

Apprentissage semi-supervisé pour les séries temporelles : l'IA au service de la médecine du futur.

1 Résumé du projet de thèse

Les cinq dernières années ont été les témoins d'une percée importante de l'IA, grâce à l'introduction de l'apprentissage profond. Des applications révolutionnaires ont été trouvées dans la plupart des domaines de la science. Cependant, il existe une limitation importante qui handicape l'extension de cet outil au domaine médical. Rappelons que l'apprentissage profond repose principalement sur l'exploitation de données en grandes quantités, pour lesquelles l'on connaît les réponses (les "labels") que le réseau doit apprendre à prédire. Ces labels sont souvent difficiles à obtenir dans le domaine médical pour diverses raisons (coût prohibitif de labélisation par des experts du domaine, problème de type RGPD, historique restreint de collecte de données propres et standardisées, nouveaux types de données obtenues grâce à de nouveaux instruments de mesures...).

Le graal dans le domaine est donc de trouver de nouvelles techniques permettant d'entraîner un réseau de neurones profond à l'aide de très peu de données labélisées. L'apprentissage semi-supervisé est une technique émergente qui a permis lors des deux dernières années, sous l'impulsion des principales équipes à la pointe du domaine (facebook research, google brain), de très gros progrès dans cette direction. Ces travaux concernaient principalement des données de type images. Cette technique tire parti, grâce à des tâches dites "pretexte", de données non labélisées mais en volume important pour entraîner un réseau qui permet d'extraire des "features" ("caractéristiques" en Français) de haut niveau. Ces features peuvent être ensuite assemblées efficacement à l'aide de réseaux très simples entraînés sur des tâches plus spécifiques ("downstream tasks") à l'aide de peu de labels.

L'objectif de cette thèse est d'adapter et appliquer ces méthodes à des séries temporelles d'origine médicale. Plus précisément, nous comptons les appliquer à trois problématiques distinctes.

1. Les données d'EEG et MEG sur lesquelles travaille l'équipe Inserm U1105 du professeur F. Wallois à l'université Picardie Jules Verne (dans le cadre du CPER MOSOPS).
2. Les données de flux et de pression de sang et de liquide cérébro-spinal de l'équipe du Docteur O. Balédent au sein de l'institut Faire Face (dans le cadre du projet Interreg REVERT et du CPER PIA4a2U-santé). Il est à noter qu'Olivier Balédent effectue conjointement une demande de bourse de thèse auprès de l'UPJV dont l'un des objectifs est de labéliser leurs données, en synergie avec le présent sujet de thèse.
3. Les données de photopléthysmographie du projet de labcom Emile APE

Ce sujet de thèse est en cohérence avec la thématique et piste de spécialisation "Santé de précision et maladies civilisationnelles" de la stratégie régionale recherche innovation, l'axe stratégique (Hub 1) "Santé de précision : vers une approche personnalisée de la prévention et des traitements" affiché par l'I-SITE Université Lille Nord-Europe et s'inscrit pleinement dans le cadre du DIM "intelligence artificielle" de l'université d'Artois. En outre, ce projet offre une dimension transdisciplinaire en croisant l'intelligence artificielle et les thématiques ci-dessus. Enfin il contribue à la fois à la structuration de de l'Alliance A2U et à la structuration de la recherche régionale grâce aux collaborations entre le LML et les laboratoires INSERM U1105 et l'institut Faire Face d'Amiens.