

Progrès récents en Intelligence Artificielle et applications

Emergence – 12/04/2021

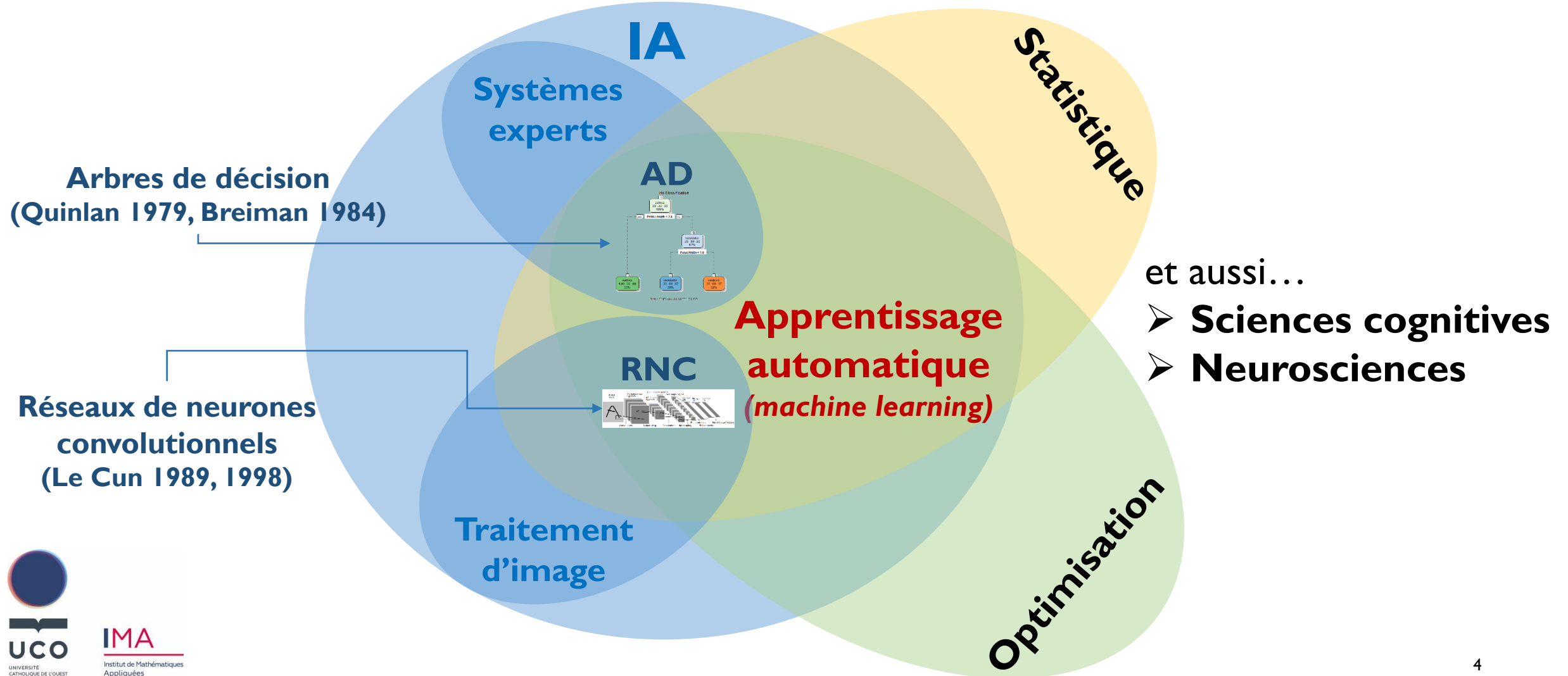
Pierre CHAUVET - Institut de Mathématiques Appliquées - Faculté des Sciences de l'UCO
<https://ima.uco.fr/>

Plan

- ❑ Apprentissage automatique
 - ❑ Réseaux neuronaux, réseaux convolutionnels
 - ❑ Arbres de décision
- ❑ Modèles ensemblistes
 - ❑ Plusieurs modèles plutôt qu'un seul : décision collective
 - ❑ Forêts aléatoires, *Gradient boosting*
- ❑ Interprétabilité & Explicabilité
 - ❑ Pourquoi blanchir la boîte ?
 - ❑ Quelques approches (par construction, a posteriori - SHAP, LIME)
- ❑ Données multicentriques : Effet de lots
 - ❑ Origine du problème
 - ❑ Harmonisation des données

Apprentissage Automatique

Apprentissage automatique et IA



Arbres de décision

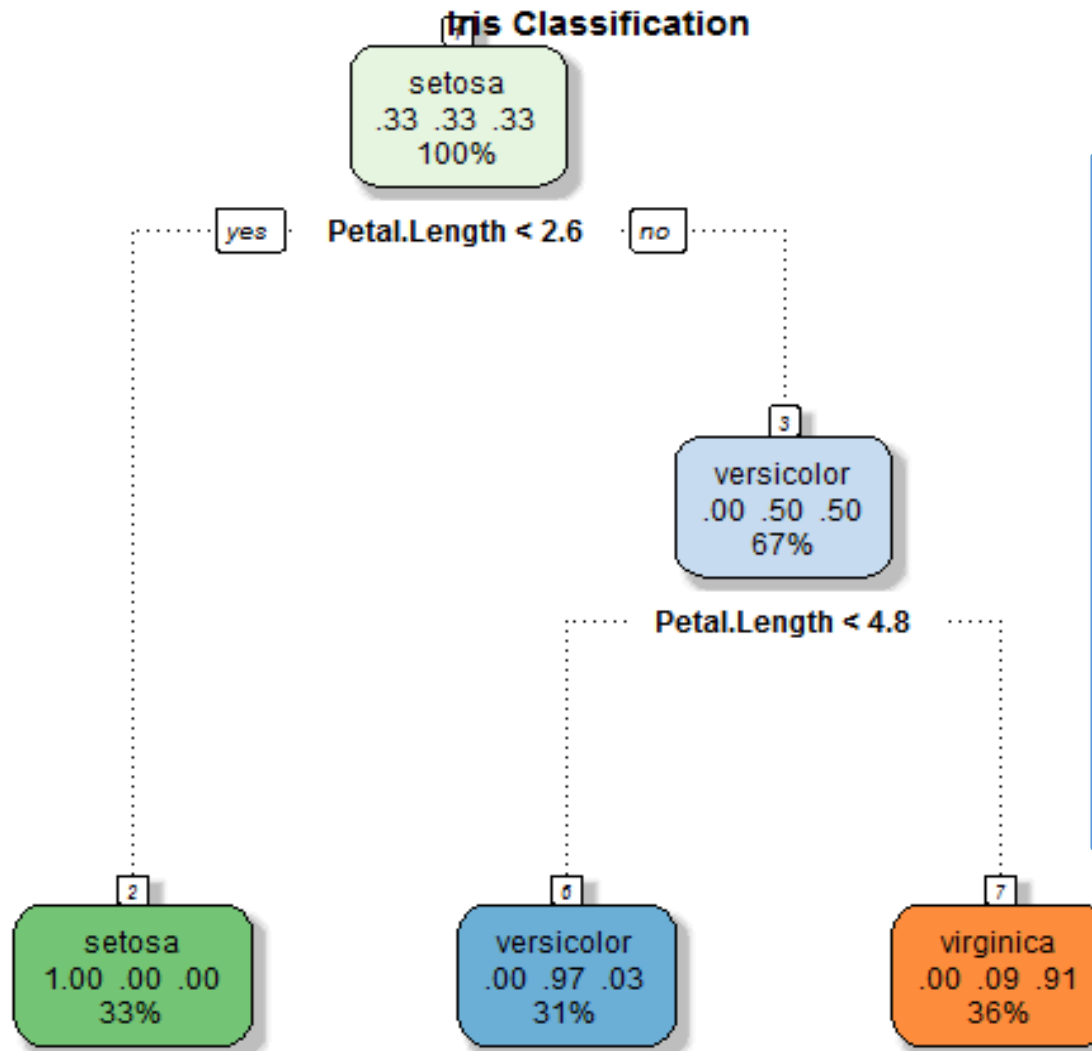
- ☐ Adapté aux données quantitatives et qualitatives (ordinales, nominales)
- ☐ Lisible pour le non spécialiste, décision explicable
- ☐ Utile à la fois pour fouiller et classer
- ☐ Mais appartient à la classe des algorithmes supervisés : nécessite des exemples labélisés/expertisés
- ☐ Extension version ensembliste : *random forest, adaboost*.

Principe de base

- ❑ Un arbre de décision n'est qu'un ensemble de « **SI** condition **ALORS...** » emboîtés, appliqué à la classification (et aussi à la régression)
- ❑ La difficulté est de le générer automatiquement à partir des données => Algorithmes de « Machine learning »
- ❑ Exemple sur la table des iris de Fisher :

```
Si (Petal.Length<2.6) Alors
    Si (Petal.Length<4.8) Alors
        classer « versicolor »
    Sinon
        classer « virginica »
Sinon
    classer « setosa »
```

Exemple - Iris de Fisher



**Arbre obtenu sous R avec
rpart() et les paramètres par
défaut.**

Traduction :

Si (Petal.Length<2.6) **Alors**
 Si (Petal.Length<4.8) **Alors**
 classer « versicolor »
 Sinon
 classer « virginica »
Sinon
 classer « setosa »

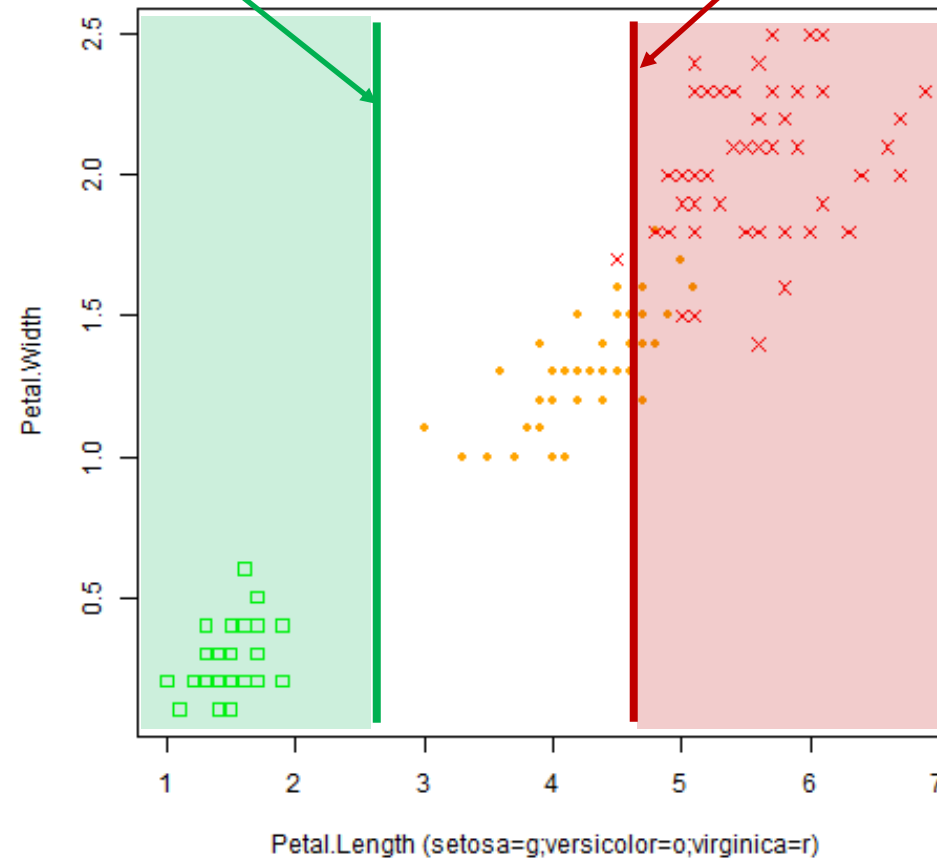
Exemple - Interprétation

Si (Petal.Length < 2.6) **Alors**
 Si (Petal.Length < 4.8) **Alors**
 classer « versicolor »
 Sinon
 classer « virginica »
Sinon
 classer « setosa »

Petal.Length=2.6

Petal.Length=4.8

IRIS PLOT (Petal Width vs Length)



Principales méthodes

☐ **CHAID** (CHi-squared Automatic Interaction Detection)

KASS G., *An exploratory technique for investigating large quantities of categorical data*, Applied Statistics, 29(2), 119-127, 1980.

☐ **CART** (*Classification And Regression Tree*)

BREIMAN L, FRIEDMAN J.H, OLSHEN R.A, STONE C., *Classification and Regression Trees*, California:Wadsworth International, 1984.

☐ **ID3** (*Induction of Decision Tree*)

QUINLAN R., *Discovering rules by induction from large collections of examples*, D. Michie ed., Expert Systems in the Microelectronic age, pp. 168-201, 1979.

QUINLAN R., *Induction of decision trees*. *Machine Learning*, 1, p81-106, 1986.

☐ **C4.5**

QUINLAN R., *C4.5: Programs for Machine Learning*, Morgan Kaufman, 1993.

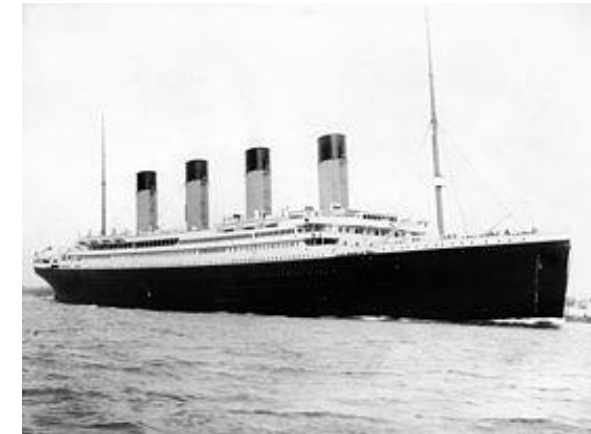
☐ **SIPINA**

ZIGHED D., AURAY J.P, DURU G., *SIPINA : méthode et logiciel*, Lacassagne, 1992.

ZIGHED D., RAKOTOMALALA R., *Graphes d'Induction : Apprentissage et Data Mining*, Hermès, 2000.

<http://sipina-arbres-de-decision.blogspot.fr/>

100



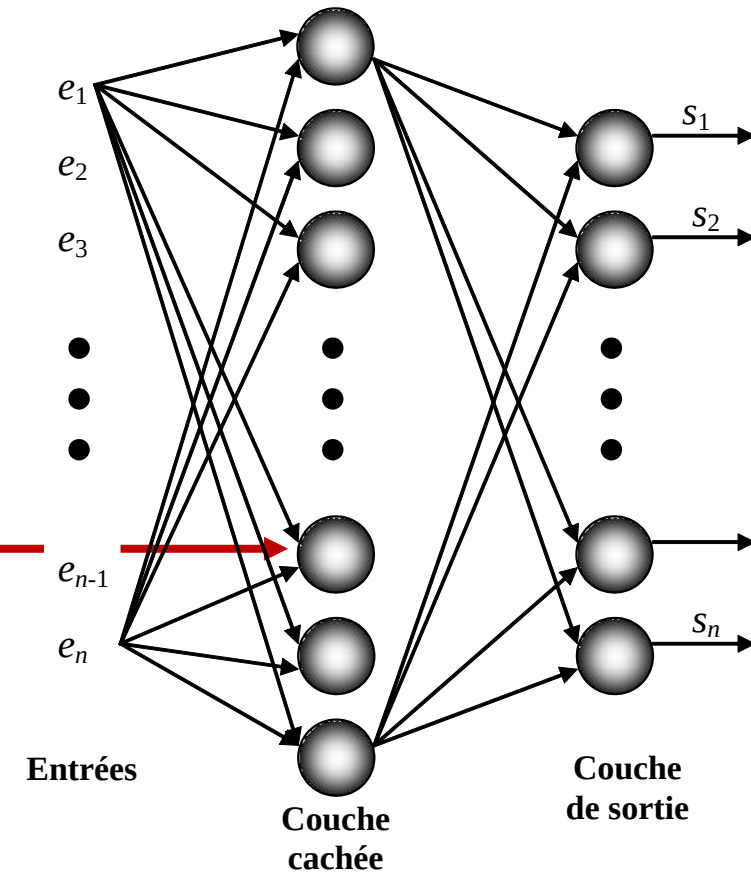
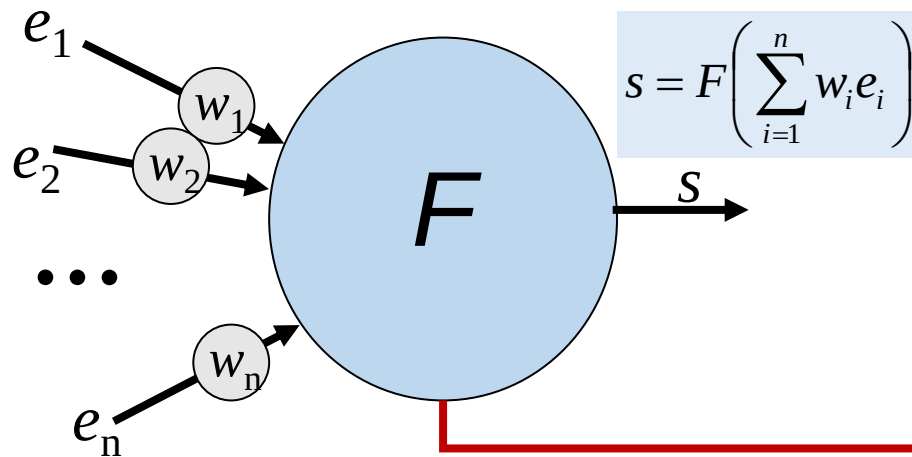
- 1 324 passagers
- 889 membres d'équipage
- Environ 1 500 morts

P_{class} = Classe (1, 2 ou 3)

Réseaux neuronaux

- ❑ Adapté aux données numériques, à l'image et au signal
- ❑ Non explicable de manière directe en général
- ❑ Versions supervisées, non supervisées, par renforcement
- ❑ Classe de modèles plutôt que modèle :
 - Perceptron multicouches
 - Réseau de neurones convolutionnels
 - Réseau de neurones récurrents
 - *Autoencoder-decoder*
 - Réseaux de Kohonen (cartes autoadaptatives)
 - etc.

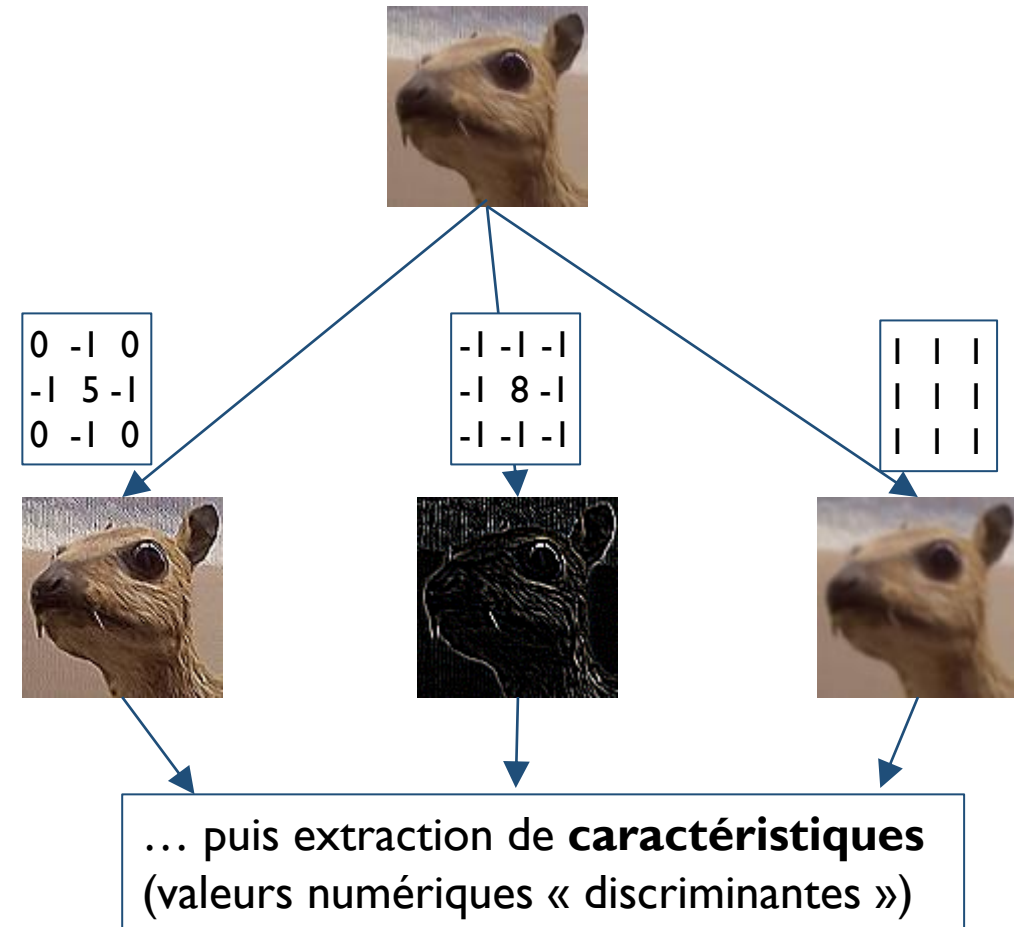
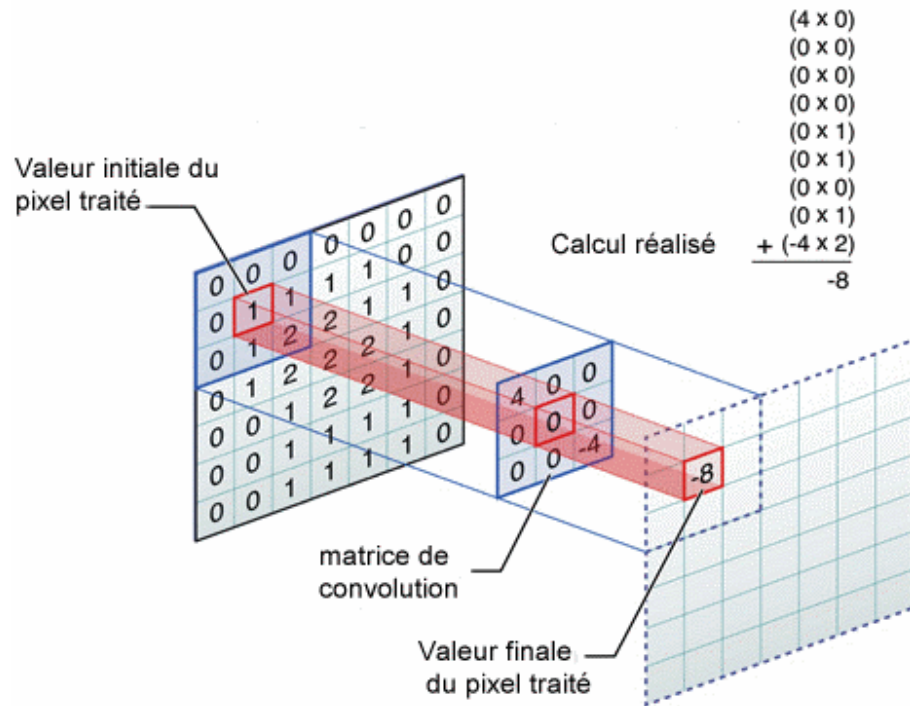
Histoire pas si récente...



- ❑ Neurone formel : Mc Culloch et Pitts 1943
- ❑ Règle de Hebb : Hebb 1949
- ❑ Perceptron : Rosenblatt 1958, Widrow et Hoff 1960
- ❑ Perceptron multicouche / algorithme de rétropropagation du gradient : Werbos 1975, ...
- ❑ Réseau de neurones convolutionnel : Le Cun et al 1998

Traitement d'image et filtres de convolution

Exemples filtres (wikipedia) →



Réseaux de neurones convolutionnels

LeNet-5 : réseau neuronal mis au point par Y. LeCun, L. Bottou, Y. Bengio et P. Haffner en 1997-1998 pour la reconnaissance de caractères manuscrits - <http://yann.lecun.com/exdb/lenet/>

PROC. OF THE IEEE, NOVEMBER 1998

7

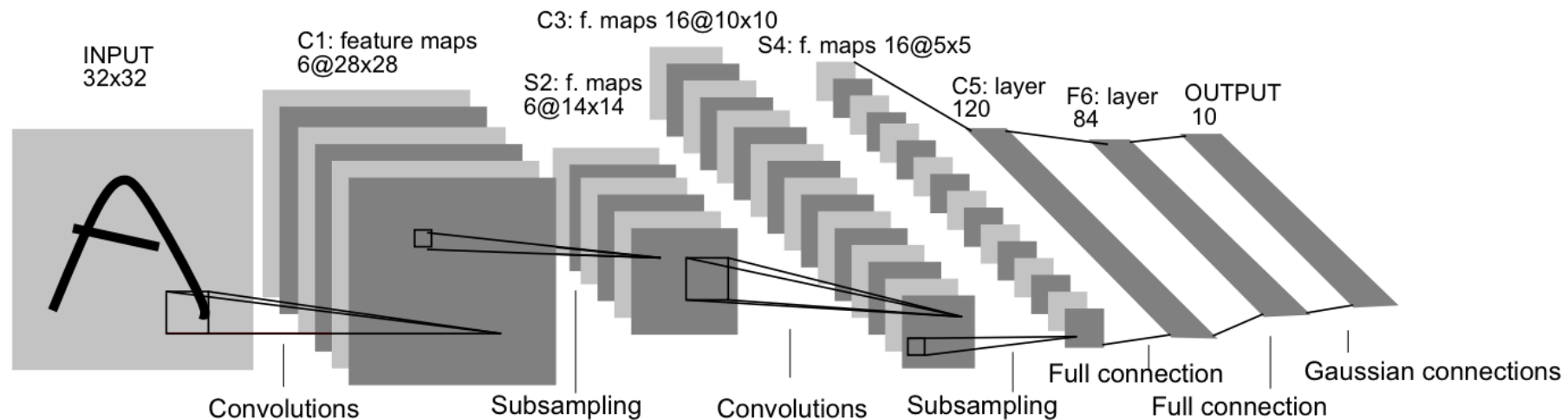


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Deep learning : fin du «feature engineering» ?

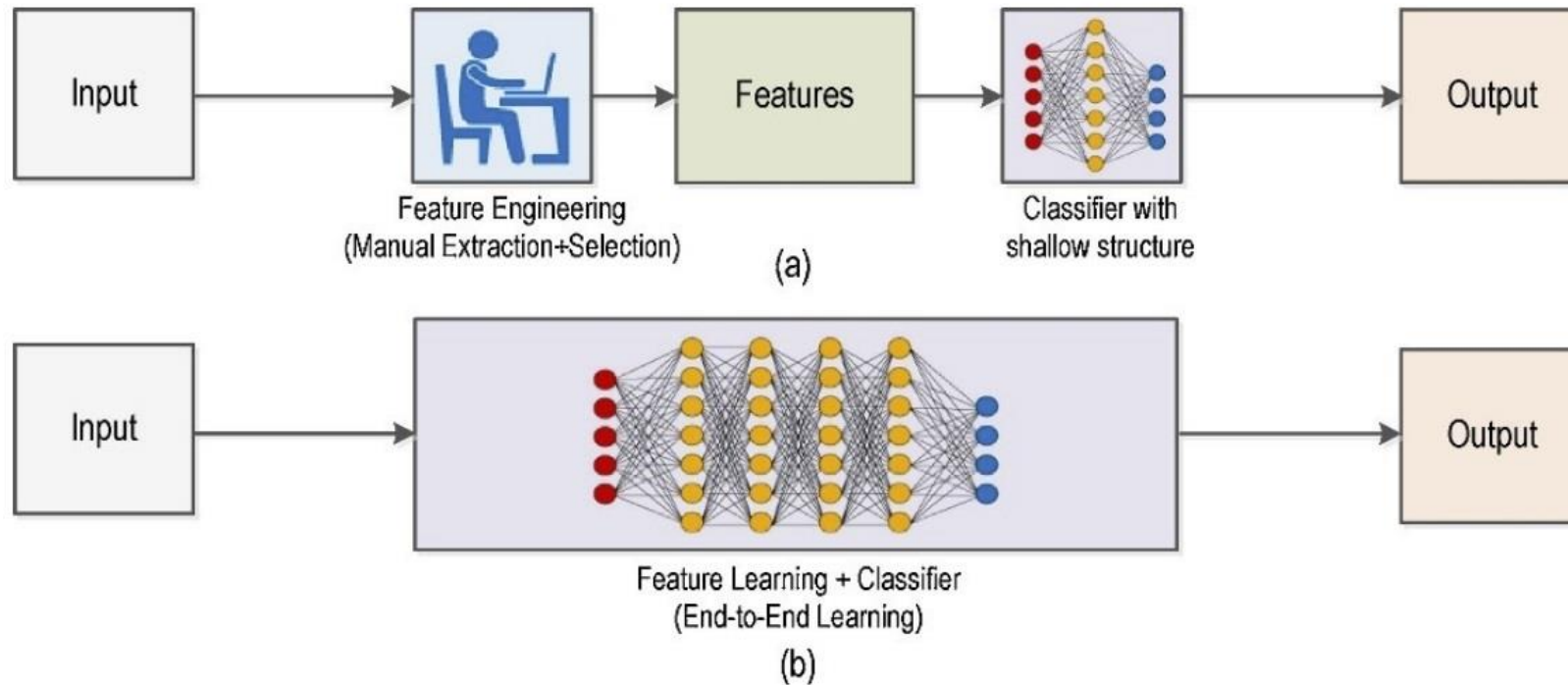


Fig. 1. (a) Traditional Computer Vision workflow vs. (b) Deep Learning workflow. Figure from [8].

Deep Learning vs. Traditional Computer Vision

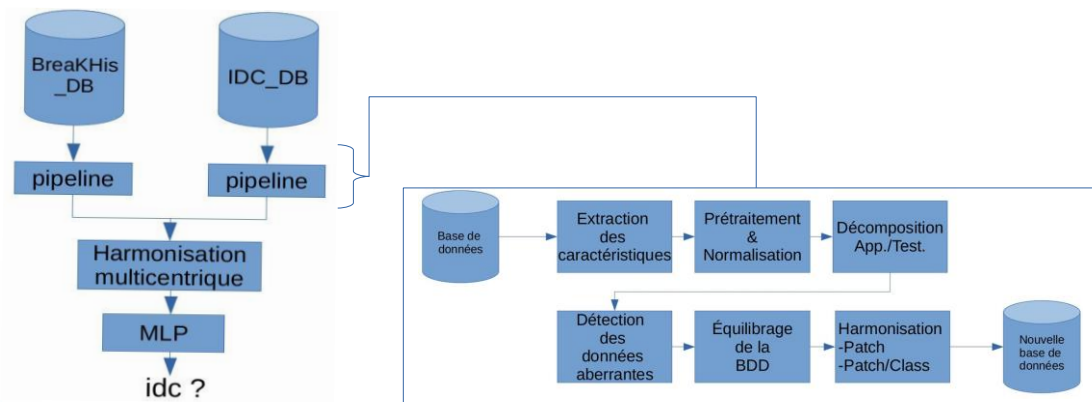
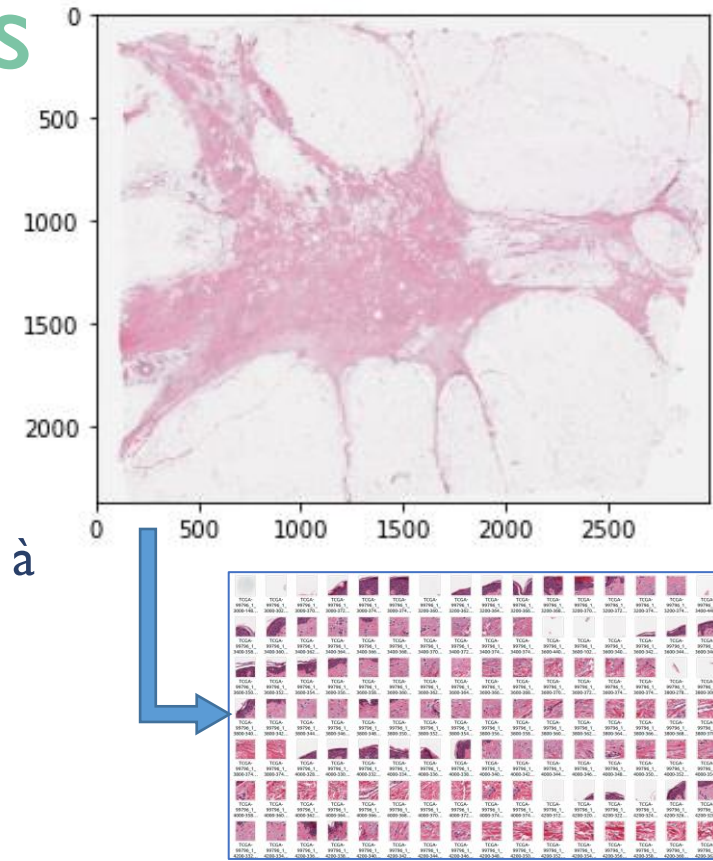
<https://arxiv.org/ftp/arxiv/papers/1910/1910.13796.pdf>

Classification coupes histologiques

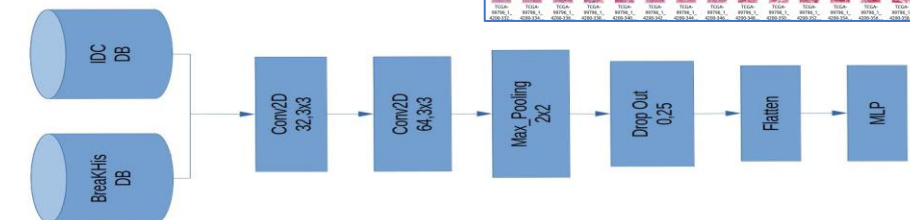
Projet HARMONY (*HARMONization methods for optimized therapY*) – Appel d'offres structurant « Numérique en Oncologie » Cancéropôle Grand Ouest / Régions 2019 – travaux Nassib Abdallah, Jean-Marie Marion.

Classification coupes histologiques de patientes atteintes ou non d'un carcinome canalaire invasif (cancer du sein le plus courant) :

- ❑ Approche de « traiteur d'image » : prétraitements, extraction de caractéristiques maîtrisées, posttraitements, classification par perceptron à une couche cachée.
- ❑ Approche *Machine learning* : entraînement d'un CNN standard.



versus

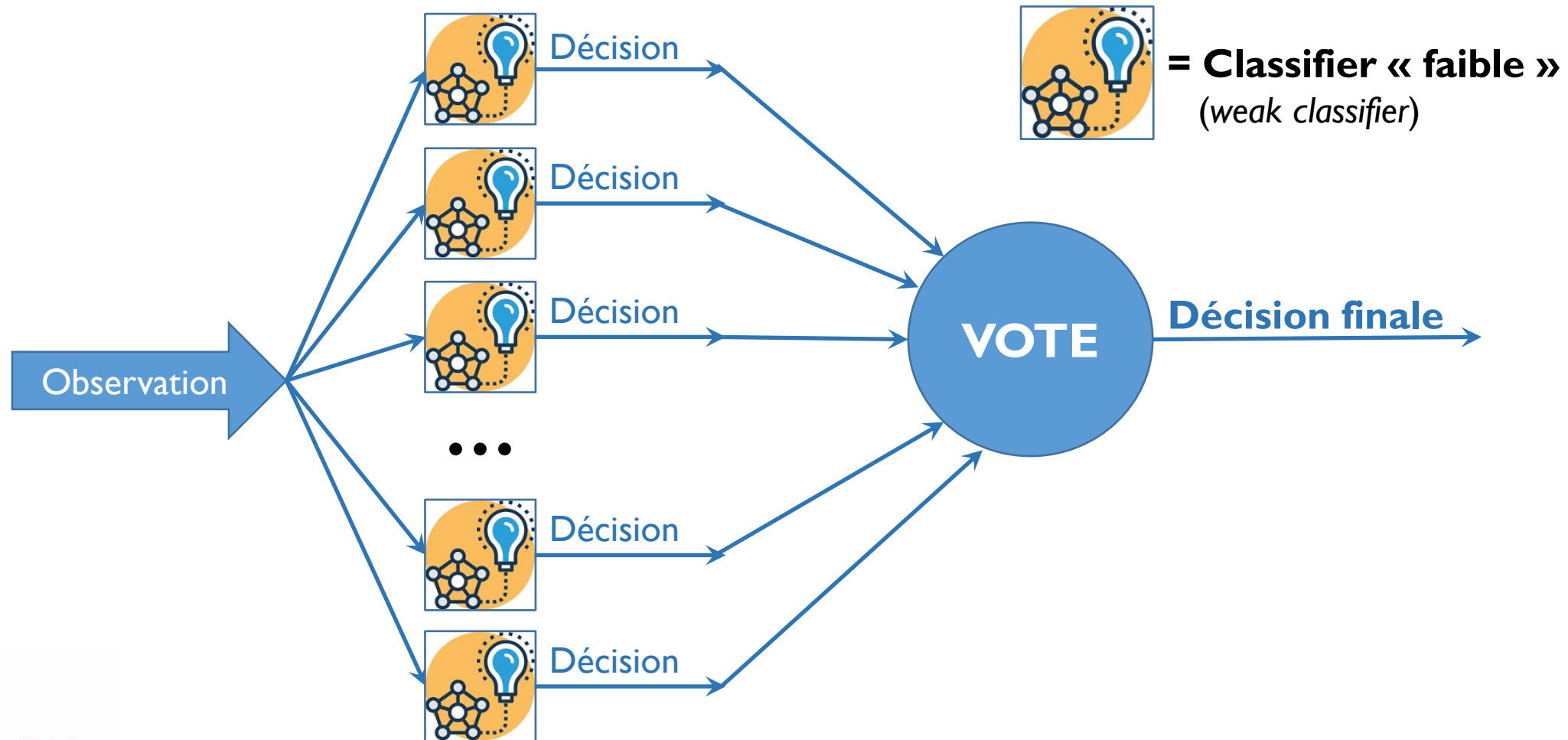


Autres projets (équipe IMA/ISISV LARIS)

- ❑ Détection de HFO dans les EEG de scalp d'enfants épileptiques
Gaëlle Milon-Harnois (thèse), Nisrine Jrad
- ❑ Construction de signatures pronostiques du cancer de l'ovaire séreux de haut grade par apprentissage automatique – Projet SignOvAA
Elena Menand (thèse), Jean-Marie Marion, Nisrine Jrad

Modèles ensemblistes

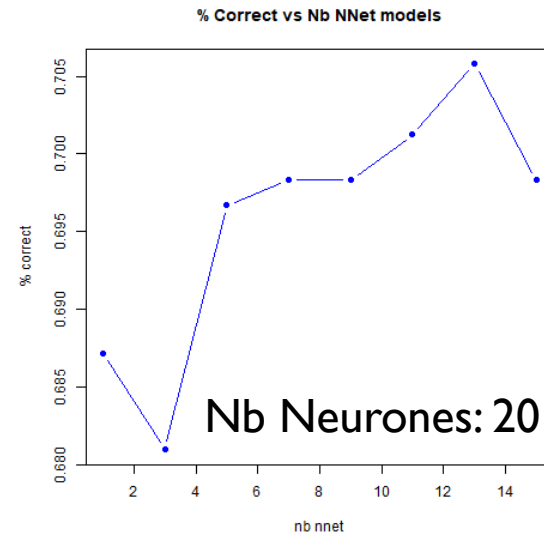
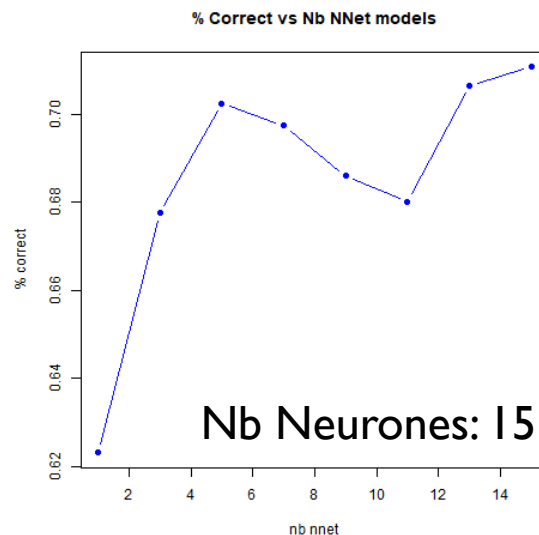
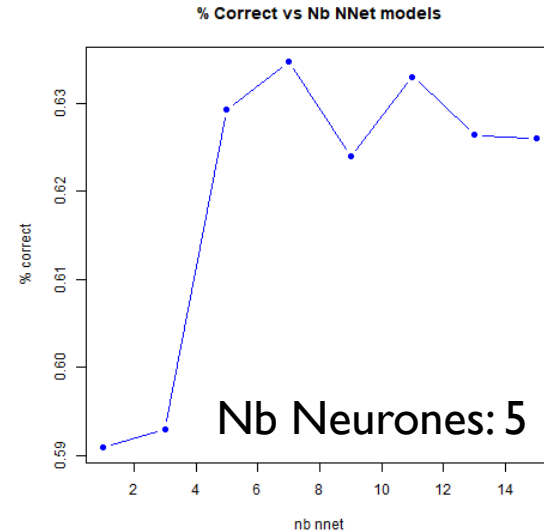
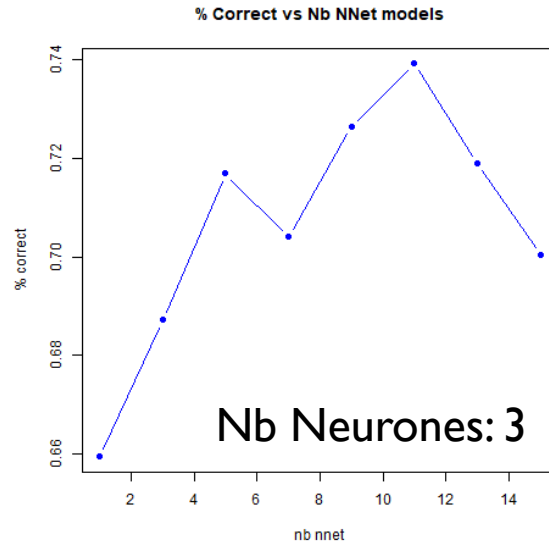
Une décision collective...



Exemple : prévision d'un risque cardiaque

- ❑ Objectif : prévoir la présence d'une maladie cardiaque à partir d'un certain nombre de variables explicatives (âge, cholestérol, pression sanguine, ...).
- ❑ Données : tableau avec 14 colonnes et 304 lignes (1 ligne pour l'entête, 303 exemples). Les 14 colonnes correspondent aux 13 variables explicatives et à la colonne « *disease* » (variable à expliquer).
- ❑ Modèle : N réseaux neuronaux à 1 couche cachée + méthode de vote (N impair $\rightarrow$ majorité absolue)

Exemple : prévision d'un risque cardiaque



Le meilleur modèle :

- 11 réseaux neuronaux 13/3/1 (74%)
- Nb poids réseau: $(13*3+3)+(3*1+1)=46$
- Nb poids modèle: $46*11 = 506$

Le plus mauvais modèle :

- 1 réseau neuronal 13/5/1 (59%)

A peu près équivalent à
un seul réseau 13/33/1

Un petit calcul...

Conclusion sous l'hypothèse que chaque classifieur faible est indépendant et à la même proba p de bon classement

Si chaque classifieur faible a un taux de classification $> 50\%$, alors le classifieur fort (décision collective par vote) a un taux de classification supérieur à chaque classifieur faible.

Considérons 3 classifieurs indépendants

Soit A_i l'événement: un individu est classé correctement par le classifieur i .

$$P(A_i) = p_i \quad P(\bar{A}_i) = 1 - p_i$$

Considérons 1 classifieur résultant de l'agrégation des ces 3 classifieurs
La décision est le résultat d'un vote majoritaire

Soit A l'événement: un individu est classé correctement par le classifieur agrégé, i.e. est correctement classé par au moins 2 classifieurs.

$$A = (A_1 \cap A_2 \cap A_3) \cup (\bar{A}_1 \cap A_2 \cap A_3) \cup (A_1 \cap \bar{A}_2 \cap A_3) \cup (A_1 \cap A_2 \cap \bar{A}_3)$$

Comme les événements A_1, A_2 et A_3 sont deux à deux indépendants, on a:

$$\begin{aligned} P(A) &= p_1 \times p_2 \times p_3 + (1-p_1) \times p_2 \times p_3 + p_1 \times (1-p_2) \times p_3 + p_1 \times p_2 \times (1-p_3) \\ &= p_1 \times p_2 \times p_3 \times (1-3) + p_2 \times p_3 + p_1 \times p_3 + p_1 \times p_2 \\ P(A) &= p_2 \times p_3 + p_1 \times p_3 + p_1 \times p_2 - 2 \times p_1 \times p_2 \times p_3 \end{aligned}$$

Le classifieur agrégé est meilleur si $P(A) > \max(p_1, p_2, p_3)$

$$P(A) > \max(p_1, p_2, p_3) \quad ?? \quad p_i \in [0; 1]$$

$$\text{ex: si } p_1 = 0,9 \quad p_2 = 0,6 \quad p_3 = 0,6 \Rightarrow P(A) = 0,732 \quad (\text{moins bon})$$

$$\text{si } p_1 = 0,9 \quad p_2 = 0,9 \quad p_3 = 0,6 \Rightarrow P(A) = 0,918 \quad (\text{meilleur})$$

$$\text{si } p_1 = p_2 = p_3 = 0,75 \Rightarrow P(A) = 0,84375 \quad (\text{meilleur})$$

• Démonstration dans le cas où $p_1 = p_2 = p_3 = p$

$$P(A) = 3p^2 - 2p^3$$

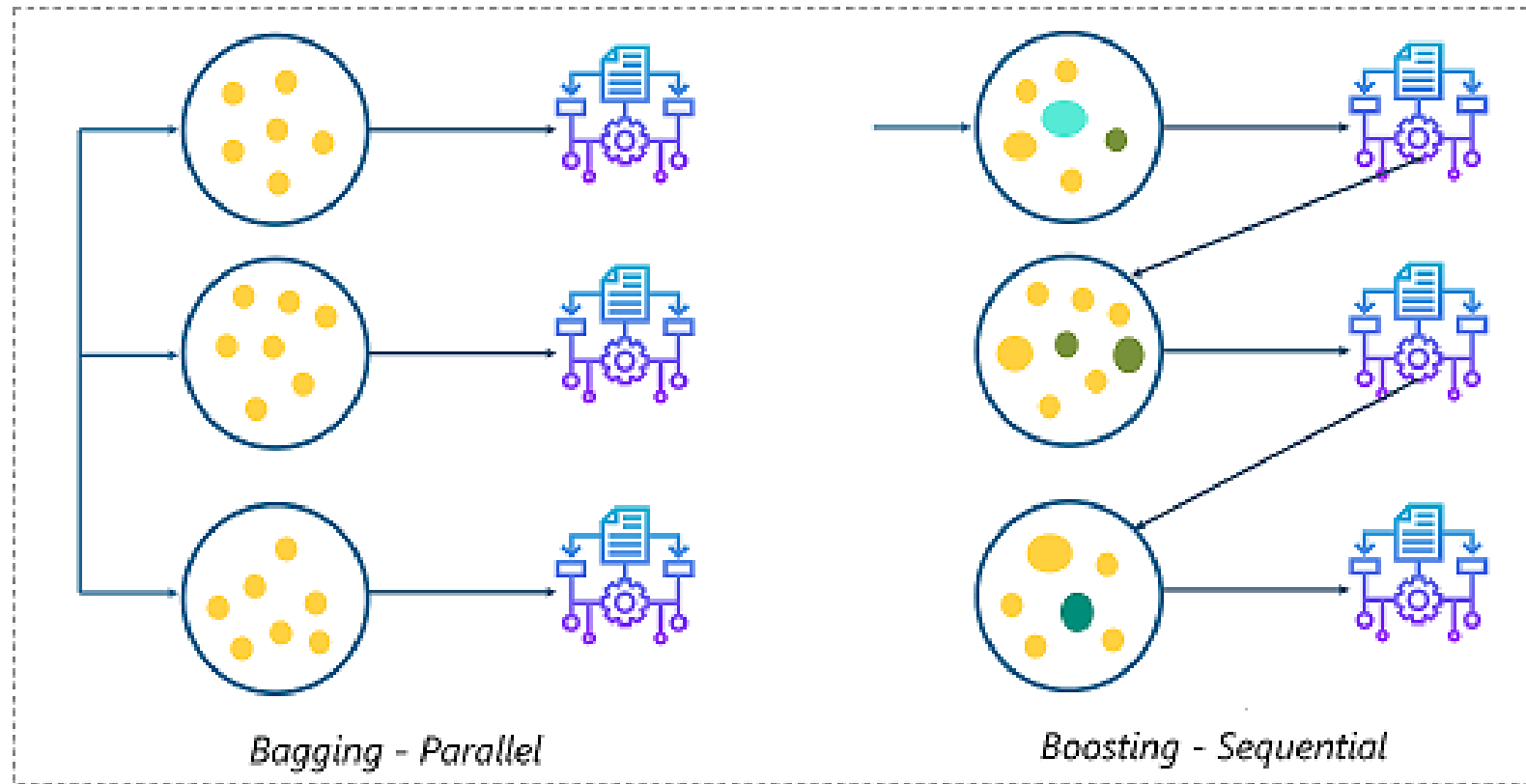
$$\begin{aligned} P(A) - p &= 0 \Leftrightarrow -2p^3 + 3p^2 - p = 0 \\ &\Leftrightarrow p(-2p^2 + 3p - 1) = 0 \end{aligned}$$

Les racines sont $0, \frac{1}{2}$ et 1

p	0	$\frac{1}{2}$	1
$P(A) - p$	0	-	0

$$P(A) > p \quad \text{ssi} \quad p \in \left] \frac{1}{2}; 1 \right[$$

Deux classes de modèles ensemblistes



Forêt aléatoire
Random forest

ADABOOST,
Gradient Boosting,
XGBOOST, etc.

Réseaux adverses génératifs

- ❑ GAN (Generative Adversarial Networks) – Ian J. Goodfellow et al.
<https://arxiv.org/abs/1406.2661>

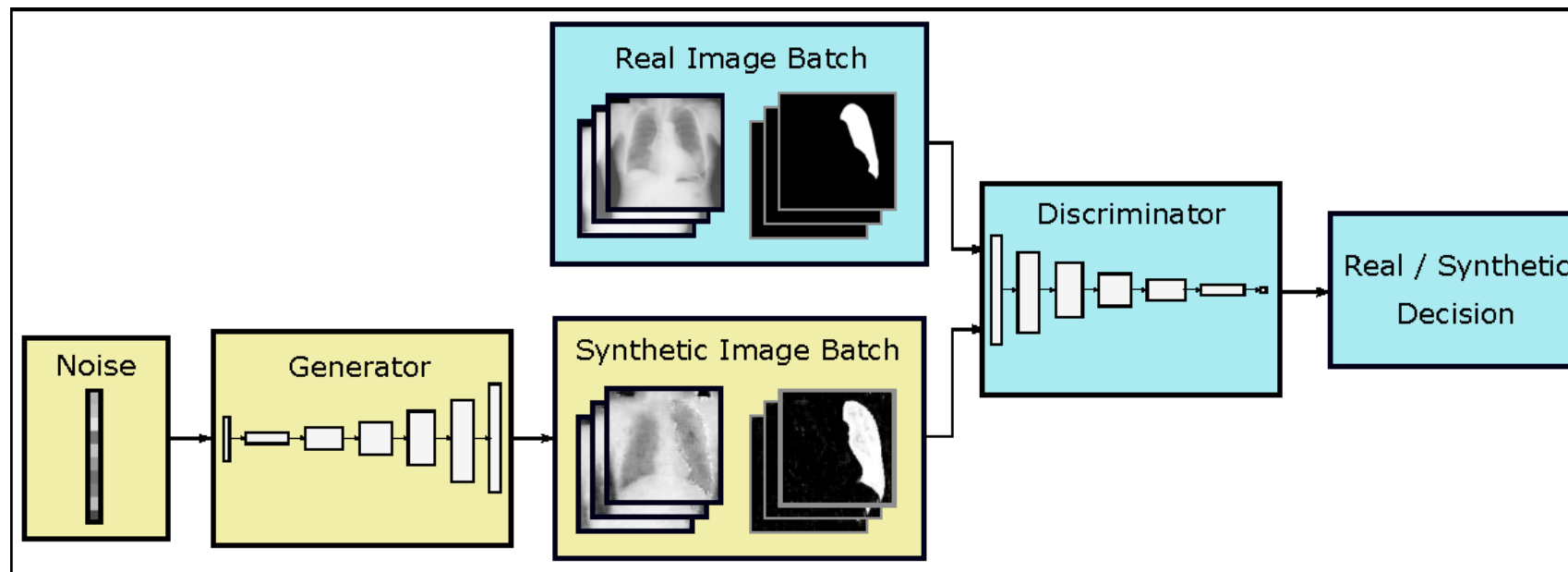


Fig. 1. Proposed GAN architecture incorporating the segmentation mask in the real and synthetic image batches

Neff, T. et al. "Generative Adversarial Network based Synthesis for Supervised Medical Image Segmentation." (2017). DOI:[10.3217/978-3-85125-524-9-30](https://doi.org/10.3217/978-3-85125-524-9-30)

Interprétabilité et Explicabilité

Loi pour une république numérique

Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique

Après l'article L. 311-3-1 du code des relations entre le public et l'administration, sont insérés les articles R. 311-3-1-1 et R. 311-3-1-2 ainsi rédigés :

« Art. R. 311-3-1-1.-...

« Art. R. 311-3-1-2.-L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes :

« 1° Le degré et le mode de contribution du traitement algorithmique à la prise de décision ;

« 2° Les données traitées et leurs sources ;

« 3° Les paramètres de traitement et, le cas échéant, leur pondération, appliqués à la situation de l'intéressé ;

« 4° Les opérations effectuées par le traitement ; ».

Explicabilité, interprétabilité, intelligibilité, ...

- ❑ Prévoir n'est pas comprendre
- ❑ Expliquer un modèle ou un algorithme $\neq$ Expliquer une décision

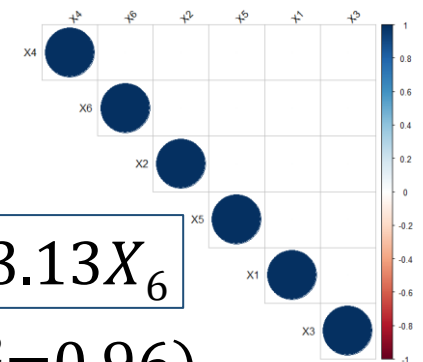
Exemple (J.-M. Marion) :

Extrait du tableau des données

obs	X1	X2	X3	X4	X5	X6	Y
1	-3,783	-0,958	-0,143	-0,130	0,020	0,217	12,2
2	-4,565	-0,991	1,101	0,040	-0,376	-0,226	14,9
3	-2,259	-1,097	-0,072	-0,477	0,065	0,006	23,1
4	-2,533	-0,287	0,079	-0,017	0,262	0,013	13,29
5	-1,021	-0,691	0,175	0,090	0,418	-0,066	28,74
...	...	...	...	...	...	...	...

Matrice des corrélations des variables explicatives

```
      X1      X2      X3      X4      X5      X6
X1  1.000000e+00  8.203945e-05  1.802859e-04 -0.0002677495 -8.875586e-05  9.734168e-06
X2  8.203945e-05  1.000000e+00 -1.412881e-04 -0.0003234629 -1.704042e-04 -8.785121e-05
X3  1.802859e-04 -1.412881e-04  1.000000e+00 -0.0004194711 -2.510295e-05 -3.190574e-04
X4 -2.677495e-04 -3.234629e-04 -4.194711e-04  1.0000000000  4.195911e-04  4.791949e-04
X5 -8.875586e-05 -1.704042e-04 -2.510295e-05  0.0004195911  1.000000e+00 -4.156241e-04
X6  9.734168e-06 -8.785121e-05 -3.190574e-04  0.0004791949 -4.156241e-04  1.000000e+00
```



Equation de régression multiple :

$$Y = 53.14 + 17.80X_1 - 19.92X_2 + 10.40X_3 + 19.85X_4 - 24.76X_5 + 33.13X_6$$

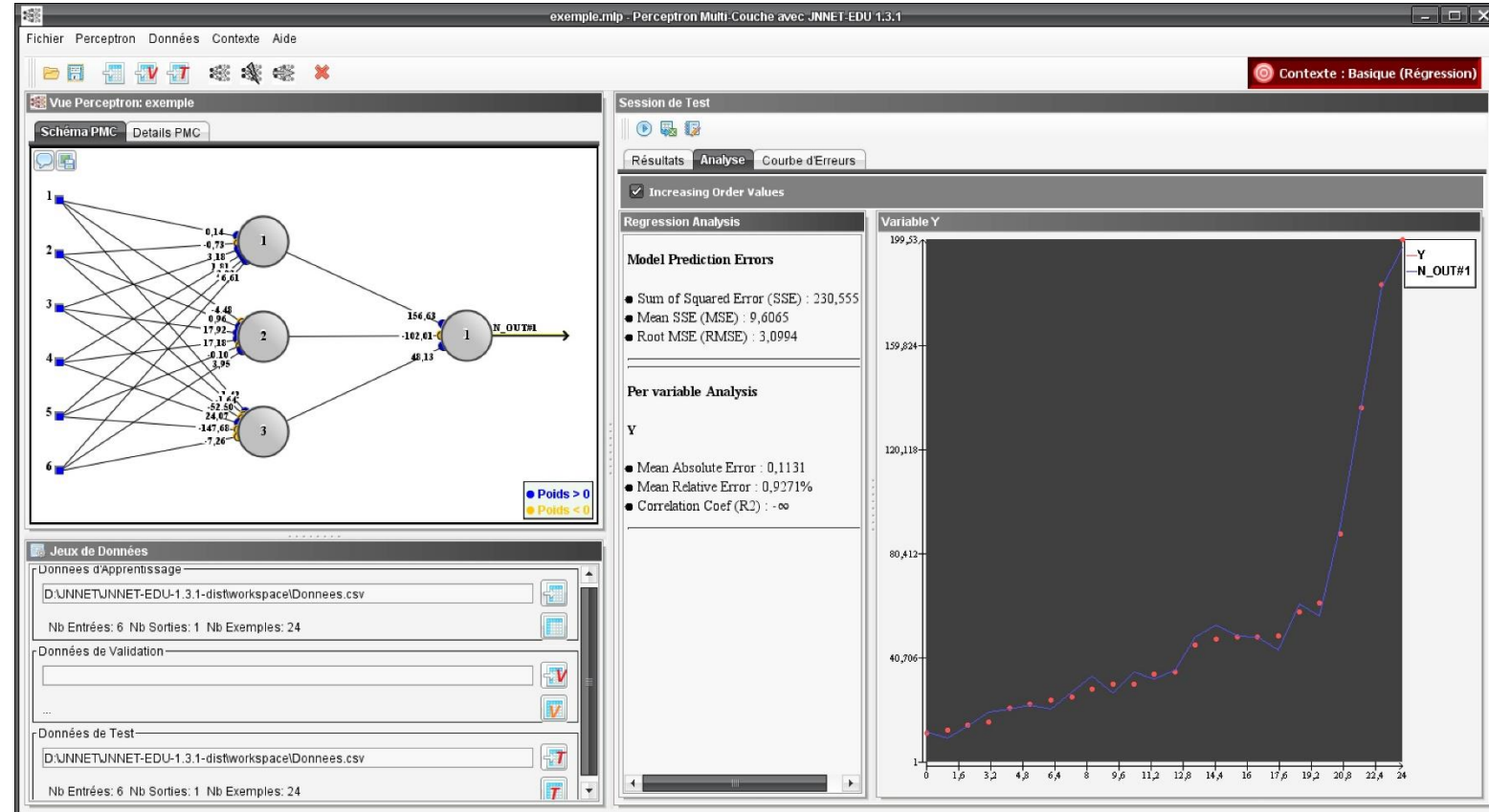
($R^2=0.96$)

Exemple modèle de régression

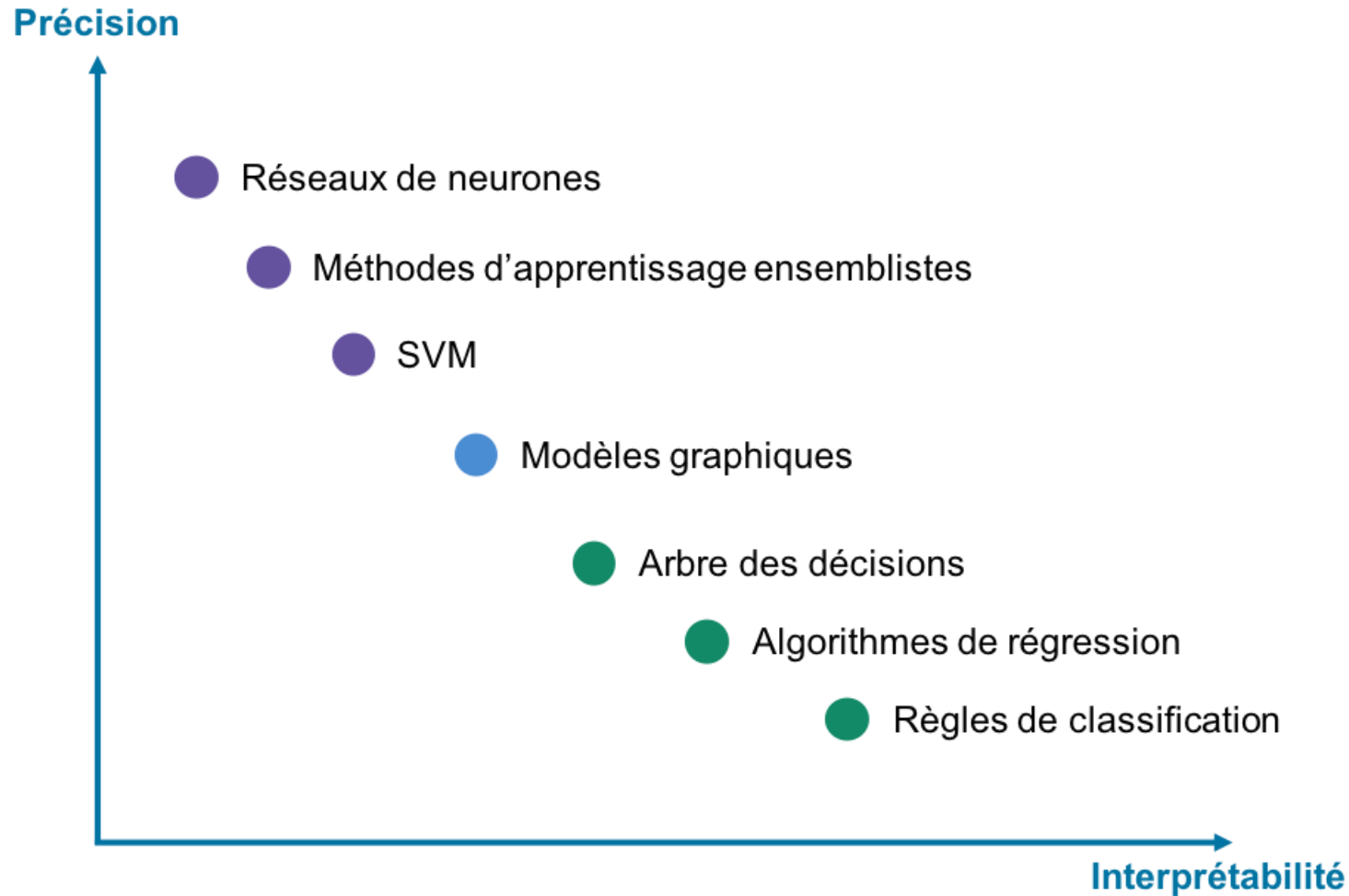
Equation de
régression multiple :

$$Y = 53.14 + 17.80X_1 - 19.92X_2 + 10.40X_3 + 19.85X_4 - 24.76X_5 + 33.13X_6$$

Perceptron à 2
couches :



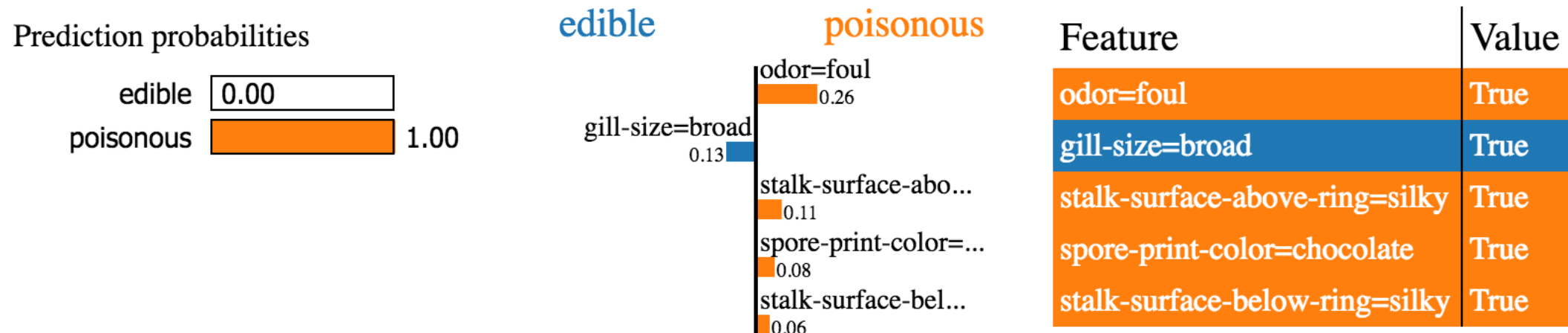
Performance versus explicabilité



<https://www.actuia.com/contribution/jean-cupe/linterpretabilite-de-lia-le-nouveau-defi-des-data-scientists/>
H. Dam, T. Tran, A. Ghose, Explainable Software Analytics. <https://arxiv.org/pdf/1802.00603.pdf>

Méthode LIME (*Local Interpretable Model-Agnostic Explanations*)

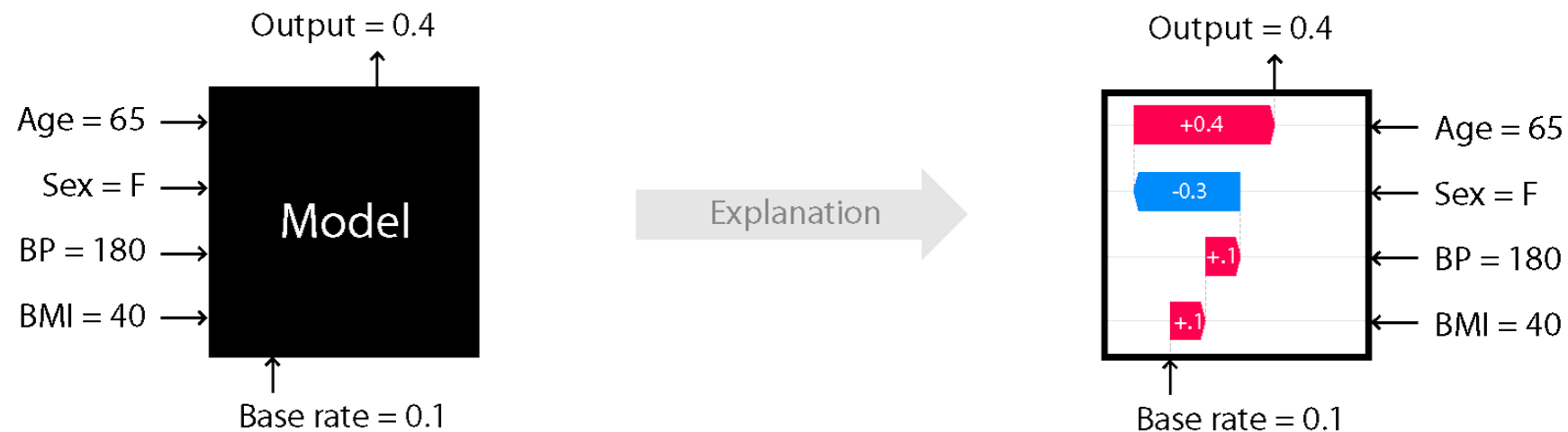
- ❑ "Why Should I Trust You?": Explaining the Predictions of Any Classifier, de Marco Tulio Ribeiro, Sameer Singh et Carlos Guestrin, 2016
- ❑ Analyse a posteriori : indépendant du modèle (« *model agnostic* »)
- ❑ Approximation locale linéaire (modèle de régression construit sur des individus similaires simulés)



D'après le site des auteurs : <https://github.com/marcotcr/lime>

Méthode SHAP (*SH*apley *Ad*ditive *ex*Planations)

- ❑ *A Unified Approach to Interpreting Model Predictions*, de Scott M. Lundberg et Su-In Lee, NIPS 2017
- ❑ Analyse a posteriori : indépendant du modèle (« *model agnostic* »)
- ❑ Pour un individu donné, indique la contribution de la valeur prise par chaque variable à la décision (valeur) prise pour cet individu.



D'après le site des auteurs : <https://github.com/slundberg/shap>

Explicabilité “par construction”

- ❑ Hybrider *deep-learning* et *feature engineering*
- ❑ Combiner arbres de décision et réseaux neuronaux :
 - *Neural Decision Tree, Deep Neural Decision Tree, Deep Neural Decision Forest*
- ❑ Exemple basique sur un problème de régression :
 - Découper en plages de valeurs -> classification
 - Construire un arbre de décision en contrôlant sa profondeur
 - Entraîner un perceptron multicouche sur chaque feuille
- ❑ L'idée n'est pas très récente !

Cid-Sueiro J., Ghattas J., Figueiras-Vidal A.R. (1997) Decision Trees Based on Neural Networks. In: Docampo D., Figueiras-Vidal A.R., Pérez-González F. (eds) Intelligent Methods in Signal Processing and Communications. Birkhäuser, Boston, MA. https://doi.org/10.1007/978-1-4612-2018-3_10

Données multicentriques : Effet de lots

Le « batch effect » (biomédical, génomique)

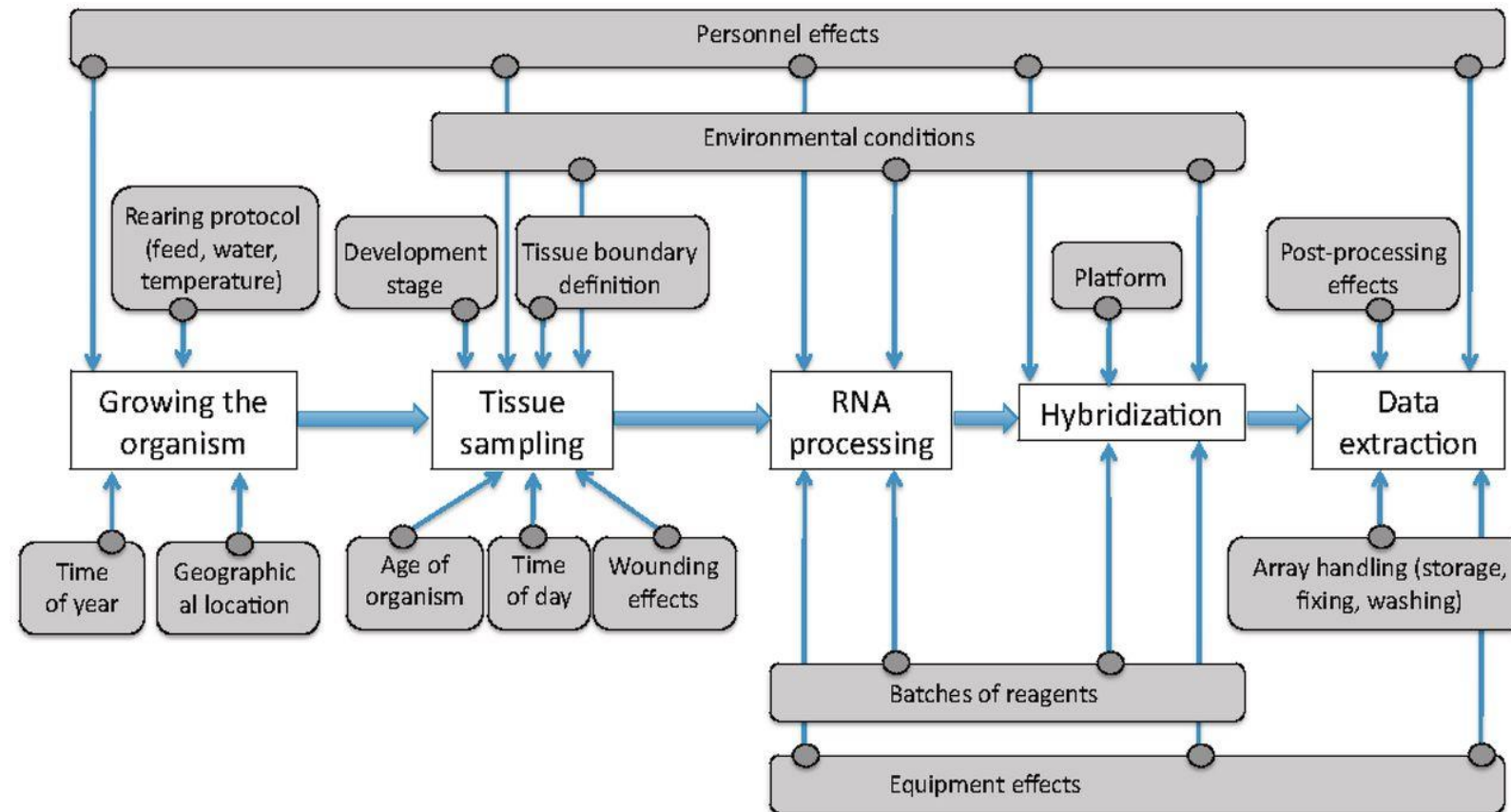
*The most well-known source of latent variation in genomic experiments are batch effects—when samples are processed on **different days**, in **different groups** or by **different people**.*

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034

*In practical data analysis, the observations included in a dataset sometimes form distinct groups—denoted as “batches”; for example, measured at **different times**, **under different conditions**, by **different persons** or even in **different labs**.*

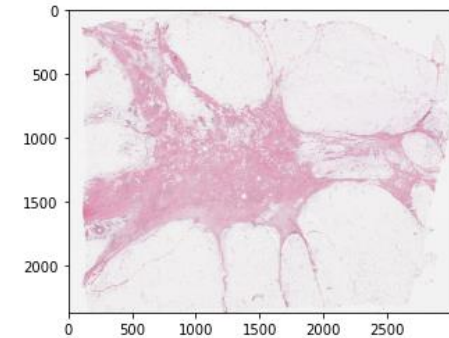
Hornung, R., Boulesteix, AL. & Causeur, D. Combining location-and-scale batch effect adjustment with data cleaning by latent factor adjustment. *BMC Bioinformatics* 17, 27 (2016). <https://doi.org/10.1186/s12859-015-0870-z>

Origines du *batch effect* en génomique



Cosmin Lazar, Stijn Meganck, Jonatan Taminiau, David Steenhoff, Alain Coletta, Colin Molter, David Y. Weiss-Solis, Robin Duque, Hugues Bersini, Ann Nowé, Batch effect removal methods for microarray gene expression data integration: a survey, *Briefings in Bioinformatics*, Volume 14, Issue 4, July 2013, Pages 469–490, <https://doi.org/10.1093/bib/bbs037>

Harmonisation ou *deep learning* ?



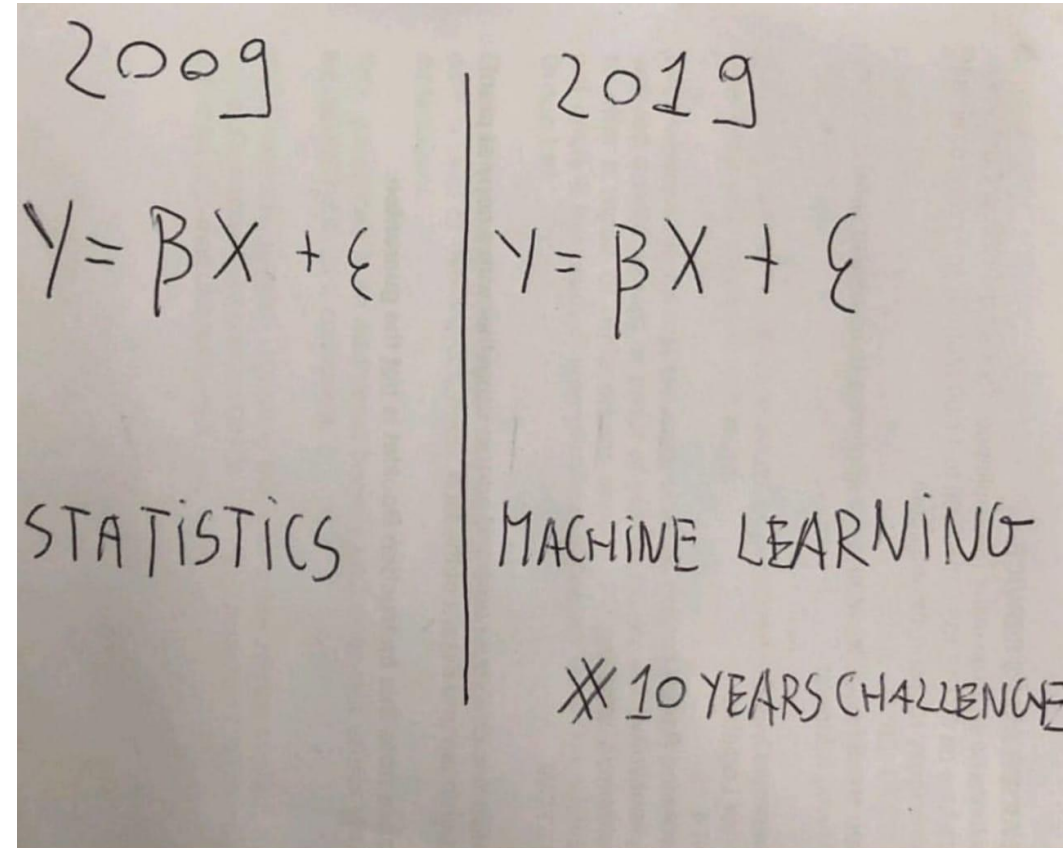
- ❑ Harmonisation avec une méthode de type ComBat, dans une approche *feature engineering* : résultats tests ~ 95% mais 6 mois de travail...

Modèle	Apprentissage	Test IDC	Test BreakHis
ANN	84,44 %	84,5 %	72,81 %
CNN-BaseModel	94,96 %	86,70 %	66 %
NassibNetv1	99,14%	89,19 %	83,89 %
Dense169	88,81 %	87,96 %	83,67 %
Dense201	91,26 %	82,97 %	85,01 %
VGG16	75,85 %	76,66 %	42,72 %
VGG19	75,85 %	76,66 %	42,73 %
ResNet50	95,64 %	85,31 %	85,46 %
ResNet50V2	100 %	88,25 %	88,14 %
Inception_ResNet	75,76 %	76,45 %	42,73 %
MobileNet	98,48 %	88,52 %	91,05 %
NasNet	86,90 %	87,23 %	68,68 %

Conclusion - perspectives

Statistique vs *Machine learning* ?

- ❑ Méthode COMBAT :
régression multilinéaire,
estimation bayésienne, loi
inverse gamma, ...
- ❑ Modèle Cox-NNet : données
de survie, Kaplan-Meier,
modèle de Cox, ...



<https://towardsdatascience.com/the-actual-difference-between-statistics-and-machine-learning-64b49f07ea3>

Attention à l'induction...

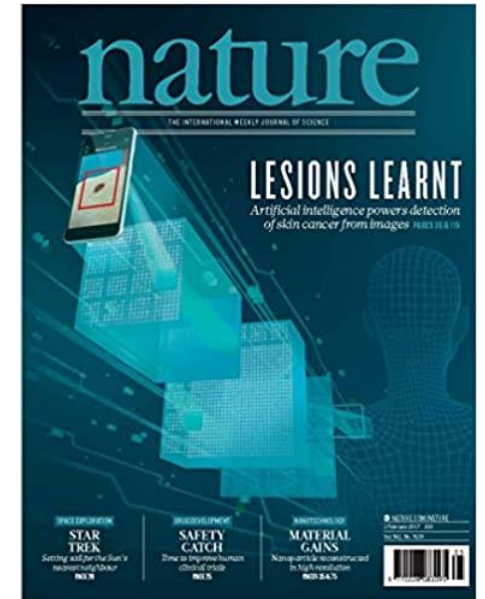
*Un poulet, ayant remarqué que le fermier
Chaque jour à manger lui apportait,
Imprudemment prédit que le fermier
Sa quotidienne pâtée toujours lui fournirait...*

*Un jour, comme d'accoutumée, le fermier arrive,
Et promptement tord le cou du volatile.
De là, se trouve justifié que l'induction
Jamais ne justifiera la moindre conclusion*

Bertrand Russel

Perspectives

- ❑ Développement des méthodes pour expliquer la décision et comprendre les erreurs grossières
- ❑ Application aux données de type graphes
- ❑ Utiliser moins de données (et soulager les experts !) :
 - Dermatologie : lésion bénigne / cancéreuse)
130000 lésions étiquetées !
- ❑ Construire des modèles de référence puis faire de l'apprentissage par transfert (*transfer learning*)
 - Dermatologie : pré-entraînement sur 1 millions d'images
 - SignOvAA : pré-entraînement sur plusieurs cancers féminins hormonaux



<https://cs.stanford.edu/people/esteva/nature/>

Remerciements

- UCO-IMA / LARIS : Dr Nisrine JRAD, Dr Jean-Marie MARION, Elena SPIRINA-MENAND, Gaëlle HARNOIS-MILON
- Univ Angers / LARIS : Dr Nassib ABDALLAH, Pr Jean-Baptiste FASQUEL (responsable équipe ISISV), Pr Patrick Van BOGAERT (PU-PH, neuropédiatrie CHU Angers)
- Univ Angers, CRCINA / ICO : Pr Alain MOREL, Dr Christophe PASSOT
- Equipe HARMONY : M. Hatt (LaTIM - UBO, coordinateur), C. Tauber (INSERM), T. Carlier (CRCINA)