

Stage M2 au CEA : Nouvelles approches par réseaux de neurones convolutifs pour le traitement des données de spectrométrie de masse en métabolomique pour la santé

Contexte : La métabolomique globale par spectrométrie de masse (MS) est un outil puissant pour la découverte de biomarqueurs en santé et la médecine de précision. Toutefois, le prétraitement et l'annotation sont actuellement des étapes limitantes dans l'analyse des données [1], qui requièrent une validation manuelle chronophage.

Les approches de deep learning en intelligence artificielle ont révolutionné ces dernières années la capacité à extraire automatiquement le signal. En particulier, les réseaux de neurones convolutifs (CNN) ont montré des performances de classification très élevées notamment dans le domaine de l'imagerie [2].

Sujet : L'objectif de ce stage est de développer des approches de traitement et d'annotation des données brutes MS (considérées comme des images) grâce aux CNNs [3], pour réduire le nombre de faux positifs et négatifs, et reconnaître automatiquement les patterns de fragmentation. Les nouveaux algorithmes et outils logiciels seront appliqués à la détection et quantification de biomarqueurs dans les échantillons de cohortes de patients.

Travail demandé : La première partie du stage consistera à extraire les signaux d'intérêt dans les données brutes sous forme d'images. Dans un deuxième temps, plusieurs architectures CNNs pour l'étiquetage automatique des signaux seront comparées. L'apprentissage sera effectué grâce à une large base de données spectrales sur des échantillons humains disponible au laboratoire.

Références :

[1] Delabriere *et al.* (2021) SLAW: A scalable and self-optimizing processing workflow for untargeted LC-MS. *Analytical Chemistry*, 93, 15024–15032.

[2] LeCun *et al.* (2015) Deep learning. *Nature*, 521, 436–444.

[3] Skarysz *et al.* (2018) Convolutional neural networks for automated targeted analysis of raw gas chromatography-mass spectrometry data. *International Joint Conference on Neural Networks (IJCNN)*.

Profil : Nous recherchons un.e candidat.e avec un très bon dossier en mathématiques appliquées (traitement d'images, statistiques, informatique), et motivé.e par les applications multidisciplinaires (chimie, santé). Le master pourra se prolonger par une thèse.

Lieu du stage : Le stage s'effectuera sur le site du CEA de Saclay, dans l'équipe de science des données métabolomiques du DMTS (<https://odisce.github.io>), en partenariat avec l'équipe d'imagerie génétique de NeuroSpin (<https://brainomics.org>).

Durée du stage : 6 mois à partir de mars 2023

Date limite de candidature : 30 novembre 2022

Contact : Merci d'adresser votre candidature (CV détaillé avec les expériences de recherche, notes Master 1 et lettre de motivation) à :

Etienne Thévenot (etienne.thevenot@cea.fr)

Equipe Sciences des Données

UMR Médicaments et Technologies pour la Santé

CEA, Centre de Saclay, F-91191 Gif sur Yvette, France

Web: <https://odisce.github.io>