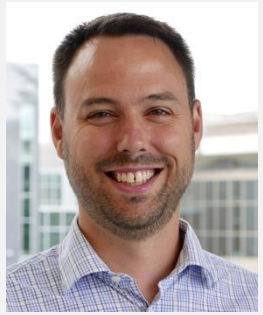
Numérique en santé:
Quelles sont les conditions gagantes ?

2025-05-26 : École d'été interdisciplinaire du numérique en santé (EINS)

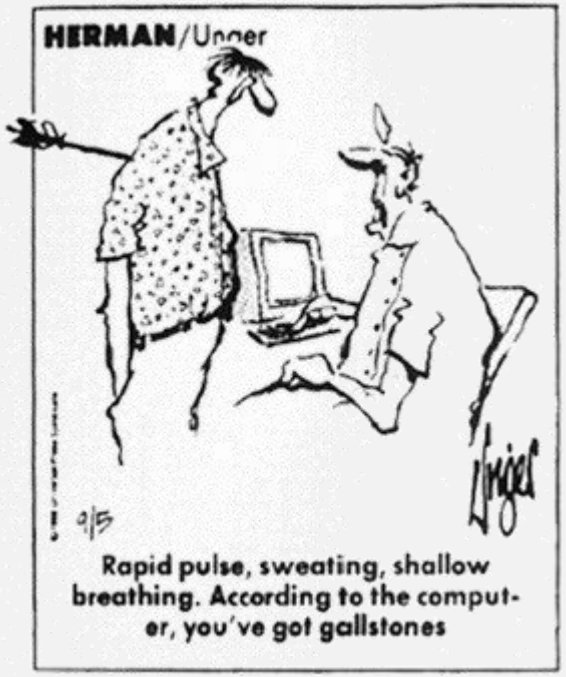
Jean-François Ethier



Professeur titulaire
Département de médecine
FMSS – Université de Sherbrooke
Chaire en informatique de la santé UdeS

Codirecteur scientifique
Groupe de recherche interdisciplinaire en informatique de la santé (GRIIS.ca)

Comité exécutif
Réseau de recherche sur les données de santé du Canada



Après la séance, vous …

- Saurez poser les questions importantes pour évaluer si un outil numérique pour la santé est :
  - pertinent;
  - sécuritaire;
  - offre une valeur ajoutée.

Aucun conflit d'intérêt

L’IA : pourquoi et comment ?

1. Avant l’IA : approches statistiques
    - Avantages
    - Limites
2. Les IA
    - Tâches
    - Apprentissage machine
    - Apprentissage profond
    - Réseau de neurones
    - Défis intrinsèques
3. IA en cliniques : défis
    - Avantages
    - Limites
4. IA en clinique : *desiderata*
    - Utilité
    - Confiance et évaluation
    - Ressources
5. IA en enseignement
    - Apprentissage vs performance

OK... Défis pour l'IA en clinique

Question

- Est-ce que vous croyez que les outils numériques seront plus utiles comme …

  a) Outils prédictifs (p. ex. : prédire le risque de faire un AVC sur 5 ans)

  b) Outils de recommandation (p. ex. : anticoaguler ou pas)

Question

- Est-ce que vous pensez qu'un outil numérique peut changer une fois qu'il est déployé dans votre hôpital ?

a) Oui

b) Non

Question

- Quel est le groupe contrôle le plus approprié pour que vous puissiez juger la performance d'un outil pour la pratique clinique ?

a) Un expert du domaine avec les données de son hôpital.

b) Un expert du domaine avec les données d'entrainement de l'outil.

c) Un expert du domaine avec les données de validation de l'outil.

d) Un expert du domaine avec des données d'un autre hôpital.

e) La réponse de ChatGPT.

En théorie

- Les **humains** ont une **variabilité** dans les performances
  - **fatigues** physiques, perte de **focus**, difficulté avec les tâches **répétitives**
- “capable of handling complex interactions in large datasets to predict outcome with greater accuracy, but the models need a greater number of input–output pairs to learn from”

Ley C *et al*. Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc. 2022 Mar;30(3):753-757.

Situation présente

- Plus de **500 outils basés** sur l'IA autorisés par la FDA

- Entre 10 % et 30 % des médecins ont utilisé un outil basé sur l'IA aux États-Unis
  - réaction allant de l'optimisme prudent au manque de confiance total

- Santé Canada
  - modèles d'IA qui pourront se modifier pendant leur utilisation en clinique

Chen, M. *et al*. Front. Med. 9, 990604 (2022).

Beaucoup d’espoirs…

- It was one of those amazing “we’re living in the future” moments. In October 2013 […] “MD Anderson is using the IBM Watson cognitive computing system for its mission to eradicate cancer.”

- Well, now that future is past. The partnership between IBM and one of the world’s top cancer research institutions is falling apart. (2017)

EDITORS' PICK

MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine

Matthew Herper Former Staff

I cover science and medicine, and believe this is biology's century.

Feb 19, 2017, 03:48pm EST

This article is more than 6 years old.

Virginia "Ginni" Rometty, chief executive officer of International Business Machines Corp. (IBM)... [+]

Défis liés à la validation et à l'implémentation

ARTICLES | VOLUME 1, ISSUE 6, E271-E297, OCTOBER 01, 2019

A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis

Xiaoxuan Liu, MBChB † · Livia Faes, MD † · Aditya U Kale, MBChB · Siegfried K Wagner, BMBCh · Dun Jack Fu, PhD · Alice Bruynseels, MBChB · et al. Show all authors · Show footnotes

Open Access · Published: September 25, 2019 · DOI: https://doi.org/10.1016/S2589-7500(19)30123-2 ·

Our review found the diagnostic **performance** of deep learning models to be **equivalent** to that of **health-care professionals**. However, a **major finding** of the review is that **few** studies presented **externally validated results** or compared the performance of deep learning models and health-care professionals using the **same sample**. Additionally, **poor reporting** is prevalent in deep learning studies, which **limits** reliable **interpretation** of the reported **diagnostic accuracy**.

- Performance souvent équivalente aux professionnels de santé
- Peu souvent validé à l'externe
- Lorsque validé, échantillon souvent différent
- Publication des résultats sous-optimale

Desiderata pour l'utilisation de l'IA en clinique

Questions que vous pourriez vouloir poser avant de choisir un outil

GRIIS
UdS Université de Sherbrooke
Réseau de recherche sur les données de santé du Canada
Health Data Research Network Canada

Question

- Étant donné la grande quantité de données utilisées pour entrainer les outils d'IA, quel avantage peuvent-ils procurer par rapport aux outils prédictifs usuels (p. ex. : le score de CHADS2) ?

a) Meilleure précision des prédictions

b) Moins de biais

c) Meilleure compréhension des mécanismes sous-jacents

d) A, B, C

e) Aucune de ces réponses

Question

- Comment identifier les outils qui seront utiles pour la pratique clinique ?

a) Bas score de Brier pour le modèle

b) Échantillon d'entrainement du modèle semblable à mes patients

c) Score d'aire sous la courbe élevé pour les patients difficiles à diagnostiquer

d) Échantillon de validation comportant au moins le double de la population couverte par mon hôpital

Utile

Dépasser l'aire sous la courbe

- Est-ce que de prédire correctement qu'un patient en soins palliatifs va mourir devrait faire partie de l'évaluation ?

Admitting service†	Points
Medicine	
General medicine	10
Cardiology	8
Gastroenterology/ nephrology/neurology	9
Palliative care	28
Hematology/oncology	14
Ante/intra/postpartum	0
Gynecology	7

Research

External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission

Carl van Walraven, Finlay A. McAlister, Jeffrey A. Bakal, Steven Hawken and Jacques Donzé

CMAJ July 14, 2015 187 (10) 725-733; DOI: https://doi.org/10.1503/cmaj.150209

Research | Open Access | Published: 02 December 2017

The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models

Melissa Assel, Daniel D. Sjoberg & Andrew J. Vickers

Diagnostic and Prognostic Research 1, Article number: 19 (2017) | Cite this article

	RF-AdminDemoDx	RF-AdminDemo	RF-Minimal	mHOMR
Internal validation[a]				
C-statistic (range)	0.90 (0.90-0.91)	0.86 (0.85-0.87)	0.85 (0.84-0.86)	0.86 (0.85-0.86)
Brier score (range)	0.068 (0.065-0.073)	0.079 (0.077-0.083)	0.082 (0.078-0.084)	0.081 (0.078-0.085)
External validation[b,c]				
C-statistic (95% CI)	0.89 (0.88-0.89)	0.85 (0.84-0.86)	0.84 (0.83-0.84)	0.84 (0.83-0.85)
Brier score (95% CI)	0.074 (0.072-0.076)	0.084 (0.081-0.086)	0.086 (0.084-0.089)	0.086 (0.083-0.088)
CDSS-eligible validation[b,d]				
C-statistic (95% CI)	0.86 (0.85-0.87)	0.81 (0.80-0.82)	0.79 (0.78-0.80)	0.80 (0.79-0.81)
Brier score (95% CI)	0.088 (0.085-0.091)	0.10 (0.097-0.10)	0.10 (0.10-0.11)	0.10 (0.099-0.11)

Expected clinical utility of automatable prediction models for improving palliative and end-of-life care outcomes: Toward routine decision analysis before implementation

Ryeyan Taseen, Jean-François Ethier

Journal of the American Medical Informatics Association, Volume 28, Issue 11, November 2021, Pages 2366–2378, https://doi.org/10.1093/jamia/ocab140

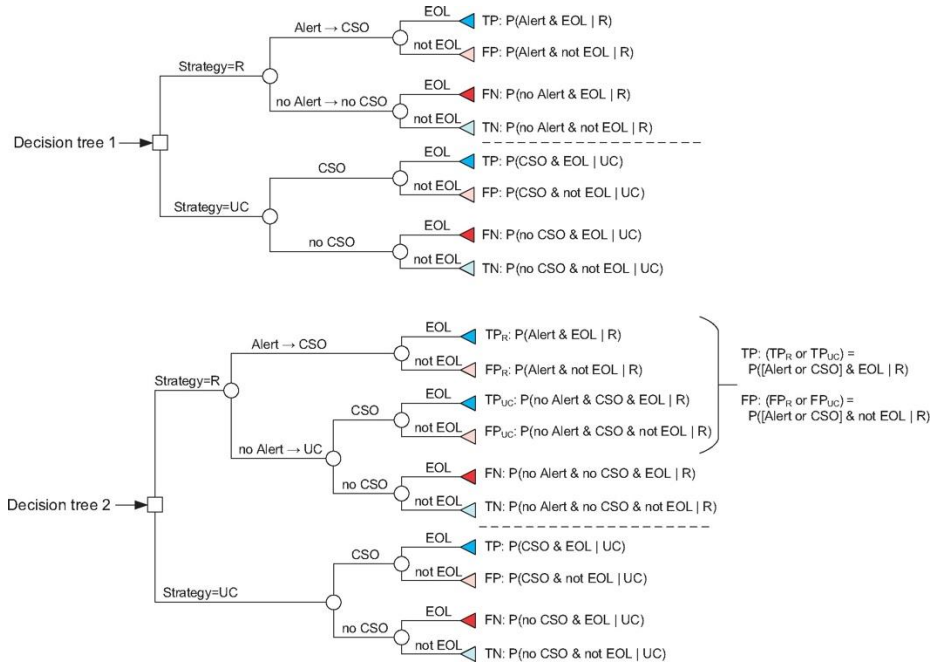
Utilité

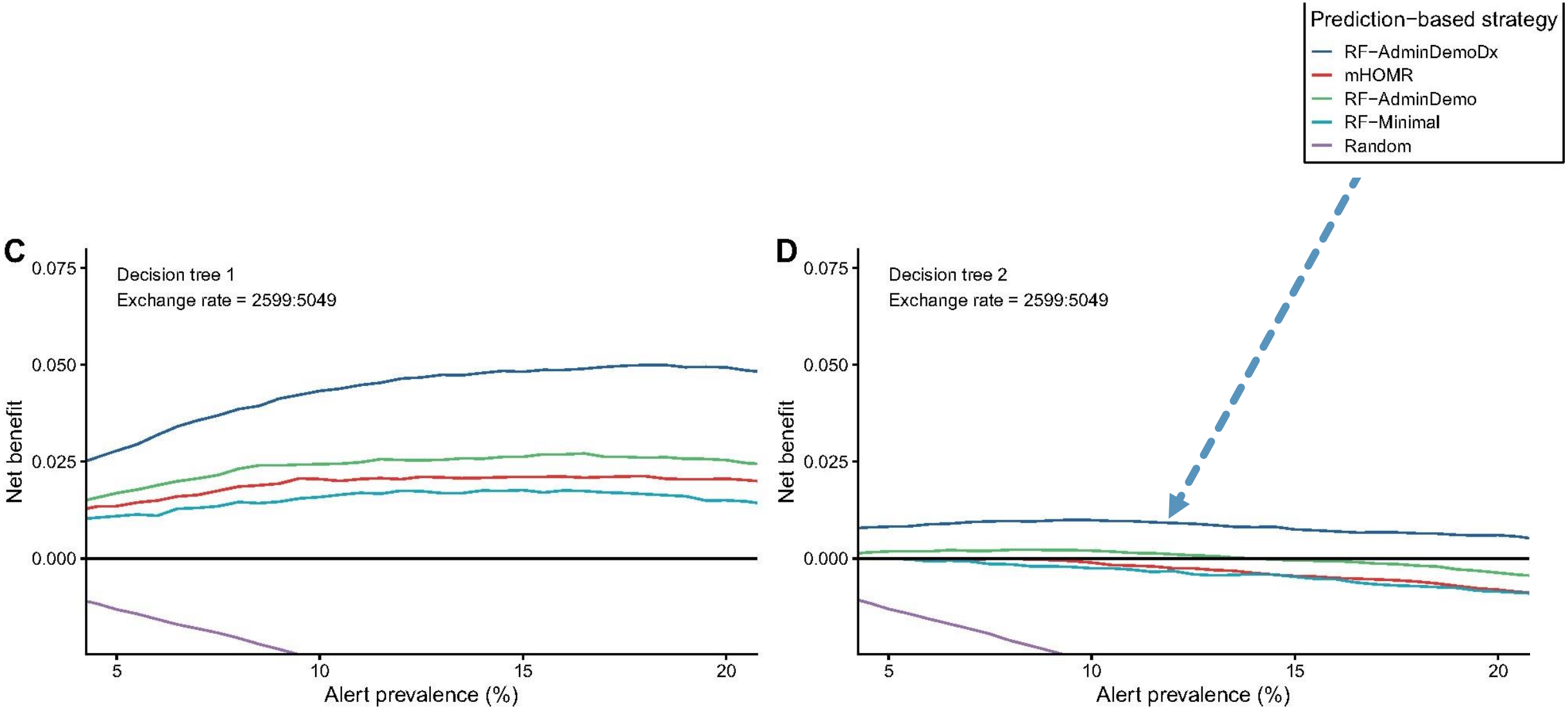
- **Qui** étaient les patients **inclus** ?
- Le modèle **fonctionnait** pour **quels patients** ?
  - Pas juste une métrique pour le modèle global
- Est-ce que les **patients** chez qui le modèle performait **moins bien** sont **exclus** de l'analyse ?
  - Intention to treat
- Est-ce que, pour les patients chez qui ça fonctionne, le modèle m'aurait aidé ?
- Est-ce que le modèle **identifie** des **patients que j'aurais manqués**
  - Exemple : des patients qui n'ont pas eu de discussions de soins de vie

Wegier P *et al.* mHOMR: a feasibility study of an automated system for identifying inpatients having an elevated risk of 1-year mortality. BMJ Qual Saf. 2019 Dec;28(12):971-979.

Anticiper l'utilité clinique

- Le **comparateur** devrait être les **soins usuels**.
  - Arbre 2 : même en cas de non-alerte, en soins usuels, des actions peuvent être posées par les cliniciens.





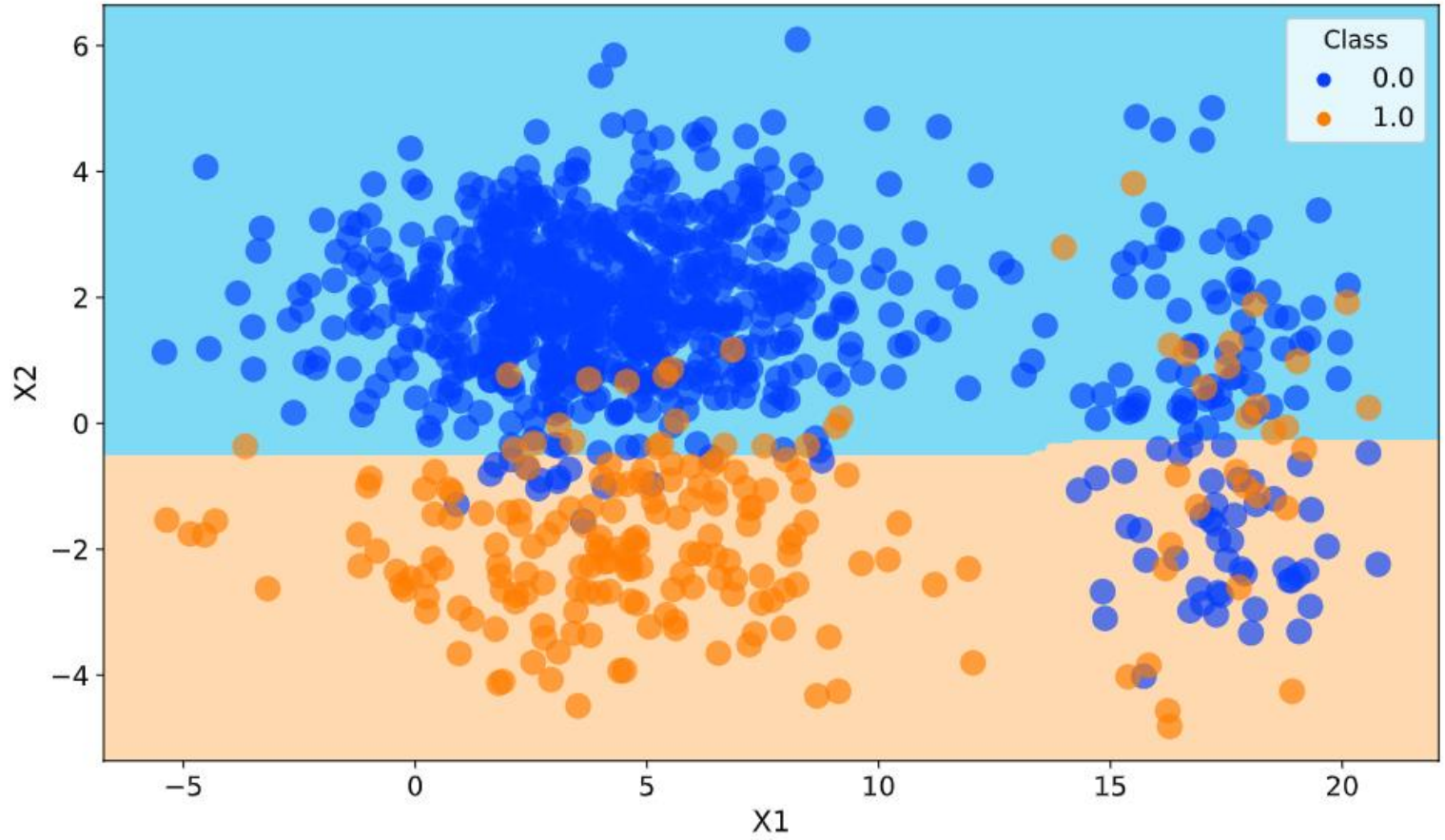
Personnalisation pour chaque patient

Biais et implications éthiques

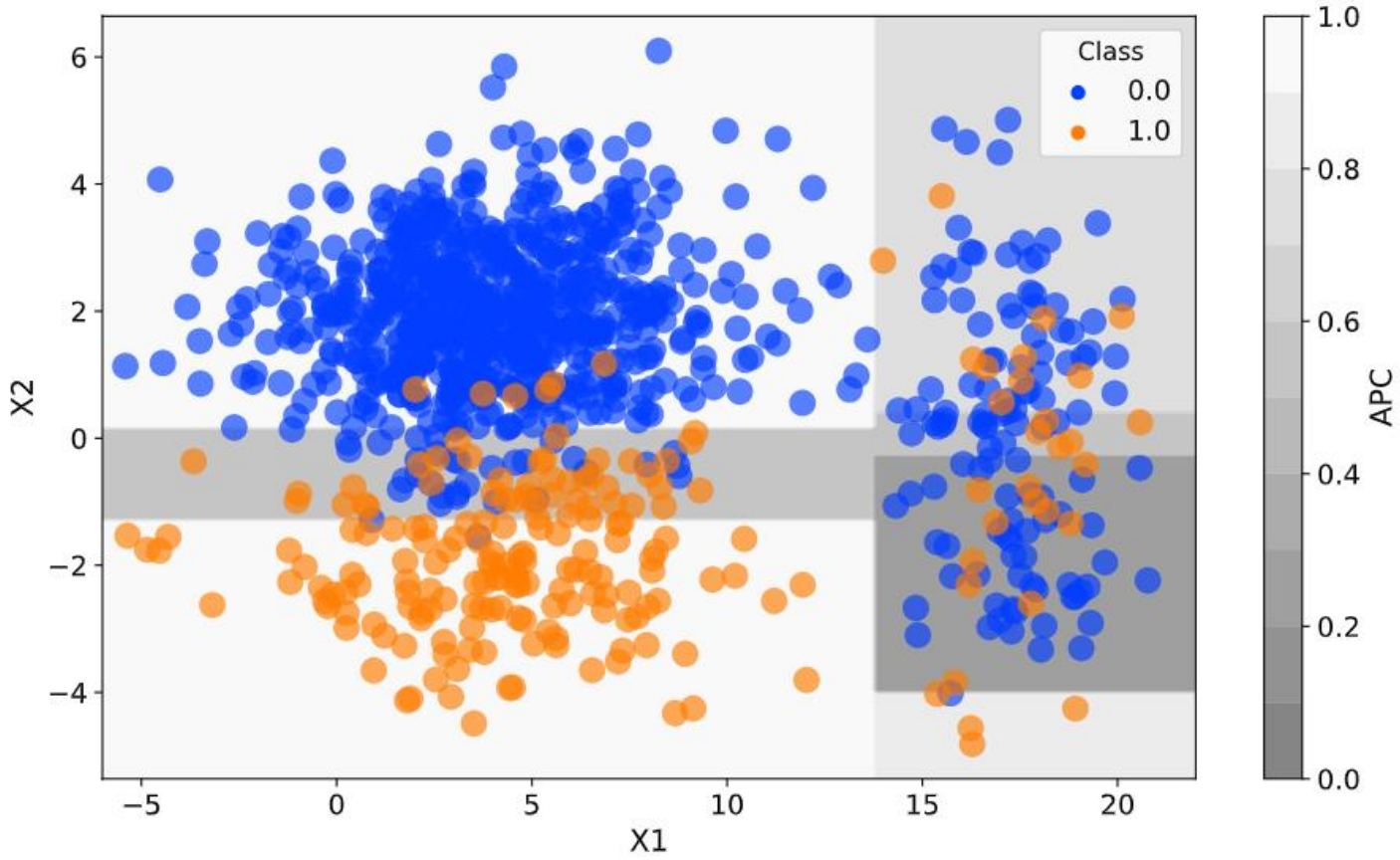
- Disons que les cliniciens **experts** ont une sensibilité et une spécificité de 88 % pour une question clinique donnée…
  - Et que, lorsqu'ils **se trompent**, c'est **au hasard**, ça peut tomber sur n'importe qui.

- Est-ce qu'un modèle avec une spécificité de 93 % et une sensibilité de 94 % pour la même question, mais qui se **trompe** presque **toujours** pour un **même groupe** (p. ex. : jeunes, hommes, roux), serait:
  - Mieux ? Pire ? Acceptable ? Désirable ?

Très bon modèle global... utile pour tous ?



Confiance différentielle



Personnalisation

- Pour un **patient donné**, sommes-nous **confiant**s du résultat **?**
- Pouvons-nous avoir une idée des **intrants** ayant eu un impact **?**
- Est-ce que **toutes les variables sont nécessaires ?**
- Comment réagit le système en cas de données manquantes ?
  - Association entre données manquantes et groupes de patients

Adaptabilité à la situation clinique

Données manquantes

- Est-ce que le système donne **toujours une réponse ?**
- Est-ce qu'une **valeur** est ajoutée **artificiellement** ?
  - Valeur semblable aux patients semblables qui en avaient ?
  - Valeur normale ?
    - Exemple : albumine et bilirubine aux soins intensifs

Quel est le résultat ?

- Est-ce seulement oui ou non ?
- Possible d'avoir une probabilité ?
- Possible d'ajuster le modèle selon la question clinique ?
  - Plus sensible ? Plus spécifique ? Meilleure VPP ?
- Est-ce que l'incertitude sur la prédiction est disponible ?

- https://res.griis.ca/divers/oym_pre.html

Déclenchement de l’outil flexible

Qu’est-ce qui fait que le score est transmis ?

- Différence entre **production** et **mise à disposition**
  - Électrolytes dans une machine de lab

- Qu’est-ce qui déclenche la mise à disposition
  - **Prescription** ? Dépistage systématique ?
    - Scan des poumons à tous les patients fumeurs hospitalisés ?

Déploiement à coût raisonnable

Déploiement

- **Combien** de **variables** sont nécessaires ?
- Est-ce des variables **déjà captées** électroniquement naturellement ?
- Est-ce que les données sont **disponibles** au moment de l'utilisation de l'outil ?
  - Toujours facile de « prévoir » ce qui se passe durant une hospitalisation… une fois que l'hospitalisation est terminée

- Comment est **connecté** l'outil aux **systèmes informatiques** de l'hôpital ?

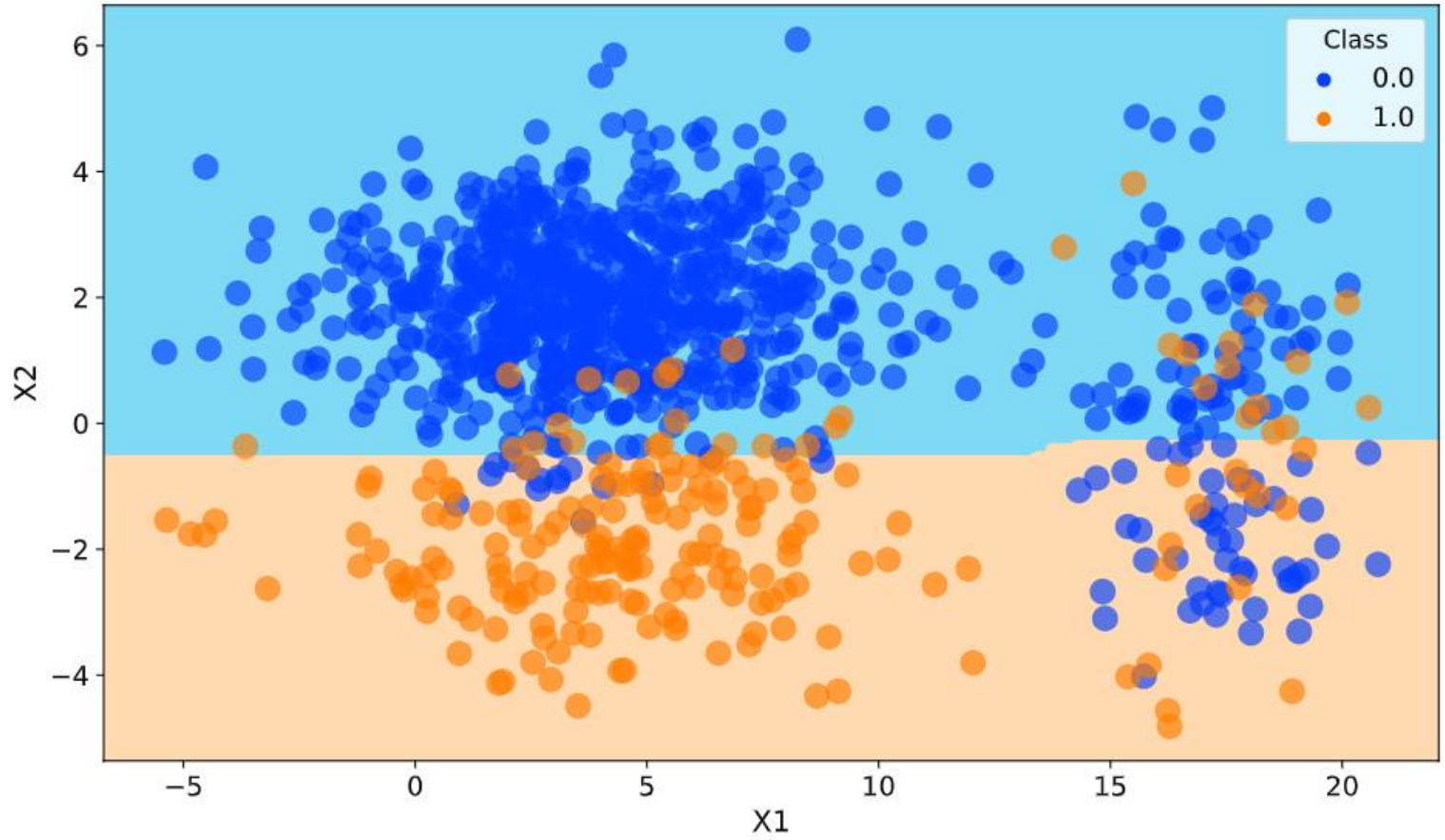
Capacité d’évaluation en continu

Une fois déployé, le défi est que l’outil demeure sécuritaire et pertinent

Évaluation en continu

- Au **départ**
- Impact de la **contamination** liée à l'utilisation du modèle
- **Changement** du **contexte** de soins
- **Extrapolation**

Au départ



Au départ

- Est-ce qu'il pourrait y avoir des **biais** ?
  - p. ex. : couleur de la peau
  - sous-traiter l'hypertension chez les femmes
- Est-ce que notre **population est similaire** ?
  - Est-ce que nous avons une sous-population qui n'était pas présente dans l'entrainement ?

Contamination positive

- Une fois que les **cliniciens** sont **sensibilisés**
  - Exemple : utilité sur les discussions de fin de vie… si tout le monde le fait

- Quand faut-il **penser à retirer un modèle** car le retour sur investissement est trop faible ?

Changement de contexte

- **Monde extérieur**
- Exemple : covid
  - « Valeur » d'une **visite à l'urgence**
    - Covid : il fallait être beaucoup plus malade pour y aller
    - Proxy pour autre chose… relation qui change ?
    - Seuil qui change ?

Extrapolation

- Suivre **l'utilisation** clinique **réelle**
- Équivalent d'un médicament utilisé hors indication
- Évaluation à faire
  - Performance, utilité, risques, bénéfices

Communication des changements du modèle

Changements au modèle

- Est-ce que vous êtes **notifié** si le modèle est **ré-entrainé** ?
  - Semblable à une nouvelle machine de scinti avec sensibilité augmentée
- Est-ce que le modèle peut **changer en continu** ?
  - Donc deux résultats différents d'une journée à l'autre pour le même patient
- Qu'est-ce qu'on va vous communiquer ?

Enjeux de gestion

Contrôle de ce qui sort de votre organisation

Contrôle ?

- Est-ce que vos **données** sont utilisées pour entrainer le modèle qui sera **revendu** ailleurs ?
- Est-ce que les données de **vos patients** sont **dans le modèle** ?
  - Les extrêmes y seront (p. ex. : très riche ou très pauvre) probablement
- Comment gérer le **consentement** ?
  - Modèle peu adapté pour un apprentissage « partiel »

Impacts sur l'organisation des soins

Qui gagne et qui perd ?

Organisation des soins

- **Impact coûts** et **ressources**

- Exemple : Analyse des Rx poumons et recommandation de scan
  - Sensibilité vs spécificité ?
  - Avalanche de scans ?
  - Retards pour d'autres indication ?

- **Documentation**
  - Qu'est-ce qui est stocké : résultats ? Intrants ? Version ?
    - Et si le modèle change toujours, comment faire un audit ?

Gagnants et perdants

- **Qui** va **gérer** les conséquences ?
  - Exemple : Fin de vie - tout le monde va se mettre à consulter les soins palliatifs ?

- Est-ce que les **gains** cliniques sont **pour ceux impactés** au niveau des ressources ?
  - Exemple : héparine faible poids moléculaire

Références additionnelles

- Smith, Barry (2023). ChatGPT: Not Intelligent. Ai: From Robotics to Philosophy the Intelligent Robots of the Future – or Human Evolutionary Development Based on Ai Foundations.
  - Systèmes complexes : narrow AI vs general AI
  - https://philpapers.org/archive/SMICNI.pdf

- Gradient descent, how neural networks learn
  - 3blue1brown : https://www.3blue1brown.com/lessons/gradient-descent

École d'été interdisciplinaire en numérique de la santé

https://eins.griis.ca/

jf.ethier@usherbrooke.ca

GRIIS Université de Sherbrooke Réseau de recherche sur les données de santé du Canada Health Data Research Network Canada