

Sujet : Détection d'anomalies cardiaques à partir de données multi-sources

I. Contexte

Ce sujet de stage s'inscrit dans le cadre du projet CARDIA.

Les observations cliniques en cardiologie ont montré que les cardiomyopathies héréditaires sont des maladies graves, complexes et mal comprises qui affectent des millions de jeunes patients à travers le monde, notamment dans les terrains de sport, pouvant provoquer la mort subite. Plus d'un millier de mutations ont été identifiées dans de nombreux gènes, indiquant que diverses molécules et voies entraînent des cardiomyopathies hypertrophiques, dilatées, restrictives et arythmogènes. Ces mutations sont souvent spécifiques à un patient, pertinentes pour une personne, ou spécifiques à un groupe, c'est-à-dire caractéristiques de populations de patients sous-spécifiques. Cependant, l'hétérogénéité génétique et la variation allélique associées à la variation de l'expression clinique posent des défis importants aux cliniciens face à un individu. De plus, le phénotypage faible limite la puissance de l'analyse génétique.

Les données récoltées sur les patients, aussi bien cliniques que biologiques, représentent un important centre d'intérêt et de préoccupation pour les professionnels de la santé. Elles constituent en effet, un gisement considérable de signaux cardiaques (ECG) et de variantes génétiques, en guise d'un phénotypage clinique pertinent sur la maladie, obtenu grâce à un profiling clinico-biologique précis pour les professionnels de la santé. Il est donc primordial pour ces derniers de disposer d'un tableau de « bord » efficace permettant d'une part, de diagnostiquer au mieux la cardiomyopathie héréditaire et d'autre part, de servir comme outil d'aide à la décision thérapeutique. Actuellement, Le diagnostic de la cardiomyopathie chez un patient donné est simple et incomplet et sera certainement amélioré en combinant des données cliniques et de génotypage. À leur tour, les profils clinico-biologiques obtenus, établis comme des empreintes, aideront à adapter la thérapie à différents patients ou groupes de patients. Cependant, l'obtention de tels profils nécessiteront un traitement automatisé et optimal face à la quantité massive de données médicales disponibles.

II. Objectifs

Le diagnostic automatique à partir de données mixtes est une tâche immense, sachant qu'il existe actuellement environ un volume considérable d'enregistrements ECG et de signatures génétiques dans le référentiel des hôpitaux, non structurés en provenance de sources hétérogènes.

De manière synoptique, l'objectif final de notre recherche est d'automatiser la création et l'enrichissement des bases de données contenant les caractéristiques et les comportements cardiaques et héréditaires, en vue d'un diagnostic optimal des pathologies sous-jacentes.

Étant donnée la nature complexe des données récoltées sur les patients, un des besoins exprimés par les professionnels de la santé couvrent les aspects suivants, qui constituent les objectifs opérationnels et scientifiques du projet:

- Besoin en détection d'anomalies: Il s'agit de détecter des signaux ECG globalement déviants ou les séquences temporelles qui posent problème afin de mettre en avant les impulsions électriques aberrantes. Ce niveau de détection est intéressant pour écarter les données « suspectes » qui pourraient drastiquement biaiser le profiling des patients. Cette suspicion sur les données peut avoir lieu aussi bien sur la partie des signaux cardiaques que sur la partie biologique à fortes mutations génétiques;

III. Démarche scientifique

Il s'agira, dans un premier temps, de dresser un état de l'art complet sur les méthodes existantes relevant de l'apprentissage multi-vues et la détection d'anomalies qui pourraient être explorées dans le cadre de ce stage. Une étude comparative de différentes approches sera menée en application sur

les données de santé et permettra de guider la suite de la recherche vers le développement et la mise en œuvre d'une approche originale et spécifiquement adaptée à la problématique abordée.

Cette démarche devrait permettre :

- D'établir un état de l'art de premier plan sur les méthodes d'apprentissage multi-vues et de détection d'anomalies.
- Les méthodes de l'état de l'art seront implémentées et testées sur des bancs d'essais et les données issues d'un projet de recherche.
- De présenter éventuellement une approche nouvelle et adaptée à cette problématique permettant de dépasser l'état de l'art.
- D'implémenter, intégration et déploiement de la méthode.

IV. Encadrants:

- Khalid Benabdeslem (Univ. Lyon1 - LIRIS): Khalid.benabdeslem@univ-lyon1.fr
- Seif-Eddine Benkabou (Univ. Poitiers, LIAS): seif.eddine.benkabou@univ-poitiers.fr