



École d'été interdisciplinaire en numérique de la santé du 10 au 14 juillet 2023

Avalanche des données, diversité des recherches

Plan d'activité pédagogique

Type de l'activité : ☐ Atelier (2 h 20 min) ☒ Présentation (45 min + 5 min de questions)

Objectif : Le présent document décrit le contenu scientifique et l'organisation de l'activité pédagogique « Avalanche des données, diversité des recherches » lors de l'école d'été.

1 Introduction

L'activité pédagogique « Avalanche des données, diversité des recherches » se déroulera dans le cadre de l'école d'été interdisciplinaire numérique de la santé (EINS) et s'inscrit dans le thème « Modélisation de données et des connaissances en santé » :

Cet évènement scientifique s'adresse aux personnes effectuant des études universitaires, aux personnes professionnelles ainsi qu'aux patientes partenaires et patients partenaires qui désirent s'initier aux défis de mise en place de projets en numérique de la santé.

1.1 Renseignements sur le présentateur

Prénom, nom : Anita Burgun

Affiliation principale :

Professeure en informatique biomédicale, Université Paris Cité

Professeure associée, Université de Sherbrooke

Courriel : Anita.Burgun@aphp.fr

Biographie

La professeure Anita Burgun est cheffe du service de santé numérique dans 2 hôpitaux (Necker et HEGP) de l'AP-HP. Titulaire d'une chaire sur le numérique en santé à l'Université de Sherbrooke et d'une chaire en IA pour la santé de l'Institut PR[AI]RIE (PaRis AI Research InstitutE) en France, sa thématique de recherche porte sur l'intégration des données de santé et les systèmes d'aide à la décision dans le cadre de la médecine de précision. Elle est impliquée dans plusieurs projets européens et réseaux internationaux sur ces thématiques, et s'intéresse plus particulièrement à la conception de méthodes d'IA hybrides pour l'aide au diagnostic des maladies rares.

2 Description

Cette section présente le contenu de l'activité et les principaux objectifs.

2.1 Contenu

Les différents types de données utiles dans le secteur de la santé sont multiples et ne se limitent ni aux essais cliniques ni aux bases médicoadministratives. La crise provoquée par la COVID-19 a encore accéléré la réutilisation des données de vie réelle, qui intègrent de plus en plus largement des textes, les géodata, etc. De nouveaux modèles intégratifs, les *Master Observational Trials*, sont aussi de plus en plus utilisés en médecine pour casser les silos traditionnels. Quelques problèmes inhérents à la qualité des données seront évoqués, comme la temporalité et les biais. Au-delà d'un objectif commun de *fairification* des données en recherche, la réutilisation de données de vie réelle dans un cadre coopératif repose sur une approche fédérée dans laquelle les bases de données sont distribuées et où seuls les modèles sont partagés. La recherche sur données a montré quelques découvertes médicales importantes et on l'illustrera par des exemples concernant l'augmentation de risques de cancer liée à des facteurs génétiques, ou le risque accru de cancer lié à la prise de certains médicaments. La recherche technologique est également très active et les modèles hybrides qui nous semblent à ce jour très prometteurs seront présentés.

2.2 Objectifs de formation

Cette activité permettra à une personne participante de :

- O1. Connaître les différents types de données utiles dans le domaine de la santé.
- O2. Appréhender les problèmes inhérents aux données de santé.
- O3. Appréhender les types de recherche sur données.
- O4. Découvrir quelques modèles de base pour la réutilisation des données et la prise de décision médicale.

3 Références

Cette section présente les principales références documentaires utilisées pour construire l'activité et les références pour approfondir des concepts présentés.

3.1 Références essentielles

M. J. Sheller et al., Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data, *Sci. Rep.*, vol. 10, no 1, p. 12598, 2020.

Masison J et al.. A modular computational framework for medical digital twins. *Proc Natl Acad Sci U S A*. 2021.

Ofstad E. et al. What is a medical decision? A taxonomy based on physician statements in hospital ...*BMJ Open* 2016C.

Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. *NPJ Digit Med*. 2020 Mar 23;3:41. doi: 10.1038/s41746-020-0253-3. PMID: 32219182; PMCID: PMC7090057.

3.2 Références complémentaires

Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018 Apr 30;361:k1479.