

Lire et comprendre un article d'intelligence artificielle en santé

Dr Benjamin Popoff

LTSI, UMR 1099, Université de Rennes

Service de Réanimation Chirurgicale, CHU de Rouen

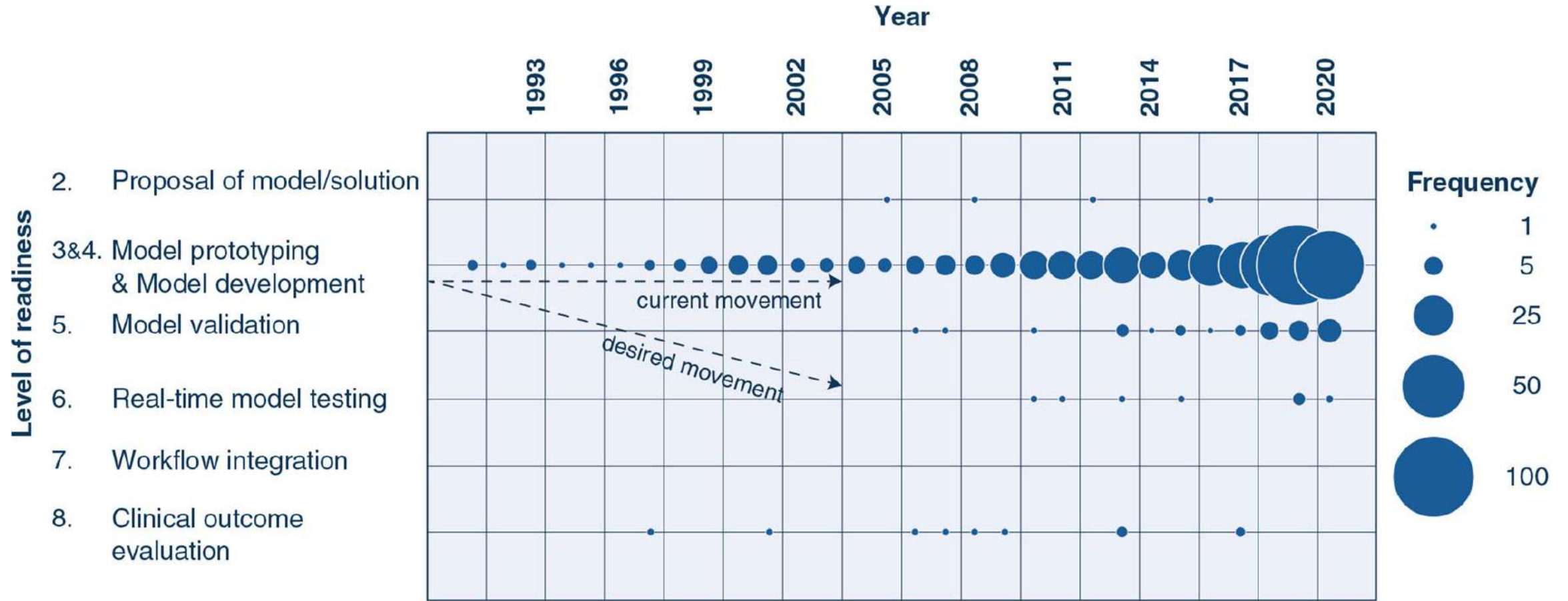


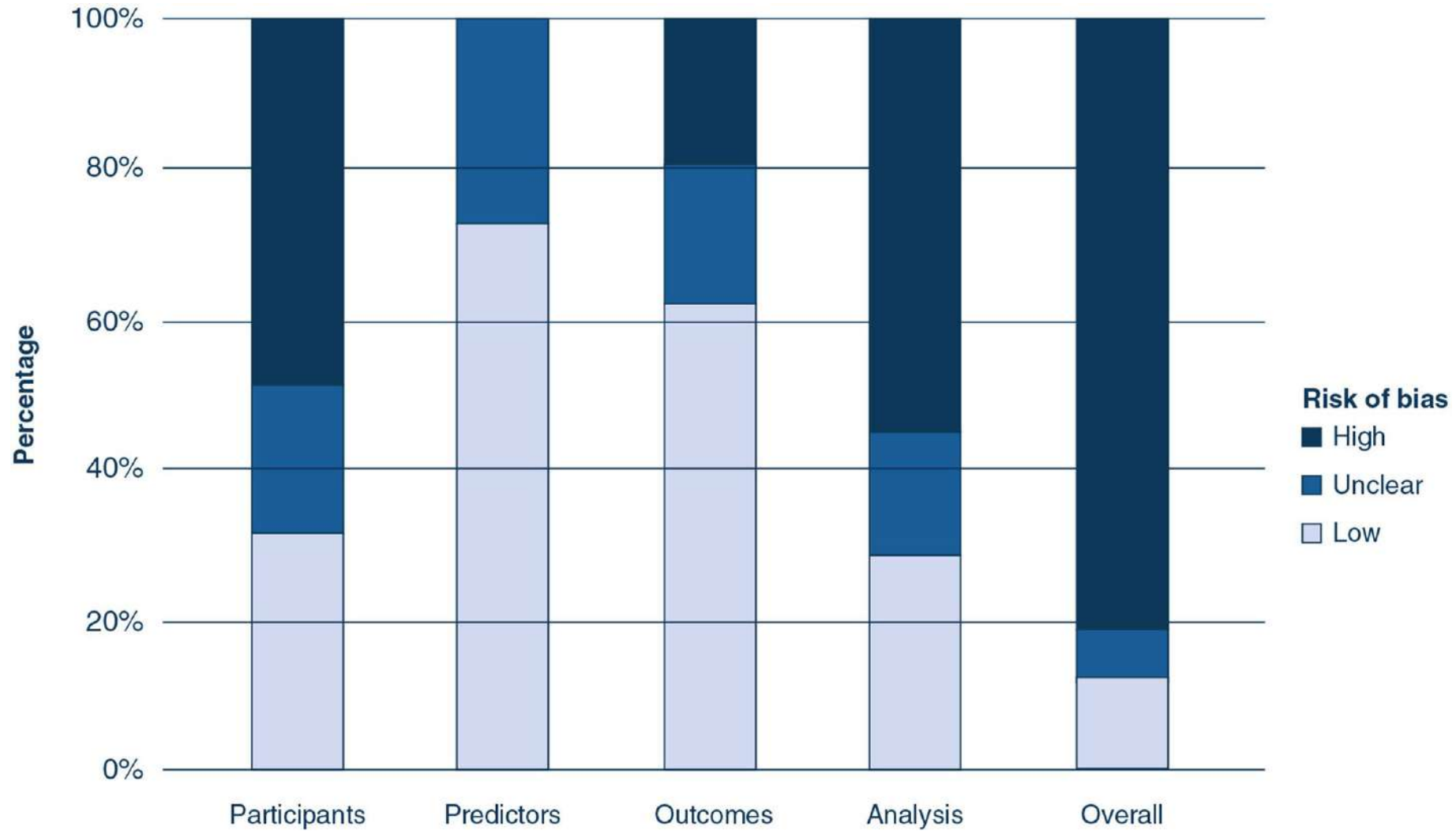
Normandie Université



Objectifs

- Connaître les bases des algorithmes d'IA/ML, leur cycle de développement et leur évaluation
- Savoir lire et critiquer un article d'IA/ML en santé, particulièrement en anesthésie-réanimation





Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu · Zhiyong Peng

Delays in identifying clinical deterioration in critically ill patients usually result in increased morbidity and mortality [17] while, conversely, early prediction of deterioration could provide a critical window of time for clinical interventions that might reduce severity [18]. Recently, a considerable body of literature has grown up around the theme of early detection of clinical outcomes in sepsis by utilizing advanced technologies [19–22], such as machine learning (ML) and artificial intelligence. However, most studies only included the first qualifying ICU admission during a single hospitalization, and excluded multiple qualifying ICU admissions [20–22]. Therefore, to our knowledge, no study has yet developed a specific ML model for predicting in-hospital mortality in septic patients requiring ICU readmission.

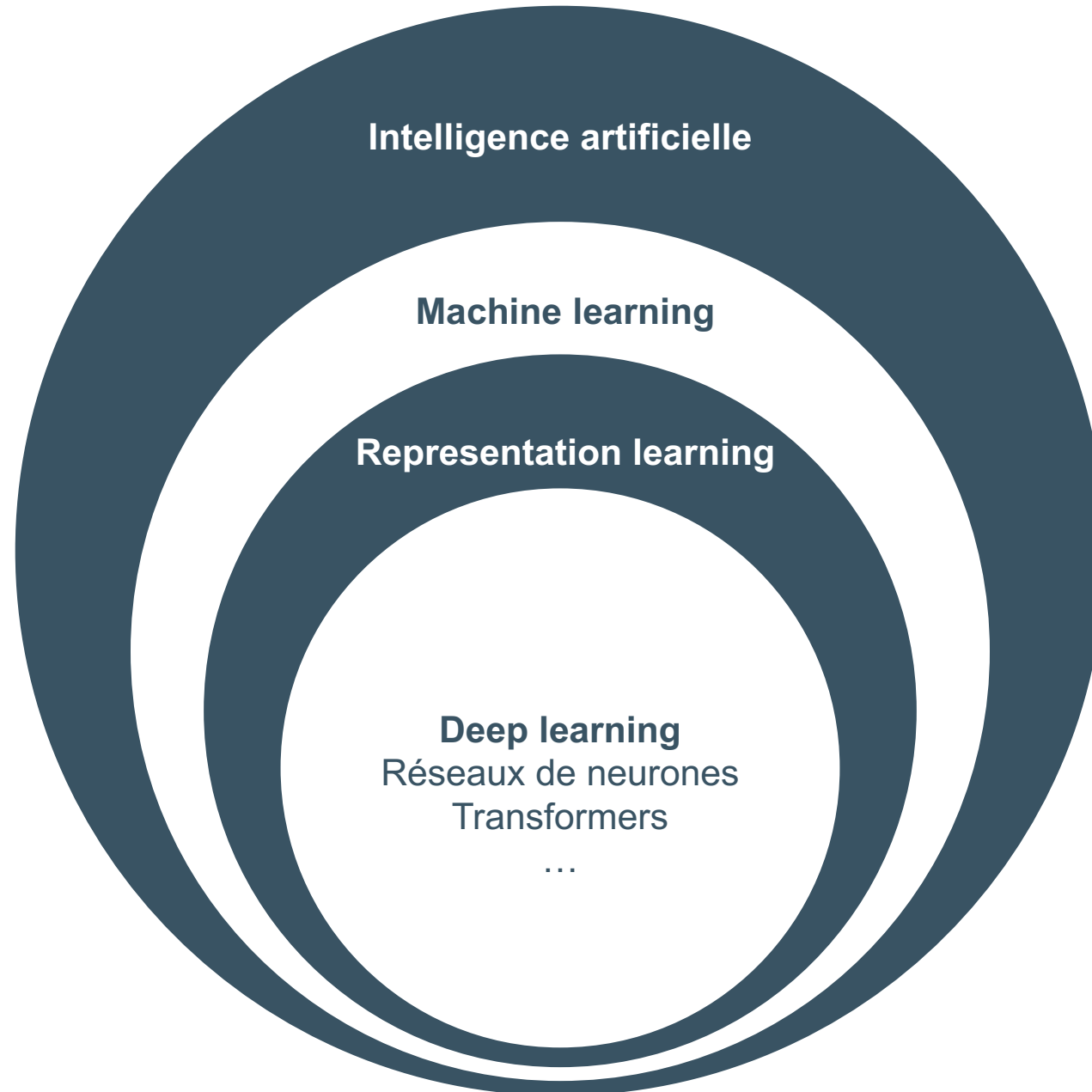
- **Objectif** : développement d'un modèle d'apprentissage automatique prédictif de la mortalité en réanimation chez les patients septiques nécessitant une réadmission en réanimation

Intelligence artificielle

« Ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine » (Larousse)

« La construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique » (Marvin Lee Minsky)

Machine learning (apprentissage automatique) : champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d'« apprendre » à partir de données



Intelligence artificielle

Machine learning

Representation learning

Deep learning

Réseaux de neurones

Transformers

...

Types de modèle

Classifier

Prédire

Agir

Méthodes

- **Source de données**

- Base MIMIC IV
- Hospitalisations en réanimation au Beth Israel Deaconess Medical Center (Boston, MA) de 2008 à 2019
- Données des 24 premières heures de la réadmission

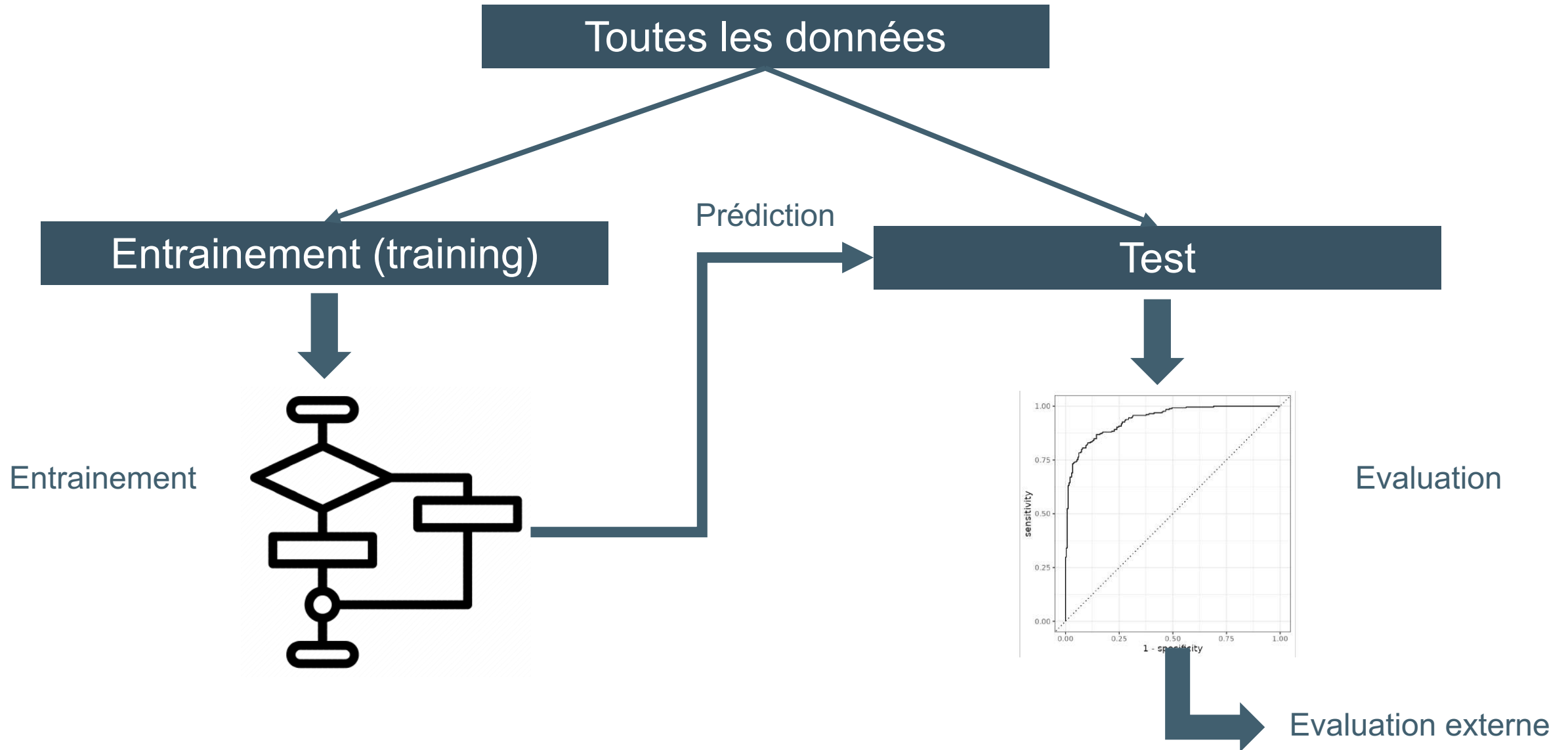
- **Critères de sélection**

- Patients adultes, réadmis en réanimation pour sepsis (critères Sepsis 3)
- Non inclusion des séjours $< 24h$, $\geq 3^{\text{ème}}$ admission

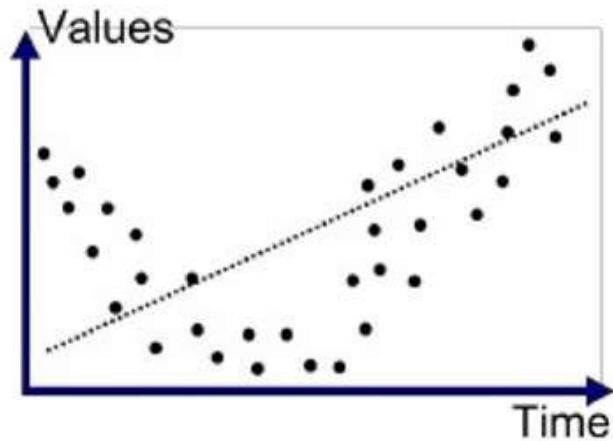
- **Séparation en deux jeux de données**

- Entraînement (75%)
- Test/validation (25%)

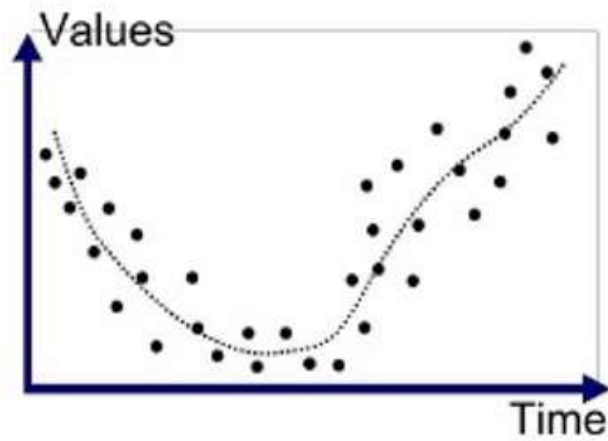
Séparation des données



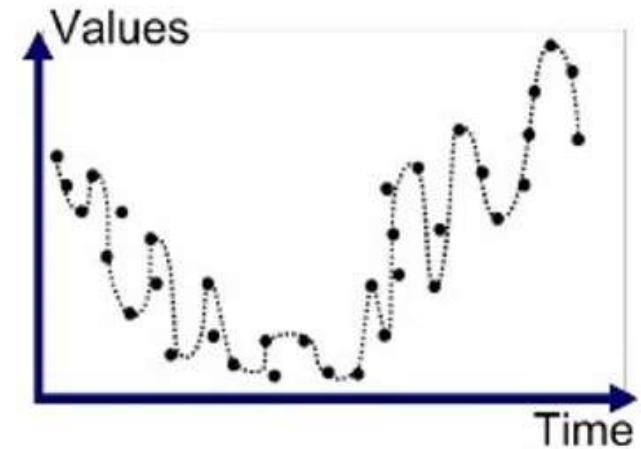
Surapprentissage



Underfitted



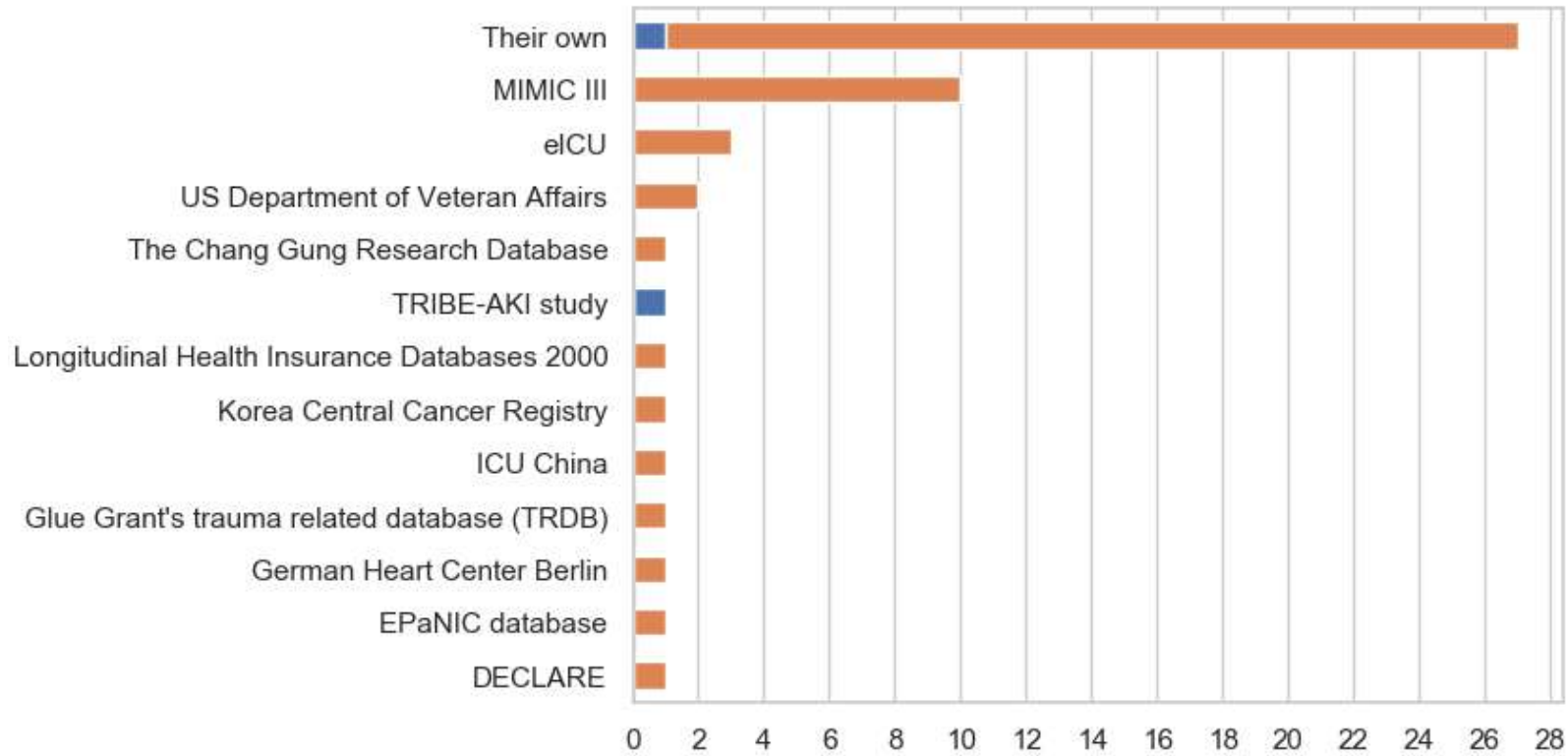
Good Fit/Robust



Overfitted

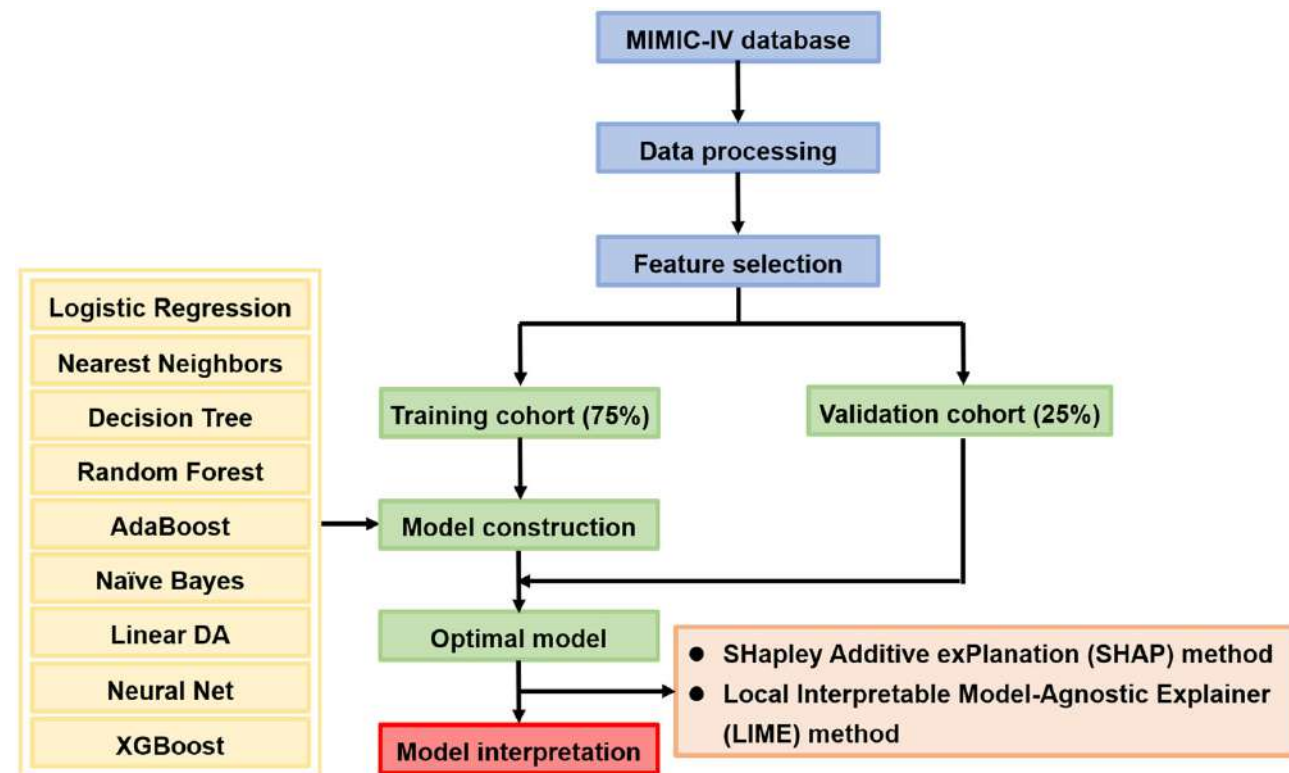
Données d'apprentissage

- **Réglementation** sur des données (RGPD, CNIL)
- **Taille de l'échantillon** suffisamment large
- **Représentativité** : population représentative de la population et du contexte de soins cibles et suffisamment hétérogène
- **Qualité des données** : données manquantes (mécanisme ?), méthodes de collecte, tests de validité sur échantillon, ...
« garbage in, garbage out »
- **Prétraitement** (preprocessing) : recodage de variable, retrait d'outliers, imputation de données, augmentation, séparation des données (train, test, validation)



Méthodes

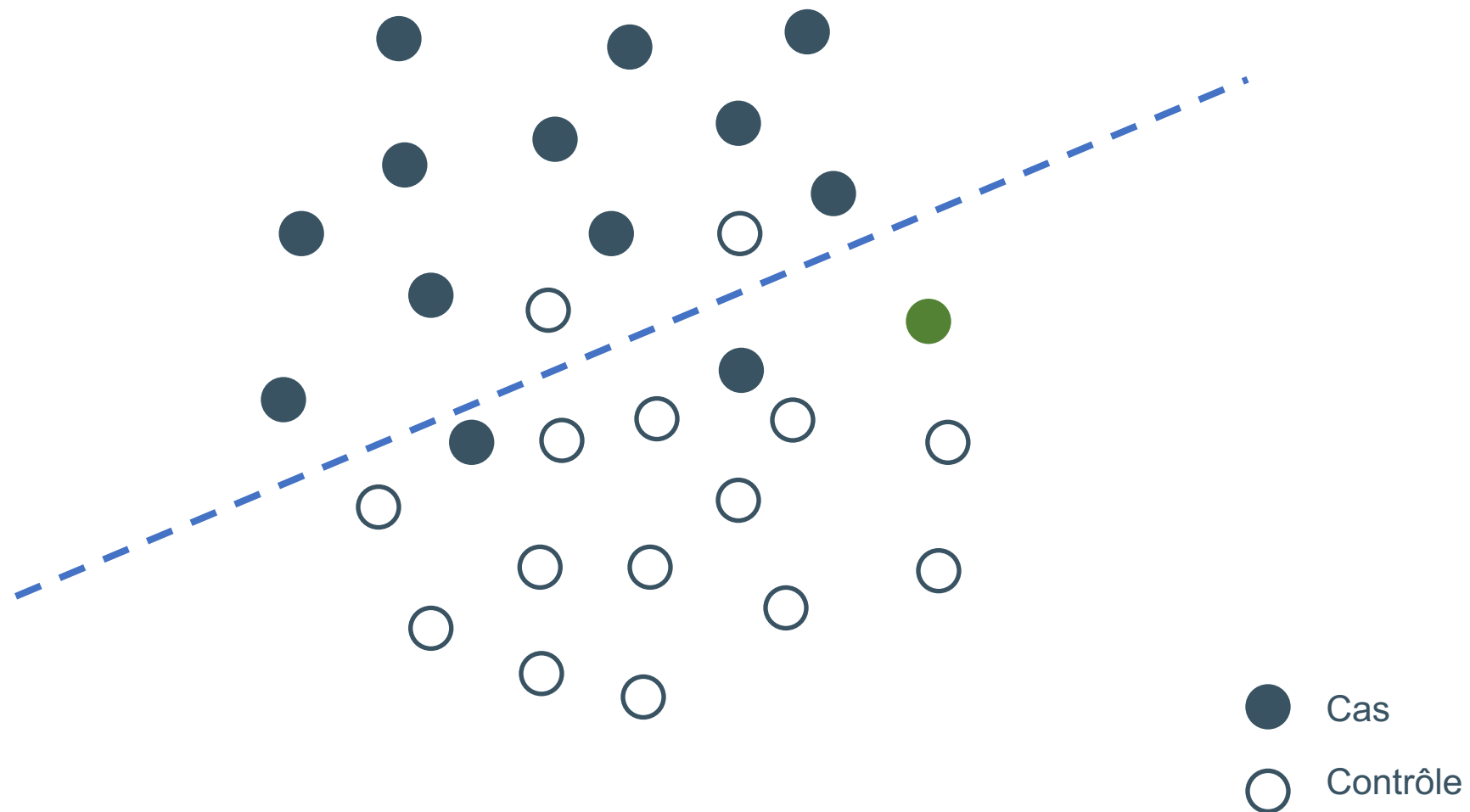
- **Sélection des variables** (feature selection)
- **Prétraitement** (preprocessing)
- **Développement des modèles**



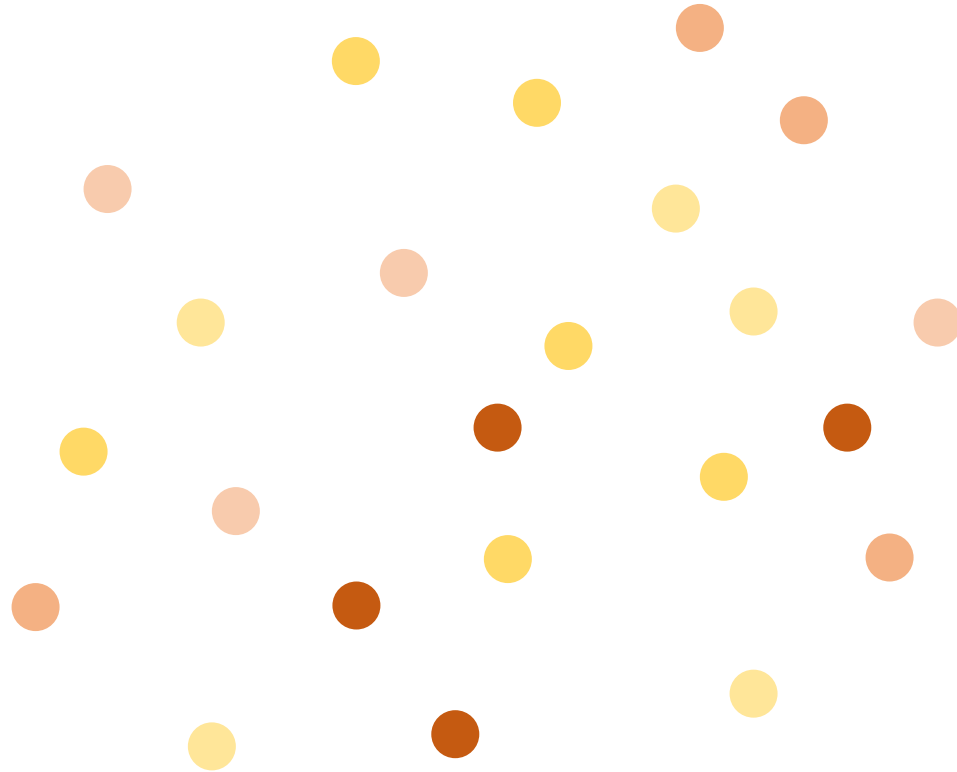
« No free lunch theorem »

- «There is no such thing as a free lunch »
- Tous les algorithmes d'optimisation fonctionnent également bien lorsque leurs performances sont moyennées sur tous les problèmes possibles
- → Il n'existe pas de meilleur algorithme d'apprentissage automatique unique

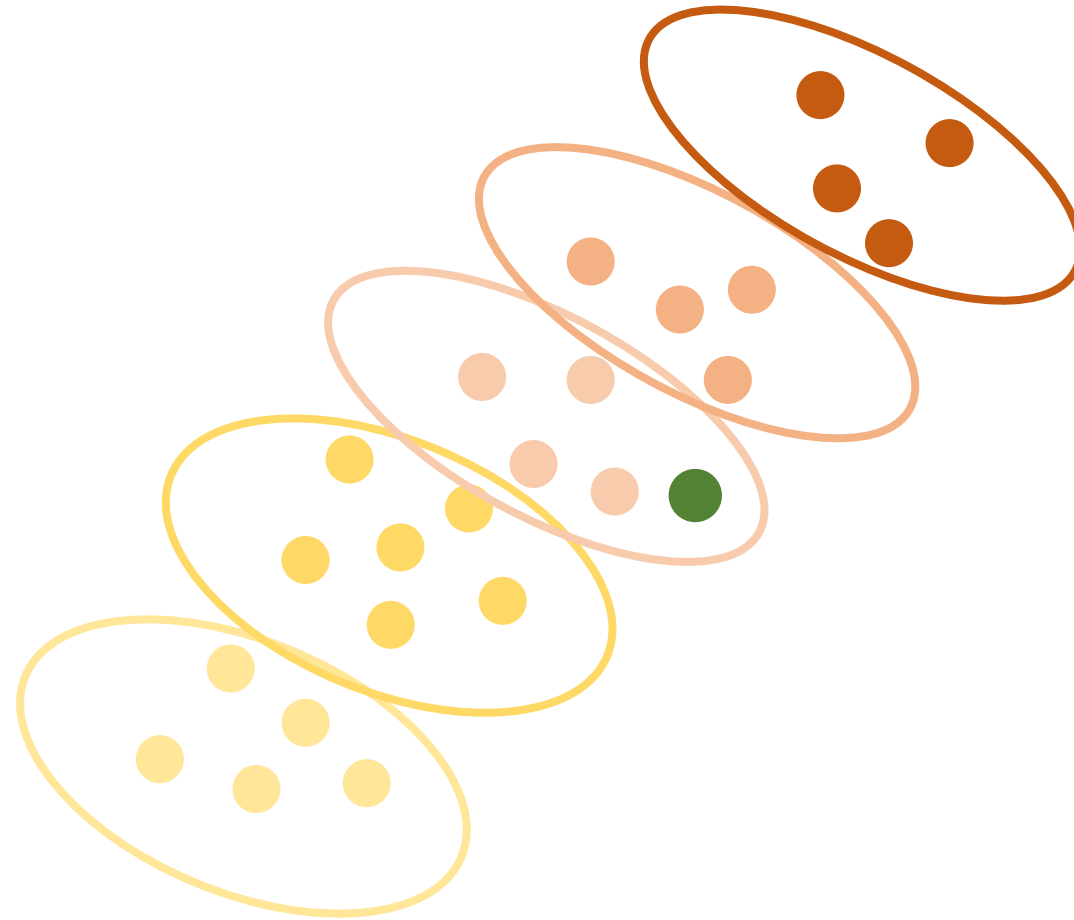
Apprentissage supervisé



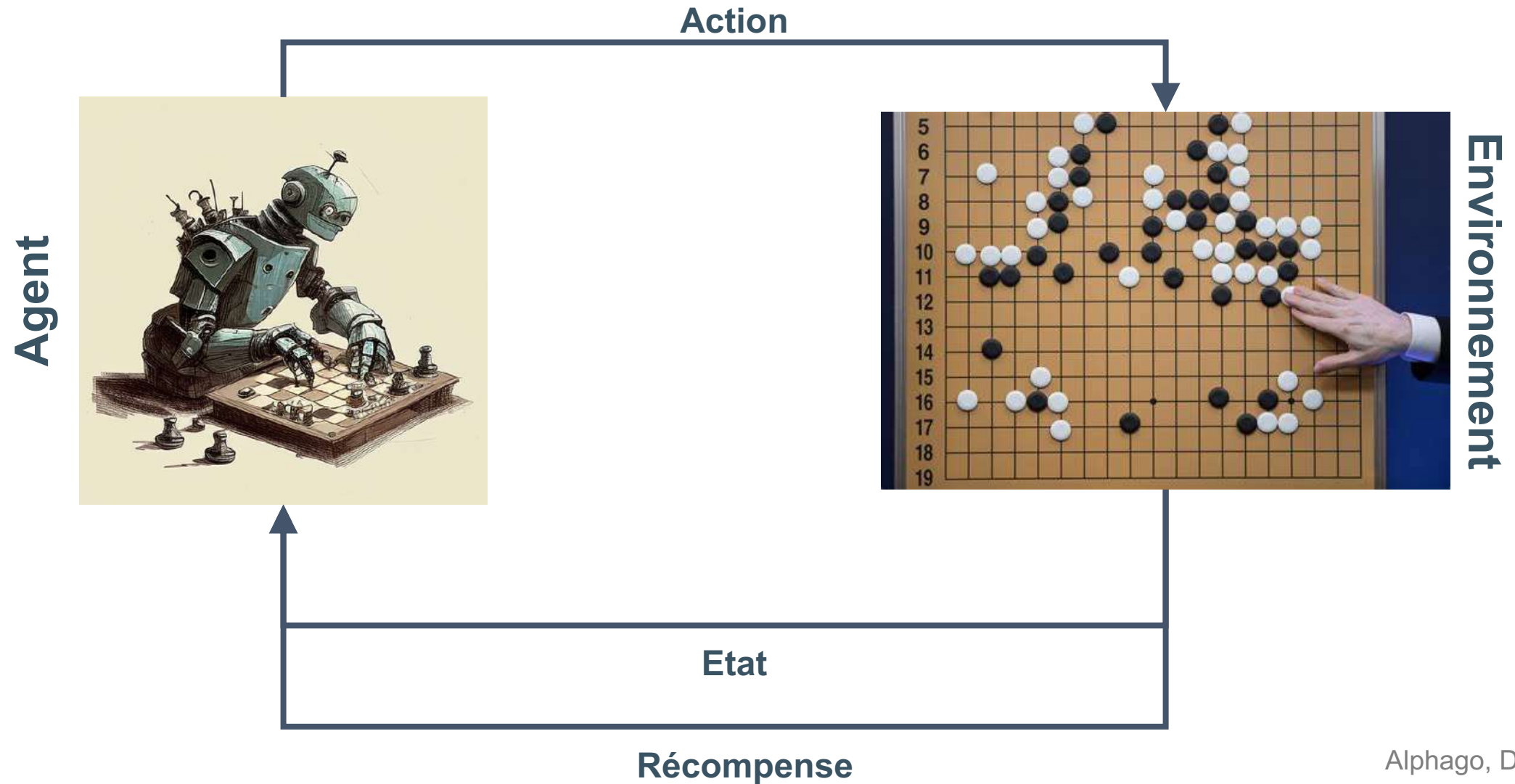
Apprentissage non supervisé



Apprentissage non supervisé

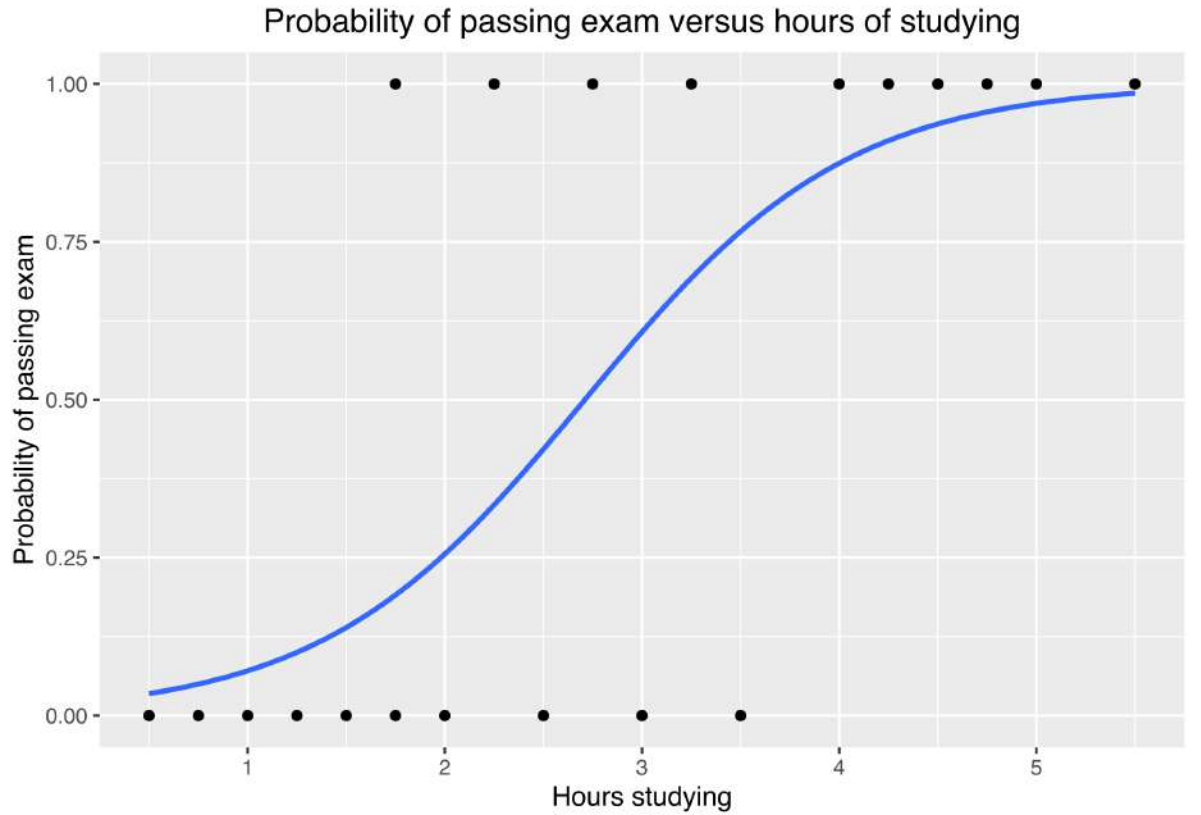
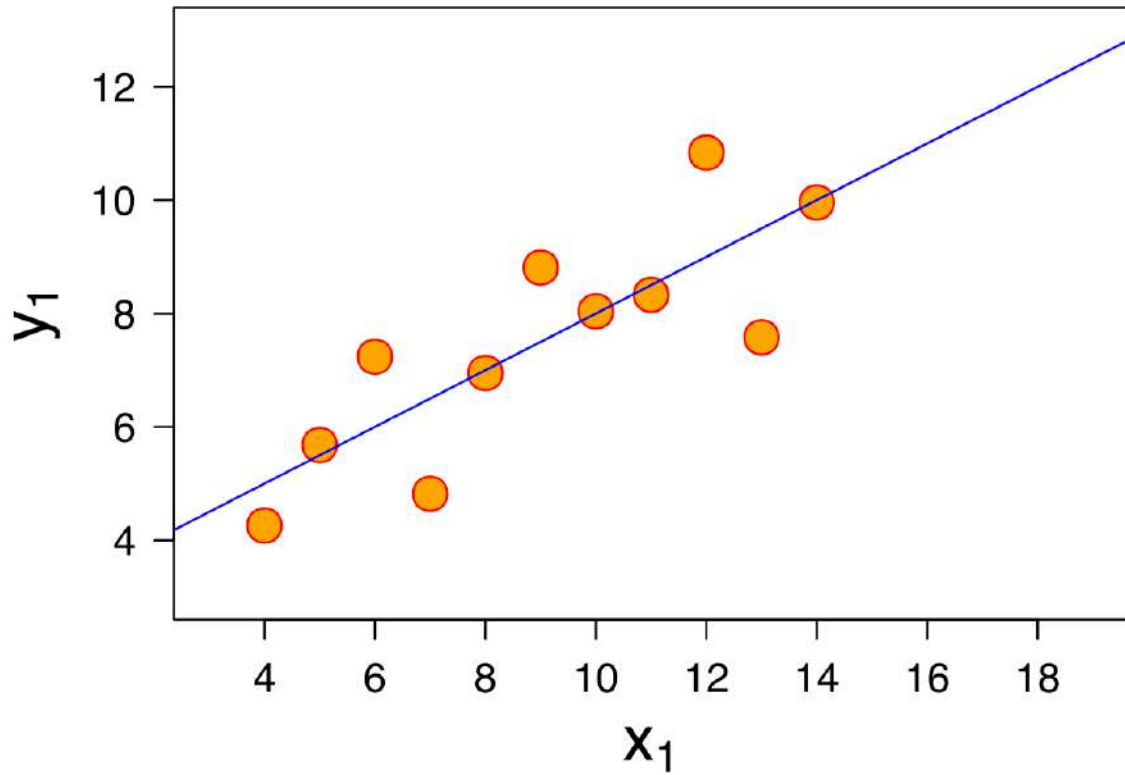


Apprentissage par renforcement



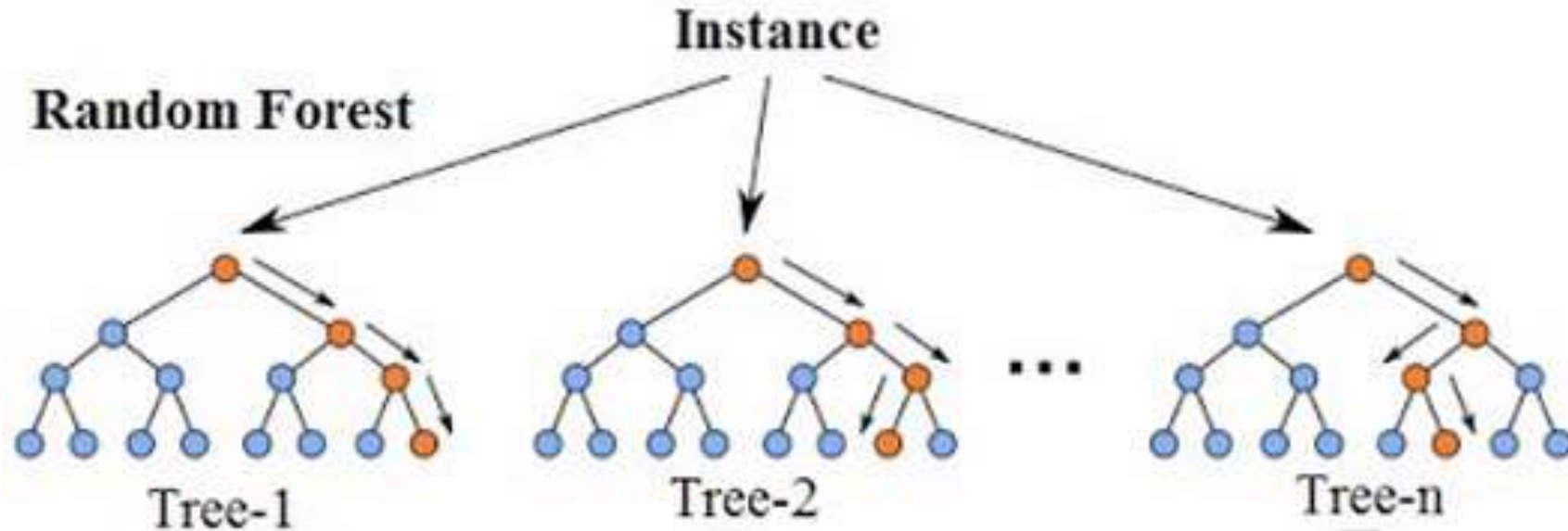
Principaux algorithmes

Régression



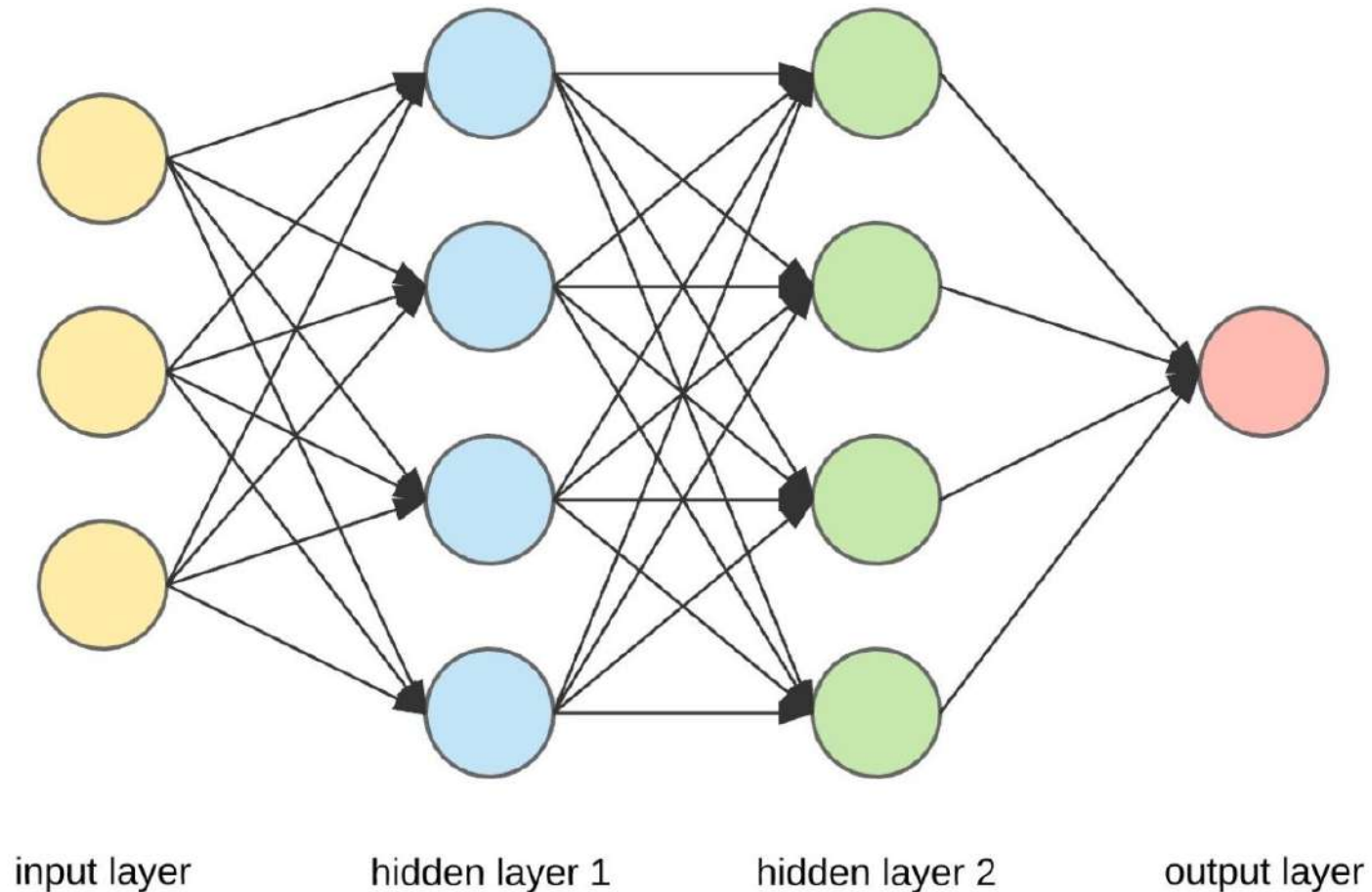
Principaux algorithmes

Arbres de décision



Principaux algorithmes

Réseaux de neurones



Méthodes

Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu · Zhiyong Peng

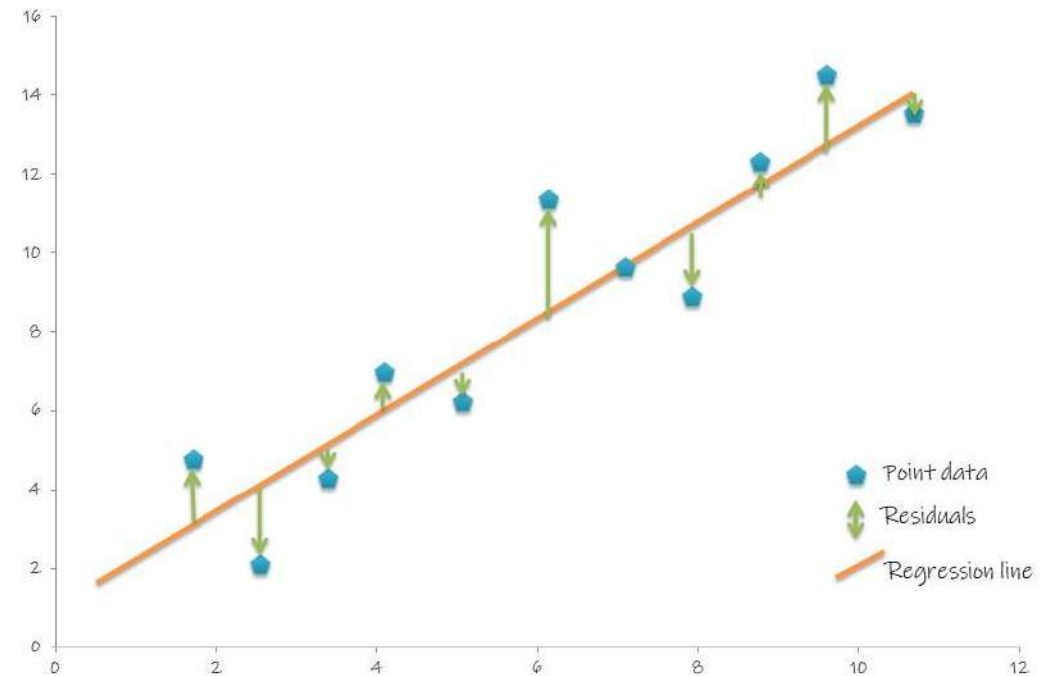
- **Evaluation des modèles**

- Critère principale : aire sous la courbe ROC (AUROC)
- Critères secondaires : précision, rappel, score F1

Méthodes d'évaluation

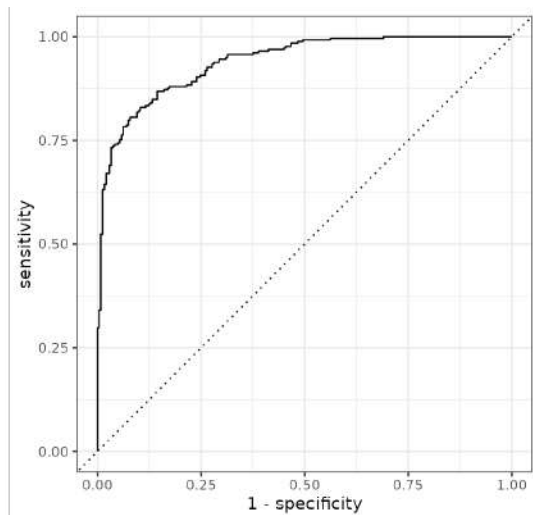
- **Régression**

- Racine de l'erreur quadratique moyenne (RMSE)
- Ecart absolu moyen (MAE)
- ...



Méthodes d'évaluation : classification

		Réalité	
		Positif	Négatif
Prédiction	Positif	Vrais positifs (VP)	Faux positifs (FP)
	Négatif	Faux négatifs (FN)	Vrai négatifs (VN)



- Accuracy

$$accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

- Précision (VPP)

$$precision = \frac{VP}{VP + FP}$$

- Rappel (sensibilité)

$$rappel = \frac{VP}{VP + FN}$$

- Score F1

$$F_1 = 2 \cdot \frac{precision \cdot rappel}{precision + rappel}$$

- AUROC

Résultats

- 1117 patients inclus (837 entraînement, 280 test)
- Comprendre la population d'apprentissage

Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu · Zhiyong Peng

Table 1 Baseline characteristics of the study cohort

Characteristics	Survivors (<i>N</i> = 896)	Non-survivors (<i>N</i> = 221)	<i>P</i> value
<i>Demographic</i>			
Age, year	67.9 (57.7–78.2)	71.9 (62.4–83.1)	< 0.001
Sex			0.118
- Male, <i>n</i> (%)	558 (62.3)	125 (56.6)	
- Female, <i>n</i> (%)	338 (37.7)	96 (43.4)	
Weight, kg	80 (68–94)	78 (67–93)	0.433
Height, cm	170 (163–178)	169 (163–175)	0.232
<i>Comorbidities</i>			
Charlson Comorbidity Index	6 (4–8)	7 (6–9)	< 0.001
Hypertension, <i>n</i> (%)	355 (39.6)	79 (35.7)	0.290
Diabetes, <i>n</i> (%)	304 (33.9)	67 (30.3)	0.307
Congestive heart failure, <i>n</i> (%)	342 (38.2)	88 (39.8)	0.652
Cerebrovascular disease, <i>n</i> (%)	146 (16.3)	47 (21.3)	0.080
Chronic pulmonary disease, <i>n</i> (%)	264(29.5)	56 (25.3)	0.224
Liver disease, <i>n</i> (%)	171 (19.1)	68 (30.8)	< 0.001
Renal disease, <i>n</i> (%)	215 (24.0)	77 (34.8)	0.001
Tumor, <i>n</i> (%)	153 (17.1)	62 (28.1)	< 0.001
<i>Vital signs on day 1</i>			
Heart rate, bpm	88 (79–99)	93 (83–103)	0.001
Systolic blood pressure, mmHg	114 (105–124)	109 (100–120)	< 0.001
Diastolic blood pressure, mmHg	60 (54–67)	58 (51–64)	0.001
Mean arterial pressure, mmHg	75 (70–82)	71 (67–78)	< 0.001
Respiratory rate	19 (17–23)	20 (18–24)	0.003
Body temperature, °C	36.9 (36.7–37.2)	36.7 (36.4–37.1)	< 0.001
SpO ₂ , %	93 (90–95)	91 (87–94)	< 0.001
<i>Laboratory findings on day 1</i>			
White blood cell, × 10 ³ /μL	14.3 (10.3–20.1)	15.0 (10.6–20.7)	0.519
Platelets, × 10 ³ /μL	193 (118–275)	142 (68–222)	< 0.001
Hematocrit, %	26.9 (23.8–30.5)	25.6 (22.7–30.1)	0.033
Hemoglobin, g/dL	8.8 (7.8–10.1)	8.2 (7.3–9.9)	0.001
International normalized ratio	1.4 (1.2–1.7)	1.6 (1.4–2.4)	< 0.001
Prothrombin time, s	15.5 (13.5–18.6)	18.1 (14.8–25.3)	< 0.001
Partial thromboplastin time, s	35 (29–54)	44 (32–75)	< 0.001

Résultats

Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu · Zhiyong Peng

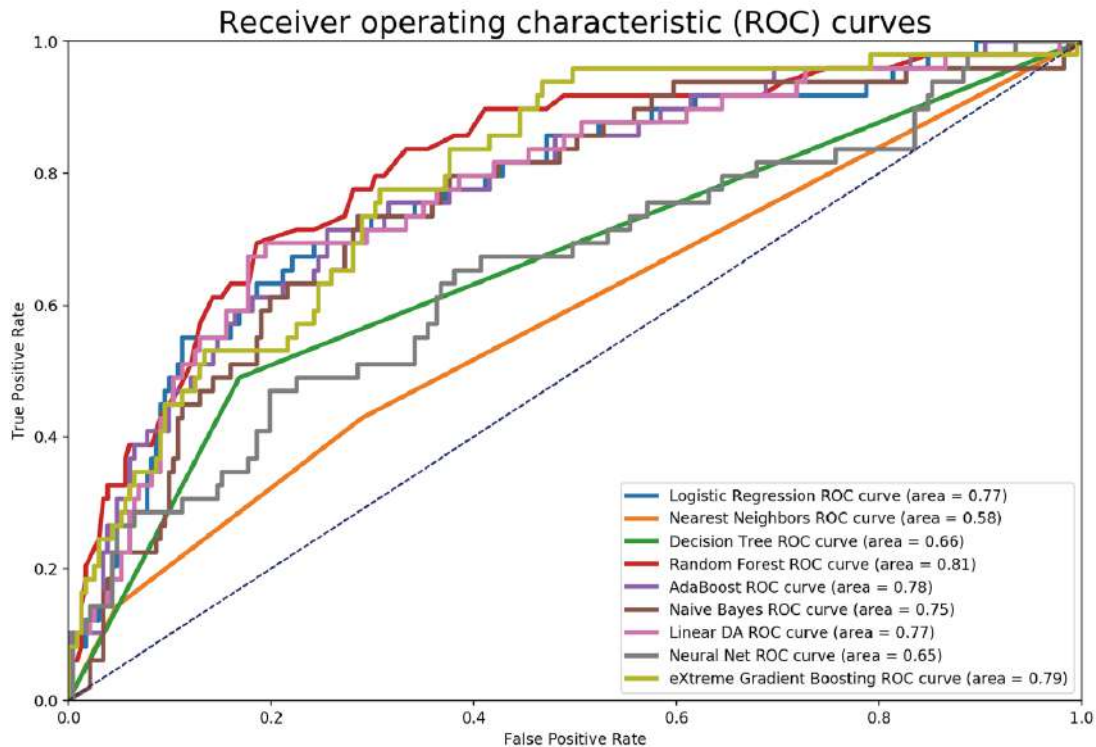


Table 2 Performance of the nine machine-learning classifiers for predicting in-hospital mortality in the validation set

Classifiers	A.C.	Accuracy	Precision	Recall	F1
Random Forest	0.81	84.64%	62%	0.31	0.41
eXtreme Gradient Boosting	0.79	81.79%	47%	0.35	0.40
AdaBoost	0.78	81.79%	48%	0.41	0.44
Logistic Regression	0.77	81.07%	44%	0.29	0.35
Linear DA	0.77	82.14%	49%	0.41	0.44
Naive Bayes	0.75	80.36%	44%	0.45	0.44
Decision Tree	0.66	77.14%	38%	0.49	0.43
Neural Net	0.65	73.93%	30%	0.37	0.33
Nearest Neighbors	0.58	81.07%	39%	0.14	0.21

A.C. Area under curve, *AdaBoost* Adaptive Boosting

Résultats

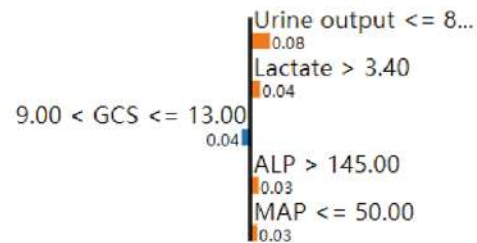
- Effet « boîte noire »
- Explicabilité
 - Valeurs de Shapley
 - A quelle point une variable participe à la prédiction ?
 - De façon globale sur l'ensemble du modèle
 - Pour un patient en particulier

Prediction probabilities



Survive

Death

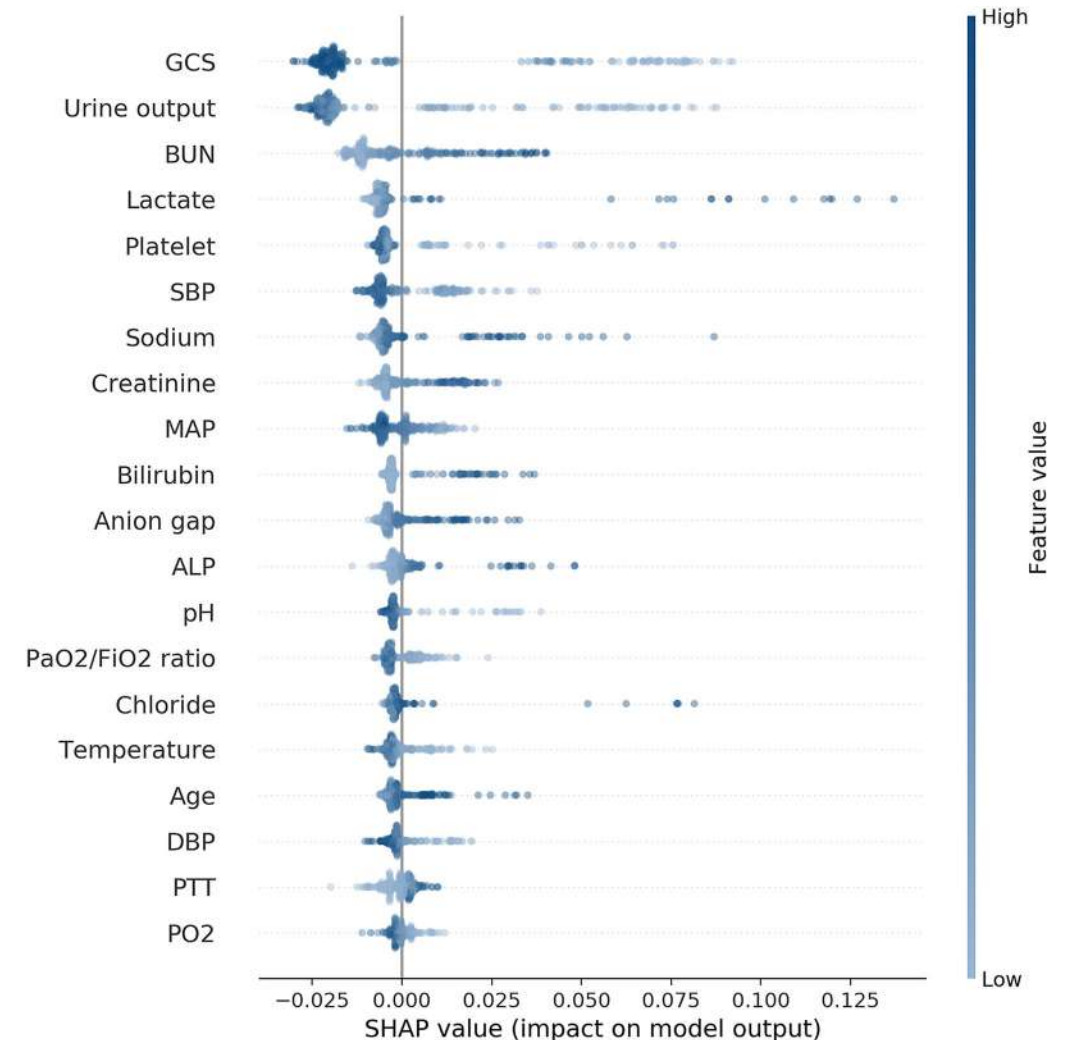


Feature Value

Urine output	222.00
Lactate	5.60
GCS	13.00
ALP	196.00
MAP	41.00

Explainable Machine-Learning Model for Prediction of In-Hospital Mortality in Septic Patients Requiring Intensive Care Unit Readmission

Chang Hu · Lu Li · Yiming Li · Fengyun Wang · Bo Hu · Zhiyong Peng



Et ensuite ?

- Pertinence informatique/pertinence clinique
- Validation externe : comment le modèle va se comporter sur une autre population ?
- Comment le modèle va s'intégrer dans une pratique clinique ?

Cycle de développement

Pré-développement

- Récupération des données
- Stockage
- Qualité
- Données manquantes
- Question clinique

Développement

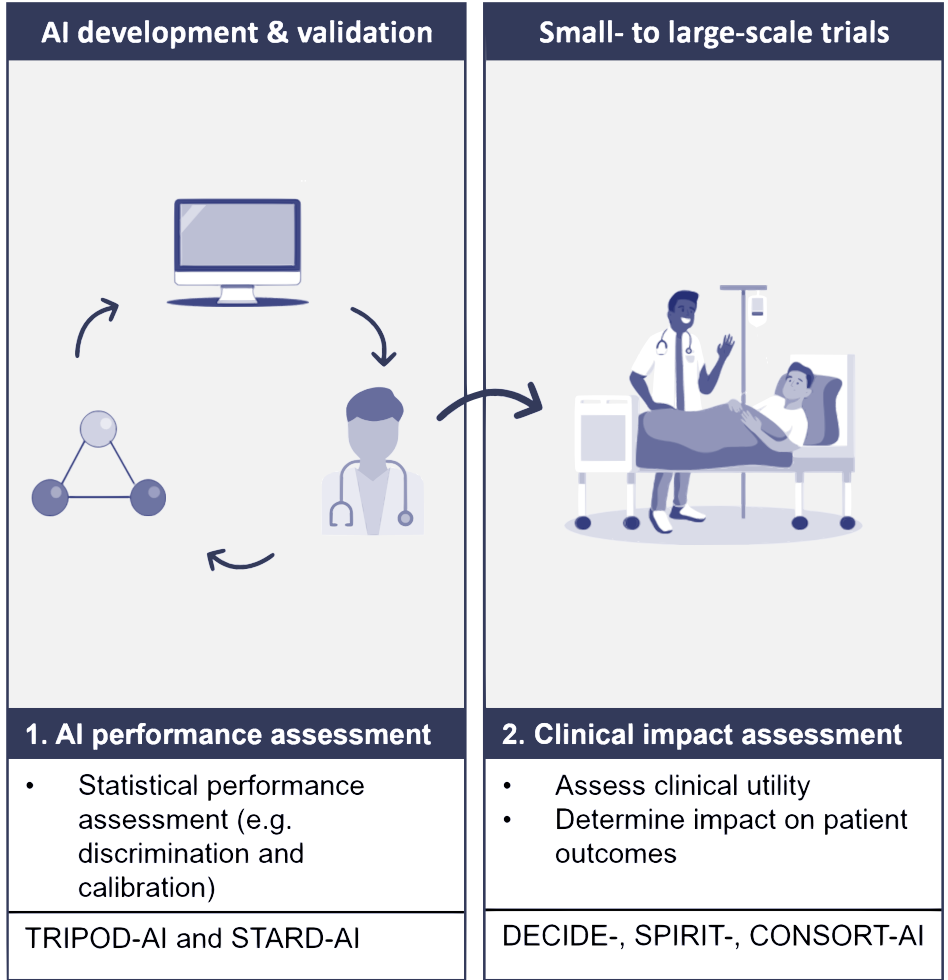
- Choix du modèle
- Entraînement
- Evaluation rétrospective

Post-développement

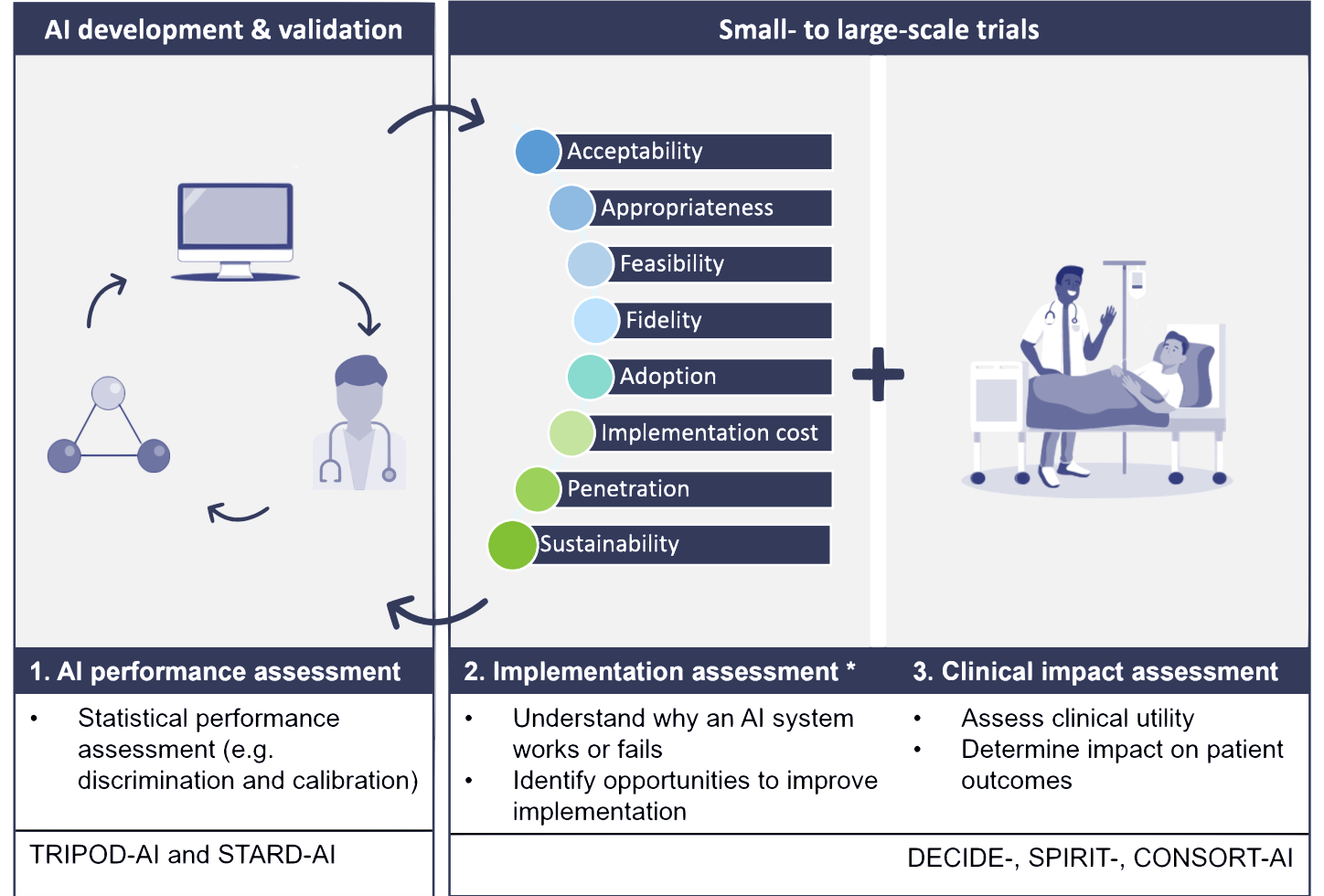
- Industrialisation (échelle)
- Intégration, interface
- Déploiement
- Evaluation prospective
- Monitoring
- Autorisation

- Acceptabilité
- Interprétabilité
- Facteurs humains

a. current situation



b. integration of implementation outcomes



van de Sande D, Chung EFF, Oosterhoff J, et al (2024) To warrant clinical adoption AI models require a multi-faceted implementation evaluation. npj Digit Med 7:1–5.