Intitulé de l'offre de stage

Classification de tumeurs cérébrales à partir de données frugales

Stagiaire en	Informatique (Deep Learning)
Affectation	CHU Toulouse
Durée	5 à 6 mois
Rémunération	environ 670 € net mensuel
Date de la publication	15/11/2026
Date d'embauche prévue	mars 2027 suivant disponibilités
Lieu	IUCT Oncopole, 1 avenue Irène Joliot-Curie, Toulouse

L'IUCT Oncopole

L'Oncopole regroupe sur un même site plus de 2000 professionnels engagés dans la lutte contre le cancer. Ils sont répartis entre l'Institut Universitaire du Cancer de Toulouse - Oncopole (IUCT-O)) et le Centre de Recherche en cancérologie de Toulouse (CRCT). L'IUCT-Oncopole réunit l'expertise et l'excellence de l'Institut Claudius Regaud (Centre de Lutte Con tre le Cancer) et de plusieurs services d'oncologie du CHU de Toulouse. Ses missions sont le soin, la recherche et l'enseignement.

L'Institut de Recherche en Informatique de Toulouse

L'Institut de Recherche en Informatique de Toulouse (IRIT), une des plus imposantes Unité Mixte de Recherche au niveau national, est l'un des piliers de la recherche en Occitanie avec ses 700 membres, permanents et non-permanents. De par son caractère multi-tutelle (CNRS, Universités toulousaines), son impact scientifique et ses interactions avec les autres domaines, le laboratoire constitue une des forces structurantes du paysage de l'informatique et de ses applications dans le monde du numérique, tant au niveau régional que national.

L'équipe d'accueil

L'informatique et l'intelligence artificielle sont appelés à prendre une place croissante dans le monde de la recherche médicale, et en particulier dans le monde de la recherche contre le cancer. L'équipe commune IRIT/CRCT, co-localisée sur le site de l'Oncopole de Toulouse, a pour but de faire travailler ensemble et sur le même site chercheurs en informatique, chercheurs dans le domaine du cancer et médecins. Elle peut s'appuyer sur le plateau technique du CRCT (analyses biologiques, séquençages génomique,...) et sur les moyens de calcul de la région Occitanie (CALMIP).

Le stagiaire sera accueilli au sein de l'équipe d'Emmanuelle Uro-Costes, professeur de médecine au CHU de Toulouse, chercheuse au Centre de Recherche en Cancérologie de Toulouse (CRCT/INSERM), et sera co-encadré par des chercheurs de l'IRIT dans le cadre de l'équipe commune IRIT-CRCT.

Objet du stage

Les tumeurs humaines sont très variées. Par exemple, derrière le terme grand public « cancer du sein » se cachent plus d'une vingtaine d'entités tumorales différentes, fréquentes ou rares. L'identification de chaque entité est capitale, car elle a des implications pronostiques et thérapeutiques. L'identification des types tumoraux repose sur l'analyse au microscope de la morphologie des cellules anormales, effectuée par des médecins spécialistes, les pathologistes. Ces dernières années, l'OMS a intégré des mutations de l'ADN tumoral au diagnostic microscopique, rendant la classification plus fiable et plus précise (exemple de diagnostic intégré : astrocytome IDH muté). Plus récemment encore, l'OMS a intégré une technique de classification des tumeurs cérébrales (brain classifier), développée par l'Université de Heidelberg, se basant sur le profil de méthylation de l'ADN tumoral et utilisant le Random Forest.

Le principe du méthylome est le suivant : sur les milliards de bases (A,T, C, G) de notre ADN, 28 millions de Cytosine (C), toujours à côté d'une Guanine (G) (on parle de doublets CpG) peuvent-être méthylées ou non. Cette méthylation régule la transcription des gènes. En particulier, si plusieurs CpG au niveau d'un promoteur (séquence d'ADN qui peut initier l'expression d'un gène) portent des cytosines méthylées, un blocage de la transcription du gène est généralement observé. Les cellules tumorales détournent ce système pour inactiver des anti-oncogènes et activer des oncogènes. A partir d'un set de gènes choisis pour leur implication dans le cancer, des puces d'hybridation pour étudier la méthylation des cellules tumorales ont été développées. Le pattern de méthylation de 450 000 CpG suffit amplement à identifier un type tumoral. Voici le lien vers une conférence de vulgarisation que nous avions faite sur ce sujet et qui est disponible en replay sur inscription.

Cette démarche de classification par le méthylome est en train de se développer pour d'autres tumeurs (tissus mous, leucémies, tumeurs cutanées etc...). Des fichiers informatiques (.idat) correspondant aux données brutes du pattern de méthylation sont disponibles en open source pour de nombreux cancers. À partir de ces bases de données on peut reconstruire des classifiers grâce à l'IA. Pour déterminer le profil de méthylation de l'ADN d'une tumeur, une nouvelle technologie ultra rapide, le séquençage par nanopore s'est développé ces dernières années. Cependant ce séquençage ne permet d'accéder qu'à quelques dizaines de milliers de sites CPG et non à la totalité des 450 000 sites renseignés dans la base de données. Vermeulen et collaborateurs ont eu l'idée d'éduquer des réseaux de neurones imbriqués pour prédire le diagnostic correctement avec uniquement 20 000 sites CPG (article Ultra-fast deep-learned CNS tumour classification during surgery). Ils ont pour cela

fait de multiples simulations en tirant au sort 20 000 sites CPG dans la base de données de 3000 tumeurs et leurs 450 000 sites CpG. Ce séquençage nanopore très rapide, couplé à cette IA peut donner des résultats pendant le temps opératoire permettant d'orienter le geste du neurochirurgien. Une telle méthodologie a également été utilisée pour classer les leucémies par technique nanopore (Rapid epigenomic classification of acute leukemia). La base de données publiques qui a été utilisée par Vermeulen et collaborateurs correspond à environ 3000 tumeurs et aux 81 classes de tumeurs cérébrales, qui étaient connues en 2018. Or de nombreuses nouvelles entités tumorales ont été décrites depuis et nous souhaitons éduquer une IA avec une base de données actualisée comportant 120 entités. De plus nous aimerions appliquer le même principe pour construire des classifiers adaptés à la technologie nanopore (c'est-à-dire des données tronquées par rapport à la base de données) pour d'autres tumeurs, comme les tumeurs ORL. En effet, les tumeurs ORL rares sont parfois difficiles à classer et pourraient bénéficier d'une classification par méthylome. Dans le cadre d'un projet européen (TRANSCAN SPELCASTER Pr D. Capper) puis d'un projet national PRTK fédérateur (REFCORomics Dr F-R Ferrand), nous allons réaliser 400 méthylomes correspondant à des tumeurs ORL rares. Environ 1000 cas sont déjà disponibles en open source sur des bases de données publiques. Le but sera donc :

> d'utiliser des réseaux de neurones imbriqués mais aussi de les comparer avec d'autres algorithmes concurrents
> d'optimiser le fonctionnement (i) en améliorant la base de données : équilibrer les classes, (ii) en parallélisant l'utilisation des CPU avec Spark (iii) en enrichissant la dataframe avec les anomalies chromosomiques.
> de valider les résultats en validation croisée et comparer les résultats obtenus avec la clusterisation en t-SNE, u-MAP ou ACP.

Formation

École d'ingénieur, de préférence avec spécialisation en informatique ou mathématiques (en année de césure ou stage long). Master 2 informatique ou mathématiques appliquées.

Compétences attendues

Des connaissances en science des données, apprentissage et réseaux de neurones seront appréciées.

Modalité de candidature

> CV à envoyer à: cup4@stages-medecine-numerique.fr
> Date limite de candidature : 01/01/2027
> Encadrants: Professeur Emmanuelle Uro-Coste (CHU de Toulouse/CRCT), Yvan Nicaise (CRCT) Sandrine Mouysset (IRIT), Daniel Ruiz (IRIT)

Une première sélection sera effectuée sur la base des CVs reçus. Les candidats seront immédiatement informés du résultat, et ceux qui seront retenus à l'issue de la première sélection seront invités

à un ou deux entretiens individuels en téléconférence avant sélection définitive.

L'ensemble des offres de stage est disponible sur http://www.stages-medecine-numerique.fr.