Institut de Recherche en Informatique de Toulouse

CHU TOULOUSE | CENTRE HOSPITALIER UNIVERSITAIRE

Intitulé de l'offre de stage

Modélisation multimodale non supervisée et explicable de données biologiques issues de DATAML dans le cadre de la leucémie aigüe myéloïde

Stagiaire en	Intelligence Artificielle (Deep Learning)
Affectation	IRIT/UT et CHU/CRCT
Durée	5 à 6 mois
Rémunération	environ 670 € net mensuel
Date de la publication	12/10/2026
Date d'embauche prévue	février ou mars 2027 suivant disponibilités
Lieu	Université de Toulouse, Laboratoire IRIT, 118 route de Narbonne, 31062 Toulouse cedex 9

L'Institut de Recherche en Informatique de Toulouse – IRIT

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. Ce stage concerne deux équipes de recherche : SAMOVA (spécialiste en traitement automatique de l'audio et en applications médicales) et APO (spécialiste en modélisation et optimisation et applications médicales).

Le Centre de Recherche en Cancérologie de Toulouse - CRCT

Le CRCT est une unité de recherche conjointe entre l'Inserm et l'Université Toulouse III Paul Sabatier. Au cœur de l'Oncopole de Toulouse, le CRCT, avec tous ses partenaires (institutionnels, universitaires, cliniques, industriels, caritatifs. . .), stimule l'innovation en termes de recherche et d'enseignement dans la lutte contre le cancer. Le CRCT conduit une approche intégrée entre la recherche, les soins et l'enseignement, dans une logique transversale et multidisciplinaire.

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).

Objet du stage

La leucémie aiguë myéloïde (LAM) est un type de cancer du sang caractérisé par une croissance rapide et incontrôlée de cellules myéloïdes anormales dans la moelle osseuse et le sang. Le pronostic de la LAM repose souvent sur l’analyse minutieuse des cellules de la moelle osseuse à partir de frottis et sur la biologie moléculaire. Ces dernières sont enregistrées dans un registre appelé DATAML depuis 2000. Ces données sont de natures différentes mêlant données démographiques, biologiques, cytogénétiques, moléculaires et données concernant le traitement et le pronostic.

Cette démarche répond au besoin d'analyser conjointement des données de natures différentes, jusqu'ici jamais étudiées ensemble. Face à l'absence de labels et à l'objectif potentiel de découvrir une nouvelle stratification de patients, une approche non supervisée et novatrice s'avère indispensable.

Cette approche vise à détecter des structures explicables sur chaque modalité en utilisant des outils basés graphes comme la modularité et les algorithme de détection de communauté (Leiden). Elle est divisée en deux parties :

- Apprentissage intra-modalité auto-supervisé par l’extraction de structures communautaires propres explicables à chaque modalité
- Intégration multi-vue pondérée reposant sur une combinaison linéaire pondérée par les modularités issues de chaque modalité. Cette approche permet de favoriser automatiquement la source de données la plus structurée, et ce, sans aucune intervention humaine.

Ce stage consisterait à développer et tester cette approche en utilisant les données SHAP comme opérateur de raffinement topologique. Ce stage offre une opportunité unique de contribuer au domaine de la santé en développant des outils innovants pour le pronostic précoce et précis de la leucémie aiguë myéloïde, tout en acquérant des compétences avancées en machine learning.

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 programmation (Python notamment), en traitement du signal et en IA apprentissage automatique seront appréciées.

Modalité de candidature

> CV à envoyer à : sandrine.mouysset@irit.fr et Bertoli.Sarah@iuct-oncopole.fr
> Date limite de candidature : 1/01/2027
> Encadrants : Sandrine Mouysset (IRIT/UT), Sarah Bertoli (CRCT/CHU)

Une première sélection sera effectuée sur la base des CVs et lettres de motivation 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.