

**Apprentissage Artificiel et Interprétabilité pour
 la médecine de précision**

JEAN-DANIEL ZUCKER
 DR IRD

Travail en collaboration Edi Prifti (ICAN), Yann Chevalyre (Dauphine),
 Blaise Hanczar (Evry) & Nataliya Sokolovska (SU)

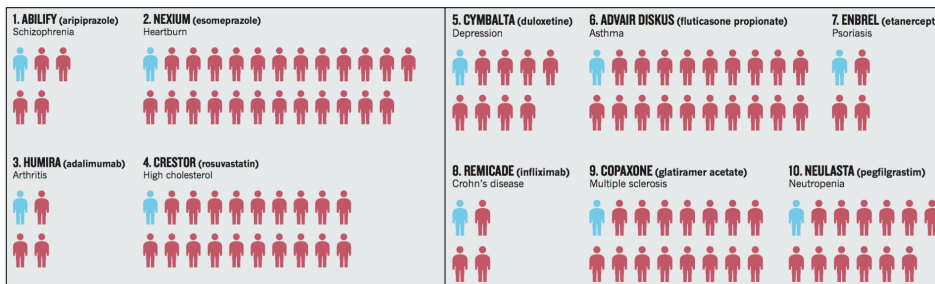


Plan

- I. Médecine de précision et big (Omics) data
- II. Apprentissage Artificiel et médecine de précision
- III. Pourquoi des modèles interprétables en médecine ?
- IV. Deux (ou un) exemples de modèles interprétables
- V. Conclusion

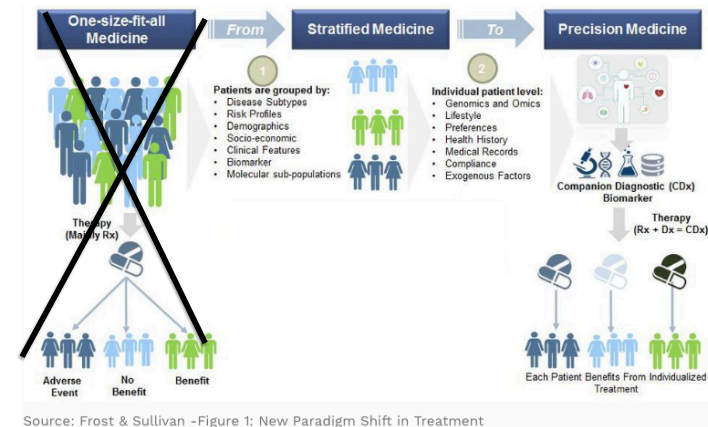
Malgré ses succès la médecine reste « imprécise »...

Pour **chaque** personne qu'ils **aident** (bleu), les **dix médicaments** les plus lucratifs aux États-Unis ne parviennent pas à améliorer les conditions d'entre **3** et **24** personnes (rouge).



Schorf, N.J., 2015. Personalized medicine: time for one-person trials. Nature.

... espoir d'une fin de la médecine « one-size-fit all »

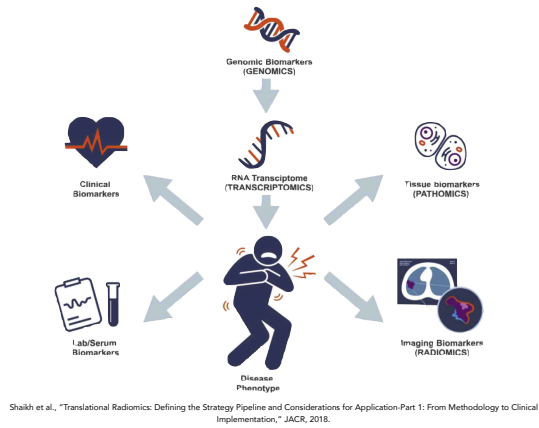


Source: Frost & Sullivan - Figure 1: New Paradigm Shift in Treatment

→ la bonne intervention au bon patient au bon moment.

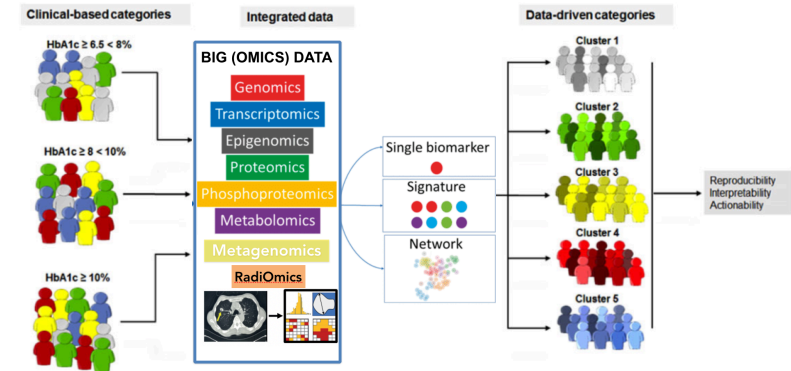
... et l'émergence de la médecine de précision...

→ fournir les meilleurs soins disponibles à chaque patient, sur la base d'une stratification en sous-classes de maladies sur une base biologique commune.



Par exemple: médecine de précision pour le diabète

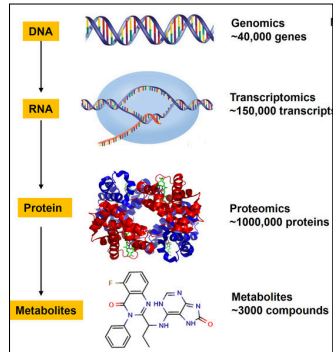
En Afrique, le diabète (5,7% de la population adulte est maintenant affectée) et les maladies cardiovasculaires tuent plus que le SIDA.



J. Merino and J. C. Florez, "Precision medicine in diabetes: an opportunity for clinical translation," Ann. N.Y. Acad. Sci., vol. 1411, no. 1, pp. 140-152, Jan. 2018.

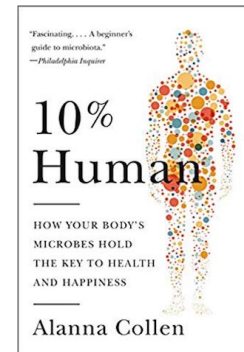
Les données « Omiques » permettent de nous caractériser très

Analyser nos propre cellules



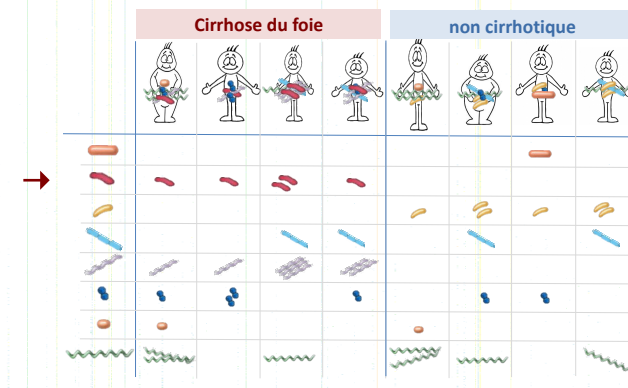
F. S. MD, et al., "Translational Radiomics: Defining the Strategy Pipeline and Considerations for Application-Part 1: From Methodology to Clinical Implementation," Journal of the American College of Radiology, 2018.

Analyser nos bactéries



Metagenomics
~10,000,000 de gènes

Vers une médecine de précision des maladies intégrant la métagénomique.

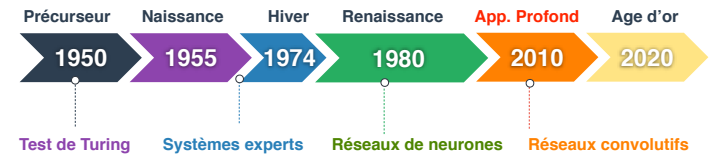
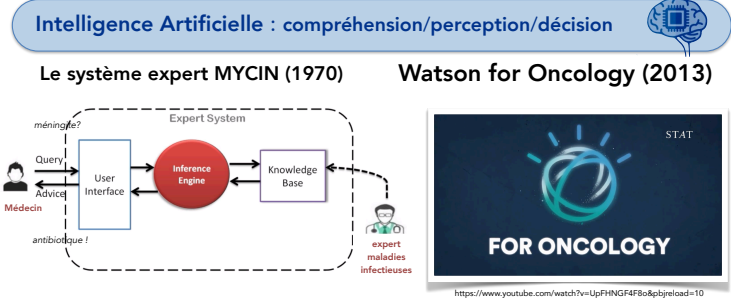


Signature géniques impliquant plusieurs gènes bactériens parmi des millions de gènes ...

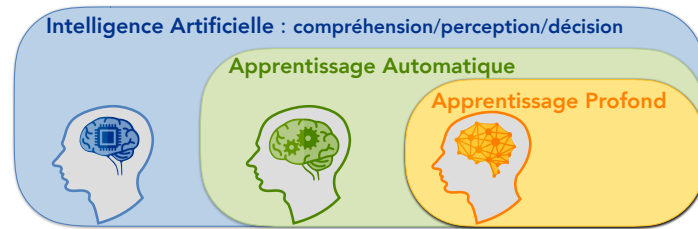
Plan

- I. Médecine de précision et big (Omics) data
- II. Apprentissage Artificiel et médecine de précision
- III. Pourquoi des modèles interprétables en médecine ?
- IV. Deux (ou un) exemples de modèles interprétables
- V. Conclusion

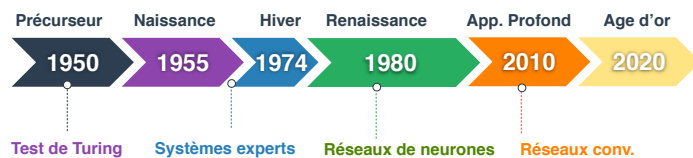
IA et médecine... une longue histoire



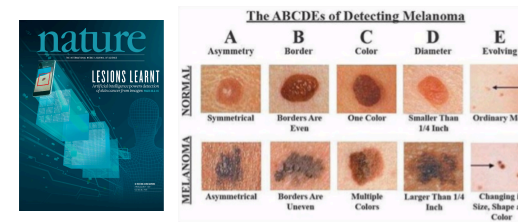
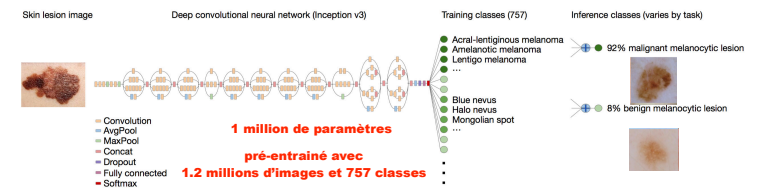
...mais l'IA et les données massives



... transforme la médecine... c'est déjà presque une vieille nouvelle !



Classification des cancer de la peau du niveau d'un expert dermatologue (Nature, 2017)



A. Esteva, et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, 2017.

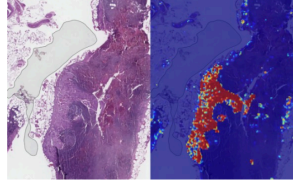
Des systèmes basés sur une collaboration homme-machine peuvent faire mieux que l'IA seule...

INTELLIGENCE ARTIFICIELLE SCIENCE

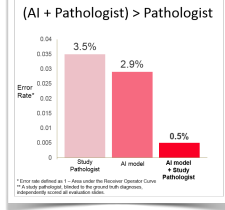
Google détecte le cancer du sein métastatique avec une précision de 99%

Google a largement investi dans le développement des applications de santé.

PAR VALENTIN CHENAO - @CHENAO - 13 OCTOBRE 2018



<https://www.ledevoir.com/technologie/532222/google-detecte-le-cancer-du-sein-metastatique-avec-une-precision-de-99>



INTELLIGENCE ARTIFICIELLE

Google a créé une intelligence artificielle capable de détecter le cancer du poumon

Les chercheurs de Google et l'hôpital universitaire de Northwestern se sont unis pour élaborer une IA capable de détecter le cancer du poumon

PAR EMILIE DUBOIS - @EMILIEDUBOIS - 13 OCTOBRE 2018



Deep Learning Drops Error Rate for Breast Cancer Diagnoses by 85%

JAMA, vol. 318, no. 22, pp. 2199-2210, Dec. 2017.

13
JDZucker

L'approbation des usages médicaux de l'IA est en marche... forte des performances en prédiction...

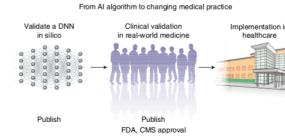


Table 1 | Peer-reviewed publications of AI algorithms compared with doctors

Specialty	Images	Publication
Radiology/ neurology	CT head, acute neurological events	Titano et al. ¹⁹
Pathology	Breast cancer	Ehteshami Bejnordi et al. ¹⁵
	Lung cancer (+ driver mutation)	Coudroy et al. ²⁰
	Brain tumors (+ methylation)	Capper et al. ¹⁶
Dermatology	Skin cancers	Esteva et al. ¹⁷
	Melanoma	Haenssle et al. ¹⁸
Ophthalmology	Skin lesions	Hon et al. ²¹
	Diabetic retinopathy	Gulshan et al. ²²
Gastroenterology	Polyps at colonoscopy*	Mori et al. ²³
	Polyps at colonoscopy	Wang et al. ²⁷
Cardiology	Echocardiography	Madani et al. ²⁴
	Echocardiography	Zhang et al. ²⁵

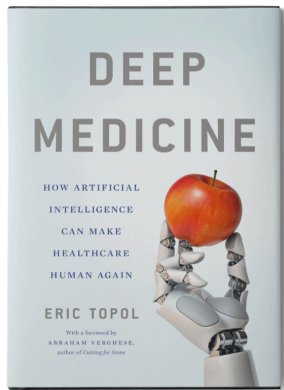
E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat Med, pp. 1-13, Jan. 2019.

14
JDZucker

Table 2 | FDA AI approvals are accelerating

Company	FDA Approval	Indication
Apple	September 2018	Atrial fibrillation detection
Aidoc	August 2018	CT brain bleed diagnosis
iCAD	August 2018	Breast density via mammography
Zebra Medical	July 2018	Coronary calcium scoring
Bay Labs	June 2018	Echocardiogram EF determination
Neural Analytics	May 2018	Device for paramedic stroke diagnosis
IDx	April 2018	Diabetic retinopathy diagnosis
Icometrix	April 2018	MRI brain interpretation
Imagen	March 2018	X-ray wrist fracture diagnosis
Viz.ai	February 2018	CT stroke diagnosis
Arterys	February 2018	Liver and lung cancer (MRI, CT) diagnosis
MaxQ-AI	January 2018	CT brain bleed diagnosis
Alivecor	November 2017	Atrial fibrillation detection via Apple Watch
Arterys	January 2017	MRI heart interpretation

Mais ce n'est pas la fin des radiologistes ! ... juste le début d'une médecine plus humaine ?



« Pour l'instant, **remplacer un radiologiste n'est juste pas possible**. Gregory Moore, vice-président Healthcare chez Google puis Microsoft (!) et lui-même radiologue, a observé qu'**il faut littéralement des milliers d'algorithmes pour pouvoir reproduire ce qu'un radiologiste peut faire un jour donné.** »

« En **libérant les médecins** des tâches qui interfèrent avec la connexion humaine, **l'IA créera un espace pour la véritable guérison qui a lieu entre un médecin qui peut écouter et un patient qui doit être entendu.** »

(Topol, 2019)

15
JDZucker

Plan

- I. Médecine de précision et big (Omics) data
- II. Apprentissage Artificiel et médecine de précision
- III. Pourquoi des modèles interprétables en médecine ?
- IV. Deux (ou un) exemples de modèles interprétables
- V. Conclusion

16
JDZucker

RGPD et modèles interprétables : droit et confiance

- Règlements de l'UE (Règlement général sur la protection des données (GDPR) en vigueur le 25 mai 2018) sur la prise de décision algorithmique et un "droit d'explication".

Goodman, B. & Flaxman, S. R. European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". AI magazine, 2017

- Une explication de la prédiction est désirée par médecins et patients lorsque un **modèle** doit être validé avant d'être déployé en routine → **confiance**

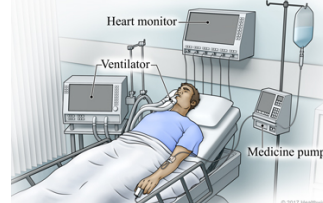
Vanthienen, et al. Performance of classification models from a user perspective. *Decision Support Systems* 51, 782-793,(2011).

erreur pas (trop) grave...



Can you smell the perfect partner?
If Tim Dowling married the right woman? A new book suggests that we unconsciously select the perfect partner by sniffing out their compatibility genes. He talks to the author about MIT, genetics and alleles - then nervously asks his wife to take a DNA test

erreur gravissime...



A. Vellido, et al., "Machine learning in critical care: state-of-the-art and a sepsis case study," BMEO,2018.

17
JDZucker

Equité/Fairness : l'IA est biaisée par les données...



Fair

- Demographic fairness
- Fairness in design
- Fairness in data
- Fairness in algorithms
- Fairness in outcomes

... L'application basée sur l'IA peut être exposée, en quelques minutes, à plus de cas qu'un clinicien ne pourrait en voir dans de nombreuses vies. L'application basée sur l'IA est capable de surpasser les dermatologues ...

Buch, V. H. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract* (2018).

OUI MAIS



Une **étude** récente a révélé que certains programmes de reconnaissance faciale **classent incorrectement moins de 1 % des hommes à la peau claire**, mais plus d'un **tiers des femmes à la peau foncée**.

Larry Hardesty | MIT News Office
February 11, 2018

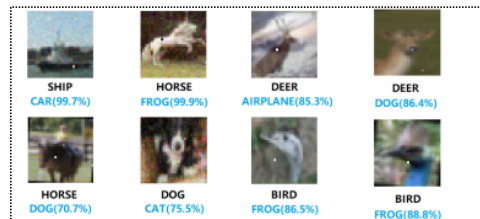
Que se passe-t-il lorsque l'on se fie à de tels algorithmes pour diagnostiquer le mélanome sur une peau claire ou foncée... ?



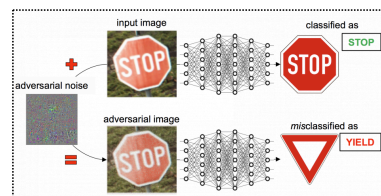
Un programme apprend à partir des données qu'on lui donne et qui peuvent donc être ... biaisées

18
JDZucker

On sait aussi que l' « IA » peut aussi être menacée par des « attaques adversariales »...



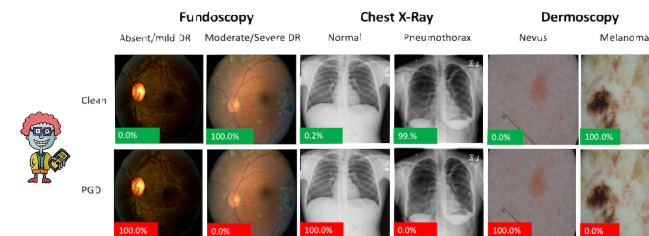
J. Su, et al. "One pixel attack for fooling deep neural networks," CoRR, 2017.



Akhtar & Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *arXiv.org*, 02-Jan-2018.

19
JDZucker

...se pose la problème de la « responsabilité » des algorithmes



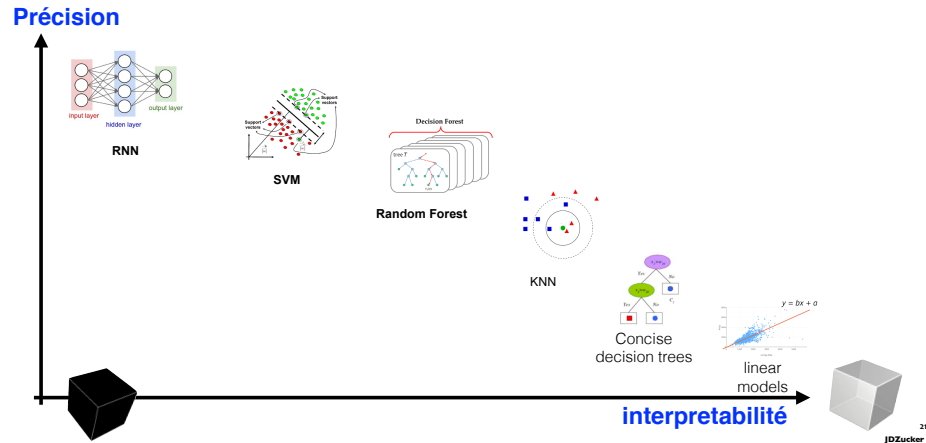
Finlayson, et al., "Adversarial Attacks Against Medical Deep Learning Systems.," *arXiv*, 2018.

Qui est responsable en cas d'erreur ?

Expliquer la décision de la machine peut éviter des erreurs !

20
JDZucker

En Machine Learning il y a souvent un compromis entre performance/précision et interprétabilité



Interprétabilité et Machine Learning

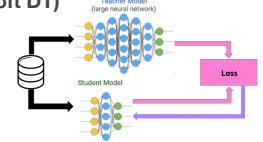
Type A - Interpréter les modèles boîtes noires

- Saliency Maps



Rajpurkar, et al. (2018).

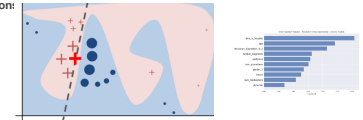
Model distillation (soft DT)



Type B - Interpréter les prédictions des modèles boîtes noires

Stratégie basée sur les perturbations pour générer des explications locales pour les prédictions:

- Feature Attribution methods: LIME (Ribeiro et al. '16)
- Feature Importance : SHAP values (Lundberg, NIPS'17)



Type C - Apprendre des modèles intrinsèquement interprétables

- Arbres de décision, Modèles de score, Modèles linéaire, Régression régularisée, ...
- « Désenchevêtrement de représentation » Fonction objectif (InfoGAN, Chen&al.16)

A- Interpréter les boîtes noires

Prediction des facteurs de risque cardiovasculaire à partir de photographies du fond de la rétine par l'apprentissage profond Poplin, et al. Nature Medicine 2018.

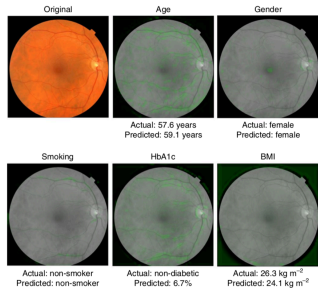


Table 6 | Percentage of the 100 attention heat maps for which doctors agreed that the heat map highlighted the given feature

Risk factor	Vessels (%)	Optic disc (%)	Non-specific features
Age	95	33	38
Gender	71	78	50
Current smoker	91	25	38
HbA1c	78	32	46
SBP	98	14	54
DBP	29	5	97
BMI	1	6	99

Heat maps (n=100) were generated for each risk factor and then presented to three ophthalmologists who were asked to check the features highlighted in each image (n=300 responses for each risk factor). The images were shuffled and presented as a set of 700, and the ophthalmologists were blinded to the output prediction of the heat maps and the ground-truth label. For the variables that were present in both datasets (age and gender), the most commonly highlighted features were identical in both datasets.

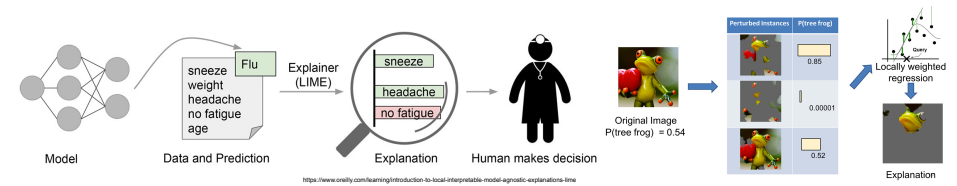
Prediction de facteurs de risque cardiovasculaire que l'on ne croyait pas présents/quantifiables ds les images rétiniennes

- l'âge (erreur absolue moyenne en 3.26 ans),
- le sexe (aire sous la courbe caractéristique de fonctionnement du récepteur (AUC) = 0,97),
- le statut tabagique (AUC = 0,71),
- la tension artérielle systolique (erreur absolue moyenne inférieure à 11,23 mmHg) et
- les événements cardiaques indésirables majeurs (AUC = 0,70).

B. Interpreter les prédictions des boîtes noires

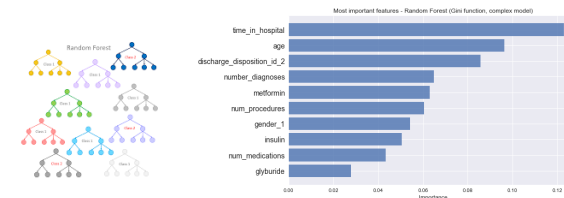
LIME (Local Interpretable Model agnostic Explanations) Ribeiro&al. "Why Should I Trust You?" 2016

→ convertir les prédictions en un point en un modèle interprétable : séparateur linéaire.



Feature Importance

→ La diminution totale de l'impureté du noeud (pondérée par la probabilité d'atteindre ce noeud) calculée en moyenne sur tous les arbres de l'ensemble.



C- Modèles « Intrinsèquement » interprétables

► Critère d'interprétabilité

- Concision (nombre de noeuds d'un arbre)
- Modèles applicables " manuellement " pour obtenir une décision
- Opérations simples (+, -, *, opérateurs logiques), valeurs entières

► **Exemple 1:** Modèle linéaire discret $y \sim x_1 - x_4 + x_5 + x_8 - x_{14}$

► **Exemple 2:** Scores

Table 2. The HOTEL score.

Variable	Points
Systolic blood pressure (mmHg) <100	1
Oxygen saturation (%) <90	1
Temperature <35 C	1
Abnormal ECG	1
Unable to stand unaided	1

doi:10.1371/journal.pone.0105695.t002

Kalish et al 2008

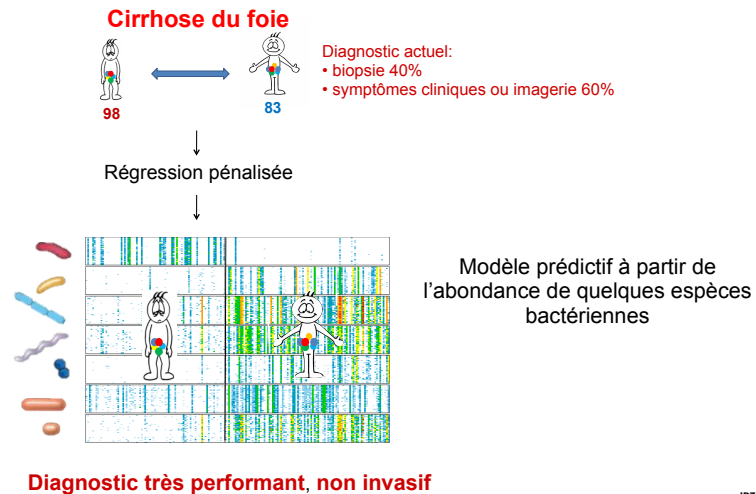
25
JDZucker

Plan

- I. Médecine de précision et big (Omics) data
- II. Apprentissage Artificiel et médecine de précision
- III. Pourquoi des modèles interprétables en médecine ?
- IV. Deux (ou un) exemples de modèles interprétables
- V. Conclusion

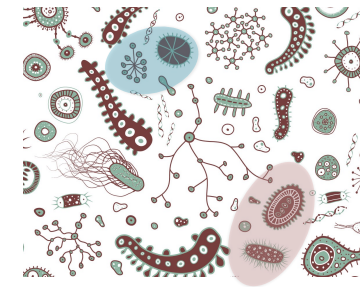
26
JDZucker

Exemple 1: diagnostic métagénomique interreprétable et précis



27
JDZucker

Inspirés par les relations écologiques, nous avons proposé les modèles Bin/Ter/Ratio (BTR), qui quantifient les interactions en comparant l'abondance des taxons.



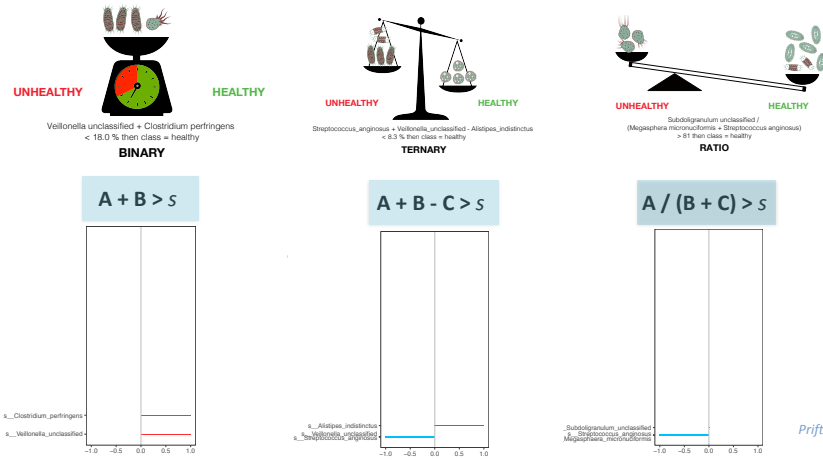
#Veillonella unclassified + #Clostridium perfringens
< 18.0 % then class = healthy

Prijfti et al. (in press)

28
JDZucker

Les modèles BTR sont extrêmement simple ...

JDZucker

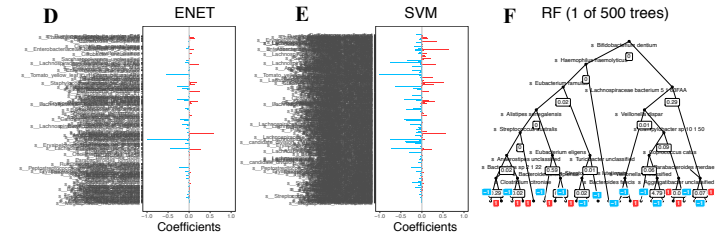


29

JDZucker

... comparé à l'état de l'art (SOTA)

JDZucker



- ENET sélectionne 159 variables sur 462.
- SVM avec noyau linéaire utilise les 462 variables.
- Seulement 1 des 500 arbres utilisés dans Random Forrest est représenté ici.

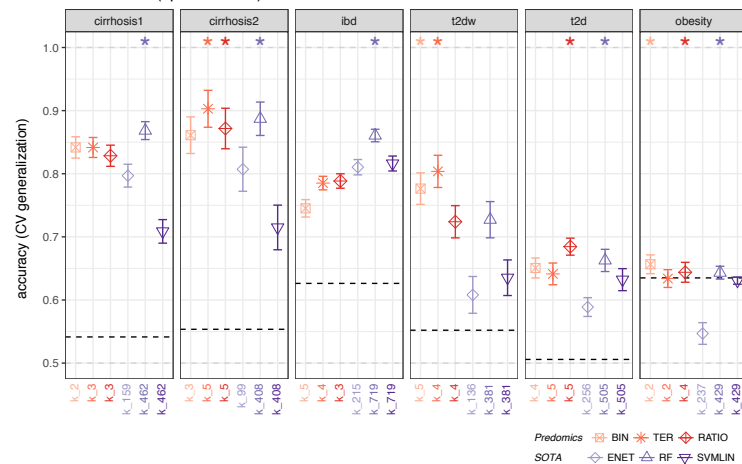
Prifti et al. (in press)

30

JDZucker

Les modèles BTR affichent des performances similaires avec SOTA pour tous les ensembles de données.

JDZucker



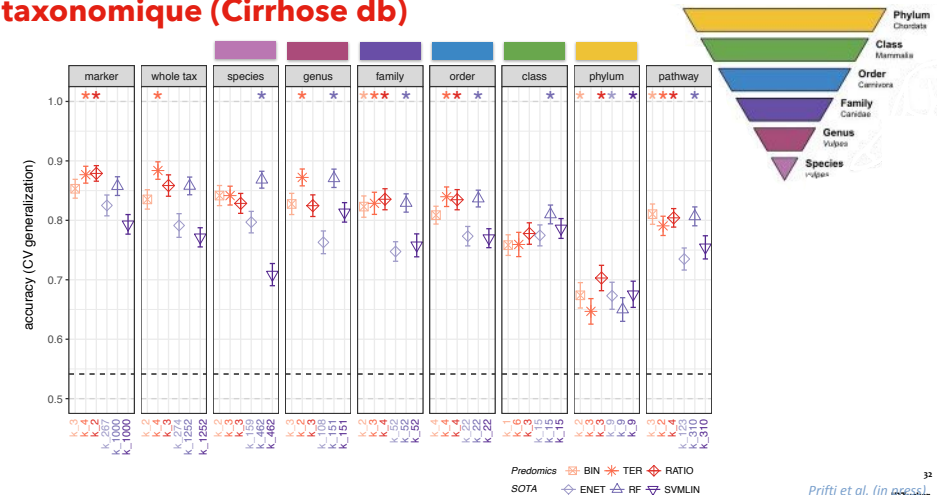
Prifti et al. (in press)

31

JDZucker

La performance prédictive diminue avec la spécificité taxonomique (Cirrhose db)

JDZucker

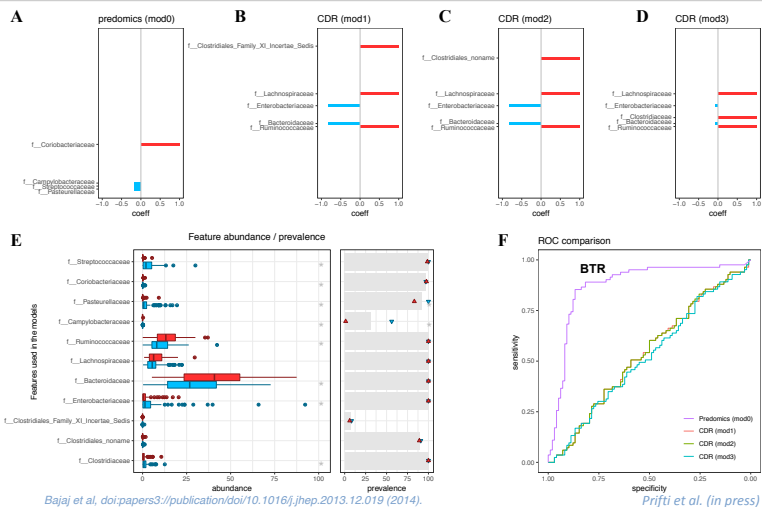


Prifti et al. (in press)

32

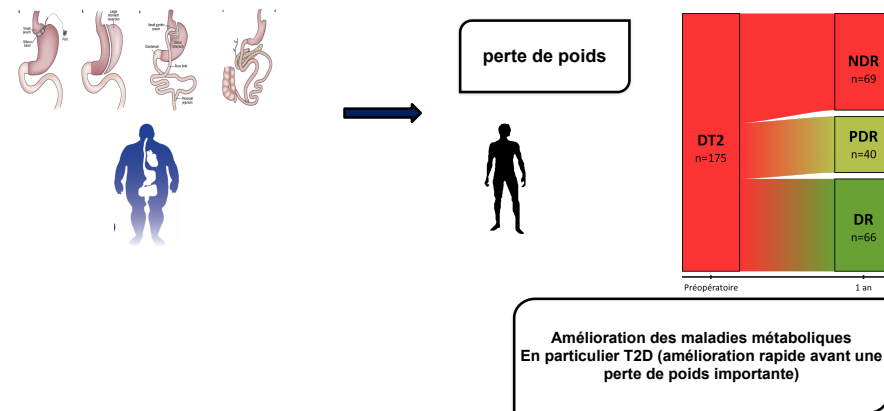
JDZucker

BTR model outperforms Cirrhosis Dysbiosis Ratio (from expert knowledge)



33
JDZucker

2-ème exemple : le score DiaRem dans la chirurgie bariatrique améliore le diabète de type 2 et le est utilisé pour prédire sa rémission



34
JDZucker

Score Diarem : meilleur score (2013) ; validé sur des cohortes indépendantes.

	Score
Age (years)	
<40	0
40-49	1
50-59	2
≥60	3
HbA_{1c} (%)	
<6.5%	0
6.5-6.9%	2
7.0-8.9%	4
≥9.0%	6
Other diabetes drugs	
No sulfonylureas or insulin-sensitising agent other than metformin	0
Sulfonylureas and insulin-sensitising agent other than metformin	3
Treatment with insulin	
No	0
Yes	10

Total score calculated by adding scores for each of the four variables.

Table 5: Calculation of DiaRem score for prediction of the probability of diabetes remission after Roux-en-Y gastric bypass surgery

Still et al Lancet endocrinology 2013;

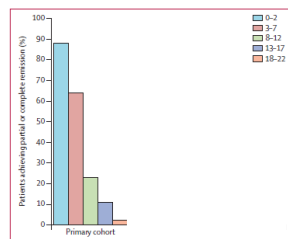


Figure 4: Proportion of patients in each cohort achieving partial or complete remission at 14 months after surgery, by DiaRem score

Table 2: Prediction errors of the different models in percentage

	Train error	Test error	Cross-validation error
Hayes - SL	13.4	31	
Hayes - J48DT	12.6	34.5	
Dixon - LR6	16	40.5	
Dixon - LR7	12	22.6	
Liou - Score	27.3	16.7	16.7
Still - Score	19.4	15.5	15.9
LR	7.1	19.9	
DT	13.1	17.6	
Lasso	14.3	18.9	
EN	13.1	17.2	

35
JDZucker

Reconstruction automatisée de scores de Diarem

1. Identification of related clinical variables

age | glycated hemoglobin | insulin | other drugs

2. Meaningful thresholds for clinical variables

<40 40-49 50-59 >60 | <6.5 6.5-6.9 7-8.9 >9 | insulin | other drugs
yes no | yes no

3. Optimization of weights for sub-groups of the variables

<40 40-49 50-59 >60 | <6.5 6.5-6.9 7-8.9 >9 | insulin | other drugs
0 1 2 3 | 0 2 4 6 | 10 0 | 3 0

4. Find an optimal separator between two classes

Classify as Remission if sum of scores < 7

Classify as Non-remission if sum of scores ≥ 7

Sokolovska, N., Y. Chevalleyre and J.-D. Zucker (AISTATS 2018).
Sokolovska, N., Y. Chevalleyre and J.-D. Zucker (DA2PL2016)

36
JDZucker

Construction de score formalisée comme un problème d'optimisation

We minimise the hinge loss penalized by the Fused Lasso:

$$\sum_{i=1}^N \ell(y_i, \theta \cdot \bar{x}_i + b) + \lambda \sum_{j=1}^{\bar{d}-1} |\theta_j - \theta_{j+1}|.$$

The linear programming formulation of the problem:

$$\begin{aligned} \min & \left(\sum_{i=1}^N \xi_i + \sum_{j=1}^{\bar{d}} \eta_j \right), \text{ such that} \\ & \text{for all } i, y_i(\theta \cdot \bar{x}_i + b) \geq 1 - \xi_i, \\ & \text{for all } j, -\lambda \eta_j \leq \theta_j - \theta_{j+1} \leq \lambda \eta_j, \\ & \xi_i \geq 0, \theta_i \in \mathbb{N} \text{ for all } i. \end{aligned}$$

Fully Corrective Binning (FCB) algorithm

Sokolovska, N., Y. Chevaleyre and J.-D. Zucker (AISTATS 2018).
Sokolovska, N., Y. Chevaleyre and J.-D. Zucker (DA2PL'2016)

37
JDZucker

Le score DiaRem avancé: l'AdDiaRem

En utilisant notre algorithme Fully Corrective Binning (FCB)

Age		Other glucose-lowering drugs	
[15 – 41]	0	No	0
(41 – 52]	3	Yes	1
(52 – 69]	5	Number of glucose-lowering drugs	
HbA1c		0	0
[4.5 – 6.9]	0	1	1
(6.9 – 7.4]	2	2	2
(7.4 – 18.4]	4	≥ 3	3
Insuline		Diabetes duration	
No	0	[0 – 6.9]	0
Yes	3	(6.9 – 14]	3
		≥ 14	5

La procédure d'apprentissage s'appuie sur IBM ILOG CPLEX Optimization Studio2 qui effectue l'optimisation sous contrainte efficacement. En particulier, des contraintes d'intégrité sont ajoutées au problème d'optimisation pour obtenir des solutions entières.

ANR DiagnoLearn N. Sokolovska (PI)

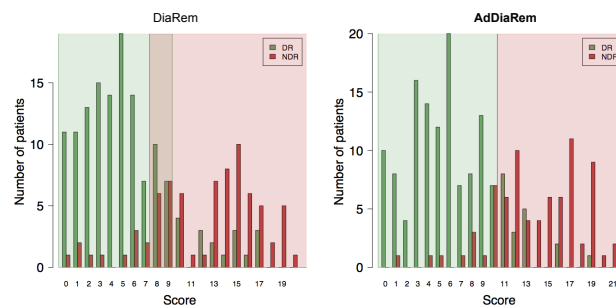


2018-2020

38
JDZucker

L'AdDiaRem est aujourd'hui utilisé en routine

- ▶ New biomarkers (diabetes duration, number of drugs taken)



The distributions of the DiaRem and AdDiaRem scores

J. Aron-Wisniewsky et al., *Diabetologia*, 2017

DiaRem AUC=81%
Ad-DiaRem AUC=84%

39
JDZucker

Plan

- I. Médecine de précision et big (Omics) data
- II. Apprentissage Artificiel et médecine de précision
- III. Pourquoi des modèles interprétables en médecine ?
- IV. Deux (ou un) exemples de modèles interprétables
- V. Conclusion

40
JDZucker

2 vues sur l'interprétabilité et le machine learning

JDZucker

PERSPECTIVE

<https://doi.org/10.1038/nature.2019.0048>

nature
machine intelligence

Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin  May 2019

Black box machine learning models are currently being used for high-stakes decision making throughout society, causing problems in healthcare, criminal justice and other domains. Some people hope that creating methods for explaining these black box models will alleviate some of the problems, but trying to explain black box models, rather than creating models that are interpretable in the first place, is likely to perpetuate bad practice and can potentially cause great harm to society. The way forward is to design models that are inherently interpretable. This Perspective clarifies the chasm between explaining black boxes and using inherently interpretable models, outlines several key reasons why explainable black boxes should be avoided in high-stakes decisions, identifies challenges to interpretable machine learning, and provides several example applications where interpretable models could potentially replace black box models in criminal justice, healthcare and computer vision.

NATURE MACHINE INTELLIGENCE | VOL 1 | MAY 2019 | 206-215 | www.nature.com/natmachintell

Prise de décisions à enjeux élevés santé, justice pénale, etc.

→ La voie à suivre est de concevoir des modèles intrinsèquement interprétables

Nous ne pouvons pas utiliser boîte noire AI pour trouver le lien de causalité, ou de compréhension
→ Acceptons les boîtes noires en ce qu'elles optimisent des résultats et suscitent l'inspiration.

ARTIFICIAL INTELLIGENCE

In defense of the black box

Black box algorithms can be useful in science and engineering **Avril 2019**
sciencemag.org SCIENCE

By Elizabeth A. Holm





The science fiction writer Douglas Adams imagined the greatest computer ever built, Deep Thought, programmed to answer the deepest question ever asked: the Great Question of Life, the Universe, and Everything. After 7.5 million years of processing, Deep Thought revealed its answer: Forty-two (2). As artificial intelligence (AI) systems enter every sector of human endeavor—including science, engineering, and health—humanity is confronted by the same conundrum that Adams encapsulated so succinctly: What good is knowing the answer when it is unclear why it is the answer? What good is a black box?

In an informal survey of my colleagues in the physical sciences and engineering, the top reason for not using AI methods such as deep learning, voiced by a substantial majority, was that they did not know how to interpret the results. This is an important objection, with implications that range from practical to ethical to legal (2). The goal of scientists and the responsibility of engineers is not just to predict what happens but to understand why it happens. Both an engineer and an AI system may learn to predict whether a bridge will collapse, but only the engineer can explain that decision in terms of physical models that can be communicated to and evaluated by others. Whose bridge would you rather cross? 3 APRIL 2019 • VOL 364 ISSUE 6435

41
JDZucker

Conclusion

JDZucker

-  **La médecine de précision** annonce un bouleversement dans la prise en charge des patients, leur parcours de soin et leur suivi grâce à l'IA.
-  **Nouveaux diagnostics moléculaires (omiques) et d'imagerie**
 - stratification des maladies → meilleurs diagnostic,
 - aide au pronostic → meilleurs choix des traitements,
 - désert médicaux → tri des patients les plus à risques.
-  Progrès de l'IA et du **Deep Learning** posent des **questions éthiques** sur son adoption en médecine : équité/confiance/transparence/**interprétabilité**
-  **L'IA doit aider les cliniciens** à être plus efficace mais **l'interprétabilité** joue un rôle clé pour la **confiance**, la compréhension des **biais**, et contribuer à la recherche de l'étiologie des maladies → « XAI » médicale

42
JDZucker

Merci à



Dr. Edi Prifti



Dr Nataliya Sokolovska



43