



## Intitulé de l'offre de stage

Développement d'un Pipeline de Machine Learning pour la Classification de Cellules de Moelle Osseuse dans le Contexte de la Leucémie Aiguë Myéloïde

<b>Stagiaire en</b>	Informatique (Deep Learning)
<b>Affectation</b>	Équipe commune CRCT/IRIT
<b>Durée</b>	5 à 6 mois
<b>Rémunération</b>	environ 650 € net mensuel
<b>Date de la publication</b>	1/12/2023
<b>Date d'embauche prévue</b>	mars / avril 2024 suivant disponibilités
<b>Lieu</b>	CRCT, 2 Avenue Hubert Curien – 31100 Toulouse

## Le Centre de Recherche en Cancérologie de Toulouse

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'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

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. Dans ce contexte, l'automatisation de la classification des cellules de moelle osseuse par le biais de techniques de machine learning offre un potentiel significatif pour améliorer l'efficacité et la précision du pronostic.

Le stage vise à développer un pipeline de machine learning robuste pour la classification automatique des cellules de moelle osseuse à partir de lames numérisées de frottis. L'objectif principal sera d'explorer différentes approches de classification, telles que la classification supervisée, non supervisée, ainsi que l'intégration de techniques d'apprentissage actif pour optimiser la performance du modèle au fil du temps. Tâches principales :

- > Collecte et prétraitement des données : Prise en main d'un outil existant pour l'extraction d'images de cellules à partir de lames numérisées de frottis de moelle osseuse. Construire une base de données d'images, et mettre en place des techniques de prétraitement pour améliorer la qualité des images.
- > Exploration des méthodes de classification : Comparer les performances des approches de classification supervisée et non supervisée en utilisant des algorithmes de machine learning adaptés au contexte spécifique de la LAM.
- > Intégration de l'apprentissage actif : Implémenter des stratégies d'apprentissage actif pour maximiser l'efficacité du modèle en sélectionnant de manière intelligente les échantillons les plus informatifs.
- > Évaluation et optimisation : Évaluer la performance du modèle sur des ensembles de données de validation et de test, et optimiser le pipeline en conséquence.

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 et en analyse d'images médicales.

## 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

- > Connaissances approfondies en machine learning, en particulier en classification d'images.
- > Maîtrise de langages de programmation tels que Python.
- > Familiarité avec les bibliothèques de machine learning comme scikit-learn, TensorFlow ou PyTorch.
- > Aptitude à travailler avec des données d'imagerie médicale.

## Modalité de candidature

- > CV à envoyer à: [dlam@stages-medecine-numerique.fr](mailto:dlam@stages-medecine-numerique.fr)
- > Date limite de candidature : 31/01/2024
- > Encadrants: Dr Sarah Bertoli (CHU de Toulouse/IUCT-O/CRCT), David Simoncini (IRIT), Sandrine Mouysset (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>.