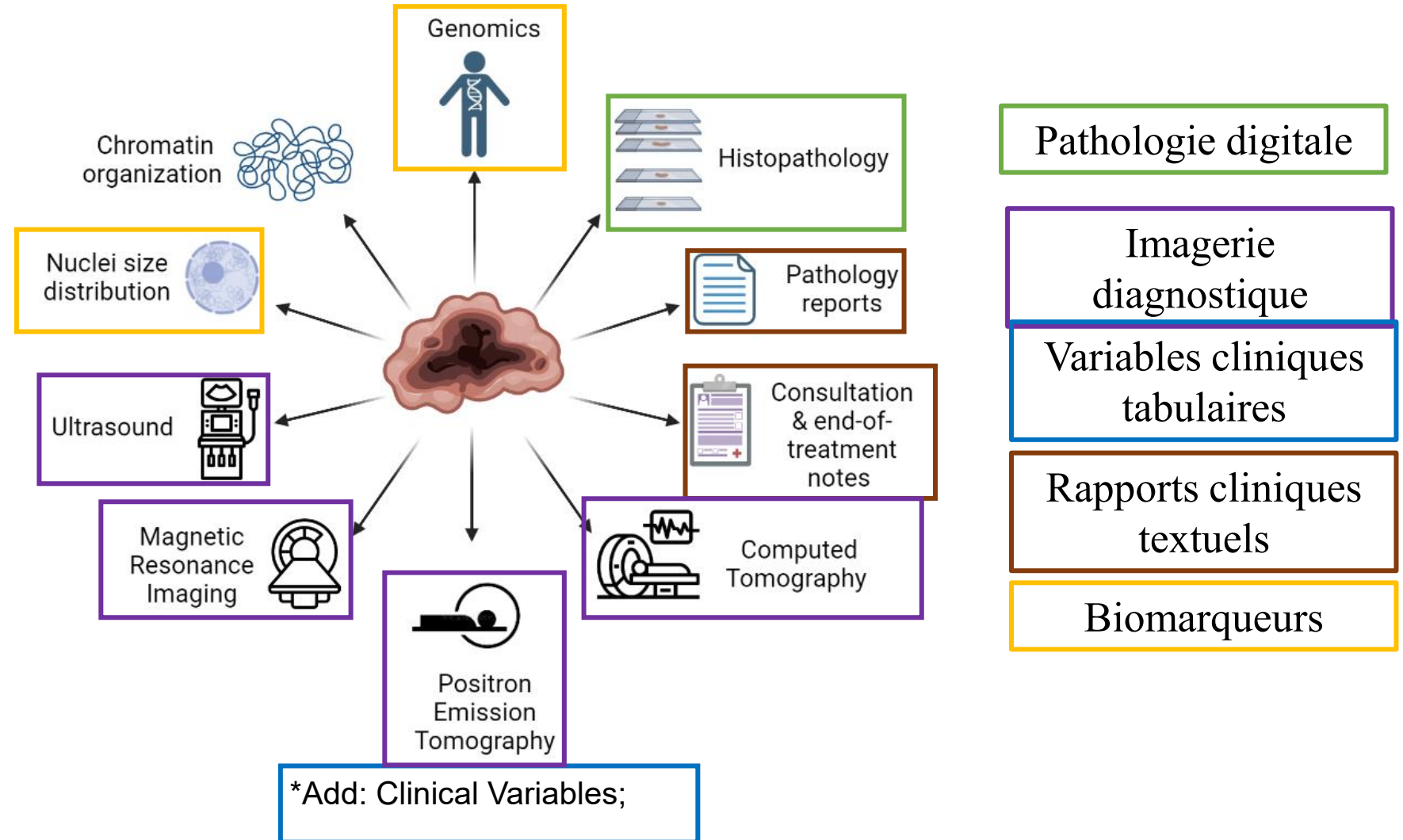
- Extraire, transformer et charger des données provenant de multiples sources hospitalières.
- Planifier le traitement pendant les heures creuses afin de minimiser la charge du serveur.
- Anonymiser automatiquement les données des patients pour assurer la confidentialité.
- S'adapter facilement aux nouvelles cohortes.
- Réduire l'intervention humaine et les coûts associés.



OPHELIA-JGH: Apprentissage et intégration de données hétérogènes pour le pronostic oncologique - Cohorte pilote canadienne multimodale (radiologie, pathologie, rapports, données cliniques)

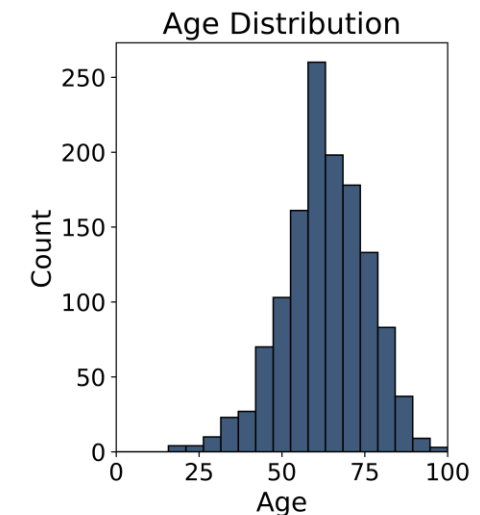
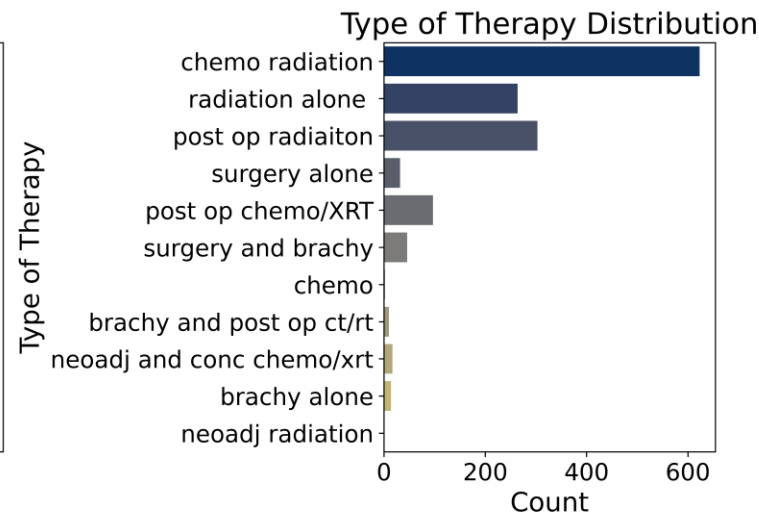
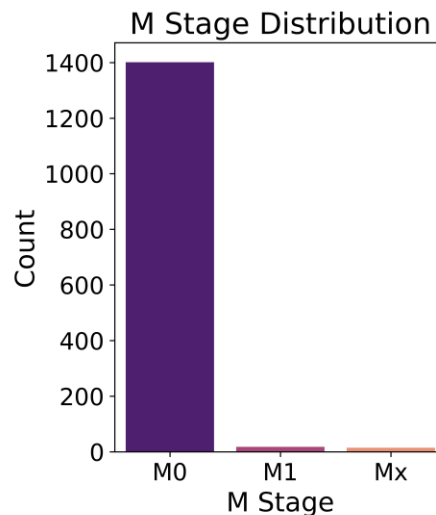
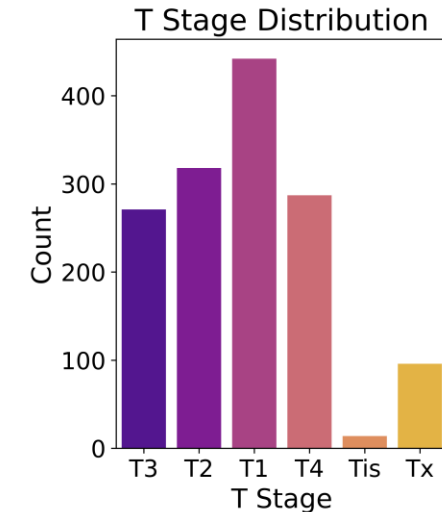
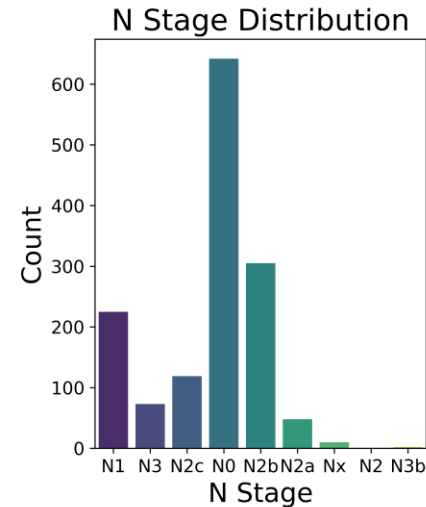
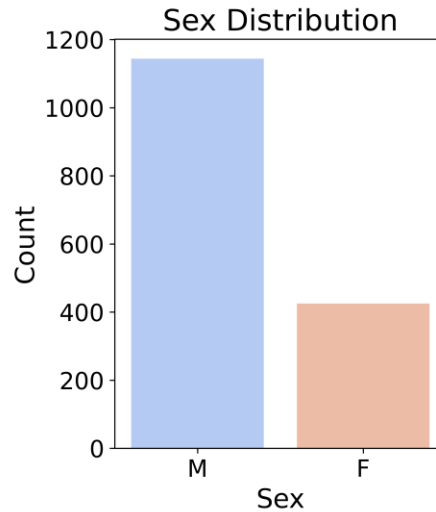
Gestion à grande échelle de données multimodales en temps réel sur les cancers de la tête et du cou à l'Hôpital général juif

Un effort pluriannuel pour constituer un ensemble de données multimodales complet comprenant des patients atteints de cancers de la tête et du cou traités rétrospectivement de 1994 à 2023 (~1 500 patients).



OPHELIA-JGH: Apprentissage et intégration de données hétérogènes pour le pronostic oncologique - Cohorte pilote canadienne multimodale (radiologie, pathologie, rapports, données cliniques)

Données tête et cou – variables cliniques : sexe, stade TNM, type de traitement, âge, etc.



OPHELIA-JGH: Apprentissage et intégration de données hétérogènes pour le pronostic oncologique - Cohorte pilote canadienne multimodale (radiologie, pathologie, rapports, données cliniques)

Curation à grande échelle des données pan-cancer de pathologie numérique et moléculaire du JGH

Biopsies diagnostiques
2018–2023 provenant des
archives du département
de pathologie (~15,000
WSIs).



**Sélection
manuelle : 3–5
WSI H&E +
toutes les WSI
moléculaires**



**Numérisation Aperio
(pathologie numérique)**



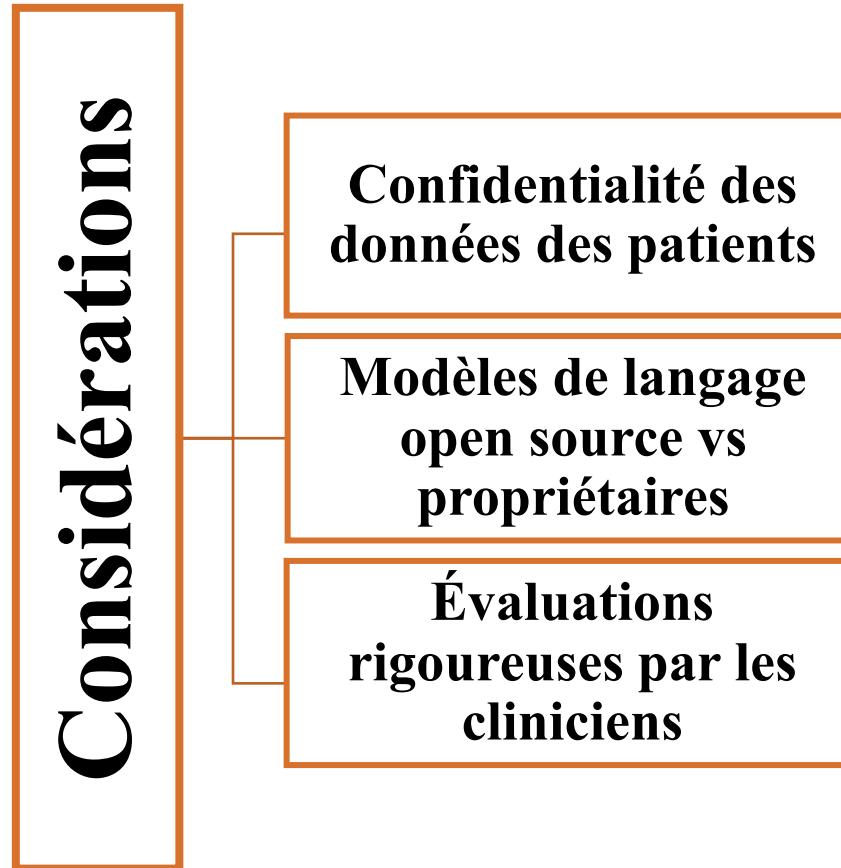
Slide digitization



66 To de stockage local pour les
WSI numérisées (H&E et
immunohistochimie)



Extraction de données pronostiques via un modèle de langage confidentiel: évaluation rétrospective clinicien–IA en oncologie tête et cou



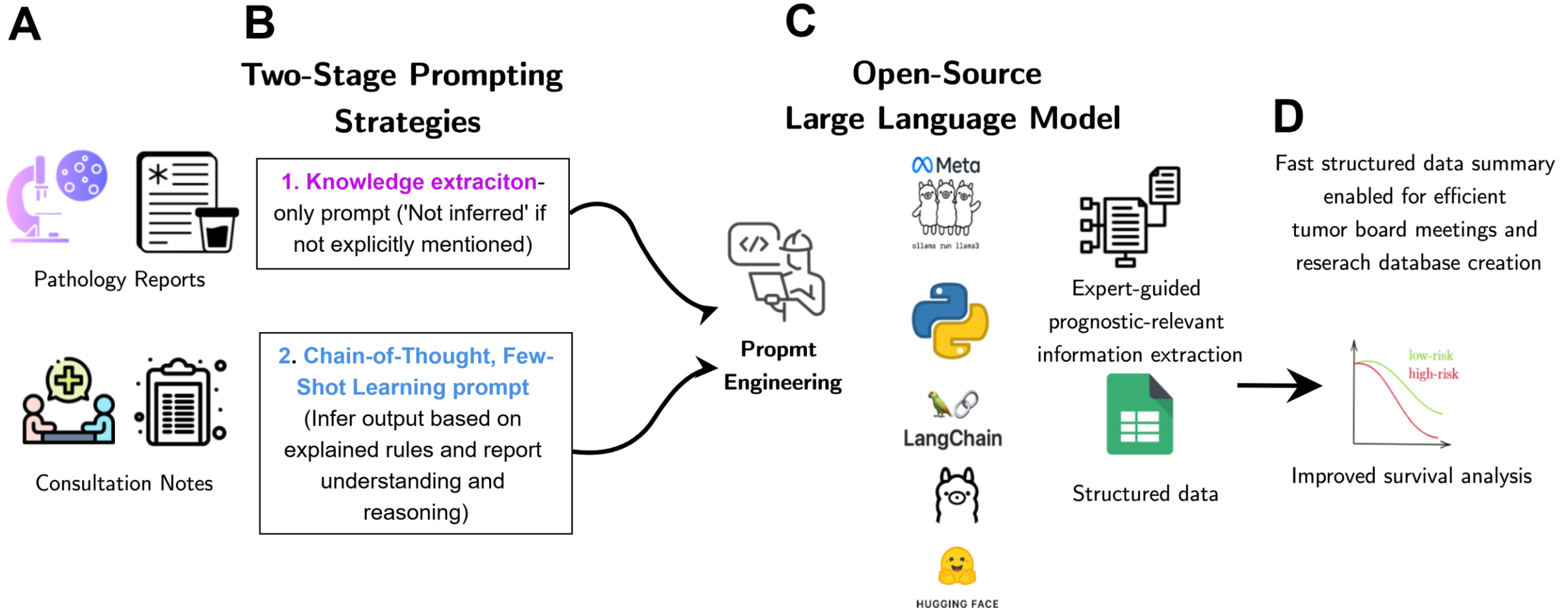
Objectif: Développer un pipeline LLM préservant la confidentialité pour l'extraction de données pronostiques en oncologie de la tête et du cou

•Jeanselme et.al.: Review of Language Models for Survival Analysis ([link](#))

•Lammert et.al.: Expert-Guided Large Language Models for Clinical Decision Support in Precision Oncology ([link](#))

Extraction de données pronostiques à l'aide d'un modèle de langage préservant la confidentialité

LLM itératif guidé par des experts pour les rapports cliniques : notes de consultation et rapports de pathologie



Extraction de données pronostiques à l'aide d'un modèle de langage préservant la confidentialité

Analyses multi-évaluateurs : définitions de la matrice de confusion pour les champs extraits par LLM comparés à la révision clinique : Vrai positif (VP), Vrai négatif (VN), Faux positif (FP), Faux négatif (FN)

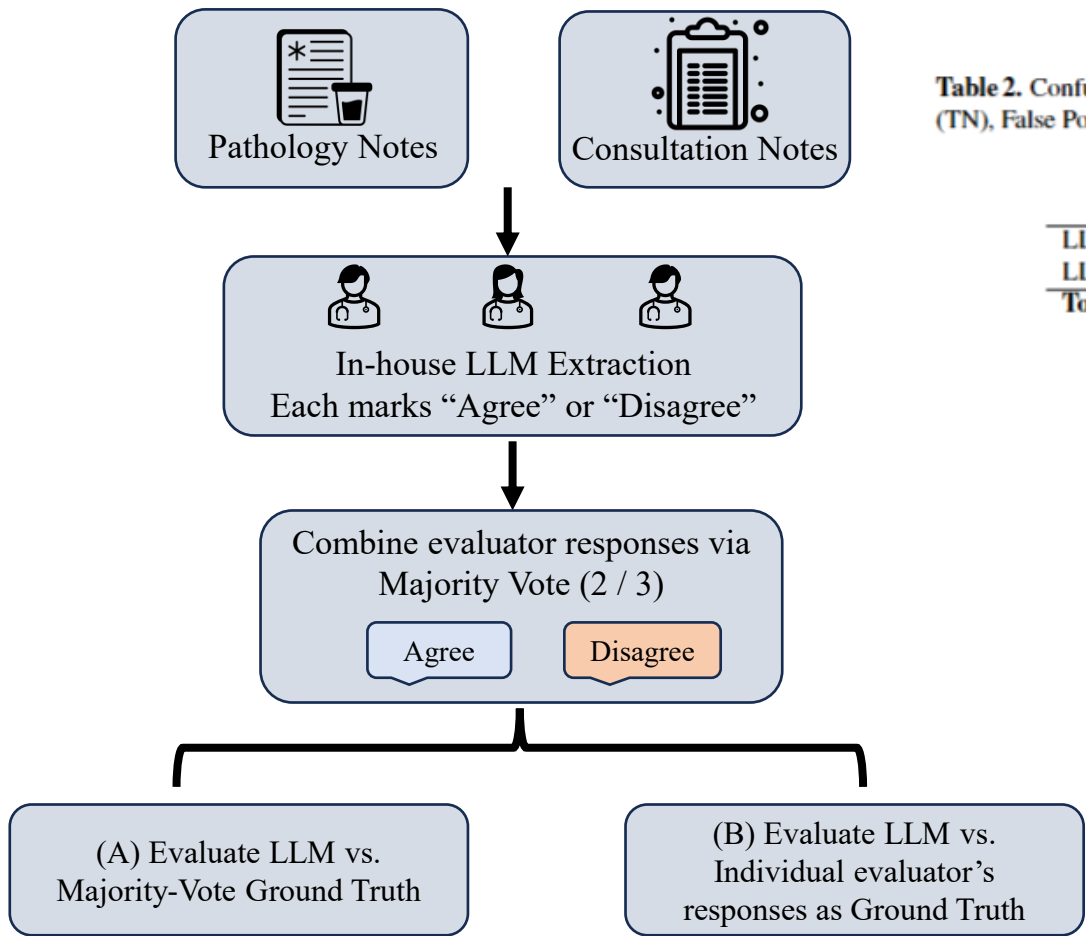


Table 2. Confusion Matrix definition for LLM Extracted Fields vs. Clinician Review: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

	Clinician Evaluation		Totals
	Agree (Yes)	Disagree (No)	
LLM: Some value (Yes)	TP	FP	(TP+FP)
LLM: Not inferred (No)	TN	FN	(TN+FN)
Totals	(TP+TN)	(FP+FN)	<i>n</i> = 1500 responses per evaluator

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{2}$$

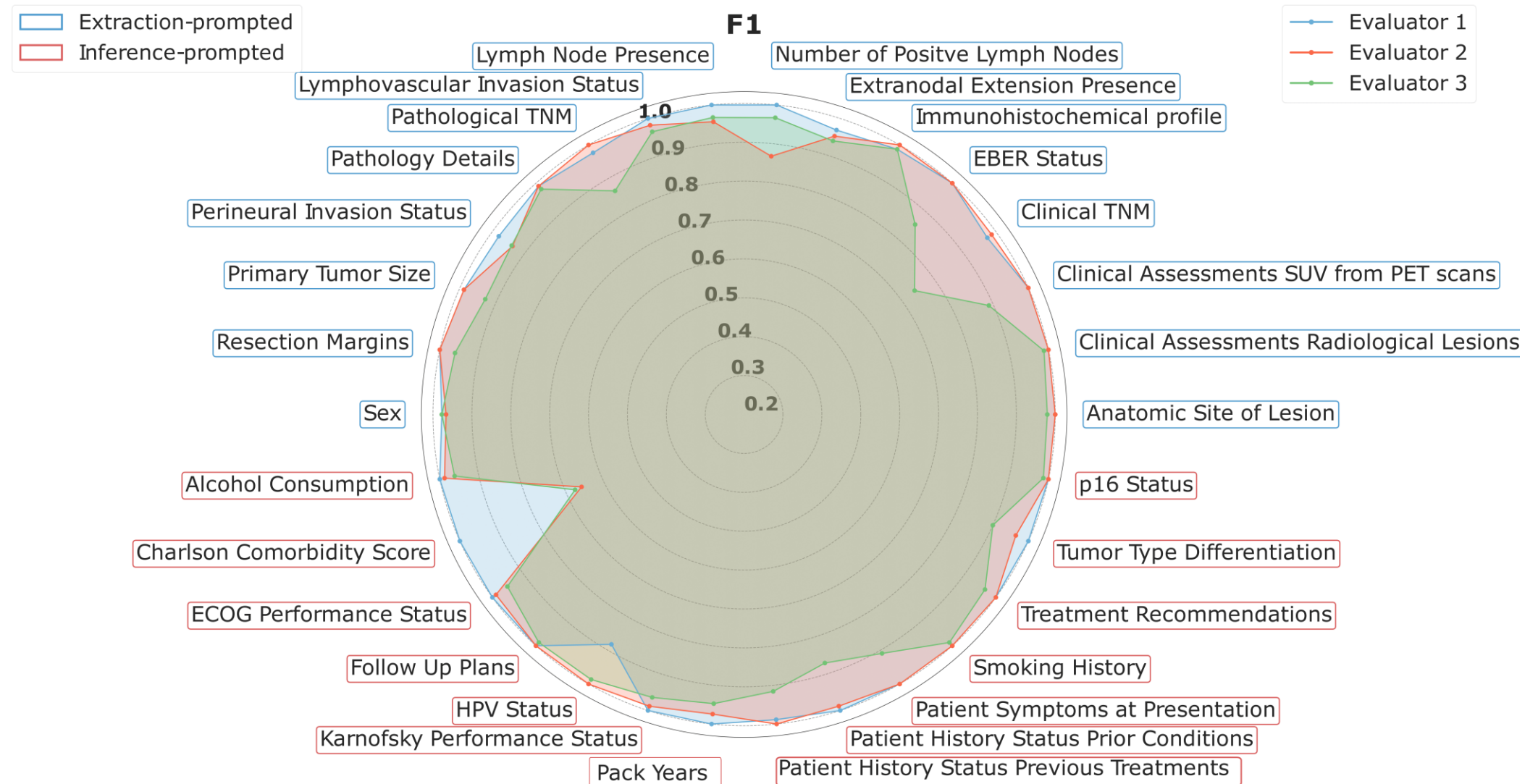
$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{4}$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{TP + FN}, \tag{5}$$

Extraction de données pronostiques à l'aide d'un modèle de langage préservant la confidentialité

Score F1 calculé par évaluateur clinique individuel, utilisé comme vérité terrain



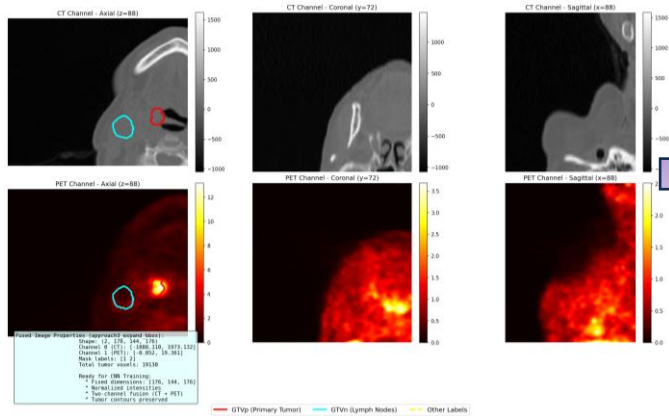
Résultats : Extraction pronostique automatisée

- 30 champs pronostiques clés : caractéristiques initiales, TNM, différenciation, statut p16/HPV, comorbidités, performance ECOG/Karnofsky
- Validé par trois radio-oncologues seniors (4500 évaluations)
- Performance du modèle :— F1 0.986 • Accord 0.922 • AC1 0.914
- 6 à 10× plus rapide que l'abstraction manuelle
- Niveau expert, déployé localement, sans cloud, réutilisable pour la fusion multimodale

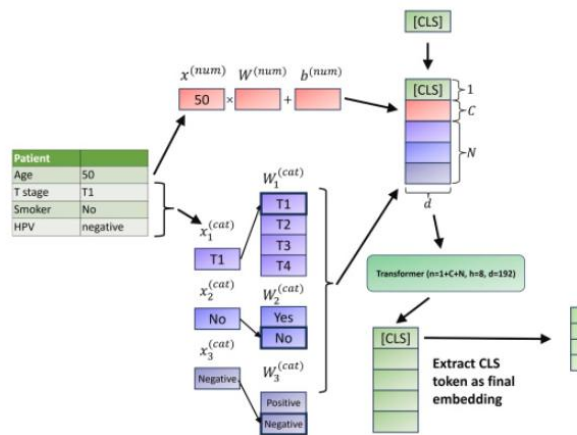
Intégration de la TEP-TDM dans un modèle multimodal de prédiction du risque (HPV+ OPSCC)

GaMMA-Surv : Fusion conjointe multimodale par auto-encodeur avec mécanisme de gating pour la TEP-TDM et les données cliniques, appliquée à la prédiction de survie dans les carcinomes épidermoïdes oropharyngés multicentriques (n=8).

TEP-TDM



les données cliniques



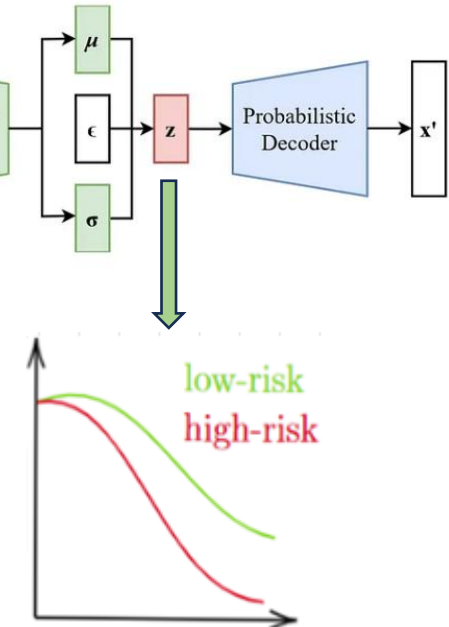
FTTransformer

Représentation d'image

Représentation conjointe (multimodale)

Représentation clinique

Fusion conjointe par VAE



<https://hecktor25.grand-challenge.org/>

Résultats: la fusion multimodale améliore nettement la prédiction du risque

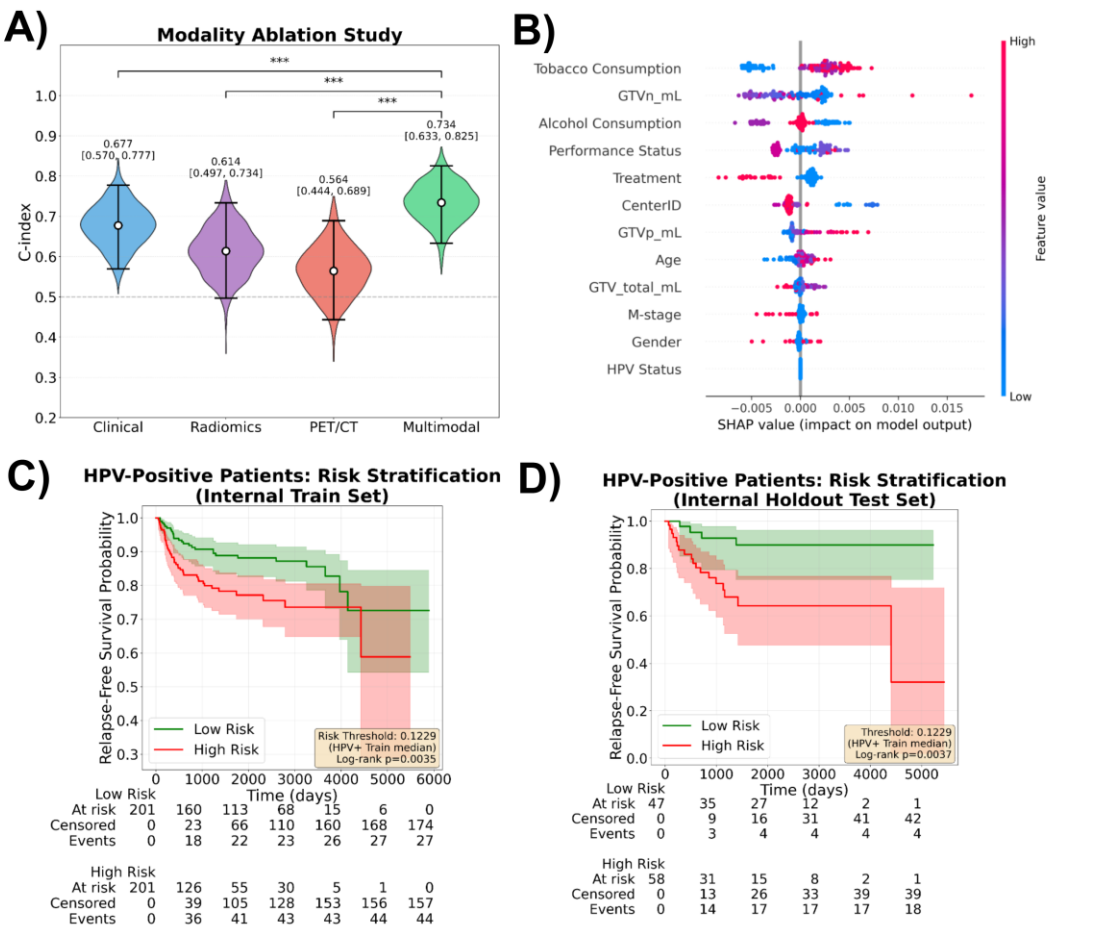


Figure 1: GaMMA-Surv model performance using Variational Autoencoder-based multimodal joint fusion of PET/CT imaging and tabular clinical variables. **A)** Modality ablation study comparing 1000-sample bootstrap resampling on the internal holdout test set concordance indices across clinical variables only, radiomics features (GTVp, GTVn, GTVtotal), PET/CT deep learning features, and multimodal fusion GaMMA-Surv. A Kruskal–Wallis H-test first determined statistical significant difference among each model, then two-sided Mann-Whitney U tests were performed between GaMMA-Surv and the unimodal models, followed by Holm–Bonferroni correction. *** indicated $p < 0.001$. **B)** SHapley Additive exPlanations (SHAP) beeswarm plot showing feature importance and directional impact on model predictions. **C-D)** Kaplan-Meier curves demonstrating risk stratification of HPV-positive patients into high- versus low-risk groups (stratified by median predicted risk from training set) on the **C)** internal training set and **D)** internal holdout test set.

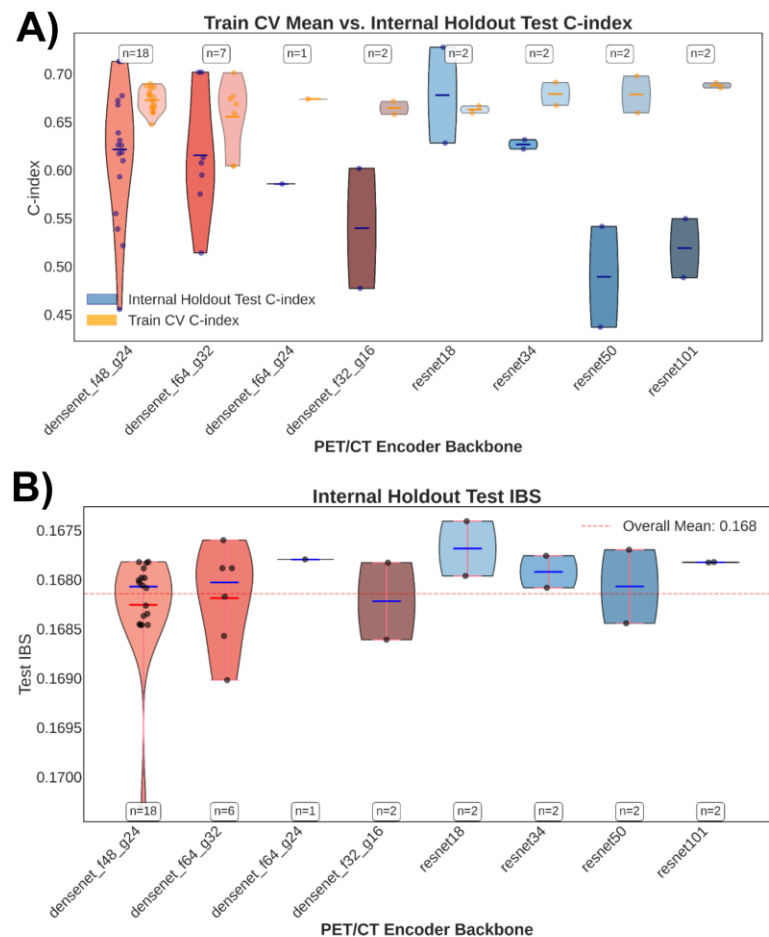
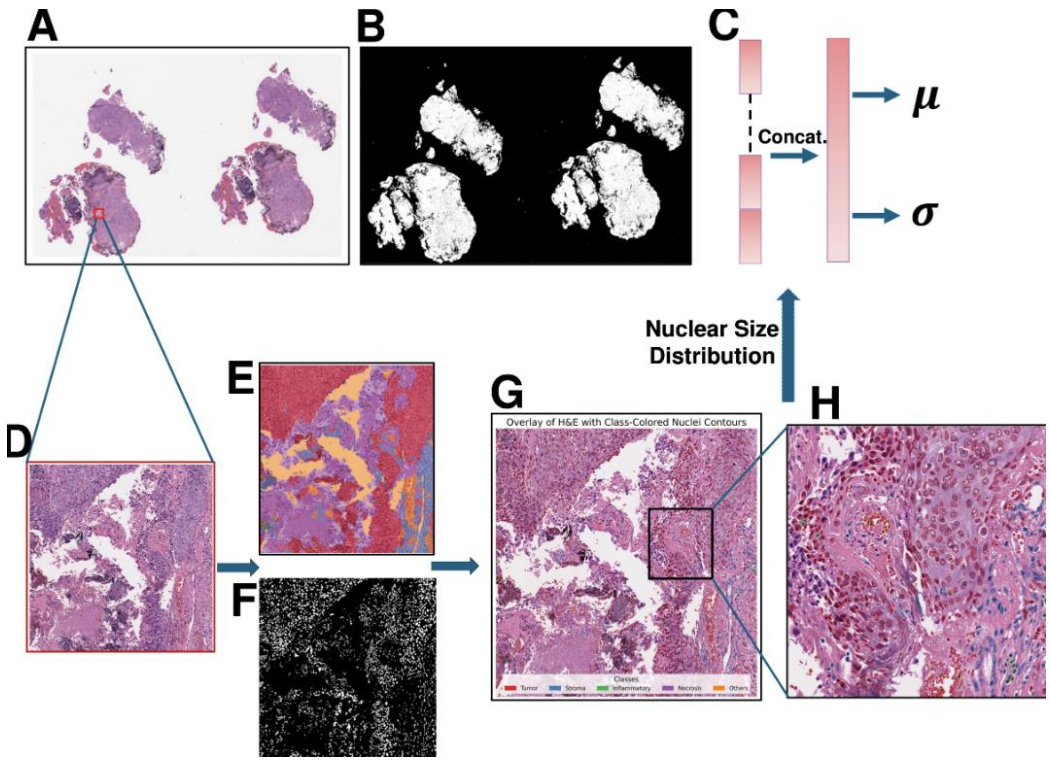
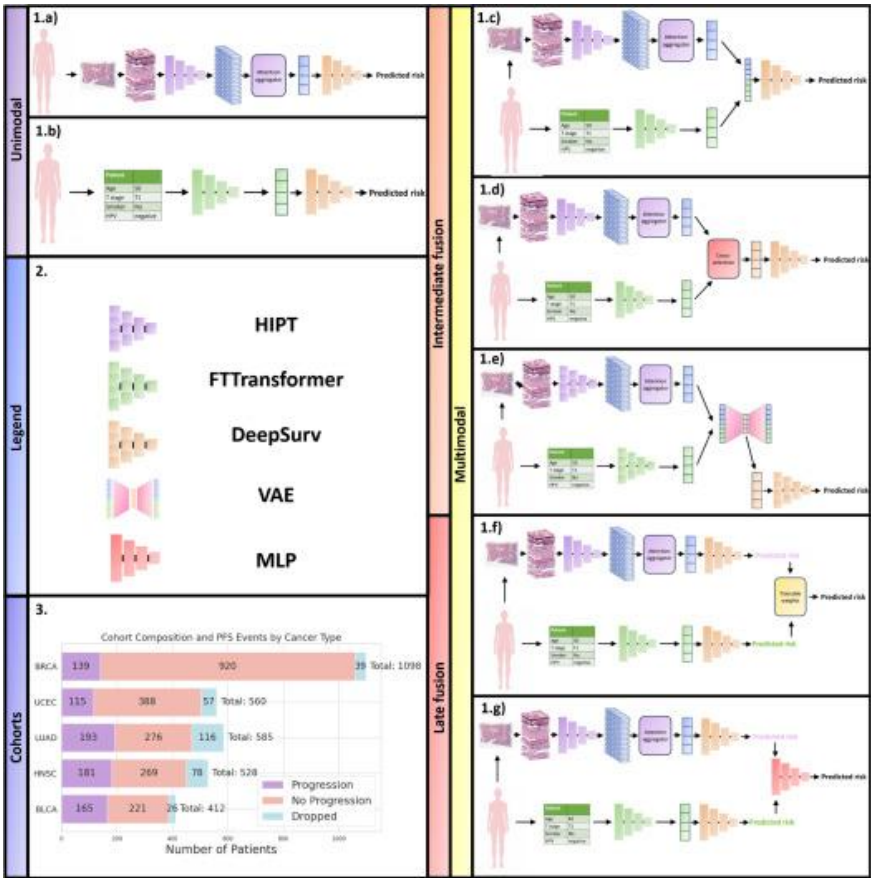


Figure 2: Comparison of PET/CT image encoder architectures using VAE multimodal joint fusion with a 192-dimensional latent space. **A)** 5-fold cross-validation mean and internal holdout test set C-index performance across different encoder architectural hyperparameters. n indicates the number of experiments for hyperparameter configuration optimization. Blue dots show the C-index of internal holdout test set (left column), and yellow dots represent the mean of 5-fold cross-validation of the train set (right column). DenseNet variants explored architectural configurations including: **(1)** baseline with 48 initial features, growth rate 24, and block configuration [8,16,32,16]; **(2)** wider shallow variant with 64 initial features, growth rate 32, and block configuration [6,12,24,16]; and **(3)** wider deep variant with 64 initial features, growth rate 32, and block configuration [8,16,32,24]. ResNet backbones of increasing depth (18, 34, 50, and 101 layers) were also evaluated. **B)** Internal holdout test set Integrated Brier Scores (IBS) for model calibration.

Pathologie numérique: un apport complémentaire pour la prédiction du risque

Fusion des lames WSI et des variables cliniques: une approche validée dans notre étude récente

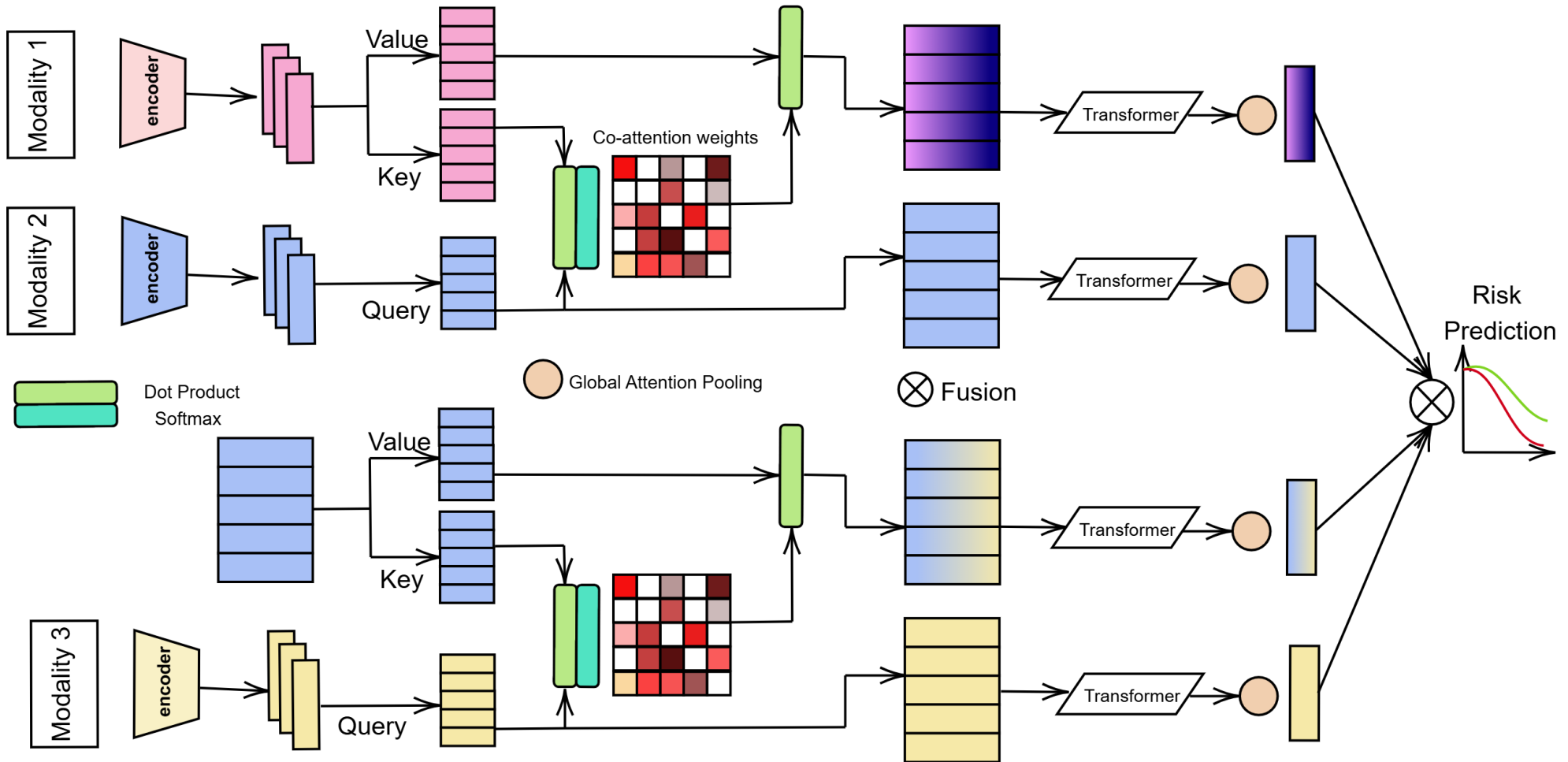


Duran J, Zou Y, Vallières M, Enger SA. Beyond single-run metrics with CP-fuse: A rigorous multi-cohort evaluation of clinico-pathological fusion for improved survival prediction in TCGA. Machine Learning with Applications. 2025 Nov 4:100789.

Zou Y, Glickman H, Pelmus M, Maleki F, Bahoric B, Lecavalier-Barsoum M, Enger SA. Tumour nuclear size heterogeneity as a biomarker for post-radiotherapy outcomes in gynecological malignancies. Physics and Imaging in Radiation Oncology. 2025 Jun 19:100793.

Aller au-delà de deux modalités : vers une fusion croisée et multi-échelle

Exploiter la richesse des données réelles : imagerie, texte, pathologie, biomarqueurs et clinique



Discussion, perspectives et collaboration à l'échelle du réseau

Vers une plateforme québécoise d'IA multimodale en oncologie

Résumé des contributions

- **OPHELIA-JGH**: Curation multimodale : clinique, textes, TEP-TDM, pathologie numérique.
- Extraction rapide de données pronostiques via un **LLM sécurisé sur site**.
- **GaMMA-Surv** : fusion PET-TDM + clinique pour la stratification du risque (OPSCC HPV+).
- **CP-Fuse** : apport complémentaire de la pathologie numérique après radiothérapie.

Perspectives

- Étendre la fusion multimodale à plus de centres et de modalités.
- Normalisation des outils de validation et de gouvernance.
- Intégration dans les flux réels des comités des tumeurs.

Collaboration

- Ouverts aux collaborations pour la **validation externe**, l'**intégration multisite** et le **déploiement clinique**.
- Contactez-nous si votre centre souhaite contribuer à une **plateforme provinciale d'IA en oncologie**.

Remerciements

Shirin Abbasinejad Enger, Ph.D.
George Shenouda, MD
Laya Rafiee, Ph.D.
Farhood Farahnak, Ph.D.
Juan Duran, M.Sc.
Harry Glickman, MD
Khalil Sultanem, MD
Marie Duclos, MD
Tomás Yokoo Teodoro, MD
Farhad Maleki, Ph.D.
Parsa Bagherzadeh, Ph.D.
Hossein Jafarzadeh, M.Sc.
Engerlab

