Projet de thèse: **Automatisation de l'analyse des causes racines et de l'analyse d'impact des changements avec des applications en santé**

Context: L'analyse des causes racines (Root Cause Analysis, RCA) est une méthode de résolution de problèmes qui vise à identifier les causes les plus lointaines des défaillances ou des dysfonctionnements. Ces méthodes sont particulièrement utiles en épidémiologie et en santé, car elles permettent de mieux comprendre les mécanismes sous-jacents des anomalies observées et d'orienter efficacement les interventions. L'analyse d'impact des changements (Change Impact Analysis, CIA) est complémentaire à la RCA car elle permet d'évaluer les effets potentiels des interventions sur les causes identifiées, en prenant en compte le contexte et les délais d'action. En santé, cette complémentarité est cruciale pour éviter des interventions inefficaces ou contre-productives. Par exemple, supposons que la méthode de RCA révèle que la fièvre observée chez plusieurs patients intubés est due à une réponse inflammatoire transitoire causée par l'intubation elle-même, plutôt qu'à une infection sous-jacente. En analysant cette situation, la CIA devra nous montrer que la fièvre disparaîtra naturellement après quelques heures sans intervention. Une décision d'administrer des antibiotiques pourrait être inutile et même contre-productive, car elle risquerait de contribuer à la résistance aux antibiotiques et de provoquer des effets secondaires chez les patients. La CIA permet ici de conclure qu'aucune intervention médicale immédiate n'est requise, évitant ainsi des complications inutiles.

L'automatisation de la RCA et de la CIA à l'aide de l'intelligence artificielle (IA) présente des avantages considérables pour le domaine de la santé, en particulier en réanimation, où les décisions doivent être prises rapidement pour optimiser la prise en charge des patients. Par exemple, en unité de soins intensifs, un médecin réanimateur équipé d'un tel outil pourrait accélérer considérablement l'identification des causes sous-jacentes d'une détérioration soudaine de l'état d'un patient. Si un patient présente une désaturation en oxygène, une hypotension sévère ou une acidose métabolique, l'IA causale pourrait aider à déterminer rapidement si ces anomalies sont dues à une infection sous-jacente, une défaillance multi-organique, ou un effet secondaire d'un traitement en cours. Il est essentiel de souligner que ces problématiques sont fondamentalement causales. Ainsi, les méthodes purement basées sur l'apprentissage automatique (e.g., deep learning) ne sont pas les plus adaptées. Ces modèles, bien qu'efficaces pour prédire l'évolution d'un patient, ne permettent pas de distinguer corrélation et causalité [Pearl, 2009, Bareinboim et al., 2022].

Pour relever ces défis, ce projet de thèse se concentre sur une approche d'IA causale [Spirtes et al., 2000, Pearl, 2009], qui peut être considérée comme une IA hybride, combinant des éléments de l'IA symbolique (raisonnement explicite, règles logiques) et des éléments de l'IA basée sur l'apprentissage automatique (extraction de modèles à partir des données). Nous nous appuierons notamment sur une nouvelle méthode d'IA causale conçue spécifiquement pour la RCA [Assaad et al., 2023], nommée SGRCA. Cette méthode repose sur des données observationnelles temporelles et un summary graph, qui représente de manière abstraite les connaissances des experts du domaine. En résumé, SGRCA consiste à identifier les effets directs à partir du summary graph en utilisant un raisonnement causal et des règles graphiques [Ferreira and Assaad, 2024b], puis à estimer ces effets directs à l'aide de méthodes d'apprentissage automatique, et comparer ces effets directs en utilisant des données normales et anormales afin de detecter les causes racines.

Un summary graph est une abstraction du graphe causal réel qui est un graphe dirigé acyclique (DAG). L'intérêt de travailler avec un summary graph, plutôt qu'avec un DAG, réside dans sa moindre spécificité, ce qui le rend plus facile à construire pour les experts du domaine. Cependant, en tant qu'abstraction, le summary graph peut contenir des cycles, ce qui rend parfois les effets directs non identifiables [Ferreira and Assaad, 2024b]. En conséquence, les causes racines ne sont pas toujours identifiables. Afin que SGRCA soit appliquée efficacement, il est essentiel de développer des solutions adaptées aux situations où les causes racines ne peuvent pas être identifiées. Une approche intéressante consisterait à exploiter les méthodes de découverte causale [Spirtes et al., 2000] pour reconstruire le véritable DAG, puisque, une fois connu, l'effet direct devient facilement identifiable. Cependant, ces méthodes reposent sur des hypothèses fortes et non testables, limitant leur applicabilité. Une alternative plus pertinente serait de développer des approches ciblées, exploitant les informations déjà présentes dans le summary graph et se concentrant uniquement sur les sous-graphes du DAG nécessaires à l'identification de l'effet direct, notamment ceux caractérisant les cycles dans le summary graph, tout en réduisant les hypothèses requises par la découverte causale pour garantir une meilleure robustesse.

En outre, SGRCA repose sur plusieurs hypothèses qui ne sont pas toujours vérifiées en épidémiologie et en santé. Elle suppose notamment que les relations causales entre les variables sont linéaires et que tous les facteurs de confusion sont mesurés. Ces hypothèses limitent considérablement le potentiel d'application de SGRCA. Afin d'élargir son champ d'application, il est nécessaire de simplifier ces hypothèses et de rendre la méthode plus flexible pour mieux répondre aux spécificités des données de santé. Pour cela, dans un cadre non linéaire, il sera essentiel d'utiliser un raisonnement contrefactuel pour identifier les effets directs [Ferreira and

Assaad, 2024a]. De plus, l'estimation de ces effets nécessitera l'emploi de méthodes avancées, telles que le deep learning [Xu et al., 2022], afin de mieux capturer la complexité des relation.

Enfin, il n'existe actuellement aucune méthode de CIA venant compléter la méthode de RCA proposée, rendant ce travail particulièrement essentiel. Pour y parvenir, il sera crucial de générer des scénarios factuels (où aucun changement n'est introduit) et des scénarios contrefactuels (où une intervention est réalisée sur les causes racines, par exemple via un ou plusieurs traitements). Les différents scenarios seront ensuite comparés afin de formuler une recommandation optimale. Avant d'utiliser des techniques d'IA générative pour générer ces scénarios, un raisonnement contrefactuel basé sur le graphe disponible sera nécessaire.

Les méthodes développées pendant cette thèse seront notamment appliquées à des données de réanimation. Pour mener à bien cette application, nous collaborerons avec le service de réanimation de l'hôpital Saint-Antoine, notamment avec Pr MAURY et Pr AIT-OUFELLA avec lesquels l'implantation hospitalière de l'IPLESP a permit de mettre en place une longue collaboration. Les données issues de l'entrepôt des données de santé de l'AP-HP ainsi que des données issues de MIMIC3 [Johnson et al., 2016] seront utilisées pour les applications.

Objectifs: L'objectif principal de cette thèse est de contribuer à la fois à la RCA et à la CIA, avec une application spécifique en santé, notamment en réanimation. Quatre objectifs principaux ont été définis : 1. Développement d'un algorithme de découverte causale locale, pour enrichir SGRCA, qui permet de détecter des relations causales suffisante pour identifier l'effet directe en question et visera à réduire les hypothèses fortes requises par les méthodes de découverte causale générale [Spirtes et al., 2000]; 2. Étendre SGRCA à des cas plus complexes, en particulier ceux impliquant des relations non linéaires et des facteurs de confusion non observés; 3. Développement d'une méthode de CIA complémentaire à SGRCA; 4. Application et évaluation des méthodes développées sur des données de réanimation ainsi que sur des données de néphrologie.

Alignment with PostGenAI@Paris project's scientific missions: Le projet est pleinement alignée avec les missions scientifiques du projet PostGenAI@Paris, car il s'inscrit directement dans le développement de technologies de rupture appliquées à la santé du futur. En particulier, le projet repose sur des avancées en IA causale, un domaine émergent qui dépasse les approches traditionnelles de l'apprentissage automatique en offrant des modèles explicables et robustes pour la prise de décision en milieu médical. L'automatisation de la RCA et la CIA en épidémiologie et en soins intensifs s'inscrit dans une vision de médecine augmentée par l'IA.

Justification of the scientific approach: L'approche scientifique adoptée dans ce projet repose sur les avancées récentes en IA causale, qui permettent d'aller au-delà des modèles prédictifs traditionnels en fournissant une compréhension des relations de cause à effet. Contrairement aux méthodes classiques d'apprentissage automatique, qui se basent principalement sur des corrélations, l'IA causale vise à identifier les mécanismes sous-jacents responsables des observations [Bareinboim et al., 2022], ce qui est fondamental pour la RCA et la CIA.

Profil du candidat: Candidat(e) hautement motivé(e), titulaire d'un M2, avec une solide formation en probabilités, apprentissage automatique et inférence causale, ainsi qu'un fort intérêt pour l'épidémiologie et la santé. Une bonne maîtrise de la programmation est également requise.

Encadrement: Cette thèse sera encadrée par Pierre-Yves Boëlle (HDR) et Charles Assaad.

References

C. K. Assaad, I. Ez-Zejjari, and L. Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8395–8404. PMLR, 25–27 Apr 2023.

E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022.

S. Ferreira and C. K. Assaad. Average controlled and average natural micro direct effects in summary causal graphs, 2024a.

S. Ferreira and C. K. Assaad. Identifiability of direct effects from summary causal graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18):20387–20394, Mar. 2024b. doi: 10.1609/aaai.v38i18.30021.

A. Johnson, T. Pollard, L. Shen, L.-w. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Celi, and R. Mark. Mimic-iii, a freely accessible critical care database. *Scientific Data*, 3:160035, 05 2016. doi: 10.1038/sdata.2016.35.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2009. ISBN 0-521-77362-8.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT press, 2nd edition, 2000.

S. Xu, L. Liu, and Z. Liu. Deepmed: Semiparametric causal mediation analysis with debiased deep learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 28238–28251. Curran Associates, Inc., 2022.