



HAL
open science

Rôle, évaluation et réglementation des intelligences artificielles appliquées au diagnostic et à la thérapeutique

Ismail Alsaïdi

► **To cite this version:**

Ismail Alsaïdi. Rôle, évaluation et réglementation des intelligences artificielles appliquées au diagnostic et à la thérapeutique. Sciences pharmaceutiques. 2023. dumas-04211274

HAL Id: dumas-04211274

<https://dumas.ccsd.cnrs.fr/dumas-04211274v1>

Submitted on 19 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Pour l'obtention du Diplôme d'État de Docteur en Pharmacie

Préparée au sein de l'Université de Caen Normandie

RÔLE, ÉVALUATION ET RÉGLEMENTATION DES INTELLIGENCES ARTIFICIELLES APPLIQUÉES AU DIAGNOSTIC ET À LA THÉRAPEUTIQUE

Présentée par
Ismail ALSAÏDI

Soutenue publiquement le 22 mai 2023
devant le jury composé de

Pr. Patrick DALLEMAGNE	Pharmacien, Professeur des Universités, Directeur du Centre d'Etudes et de Recherche sur le Médicament de Normandie (CERMN, Caen)	Président de jury
Pr. Lydia GUITTET	Praticien Hospitalier, Professeur des Universités, Responsable de l'Unité de Santé Publique au sein de la Direction de la Recherche et de l'Enseignement (CHU de Caen)	Directrice de thèse
Dr. Maxime FRANCE	Pharmacien, Data scientist (MSD, Paris)	Examineur

Thèse dirigée par le Pr. Lydia GUITTET

LISTE DES ENSEIGNANTS-CHERCHEURS

Directrice de la Faculté des Sciences Pharmaceutiques

Professeure Pascale SCHUMANN-BARD

Assesseurs

Professeur MALZERT-FREON Aurélie
Professeure Anne-Sophie VOISIN-CHIRET

Directrice administrative

Madame Sarah CHEMTOB

Directrice administrative adjointe

Madame Amélie LOPEZ

PROFESSEURS DES UNIVERSITES

BOULOUARD Michel	Physiologie, Pharmacologie
BUREAU Ronan	Biophysique, Chémoinformatique
COLLOT Valérie	Pharmacognosie
DALLEMAGNE Patrick	Chimie médicinale
DAUPHIN François	Physiologie, Pharmacologie
DELEPEE Raphaël	Chimie analytique
FABIS Frédéric	Chimie organique
FRERET Thomas	Physiologie, Pharmacologie
GARON David	Botanique, Mycologie, Biotechnologies
GIARD Jean-Christophe	Bactériologie, Virologie
MALZERT-FREON Aurélie	Pharmacie galénique
ROCHAIS Christophe	Chimie organique
SCHUMANN-BARD Pascale	Physiologie, Pharmacologie
SICHEL François	Toxicologie
SOPKOVA Jana	Biophysique, Drug design
VILLEDIEU Marie – HDR	Biologie et thérapies innovantes des cancers
VOISIN-CHIRET Anne-Sophie	Chimie médicinale

MAITRES DE CONFERENCES DES UNIVERSITES

ANDRE Véronique – HDR	Biochimie, Toxicologie
BOUET Valentine – HDR	Physiologie, Pharmacologie
CAILLY Thomas – HDR	Chimie bio-inorganique, Chimie organique
DENOYELLE Christophe – HDR	Biologie cellulaire et moléculaire, Biochimie, Cancérologie
DHALLUIN Anne	Bactériologie, Virologie, Immunologie
DUBOST Emmanuelle	Chimie organique
ELDIN de PECOULAS Philippe – HDR	Parasitologie, Mycologie médicale

GROO Anne-Claire	Pharmacie galénique
KIEFFER Charline	Chimie médicinale
KRIEGER Sophie (Praticien hospitalier) – HDR	Biologie clinique
LAPORTE-WOJCIK Catherine	Chimie bio-inorganique
LEBAILLY Pierre – HDR	Santé publique
LECHEVREL Mathilde – HDR	Toxicologie
LEGER Marianne	Physiologie, Pharmacologie
LEPAILLEUR Alban – HDR	Modélisation moléculaire
LOHARD Steven	Biologie cellulaire et moléculaire, Biochimie, Cancérologie
N'DIAYE Monique - HDR	Parasitologie, Mycologie médicale, Biochimie clinique
PAIZANIS Eleni	Physiologie, Pharmacologie
POTTIER Ivannah	Chimie et toxicologie analytiques
PREVOST Virginie – HDR	Chimie analytique, Nutrition, Education thérapeutique du patient
QUINTIN Jérôme	Pharmacognosie
RIOULT Jean-Philippe	Botanique, Mycologie, Biotechnologies
SAINT-LORANT Guillaume (Praticien hospitalier).....	Pharmacie clinique
SINCE Marc– HDR	Chimie analytique
THEAULT BRYERE Joséphine	Biostatistiques

PROFESSEUR AGREGÉ (PRAG)

PRICOT Sophie	Anglais
----------------------------	---------

PERSONNEL ASSOCIÉ À TEMPS PARTIEL (PAST)

SEDILLO Patrick	Pharmacie officinale
SEGONZAC Virginie	Pharmacie officinale

Enseignants titulaires du Diplôme d'Etat de Docteur en Pharmacie

Remerciements

Ce travail est avant tout le fruit de rencontres et de riches échanges. Je tiens à rendre hommage à toutes les personnes ayant contribué, directement ou non, au choix de mon sujet de thèse et à sa réalisation.

J'aimerais tout d'abord remercier ma directrice de thèse, le Pr. Lydia Guittet, pour sa confiance, sa patience, son investissement et son encadrement exemplaire. Je la remercie chaudement pour ses encouragements, ainsi que la compréhension et l'empathie dont elle a fait preuve au cours des périodes durant lesquelles mes avancements étaient modestes.

J'aimerais ensuite remercier le Pr. Patrick Dallemagne qui m'a accordé l'honneur et le privilège de présider mon jury de thèse. Ma sollicitation fut évidente tant mon respect est immense pour son engagement vis-à-vis de la maladie d'Alzheimer et qu'il est pour moi une véritable inspiration depuis le commencement de mon cursus universitaire.

Je voudrais également exprimer ma profonde gratitude envers le Dr. Maxime France dont le parcours passionnant et les conseils avisés m'ont grandement encouragé à poursuivre mes projets dans l'univers de la science des données. Je le remercie pour tout le temps qu'il m'a consacré et me consacre une fois de plus dans le cadre cette thèse.

Durant mes recherches bibliographiques, j'ai eu l'opportunité de recevoir un enseignement de qualité en apprentissage automatique sans lequel mon manuscrit ne serait pas ce qu'il est aujourd'hui. Je ne remercierai donc jamais assez Etienne Boisseau, Alexandre Le Fournier, Violaine Serugue, Marouan Khata ainsi que toute l'équipe de DATAROCKSTARS qui, par leur pédagogie, leur bienveillance et leur ferveur, ont intensifié ma passion pour les métiers de l'intelligence artificielle et m'ont permis d'acquérir les compétences nécessaires à leur exercice.

Enfin et surtout, je remercie mes parents, ainsi que ma sœur Mariam et mon frère Zakaria, dont l'amour, la loyauté et le soutien indéfectibles constituent le pilier central de ma détermination dans chacun de mes projets.

Un grand merci à tous.

Table des matières

Liste des abréviations	1
Liste des figures	2
Liste des tableaux	7
Introduction	9
PARTIE 1 : Définition, histoire et principes techniques de l'intelligence artificielle.....	13
I. Héritages et définitions.....	13
1. Une discipline aux racines hétérogènes.....	13
2. Pluralité et connexité des définitions.....	15
a. Intelligence artificielle et cybernétique	15
b. Intelligence artificielle et informatique	16
c. Apprentissage automatique et profond.....	17
d. Intelligence artificielle et science des données.....	19
e. Définition internationale.....	19
f. Définition européenne	20
g. Définition française	20
h. Définitions académiques	21
3. Classification macroscopique.....	21
a. Intelligences artificielles faibles	22
b. Intelligences artificielles fortes.....	22
4. Deux courants rivaux mais complémentaires.....	24
a. Intelligence artificielle symbolique	24
b. Intelligence artificielle connexionniste.....	24
c. Qu'en est-il de nos jours ?	25
II. Origines et émergence : de l'idée à la technique.....	26
1. Le mythe d'une machine intelligente (Antiquité - 1950).....	26
a. Antiquité.....	26
b. IXème siècle.....	28
c. XVIème siècle	28
d. XVIIème siècle.....	28
e. XVIIIème siècle	29
f. XIXème siècle	30
g. XXème siècle	30
2. Naissance des premiers programmes d'intelligence artificielle (1950 - années 1970).....	36
3. Premier hiver des intelligences artificielles (1974 - 1980).....	46
4. Ascension des systèmes experts et premiers réseaux de neurones (années 1980)	47

5.	Second hiver des intelligences artificielles (1987 - 1993).....	51
6.	Le développement fulgurant de l'apprentissage automatique (années 1990 à nos jours)	52
III.	L'intelligence artificielle en pratique	60
1.	Les systèmes experts	61
a.	Des systèmes fondés sur les connaissances.....	61
b.	La représentation des connaissances	63
c.	Les ontologies.....	64
d.	L'aide à la décision.....	66
e.	Le défi des connaissances implicites	67
f.	Forces et faiblesses des systèmes experts.....	68
2.	L'apprentissage automatique.....	69
a.	Une science des données et de l'apprentissage	69
b.	L'apprentissage sous toutes ses formes	70
c.	Les quatre problématiques majeures traitées par l'apprentissage automatique.....	74
d.	Exemples d'algorithmes d'apprentissage automatique	77
d.1.	La régression linéaire	78
d.2.	La régression logistique	79
d.3.	La classification naïve bayésienne	80
d.4.	Les machines à vecteurs de support	80
d.5.	La méthode des k plus proches voisins	86
d.6.	La forêt d'arbres décisionnels	87
d.7.	La méthode des k-moyennes.....	90
d.8.	L'analyse en composantes principales	92
e.	Les réseaux de neurones et l'apprentissage profond	94
e.1.	Du neurone formel au réseau de neurones.....	95
e.2.	La diversité des réseaux de neurones.....	97
e.3.	L'apprentissage profond	101
e.4.	Les principaux champs d'application	105
f.	Forces et faiblesses de l'apprentissage automatique	106
3.	L'intelligence artificielle neuro-symbolique : le meilleur des deux mondes ?.....	107
PARTIE 2 : Intelligences artificielles appliquées au diagnostic et à la thérapeutique.....		110
I.	Cas d'usage	110
1.	Définitions.....	110
2.	Addictologie	111
a.	Cas d'usage en diagnostic	111
b.	Cas d'usage en thérapeutique	111
3.	Anatomie pathologique (cas d'usage en diagnostic).....	112
4.	Andrologie.....	113

a.	Cas d'usage en diagnostic	113
b.	Cas d'usage en thérapeutique	113
5.	Anesthésie-réanimation	115
a.	Cas d'usage en diagnostic	115
b.	Cas d'usage en thérapeutique	116
6.	Angiologie	117
a.	Cas d'usage en diagnostic	117
b.	Cas d'usage en thérapeutique	118
7.	Biologie médicale (cas d'usage en diagnostic).....	119
8.	Cardiologie	119
a.	Cas d'usage en diagnostic	119
b.	Cas d'usage en thérapeutique	120
9.	Chirurgie (cas d'usage en thérapeutique).....	121
10.	Dermatologie	122
a.	Cas d'usage en diagnostic	122
b.	Cas d'usage en thérapeutique	123
11.	Endocrinologie	124
a.	Cas d'usage en diagnostic	124
b.	Cas d'usage en thérapeutique	125
12.	Gastro-entérologie	126
a.	Cas d'usage en diagnostic	126
b.	Cas d'usage en thérapeutique	127
13.	Gériatrie (cas d'usage en diagnostic)	128
14.	Génétique (cas d'usage en diagnostic)	128
15.	Gynécologie-obstétrique	129
a.	Cas d'usage en diagnostic	129
b.	Cas d'usage en thérapeutique	131
16.	Hématologie	132
a.	Cas d'usage en diagnostic	132
b.	Cas d'usage en thérapeutique	133
17.	Hépatologie	134
a.	Cas d'usage en diagnostic	134
b.	Cas d'usage en thérapeutique	135
18.	Immunologie	136
a.	Cas d'usage en diagnostic	136
b.	Cas d'usage en thérapeutique	137
19.	Infectiologie.....	138
a.	Cas d'usage en diagnostic	138

b.	Cas d'usage en thérapeutique	139
20.	Médecine d'urgence	140
a.	Cas d'usage en diagnostic	140
b.	Cas d'usage en thérapeutique	142
21.	Médecine nucléaire.....	142
a.	Cas d'usage en diagnostic	142
b.	Cas d'usage en thérapeutique	143
22.	Médecine nutritionnelle.....	145
a.	Cas d'usage en diagnostic	145
b.	Cas d'usage en thérapeutique	146
23.	Médecine palliative (cas d'usage en thérapeutique).....	147
24.	Médecine Physique et de Réadaptation (cas d'usage en thérapeutique)	148
25.	Néonatalogie	150
a.	Cas d'usage en diagnostic	150
b.	Cas d'usage en thérapeutique	151
26.	Néphrologie	152
a.	Cas d'usage en diagnostic	152
b.	Cas d'usage en thérapeutique	153
27.	Neurologie.....	154
a.	Cas d'usage en diagnostic	154
b.	Cas d'usage en thérapeutique	155
28.	Odontologie	157
a.	Cas d'usage en diagnostic	157
b.	Cas d'usage en thérapeutique	158
29.	Oncologie	160
a.	Cas d'usage en diagnostic	160
b.	Cas d'usage en thérapeutique	161
30.	Ophtalmologie.....	162
a.	Cas d'usage en diagnostic	162
b.	Cas d'usage en thérapeutique	163
31.	Orthopédie.....	165
a.	Cas d'usage en diagnostic	165
b.	Cas d'usage en thérapeutique	166
32.	Oto-rhino-laryngologie.....	167
a.	Cas d'usage en diagnostic	167
b.	Cas d'usage en thérapeutique	168
33.	Pédiatrie.....	169
a.	Cas d'usage en diagnostic	169

b.	Cas d'usage en thérapeutique	170
34.	Pharmacologie (cas d'usage en thérapeutique)	171
35.	Radiologie (cas d'usage en diagnostic)	172
36.	Pneumologie	172
a.	Cas d'usage en diagnostic	172
b.	Cas d'usage en thérapeutique	174
37.	Psychiatrie	175
a.	Cas d'usage en diagnostic	175
b.	Cas d'usage en thérapeutique	177
38.	Rhumatologie	178
a.	Cas d'usage en diagnostic	178
b.	Cas d'usage en thérapeutique	179
39.	Soins intensifs	180
a.	Cas d'usage en diagnostic	180
b.	Cas d'usage en thérapeutique	181
40.	Télémédecine	183
a.	Cas d'usage en diagnostic	183
b.	Cas d'usage en thérapeutique	184
41.	Toxicologie	185
a.	Cas d'usage en diagnostic	185
b.	Cas d'usage en thérapeutique	186
42.	Transplantations	187
a.	Cas d'usage en diagnostic	187
b.	Cas d'usage en thérapeutique	189
43.	Traumatologie	190
a.	Cas d'usage en diagnostic	190
b.	Cas d'usage en thérapeutique	191
44.	Urologie	193
a.	Cas d'usage en diagnostic	193
b.	Cas d'usage en thérapeutique	194
II.	Intérêts et perspectives	195
1.	Synthèse des bénéfices pour les patients et les professionnels de santé	195
2.	Vers une numérisation du patient ?	197
III.	Limites, défis et appréhensions	198
1.	L'applicabilité et la performance en contexte clinique	199
2.	L'accessibilité pour les patients et les professionnels de santé	201
3.	Les données de santé	202
4.	L'explicabilité	204

5. Les questions d'ordre éthique.....	206
6. Le cadre réglementaire et juridique.....	209
PARTIE 3 : Évaluation et réglementation des intelligences artificielles appliquées au diagnostic et à la thérapeutique.....	211
I. Réglementation relative aux données de santé.....	211
1. Pourquoi est-il important de réglementer les données de santé ?.....	211
2. Les réglementations en vigueur.....	213
II. Évaluation des modèles.....	221
1. Qu'est-ce qu'un modèle idéal ?.....	221
a. Le sous-apprentissage.....	221
b. Le sur-apprentissage.....	223
c. L'ajustement optimal.....	224
2. Métriques usuellement employées en apprentissage supervisé.....	225
a. Métriques employées en classification.....	226
b. Métriques employées en régression.....	230
3. Validation des modèles.....	233
a. Le principe de la validation.....	233
b. Les méthodes d'échantillonnage.....	234
4. Évaluation finale.....	237
a. La loi de Goodhart appliquée à l'apprentissage automatique.....	237
b. La validation croisée imbriquée.....	238
III. Mise sur le marché (Espace économique européen).....	240
1. La qualification des intelligences artificielles appliquées au diagnostic et à la thérapeutique.....	240
2. L'autorisation de mise sur le marché.....	243
a. Le marquage CE médical.....	243
b. Les exigences réglementaires essentielles.....	244
c. L'évaluation clinique.....	245
d. L'évaluation de la conformité au marquage CE.....	247
e. La mise sur le marché.....	248
3. La surveillance post-AMM.....	249
4. La question de la responsabilité juridique.....	250
PARTIE 4 : Mise en pratique - Méla-Safe.....	253
I. Justification du projet.....	253
1. Contexte.....	253
2. Réponse au besoin.....	254
II. Réalisation.....	255
1. Constitution d'un jeu de données.....	255

a.	Collecte des données	255
b.	Description des données	255
c.	Répartition des données.....	256
d.	Autres prétraitements des données	258
2.	Construction du modèle d'intelligence artificielle	260
a.	Méthode d'apprentissage.....	260
b.	Choix des métriques	261
c.	Entraînement et validation des modèles	262
d.	Evaluation finale.....	264
3.	Intégration du modèle au sein d'une architecture.....	264
a.	Infrastructure de développement	264
b.	Fonctionnalités majeures	265
c.	Interface et expérience utilisateur.....	268
d.	Parcours utilisateur	270
e.	Gestion des vulnérabilités	273
4.	Bilan et axes d'amélioration.....	275
III.	Aspects réglementaires (Espace économique européen).....	277
1.	Réglementation des données	277
2.	Mise sur le marché	278
Conclusion	280
Bibliographie	283

Liste des abréviations

AMM : Autorisation de Mise sur le Marché

ANSM : Agence Nationale de Sécurité du Médicament et des produits de santé

ASA : *American Society of Anesthesiologists*

AUC : *Area Under the Curve* / Aire sous la courbe

BAR : *Balance of Risk*

CE : Conformité Européenne

CESE : Comité Économique et Social Européen

CNIL : Commission Nationale de l'Informatique et des Libertés

CNRTL : Centre National de Ressources Textuelles et Lexicales

COVID-19 : *Coronavirus Disease 2019*

CPU : *Central Processing Unit* / Unité centrale de calcul

D-MELD : *Delta Model for End-Stage Liver Disease*

DARPA : *Defense Advanced Research Projects Agency*

DRI : *Donor Risk Index*

EEE : Espace Économique Européen

EEG : Électroencéphalographie

EMG : Électromyographie

GPU : *Graphics Processing Unit* / Processeur graphique

HAS : Haute Autorité de Santé

IA : Intelligence Artificielle

IBM : *International Business Machines Corporation*

IC : Intervalle de Confiance

IEC : *International Electrotechnical Commission* / Commission électrotechnique internationale

IPL : *Information Processing Language*

ISO : *International Organization for Standardization* / Organisation internationale de normalisation

k-NN : *k-Nearest Neighbors*

LASSO : *Least Absolute Shrinkage and Selection Operator*

LISP : *LISt Processing*

LPPR : Liste des Produits et Prestations Remboursables

MELD : *Model for End-stage Liver Disease*

MIT : *Massachusetts Institute of Technology*

NGAL : *Neutrophil Gelatinase-Associated Lipocalin*

NT-proBNP : *N-Terminal pro-Brain Natriuretic Peptide*

OMS : Organisation Mondiale de la Santé

P-SOFT : *Survival Outcomes Following Pediatric liver Transplantation*

PELD : *Pediatric End-stage Liver Disease*

PSA : *Prostate-Specific Antigen* / Antigène spécifique de prostate

RGPD : Règlement Général sur la Protection des Données

ROC : *Receiver Operating Characteristic*

SARS-CoV-2 : *Severe Acute Respiratory Syndrome Coronavirus 2*

SNITEM : Syndicat National de l'Industrie des Technologies Médicales

SOFT : *Survival Outcomes Following liver Transplantation*

UE : Union Européenne

UI : *User Interface* / Interface utilisateur

UX : *User Experience* / Expérience utilisateur

VGG : *Visual Geometry Group*

Liste des figures

Figure 1. Proportion de littérature relative aux intelligences artificielles répertoriée par MEDLINE au cours du temps.....	9
Figure 2. Diagramme de Venn faisant figurer les relations entre la science des données, l’intelligence artificielle au sens large, l’apprentissage automatique, l’apprentissage profond et les données massives	18
Figure 3. Subdivision du domaine de l’intelligence artificielle selon deux approches	25
Figure 4. Différents mythes faisant écho à l’intelligence artificielle de par le monde	27
Figure 5. Mécanisme interne du <i>Turc mécanique</i>	29
Figure 6. Menuet composé à partir du <i>ludus melothedicus</i>	30
Figure 7. Notation employée par Gottlob Frege.....	30
Figure 8. Schéma d’un système de communication selon Claude Shannon.....	33
Figure 9. Illustration du test de Turing	34
Figure 10. Diagramme de Venn illustrant l’une des faiblesses du test de Turing	35
Figure 11. Arthur Samuel jouant aux dames avec un ordinateur IBM.....	37
Figure 12. Carte perforée.....	39
Figure 13. Exemple de conversation avec <i>ELIZA</i>	40
Figure 14. Représentation d’une rétropropagation s’opérant sur un réseau de neurones	42
Figure 15. Exemple de représentation d’un réseau sémantique	43
Figure 16. Exemple de cadre	43
Figure 17. Exemple de script situationnel : le script du restaurant.....	44
Figure 18. Affichage de SHRDLU	45
Figure 19. Schéma d’un agent rationnel simple	45
Figure 20. Réseau de Hopfield à quatre neurones	50
Figure 21. Exemple de perceptron multicouche	50
Figure 22. Réseau bayésien modélisant des voies de signalisation cellulaire	50
Figure 23. Architecture de la plate-forme de conception du <i>Jardin des hasards</i>	53
Figure 24. Photographie d’un match de la RoboCup-97 en catégorie « robots roulants de taille moyenne »	54
Figure 25. (a) Illustration d’un capteur laser incliné vers le bas pour balayer le terrain devant le véhicule au cours de son déplacement. (b) Chaque laser acquiert un nuage de points tridimensionnel permettant l’analyse de la praticabilité du terrain et des obstacles potentiels	57
Figure 26. Schéma de fonctionnement du MapReduce	58
Figure 27. Les composantes essentielles d’un système expert.....	62
Figure 28. Ontologie convergeant vers un diagnostic d’appendicite dans le cadre du projet <i>Lerudi</i>	65
Figure 29. Représentation d’un système d’aide à la décision.....	66
Figure 30. Interface utilisateur du système expert CDSS.....	67
Figure 31. Modèle SECI de création des connaissances	67
Figure 32. Illustration d’un apprentissage supervisé réalisé à partir de photographies de tumeurs cutanées	71
Figure 33. Illustration d’un apprentissage non supervisé réalisé à partir de photographies de tumeurs cutanées	72
Figure 34. Apprentissage par renforcement	72
Figure 35. Illustration d’un apprentissage par transfert réalisé à partir d’un modèle de reconnaissance visuelle généraliste	73
Figure 36. Modèle de régression prédisant une variable cible à partir d’une variable prédictive.....	74

Figure 37. Modèle de classification prédisant une variable cible à partir de deux variables prédictives	75
Figure 38. Partitionnement de données basé sur deux variables prédictives.....	76
Figure 39. Réduction de dimensionnalité conservant deux dimensions appliquée à des données en 3 dimensions.....	76
Figure 40. Réduction de dimensionnalité appliquée à une problématique de classification à deux variables prédictives	77
Figure 41. Droite représentant un modèle de régression linéaire simple ajustée selon la méthode des moindres carrés	78
Figure 42. Courbe représentant un modèle de régression logistique ajustée selon une fonction logit.	79
Figure 43. Machine à vecteur de support prédisant une variable cible à partir de deux variables prédictives	81
Figure 44. Machine à vecteur de support prédisant une variable cible à partir d'une variable prédictive	81
Figure 45. Distribution permettant une séparation linéaire vs. distribution ne permettant pas une séparation linéaire	82
Figure 46. Effets d'une observation aberrante sur un classifieur à marge maximale	83
Figure 47. Marges obtenues en faisant intervenir un classifieur à marge souple	83
Figure 48. Effets de la variation de l'hyperparamètre C sur la souplesse des marges.....	84
Figure 49. Illustration du biais et de la variance.....	84
Figure 50. Projection d'une distribution d'observations en deux dimensions dans un espace en trois dimensions.....	84
Figure 51. Application d'un noyau polynomial de second degré sur une distribution d'observations en une dimension	85
Figure 52. k-NN à $k=3$ prédisant la malignité d'une tumeur à partir de deux variables prédictives (volume et densité de la tumeur)	86
Figure 53. Illustration de l'influence de l'hyperparamètre k sur une classification par k-NN	87
Figure 54. Apprentissage supervisé exécuté par un algorithme d'arbre décisionnel à partir d'un jeu de données comportant deux variables prédictives, la variable cible étant la malignité de la tumeur	88
Figure 55. Exemple de frontière tracée entre les différentes classes d'un jeu de données au cours d'un apprentissage supervisé par un algorithme de forêt aléatoire.....	89
Figure 56. Illustration du <i>bootstrap aggregation</i>	89
Figure 57. Illustration du fonctionnement d'un modèle de forêt aléatoire	90
Figure 58. Partitionnement exécuté par un <i>algorithme de Lloyd</i> à partir d'un jeu de données comportant deux variables.....	92
Figure 59. Composantes principales tracées sur un jeu de données en deux dimensions	93
Figure 60. Réduction de dimensionnalité exécutée par un algorithme d'analyse en composantes principales sur un jeu de données en deux dimensions	93
Figure 61. Parallèle entre un neurone biologique et son modèle mathématique	95
Figure 62. Représentation schématique d'un neurone artificiel	96
Figure 63. Représentation schématique d'un réseau de neurones artificiels	96
Figure 64. Composants d'un réseau de neurones.....	97
Figure 65. Variété d'architectures de réseaux de neurones artificiels	98
Figure 66. Représentation schématique d'une architecture de réseau de neurones profond convolutif	99
Figure 67. Représentation schématique d'une architecture de réseau de neurones récurrent	100
Figure 68. Effet du décrochage sur l'entraînement	103
Figure 69. Descente de gradient selon le taux d'apprentissage	104
Figure 70. Vue d'ensemble de l'explicabilité et de la performance des algorithmes d'apprentissage automatique	108
Figure 71. Comparaison des courants symbolique et connexionniste de l'intelligence artificielle	108

Figure 72. Processus d'apprentissage automatique.....	114
Figure 73. Exemple de caractéristiques anatomiques et morphologiques constituant les données d'entraînement.....	115
Figure 74. Plans de coupe échographique d'une apophyse transverse.....	116
Figure 75. Représentation schématique de l'architecture présentée (<i>ConvNet₁</i> et <i>ConvNet₂</i>).....	117
Figure 76. Points de cartographie issus des données d'entraînement.....	118
Figure 77. Représentation schématique de l'architecture présentée (<i>SleeveNet</i>).....	121
Figure 78. Représentation schématique du modèle présenté avec un exemple de prédiction de lésion mélanocytaire maligne	122
Figure 79. Méthode d'évaluation et performances des différents modèles présentés	124
Figure 80. Représentation schématique du modèle présenté.....	125
Figure 81. Logigramme du parcours des données d'entraînement.....	130
Figure 82. Interface web du modèle	131
Figure 83. Données en entrée et classes prédites en sortie (<i>AISACS</i>)	133
Figure 84. Représentation schématique du méta-apprentissage à l'origine du modèle présenté.....	135
Figure 85. Processus de prédiction de structures d'épitopes par le modèle présenté.....	137
Figure 86. Représentation schématique du modèle présenté.....	138
Figure 87. Représentation schématique de l'architecture présentée (<i>IDDAP</i>)	139
Figure 88. Représentation schématique du modèle présenté et liste des 19 variables d'entraînement.....	141
Figure 89. Représentation schématique du modèle présenté.....	143
Figure 90. Représentation schématique du modèle présenté (<i>SResCNN</i>)	144
Figure 91. Illustration des différentes étapes de la démarche.....	145
Figure 92. Représentation schématique de l'architecture présentée.....	147
Figure 93. Rééducation du haut du corps assistée par un exosquelette actif.....	148
Figure 94. Collecte, pré-traitement et labellisation des données d'entraînement.....	149
Figure 95. Représentation schématique du second modèle présenté.....	153
Figure 96. Test de tapotement rapide des doigts	154
Figure 97. Méthodologie de réseau basée sur les études d'association pangénomique portant sur la maladie d'Alzheimer.....	156
Figure 98. Représentation schématique des modèles présentés	157
Figure 99. (a) Arbre faisant figurer les différentes modalités organisées selon leur ordre d'intervention dans le processus de planification d'extraction dentaire. (b) Représentation schématique du second modèle	158
Figure 100. Vue d'ensemble de l'entraînement des modèles et de leur application en contexte clinique	159
Figure 101. Vue d'ensemble de l'architecture présentée (<i>CDRscan</i>).....	161
Figure 102. Représentation schématique du modèle présenté.....	162
Figure 103. Représentation schématique du modèle présenté (<i>CADNet</i>).....	164
Figure 104. Représentation schématique de l'architecture présentée	167
Figure 105. Représentation schématique du modèle présenté.....	169
Figure 106. Exemples de signaux acoustiques pulmonaires associés à une obstruction pathologique des voies respiratoires, une pathologie parenchymateuse ou l'absence de pathologie respiratoire	173
Figure 107. Sélection et ajustement des faisceaux	174
Figure 108. Schéma des extractions de caractéristiques réalisées par segmentation sur des images par résonance magnétique fonctionnelle de repos.....	176
Figure 109. Représentation schématique du modèle présenté (<i>ARPN</i>) et de son entraînement.....	177
Figure 110. Comparaison d'un flux de travail traditionnel et d'un flux de travail intégrant le modèle présenté, dans le cadre d'une admission d'un(e) patient(e) gravement brûlé(e).....	181
Figure 111. Représentation schématique du modèle présenté.....	182

Figure 112. Exemples de graphiques représentant la pression inspiratoire au cours d'un cycle respiratoire associées à leurs aires sous la courbe respectives	182
Figure 113. (a) système de rééducation de la cheville avec le pied du sujet monté sur repose-pieds et plate-forme gyroskopique ; (b) signal émis par le gyroscope durant la dorsiflexion d'une cheville atteinte par une hémiplégie (cinq répétitions) ; (c) signal émis par le gyroscope durant la dorsiflexion d'une cheville saine (cinq répétitions).....	184
Figure 114. Méthodologie de collecte, segmentation et labellisation des données d'entraînement ...	186
Figure 115. Exemples de prédictions effectuées sur des foies transplantés et non transplantés	188
Figure 116. Représentation schématique des modèles présentés	189
Figure 117. Exemples de données d'entraînement.....	190
Figure 118. Exemples de données d'entraînement.....	192
Figure 119. Représentations schématiques de deux modèles parmi ceux présentés	193
Figure 120. Application du concept de jumeau numérique aux patients atteints de sclérose en plaques	198
Figure 121. Exemple de carte de saillance réalisée lors d'une prédiction de démence à partir d'une image par résonance magnétique	206
Figure 122. Exemple de cartes de saillance générées sur différentes couches d'un réseau de neurones lors d'une prédiction de trouble de la conduite à partir d'une image par résonance magnétique.....	206
Figure 123. Cycle de vie d'un modèle d'apprentissage automatique.....	212
Figure 124. Synthèse des questions à se poser pour chaque nouveau traitement de données de santé	217
Figure 125. La protection des données dans le monde.....	220
Figure 126. Exemple de modèle de classification ayant sous-appris	222
Figure 127. Exemple de modèle de classification ayant sur-appris.....	223
Figure 128. Illustration de l'évolution de l'ajustement au cours de l'entraînement d'un modèle	225
Figure 129. Exemples de courbe de ROC associées à leur aire sous la courbe.....	228
Figure 130. Exemple d'indice de Youden situé sur une courbe de ROC	228
Figure 131. Exemple schématique d'évaluation d'un modèle de classification binaire.....	230
Figure 132. Illustration schématique du processus de validation d'un modèle d'apprentissage automatique	234
Figure 133. Exemple de validation croisée à cinq blocs	235
Figure 134. Exemple de validation croisée d'un contre tous.....	235
Figure 135. Exemple de validation croisée stratifiée	236
Figure 136. Exemple de <i>bootstrapping</i>	236
Figure 137. Exemple de <i>hold-out</i>	237
Figure 138. Exemple de validation croisée imbriquée à cinq blocs pour la validation externe et quatre blocs pour la validation interne.	239
Figure 139. Arbre de décision destiné à éclairer les fabricants quant à la qualification de leur logiciel en dispositif médical.....	242
Figure 140. Cycle des exigences relatives à la mise sur le marché d'un dispositif médical	249
Figure 141. Incidence de cancer au États-Unis entre 1975 et 2019.....	253
Figure 142. Finalité du projet <i>Méla-Safe</i>	254
Figure 143. Arbres des jeux de données retenus.....	256
Figure 144. Constitution et répartition du jeu de données final	257
Figure 145. Redimensionnement des photographies.....	258
Figure 146. Exemple d'augmentation de données réalisé sur une photographie de tumeur cutanée bénigne	259
Figure 147. Importation, augmentation, transformation en tenseurs et normalisation des données d'entraînement.....	259

Figure 148. Importation, transformation en tenseurs et normalisation des données de validation et de test	259
Figure 149. Importation, redimensionnement, transformation en tenseurs et normalisation des données d'entrée utilisateur	260
Figure 150. Courbes de ROC des deux modèles retenus à l'issue de la phase de validation.....	264
Figure 151. Extrait de la maquette de <i>Méla-Safe</i> réalisée avant son développement : importation d'une photographie.....	265
Figure 152. Exemple de codage d'un fichier jpeg en Base64 exécuté par le composant « dcc.Upload »	266
Figure 153. Fonction de décodage d'une série de caractère en Base64 vers un fichier <i>jpeg</i> ou <i>png</i> ..	266
Figure 154. Fonction de chargement du modèle et d'initiation de l'inférence.....	267
Figure 155. Fonction de préparation de la donnée d'entrée et d'inférence	267
Figure 156. Fonction de rappel dédiée à la prédiction	267
Figure 157. Vue d'ensemble de l'architecture de <i>Méla-Safe</i>	269
Figure 158. Fonction de rappel dédiée au changement de langue.....	269
Figure 159. Page d'accueil (<i>Méla-Safe</i>).....	270
Figure 160. Page de prédiction (<i>Méla-Safe</i>).....	270
Figure 161. Importation d'un fichier à partir d'un périphérique de stockage (navigateur pour ordinateur)	271
Figure 162. Menu déroulant d'importation de fichier à partir d'un périphérique de capture ou de stockage (navigateur pour smartphone).....	271
Figure 163. Message d'alerte consécutif à la tentative de prédiction sans upload préalable d'un fichier conforme.....	272
Figure 164. Page de prédiction affichant un indicateur de chargement (<i>Méla-Safe</i>)	272
Figure 165. Résultat affiché selon que la photographie ait donné lieu à une prédiction de tumeur bénigne ou de tumeur maligne	273
Figure 166. Message d'alerte relatif à la fiabilité des prédictions.....	273
Figure 167. Menu déroulant de changement de langue.....	273
Figure 168. Extrait de la maquette de <i>Méla-Safe</i> réalisée avant son développement : page d'accès aux prédictions sauvegardées.....	276

Liste des tableaux

Tableau 1. Classement des 20 pays les plus avancés en matière d'IA selon le Global AI Index (avril 2023).....	214
Tableau 2. Résultats les plus significatifs obtenus à partir du jeu de validation	263
Tableau 3. Résultats obtenus à partir du jeu de test	264

Introduction

Introduction

L'intelligence artificielle fait assurément partie des sujets les plus discutés de notre époque. En témoigne la mise en avant sans cesse croissante de cette thématique par les revues scientifiques et médias généralistes. Sur MEDLINEⁱ notamment, les termes relatifs à l'intelligence artificielle ont été de plus en plus employés à partir des années 80 [2] (figure 1).

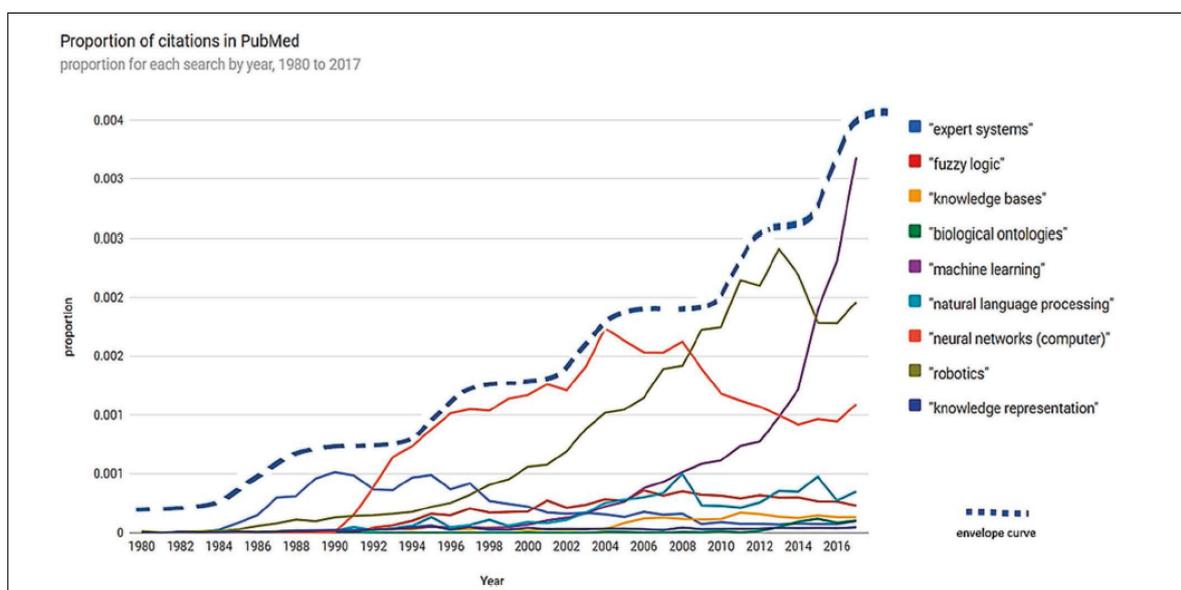


Figure 1. Proportion de littérature relative aux intelligences artificielles répertoriée par MEDLINE au cours du temps (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Laszlo Balkanyi et Ronald Cornet.) [2]

Cette discipline de l'informatique, ayant comme parents les automates et la cybernétique, a émergé dans les années 50 dès lors que les machines furent capables de réaliser des tâches nécessitant une forme d'intelligence [3]. Selon Andrew Ng, chercheur en informatique spécialisé dans l'apprentissage automatique et la robotique, « *Tout comme la révolution industrielle a grandement libéré l'humanité d'un fardeau physique, l'intelligence artificielle a le potentiel de grandement libérer l'humanité d'un fardeau mental* » [4]. En effet, l'intelligence artificielle ne se limite pas à un sujet d'actualité brûlant, puisque celle-ci est déjà à l'œuvre dans des secteurs très divers : le transport, la finance, la communication, le divertissement, mais aussi la santé [5]. Ce potentiel, dont les limites nous sont encore partiellement inconnues, suscite notamment l'intérêt des pouvoirs publics qui perçoivent la nécessité de mener une réflexion de fond concernant les capacités, les perspectives et le cadre réglementaire relatif aux intelligences artificielles [6]. En l'occurrence, en septembre 2017, le Premier ministre Édouard Philippe

ⁱ MEDLINE (*Medical Literature Analysis and Retrieval System Online*) est une base de données bibliographiques, gérée par la bibliothèque nationale américaine (*United States National Library of Medicine*) qui couvre tous les domaines médicaux et bio médicaux de l'année 1966 à nos jours [1].

confié à Cédric Villani, mathématicien et député de la République en marche, cette tâche lourde mais fondatrice [7]. La santé fait partie des quatre axes prioritaires identifiés par Cédric Villani dans son rapport *Donner un sens à l'intelligence artificielle* remis en mars 2018 [8]. Selon ce rapport, les applications de l'intelligence artificielle en santé permettraient notamment d'améliorer la qualité, la sécurité et l'accessibilité des soins. Consécutivement à cette mise en lumière, Cédric Villani et Bernard Nordlinger, chef du service de chirurgie générale digestive et oncologie de l'hôpital Ambroise-Paré, publient en octobre 2018 un ouvrage intitulé *Santé et intelligence artificielle* [9]. Cet ouvrage, fruit d'une collaboration entre l'Académie nationale de médecine et l'Académie des sciences, présente un état de l'art au sein duquel certaines spécialités médicales se distinguaient déjà en 2018 comme étant au centre des expérimentations et avancées : la radiologie, l'oncologie, la dermatologie et les maladies génétiques. Les autres spécialités médicales ne sont pas en reste, celles-ci semblent avoir tous les atouts pour se prêter aux mêmes innovations. Les pistes de travail sont nombreuses et ne manquent pas de motiver les nombreuses start-ups et fabricants de dispositifs médicaux voyant en cette technologie l'opportunité de développer de nouvelles solutions thérapeutiques et diagnostiques [10].

L'appropriation de ces systèmes par les professionnels de santé est essentielle pour garantir leurs développement et usage éthiques [11]. Appréhender leur fonctionnement requiert tout d'abord de maîtriser le vocabulaire propre à l'intelligence artificielle, car ses différents concepts tels que le *big data*, le *machine learning* ou encore le *deep learning* sont souvent amalgamés. Se familiariser avec les différentes techniques, leurs méthodologies d'évaluation, ainsi que leurs capacités et limites connues, est primordial pour établir la façon dont leur intégration à la pratique des soignants peut servir au mieux l'intérêt des patients. Néanmoins, ces technologies sont sujettes à des limitations lorsque leurs applications concernent le diagnostic ou la thérapeutique [12]. Outre la complexité de leurs aspects techniques, leurs éventuelles conséquences néfastes sont susceptibles d'inquiéter les soignants et leurs patients. En effet, l'évolution des pratiques que cela engendre et pourrait engendrer à l'avenir suscite un certain nombre de questions éthiques et juridiques. Au cœur de ces dernières réside l'enjeu de la réglementation, dont la réflexion nécessite un dialogue interdisciplinaire faisant intervenir non seulement des professionnels de santé et de la donnée, mais également des juristes et philosophes spécialistes de l'éthique [13].

Ce manuscrit, en naviguant à travers ces différents sujets, a pour objectif principal de répondre à la problématique suivante : quelles sont les applications actuelles et potentielles des intelligences artificielles en diagnostic et en thérapeutique, et peut-on y transposer la

réglementation et les méthodes d'évaluation appliquées aux produits de santé ? Nous nous fixons comme objectif secondaire de proposer aux lecteurs, notamment aux novices, une vue historique et technique suffisamment large pour mener une réflexion sur le sujet des intelligences artificielles.

PARTIE 1 : Définition, histoire et principes techniques de l'intelligence artificielle

PARTIE 1 : Définition, histoire et principes techniques de l'intelligence artificielle

Cette première partie explorera les origines de l'intelligence artificielle et dressera un schéma général de ce que représente cette discipline en présentant notamment ses aspects techniques majeurs. En premier lieu, celle-ci proposera un panorama des disciplines connexes, définitions et classifications de l'intelligence artificielle.

I. Héritages et définitions

1. Une discipline aux racines hétérogènes

Bien qu'elle fasse partie des champs les plus récents des sciences et de l'ingénierie, l'intelligence artificielle en tant que concept est antérieure à l'Époque contemporaine. En effet, l'intelligence artificielle trouve ses racines les plus anciennes dans la philosophie [14]. Nous verrons qu'au cours du temps, cette discipline s'est caractérisée par son universalité, de fait elle hérite aujourd'hui de millénaires de réflexions et de découvertes. De par sa versatilité et son potentiel, l'intelligence artificielle a su s'adapter à ces différents domaines et élargir en retour leurs champs des possibles. Les disciplines ayant bénéficié et contribué à faire progresser les intelligences artificielles vont du plus général (apprentissage, perception) au plus spécifique (jeu d'échecs, démonstration de théorèmes mathématiques, écriture, traduction, conduite de véhicules, diagnostic médical, etc.) [15]. Parmi ces disciplines fondatrices de l'intelligence artificielle, nous pouvons notamment citer :

- La philosophie [16,17] :
 - La logique et les méthodes de raisonnement (formel, procédural, géométrique, par analogie, par généralisation et abstraction) ;
 - La vision de l'esprit comme étant un système physique opérant selon un ensemble de lois ;
 - L'apprentissage, le langage, le principe de rationalité.

- Les mathématiques [18] :
 - Les représentations formelles (représentation des connaissances par des objets logiques reliés par des propriétés, axiomes et règles) ;
 - Les calculs et algorithmes ;
 - Les probabilités.

- Les neurosciences [19] :
 - L'étude du fonctionnement du cerveau ;
 - La comparaison entre les processus cérébraux et les processus informatiques.

- La psychologie [20] :
 - L'étude des pensées, des actions et de leurs mécanismes ;
 - La psychologie cognitive (étude des fonctions cognitives humaines) ;
 - Les sciences cognitives (comment un modèle informatique pourrait être mis à l'œuvre dans l'étude de la mémoire, du langage et de la pensée ?).

- La linguistique [21] :
 - L'étude du lien entre le langage et la pensée ;
 - Le développement d'intelligences artificielles dans le domaine de la linguistique moderne (traitement automatique du langage naturel).

- L'économie [22] :
 - La théorie de la décisionⁱⁱ ;
 - La théorie de la décision probabiliste (combinaison de la théorie de la décision et de la théorie des probabilitésⁱⁱⁱ en prise de décision) ;
 - La théorie des jeux^{iv}.
 - Le processus de décision markovien^v.

ⁱⁱ Domaine de l'économie étudiant les prises de décision des individus dans des situations où les conséquences sont incertaines [23].

ⁱⁱⁱ Branche des mathématiques traitant de l'analyse de situations aléatoires et incertaines. En économie, elle est utilisée pour modéliser et quantifier les risques et les incertitudes associés aux décisions économiques [24].

^{iv} Domaine des mathématiques s'intéressant aux interactions stratégiques des agents (appelés « joueurs »). Cette théorie a pour objectif de formaliser des situations conflictuelles inhérentes à une communauté composée de ces individus en interaction, de discuter puis de proposer des solutions à ces conflits. La conception des solutions est notamment guidée par des critères d'optimalité individuelle ou collective, de cohérence temporelle, de justice distributive [25].

^v Modèle stochastique où un agent, dont les résultats de ses actions sont aléatoires, prend des décisions [26].

- Le génie informatique [27] :
 - Conception et fabrication de systèmes informatiques adaptés et suffisamment puissants pour permettre le développement et le fonctionnement des intelligences artificielles.

- La cybernétique et la théorie du contrôle^{vi} [29] :
 - Conception d'agents en mesure de recevoir des informations provenant de leur environnement ;
 - Conception de systèmes en mesure d'identifier l'optimum d'une fonction objectif^{vii} au cours du temps.

2. Pluralité et connexité des définitions

La diversité des disciplines ayant contribué à développer les intelligences artificielles a donné lieu à une pluralité de ses définitions. Bien que celles-ci soient similaires sémantiquement, ces définitions présentent des subtilités selon la perspective adoptée.

a. Intelligence artificielle et cybernétique

Dans les années 50, nous assistons à l'émergence d'un nouveau domaine de recherche portant sur l'étude des processus cognitifs et leur simulation. John Bates fonde le *Ratio Club*, un club anglais informel qui réunit des psychologues, des médecins, des mathématiciens et des ingénieurs entre 1949 et 1958 pour discuter de ce nouveau domaine de recherche qu'ils nomment « cybernétique » [31]. En parallèle, en juillet 1956 lors de la conférence de Dartmouth aux Etats-Unis, John McCarthy choisit un terme différent pour désigner ce domaine de recherche : « intelligence artificielle » [15]. Nous assistons alors à une rivalité sémantique et 2 communautés se forment sur la base de cette polarité. Celles-ci, malgré des niveaux de financement qui varient avec le temps, seront en interaction permanente. Cette rivalité a cependant conduit ces deux communautés à diverger et a, par la suite, donné lieu à deux disciplines distinctes.

^{vi} Branche de l'ingénierie et des mathématiques appliquées ayant comme objet la conception et le développement de systèmes de contrôle automatique pour réguler le comportement des systèmes dynamiques [28].

^{vii} Fonction mathématique utilisée pour évaluer la qualité ou la performance d'une solution ou d'un système, par rapport à un ensemble de critères ou d'objectifs spécifiques. Elle est souvent utilisée dans des problèmes d'optimisation, où l'objectif est de trouver la meilleure solution possible en maximisant ou en minimisant la valeur de la fonction objectif [30].

L'édition actuelle du *Dictionnaire de l'Académie française* définit la cybernétique comme étant la « science des systèmes dans lesquels l'effet obtenu agit à son tour, par rétroaction, sur le mécanisme provoquant cet effet, afin d'obtenir un résultat constamment adapté au but désiré » tandis qu'elle définit l'intelligence artificielle comme un « ensemble de propriétés rapprochant du cerveau humain certains systèmes informatiques très évolués. » [32,33]. L'*Encyclopédie Larousse* propose, pour l'intelligence artificielle, la définition suivante : « ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. » [34]. Cependant, cette dernière ne précise pas ce qu'elle entend par « intelligence », un thème plus vaste encore et sujet à débat que celui de l'intelligence artificielle. Parmi les définitions de l'intelligence proposées par l'Académie française, nous pouvons par exemple citer les suivantes : « Faculté de comprendre, de concevoir, de connaître, et notamment faculté de discerner ou d'établir des rapports entre des faits, des idées ou des formes pour parvenir à la connaissance. » et « Aptitude à adapter son comportement à une situation nouvelle » [35]. Ces deux définitions de l'intelligence paraissent complémentaires tant elles en décrivent des aspects différents. Nous pouvons en retenir les notions de conceptualisation, de connaissance et surtout d'adaptation.

Si de nos jours la cybernétique s'intéresse aux systèmes et aux interactions, l'intelligence artificielle est davantage orientée vers les sciences cognitives en se fixant comme objectif de modéliser l'intelligence à l'instar, par exemple, de la physique qui vise à modéliser les phénomènes naturels de l'univers [36]. L'intelligence artificielle n'est donc aujourd'hui plus à confondre avec la cybernétique.

b. Intelligence artificielle et informatique

Il est important de préciser que l'intelligence artificielle est aussi à distinguer de l'informatique. L'Académie française définit l'informatique comme étant la « science du traitement rationnel et automatique de l'information » ou encore comme « l'ensemble des applications de cette science. » [37]. Il s'agit donc d'une science qui englobe une large gamme de technologies et qui, comme nous l'avons énoncé précédemment, permet à l'intelligence artificielle d'exister en étant son support. En retour, l'intelligence artificielle a permis le progrès de l'informatique (traitement de données, optimisation des processus, etc.) et promet de révolutionner ce domaine à l'avenir [38].

Classiquement, les méthodes en informatique permettent le traitement de nombres ou de textes spécifiques, de façon machinale et par le calcul. Celles-ci sont capables de suivre des algorithmes^{viii} rigides et achevés et ne sont donc généralisables qu'à une classe de problèmes semblables. En comparaison, les méthodes propres aux intelligences artificielles tentent de mimer le raisonnement humain et sont par conséquent généralisables à des domaines et contextes plus variables. Bien qu'elle emprunte à l'informatique ses procédés, une intelligence artificielle est supposée affranchir un programme de la nécessité de suivre une méthode simple et précise. Les intelligences artificielles sont capables de traiter des symboles, d'utiliser des inférences, des heuristiques^{ix} et des raisonnements y compris dans le cas où ces derniers sont incertains (ex. : jeu d'échecs, compréhension et traduction de texte, reconnaissance d'images, diagnostic médical, etc.). En nous permettant un néologisme, nous pourrions presque nommer l'intelligence artificielle : « informatique heuristique » [41].

c. Apprentissage automatique et profond

Un éventail aussi large de finalités a engendré un vaste spectre de techniques pour parvenir à celles-ci. Qui-plus-est, les mécanismes intellectuels qu'une tâche implique peuvent s'avérer nombreux, complexes, évolutifs et imprévisibles [42]. Les intelligences artificielles ont donc dû s'adapter en reposant sur des techniques de plus en plus élaborées et capables d'apprendre, c'est ce que l'on appelle l'apprentissage automatique ou encore l'apprentissage profond. Ces concepts sont souvent sujets aux amalgames alors qu'en réalité ils désignent, bien qu'ils soient imbriqués, des choses différentes [43] :

- L'apprentissage automatique (ou *machine learning*) fait partie des techniques employées dans le fonctionnement d'une forme particulière d'intelligences artificielles. Celui-ci permet à un programme d'apprendre à réaliser des tâches pour lesquelles il n'est pas programmé de façon formelle. Au lieu d'indiquer explicitement des règles et des actions à une architecture, l'apprentissage automatique permet à cette architecture de prendre des décisions basées sur un apprentissage à partir de données. Cela présente

^{viii} L'Académie française définit un algorithme comme suit : « Méthode de calcul qui indique la démarche à suivre pour résoudre une série de problèmes équivalents en appliquant dans un ordre précis une suite finie de règles. » [39].

^{ix} Selon le CNRTL (Centre National de Ressources Textuelles et Lexicales), une méthode heuristique procède par approches successives en éliminant progressivement les alternatives et en ne conservant qu'une gamme restreinte de solutions tendant vers celle qui est optimale [40].

l'avantage d'accroître l'efficacité, l'adaptabilité et la scalabilité^x de ce programme. Nous pouvons distinguer plusieurs formes d'apprentissage automatique que nous développerons par la suite.

- L'apprentissage profond (ou *deep learning*) quant à lui désigne des techniques employées dans le fonctionnement d'une forme plus récente et spécifique d'apprentissage automatique. Ces techniques attribuent à un programme la capacité de s'appuyer sur des neurones artificiels organisés en couches. Ces couches, en étant interconnectées, forment ce que l'on nomme un « réseau de neurones artificiels » et tentent ainsi d'imiter le fonctionnement du cerveau humain. Cela permet à une architecture d'effectuer des tâches souvent plus complexes qu'en apprentissage automatique traditionnel, comme par exemple la reconnaissance d'image, de son, ou le traitement du langage. Selon les données que l'on fournit à ce réseau de neurones pour son apprentissage, celui-ci sera plus ou moins adapté à une tâche ou une autre.

Ces techniques sont aussi à distinguer des intelligences artificielles qui englobent non seulement l'apprentissage automatique et l'apprentissage profond mais également l'ensemble des théories mathématiques et scientifiques ayant amené à leur élaboration. Schématiquement, les intelligences artificielles incluent les techniques d'apprentissage automatique qui elles-mêmes incluent les techniques d'apprentissage profond (*figure 2*) [45].

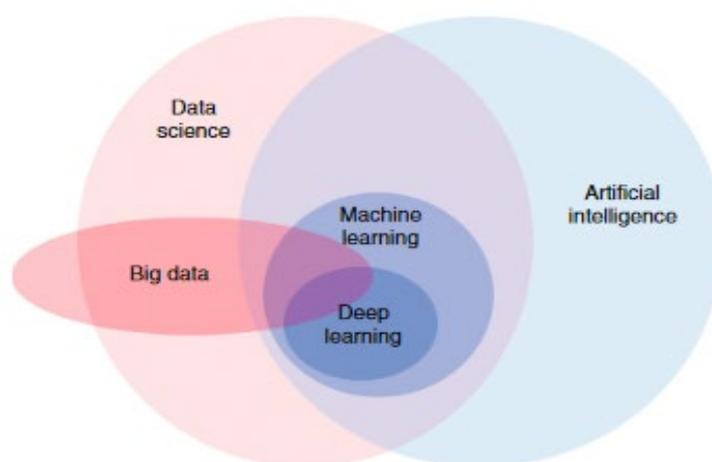


Figure 2. Diagramme de Venn faisant figurer les relations entre la science des données, l'intelligence artificielle au sens large, l'apprentissage automatique, l'apprentissage profond et les données massives.
(Reproduit avec la permission de l'éditeur Elsevier. Auteurs : P. Ignacio Dorado-Díaz et coll.) [46]

^x Selon la Commission d'enrichissement de la langue, la scalabilité (ou extensibilité) désigne « l'aptitude d'un produit ou d'un système à fonctionner correctement, sans perdre ses propriétés essentielles, lors d'un changement d'échelle d'un ou plusieurs paramètres » [44].

d. Intelligence artificielle et science des données

La science des données, communément désignée par son anglicisme *data science*, regroupe également un ensemble d'outils et de techniques mais leur périmètre s'étend au-delà de celui des intelligences artificielles [47]. Comme son appellation l'indique, il s'agit d'une science s'intéressant aux données et à leurs traitements. L'apprentissage automatique s'appuie aussi en partie sur les données, c'est pourquoi la science des données y est souvent associée.

Il ne faut cependant pas les confondre car, si l'apprentissage automatique est tourné vers les mathématiques et la programmation spécifiques aux intelligences artificielles, la science des données couvre un tout autre spectre. Ce spectre comprend notamment la collecte, l'exploration, l'analyse, la visualisation ainsi que les manipulations de données telles que l'intégration ou la modification, mais aussi les techniques d'apprentissage automatique. Quoiqu'il en soit, nous pouvons assurément affirmer que l'apprentissage automatique et la science des données ont mutuellement bénéficié des avancées de l'un et de l'autre.

e. Définition internationale

Au cours des années 90, L'ISO^{xi} (*International Organization for Standardization*) a publié la série de normes ISO IEC 2382 visant à définir le vocabulaire relatif aux technologies de l'information [48,49]. Parmi les définitions publiées, nous pouvions retrouver pour la première fois celle de l'intelligence artificielle : « *Capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage.* » [50]. La dernière mise à jour de cette définition par l'ISO a été publiée en 2015 : « *Discipline qui traite des systèmes informatiques capables d'exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement, l'apprentissage et l'auto-amélioration.* » [51].

^{xi} L'ISO (*International Organization for Standardization*) est une organisation internationale non gouvernementale, créée en 1947, dont les 167 membres sont les organismes nationaux de normalisation. Par ses membres, l'organisation réunit des experts qui mettent en commun leurs connaissances pour élaborer des normes internationales d'application volontaire, fondées sur le consensus, pertinentes pour le marché, soutenant l'innovation et apportant des solutions aux enjeux mondiaux [48].

f. Définition européenne

En 2017, le CESE^{xii} (Comité Économique et Social Européen) publie dans le Journal officiel de l'Union européenne un avis concernant les enjeux économiques et sociétaux relatifs aux intelligences artificielles. Au sein de celui-ci, le CESE propose une définition de l'intelligence artificielle : « *Il n'existe pas de définition précise unique de l'IA. Il s'agit d'un concept qui englobe un grand nombre de (sous-)domaines, tels que l'informatique cognitive (cognitive computing: algorithmes qui raisonnent et apprennent à un niveau supérieur, c'est-à-dire plus humain), l'apprentissage automatique (machine learning: algorithmes qui apprennent de manière autonome à réaliser des tâches), l'intelligence augmentée (augmented intelligence: coopération entre l'homme et la machine) et la robotique intelligente (IA intégrée dans des robots). L'objectif principal de la recherche et du développement en matière d'IA est toutefois d'automatiser les comportements intelligents, entre autres la capacité de raisonner, de collecter des informations, de planifier, d'apprendre, de communiquer, de manipuler, de signaler et même de créer, de rêver et de percevoir.* » [53].

g. Définition française

Dans le *Journal officiel de la République française*, un arrêté définissant plusieurs termes en informatique est publié le 27 juin 1989. Parmi les termes définis, nous pouvons retrouver l'intelligence artificielle : « *Discipline relative au traitement par l'informatique des connaissances et du raisonnement.* » Ce même arrêté définit également les systèmes experts, une forme d'intelligence artificielle capable de raisonner à partir de faits et de règles connues : « *Ensemble de logiciels exploitant dans un domaine particulier des connaissances explicites et organisées, pouvant se substituer à un expert humain.* » [54].

^{xii} Le CESE est un organe consultatif de l'UE qui se compose de représentants d'organisations de travailleurs et d'employeurs et d'autres groupes d'intérêts. Il transmet des avis sur des questions européennes à la Commission, au Conseil de l'UE et au Parlement européen et sert ainsi de lien entre les instances de décision et les citoyens de l'UE [52].

h. Définitions académiques

Les enseignants-chercheurs n'ont pas été en reste lorsqu'il s'est agi de proposer des définitions de l'intelligence artificielle. En voici quelques-unes :

- « *La science et l'ingénierie visant à créer des machines intelligentes, particulièrement des programmes informatiques intelligents.* » par John McCarthy (1927-2011) [55] ;
- « *L'étude et la conception d'agents intelligents ; c'est-à-dire des systèmes susceptibles d'analyser leurs environnements pour faire des choix ou mener des actions dans le but de maximiser leurs performances pour un objectif donné.* » par Stuart Russell (né en 1962) et Peter Norvig (né en 1956) [56] ;
- « *La construction de programmes informatiques capables d'accomplir des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains.* » par Marvin Minsky (1927-2016) [57] ;
- « *Etude des activités intellectuelles de l'homme pour lesquelles aucune méthode n'est a priori connue.* » par Jean-Louis Laurière (1945-2005) [58] ;
- « *Faire faire par une machine des tâches que l'homme accomplit en utilisant son intelligence* » par Dominique Pastre [41].

3. Classification macroscopique

A l'instar des définitions, les classifications sont multiples et nous aborderons au cours de ce manuscrit plusieurs d'entre elles. Parmi celles-ci, la classification qui caractérise les intelligences artificielles selon l'axe faible/forte (ou étroite/générale) est certainement la plus simple à appréhender, celle-ci vise à différencier les intelligences artificielles selon leurs capacités.

a. Intelligences artificielles faibles

Les intelligences artificielles dites faibles (ou étroites) correspondent aux intelligences artificielles telles que nous les connaissons à l'heure actuelle. Elles peuvent être de complexité très variable, les systèmes faisant intervenir de l'apprentissage automatique ou profond entrent notamment dans cette catégorie [59]. Celles-ci représentent une technologie, dite « automatique », qui tente de simuler l'intelligence humaine dans un domaine d'expertise spécifique. Elles ont vocation à assister les humains dans leurs tâches en exécutant des fonctions bien définies et en nombre limité [60]. Leur fonctionnement repose sur la transformation de données en informations exploitables par le biais de calculs et de prédictions effectués par des modèles. Elles sont en particulier employées pour automatiser des tâches fastidieuses ou effectuer des analyses complexes. Bien qu'autonomes, ces intelligences artificielles sont incapables d'avoir une compréhension globale de ce qu'elles effectuent ou de développer une conscience. Nous sommes, à l'heure actuelle, incapables de développer des réseaux de neurones équivalents à un cerveau humain tel qu'il est décrit par les sciences cognitives et les neurosciences. Par conséquent, elles présentent des capacités d'adaptation limitées et exigent d'être réajustées pour accomplir des tâches allant au-delà de leur rôle initial. Si certaines d'entre-elles semblent parfois atteindre un niveau de raisonnement comparable à celui des humains, il ne s'agit en réalité que de l'exécution d'un ensemble limité de fonctions pré-programmées [61].

A ce jour, seule cette forme d'intelligences artificielles a pu être créée et déployée (ex. : *AlphaGO*, *Siri*, *ChatGPT*, voitures autonomes, fils d'actualité et suggestions d'amis sur les réseaux sociaux, filtres anti-spam des boîtes mails, aide à la décision, domotique^{xiii}, etc.). Si des systèmes aussi complexes et performants sont qualifiés de « faibles », à quoi pourrait ressembler une intelligence artificielle « forte » ?

b. Intelligences artificielles fortes

L'intelligence artificielle dite « forte » (appelée aussi générale ou encore véritable) est un type d'intelligence artificielle polyvalente, capable d'exécuter des tâches intellectuelles variées et couvrant un large périmètre de domaines avec la même approche intellectuelle qu'un humain [61]. L'un des avantages majeurs des intelligences artificielles fortes est qu'elles nécessiteraient

^{xiii} Le dictionnaire Larousse définit la domotique comme suit : « Ensemble des techniques visant à intégrer à l'habitat tous les automatismes en matière de sécurité, de gestion de l'énergie, de communication, etc. » [62].

peu de temps pour être formées à accomplir leurs tâches et qu'elles seraient donc capables d'atteindre rapidement un haut niveau de qualification. Si l'intelligence artificielle faible tente de simuler la cognition humaine, l'intelligence artificielle forte, quant à elle, vise à atteindre la véritable cognition humaine [63]. Théoriquement, celles-ci seraient en mesure, tel un humain, de mener une réflexion approfondie et globale, de raisonner analogiquement et métaphoriquement, d'apprendre, de percevoir, de planifier, d'interagir avec leur environnement, de communiquer, de résoudre des énigmes, d'être créatives, d'anticiper, de porter des jugements, de prendre du recul sur leurs actions, de plaisanter ou même de ressentir des émotions et développer une conscience (machines à prise de conscience de soi) voire une sagesse^{xiv} [65]. Nous employons ici le terme « théoriquement » car, bien que de nombreux chercheurs de par le monde tentent de concevoir une intelligence artificielle forte, il n'en existe à ce jour aucun exemple réel [66]. En effet, c'est surtout par le biais de la science-fiction que nous pouvons entrevoir ce à quoi elle pourrait ressembler (ex. : *Blade Runner*, *Ex machina*, *Her*, *Black Mirror*, *Westworld*, *l'Odyssée de l'espace*, *Terminator*, *A.I. Intelligence artificielle*, *Real Humans*). Ces œuvres entretiennent le mythe de l'intelligence artificielle forte, en abordant notamment les enjeux éthiques et moraux que celle-ci impliquerait, et vont même parfois jusqu'à imaginer une machine supérieure à l'homme : la superintelligence artificielle [67].

Selon certains spécialistes, la première intelligence artificielle forte pourrait voir le jour entre 2030 et 2045 [68]. Cependant, d'autres experts prédisent que cela arrivera plutôt au cours du prochain siècle et d'autres encore estiment qu'atteindre cet objectif pourrait s'avérer impossible [69]. Enfin, certains théoriciens affirment qu'une intelligence artificielle forte aurait probablement besoin de suivre le même processus de développement que les humains avant d'atteindre son plein potentiel. En passant par plusieurs stades de développement, celle-ci développerait ses capacités par l'apprentissage au fil des interactions et des expériences, ce qui nécessiterait un temps considérable. Il s'agit manifestement d'un sujet clivant, d'autant plus qu'il n'existe pour le moment aucun consensus sur les éléments permettant d'identifier avec certitude qu'une intelligence artificielle est forte.

^{xiv} Le CNRTL définit la sagesse comme suit : « Sagesse de celui/celle qui possède le savoir, la science à un degré élevé ainsi que les qualités de jugement, d'habileté, de raison, de prudence. » [64].

4. Deux courants rivaux mais complémentaires

En nous focalisant sur l'intelligence artificielle qui opère de nos jours, c'est-à-dire l'intelligence artificielle faible, nous pouvons distinguer deux courants qui s'affirment au cours de son histoire : l'intelligence artificielle symbolique et l'intelligence artificielle connexionniste.

a. Intelligence artificielle symbolique

L'intelligence artificielle est vue par certains comme étant une science de l'ingénieur. Selon cette vision, le propos de l'intelligence artificielle est de parvenir à habilitier avec succès des machines à effectuer des tâches ou des activités considérées comme intelligentes en y intégrant des agents s'articulant autour de connaissances acquises et d'un raisonnement formel. Nous parlons ici de l'intelligence artificielle symbolique (approche logiciste), que l'on retrouve notamment dans les systèmes experts et les moteurs de règles^{xv}. Ces programmes peuvent s'avérer très performants lorsqu'il s'agit de reproduire des tâches intellectuelles suivant une logique spécifique bâtie selon des règles prédéfinies, leurs capacités d'adaptation et d'apprentissage sont en revanche très limitées voire nulles. L'un des avantages notables des intelligences artificielles symboliques est que celles-ci bénéficient d'une bonne explicabilité car leur logique peut être décortiquée [71].

b. Intelligence artificielle connexionniste

D'autres voient l'intelligence artificielle davantage comme une science cognitive dans le sens où celle-ci se rattacherait à l'étude des mécanismes de l'intelligence, à l'identification et à la reproduction du raisonnement humain. Il s'agit d'une approche visant à émettre des théories et modèles capables de mimer le raisonnement tels que les modèles mentaux ou les réseaux de neurones. Le terme « cognitif » se rapporte à la notion d'apprentissage, à ses processus et à son rôle dans l'intelligence. Il s'agit de l'intelligence artificielle connexionniste (approche neuronale), aussi nommée intelligence artificielle numérique, qui se concrétise à travers l'apprentissage automatique. Il s'agit d'outils puissants, adaptables et capables d'apprendre mais dont l'explicabilité est limitée bien que cela soit très variable selon les modèles [72].

^{xv} L'entreprise IBM définit un moteur de règles comme suit : « *Un moteur de règles est un module qui automatise la gestion de certains processus très variables. Le concept fondamental consiste à séparer les objets concernés par les processus de la logique qui met en œuvre ces processus. La logique est définie en écrivant des règles. Pour chaque processus, le moteur de règles reconnaît les règles à appliquer et les objets sur lesquels elles fonctionnent. En cas de variation de la logique, les règles peuvent varier sans intervention dans l'architecture du système.* » [70].

c. Qu'en est-il de nos jours ?

Ces deux approches ont vécu leurs heures de gloire et leurs heures difficiles de façon alternée depuis les années 50 [73]. A l'aube des premières intelligences artificielles, c'est l'approche symbolique qui avait le vent poupe, la tendance s'est inversée par la suite. En effet, bien qu'elles persistent toutes deux aujourd'hui, nous constatons une perte de vitesse significative pour l'intelligence artificielle symbolique depuis les années 90. La hausse exponentielle des puissances de calcul et des quantités de données a permis à l'intelligence artificielle connexionniste de se développer à pas de géant, notamment les réseaux de neurones dans toutes leurs applications. Pourtant, les intelligences artificielles symboliques et les intelligences artificielles connexionnistes ne sont pas antagonistes en soi, elles furent davantage des rivales du point de vue des financements que pour des raisons fonctionnelles (*figure 3*). De fait, nous pouvons déjà observer les prémices d'une intelligence artificielle « neuro-symbolique » qui réunirait le meilleur de ces deux approches [74].

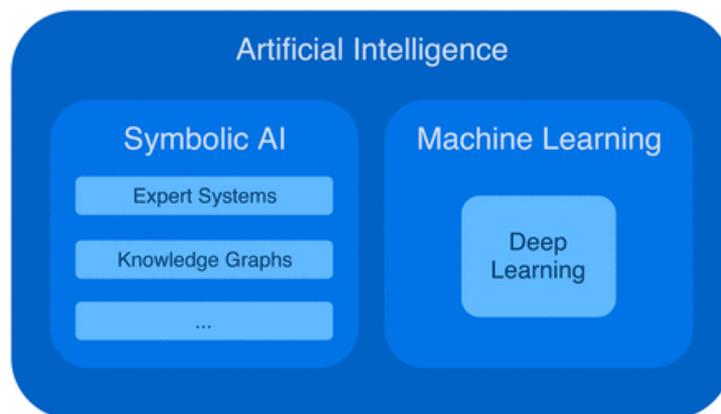


Figure 3. Subdivision du domaine de l'intelligence artificielle selon deux approches : symbolique et connexionniste (Reproduit avec la permission de la revue *The Journal of Nuclear Medicine*. Auteurs : Felix Nensa et coll.) [75]

Le chemin fut long avant que nous puissions nommer, définir l'intelligence artificielle et en établir une classification. Et, bien que celles que nous avons présentées soient considérées aujourd'hui comme pertinentes, nous constaterons que les méthodes employées en intelligences artificielle sont très diverses et font également l'objet de catégories. Voyons à présent comment les contours de l'intelligence artificielle se sont dessinés jusqu'à notre époque.

II. Origines et émergence : de l'idée à la technique

Durant des millénaires, les notions de raisonnement et d'intelligence ont été le sujet de nombreuses réflexions et débats [76]. Les imaginaires de toutes civilisations, nourris par la fascination, n'ont cessé de se questionner sur les origines et le fonctionnement de cette capacité dont l'humain est doté. Ces questionnements ont inmanquablement marqué les mythes, les croyances et les aspirations de ces civilisations, et ainsi l'idée d'objets doués d'intelligence ou d'autonomie fit surface. Nous verrons lors de cette première partie que le domaine de l'intelligence artificielle va au-delà de l'imaginaire puisqu'il tente non seulement de comprendre les entités intelligentes mais également d'en construire. Nous verrons que malgré son caractère moderne, le concept d'intelligence artificielle a des racines ancrées dans notre antiquité. Aussi, à travers cette rétrospective, nous constaterons qu'au fil des années il s'est dégagé quatre écoles de pensées majeures concernant la définition d'une intelligence artificielle. Ces quatre approches, différentes mais complémentaires, s'articulent autour de deux dimensions : penser/agir et performances humaines/rationalité [77].

1. Le mythe d'une machine intelligente (Antiquité - 1950)

a. Antiquité

À l'Antiquité déjà, nous retrouvons des mythes et croyances faisant écho au concept d'intelligence artificielle (*figure 4*). Dans *l'Iliade*, épopée de la Grèce antique dont la création est estimée entre 750 et 700 av. J.-C., le dieu du feu *Héphaïstos* avait créé des objets capables d'autonomie (ex. : tables à trois pieds). Parmi la multitude de personnages ayant insufflé cette capacité, nous pouvons également citer *Dédale* (créateur de *Talos*, un géant de bronze défenseur de la Crète), *Pygmalion* (créateur d'une statue qu'*Aphrodite* amena à la vie), *Zeus* (créateur de *Pandore*, la première femme) ou encore *Prométhée* (créateur des humains) [78].

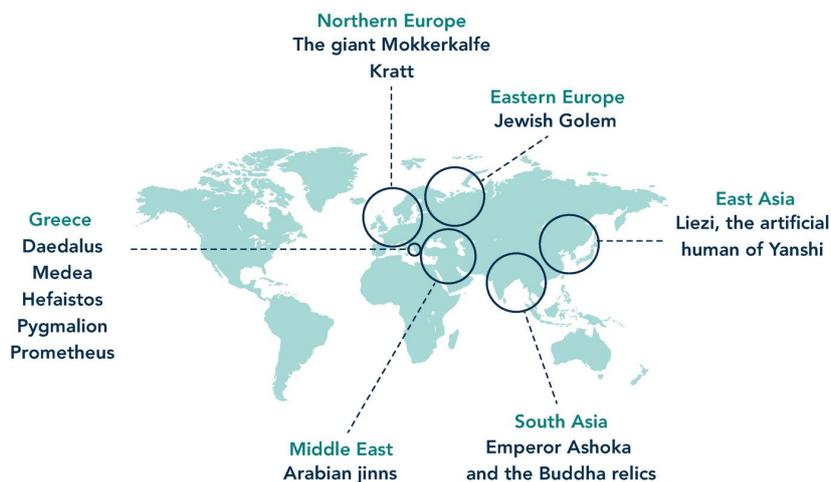


Figure 4. Différents mythes faisant écho à l'intelligence artificielle de par le monde (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Haroon Sheikh et coll.) [78]

La légende du Golem, un être artificiel fait d'argile mais capable de s'animer et se mouvoir, est mentionnée pour la première fois dans la Bible hébraïque (psaumes 139:16) dont la datation traditionnelle estime l'écriture vers 460 av. J.-C. [79]. Dans un registre similaire, la mythologie nordique conte la légende d'*Hrungnir* qui aurait combattu *Thor* en étant assisté de *Mökkurkalfe*, une forme de vie artificielle faite d'argile [80]. *Le vrai classique du vide parfait*, recueil de fables philosophiques et d'aphorismes attribué au philosophe chinois Lie Zi, relate notamment une histoire mettant en scène un automate capable entre autres de marcher, se baisser, se redresser et chanter sur commande [81]. La *Lokapannatti* (« Description du Monde » en pali), un texte issu de la tradition bouddhiste, évoque les reliques de *Buddha* que l'empereur Ashoka souhaite réunir, celles-ci sont protégées par des gardes mécaniques [82]. Les *djinns* (« génies » en arabe) sont des créatures décrites dans la mythologie arabe et la théologie islamique comme étant douées d'un libre-arbitre et d'une intelligence similaire à celle des humains [83].

Dans la Grèce antique, le philosophe Aristote (384-322 av. J.-C.) et le mathématicien Euclide (environ 300 av. J.-C.) formalisent les fondements de la logique telle qu'elle sera employée dans le domaine des intelligences artificielles à partir du XX^{ème} siècle. En étudiant les procédés des raisonnements et en proposant des structures argumentatives irréfutables (ex. : syllogismes^{xvi} dont la conclusion est vraie si la prémisse est vraie), Aristote est à l'origine de cette grande discipline de la philosophie que l'on nomme « la logique » [85]. « *Socrate est un homme, tous les hommes sont mortels, donc Socrate est mortel.* » est probablement le plus célèbre des syllogismes [86]. Les apports d'Euclide quant à eux sont spécifiques aux mathématiques, ceux-ci sont présentés dans son traité mathématique et géométrique *Éléments* qui présente notamment le premier algorithme connu [87].

^{xvi} L'Académie française définit un syllogisme comme suit : « *Raisonnement composé de trois propositions, la majeure, la mineure et la conclusion.* » [84].

b. IX^{ème} siècle

Au cours du IX^{ème} siècle, le mathématicien Al-Khwârizmî systématise les algorithmes en réunissant et en classifiant ceux-ci. Le terme « algorithmes » tient d'ailleurs ses origines de son nom (Al-Khwârizmî), qui signifie « l'homme de Khwârizm », une ville située dans l'actuel Ouzbékistan [88]. Les algorithmes deviendront un outil primordial dans de nombreux domaines dont l'intelligence artificielle.

c. XVI^{ème} siècle

Au XVI^{ème} siècle, la compréhension des mécanismes physiques et des fonctions corporelles s'accélère grâce aux avancées dans les domaines de la mécanique et de la médecine. Les découvertes et les expériences menées par des scientifiques tels qu'Andreas Vesalius, William Harvey et Leonardo da Vinci, contribuent à une meilleure compréhension de la structure et du fonctionnement des organes du corps humain [89,90]. Ces avancées induisent de nouvelles réflexions concernant la conception de machines intelligentes capables de reproduire le fonctionnement des organes humains.

Thomas Harriot, mathématicien et astronome anglais, met au point des systèmes de numération non décimaux, dont le binaire qui sera le fondement du développement des ordinateurs modernes [91].

d. XVII^{ème} siècle

Au XVII^{ème} siècle, René Descartes donne naissance à la thèse métaphysique d'un « animal-machine », selon laquelle le comportement des animaux serait semblable aux mécanismes des machines [92].

Gottfried Wilhelm Leibniz, un philosophe, mathématicien et scientifique allemand, développe en 1671 l'ancêtre de la calculatrice moderne, une machine à calculer fonctionnant à l'aide d'un système de roues dentées [93]. En outre, il met en avant l'idée du codage binaire en 1679 [94].

e. XVIIIème siècle

En 1737, Jacques Vaucanson crée le *fluteur*, premier automate^{xviii} connu de l'histoire. Celui-ci était capable de jouer automatiquement plusieurs airs de musique différents. En 1739, il présente au public une autre de ses inventions : le *canard digérant* capable de se déplacer, s'alimenter, digérer et cancaner à l'image d'un véritable canard [96].

Julien Offray de La Mettrie, médecin et philosophe, publie en 1748 un ouvrage nommé *l'homme-machine* inspiré par l'animal-machine de René Descartes mais adoptant une vision plus déterministe et matérialiste. Dans cet ouvrage, Julien Offray de La Mettrie revendique le parallèle entre les êtres vivants et la mécanique des machines [97].

En 1769, le baron Wolfgang von Kempelen conçoit l'un des plus célèbres tours de prestidigitation : l'automate joueur d'échecs (aussi appelé *Turc mécanique*), un automate qui impressionne par sa capacité à affronter et vaincre un adversaire aux échecs (*figure 5*). En réalité une personne était dissimulée dans le mécanisme, il s'agissait donc d'un leurre mais celui-ci a contribué à susciter l'idée des automates dans l'imaginaire collectif [98].

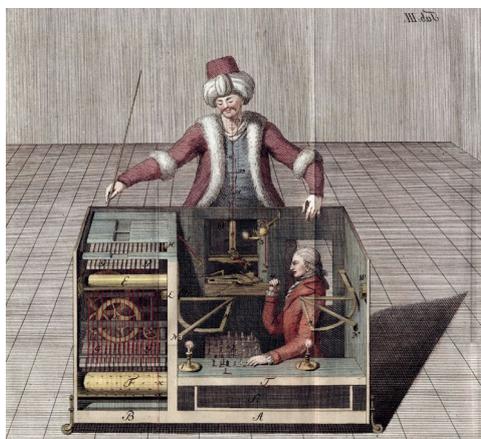


Figure 5. Mécanisme interne du *Turc mécanique*

Au cours du XVIIIème siècle, un document nommé *ludus melothedicus* dont l'auteur est anonyme est découvert à la Bibliothèque Nationale de France (*figure 6*). Ce manuscrit présente une méthode mathématique permettant d'automatiser la composition de menuets par le biais d'un ensemble de règles et de lancements de dés [99].

^{xviii} Le dictionnaire Larousse définit un automate comme suit : « Machine qui, par le moyen de dispositifs mécaniques, pneumatiques, hydrauliques, électriques ou électroniques, est capable d'actes imitant ceux des corps animés. (pris souvent dans le sens d'androïde.) » ou « Machine et mécanisme automatiques, utilisés par exemple pour la peinture et le soudage dans l'industrie automobile. » [95].

⑥ MENUET composé à partir du ludus melothedicus.



Figure 6. Menuet composé à partir du *ludus melothedicus*, extrait d'un compte rendu de réunion tenue le 31 janvier 1964 et portant sur la composition de la musique à l'aide de machines à calculer [100]

f. XIX^{ème} siècle

C'est au XIX^{ème} siècle que la logique, discipline philosophique initiée durant la Grèce antique, se modernise avec notamment l'aboutissement de l'idéographie, un langage visant à représenter la logique mathématique. C'est Gottlob Frege, un mathématicien, logicien et philosophe allemand, qui publie pour la première fois en 1879 un ouvrage portant sur l'idéographie (*figure 7*). Celui-ci sera suivi en 1884 par *Les Fondements de l'arithmétique*, du même auteur, qui approfondit ce concept [101].

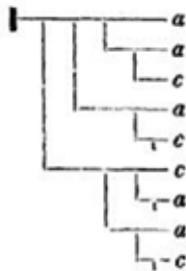


Figure 7. Notation employée par Gottlob Frege (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteur : Bruno Leclercq.) [102]

g. XX^{ème} siècle

En 1912, Leonardo Torres Quevedo, ingénieur des travaux publics, mathématicien, physicien et inventeur espagnol, rend réalité la mise en scène de Wolfgang von Kempelen en inventant *El Ajedrecista*, considéré comme étant le premier véritable automate joueur d'échecs [103].

Neuf ans plus tard, le terme « robot » est employé publiquement pour la première fois lorsque la pièce de théâtre nommée R.U.R (*Rossum's Universal Robots*) est jouée à Prague [104]. Cette œuvre, écrite par l'auteur tchécoslovaque Karel Čapek, met en scène un futur dystopique où

des machines à l'apparence humaine prennent le dessus sur l'humanité. Le mot « robot » ne fut pas choisi au hasard par son auteur car *robota* signifie « corvée » en tchèque [105].

Durant les années 30, Kurt Gödel, Alonzo Church, Jacques Herbrand et Alan Turing, à travers leurs travaux en tant que mathématiciens, progressent dans l'automatisation des calculs et raisonnements [106].

Konrad Zuse, ingénieur allemand, crée en 1941 le Z3 (*Zuse 3*), premier ordinateur programmable en binaire et à virgule flottante fonctionnel. Il convertira les règles du jeu d'échecs en programme pour la première fois en 1945 [107].

Isaac Asimov et John Campbell, auteurs de science-fiction, formulent des lois auxquelles les robots présents dans leurs romans doivent obéir. Celles-ci, que l'on nomme aujourd'hui les « trois lois de la robotique » ou encore les « trois lois d'Asimov », sont présentées pour la première fois en 1942 dans la nouvelle *Runaround* [108] :

- Un robot ne peut porter atteinte à un être humain ni, restant passif, permettre qu'un être humain soit exposé au danger ;
- Un robot doit obéir aux ordres que lui donne un être humain, sauf si de tels ordres entrent en conflit avec la première loi ;
- Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi.

En mai 1942, Frank Fremont-Smith souhaitant encourager l'interdisciplinarité, notamment entre les sciences dites « exactes » et les sciences humaines et sociales, organise une réunion sur le sujet de l'inhibition cérébrale. Cette réunion interdisciplinaire s'avèrera fondatrice pour le domaine de l'intelligence artificielle car celle-ci rassemble plusieurs personnalités qui contribueront grandement à l'émergence de cette discipline : Warren McCulloch et Arturo Rosenblueth (neurobiologistes et physiciens), Margaret Mea et Gregory Bateson (anthropologues), Lawrence Kubie (psychiatre et psychanalyste), Milton Erickson (psychiatre et psychologue), Howard Liddell (psychologue), et bien d'autres. Ce premier événement, qui préfigurera les conférences de Macy, a donné lieu à des échanges informels autour de plusieurs sujets tels que le fonctionnement de la pensée humaine, l'hypnose ou le réflexe conditionné [109].

Warren McCulloch, présent à cette réunion, ainsi que Walter Pitts proposent en 1943 le tout premier modèle mathématique et informatique du neurone biologique basé sur leurs propriétés connues à l'époque. Il s'agissait là d'un modèle simple, binaire, c'est-à-dire dont la sortie vaut 0 ou 1. Cette proposition sera à l'origine des premières constructions de perceptrons et réseaux de neurones artificiels [110].

En conséquence de la réunion sur l'inhibition cérébrale, s'est déroulée une très célèbre série de conférences portant sur la cybernétique appelées aujourd'hui « Conférences de Macy » [111]. Celles-ci, dont la coordination scientifique était confiée à Warren McCulloch, furent au nombre de dix et se sont tenues entre 1946 et 1953, rassemblant ainsi la plupart des acteurs qui allaient jouer un rôle important en ingénierie de l'intelligence et en sciences cognitives [112]. Ces participants étaient issus des sciences « exactes » et des sciences humaines et sociales à l'instar de la réunion sur l'inhibition cérébrale. L'objectif de ces conférences était d'élaborer et alimenter une théorie générale du fonctionnement de la pensée humaine. Cet objectif de départ ne sera pas atteint, néanmoins ces conférences thématiques ont malgré tout permis de grandement approfondir les sujets qui y ont été abordés. Ces conférences ont également eu comme conséquence de former deux communautés aux visions différentes. La première était enthousiaste à l'idée de simuler les processus cognitifs par le biais de machines digitales tandis que l'autre souhaitait analyser ces processus en tenant compte des dimensions émotionnelles et sociales [113]. Dès lors, trois questions majeures sont soulevées :

- Comment un de nos sens pourrait-il être simulé, en particulier la vue (reconnaissance et traitement des images) et l'ouïe (reconnaissance et traitement des sons) ?
- Comment un joueur pourrait-il devenir imbattable ? Les échecs en particulier ont été et seront un terrain ludique approprié à l'étude de cette problématique. Parmi les notions soulevées par celle-ci, nous pouvons citer la théorie des jeux dont les fondements ont été décrits durant les années 1920 et qui est toujours appliquée actuellement à des domaines divers (ex. : l'analyse comportementale durant la crise sanitaire due à la COVID-19) [114].
- Comment le processus de recherche d'un scientifique pourrait-il être simulé ? Cette question s'appliquera dans un premier temps au calcul formel, c'est-à-dire à la démonstration de théorèmes mathématiques.

En 1943, l'armée américaine commande un calculateur entièrement électronique afin d'effectuer des calculs de balistique [115]. Celui-ci sera achevé en 1945 et nommé ENIAC (*Electronic Numerical Integrator and Computer*) [116].

En 1948, Claude Shannon publie un article nommé *A Mathematical Theory of Communication* dans lequel il expose la théorie de l'information. Appliquée à la télécommunication, cette théorie permet de quantifier, à travers une approche probabiliste, le contenu moyen en information d'un ensemble de messages. Dans cet article, Claude Shannon identifie les éléments basiques d'un système de communication : source d'information, transmetteur, canal, bruit, récepteur, destinataire (*figure 8*). Il introduit également le *bit* comme unité d'information [117].

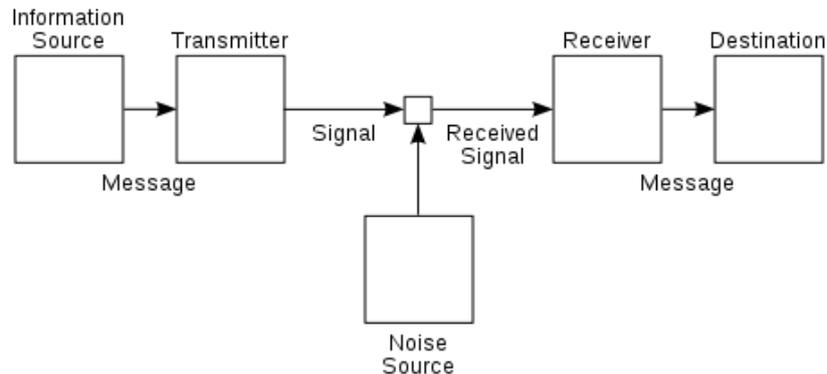


Figure 8. Schéma d'un système de communication selon Claude Shannon [117]

En 1949, il publie un autre article nommé *Programming a computer for playing chess* dans lequel il analyse et met en avant des techniques et algorithmes formant la base des programmes d'échecs développés par la suite [118]. Il construit également un automate composé de 150 relais électromécaniques permettant de jouer aux échecs [119].

Ces avancées de la cybernétique ont jusqu'alors grandement contribué à l'industrialisation. Dans les années 50, les chaînes de montage étaient déjà en mesure de saisir des objets (outils, vis, etc.) et de les assembler (ex. : production de véhicules motorisés) [120]. Mais ces systèmes présentaient des limites puisque ceux-ci étaient incapables de s'adapter aux variations de leur environnement, par exemple si la position des objets à saisir était différente de celle prévue par le programme. Les besoins industriels des uns ainsi que la curiosité scientifique des autres ont suscité des questionnements à ce sujet, en l'occurrence s'il était envisageable de permettre aux programmes informatiques contenus au sein de ces automates de prendre en considération ces variations de l'environnement et de s'adapter à celles-ci. Cette perspective du passage de la robotique que l'on connaissait à une machinerie douée d'adaptabilité a motivé les discussions autour des techniques susceptibles de doter les machines de comportements intelligents et de l'évaluation de cette intelligence. Les premières idées consistaient à apporter de la connaissance sous forme de données autour de la tâche à résoudre afin de permettre au programme d'effectuer une forme d'apprentissage [121].

Alan Turing, fameux mathématicien britannique ayant déchiffré le code *Enigma* utilisé par l'Allemagne durant la seconde guerre mondiale, publie en 1950 *Computing Machinery and Intelligence* [122]. Cet article a notamment pour but de répondre à la question « *Les machines peuvent-elles penser ?* » et propose ainsi une définition opérationnelle de l'intelligence à travers ce que l'on appelle aujourd'hui le test de Turing [123]. Le test de Turing évalue la capacité d'une machine à imiter le comportement humain, telle que jugée par un autre humain en aveugle. Une machine (A) réussit le test de Turing lorsqu'un humain (C) n'est plus en mesure de distinguer la différence entre le comportement de cette machine (A) et celui d'un humain (B) sur une tâche similaire (*figure 9*). Ici, c'est donc la vision « *agir comme les humains* » qui est mise en abstraction en souhaitant faire appliquer par une machine la gestion de connaissances, le raisonnement, le langage, l'apprentissage, la perception et la manipulation d'objets [77]. Les modalités de réussite du test varient donc selon la tâche en question [124]. Par exemple, pour réussir le test dans le cadre d'un échange verbal, une machine doit être capable de répondre aux questions posées par un contrôleur humain et ce dernier doit être dans l'incapacité d'affirmer si les réponses fournies par la machine proviennent d'un humain ou d'un programme informatique. Cela impliquerait que le programme possède les capacités suivantes :

- La représentation des connaissances, permettant au programme de stocker les informations qui lui sont fournies ou qu'il acquière ;
- Le raisonnement automatisé, permettant au programme de faire usage des informations mémorisées afin d'aboutir à des conclusions et de répondre aux questions ;
- Le traitement du langage naturel, permettant au programme de communiquer ;
- L'apprentissage, conférant au programme la capacité de s'adapter à différentes situations, d'identifier des constantes et de les extrapoler.

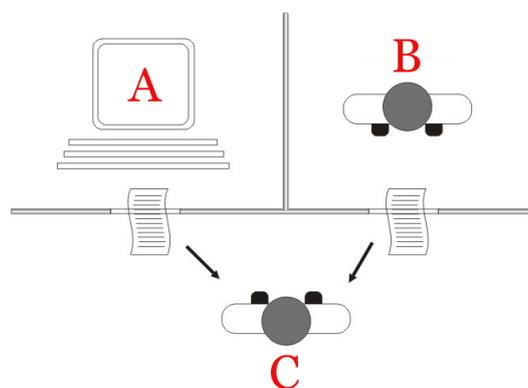


Figure 9. Illustration du test de Turing (Reproduit selon les termes de la licence CC BY 2.5. Auteurs : Juan Alberto Sánchez Margallo. https://commons.wikimedia.org/wiki/File:Turing_test_diagram.png)

Il est important de noter qu'afin d'assurer que le jugement du contrôleur humain ne soit pas influencé, le test de Turing proscrit les interactions physiques directes entre ce contrôleur et la machine passant le test. Ceci est acceptable méthodologiquement car la stimulation physique n'est pas considérée comme indispensable pour évaluer l'intelligence. Cependant, afin de satisfaire l'ensemble des critères évalués par le test de Turing, la machine testée devrait également être capable de :

- Percevoir des objets à l'aide d'un dispositif de vision artificielle ;
- Manipuler des objets et se déplacer (robotique).

Pour ce faire, le test de Turing complet peut être mis en œuvre en permettant au contrôleur d'émettre un signal vidéo et de remettre des objets matériels, ce qui lui permettrait de tester les capacités perceptives et motrices du sujet [125].

Le test de Turing a donc été considéré pour évaluer la capacité de penser et l'intelligence d'une machine. Cependant, lorsqu'une machine réussit le test, peut-on pour autant affirmer de façon absolue que cette machine est douée d'intelligence ? En réalité, le test de Turing présente des faiblesses. De nombreux experts du domaine, tels que Gary Marcus par exemple, ont vivement critiqué la fiabilité de ce test en affirmant que celui-ci était essentiellement en mesure de mettre en évidence la disposition de l'esprit humain à être dupé [126]. Le test suppose qu'un contrôleur humain est capable de déterminer si une machine peut « penser » en comparant ses réactions aux réactions d'un humain. Cette supposition sur laquelle le test se base est sujette à plusieurs problématiques : la fiabilité du jugement du contrôleur, la pertinence d'une comparaison entre une machine et un humain (certains comportements humains peuvent être considérés comme non intelligents et certains comportements considérés comme intelligents sont inhumains), et le fait de comparer uniquement les réactions (*figure 10*).

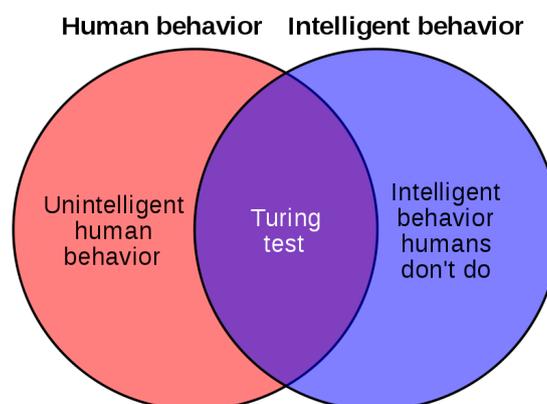


Figure 10. Diagramme de Venn illustrant l'une des faiblesses du test de Turing

Malgré ces faiblesses, Alan Turing a accompli la prouesse de concevoir à l'aube des premières intelligences artificielles un test qui est encore aujourd'hui considéré comme une référence. En effet, les six capacités qu'évalue le test constituent actuellement les composantes majeures de l'intelligence artificielle appliquée [41].

2. Naissance des premiers programmes d'intelligence artificielle (1950 - années 1970)

Si durant la première moitié du 20^{ème} siècle le concept d'intelligence artificielle s'est dessiné, c'est durant la deuxième moitié que le terme sera érigé et que l'on verra naître les premiers programmes d'intelligence artificielle ayant pour objectif de faire réaliser à des machines, en mimant l'activité cérébrale, des tâches jusqu'alors accomplies uniquement par des humains.

C'est en 1951 que Marvin Minsky et Dean Edmonds construisent le SNARC (*Stochastic Neural Analog Reinforcement Calculator*), le premier réseau de neurones artificiels. Celui-ci vise à simuler, à l'aide de diodes, un rat à la recherche de nourriture dans un labyrinthe. Elaboré à l'aide de 3000 tubes électroniques, ce réseau de neurones est composé de 40 neurones et de synapses capables de modifier leur pondération selon qu'un choix aboutisse sur un succès ou un échec [127].

En 1952, Arthur Samuel écrit un programme de jeu de dames fonctionnel sur l'*IBM 701*, le premier ordinateur commercialisé par la compagnie IBM (*International Business Machines Corporation*) [128]. Il élaborera par la suite une version capable d'apprendre qu'il terminera en 1955 et présentera à la télévision en 1956 [129]. Arthur Samuel se lance alors le défi d'apprendre à l'*IBM 701* à le battre au jeu de dames. Malgré la simplicité des règles du jeu, il existe près de 500 milliards de milliards (5×10^{20}) de situations différentes [130]. Il serait très long et difficile de programmer chacune de ces situations ainsi que les différentes réponses de l'ordinateur à celles-ci. Pour pallier à cette complexité, Arthur Samuel eut l'idée de jouer contre l'ordinateur un très grand nombre de fois puis d'utiliser le résultat des parties (victoire/défaite) afin d'attribuer une probabilité de victoire en jouant un coup donné dans une situation donnée (*figure 11*). L'aboutissement de ce programme sera atteint en 1959 puisque, cette année-là, Arthur Samuel réalise un programme capable de jouer aux dames et d'apprendre en ajustant ses coefficients au fil de l'expérience. Sa consécration aura lieu en 1962 puisque ce programme vaincra le champion du Connecticut [131]. Il s'agira de la première fois qu'un programme est capable de concurrencer l'intellect humain. En raison de ces travaux, Arthur Samuel est aujourd'hui considéré comme l'un des pères fondateurs de l'apprentissage automatique.



Figure 11. Arthur Samuel jouant aux dames avec un ordinateur IBM (© IBM) [129]

Parallèlement, en 1954, Allen Newell a pour projet de développer un programme de jeu d'échecs, plus complexe encore à programmer que le jeu des dames [132]. Pour atteindre son objectif, il s'entourera de John Clifford Shaw, Herbert Simon et de psychologues, dont Adriaan de Groot qui a étudié de grands joueurs d'échecs. Cette collaboration aboutira en 1956 à la création du langage IPL (*Information Processing Language*), intégrant comme fonctionnalité la manipulation des informations symboliques^{xviii} [134]. Ce langage a notamment initié le concept de traitement de listes et peut être considéré comme le prédécesseur d'un langage qui sera grandement employé dans la recherche en intelligence artificielle : le langage LISP (*list processing*), créé par John McCarthy et dont la première version date de 1958 [135].

Allen Newell, John Cliff Shaw et Herbert Simon vont également écrire un programme informatique entre 1955 et 1956 nommé le *Logic Theorist* [136]. Il s'agira là du premier programme de démonstration en logique des propositions^{xix}, considéré ainsi comme étant le premier programme d'intelligence artificielle [127]. Celui-ci est capable de résoudre des problèmes en employant des raisonnements similaires à ceux d'un être humain. Il a démontré ces capacités en prouvant 38 des 52 théorèmes des *Principia Mathematica*^{xx} [139]. Ce programme fut présenté lors de la conférence de Dartmouth, une conférence considérée par beaucoup comme étant le lieu de naissance de l'intelligence artificielle.

Initiée par des chercheurs en mathématiques et en informatique (John McCarthy, Marvin Minsky, Nathaniel Rochester et Claude Shannon), c'est lors de l'été 1956 que cet atelier

^{xviii} En programmation informatique, un symbole est une étiquette apposée sur certains éléments du code objet ou d'un arbre syntaxique abstrait, permettant de les identifier sous cette forme transformée du code source [133].

^{xix} La logique des propositions est un langage formel constitué d'une syntaxe et d'une sémantique. La syntaxe décrit l'ensemble des formules qui appartiennent au langage. La sémantique permet de donner un sens aux formules du langage [137].

^{xx} Les *Principia Mathematica* sont une œuvre d'Alfred North Whitehead et Bertrand Russell publiés en trois volumes entre 1910 et 1913. Cette œuvre a pour sujet les fondements des mathématiques [138].

scientifique a lieu à Dartmouth aux Etats-Unis [140]. Durant huit semaines, cette réunion de travail réunit 20 participants issus de différentes disciplines, dont quatre participants majeurs des conférences de Macy (Warren McCulloch, Julian Bigelow, Claude Shannon et Ross Ashby). C'est lors de cette conférence que John McCarthy convainc l'auditoire d'adopter l'expression « intelligence artificielle » pour désigner une machine dotée d'un « esprit » semblable à l'humain. La lumière est ainsi projetée vers les sciences cognitives qui sont alors perçues comme le carrefour de rencontre entre la psychologie expérimentale et les modèles informatiques algorithmiques, ceci dans le but de décrire avec précision et rigueur le fonctionnement de l'esprit humain : « *penser comme les humains* » [77]. John McCarthy prononcera également en amont de cette conférence une phrase qui cristallisera cette approche cognitive : « *Toute activité intellectuelle peut être décrite avec suffisamment de précision pour être simulée par une machine* » [141]. C'était une vision avant-gardiste et très ambitieuse car celle-ci implique d'explorer l'activité cérébrale à une époque où les connaissances sur ce sujet sont très limitées. Quoiqu'il en soit la discipline était officiellement née, une discipline réunissant l'informatique, l'électronique et les sciences cognitives, avec ses références bibliographiques et comme toute discipline qui s'affirme ses controverses. De nos jours, les sciences cognitives et les intelligences artificielles ne sont plus confondues et sont considérées comme étant des disciplines distinctes tant les progrès réalisés dans ces deux domaines sont importants.

John Mc Carthy présente également, à la conférence de Dartmouth, le principe d'élagage alphabète, un algorithme d'évaluation. C'est en 1958 que ce principe est mis en pratique pour la première fois par NSS (*Newell-Shaw-Simon*), un programme de jeu d'échecs présentant la même structure que le *Logic Theorist* en mettant en œuvre la notion d'heuristique [142]. Les progrès sont tels qu'Allen Newell et Herbert Simon estiment qu'un programme atteindra le niveau d'un champion d'échecs avant 1968 [143]. Cela ne se concrétisera en réalité que beaucoup plus tard en 1997, néanmoins cette affirmation témoigne de l'enthousiasme et de l'émulation suscités par ces travaux.

En 1957, Frank Rosenblatt, psychologue, crée le premier perceptron, un neurone artificiel capable de classer des données de façon autonome. Son modèle d'apprentissage prend la forme d'un logiciel conçu tout d'abord sur un *IBM 704* puis sur une machine destinée à la reconnaissance d'image, le *Mark I* [144]. Ce perceptron simple, bien que limité en possibilités de calcul, signe la naissance de l'intelligence artificielle connexionniste et ouvrira plus tard la voie à l'apprentissage profond tel que nous le connaissons aujourd'hui.

En 1958, Hans Peter Luhn publie un article intitulé *A Business Intelligence System* dans lequel il décrit sa théorie d'un système d'informatique décisionnelle (ou *business intelligence*). Ce système a pour but d'automatiser le traitement et la diffusion d'informations à partir des données qui lui sont fournies [145].

Dans la continuité du *Logic Theorist*, Herbert Simon, Cliff Shaw et Allen Newell créent en 1959 le *General Problem Solver*, un programme informatique développé en langage IPL et visant à résoudre tous types de problèmes. Pour concrétiser ce projet, ils ont tenté de comparer les étapes de raisonnement du programme à celles de sujets humains face aux mêmes problèmes à résoudre. Il s'agit du premier programme à différencier sa base de données et sa stratégie de résolution de problèmes, il fonctionne sur un principe d'analyse moyens-fins en analysant les différences entre les situations et en construisant des objectifs et sous-objectifs [146].

En 1960, Herbert Gelernter développe le programme *geometry machine* capable de démontrer des théorèmes en géométrie en utilisant des raisonnements plus ou moins rigoureux [147].

En 1961, Bert Green et son équipe s'intéressent à un programme nommé *Baseball* (conçu par Fredrick Frick, Oliver Selfridge et Gerald Dineen). Il s'agit d'un programme en langage IPL capable de répondre, à partir d'une base de données, à des questions posées en anglais. Le programme fonctionne en deux temps : premièrement, la partie linguistique du programme interprète la question à partir d'une carte perforée en analysant son contenu et sa syntaxe et en identifiant les informations contenues dans la question ainsi que les données à mobiliser pour y répondre (*figure 12*). Ensuite, le processeur se charge de rechercher les informations correspondantes dans la base de données. Après application des traitements nécessaires, la réponse est imprimée. Ce programme était spécialisé dans l'apport de réponses concernant les matchs de Baseball mais la technique employée laissait entrevoir la possibilité de développer des programmes de traduction automatique [148].

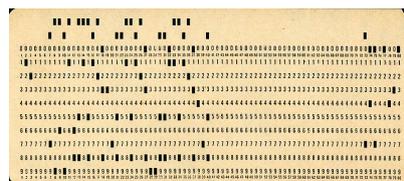


Figure 12. Carte perforée (*Reproduit selon les termes de la licence CC BY 2.0. Auteur : Pete Birkinshaw, <https://flickr.com/photos/93001633@N00/5151286161>*)

Lofti Zadeh formalise en 1965 la logique floue, une technique largement utilisée en intelligence artificielle et de façon plus générale en informatique. Contrairement à la logique booléenne qui admet deux possibilités pour une proposition (0 pour fausse, 1 pour vraie), la logique floue propose d'attribuer à une proposition un état pouvant prendre une valeur comprise entre 0 et 1 [149].

L'année 1965 marque un tournant pour l'intelligence artificielle car c'est en cette année qu'est créé le premier système expert, une forme d'intelligence artificielle qui sera la figure de proue de nombreux chercheurs du domaine. Ce sont les informaticiens Edward Feigenbaum et Bruce Buchanan, le médecin Joshua Lederberg et le chimiste Carl Djerassi qui sont à l'origine de ce programme nommé Dendral (acronyme de *Dendritic Algorithm*). Celui-ci est capable d'identifier les structures chimiques d'un composé à partir de données telles que la spectrométrie de masse et la spectroscopie de résonance magnétique nucléaire (RMN). Dendral, de par ses performances, va être à l'origine d'une vague de systèmes experts [150].

Entre 1964 et 1966, Joseph Weizenbaum réalise l'exploit d'écrire seul le programme informatique *ELIZA*, le premier *chatbot* fonctionnel de l'histoire. Celui-ci tente de simuler un psychologue en entretenant un dialogue avec l'utilisateur. *ELIZA* fonctionne en reconnaissant les mots et expressions clés présents dans la phrase de l'utilisateur puis en construisant une réponse incluant plus ou moins ces mots et expressions (*figure 13*) [151].

```
Welcome to
                EEEEE LL    IIII ZZZZZZ  AAAAA
                EE   LL    II     ZZ   AA  AA
                EEEEE LL    II     ZZZ  AAAAAA
                EE   LL    II     ZZ   AA  AA
                EEEEE LLLLL IIII  ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

Figure 13. Exemple de conversation avec *ELIZA*

La pertinence des réponses formulées par le programme est très variable selon les termes et la syntaxe choisis par l'utilisateur. Durant certaines conversations, *ELIZA* parvenait à être suffisamment convaincant pour réussir le test de Turing. Bien que « *Je comprends* » fasse partie de ses réponses éventuelles, *ELIZA* n'a pas la capacité de comprendre le sens des phrases qu'il traite et c'est à ce niveau que résident ses limites [152]. Malgré cela, ce programme représente à cette époque un pas de géant dans l'histoire du traitement du langage naturel.

En 1967, Richard Greenblatt conçoit le premier programme d'échecs suffisamment performant pour vaincre un joueur de niveau moyen [153].

En 1969 se tient la toute première IJCAI (*International Joint Conference on Artificial Intelligence*) durant laquelle sont présentées les recherches les plus plébiscitées portant sur les intelligences artificielles [154]. Celle-ci se tiendra tous les deux ans entre 1969 et 2015 puis tous les ans [155].

En 1970, Edgar Frank Codd dévoile ses travaux relatifs au modèle relationnel des bases de données. En se basant sur la théorie mathématique des ensembles, ce modèle permet d'attribuer aux données des relations entre elles de façon logique et mathématique en structurant celles-ci sous la forme de matrices. Ce modèle se révélera bien plus tard révolutionnaire pour les systèmes de gestion de bases de données et par extension pour les intelligences artificielles [156].

Toujours en 1970, Seppo Linnainmaa met au point la rétropropagation de gradient qu'il décrit dans son mémoire de maîtrise [157]. Il s'agit d'un aspect fondamental de l'entraînement des réseaux de neurones tel qu'il sera pratiqué jusqu'à nos jours. Celle-ci opère en modifiant les poids (ou coefficients) associés à chaque synapse en commençant par la dernière couche puis en parcourant l'ensemble du réseau de neurones jusqu'à sa première couche (*figure 14*). La modification des poids s'effectue selon les erreurs mesurées, c'est-à-dire que plus une synapse est responsable d'une erreur au niveau de la couche de sortie, plus le poids associé à cette synapse est susceptible d'être modifié en conséquence. Ce concept sera ensuite repris et développé par plusieurs pionniers en matière d'apprentissage automatique tels que Paul Werbos par exemple qui, dès 1974, communiquera à plusieurs reprises sur le sujet [158]. Cette découverte représente une avancée décisive pour l'approche connexionniste des intelligences artificielles.

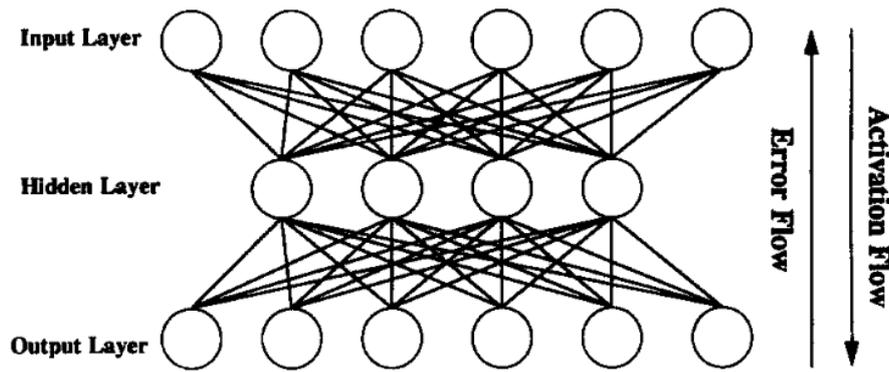


Figure 14. Représentation d'une rétropropagation s'opérant sur un réseau de neurones. Le flux d'activation propage les données d'entraînement de la couche d'entrée vers la couche de sortie. Le flux d'erreur se propage ensuite à son tour de la couche de sortie vers la couche d'entrée. (Reproduit avec la permission de l'éditeur Elsevier. Auteur : Barry J. Wythoff.) [159]

L'approche symbolique, sa rivale, n'est pas en reste durant les années 70 puisque les systèmes experts vont y avoir le vent en poupe [160]. Les chercheurs du secteur émettent l'hypothèse que les programmes d'intelligence artificielle seraient significativement plus performants en étant dotés d'une connaissance approfondie du domaine qu'ils traitent. Leur objectif va alors être d'extraire ces connaissances, de les formaliser en règles puis de les programmer. Cependant, plusieurs questions se posent : Comment sélectionner les connaissances à extraire ? Comment représenter celles-ci ? Comment les programmer ? Ces idées et ces questionnements vont, au fil des avancées, mener à la création de systèmes basés sur la connaissance, les systèmes experts. Dans le même temps, de nombreux chercheurs défendent la thèse selon laquelle la perception et la réflexion sont facilitées par des structures cognitives complexes regroupant et organisant les informations de façon rationnelle. L'un des arguments principaux à cette thèse est que la mémoire humaine contiendrait plus d'informations que celle-ci ne pourrait gérer sans avoir recours à une forme ou une autre de structuration. Dans une volonté d'appliquer ces théories, le traitement du langage naturel va alors introduire différentes techniques de programmation et méthodes de représentation des connaissances. Le traitement du langage naturel va faire un bond en avant grâce notamment aux réseaux sémantiques qui fleurissent durant cette période [161]. Ceux-ci, nommés ainsi par Ross Quillian en 1968, ont pour objectif de représenter graphiquement des connaissances telles qu'elles pourraient être organisées dans la mémoire humaine (*figure 15*) [162].

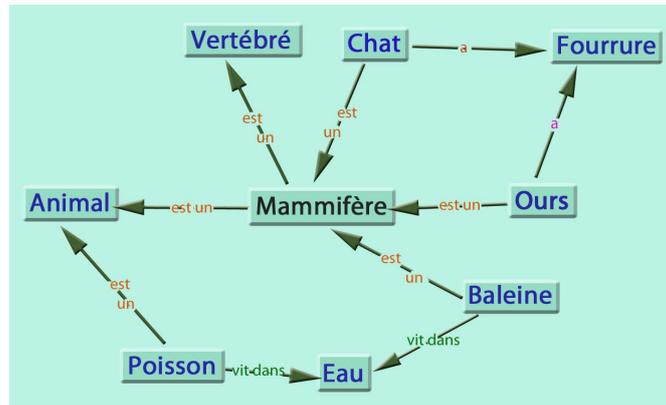


Figure 15. Exemple de représentation d'un réseau sémantique (Reproduit selon les termes de la licence CC BY-SA 3.0. Auteur : Lamiot. https://commons.wikimedia.org/wiki/File:SemanticNetArbre_s%C3%A9mantique_fr.jpg)

Marvin Minsky propose en 1974 le concept de cadre qui est un dérivé du réseau sémantique. En intelligence artificielle, un cadre désigne une structure de données visant à subdiviser les connaissances en sous-structures (*figure 16*). Contrairement aux réseaux sémantiques, les représentations sous forme de cadre permettent de préciser des exceptions donc d'être à la fois plus flexibles et plus en adéquation avec la réalité [163].

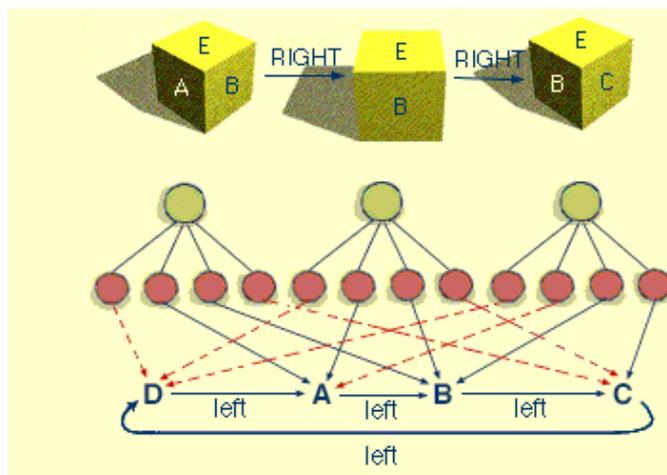


Figure 16. Exemple de cadre (Reproduit selon les termes de la licence CC BY-NC 4.0. Auteur : Marvin Minsky.) [163]

Dès 1975, Roger Schank et Robert Abelson vont mettre la lumière sur un type de structure qui, à l'instar des cadres, va constituer une avancée significative dans le monde des intelligences artificielles : les scripts. Un script se définit par un ensemble de séquences d'évènements associés à un contexte. Essentiellement, il s'agit d'une structure de connaissances caractérisant des évènements et des actions nécessaires, auxiliaires et optionnelles dans le cadre d'une situation définie et le plus souvent d'un point de vue (*figure 17*). En tant que structure, un script est composé d'emplacements (ou *slots*) et de règles conditionnant les contenus de ces emplacements. Ces derniers sont interconnectés, ce qui implique que le contenu des uns peut influencer le contenu des autres [164].

<p>Script: Restaurant Entry Conditions: Customer is hungry. Customer has money. Scenes:</p> <p>1. Entering Customer goes into restaurant. (E₁) Customer looks around. (E₂) Customer decides where to sit. (E₃) Customer goes to a table and sits down. (E₄)</p> <p>2. Ordering Customer picks up a menu. (E₅) Customer decides on food. (E₆) Customer orders food from waiter. (E₇) Waiter tells cook the order. (E₈) Cook prepares food. (E₉)</p> <p>3. Eating Cook gives food to waiter. (E₁₀) Waiter gives food to customer. (E₁₁) Customer eats food. (E₁₂)</p> <p>4. Exiting Waiter writes out check. (E₁₃) Waiter brings check to customer. (E₁₄) Customer gives tip to waiter. (E₁₅) Customer goes to cash register. (E₁₆) Customer gives money to cashier. (E₁₇) Customer leaves restaurant. (E₁₈)</p>

Figure 17. Exemple de script situationnel : le script du restaurant (Reproduit avec la permission de l'auteur. Sabah Al-Fedaghi.) [165]

Concomitamment à ces avancées, Terry Winograd met au point en 1970 la première version de *SHRDLU*, un programme informatique capable de simuler la manipulation de blocs géométriques à partir d'instructions données en anglais [166]. Par le biais d'un dialogue interactif, l'utilisateur peut demander à *SHRDLU* de déplacer ces blocs ou lui poser des questions sur ses manipulations antérieures (*figure 18*). A l'aide de la mémoire qui lui est pourvue, *SHRDLU* peut y conserver ses actions passées ainsi que leurs résultats et ainsi, dans une certaine mesure, interpréter la demande de l'utilisateur lorsque celle-ci n'est pas entièrement explicite. Par exemple, si l'utilisateur lui demande de retirer un cube sans préciser lequel, *SHRDLU* considérera dans la plupart des cas qu'il s'agit du dernier cube manipulé. *SHRDLU* peut aussi, grâce à sa mémoire et aux règles dont il est pourvu, identifier si une action est possible ou non, ou demander à l'utilisateur de reformuler sa demande si celle-ci lui paraît incongrue. C'est ici l'approche « *agir rationnellement* » qui est mise à l'épreuve, il est attendu de *SHRDLU* qu'il présente les qualités d'un agent rationnel [77]. Un agent rationnel doit être en mesure de réagir à un stimulus (ici une requête sous forme de texte) et de procéder aux actions les plus efficaces compte tenu des informations dont il dispose, en vue d'atteindre ou de s'approcher le plus possible du résultat escompté, et ce de façon autonome (*figure 19*). Cette approche repose non seulement sur la qualité des inférences produites par l'agent mais également sur ses raisonnements ainsi que sur la représentation des connaissances dont il dispose [167]. *SHRDLU* est considéré par certains comme étant le premier programme capable de véritablement comprendre et interpréter le langage naturel [168].

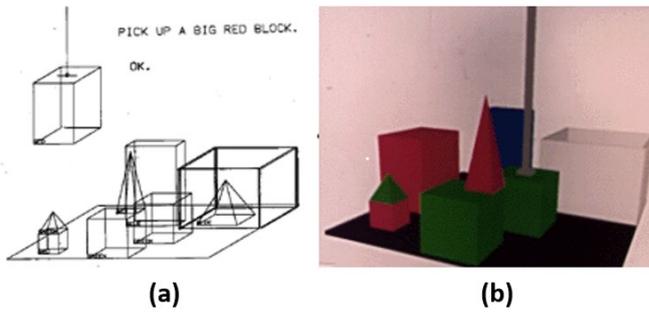


Figure 19. Affichage de SHRDLU : (a) version originale ; (b) version ultérieure intégrant des textures colorées [166]

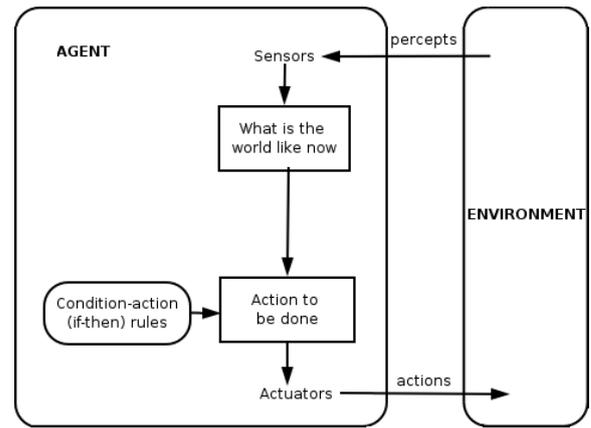


Figure 18. Schéma d'un agent rationnel simple

En 1972 vont naître deux langages de programmation notables et dont les apports seront essentiels pour le développement des intelligences artificielles. Le premier est le langage Prolog, acronyme de « Programmation en logique », créé par Alain Colmerauer et Philippe Roussel. Comme son nom l'indique, il s'agit d'un langage de programmation logique, c'est-à-dire s'appuyant sur des faits élémentaires, des règles de logique associant des conséquences à ces faits, et un moteur d'inférences. Le langage Prolog est encore aujourd'hui considéré comme l'une des principales innovations en informatique de ces 50 dernières années [169]. Le second est le langage *Smalltalk*, créé par Alan Kay, Dan Ingalls, Ted Kaehler et Adele Goldberg. Il s'agit d'un langage de programmation innovant pour l'époque en raison de plusieurs des concepts qu'il concrétise comme notamment celui de la programmation orientée objet. Celle-ci désigne un modèle de langage de programmation qui, au lieu de se focaliser uniquement sur la logique, permet de définir des données en tant qu'objets manipulables. La nature et la hiérarchie de ces objets peuvent alors être organisées selon des classes et des sous-classes [170]. Inspiré des langages LISP et *Simula*, *Smalltalk* constitue un apport fondamental pour les futurs langages de programmation puisqu'il en inspirera à son tour plusieurs actuellement en vogue tels que *Java* et *Ruby*.

Ces différentes prouesses ont assurément fait grandir l'enthousiasme autour des intelligences artificielles, mais ces nouvelles technologies vont aussi susciter de vives critiques et être confrontées à des difficultés techniques qui feront traverser à l'intelligence artificielle son premier hiver [171].

3. Premier hiver des intelligences artificielles (1974 - 1980)

Durant les années 60, l'optimisme vis-à-vis des intelligences artificielles est tel que les chercheurs aspirent à tendre vers des programmes doués d'une pensée rationnelle. Il s'agit de l'approche « *penser rationnellement* » qui vise à codifier la pensée et se veut comme une mise en pratique des écoles grecques philosophiques de la logique et de leurs héritages [77]. Seulement, deux problèmes majeurs se posent. Tout d'abord, les notations logiques ne permettent pas d'exprimer l'entièreté de la connaissance, en particulier lorsqu'il s'agit de connaissances informelles ou incertaines. Le deuxième obstacle se dressant devant cette approche est la limite technologique. Plus le nombre de règles et de faits dont le programme doit tenir compte est élevé, plus ce programme sera exigeant en ressources de calcul, en particulier si aucune directive n'est donnée au programme concernant les raisonnements à prioriser. En outre, la vision par ordinateur et le traitement automatique du langage naturel requièrent plus de données qu'il n'est possible de traiter pour les ordinateurs de l'époque [172]. Ces problématiques vont prendre du relief avec les critiques émises de part et d'autre.

Dès 1965, Hubert Dreyfus, enseignant au MIT (*Massachusetts Institute of Technology*), pointe du doigt le fait que les estimations exprimées par les chercheurs quant aux progrès de l'intelligence artificielle se sont avérées pour beaucoup erronées. L'approche symbolique est, selon lui, à remettre en question [173].

En 1969, Marvin Minsky et Seymour Papert démontrent les limites du perceptron dans leur livre intitulé *Perceptrons*. Ils soulignent également le décalage entre ses capacités réelles et les prédictions formulées par Frank Rosenblatt, créateur du perceptron [174]. L'acuité de la critique est telle que celle-ci va entraîner une traversée du désert pour le connexionnisme, aucune recherche ne sera menée dans ce domaine durant dix ans.

Le mathématicien James Lighthill publie en 1973 l'article intitulé *Intelligence artificielle : une enquête globale*, plus communément appelé *Le rapport de Lighthill*. L'objectif de ce rapport est d'évaluer la recherche académique en intelligence artificielle. Celui-ci cristallise les déceptions relatives au manque d'avancées exploitables, en particulier en robotique et en traitement du langage naturel [175].

En 1976, Joseph Weizenbaum, lui aussi enseignant au MIT et créateur du programme *ELIZA*, publie un ouvrage nommé *Puissance informatique et raison humaine* dans lequel il expose son

point de vue concernant les conséquences néfastes que peut entraîner un usage à mauvais escient de l'intelligence artificielle [176].

John Searle publie en 1980 un article dans lequel il expose sa critique du test de Turing qu'il étaye par l'argument de la chambre chinoise. Celui-ci met en scène deux interlocuteurs, l'un à l'intérieur d'une chambre ne disposant pas de compétences linguistiques en chinois (incarné par John Searle), l'autre à l'extérieur maîtrisant le chinois. Bien qu'elle ne sache pas parler chinois, la personne à l'intérieur de la pièce est en mesure d'entretenir une conversation dans cette langue avec la personne à l'extérieur en suivant des instructions précises, donnant ainsi l'impression d'être sinophone. La personne à l'intérieur serait, dans cette configuration, en mesure de réussir le test de Turing alors que pourtant elle serait incapable d'entretenir une conversation en chinois sans avoir accès aux instructions ou si la conversation s'étend au-delà de celles-ci. Le test de Turing est donc, selon John Searle, insuffisant pour qualifier un programme d'intelligent car incapable d'évaluer la compréhension et la capacité de raisonnement de ce programme. C'est ainsi qu'est établi le qualificatif « faible » pour désigner les intelligences artificielles [177].

La DARPA (*Defense Advanced Research Projects Agency*), déçue par les résultats en deçà de ce qui était attendu, rédige un rapport qui rejoint celui de James Lighthill et décide d'interrompre leurs financements de recherche en intelligence artificielle [178]. Il en sera de même pour la NRC (*Nuclear Regulatory Commission*) et le gouvernement britannique [127, 179]. Tout cela va provoquer le premier hiver de l'intelligence artificielle durant lequel les financements alloués à ce domaine se feront plus rares. Durant cette période, l'intelligence artificielle s'industrialise mais évolue peu du point de vue de la recherche, il s'agit donc non seulement d'un hiver économique mais surtout d'un hiver scientifique [171].

4. Ascension des systèmes experts et premiers réseaux de neurones (années 1980)

Cet hiver ne durera pas puisque l'intérêt autour des intelligences artificielles fait son retour dès les années 80. Ce sont dans un premier temps des projets japonais qui seront à l'origine de ce vent nouveau. Durant cette période, le ministère japonais de l'Économie, du Commerce et de l'Industrie nage à contre-courant en attribuant des financements importants à la recherche sur l'intelligence artificielle. Ils attribuent notamment 850 millions de dollars au projet des ordinateurs de cinquième génération qui débute en 1982. Ce projet, réalisé essentiellement en

langage Prolog, a pour objectif de concevoir des ordinateurs capables d'atteindre les performances d'un superordinateur en utilisant plusieurs processeurs en même temps et de faire usage à la fois d'une base de données massive et de programmation logique pour traduire, entretenir une conversation, reconnaître le contenu d'une image et raisonner. Il s'agit d'un projet révolutionnaire et en avance sur son temps qui n'atteindra pas ses objectifs à la fin des années 80 [180]. Cependant, en raison de la réputation forgée par l'industrie japonaise durant les années 70, notamment dans les secteurs de l'électronique et de l'automobile, ce financement massif entraîne une remise en question de la part des gouvernements et financeurs. En réponse au projet financé par le gouvernement japonais, la MCC (*Microelectronics and Computer Technology Corporation*) est fondée en 1982 par un groupement d'entreprises américaines afin de financer des projets d'ampleur en intelligence artificielle [181]. Peu de temps après, la DARPA fait machine arrière en investissant à nouveau dans l'intelligence artificielle et en fondant la *Strategic Computing Initiative*, un programme gouvernemental prévoyant de renforcer la recherche en intelligence artificielle entre 1983 et 1993 [182]. A l'instar des Etats-Unis, le Royaume-Uni investit 350 millions de livres dans un programme similaire entre 1983 et 1987, le programme *Alvey* [183].

Les obstacles rencontrés durant les années 70 ont amené les chercheurs à penser que les comportements intelligents étaient dépendants du traitement de la connaissance [184]. C'est pourquoi l'un des domaines les plus en vogue dans la recherche en intelligence artificielle des années 80 est l'ingénierie des connaissances. Cela a fait la part belle aux systèmes experts qui avaient déjà commencé à s'affirmer durant les années 70. L'exemple le plus réputé est le programme *MYCIN*, dont le développement a démarré au début des années 70 et a duré six ans. Il s'agit d'un système expert d'aide au diagnostic capable, à partir d'une symptomatologie et de résultats biologiques, d'identifier des bactéries responsables d'une infection et de recommander un traitement avec une posologie adaptée au poids du/de la patient(e). *MYCIN* est capable d'expliquer, à la demande, le raisonnement l'ayant amené à son diagnostic et à ses recommandations. Le programme fonctionne à partir d'un moteur d'inférence et d'une base de connaissances constituée de 600 règles. Pour des raisons légales, éthiques et surtout du fait du parc technologique encore restreint de l'époque, *MYCIN* ne sera jamais employé en pratique [185]. Malgré cela, *MYCIN* a ouvert la voie aux systèmes experts qui furent ensuite nombreux à suivre. Nous pouvons par exemple citer *Xcon* (*eXpert CONfigurer*) en informatique, *Hearsay-II* en linguistique ou encore *PROSPECTOR* en géologie [186,187,188]. Les systèmes experts amènent le secteur de l'intelligence artificielle à s'industrialiser et permet ainsi à celle-ci de recouvrer de la crédibilité et de l'audience.

En 1984, Douglas Lenat démarre le projet Cyc dont le nom est dérivé du mot « *Encyclopedia* ». Comme sa dénomination l'indique, il s'agit d'une base de connaissance générale ayant pour objectif de doter l'intelligence artificielle d'une culture générale similaire à celle d'un humain. Douglas Lenat émet l'opinion que, pour qu'une machine puisse comprendre les concepts humains, il est nécessaire de lui apprendre chacun de ces concepts individuellement. Ce projet s'appuie notamment sur les ontologies, que nous développerons plus tard dans ce manuscrit (*voir section P1.III.1.c.*) [189]. Cyc continuera d'être développé et deviendra plus tard *ResearchCyc* puis enfin *OpenCyc* accessible au grand public dès 2001 [190]. Sa dernière version à ce jour, la 4.0, regroupe 239 000 concepts et 2 093 000 faits [191].

Les bases de données de séquences génétiques s'enrichissent également durant les années 80 et suscitent le besoin de développer des programmes capables de détecter des motifs. Cette démarche va donner naissance à la bio-informatique qui va grandement bénéficier des moteurs d'apprentissage à base de règles, des méthodes statistiques usuelles et de la programmation logique inductive. Cette dernière est à l'intersection de la programmation logique et de l'apprentissage automatique. En fouille de données, celle-ci est en mesure de rechercher des régularités et d'établir des règles générales à partir de connaissances spécifiques [192].

L'intelligence artificielle connexionniste, dont la recherche était à l'arrêt durant les années 70, est à nouveau mise en avant suite aux travaux du physicien John Hopfield. En 1982, il découvre un type de réseau neuronal capable d'apprendre et traiter l'information suivant le modèle de la mémoire associative, appelé aujourd'hui *réseau de Hopfield* (*figure 20*) [193]. La rétropropagation du gradient de l'erreur est remise sur le devant de la scène par David Parker en 1985 puis par Yann LeCun ainsi que par David Rumelhart, Geoffrey Hinton et Ronald Williams en 1986 [194,195,196]. Bien que découverte par Paul Werbos 10 ans plus tôt, ce furent les travaux et publications des chercheurs sus-cités qui rendirent cette méthode populaire et opérationnelle. Celle-ci a permis notamment la mise au point des perceptrons multicouches, un type de réseau de neurones artificiel (*figure 21*). Les perceptrons étaient dès lors, grâce à la rétropropagation du gradient, capables de traiter des problèmes non linéaires.

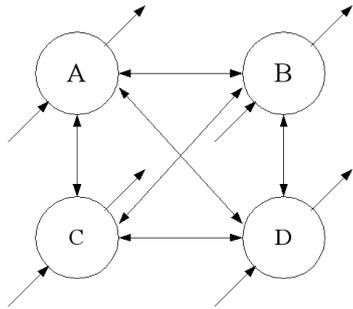


Figure 20. Réseau de Hopfield à quatre neurones (Reproduit selon les termes de la licence CC BY-SA 3.0. Auteur : Zeno Gantner. https://commons.wikimedia.org/wiki/File:Hopfield_net.png)

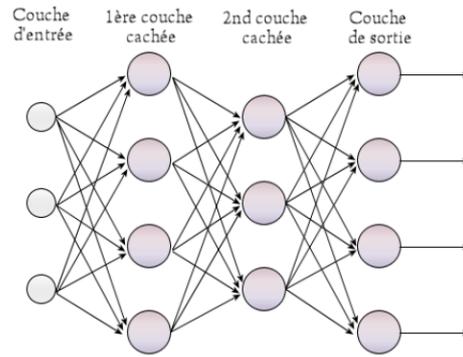


Figure 21. Exemple de perceptron multicouche

En 1988, Judea Pearl initie les réseaux bayésiens, des outils permettant de modéliser des incertitudes en se fondant sur la théorie des probabilités. Appliqués à la prise de décision, les réseaux bayésiens permettent de choisir, parmi plusieurs hypothèses, la plus performante (figure 22). Ceux-ci sont encore aujourd'hui très populaires en intelligence artificielle [197].

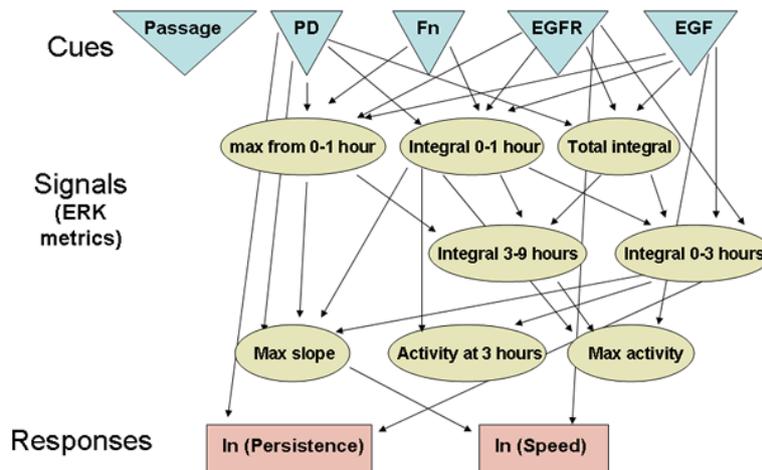


Figure 22. Réseau bayésien modélisant des voies de signalisation cellulaire (Reproduit selon les termes de la licence CC BY-SA 4.0. Auteurs : Sachs K. et coll. https://commons.wikimedia.org/wiki/File:Bayesian_network.png)

En 1989, David Goldberg popularise les algorithmes génétiques dans son ouvrage *Genetic Algorithms in Search, Optimization, and Machine Learning*. Les algorithmes génétiques appartiennent à la famille des algorithmes évolutionnistes, il s'agit de méthodes de calcul bioinspirées puisqu'elles simulent le processus de sélection naturelle pour résoudre un problème. En suivant le principe de sélection naturelle telle qu'il s'applique à la génétique, le principe des algorithmes génétiques est de mettre en concurrence des solutions pour un problème donné puis, suite à une itération de processus aléatoires, d'identifier les solutions optimales [198].

5. Second hiver des intelligences artificielles (1987 - 1993)

Entre la fin des années 80 et le début des années 90, l'intelligence artificielle vit son second hiver, cette fois-ci essentiellement économique. En effet, malgré la perte temporaire d'intérêt de la part des investisseurs et des agences gouvernementales, la recherche sur l'intelligence artificielle parvient à se maintenir [171].

Cette période difficile est due, une nouvelle fois, à la déception consécutive aux attentes trop élevées suscitées par l'enthousiasme autour des systèmes experts. Les premiers systèmes experts développés, pour certains ne trouvent pas leur place dans l'écosystème professionnel pour lequel ils sont conçus (ex. : *MYCIN*), pour d'autres rencontrent un succès commercial mais coûtent trop chers en maintenance (ex. : *Xcon*) [127]. De plus, leurs défauts classiques, à savoir leur incapacité à apprendre ainsi que leur difficulté à s'adapter aux situations inhabituelles, se révèlent et contribuent à ralentir les systèmes experts dans leur course. Ce phénomène est accentué par l'échec partiel du projet d'ordinateurs de cinquième génération lancé au Japon qui, trop ambitieux pour l'époque, n'atteint pas l'ensemble de ses objectifs [199].

Dans le même temps, le marché de l'informatique spécialisé dans l'intelligence artificielle (ex. : *machines Lisp* de Symbolics) s'effondre dès 1987 face aux progrès technologiques véloces des ordinateurs grand public. La vitesse et la puissance atteintes par les ordinateurs de bureau (ex. : *Macintosh II* d'Apple, *Personal System 2* d'IBM), qui-plus-est en étant moins coûteux que les *machines Lisp*, rendent ceux-ci plus intéressants aux yeux des consommateurs et des industriels [178].

Devant ce déclin de popularité, la *Strategic Computing Initiative* de la DARPA, dont la direction a changé depuis sa création, fait le choix de réduire les fonds alloués à la recherche sur l'intelligence artificielle et de les rediriger vers des projets à plus court terme et moins risqués [199].

La recherche en intelligence artificielle traverse alors une période de remise en question et une théorie émerge en particulier : celle de l'*embodiment*. De nombreux chercheurs, tels que Rodney Brooks et Hans Moravec, émettent l'hypothèse que l'intelligence d'une machine serait conditionnée par ses facultés sensori-motrices. Selon cette théorie, le rôle du raisonnement abstrait dans l'intelligence est d'une importance moindre que celui acquis par l'expérience des sens, pour acquérir une intelligence véritable une machine doit être capable de percevoir, se

déplacer, évoluer dans son environnement et enfin avoir conscience de son propre corps [200,201]. En renouant avec la cybernétique et en conviant la robotique à la fête, cette approche de l'intelligence artificielle contribuera grandement à dynamiser les travaux sur la vision par ordinateur.

6. Le développement fulgurant de l'apprentissage automatique (années 1990 à nos jours)

Bien que le monde des intelligences artificielles fut marqué par un court hiver au début des années 90, le long printemps qui s'en suivra verra plus que jamais fleurir le connexionnisme. L'augmentation de la puissance des ordinateurs et de la quantité des données traitées vont permettre à l'intelligence artificielle d'accéder à des niveaux de performance encore jamais atteints et ainsi d'accomplir certains de ses objectifs les plus anciens [202]. En outre, la recherche en intelligence artificielle va acquérir de la rigueur méthodologique de par le rapprochement que celle-ci va opérer avec le domaine des statistiques. Stuart Russell et Peter Norvig, dans leur deuxième édition de leur ouvrage *Intelligence artificielle : une approche moderne*, qualifieront cette nouvelle approche de révolutionnaire [56]. De grandes avancées vont être faites en planification, analyse de données, traitement du langage naturel, reconnaissance visuelle et sonore, et bien d'autres [203,204]. Cette consécration s'accompagnera en revanche d'une augmentation de la complexité des algorithmes et des modèles, avec notamment les réseaux de neurones dont l'explicabilité se révélera de plus en plus hors de portée [205]. De plus, les fantasmes d'une intelligence artificielle forte entretenus par les œuvres de la culture populaire affectent la réputation de cette discipline. En comparaison, l'intelligence artificielle faible fait pâle figure selon certains, par conséquent de nombreux chercheurs éviteront d'employer les termes « intelligence artificielle » dans le cadre de leurs projets afin de faciliter l'obtention de financements. C'est ce qui entraînera, conjointement à la complexité grandissante, l'émergence de sous-disciplines de l'intelligence artificielle propres à des secteurs spécifiques, tels que la robotique intelligente par exemple. Sans pour autant s'annoncer par son nom, l'intelligence artificielle va donc s'ancrer dans les pratiques industrielles, professionnelles, judiciaires, personnelles et même artistiques [206].

En 1994, le Laboratoire d'informatique de Paris-VI conçoit une installation vidéonumérique affichant un écosystème virtuel en deux dimensions composé de diverses formes évoluant avec le temps et interagissant entre-elles. Les couleurs et tailles de ces formes évoluent continuellement selon des données temporelles et météorologiques que l'écosystème reçoit par

modem. Ce projet à la croisée de l'intelligence artificielle et de l'art, nommé *Jardins des hasards*, inspirera plus tard de nombreuses architectures d'intelligence artificielle à but artistique (*figure 23*) [207].

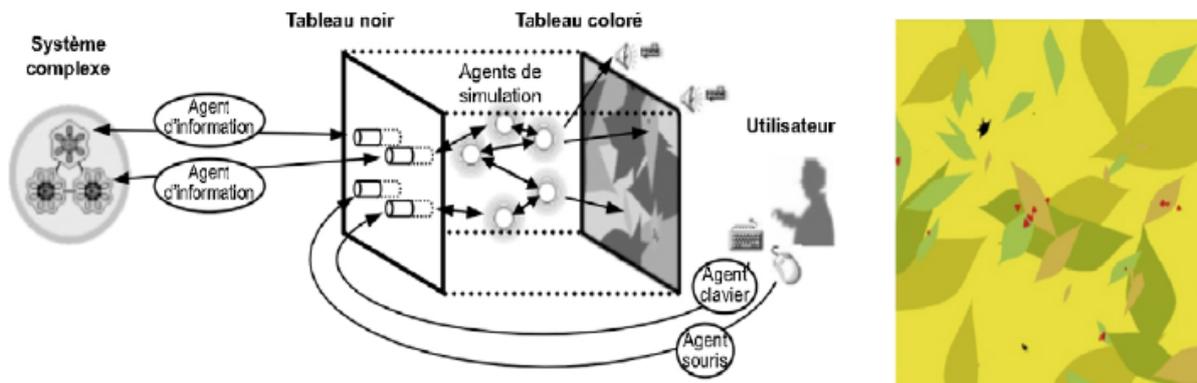


Figure 23. Architecture de la plate-forme de conception du *Jardin des hasards* (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteur : Jean-Paul Fourmentraux.) [207]

A partir de 1985, le laboratoire de robotique de l'université Carnegie-Mellon travaille sur l'automatisation de la conduite automobile. En un peu plus de dix ans, ils conçoivent dix véhicules automobiles à conduite automatique nommés *Navlab*. En 1995, grâce au réseau de neurones de reconnaissance visuelle ALVINN (*Autonomous Land Vehicle In a Neural Network*), le *Navlab 5* parvient à traverser les Etats-Unis (de Pittsburgh à San Diego) de façon quasi-autonome, seuls les freins et l'accélérateur étaient gérés par des opérateurs humains [208].

En 1997, Garry Kasparov, champion du monde aux échecs, est battu pour la première fois par un système informatique. Cet honneur revient à *Deep Blue*, un superordinateur conçu par IBM. Dans la même année, *Logistello*, un programme spécialisé en Othello (jeu de stratégie opposant deux joueurs) parvient à vaincre Takeshi Murakami le champion du monde [209].

Toujours en 1997 se tient la première édition de la RoboCup à Nagoya (Japon), un championnat international de football durant lequel les matchs sont joués uniquement par des robots et en autonomie par ceux-ci (*figure 24*) [210]. Cet événement réunira 38 équipes issues de 11 pays différents et donnera lieu par la suite à une édition chaque année [211]. Outre la dimension compétitive, la RoboCup sera aussi un lieu de partage de connaissances et d'échange où se tiendront des symposiums internationaux traitant de la robotique et de l'intelligence artificielle [212].



Figure 24. Photographie d'un match de la RoboCup-97 en catégorie « robots roulants de taille moyenne »
(Reproduit selon les termes de la licence CC BY-ND 2.0. Auteur : Itsuki Noda.) [212]

Le traitement du langage naturel poursuit sa course et se consolide grâce à l'enrichissement des bases de connaissances, telles que Cyc que nous avons déjà évoqué, et des bases de données lexicales telles que *WordNet* créée en 1998 [213]. Ces avancées ont permis, dans la continuité des réseaux sémantiques, la création d'ontologies dans divers domaines. Nous pouvons par exemple citer l'UMLS (*Unified Medical Language System*), une compilation faisant office de thesaurus et d'ontologie dans le domaine biomédical [214].

Les années 90 signent également le début de la démocratisation d'internet avec la naissance du *World Wide Web*, une multitude de pages accessibles via leur adresse web (ou URL, *Uniform Resource Locator*) [215]. Internet va permettre un élargissement de la communication et du partage de connaissances, mais cela pose quelques questions. Comment ces masses d'information promises à un accroissement continu peuvent et doivent être organisées ? Comment permettre aux utilisateurs de rechercher et d'avoir accès à l'information souhaitée ? C'est en 1998 que Sergey Brin and Lawrence Page proposent une réponse à ces questions à travers leur prototype de moteur de recherche à large échelle qu'ils nomment *Google*. Leur projet est conçu pour explorer et indexer les pages du web de façon à produire des résultats de recherche plus pertinents et satisfaisants que les moteurs de recherche préexistants. Le procédé repose sur des techniques avancées de recherche d'information, d'exploration de données (ou *data mining*) et d'intelligence artificielle. Ils partagent librement leur prototype donnant accès, en 1998, à 24 millions de pages [216].

L'intelligence artificielle devient de plus en plus une composante de projets plus larges au lieu d'en être le cœur. En 1999, la NASA (*National Aeronautics and Space Administration*) confie à un programme d'intelligence artificielle nommé *Remote Agent Experiment* (RAX) le pilotage durant deux jours du satellite *Deep Space* à une distance d'environ 1 à 100 millions de kilomètres de la Terre. Remote Agent avait pour rôle principal de diagnostiquer les anomalies et de les corriger en toute autonomie. Il s'agit de la première intelligence artificielle à contrôler un vaisseau spatial sans la supervision d'un humain et elle le fit avec succès [217].

Également en 1999, Sony commercialise AIBO (*Artificial Intelligent Robot*), le premier robot animal de compagnie doté d'une intelligence artificielle. Celui-ci est capable de se mouvoir et d'intégrer de la reconnaissance visuelle et sonore. Ces capacités lui permettent d'être autonome, d'apprendre à partir de ce qu'il perçoit dans son environnement et de développer une personnalité évoluant au fil du temps [218].

L'intelligence artificielle commencera à être appliquée aux services durant les années 2000. Ce sera par exemple le cas de l'apprentissage en ligne (ou *e-learning*) qui, grâce à l'intelligence artificielle, devient plus adaptatif et donc plus efficace. Sur la base des données collectées concernant les intérêts, le niveau, et les objectifs de l'apprenant, le modèle d'intelligence artificielle est en mesure d'ajuster le rythme, la méthode d'apprentissage et le contenu de la formation [219]. Les systèmes de recommandations sur le web se développent également et se généralisent : e-commerces, multimédia, littérature, restauration, voyage, etc. Ceux-ci se fondent également sur les données collectées auprès des utilisateurs concernant entre autres leur usage du web. Les systèmes de recommandations mobilisent eux aussi des techniques d'intelligence artificielle tels que le raisonnement à partir de cas, le filtrage de contenu ou le filtrage collaboratif [220].

La popularisation d'internet ayant amené son lot de comportements malveillants, des solutions ont dû être pensées pour protéger le web et ses utilisateurs [221]. Au début des années 2000, Luis von Ahn et Manuel Blum développent et publient le test CAPTCHA (*completely automated public turing test to tell computers and humans apart*), un système visant à différencier un humain d'une machine. Comparable au test de Turing donc, le test CAPTCHA a été développé pour protéger un site web des spams exécutés par des programmes. Les tests CAPTCHA se sont petit à petit perfectionnés grâce notamment à l'ajout de techniques d'intelligence artificielle (ex. : indication et images présentées à l'utilisateur, celui-ci devant indiquer quelles images correspondent à l'indication) [222].

Le chercheur Sargur Srihari et son équipe mettent au point un système impliquant des algorithmes d'apprentissage automatique capables d'entraîner un modèle à déceler si deux documents écrits à la main l'ont été par la même personne. Leur étude, qu'ils publient en 2002, évalue sa fiabilité à 98% sur un échantillon de 1500 sujets. Ce système contribuera notamment à moderniser les méthodes d'analyse de la police scientifique lorsqu'il s'agit d'identifier la personne à l'origine d'un texte manuscrit [223].

En 2002, David Hanson présente *K-Bot*, une tête d'androïde capable d'analyser les visages, et de détecter et suivre les mouvements. Doté d'une peau en polymère et de 24 muscles mécaniques, *K-Bot* peut aussi simuler 28 expressions faciales [224]. Les retours dithyrambiques quant à sa réalisation pousseront David Hanson à fonder en 2003 l'entreprise Hanson Robotics qui sera notamment à l'origine de *Sophia*, un androïde encore plus confondant [225].

En 2003, l'entreprise Mitsubishi présente *Wakamaru*, un robot domestique destiné essentiellement à l'aide aux personnes âgées [226]. *Wakamaru* comprend plus de 10000 mots et est capable de communiquer par la parole. Il peut aussi reconnaître jusqu'à 10 personnes différentes via une reconnaissance vocale et faciale, se déplacer, se recharger en autonomie, diagnostiquer des symptômes simples ou encore joindre un hôpital ou les secours si nécessaire [227].

Les intelligences artificielles spécialisées dans les jeux continuent de s'améliorer durant les années 2000. En 2003, Garry Kasparov se mesure à *Deep Junior*, un programme d'échecs plus performant encore que *Deep Blue*. *Deep Junior* est capable de calculer 3 millions de positions par seconde. Après un entraînement au préalable sur sa version commerciale, Garry Kasparov parvient à égaliser avec *Deep Junior* (une victoire, une défaite et quatre égalités) [228]. Ce domaine des intelligences artificielles fera à nouveau parler de lui en 2011 lorsque l'*IBM Watson* gagnera le jeu télévisé *Jeopardy!* face à Ken Jennings et Brad Rutter, les deux champions en titre, puis en 2017 lorsqu'*AlphaGO* vaincra Ke Jie, le champion du monde au jeu de go [229,230].

En 2004, la société canadienne Everon met au point un système d'alarme cardiaque portatif qu'elle nomme *Vital Positioning System*. Celui-ci prend la forme d'une ceinture à porter au niveau de la taille ou de l'abdomen et comporte un électrocardiographe numérique, des capteurs d'activité cardiaque, un module GPS et un transmetteur Bluetooth se connectant à un téléphone. Grâce à ses composants et à l'intelligence artificielle qu'il intègre, ce système est capable de détecter l'imminence d'un infarctus du myocarde, d'en informer les services d'urgence et d'envoyer à ceux-ci la localisation GPS du/de la patient(e) [231].

A partir de 2004, la DARPA organise un événement durant lequel plusieurs véhicules terrestres entièrement autonomes sont mis en compétition, il s'agit du DARPA Grand Challenge. Lors de la deuxième édition en 2005 qui se tient sur une piste de désert (désert des Mojaves), 5 voitures compétitrices sur 23 parviennent à conduire en autonomie sur l'ensemble du parcours qui

mesure environ 228 kilomètres (*figure 25*) [232]. Durant la troisième édition en 2007, qui se tiendra cette fois-ci en milieu urbain (George Air Force Base à Victorville, Californie), 6 véhicules sur les 11 qualifiés parviendront à circuler en autonomie sur environ 96 kilomètres tout en respectant le code de la route et les conditions usuelles de trafic [233].

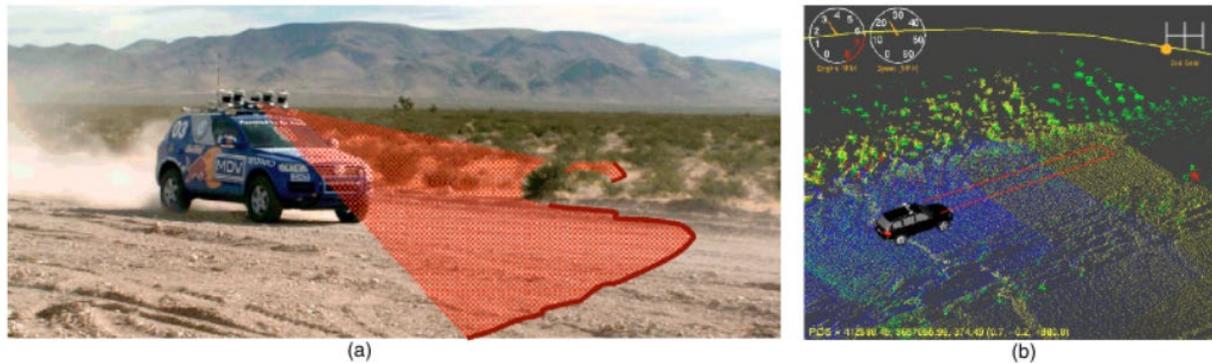


Figure 25. (a) Illustration d'un capteur laser incliné vers le bas pour balayer le terrain devant le véhicule au cours de son déplacement. (b) Chaque laser acquiert un nuage de points tridimensionnel permettant l'analyse de la praticabilité du terrain et des obstacles potentiels. (Reproduit avec la permission de l'éditeur John Wiley and Sons. Auteurs : Sebastian Thrun et coll.) [232]

Le développement d'internet et du *World Wide Web* a considérablement accéléré le phénomène des données de masse (ou *big data*), ouvrant ainsi une gigantesque fenêtre de tir aux techniques d'apprentissage automatique qui, rappelons-le, s'appuient en grande partie sur des données [234]. Cependant, un problème se pose : le traitement et l'analyse de données sont de plus en plus chronophages et énergivores à mesure que le volume de ces données est croissant. Des outils spécifiques à cette problématique ont dû alors être développés, c'est le cas notamment de *MapReduce*. Développé par l'entreprise Google en 2004, Il s'agit d'un modèle de programmation consistant à paralléliser les calculs effectués en subdivisant ces calculs entre différents nœuds interconnectés au sein d'un cluster (*figure 26*). *MapReduce* réduit ainsi le temps et la puissance de calcul nécessaires lorsqu'il s'agit d'effectuer des opérations sur des grandes quantités de données [235]. Inspiré par *MapReduce*, Doug Cutting accessibilise ce modèle en développant la première version de l'infrastructure logicielle *Hadoop* en 2006. En plus d'inclure la parallélisation de traitement, *Hadoop* comporte d'autres modules tels que le HDFS (*Hadoop Distributed File System*), un système de stockage visant notamment à garantir l'accessibilité des données en cas de panne matérielle [236]. *Hadoop* sera rendu *open source* par l'entreprise Yahoo! en 2008 [237]. Un an plus tard, Matei Zaharia conçoit l'infrastructure logicielle *Spark* ayant pour but d'améliorer la vitesse de traitement des données par *Hadoop*. Contrairement à *Hadoop* qui procède par étapes, *Spark* est en mesure de travailler sur l'ensemble des données en même temps, ce qui accélère considérablement le processus [238].

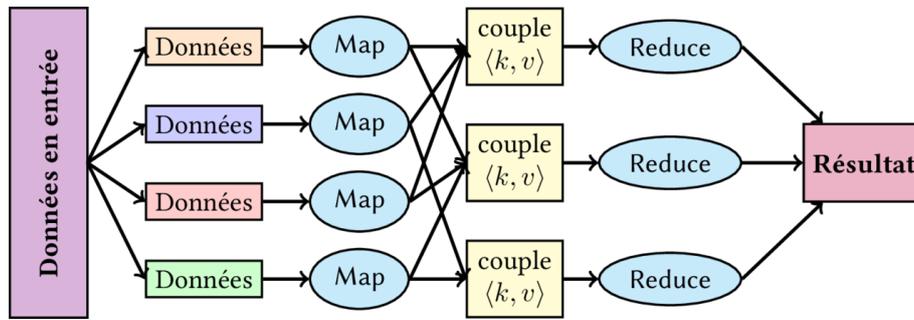


Figure 26. Schéma de fonctionnement du MapReduce (Reproduit selon les termes de la licence CC BY-SA 3.0. Auteur : Clém IAGL. <https://commons.wikimedia.org/wiki/File:Mapreduce.png>)

Ces outils, associés à la puissance de calcul sans cesse croissante des processeurs CPU (*Central Processing Unit*), permettent alors à l'apprentissage automatique d'entamer sa grande ascension. Les domaines d'application étant de plus en plus spécifiques, l'apprentissage automatique a pour challenge de proposer des méthodes tout autant spécifiques. C'est en 2007 que David Cournapeau lance le projet *Scikit-learn*, une librairie regroupant un ensemble d'algorithmes d'apprentissage automatique programmés en *Python*. Trois ans plus tard, l'institut national de recherche en sciences et technologies du numérique (Inria) prend les rênes de ce projet et le rend, pour la première fois, public [239]. *Scikit-learn* sera maintenu et enrichi avec le temps par une multitude de développeurs [240]. Dans le même temps, Geoffrey Hinton pose à nouveau sa pierre à l'édifice des réseaux de neurones en publiant un article dans lequel il expose des éléments fondamentaux quant à leur performance et leur explicabilité. Dans cet article, il emploie le terme « profond » à plusieurs reprises pour désigner des réseaux de neurones ou l'apprentissage que ceux-ci accomplissent, ce qui préfigurera la dénomination « apprentissage profond » [241]. En 2010 est lancée la première édition de l'ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*), une compétition qui se tiendra chaque année jusqu'en 2017. Durant l'ILSVRC, les participants doivent mettre à l'épreuve leurs modèles de reconnaissance d'images et tenter d'obtenir les meilleures performances avec le jeu de données *ImageNet*. En 2012, le modèle qui remporte la victoire est un réseau de neurones à 60 millions de paramètres (ou coefficients). La prouesse est notable pour deux raisons : tout d'abord, la performance atteinte est telle qu'elle surclasse de loin le modèle ayant remporté la deuxième place. De plus, les vainqueurs ont fait preuve d'ingéniosité en entraînant leur réseau de neurones avec des méthodes innovantes, comme par exemple en utilisant des processeurs graphiques (GPU) qui s'avéreront plus adaptés pour ce type de tâches que les processeurs habituellement employés en apprentissage automatique (CPU). Cette victoire va propulser au-devant de la scène les réseaux neuronaux de classification d'images et de sons [242].

Basés sur ces technologies ainsi que sur le traitement du langage naturel, de nombreux outils grand public verront le jour durant les années 2010. Le réseau social *Facebook* propose à ses utilisateurs, dès 2010, un système de reconnaissance automatique des visages [243]. Cependant, suite à plusieurs scandales médiatiques, l'entreprise Facebook décidera de mettre un terme, entre autres, à cette fonctionnalité en 2021 [244]. En 2011, l'entreprise Apple présente *Siri*, un assistant virtuel intégré à leur produit phare l'*iPhone*. *Siri* est capable d'interpréter des questions ou instructions formulées à l'oral et d'y apporter une réponse appropriée. Sur le même principe, des systèmes similaires verront le jour les années suivantes : *Google Assistant* par Google en 2012, *Cortana* par Microsoft et *Alexa* par Amazon tous deux en 2014 [245].

Les agents conversationnels autonomes (ou *chatbots*) progressent également mais se confrontent à quelques problématiques. En 2014, l'Université de Reading (Grande-Bretagne) affirme avoir fait passer avec succès le test de Turing auprès de 33% des juges. L'intelligence artificielle concernée est *Eugene Goostman*, un agent conversationnel simulant un jeune garçon ukrainien de 13 ans [246]. Cette déclaration sera cependant vivement remise en question par la communauté scientifique au vu du pourcentage de réussite jugé trop faible par certains [247]. En 2016, Microsoft et Bing tentent une expérience en introduisant sur *Twitter* l'intelligence artificielle nommée *Tay (Thinking About You)* afin que celle-ci échange avec les utilisateurs du réseau social. Cependant, suite aux actions volontaires et coordonnées de la part d'utilisateurs, les propos tenus par *Tay* devenaient injurieux et immoraux. Par conséquent, Microsoft choisit de définitivement déconnecter *Tay* après une seconde tentative une semaine plus tard aboutissant de nouveau à une situation similaire [248]. En avril 2015, l'entreprise Hanson Robotics met en service *Sophia*, une androïde conçue pour apprendre au contact des êtres humains. En effet, *Sophia* est dotée d'une intelligence artificielle évoluant sur la base du mimétisme anthropomorphique. *Sophia* peut être vue comme incarnant la synthèse des technologies présentées dans cet historique dans la mesure où, au-delà d'être un agent conversationnel, celle-ci embarque de la reconnaissance visuelle et sonore, de la robotique poussée lui conférant une capacité à imiter les expressions du visage, la gestuelle et à effectuer des actions typiquement humaines telles que peindre un portrait par exemple. Les conversations qu'elle a pu tenir depuis sa création ont mis en évidence que, malgré le bond en avant en matière d'intelligence artificielle conversationnelle qu'elle représente, *Sophia* est très loin de pouvoir prétendre au titre d'intelligence artificielle forte. En octobre 2017, *Sophia* obtient la nationalité saoudienne, ce qui fait d'elle le premier androïde à être nationalisé [249].

Par la suite, nous assistons à une stagnation des performances de l'intelligence artificielle à destination du grand public jusqu'à l'année 2022 qui signe le retour en force des intelligences artificielles génératives (*Midjourney*, *DALL-E 2*, et *ChatGPT*) [250,251]. La recherche se poursuit, l'intelligence artificielle expérimentale se penche sur des idées novatrices tandis que l'intelligence artificielle théorique tend davantage vers une formalisation de sa méthodologie [252]. L'étude des problèmes rencontrés donne lieu très régulièrement à de nouvelles techniques en apprentissage automatique [253]. Le *Python* fait désormais partie des langages de programmation les plus utilisés mondialement et sa librairie *Scikit-learn* est une référence dans le monde de l'apprentissage automatique appliqué [254]. La puissance des processeurs GPU augmente plus vite encore que celle des CPU et permet d'accroître d'un cran les performances atteintes en apprentissage profond [255,256]. Les solutions offertes par l'intelligence artificielle bénéficient désormais à l'ensemble du secteur technologique et ses applications deviennent trop nombreuses pour toutes être citées. Se pose aussi de plus en plus la question des implications éthiques et de la réglementation à bâtir autour des intelligences artificielles, notamment au sujet des droits applicables aux androïdes, de la question de la responsabilité, ou encore du statut juridique des œuvres d'art générées via intelligence artificielle [257]. Bien qu'à ce jour le test de Turing n'ait pas encore été passé avec certitude et que l'intelligence artificielle forte reste un mythe, aborder ces questions paraît urgent tant les progrès réalisés par les algorithmes d'apprentissage profond semblent coïncider avec la loi de Moore [258]. Nous aborderons ces questions par la suite en nous limitant au périmètre médical, mais avant cela ce manuscrit propose à son lecteur, s'il le souhaite, un tour d'horizon très synthétique des techniques employées en intelligence artificielle.

III. L'intelligence artificielle en pratique

Ce volet propose à son lecteur une synthèse des techniques majeures employées en intelligence artificielle. L'objet de la démarche est de fournir les éléments théoriques et techniques nécessaires pour appréhender confortablement des cas d'usage. Bien que la grande tendance actuelle soit au connexionnisme, nous verrons que le symbolisme demeure une approche à ne pas négliger.

1. Les systèmes experts

a. Des systèmes fondés sur les connaissances

Les systèmes experts s'inscrivent dans le symbolisme, l'approche la plus ancienne de l'intelligence artificielle. Il s'agit de systèmes qui tentent de reproduire les mécanismes cognitifs d'un expert en se basant sur les connaissances spécifiques à son domaine. Pour ce faire, ces connaissances suivent une logique et sont organisées en faits, règles de décision et inférences. Pour un problème donné dans des conditions données, un système expert fournira dogmatiquement la même solution. De par sa dimension déterministe, il s'agit du type d'intelligence artificielle le plus proche de l'informatique traditionnelle. Il est à noter tout de même que certains systèmes experts sont en mesure de suivre un raisonnement incertain, appelé aussi logique floue. Dans ce cas, les connaissances peuvent être associées à des coefficients de vraisemblance. Depuis Dendral, le premier système expert, ces programmes se sont développés et nous les retrouvons aujourd'hui dans un grand nombre de secteurs. Le secteur de la santé n'est pas en reste, nous avons déjà abordé dans notre historique *MYCIN*, le premier système expert à visée diagnostique [259].

Les systèmes experts peuvent se décomposer en quatre éléments importants [260] (*figure 27*) :

- Un module d'acquisition des connaissances permettant de saisir, collecter, organiser et structurer les connaissances afin de formaliser l'expérience des experts dans un langage exploitable par le système.
- Une base de connaissances devant regrouper l'ensemble des informations nécessaires à un expert. Nous y retrouvons des connaissances expertes, une base de règles, ainsi qu'une base de faits. La base de règles est comparable au savoir-faire de l'expert, celle-ci est fixe et contient l'ensemble des règles logiques utilisées par le moteur d'inférence. La base de faits est comparable à l'expérience de l'expert, celle-ci évolue au cours de l'expertise et réunit, sous la forme de données numériques et symboliques, les faits en lien avec le problème en cours de traitement.
- Un moteur d'inférence qui, à partir du contenu de la base de connaissance, est capable de générer des règles et faits nouveaux. Il existe plusieurs types de moteurs d'inférence (ex. : à chaînage avant, arrière ou mixte) mais leur principe est commun. A partir de règles d'inférence, le moteur d'inférence a pour rôle d'apporter une réponse à une requête de l'utilisateur. Ces règles d'inférence sont régies par des méta-règles chargées de guider le moteur d'inférences suivant une stratégie de résolution. Le moteur d'inférences peut aussi parfois contenir des métaconnaissances, des connaissances

apportant des informations sur une ou plusieurs autres connaissances. Les métaconnaissances permettent au moteur d'inférences d'améliorer les connaissances du système ou d'en acquérir de nouvelles.

- Une interface permettant la communication entre le système et son utilisateur.

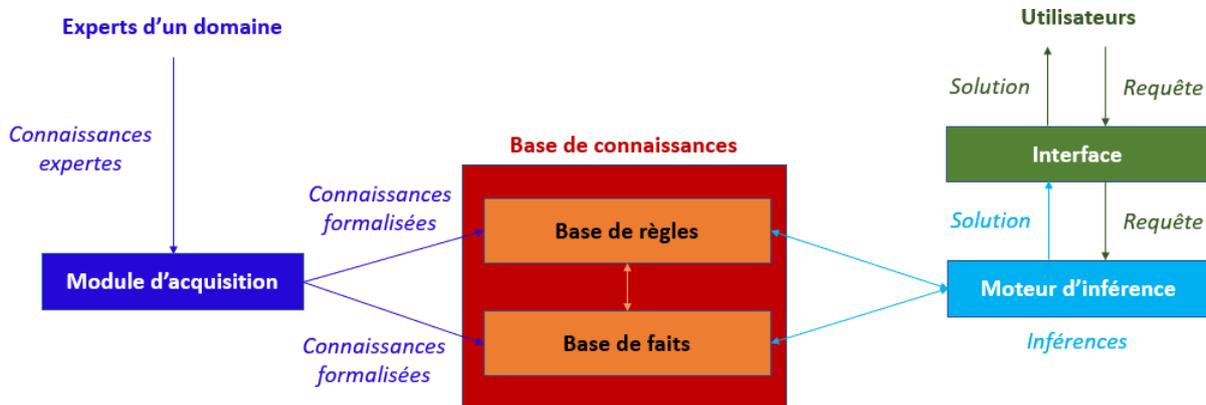


Figure 27. Les composants essentiels d'un système expert

Appliqués à la santé, l'une des tâches des systèmes experts peut donc être de proposer un diagnostic et un traitement adapté en s'appuyant sur les connaissances d'une spécialité médicale ainsi que sur une formalisation des raisonnements émis par les spécialistes. Par exemple, considérons un système expert disposant, au sein de sa base de connaissances, des symptômes relatifs à la tuberculose. Très schématiquement, si les symptômes déclarés à ce système indiquent la présence d'une tuberculose, alors son moteur d'inférence émettra la conclusion « présence d'une tuberculose » [261]. La performance d'un système expert est donc grandement dépendante de la qualité du contenu de sa base de connaissances.

Le moteur d'inférence et la base de connaissances sont bien à distinguer car le premier peut être conçu dans un langage de programmation quelconque tandis que la seconde a pour nécessité d'être écrite dans un langage déclaratif. En effet, la base de connaissances doit être accessible à un expert ne maîtrisant pas l'informatique afin que son contenu puisse être aisément consulté et modifié. Cependant, par souci d'efficacité, les connaissances déclaratives apportées par l'expert sont souvent converties en connaissances procédurales ou sous la forme de programme [259].

De manière générale, les difficultés principales relatives aux systèmes experts sont liées à la modélisation des connaissances. C'est notamment le cas dans le domaine de la santé en raison de la quantité et la complexité des connaissances à modéliser. Cela nécessite un travail approfondi de la part des spécialistes devant décrire avec précision leurs connaissances mais

également leurs raisonnements. Ceux-ci doivent travailler conjointement avec les ingénieurs chargés de concevoir les systèmes experts qui doivent veiller à ce que différents aspects soient étudiés : la nature des données, les connaissances à acquérir, la relation entre ces connaissances, les résultats produits par le système, les types de problèmes à résoudre, la précision et autres contraintes à observer quant aux données et aux résultats, ou encore les connaissances implicites de l'expert parfois considérées comme évidentes par celui-ci. La représentation et la formalisation des connaissances ainsi que les méthodes qui en découlent sont donc un enjeu important dans la réalisation des systèmes experts d'aide à la décision diagnostique ou thérapeutique [71].

b. La représentation des connaissances

Pour pallier à ces difficultés, l'ingénierie des connaissances s'attèle, depuis les années 70, à la lourde tâche qu'est la représentation des connaissances destinées aux systèmes experts [262]. Son principe est de déterminer les connaissances explicites nécessaires au système expert puis d'adapter celles-ci dans un langage interprétable par le système [263]. Les méthodes en ingénierie des connaissances reposent sur les étapes fondamentales suivantes [264] :

- 1) L'identification et l'acquisition des connaissances à représenter ;
- 2) La représentation (incluant la formulation et la structuration), qui consiste à bâtir un réseau de connaissances et à coder celui-ci dans la base de connaissances ;
- 3) L'interconnexion des connaissances par des inférences ;
- 4) L'explication et la justification qui permettent d'assurer l'explicabilité des solutions fournies par le système ;
- 5) Dans un temps ultérieur, après leur utilisation, les connaissances représentées doivent être validées, améliorées et mises à jour.

L'ingénierie des connaissances peut notamment mettre en évidence des règles de décision pertinentes à partir de données textuelles, en particulier si ces données comportent des informations relatives aux décisions à prendre en pratique. Pour ce faire, des méthodes empruntées à l'apprentissage automatique, telles que le traitement automatique du langage naturel, peuvent être employées. Ces règles de décision peuvent ensuite être formalisées afin d'être interprétables par le système expert. Il existe plusieurs formalismes : les logiques basées sur des prédicats, les logiques non classiques (ex. : logique floue), les règles de déduction ou sujettes à des conditions, le formalisme objet (basé sur la programmation objet) [265].

Ces méthodes de formalisation ont cependant des limites. Bien que celles-ci permettent à un programme de lire le contenu de la base de connaissances, elles ne garantissent pas à ce programme d'en saisir le sens. Si nous reprenons notre exemple de système expert comportant la règle de décision « si symptômes *X*, alors diagnostic *Y* », les termes correspondant aux symptômes ne sont en l'état pas substituables par d'autres chaînes de caractères. Si le terme soumis à cette partie du système est différent de ce qui est attendu, la requête ne sera pas en mesure d'aboutir sur le diagnostic approprié. Pour rendre cela possible, il faudrait que le système dispose de l'ensemble des termes pouvant être employés et que ceux-ci puissent être interconnectés selon leurs significations et leurs implications mutuelles.

A défaut de pouvoir fournir et relier l'entièreté du vocabulaire existant à un système, une autre approche consiste à standardiser les termes employés pour désigner les sujets d'une discipline. Dans le domaine de la santé, des initiatives de nomenclature ont été menées visant à standardiser et interconnecter les termes médicaux à travers des terminologies hiérarchisées. Nous pouvons par exemple citer la CIM (Classification internationale des maladies) pour les pathologies, MedDRA (*Medical Dictionary for Regulatory Activities*) pour les traitements médicamenteux ou encore LOINC (*Logical Observation Identifiers Names & Codes*) pour les examens de biologie médicale [266,267,268]. Cependant, cette approche ne répond que partiellement au problème car la hiérarchie qui érige ces nomenclatures ne permet pas aux termes d'être porteurs de sens pour un système informatique. L'ingénierie des connaissances souhaite dépasser ces limites en attribuant véritablement du sens aux données, et l'un des outils essentiels qu'elle développera dans ce but sont les ontologies.

c. Les ontologies

Le concept d'ontologie nous vient du XVII^{ème} siècle, il s'agit d'une branche de la philosophie qui étudie le sens de l'être [269]. En informatique, une ontologie désigne une forme de représentation des connaissances, dans un domaine ou sous-domaine, tenant compte des relations sous-tendant les concepts propres à ces connaissances. Ces relations, souvent de nature sémantique, forment ainsi un réseau structuré (*figure 28*). Un terme n'est donc pas seulement défini par ses synonymes, au sein d'une ontologie celui-ci est défini par un ensemble de connaissances dont il est l'objet. Ces architectures permettent aux termes d'acquérir du relief et, dans une certaine mesure, du sens. Les ontologies sont, par nécessité, la résultante d'un consensus entre plusieurs collaborateurs car celles-ci doivent s'approcher le plus possible d'une représentation objective. Elles sont souvent représentées sous la forme d'un graphique, facilitant ainsi leur conception et leur amélioration [270].

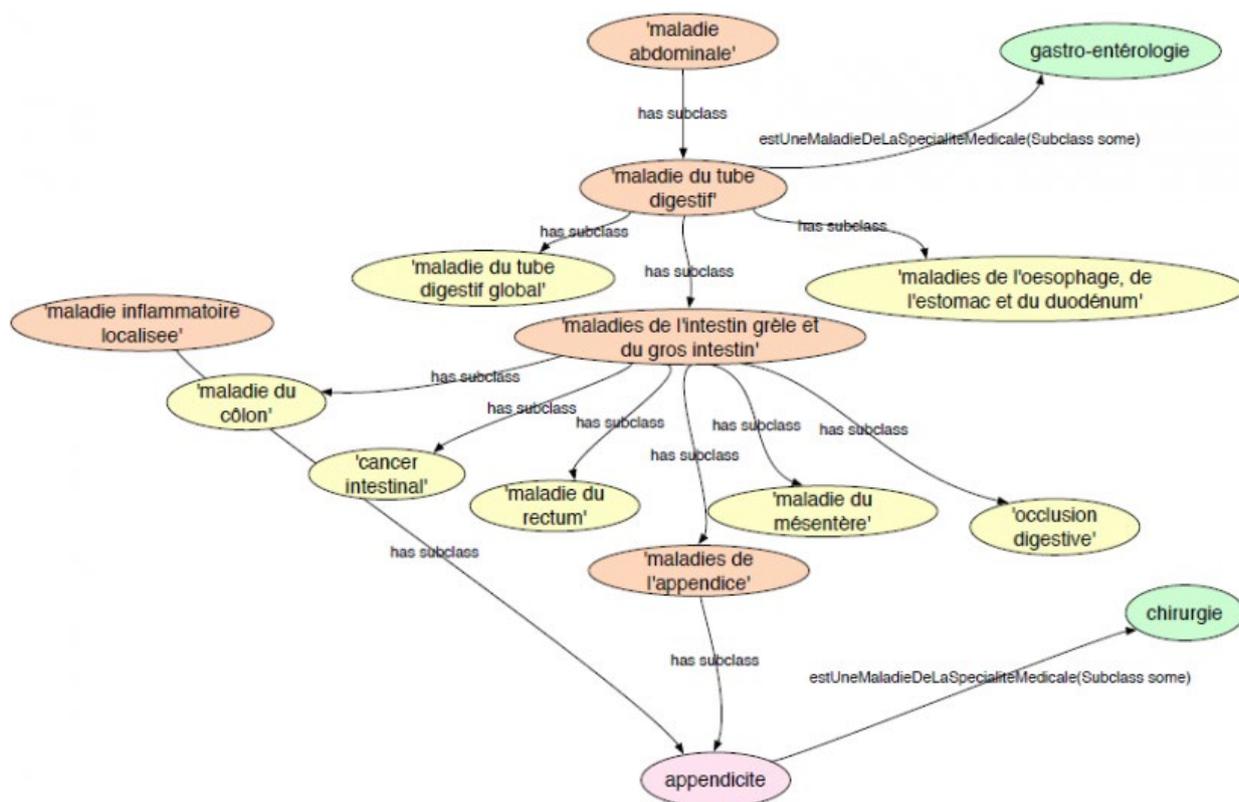


Figure 28. Ontologie convergeant vers un diagnostic d'appendicite dans le cadre du projet *Lerudi*
 (Reproduit avec la permission de l'auteur, Emmanuel Coron) [271]

En santé, la provenance des connaissances peut être très hétérogène : bases de données, littérature scientifique, recommandations de bonnes pratiques, comptes rendus médicaux, expertises, etc [272]. Cette diversité des sources s'accompagne de surcroît d'une hétérogénéité de la nature des données : cliniques, génétiques, biologiques, etc. Qui-plus-est, celles-ci ne sont pas uniquement sous forme de texte, elles peuvent notamment être des images lorsqu'il s'agit de données d'imagerie médicale [273]. Il est donc essentiel pour un système expert d'être capable de considérer et traiter des données de natures et formats variés. Cette problématique peut également être résolue à travers les ontologies par interopérabilité sémantique. En effet, les ontologies sont particulièrement utiles lorsqu'il s'agit de croiser des données, y compris lorsque celles-ci sont intégrées au sein d'ontologies différentes. Les relations sémantiques sont en mesure d'interconnecter les concepts au-delà des frontières définies par leurs ontologies, et permettent de ce fait au système expert d'en avoir une vue plus globale [274].

A l'instar des travaux de nomenclature, des projets visant à développer et collectiviser les ontologies en santé sont actuellement actifs. Nous pouvons notamment citer le projet *Presence* mené par le *National Center for Biomedical Ontology* de Stanford, *BioPortal* ou encore la *SNOMED CT* (*Systematized Nomenclature of Medicine Clinical Terms*) [275,276,277].

d. L'aide à la décision

Grâce aux ontologies, les systèmes experts acquièrent leurs lettres de noblesse car leurs règles de décision se fondent dorénavant sur une large toile sémantique. Les systèmes experts obtiennent ainsi une forme de raisonnement et la capacité d'aborder des problèmes médicaux de façon holistique. En s'intégrant dans des systèmes d'aide à la décision, les systèmes experts deviennent capables d'épauler les professionnels de santé dans leurs démarches diagnostiques et thérapeutiques (*figure 29*) [278].

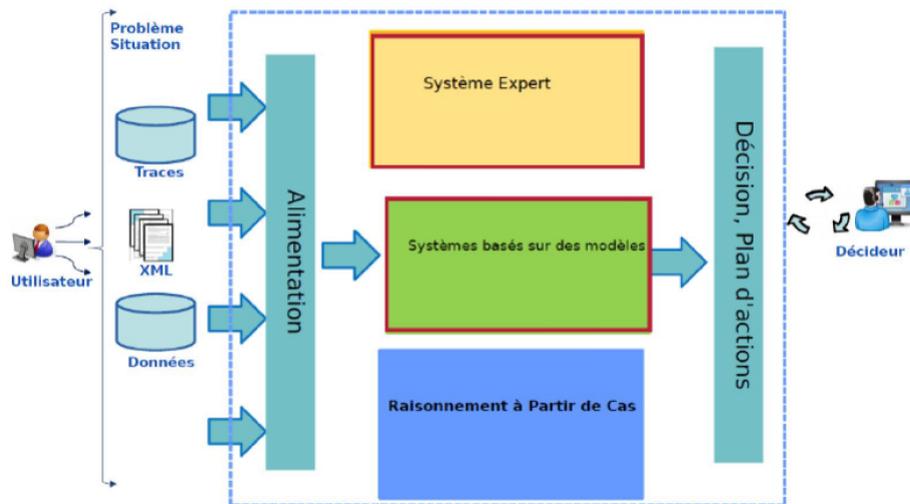


Figure 29. Représentation d'un système d'aide à la décision (Reproduit avec la permission de l'auteur, Jean Grieu) [279]

En fournissant à un système expert les ontologies représentant chacune des pathologies et chacun des traitements connus à ce jour, celui-ci serait théoriquement capable de tenir compte des subtilités symptomatologiques, des contre-indications et interactions médicamenteuses pour proposer les diagnostics, prescriptions et recommandations les plus précis possibles. Ces systèmes prennent tout leur sens lorsque la démarche diagnostique ou thérapeutique concerne un(e) patient(e) souffrant de multiples pathologies (*figure 30*) [280]. L'enjeu fondamental des systèmes experts appliqués au diagnostic et à la thérapeutique semble donc définitivement résider dans la représentation des connaissances médicales, et notamment dans la conception d'ontologies.

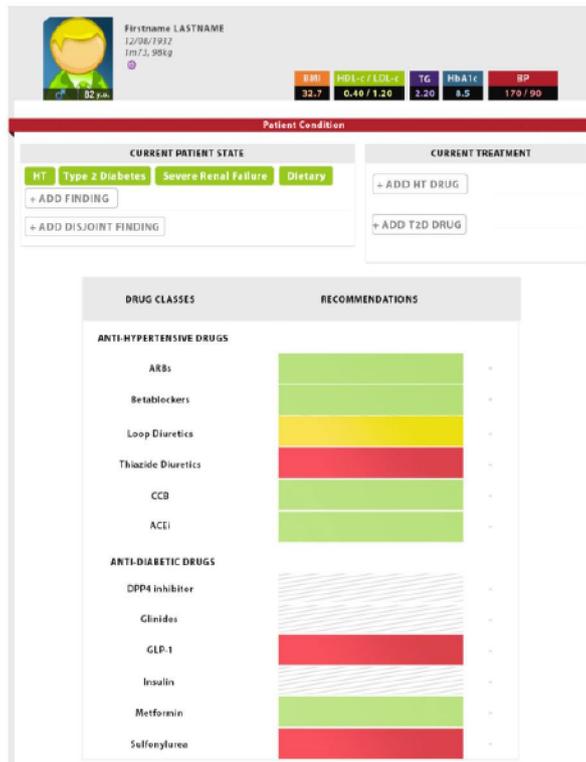


Figure 30. Interface utilisateur du système expert CDSS (*Clinical Decision Support System*) (Reproduit selon les termes de la licence CC BY-NC 3.0. Auteurs : Alexandre Galopin et coll.) [280]

e. Le défi des connaissances implicites

Une représentation des connaissances aboutie implique de pouvoir formaliser l'ensemble des connaissances propres au sujet concerné. Cela induit une problématique complexe, celle des connaissances implicites. A l'inverse des connaissances explicites qui sont disponibles sur des supports d'information écrits ou électroniques, les connaissances implicites désignent des compétences innées ou acquises par l'expérience (*figure 31*). Celles-ci comportent une dimension intuitive, sont spécifiques au contexte et généralement communiquées à travers des exemples [281].

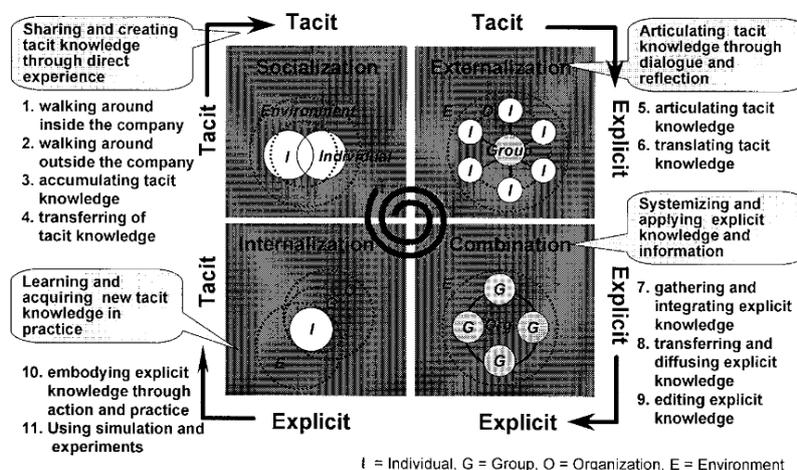


Figure 31. Modèle SECI (Socialisation, Extériorisation, Combinaison, Intériorisation) de création des connaissances (Reproduit avec la permission de l'éditeur Taylor & Francis. Auteurs : Ikujiro Nonaka et Ryoko Toyama.) [281]

Les connaissances implicites, de par leur caractère tacite, sont plus difficiles à spécifier et donc à formaliser. Pourtant, il s'agit de connaissances intervenant profitablement dans le processus de décision. Celles-ci peuvent être assimilées à l'idée de sens commun ou de culture générale, que le projet Cyc tente de représenter à travers une ontologie globale et une base de connaissance générale.

f. Forces et faiblesses des systèmes experts

La grande force des systèmes experts réside dans leur explicabilité. Leur fonctionnement intrinsèque est connu, compris et maîtrisé, ce qui permet aux équipes en charge de tenir compte de la complexité des cas, notamment prévoir les exceptions. Ce sont des systèmes ayant montré leur efficacité lorsqu'il s'agit de résoudre des problèmes suivant un protocole, d'apporter des réponses à des questions de logique formelle ou de manipuler des symboles clairement définis [259].

En revanche, les systèmes experts souffrent d'une rigidité de raisonnement qu'ils doivent justement à leur formalisme. Ils se révèlent inefficaces lorsque le problème à traiter leur est inconnu, en particulier lorsque celui-ci est atypique. Ils perdent également en puissance lorsque les données réelles relatives à ce problème sont évolutives ou trop contextuelles. Cela rend, par conséquent, les systèmes experts inadaptés à certains secteurs tels que les ressources humaines, le commerce, le marketing ou encore la médecine dans une certaine mesure. Ce problème est accentué par la difficulté inhérente à la formalisation des connaissances qui nécessite des compétences en informatique. Cela engendre souvent l'impossibilité pour l'utilisateur d'adapter ou améliorer le programme lui-même. Qui-plus-est, comme nous l'avons expliqué précédemment, certaines connaissances dites implicites sont par essence complexes à représenter [282].

Plus spécifiquement en santé, nous sommes actuellement loin d'avoir à disposition des ontologies qui couvrent l'ensemble du domaine [283]. De plus, la conception d'une ontologie dans les règles de l'art est très chronophage et coûteuse [284]. Le processus exige l'implication d'experts du domaine traité et d'un nombre souvent élevé d'ingénieurs chargés de formaliser et développer les outils mathématiques et techniques requis. Une solution à ce problème serait de concevoir des programmes capables de mettre au point et ajuster eux-mêmes leurs modèles d'analyse à partir d'un entraînement adapté et des données fournies par les experts. Une telle approche existe, il s'agit de l'apprentissage automatique.

2. L'apprentissage automatique

a. Une science des données et de l'apprentissage

Le *machine learning*, anglicisme d'apprentissage automatique, s'inscrit dans le connexionnisme. Il s'agit d'une approche dont la théorie est pratiquement aussi ancienne que les termes « intelligence artificielle » mais qui a dû attendre les années 90 pour se développer. Jusqu'à nos jours, les techniques d'apprentissage automatique poursuivent leur amélioration qui semble s'indexer à la hausse de la puissance des processeurs et des quantités de données collectées, car en effet l'apprentissage automatique repose sur deux dimensions essentielles : les données et, comme son appellation l'indique, l'apprentissage [285]. Comme nous le verrons au cours de ce manuscrit, les applications de l'apprentissage automatique sont très nombreuses dans le domaine de la santé, notamment en diagnostic et en thérapeutique.

Le principe de l'apprentissage automatique est d'identifier des règles (explicites ou implicites) contenues dans un jeu de données, à travers une méthodologie probabiliste dirigée par des algorithmes plus ou moins complexes. Cette approche prend le contre-pied du symbolisme en partant du principe que, au lieu de formaliser les connaissances, il est plus simple et plus rapide de confier cette tâche à des algorithmes capables de généraliser par induction des lois à partir du jeu de données fourni [285]. Les données, que certains nomment parfois « *l'or noir du 21^{ème} siècle* », sont donc ici la matière première, d'où la place centrale que celles-ci occupent actuellement dans le paysage des intelligences artificielles [286].

En 2020, le volume total des données créées, copiées et consommées a été estimé à 64,2 zettabytes (10^{21} bytes) mondialement, les données de santé n'échappant pas à la règle. La pandémie de COVID-19, la démocratisation de l'internet des objets (objets connectés) ainsi que le déploiement de la fibre et de la 5G ont largement concouru à accélérer ce phénomène, amenant ce chiffre à pratiquement doubler d'ici 2025 selon les prévisions [287]. Cette tendance élargit davantage le champ des possibles pour l'apprentissage automatique. D'ailleurs, l'apprentissage automatique s'inscrit généralement dans une démarche plus large que l'on nomme souvent « sciences des données » et qui réunit, en plus de l'apprentissage automatique en lui-même, la collecte de données, leur manipulation, leur analyse descriptive et leur visualisation [288].

L'apprentissage automatique est parfois apparenté aux modèles statistiques, à raison car celui-ci s'appuie en partie sur des théories statistiques. Cependant, il s'en différencie par les outils qu'il emploie et par la notion d'apprentissage qui le caractérise. Tom Mitchell, actuel président du département apprentissage automatique à l'Université de Carnegie Mellon, nous livre en 1997 sa définition de l'apprentissage automatique : « *On dit qu'un programme informatique apprend de l'expérience E par rapport à une catégorie de tâches T et à une mesure de performance P , si sa performance aux tâches T , mesurée par P , s'améliore avec l'expérience E .* » [289]. Cette description, aussi large qu'abstraite, laisse envisager une multitude de méthodes d'apprentissages.

b. L'apprentissage sous toutes ses formes

Le principe général des algorithmes d'apprentissage automatique est d'entraîner un modèle à partir de données de sorte que les paramètres (ou coefficients) de ce modèle tendent vers une minimisation de l'erreur générale. Usuellement, l'apprentissage automatique peut opérer suivant six méthodes d'apprentissage distinctes :

- L'apprentissage supervisé consiste à entraîner un modèle en dirigeant celui-ci vers une échelle numérique ou des classes préétablies. En d'autres termes, l'entraînement est réalisé à partir de données dont on précise la classe ou la valeur numérique associée, ces données sont alors dites « labellisées ». Par exemple, si l'on souhaite entraîner un modèle à détecter des tumeurs malignes à partir de photographies, l'apprentissage supervisé implique de labelliser chaque photographie du jeu de données utilisé pour son entraînement, c'est-à-dire d'indiquer pour chacune de ces photographies si celle-ci présente une tumeur maligne ou non (*figure 32*) [285].

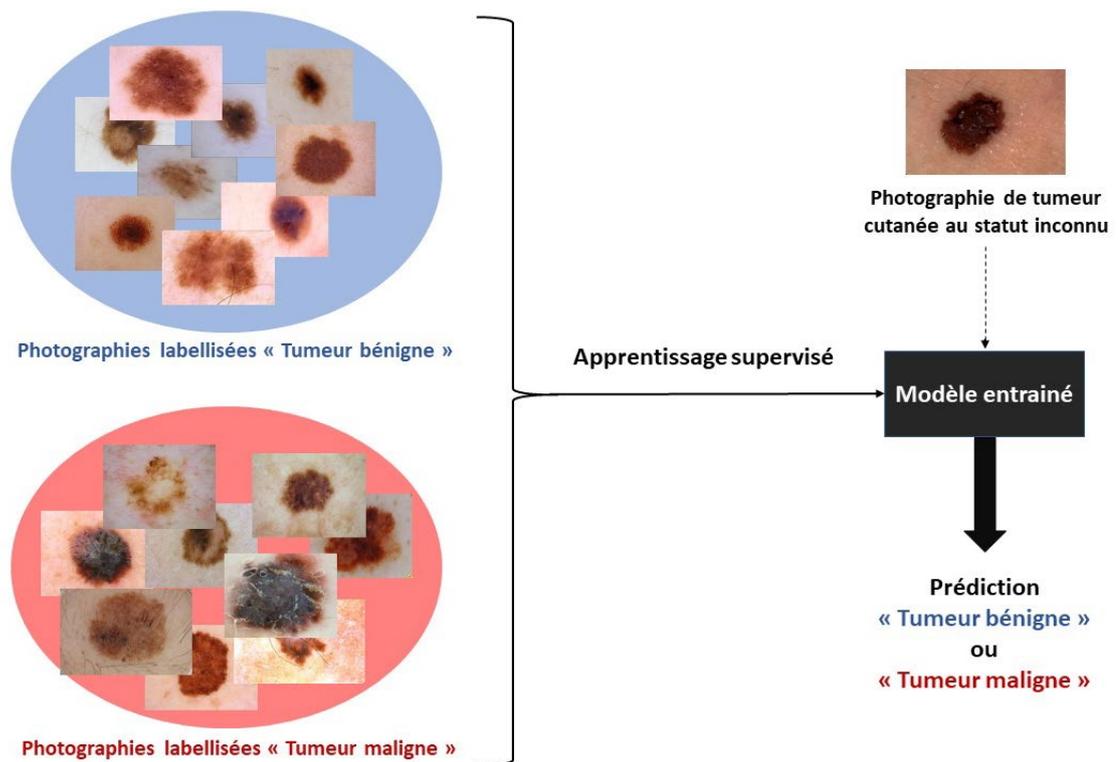


Figure 32. Illustration d'un apprentissage supervisé réalisé à partir de photographies de tumeurs cutanées

- A l'inverse, l'apprentissage non supervisé est réalisé à partir de données vierges de toute classe ou valeur numérique associée, donc non labellisées. Dans le cadre de l'apprentissage non supervisé, l'algorithme est en quelques sortes autodidacte et juge, celui-ci doit déduire lui-même les éventuelles classes ou échelles numériques se profilant des données et former des groupes à partir de celles-ci. Le degré d'autodidactie est plus ou moins élevé selon les algorithmes d'apprentissage non supervisé car, par exemple, certains opèrent en requérant que nous leur précisions le nombre de classes à conjecturer tandis que d'autres ne requièrent pas cette information. En reprenant l'exemple illustrant l'apprentissage supervisé mais cette fois-ci en effectuant un apprentissage non supervisé, notre algorithme serait chargé d'établir lui-même des catégories à partir des photographies. En l'occurrence, nous attendrions de l'algorithme qu'il identifie au moins deux groupes distincts correspondant aux tumeurs bénignes et aux tumeurs malignes (*figure 33*) [285].

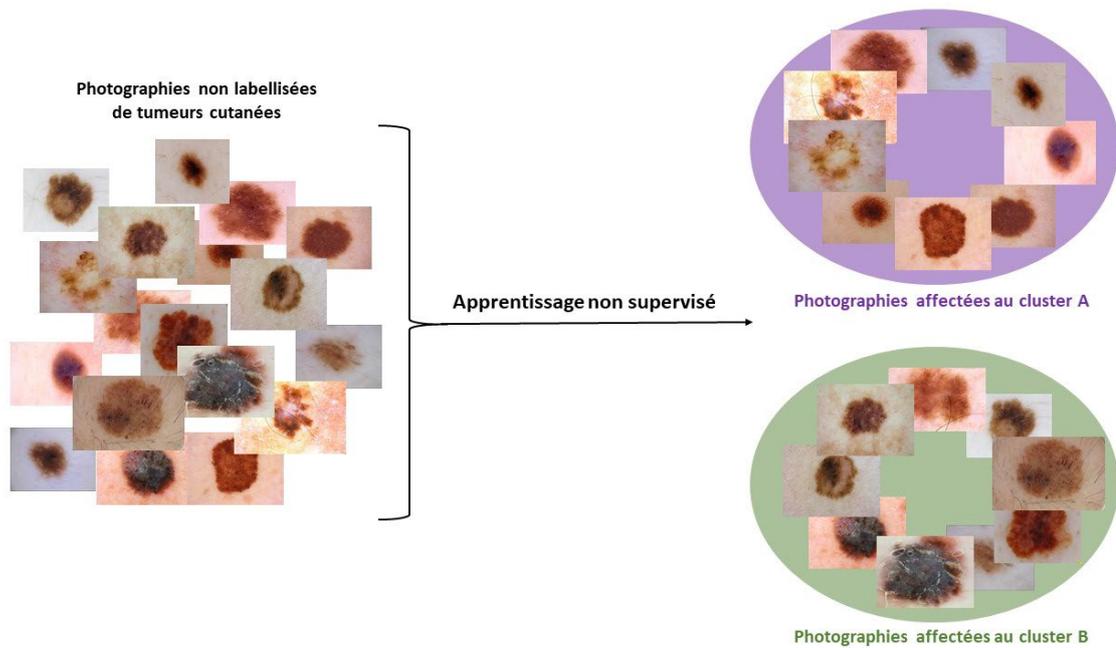


Figure 33. Illustration d'un apprentissage non supervisé réalisé à partir de photographies de tumeurs cutanées

- L'apprentissage semi-supervisé est un hybride des deux apprentissages sus-cités. L'entraînement s'effectue sur un mélange de données labellisées et non labellisées [290].
- L'apprentissage auto-supervisé est une forme avancée de l'apprentissage supervisé car celui-ci peut traiter des données non labellisées. Le modèle est entraîné en devant prédire une partie des données qui est masquée (ex. : mots retirés d'une phrase ou pixels manquants sur une image) à partir des données visibles. Au cours de son entraînement, le modèle est ainsi supposé acquérir la capacité d'attribuer lui-même des labels aux données non labellisées [291].
- L'apprentissage par renforcement se distingue des apprentissages précédemment présentés dans la mesure où celui-ci se fonde sur un système de récompense. Cette méthode d'apprentissage concerne des modèles qui apprennent à partir de leur environnement et du résultat de leurs actions. En itérant des expériences puis en tenant compte de leurs valeurs de retour basées sur ce qui est attendu ou non, ces modèles sont en mesure de tendre vers un comportement décisionnel optimal (*figure 34*) [285].

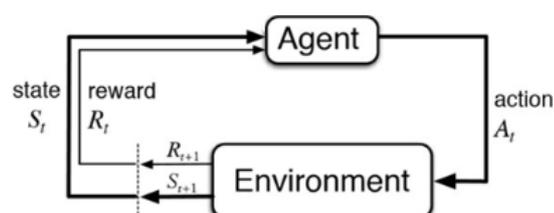


Figure 34. Apprentissage par renforcement (Reproduit selon les termes de la licence CC BY 2.0. Auteur : Batta Mahesh.) [292]

- L'apprentissage par transfert est une méthode d'apprentissage fondée sur, comme son nom l'indique, le transfert de connaissances. En pratique, il s'agit de transvaser les paramètres d'un modèle ayant déjà été entraîné au préalable sur une tâche similaire à ce que nous souhaitons accomplir. L'intérêt principal est de disposer d'un modèle bénéficiant d'une performance initiale, que l'on peut ensuite réentraîner à partir de notre jeu de données pour le spécialiser davantage. Nous pourrions, par exemple, effectuer un apprentissage par transfert à partir d'un modèle de reconnaissance visuelle généraliste réputé performant, celui-ci serait déjà possiblement efficace pour détecter des tumeurs malignes à partir de photographies. Il serait alors possible d'améliorer ses performances pour cette tâche spécifique en le réentraînant sur un jeu de données comportant des photographies labélisées de tumeurs bénignes et malignes (figure 35) [293].

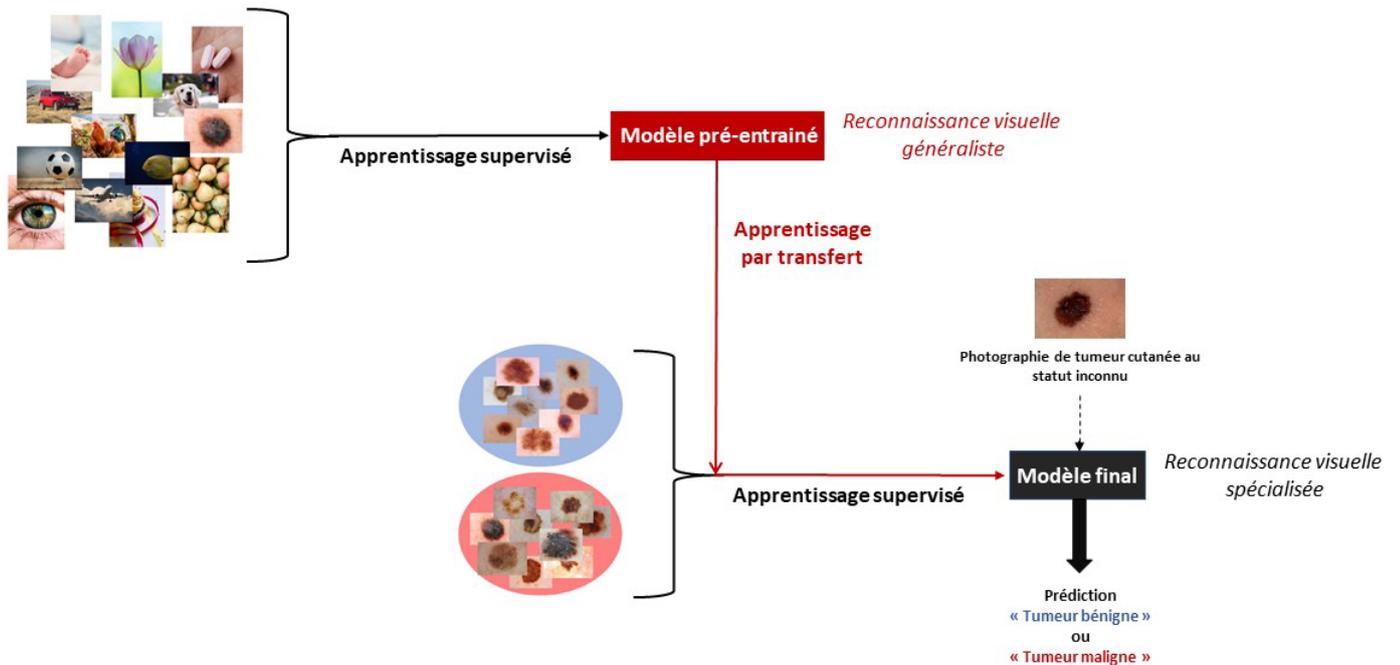


Figure 35. Illustration d'un apprentissage par transfert réalisé à partir d'un modèle de reconnaissance visuelle généraliste

c. Les quatre problématiques majeures traitées par l'apprentissage automatique

L'apprentissage automatique s'inscrit donc dans des finalités pouvant varier selon la méthode d'apprentissage adoptée. L'apprentissage supervisé vise in fine à doter un modèle de facultés de prédiction. Selon les données en entrée et ce qui est attendu en sortie, cette démarche peut être subdivisée en deux grandes problématiques : la régression et la classification.

En statistiques, la régression désigne un ensemble de méthodes permettant d'étudier le degré de corrélation entre des variables quantitatives. Par extension, l'objet de la régression en apprentissage automatique est de doter un modèle de la capacité à prédire la valeur d'une variable aléatoire quantitative dite cible (ou *target*). Suite à un apprentissage supervisé, le modèle devient capable de prédire, avec plus ou moins de précision, une valeur numérique Y inconnue en se basant sur les valeurs X connues d'une ou plusieurs variables prédictives, nommées aussi « caractéristiques » (ou *features*) [294]. Par exemple, un modèle de régression pourrait être en mesure de prédire l'espérance de vie d'un(e) patient(e) suite à un diagnostic d'une pathologie en se basant sur ses facteurs pronostiques (*figure 36*).

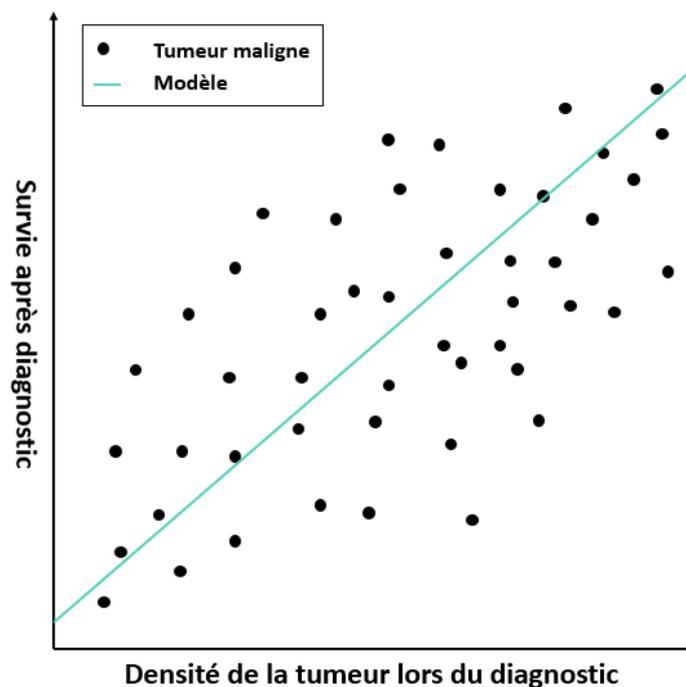


Figure 36. Modèle de régression prédisant une variable cible (survie du ou de la patiente) à partir d'une variable prédictive (densité de la tumeur)

Les problèmes de classification sont assez similaires aux problèmes de régression à ceci près que leurs prédictions concernent une variable cible qualitative. L'apprentissage est également supervisé et, à l'issue de celui-ci, un modèle obtient la capacité de prédire une qualité (valeur non numérique) Y inconnue en se basant sur les valeurs X connues d'une ou plusieurs variables

prédictives^{xxi} [296]. Par exemple, un modèle de classification pourrait être en mesure de prédire la malignité d'une tumeur en se basant sur ses caractéristiques morphologiques (figure 37).

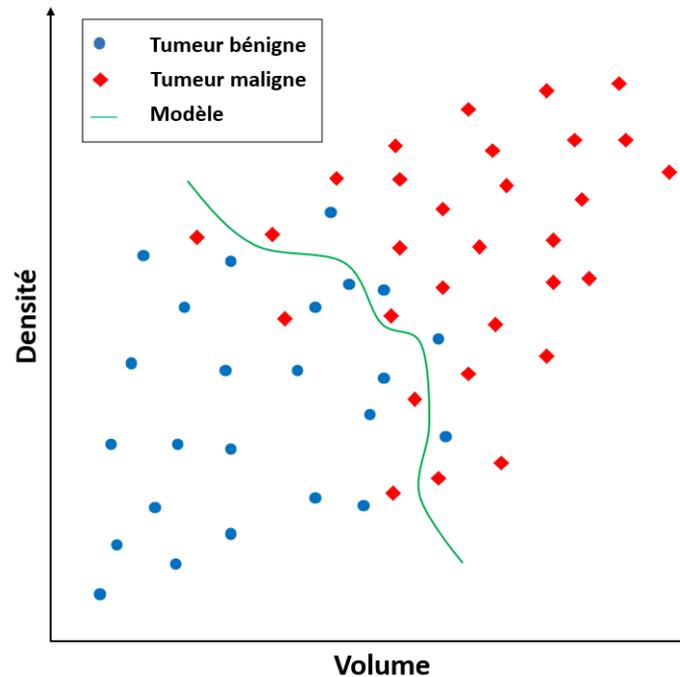


Figure 37. Modèle de classification prédisant une variable cible (malignité d'une tumeur) à partir de deux variables prédictives (volume et densité de la tumeur)

L'apprentissage non supervisé quant à lui a pour but de diriger les projecteurs vers les informations contenues dans un jeu de données. Dans ce cadre, l'apprentissage automatique peut s'inscrire dans la problématique du partitionnement de données. Le partitionnement de données (ou *clustering*) peut être vu comme un problème de classification non supervisée. Il s'agit d'une méthode permettant de segmenter un jeu de données en sous-groupes (ou clusters) sur la base des valeurs X connues d'une ou plusieurs variables. Aucune variable Y n'est considérée dans un problème de *clustering*, toutes les variables pertinentes, qu'elles soient quantitatives ou qualitatives, sont mobilisées dans un apprentissage non supervisé afin que l'algorithme puisse constituer les clusters les plus précis. La constitution des clusters se fonde sur les similarités mesurées entre les observations sur la base de deux critères de proximité : l'inertie intra-classe que l'algorithme tente de minimiser afin d'obtenir les clusters les plus homogènes possibles, et l'inertie inter-classe que l'algorithme tente de maximiser afin que les centroïdes des clusters soient les plus éloignés possibles. Les algorithmes de partitionnement peuvent également être utilisés pour mettre en évidence des éléments aberrants (ou *outliers*), c'est-à-dire des observations éloignées des clusters mis en évidence [297]. A titre d'exemple,

^{xxi} Remarque : les régressions et classifications attribuant des labels à des segments de données (ou « zones d'intérêt ») sont nommées segmentations. Par exemple, une segmentation d'image attribue des labels à des pixels ou groupes de pixels [295].

un algorithme de partitionnement pourrait être en mesure d'identifier des clusters de patients en se basant uniquement sur les symptômes qu'ils présentent, sans connaître ni préjuger des pathologies dont ils souffrent (*figure 38*).

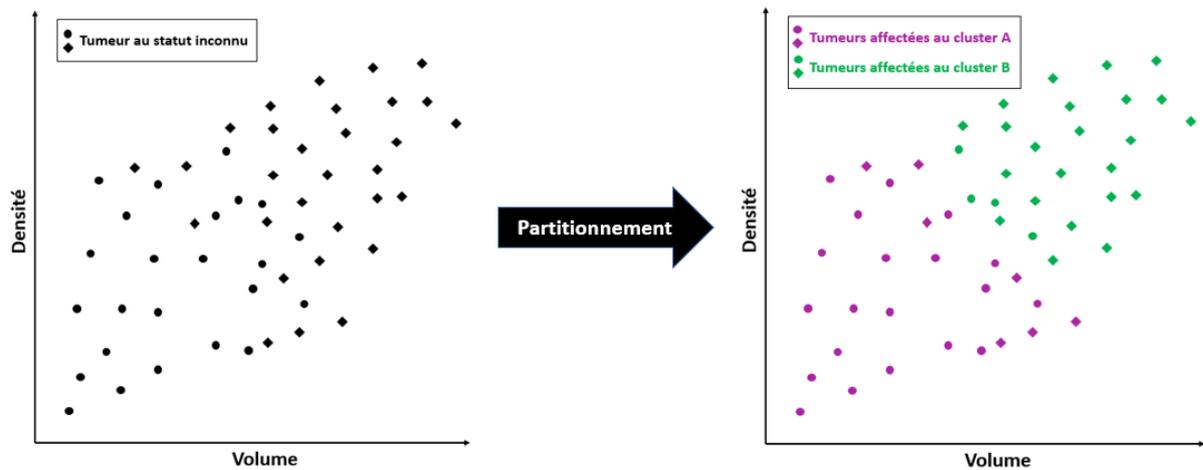


Figure 38. Partitionnement de données basé sur deux variables prédictives (volume et densité de la tumeur)

La quatrième problématique majeure de l'apprentissage automatique est la réduction de dimensionnalité. Celle-ci peut s'opérer en apprentissage supervisé comme en non supervisé, bien qu'en l'occurrence la plupart du temps celle-ci soit non supervisée. La réduction de dimensionnalité est un processus visant à réduire le nombre de dimensions au sein d'un jeu de données (*figure 39*). Cela se traduit par un sacrifice des variables dont les variances sont les plus faibles, c'est-à-dire les variables les moins porteuses en informations. Un nombre important de variables, et donc de dimensions, peut dans certains cas complexifier l'exploitation d'un jeu de données. L'intérêt de cette démarche est donc de faciliter la représentation graphique, l'analyse de ces données ou l'entraînement par un algorithme de régression, classification ou partitionnement. Plutôt que de sacrifier des variables, un algorithme de réduction de dimensionnalité procède en transposant les valeurs prises par ces variables sur les axes des nouvelles dimensions (*figure 40*) [298].

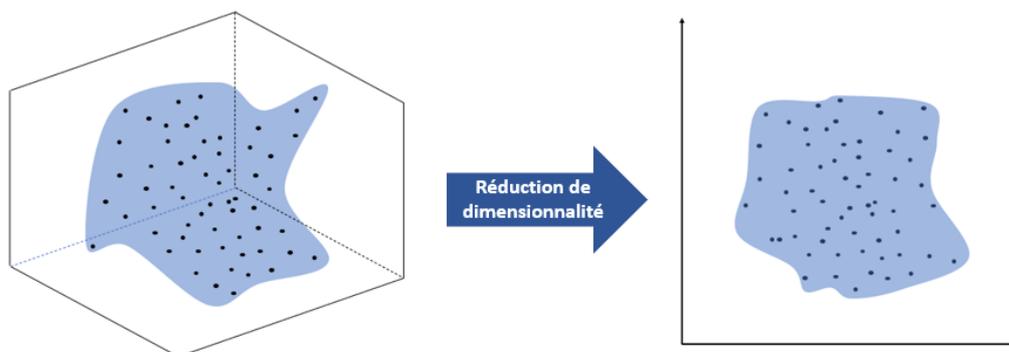


Figure 39. Réduction de dimensionnalité conservant deux dimensions appliquée à des données en 3 dimensions

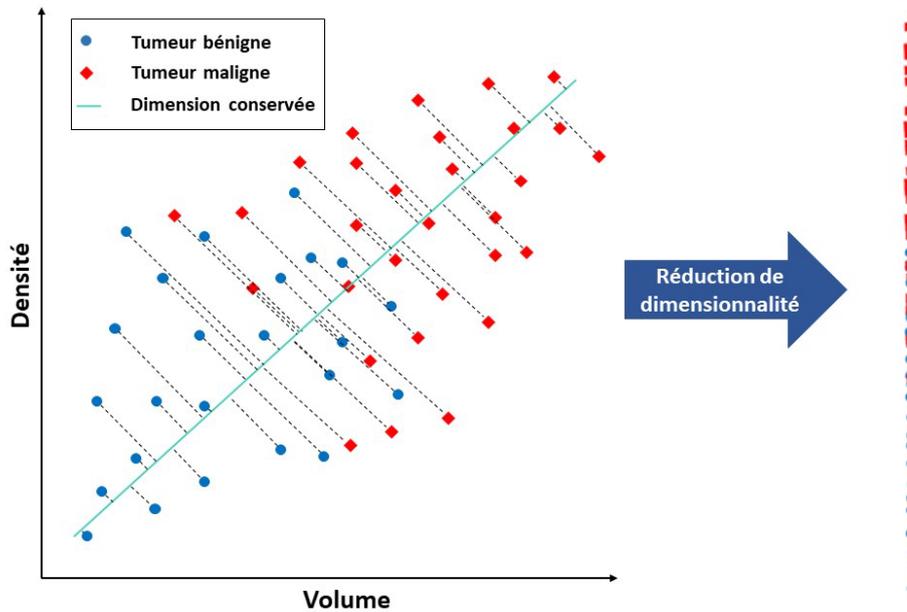


Figure 40. Réduction de dimensionnalité appliquée à une problématique de classification à deux variables prédictives (volume et densité de la tumeur)

d. Exemples d'algorithmes d'apprentissage automatique

Pour traiter les problèmes exposés précédemment, l'apprentissage automatique a recouru à des techniques diverses qui, en pratique, prennent la forme d'algorithmes. Le choix de l'algorithme à employer est assujéti à plusieurs critères dont notamment sa vélocité, son coût computationnel, sa méthode d'apprentissage, la problématique à traiter, la quantité et la nature des données ou encore le degré d'explicabilité qui sied à nos exigences. Chaque algorithme d'apprentissage automatique est caractérisé par un fonctionnement spécifique, une explicabilité plus ou moins aisée, des forces et des faiblesses qui conditionnent ses performances et son champ d'application. La performance d'un algorithme est aussi fortement conditionnée par ses hyperparamètres, c'est-à-dire des paramètres caractéristiques de l'algorithme sur lesquels l'utilisateur a la main. Les algorithmes d'apprentissage automatique sont trop nombreux pour être tous présentés dans ce manuscrit, c'est pourquoi nous nous en tiendrons à un échantillon d'algorithmes parmi les plus populaires.

d.1. La régression linéaire

La régression linéaire est le fer de lance des algorithmes de régression. Il s'agit d'un algorithme d'apprentissage supervisé traitant exclusivement des problèmes de régression [299]. Celui-ci est basé sur la modélisation d'une relation linéaire entre une variable cible quantitative et une ou plusieurs variables prédictives^{xxii} [301] : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$. A partir des données d'entraînement, l'algorithme de régression linéaire établit cette équation de façon à ce que celle-ci limite le plus possible l'erreur générale (la distance entre l'équation et chacune des observations), ou autrement dit qu'elle corresponde au mieux à l'ensemble des observations. Graphiquement, la relation linéaire peut être représentée par une droite ou un hyperplan traversant un nuage de points (observations) de manière à être le moins éloigné possible de l'ensemble des points (*figure 41*). Une droite correspond à une régression linéaire simple (une seule variable prédictive), un hyperplan à une régression linéaire multiple à deux variables prédictives.

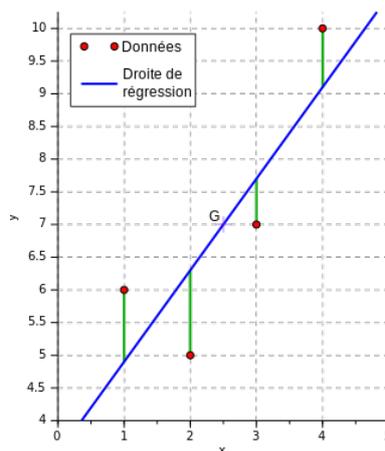


Figure 41. Droite représentant un modèle de régression linéaire simple ajustée selon la méthode des moindres carrés (Reproduit selon les termes de la licence CC BY-SA 3.0. Auteurs : Cdang, Krishnavedala. https://commons.wikimedia.org/wiki/File:Linear_least_squares_example2.svg)

La régression linéaire reste très populaire malgré sa simplicité car il s'agit d'une méthode facile à mettre en œuvre, vélocité, dont les résultats sont aisément explicables et qui ne comporte aucun hyperparamètre critique à ajuster. Cependant, la régression linéaire n'est pas adaptée à tous les jeux de données car celle-ci doit nécessairement supposer une linéarité entre les variables d'entrée et celle de sortie. Qui-plus-est, ses performances se voient fortement amoindries si les données sont peu nombreuses ou que le nombre de dimensions est important. Aussi, la régression linéaire est très sensible aux valeurs aberrantes, en particulier lorsqu'elle mobilise la somme des moindres carrés pour estimer l'erreur générale [302].

^{xxii} Y étant la variable à prédire pour une combinaison de valeurs des variables prédictives X, X_k les variables prédictives, β_k les coefficients de régression des variables prédictives, β_0 l'ordonnée à l'origine et ε l'erreur résiduelle.

d.2. La régression logistique

Bien qu'elle porte dans son nom le terme « régression », la régression logistique est en fait une méthode de classification supervisée [303]. À l'origine, il s'agit d'un modèle statistique permettant d'étudier les relations entre une variable qualitative binaire (ex. : présence/absence d'une pathologie) et une ou plusieurs variables qualitatives ou quantitatives. La régression logistique est modélisée par une relation linéaire entre les variables prédictives et le logit d'une variable cible qualitative binaire^{xxiii} [304] : $\text{logit}(\pi) = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$. En établissant cette fonction à partir des données d'entraînement ainsi qu'un seuil en guise de critère de classification, l'algorithme de régression logistique entraîne un modèle à classer des observations selon que celles-ci soient au-dessus ou en-dessous du seuil défini (*figure 42*). Classiquement, les algorithmes de régression logistique s'appuient sur une fonction sigmoïde, mais il est possible de faire intervenir d'autres types de fonctions (ex. : tangente hyperbolique, fonction softmax) [305].

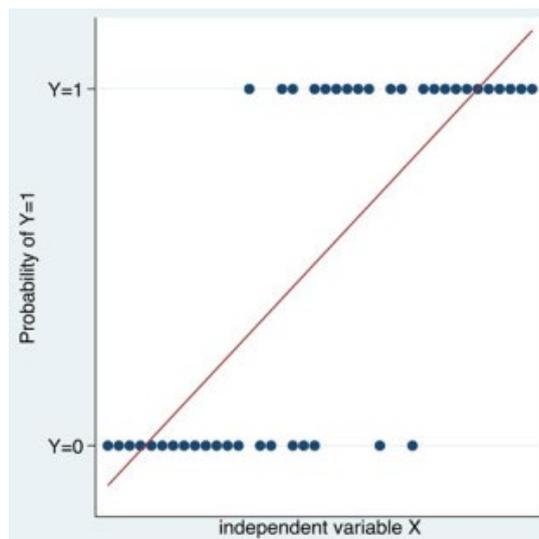


Figure 42. Courbe représentant un modèle de régression logistique ajustée selon une fonction logit
(Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Patrick Schober et Thomas R. Vetter.) [304]

La régression logistique présente les mêmes avantages que la régression linéaire. Il s'agit d'une méthode simple, vélocité, dont les résultats sont facilement explicables et qui ne comporte aucun hyperparamètre critique à ajuster. De même, ses performances sont grandement affectées si les données sont peu nombreuses ou que le nombre de dimensions est important [302].

^{xxiii} π étant la probabilité $P(Y=1 | X)$ que la variable à prédire prenne la valeur de succès (ex. : présence d'une pathologie) pour une combinaison de valeurs des variables prédictives X , X_k les variables prédictives, β_k les coefficients de régression des variables prédictives et β_0 l'ordonnée à l'origine.

d.3. La classification naïve bayésienne

La classification naïve bayésienne (ou *Naive Bayes classifier*) est un algorithme d'apprentissage supervisé probabiliste traitant, comme son nom l'indique, des problèmes de classification [306]. Celui-ci est basé sur le théorème de Bayes et admet l'hypothèse d'une indépendance des variables. Cet algorithme d'apprentissage procède en calculant, à partir des observations, les probabilités conditionnelles entre les modalités des variables prédictives et les modalités de la variable cible (ex. : probabilité que le ou la patiente soit malade sachant qu'il/elle présente une fièvre) :

$$P(\text{modalité cible} \mid \text{modalités prédictives}) = \frac{P(\text{modalités prédictives} \mid \text{modalité cible}) \times P(\text{modalité cible})}{P(\text{modalités prédictives})}$$

L'algorithme établit ensuite, à partir de l'ensemble de ces probabilités, des règles de décision minimisant l'erreur générale. La classification naïve bayésienne se décline en plusieurs algorithmes selon la distribution que l'on souhaite supposer quant aux variables prédictives : gaussienne si la distribution supposée est continue, multinomiale si la distribution supposée est discrète ou catégorielle, Bernoulli si la distribution supposée est binaire, etc [307].

La classification naïve bayésienne est un algorithme populaire en raison de sa simplicité, sa vitesse, son explicabilité et sa capacité d'adaptation aux faibles volumes de données d'entraînement [302]. Bien que l'indépendance des variables prédictives soit rare en pratique, la classification naïve bayésienne parvient à atteindre des performances élevées dans divers cas d'usage [308]. Cependant, ses performances peuvent être fortement altérées si les données d'entraînement comportent d'importants déséquilibres de classe, en particulier si cela implique que certaines probabilités conditionnelles soient nulles [309].

d.4. Les machines à vecteurs de support

Les machines à vecteurs de support (ou SVM, *support vector machines*) sont une famille d'algorithmes d'apprentissage automatique permettant de résoudre des problèmes tant de classification que de régression ou de détection de valeurs aberrantes [310,311]. En règle générale, il s'agit d'algorithmes d'apprentissage supervisé, plus rarement non supervisé (ex. : *one-class support vector machines*) [312]. Ils fonctionnent sur un principe similaire aux régressions linéaire et logistique. Comme le montre la figure ci-dessous, ils opèrent en séparant les données en sous-groupes à l'aide d'un seuil, de telle façon que la distance entre les différents sous-groupes et la frontière qui les sépare soit maximale (*figure 43*).

Cette distance est appelée « marge », c’est pourquoi les machines à vecteurs de support sont qualifiées de « séparateurs à vaste marge ». Les vecteurs de support sont en fait les observations les plus proches de la frontière [313].

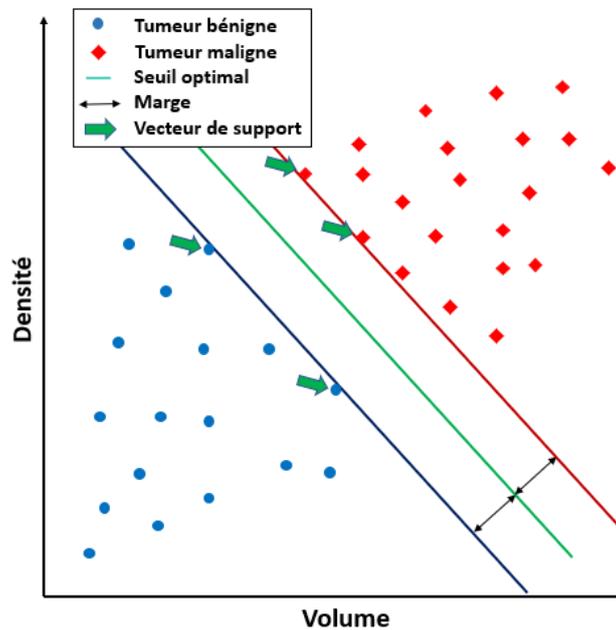


Figure 43. Machine à vecteur de support prédisant une variable cible (malignité d'une tumeur) à partir de deux variables prédictives (volume et densité de la tumeur)

Pour illustrer cela, prenons un exemple de cas d’usage en une dimension. Dans ce cas d’usage, nous souhaitons différencier les **tumeurs bénignes** des **tumeurs malignes** à l’aide d’une classification basée sur une variable prédictive (densité de la tumeur). En établissant un **seuil optimal**, il est possible d’identifier s’il s’agit d’une **tumeur bénigne** ou d’une **tumeur maligne** selon que l’observation se trouve d’un côté ou de l’autre du **seuil**. Les marges, c’est-à-dire les distances entre le **seuil** et les vecteurs de support (points de chaque classe les plus proches du seuil), sont également établies de façon à séparer les deux classes de manière équidistante (figure 44).

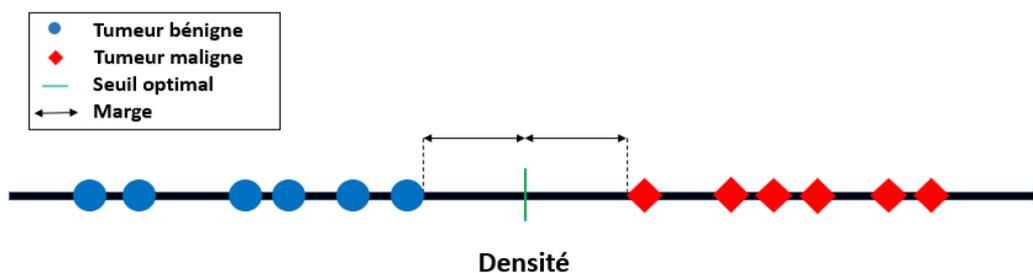


Figure 44. Machine à vecteur de support prédisant une variable cible (malignité d'une tumeur) à partir d'une variable prédictive (densité de la tumeur)

Remarque : dans un espace en deux dimensions, le seuil serait caractérisé par l'équation d'une droite séparant nos deux classes. Les marges seraient également représentées par deux droites à équidistance du seuil (figure 43). En trois dimensions ou plus, le seuil ainsi que les marges seraient caractérisés par l'équation d'un hyperplan.

Dans les rares cas où les données peuvent être séparées de façon linéaire, il est possible de faire intervenir un classifieur à marge maximale (ou *maximal margin classifier*). Celui-ci est en mesure de déterminer le seuil optimal, celui pour lequel les marges sont maximisées et les différentes classes parfaitement séparées (figures 43 et 44).

Cependant, la plupart du temps, la distribution des observations ne permet pas une séparation parfaite à l'aide d'une droite ou d'un plan linéaire (figure 45). Le problème se pose notamment si les données comportent une ou plusieurs observations aberrantes.

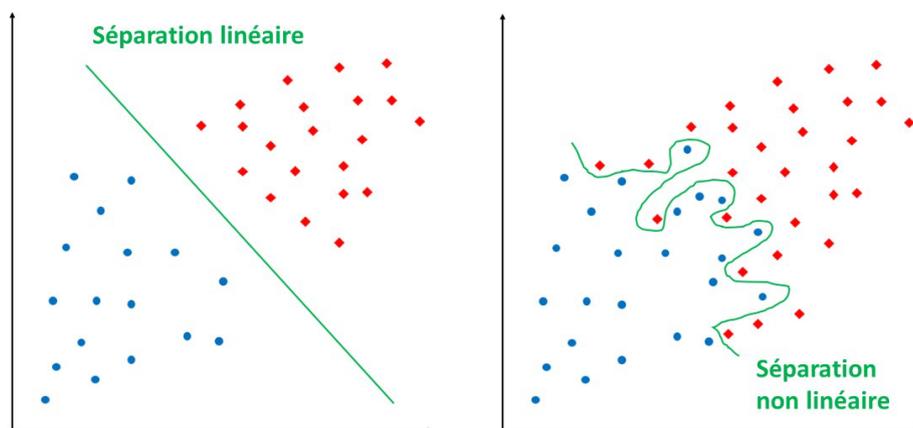


Figure 45. Distribution permettant une séparation linéaire vs. distribution ne permettant pas une séparation linéaire

Si nous reprenons notre exemple en une dimension mais cette fois-ci avec une valeur aberrante, nous observerions des marges plus petites car un classifieur à marge maximale placerait notre seuil entre l'observation aberrante et l'observation la plus proche du groupe opposé (figure 46). Dans ce cas de figure, une tumeur au statut inconnu pourrait être prédite comme étant maligne même si celle-ci compte davantage d'observations à proximité correspondant à des tumeurs bénignes (figure 46).

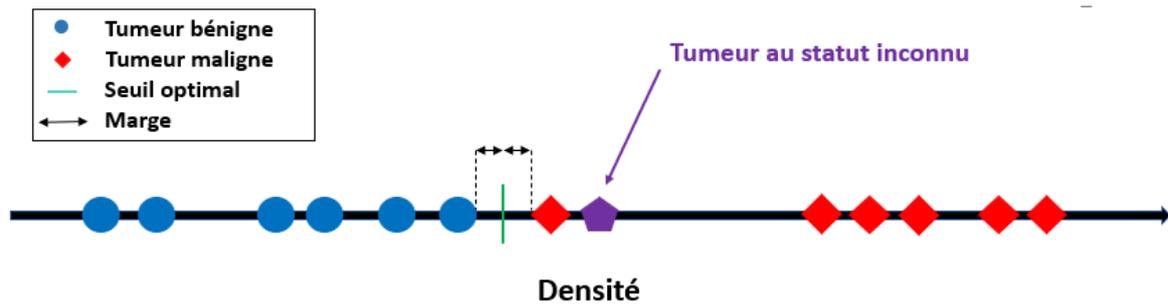


Figure 46. Effets d'une observation aberrante sur un classifieur à marge maximale

Pour éviter ce type d'erreur, un modèle d'apprentissage automatique doit être en mesure d'assouplir ses prédictions en tolérant une part d'inexactitude. En tenant compte des densités des observations et non plus uniquement de celles aux valeurs les plus extrêmes, les marges peuvent passer outre les observations aberrantes dans une certaine mesure. Celles-ci sont alors appelées marges souples (ou *soft margins*) (figure 47).

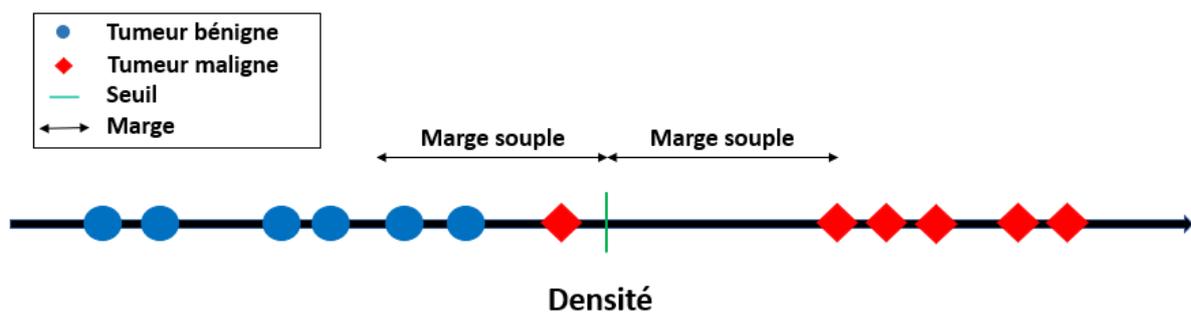


Figure 47. Marges obtenues en faisant intervenir un classifieur à marge souple

Cette approche, qui fait cette fois-ci intervenir un classifieur à marge souple (ou *soft margin classifier*), permet aux marges d'être plus ou moins permissives. L'hyperparamètre qui régit cette permissivité est le paramètre de régularisation C . Plus C est faible, plus les marges sont larges, et donc plus celles-ci englobent des observations n'appartenant pas à la classe définie. Diminuer C a pour effet de diminuer la variance mais d'augmenter le biais. Si $C = 0$, la permissivité est donc totale (figure 48). A l'inverse, augmenter C a pour effet d'étrécir les marges, donc d'augmenter la variance et de diminuer le biais. Si $C \approx +\infty$, la permissivité tend vers 0, nous revenons par conséquent à l'inflexible classifieur à marge maximale. Si le biais est trop important, les prédictions auront tendance à trop s'éloigner de la réalité (figure 49). Si la variance est trop importante, les prédictions du modèle seront imprécises. L'idée est de déterminer une valeur de C permettant un compromis [314].

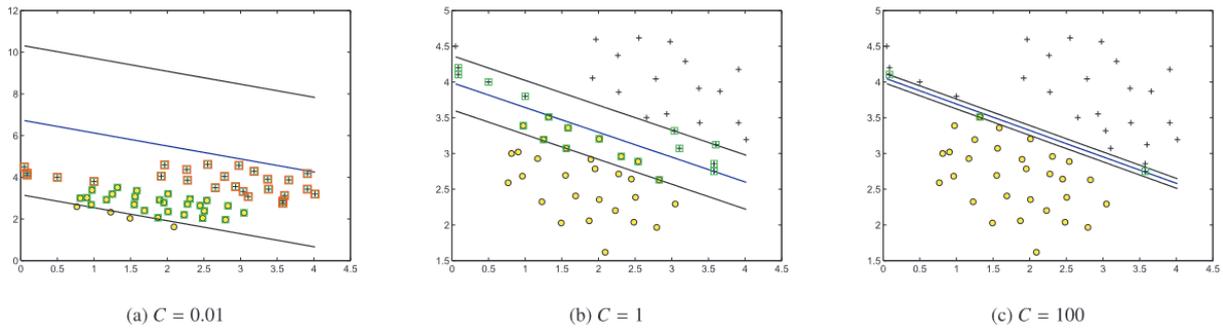


Figure 48. Effets de la variation de l'hyperparamètre C sur la souplesse des marges. Les points verts représentent les vecteurs de support tandis que les points au contour rouge représentent les observations dont la classification est erronée. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Alaa Tharwat et coll.) [314]

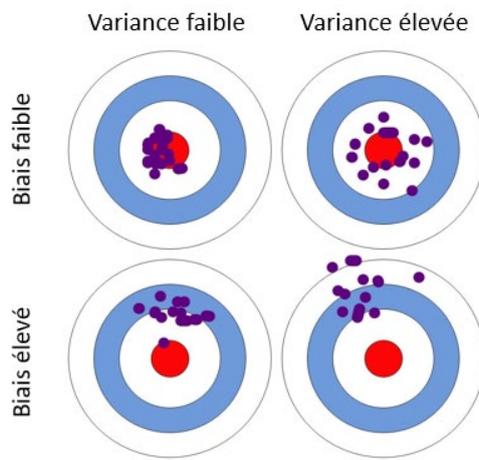


Figure 49. Illustration du biais et de la variance (© NVS Yashwanth. Reproduit selon les termes de la licence MIT. <https://nvsyashwanth.github.io/machinelearningmaster/bias-variance/>)

En pratique, les observations ne peuvent que très rarement être séparées parfaitement car celles-ci sont généralement intriquées (figure 45). Pour pallier à cela, les machines à vecteurs de support reposent souvent sur l'utilisation de noyaux de système d'exploitation (ou *kernels*). Ces fonctions mathématiques permettent de séparer les données en les projetant dans un espace vectoriel de plus grande dimension (figure 50) [315].

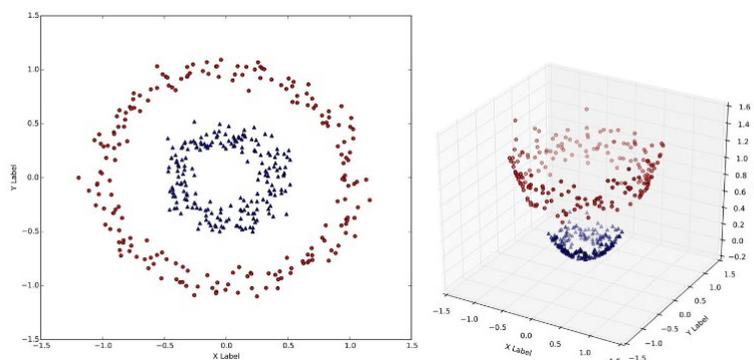


Figure 50. Projection d'une distribution d'observations en deux dimensions (à gauche) dans un espace en trois dimensions (à droite). Une fois projetées, ces observations deviennent linéairement séparables selon leur classe (par un hyperplan). (Reproduit avec la permission de l'éditeur Springer Nature. Auteur : Alaa Tharwat) [315]

Reprenons notre exemple avec cette fois-ci toutes les observations de tumeurs bénignes situées entre deux densités d'observations de tumeurs malignes (*figure 51*). En appliquant un noyau polynomial de second degré à cette distribution, nous obtiendrions un espace à deux dimensions (ayant pour ordonnée la densité au carré) dans lequel le seuil est en mesure de séparer linéairement les deux classes. Dans cette configuration, le classifieur est désormais capable de prédire la malignité d'une tumeur de façon linéaire et performante.

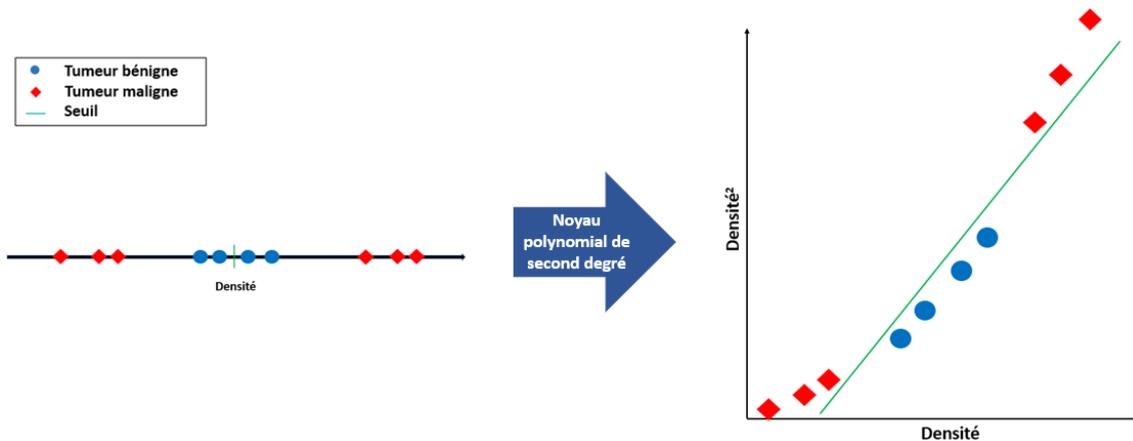


Figure 51. Application d'un noyau polynomial de second degré sur une distribution d'observations en une dimension

Différents noyaux peuvent être appliqués par les machines à vecteurs de support, le choix de celui-ci est à la discrétion de l'utilisateur. Parmi les noyaux les plus employés dans le cadre des machines à vecteurs de support, nous pouvons citer par exemple [316] :

- Le noyau polynomial^{xxiv} (comme dans notre exemple) : $k(x, x') = (x \cdot x' + 1)^d$;
- Le noyau de fonction de base radiale^{xxv} (ou *radial basis function*) : $k(x, x') = e^{-\gamma \|x - x'\|^2}$.

Les machines à vecteurs de support sont réputées pour leurs solides garanties théoriques et leur grande flexibilité. Ils sont capables d'entraîner des modèles puissants et efficaces avec une large gamme de données, y compris lorsque le nombre de dimensions est faible ou très élevé. Comparées aux régressions linéaire et logistique, les machines à vecteurs de support sont en revanche plus complexes à manœuvrer car celles-ci nécessitent un ajustement précis des hyperparamètres (paramètre de régularisation C , noyau, hyperparamètres du noyau, etc.). De plus, ces algorithmes sont particulièrement sensibles aux différences d'unités et d'échelles entre les variables, ce qui nécessite également un prétraitement soigneux des données. Enfin, les modèles construits à partir des machines à vecteurs de support sont généralement plus difficiles à expliquer [302,317].

^{xxiv} $k(x, x')$ étant le noyau (ou fonction de similarité entre deux vecteurs d'entrée x et x'), $x \cdot x'$ le produit scalaire des vecteurs d'entrée, et d l'ordre du polynôme.

^{xxv} γ (gamma) étant un hyperparamètre régulant l'influence de la distance entre les observations. Plus γ est élevé, moins la proximité entre les observations influence la classification.

d.5. La méthode des k plus proches voisins

La méthode des k plus proches voisins (ou k-NN, *k-nearest neighbors*) est un algorithme d'apprentissage supervisé permettant de résoudre des problèmes de classification ou de régression [318,319]. Celui-ci est fondé sur les distances entre notre observation pour laquelle nous souhaitons prédire une variable et les k observations qui lui sont les plus proches [320].

Reprenons un exemple de distribution d'observations correspondant à des tumeurs dont nous connaissons le statut. En considérant deux variables prédictives (densité et volume), nous pourrions représenter ces observations, dans un graphique en deux dimensions, par des points de couleur bleue ou rouge selon que ceux-ci correspondent respectivement à des tumeurs **bénignes** ou **malignes**. A partir de ces observations, nous pourrions tenter de prédire si une tumeur, dont nous ignorons le statut, est bénigne ou maligne sur la base de sa densité et son volume. Considérons trois observations de tumeurs au statut inconnu dont nous connaissons la taille et la densité, que nous représenterons par des **points mauves**. Après entraînement par un algorithme des k plus proches voisins, un modèle serait en mesure de classer nos **observations** selon que celles-ci soient plus proches de tumeurs **bénignes** ou **malignes**. Comme indiqué par le nom de l'algorithme, le nombre de voisins considérés est désigné par l'hyperparamètre k . Par exemple, si l'hyperparamètre k est égal à 3, les prédictions réalisées par notre modèle seront basées sur les modalités des trois voisins les plus proches de nos **observations** (figure 52).

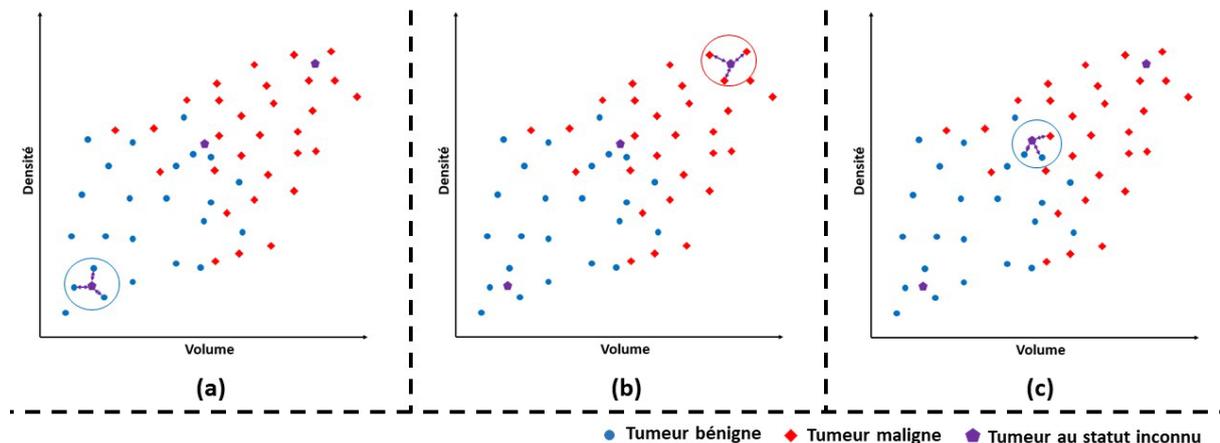


Figure 52. k-NN à $k=3$ prédisant la malignité d'une tumeur à partir de deux variables prédictives (volume et densité de la tumeur). (a) Si les trois voisins les plus proches d'une observation correspondent toutes à des tumeurs bénignes, notre modèle entraîné par un k-NN prédirait qu'il s'agirait d'une tumeur bénigne. (b) Si les trois voisins les plus proches d'une observation correspondent toutes à des tumeurs malignes, notre modèle entraîné par un k-NN prédirait qu'il s'agirait d'une tumeur maligne. (c) Si deux des trois voisins les plus proches de notre observation correspondent à des tumeurs malignes, notre modèle entraîné par un k-NN prédirait qu'il s'agirait d'une tumeur maligne car il s'agit de la modalité la plus représentée parmi les voisins considérés.

L'algorithme des k plus proches voisins présente l'avantage de produire des modèles facilement explicables, simples et généralement rapides à construire. Ceux-ci peuvent présenter de bonnes performances même avec un faible entraînement et sans ajustements, de fait ils constituent souvent un point de départ intéressant. Ses hyperparamètres ne sont néanmoins pas à négliger car l'un des problèmes majeurs de cet algorithme est sa forte sensibilité à la densité des points. De fait, selon le nombre (k) de voisins considérés et la métrique adoptée pour mesurer les distances, les résultats obtenus peuvent fortement varier (*figure 53*). Une autre faiblesse importante de l'algorithme des k plus proches voisins est sa sensibilité aux valeurs extrêmes et aux observations aberrantes. En effet, les distances calculées à partir des variables aux étendues les plus grandes auront davantage d'impact sur la prédiction. Effectuer une normalisation, c'est-à-dire uniformiser les échelles des variables, peut néanmoins résoudre ce problème. Enfin, il s'agit d'un algorithme dont les performances peuvent fortement décroître à mesure qu'augmente le nombre de variables présentes dans le jeu de données [302].

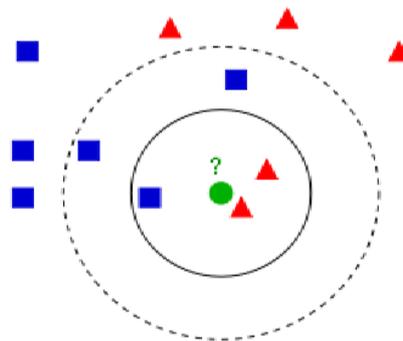


Figure 53. Illustration de l'influence de l'hyperparamètre k sur une classification par k -NN (*Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Aditya Srivastava et coll.*) [60]

d.6. La forêt d'arbres décisionnels

Les arbres de décision sont des outils d'aide à la décision représentant graphiquement des choix possibles et leurs conséquences [321]. En apprentissage automatique, il existe des algorithmes d'apprentissage supervisé capables de construire des arbres de décision optimaux pour traiter des problèmes de classification ou de régression [322,323]. Ceux-ci procèdent en divisant les différentes observations du jeu de données en sous-ensembles plus petits et plus homogènes selon la valeur de la variable cible. Cette division est réalisée en choisissant la modalité prédictive qui maximise la séparation entre les sous-ensembles, selon un critère spécifique (ex. : indice de Gini). Graphiquement, cela se traduit par des séparations rectangulaires sensées regrouper les différentes modalités ou valeurs numériques entre elles (*figure 54*). Chaque séparation répond à une condition et donne lieu à un nœud dans l'arbre de décision.

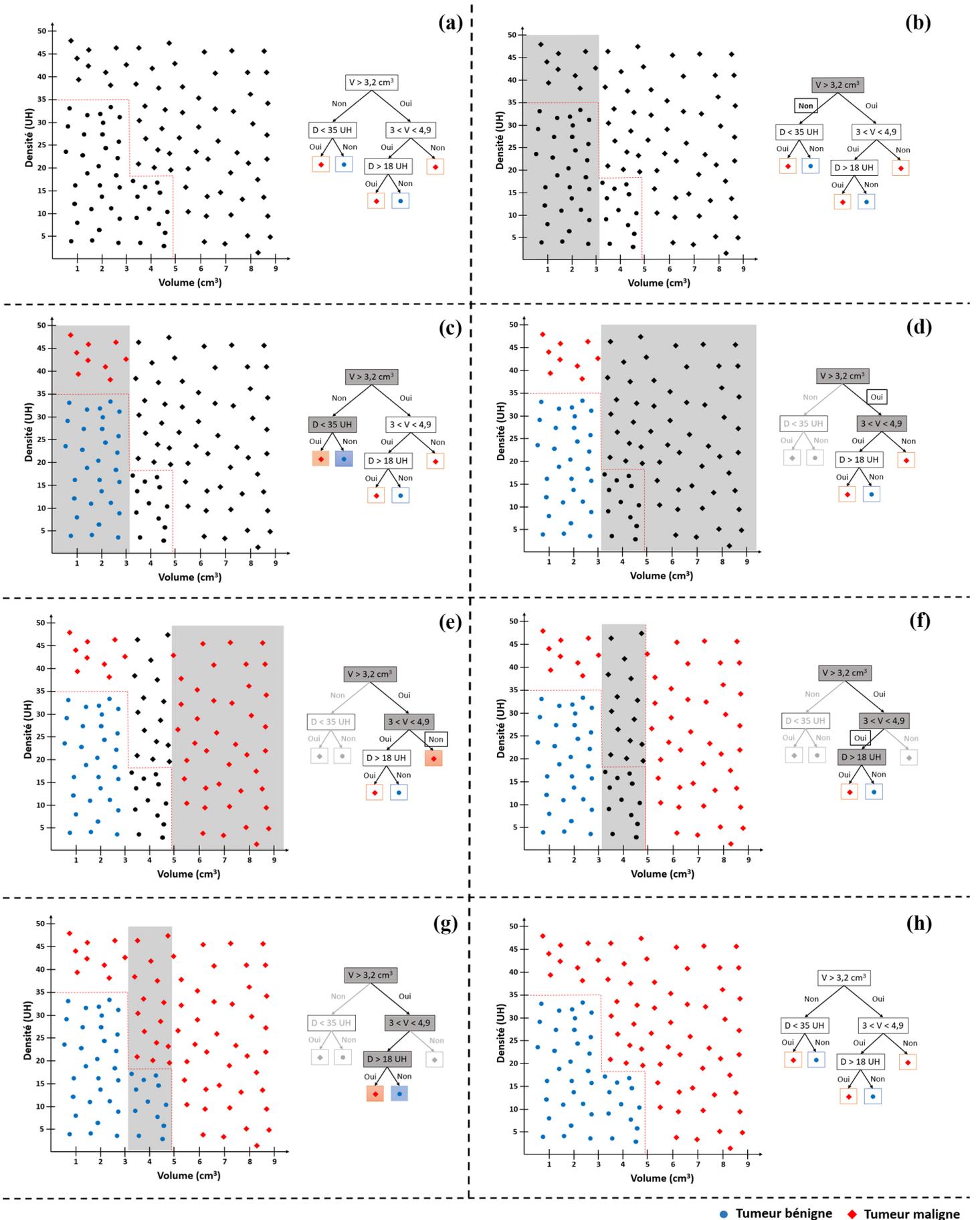


Figure 54. Apprentissage supervisé exécuté par un algorithme d'arbre décisionnel à partir d'un jeu de données comportant deux variables prédictives (densité et volume de la tumeur), la variable cible étant la malignité de la tumeur

La forêt d'arbres décisionnels, appelée aussi forêt aléatoire (ou *random forest*), est un algorithme fonctionnant sur le même principe mais en mobilisant plusieurs arbres de décision [324,325]. Cela permet d'affiner l'analyse ce qui se traduit graphiquement par des frontières moins rectangulaires (*figure 55*).

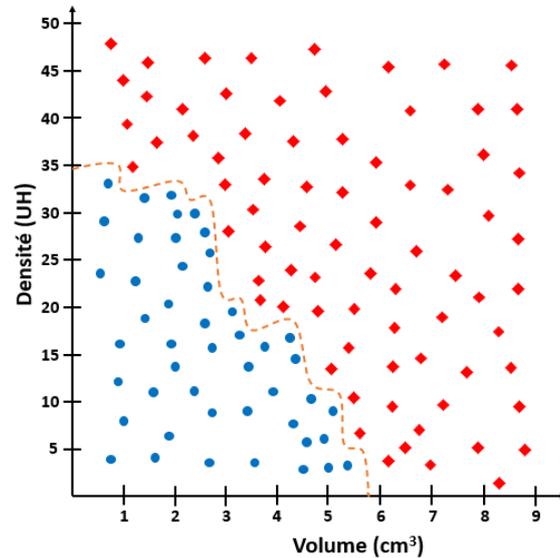


Figure 55. Exemple de frontière tracée entre les différentes classes d'un jeu de données au cours d'un apprentissage supervisé exécuté par un algorithme de forêt aléatoire

Cet algorithme procède par *bootstrap aggregation* (ou *bagging*), une technique qui, dans ce contexte, consiste à générer différents arbres de décision en considérant aléatoirement des sous-ensembles du jeu de données puis à se focaliser sur des variables prédictives spécifiques à tour de rôle (*figure 56*). Cette opération s'effectue sur l'ensemble du jeu de données [326].

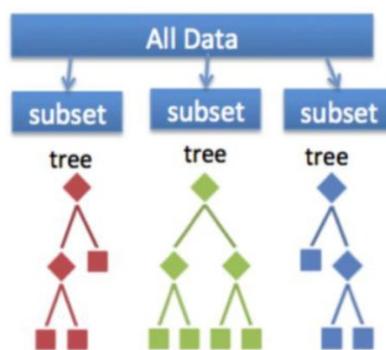


Figure 56. Illustration du *bootstrap aggregation* (Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Aditya Srivastava et coll.) [60]

Enfin, chacun de ces arbres participe dans la prédiction. Si la variable cible est quantitative, il s'agira de la moyenne de toutes les valeurs prédites par l'ensemble des arbres (*figure 57*). Si la variable cible est qualitative, il s'agira du mode de toutes les modalités prédites par l'ensemble des arbres. Comme le met en évidence la loi des grands nombres, cette approche améliore

grandement la performance de la prédiction comparée à celle d'un seul arbre de décision [327]. Il est à noter que cette augmentation de la performance se fait au détriment du coût computationnel et que celle-ci est de moins en moins significative à mesure que le nombre d'arbres augmente [328]. En outre, cet algorithme peut être adapté à toutes sortes de contextes à l'aide des nombreux hyperparamètres qui lui sont propres tels que, pour ne citer qu'eux, le nombre d'arbres et leur profondeur maximale.

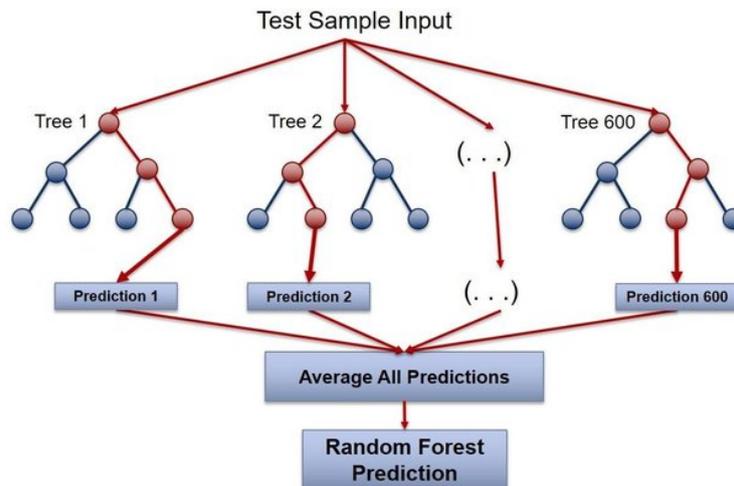


Figure 57. Illustration du fonctionnement d'un modèle de forêt aléatoire (Sandia National Laboratories) [329]

La forêt aléatoire produit des modèles puissants sans être exigeante en termes d'ajustement. De plus, ceux-ci sont peu sensibles aux différences d'échelle entre les variables. Cependant, ce que la forêt aléatoire gagne en performance, elle le perd en explicabilité. En effet, contrairement à un arbre décisionnel isolé, la forêt aléatoire produit des modèles difficilement explicables ou visualisables, et cela est d'autant plus vrai que le nombre d'arbres est important. Qui-plus-est, ses performances sont affectées lorsque le nombre de variables est important ou lorsque les données manquantes sont nombreuses. Aussi, l'entraînement d'un modèle par une forêt aléatoire peut être long et coûteux en ressources si le jeu de données est volumineux [302].

d.7. La méthode des k-moyennes

La méthode des k-moyennes (ou *k-means*) fait partie des méthodes de partitionnement de données les plus utilisées et maîtrisées à ce jour. Celle-ci repose généralement sur un algorithme d'apprentissage non supervisé, l'*algorithme de Lloyd*, mais il en existe des approches supervisées ou semi-supervisées (ex. : *k-means++*) [330,331]. Comme tous les algorithmes de partitionnement de données, le principe appliqué par l'*algorithme de Lloyd* est de constituer des clusters regroupant des observations présentant les points communs les plus significatifs.

Puisqu'il s'agit d'un algorithme d'apprentissage non supervisé, ces clusters se constituent sans que l'utilisateur ait à aiguiller l'algorithme vers ceux-ci. L'utilisateur doit cependant indiquer à l'algorithme la valeur de k qui désigne le nombre de centroïdes, ou autrement dit le nombre de clusters attendus. L'utilisateur doit veiller à choisir et ajuster soigneusement cet hyperparamètre car les résultats obtenus dépendent fortement de celui-ci. [332]

Pour illustrer la méthode des k -moyennes, considérons un jeu de données en deux dimensions dont les observations ne sont pas labellisées (*figure 58*). L'algorithme de Lloyd procéderait selon les étapes suivantes :

- 1) Au vu du nuage de points représentés, nous pouvons instinctivement envisager que trois groupes distincts s'en dégagent (*figure 58-a*). Nous fixerons donc à 3 la valeur de l'hyperparamètre k dans cet exemple.
- 2) L'algorithme procède tout d'abord en plaçant aléatoirement nos trois centroïdes (*figure 58-b*).
- 3) L'algorithme affecte ensuite chaque observation au centroïde qui lui est le plus proche, nous obtenons ainsi trois groupes correspondant à des ébauches de clusters (*figure 58-c*).
- 4) L'algorithme poursuit en calculant les barycentres des clusters constitués. Les centroïdes se déplacent alors au niveau des barycentres de leurs clusters respectifs. Nous pouvons remarquer qu'après déplacement des centroïdes, certains points verts sont plus proches du centroïde rouge (*figure 58-d*).
- 5) De nouveau, chaque observation est affectée au centroïde qui lui est le plus proche (*figure 58-e*).
- 6) Puis les positions des barycentres sont recalculées et les centroïdes se déplacent à leurs niveaux (*figure 58-f*).
- 7) Les quatrième, cinquième et sixième étapes se répètent successivement jusqu'à ce que les centroïdes soient placés de façon optimale (*figure 58-g*), nous obtenons alors les clusters finaux (*figure 58-h*). Remarque : plus une observation est proche du centroïde correspondant à son cluster, plus son appartenance à ce cluster est certaine.

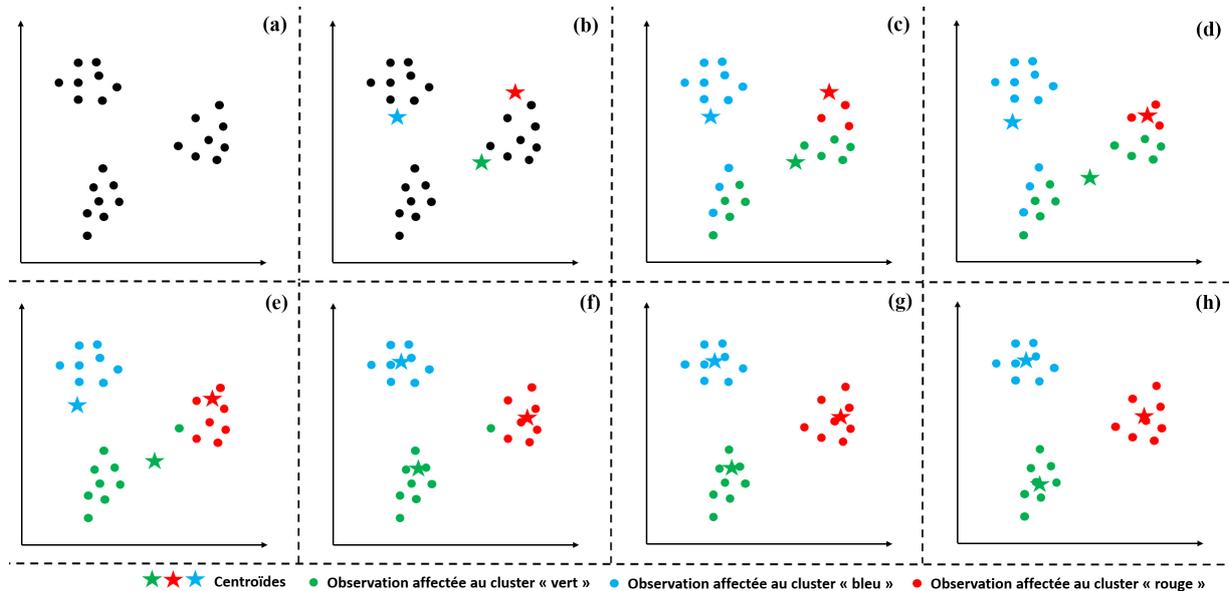


Figure 58. Partitionnement exécuté par un *algorithme de Lloyd* à partir d'un jeu de données comportant deux variables

La méthode des k-moyennes est extrêmement populaire en raison de sa simplicité, sa vitesse d'exécution et sa bonne explicabilité. De plus, l'*algorithme de Lloyd* s'adapte bien aux grands jeux de données et est globalement performant, en particulier en deux dimensions. En revanche, les résultats produits par cet algorithme sont très dépendants du nombre de centroïdes et sont sujets à un facteur aléatoire. Selon l'hyperparamètre k choisi et selon l'emplacement initial des centroïdes, les variances (distances entre les centroïdes et les observations) peuvent varier du tout au tout. Cette méthode exige donc parfois d'entraîner un modèle plusieurs fois jusqu'à obtenir des performances satisfaisantes. Enfin, une faiblesse fondamentale de l'*algorithme de Lloyd* est de ne pas être en mesure de supposer un nombre de clusters, forçant l'utilisateur à devoir le faire lui-même ou, à défaut, se tourner vers des algorithmes de partitionnement de données capables d'estimer ce paramètre (ex. : DBSCAN, *Density-Based Spatial Clustering of Applications with Noise*) [302,333].

d.8. L'analyse en composantes principales

L'analyse en composantes principales (ou PCA, *principal component analysis*) est originellement une méthode statistique permettant de réduire la dimensionnalité d'un jeu de données sans nécessiter la suppression de variables. Comme nous l'avons précédemment expliqué (*voir section P1.III.2.c.*), cette méthode procède en transposant les valeurs prises par les variables les moins informatives sur celles des dimensions conservées. Cette démarche est particulièrement utile lorsque la résolution du problème traité exige, ou est facilitée, par un faible nombre de dimensions. C'est le cas notamment des algorithmes de partitionnement de

données dont la performance a tendance à chuter lorsque le nombre de dimensions excède trois. En apprentissage automatique, l'algorithme d'analyse en composantes principales est chargé de déterminer les dimensions optimales à conserver. Il s'agit classiquement d'un algorithme d'apprentissage non supervisé mais il en existe une variante supervisée (*scalable principal component analysis*) [334,335].

L'algorithme d'analyse en composantes principales procède tout d'abord en construisant des variables synthétiques sous forme linéaire caractérisant au mieux la magnitude et la direction des variances au sein du jeu de données, ces variables synthétiques sont nommées « composantes principales ». La première composante principale correspond à la plus grande distribution au sein du jeu de données, la seconde composante principale à la seconde distribution la plus grande du jeu de données tout en étant orthogonale à la première, etc. (*figure 59*). Le nombre maximal de composantes principales pour un jeu de données est égal au nombre de dimensions (ou de variables) au sein de celui-ci [335].

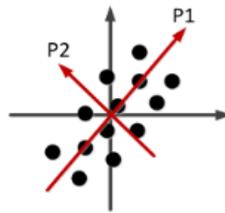


Figure 59. Composantes principales tracées sur un jeu de données en deux dimensions (*Reproduit avec la permission de l'auteur principal. Auteurs : Benjamin Ghogh et Mark Crowley.*) [335]

A titre d'exemple, considérons un jeu de données comportant deux variables, donc deux dimensions. Dans cet exemple, nous souhaiterions conserver une seule dimension (*figure 60*).

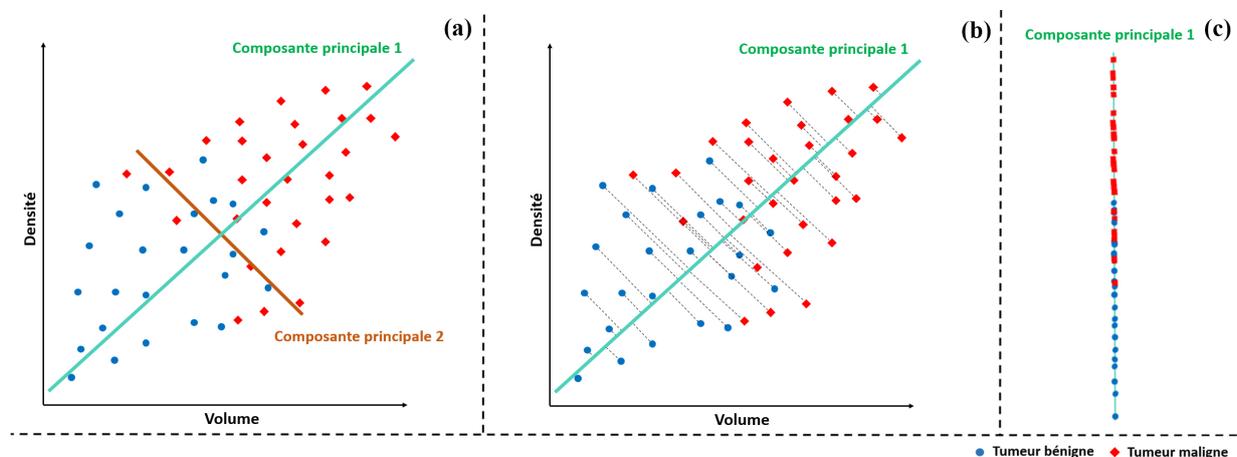


Figure 60. Réduction de dimensionnalité exécutée par un algorithme d'analyse en composantes principales sur un jeu de données en deux dimensions

L'algorithme d'analyse en composantes principales commence tout d'abord par déterminer les composantes principales qualifiant notre jeu de données. Puisque notre jeu de données comporte deux variables, nous aurons deux composantes principales représentés graphiquement par deux axes orthogonaux. En l'occurrence, la **première composante principale** tracée ci-contre représente environ 75% de la distribution tandis que la **seconde composante principale** représente environ 25% de la distribution (*figure 60-a*).

L'algorithme transpose ensuite les observations par projetés orthogonaux sur les n composantes principales les plus importantes, n étant le nombre de dimensions que nous souhaitons conserver. En privilégiant les composantes principales les plus porteuses en information, l'algorithme limite le plus possible la perte de celle-ci. Dans le cas de notre exemple, c'est donc sur la **première composante principale** que les observations seraient transposées puisque celle-ci représente la majeure partie de la distribution (*figure 60-b*). L'espace en deux dimensions devient alors un axe en une dimension (*figure 60-c*).

L'algorithme d'analyse en composantes principales est l'algorithme de réduction dimensionnelle le plus simple à paramétrer et à expliquer. Le choix du nombre de composantes principales à mobiliser est l'hyperparamètre essentiel de cet algorithme et celui-ci dépend du problème que l'on souhaite traiter. Nous chercherions, par exemple, à conserver les deux ou trois composantes principales les plus importantes s'il s'agit de préparer un jeu de données à un partitionnement. La faiblesse majeure de cet algorithme est sa forte sensibilité aux valeurs aberrantes puisque celles-ci peuvent grandement influencer les variances [302].

e. Les réseaux de neurones et l'apprentissage profond

Parmi les méthodes d'apprentissage automatique, l'apprentissage profond est sans aucun doute celle qui a le plus bénéficié des progrès réalisés en matière de puissance de calcul [336]. L'apprentissage profond s'applique sur les réseaux de neurones, des architectures qui, contrairement aux modèles d'apprentissage automatique les plus simples, permettant des analyses non linéaires [337]. En effet, au sein d'un réseau de neurones artificiels, chaque neurone est caractérisé par un seuil et une fonction d'activation qui conditionnent sa valeur de sortie. De fait, les valeurs de sortie présentées par la dernière couche du réseau ne sont pas le produit linéaire des sorties de chaque couche mais plutôt la résultante d'un ensemble complexe d'activations et de non activations [338]. Afin de mieux comprendre les tenants et aboutissants de l'apprentissage profond, commençons tout d'abord par présenter le neurone formel, l'unité élémentaire des réseaux de neurones artificiels.

e.1. Du neurone formel au réseau de neurones

Comme présenté dans notre historique, la première conceptualisation d'un neurone artificiel date de 1943 (*voir section Pl.II.1.g.*) [110]. Cette approche de l'intelligence artificielle s'est révélée être l'instigatrice de ce que représente aujourd'hui l'apprentissage automatique et par extension l'apprentissage profond. Le choix du terme « neurone », bien que discuté de nos jours, n'est à l'époque pas dû au hasard car la tentative visait à représenter mathématiquement et informatiquement un neurone biologique (*figure 61*). Petit à petit, ce paradigme biomimétique a laissé place à une approche statistique, plus robuste méthodologiquement.

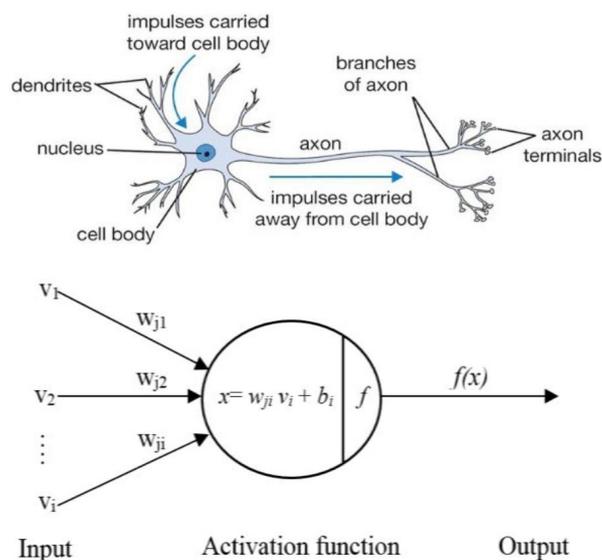


Figure 61. Parallèle entre un neurone biologique et son modèle mathématique (*Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Mian Mian Lau et King Hann Lim.*) [339]

Schématiquement, un neurone formel (ou perceptron) est une fonction dont la ou les sorties sont conditionnées par les entrées, un seuil et une fonction d'activation. Chaque entrée, généralement une grandeur numérique, est associée à une pondération (ou poids) dont la valeur varie selon la tâche à accomplir par le neurone. Chaque sortie est contrôlée par une fonction d'activation [340]. Nous pouvons schématiser le fonctionnement d'un perceptron en trois étapes (*figure 62*) :

- 1) Le calcul par une fonction d'agrégation, qui est généralement la somme des valeurs reçues en entrée (x) multipliées par leurs pondérations respectives (w) : $w_1x_1 + w_2x_2 + \dots + w_nx_n$;
- 2) Le traitement de cette somme par une fonction d'activation (ϕ) ;
- 3) La comparaison du résultat obtenu par rapport au seuil (θ) afin de définir la valeur de sortie (y) du neurone. La valeur de sortie dépend en grande partie de la fonction d'activation. Par exemple, si la fonction d'activation est une fonction de Heaviside : si

le résultat est supérieur au seuil, le neurone s'active et présente 1 comme valeur de sortie. Si le résultat est inférieur au seuil, le neurone est inactif et présente 0 comme valeur de sortie selon la fonction d'activation.

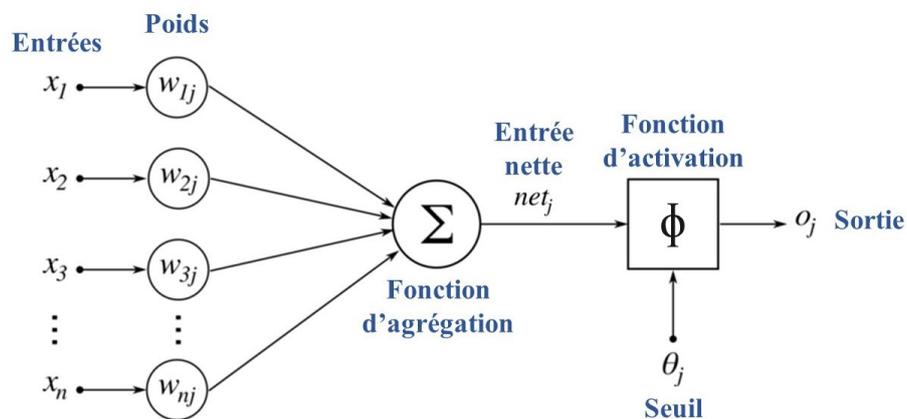


Figure 62. Représentation schématique d'un neurone artificiel (Reproduit selon les termes de la licence CC BY-SA 3.0. Auteur : Chrislb. <https://flickr.com/photos/93001633@N00/5151286161>)

En vue d'approfondir ce modèle et permettre des applications plus poussées, les travaux menés par la communauté connexionniste ont par la suite conduit aux réseaux de neurones artificiels (ou perceptrons multi-couches). Nous pouvons parler de réseau de neurones dès lors que plusieurs neurones artificiels sont interconnectés et sont organisés a minima selon une couche d'entrée, une couche de sortie et une couche intermédiaire (figure 63). La couche d'entrée a pour rôle de recevoir les données d'apprentissage ou les observations devant faire l'objet d'une prédiction. La couche de sortie, quant à elle, émet les valeurs de sortie correspondant aux prédictions effectuées. Enfin, c'est au niveau de la ou des couches intermédiaires, nommées couches cachées, que l'analyse des données d'entrée est effectuée. C'est pourquoi, l'architecture de cette partie du réseau est fondamentale. Le choix du nombre de couches, du nombre de neurones qu'elles contiennent et de la façon dont ceux-ci sont interconnectés doit être soigneusement calibré selon la problématique traitée et la nature des données d'entrée [338].

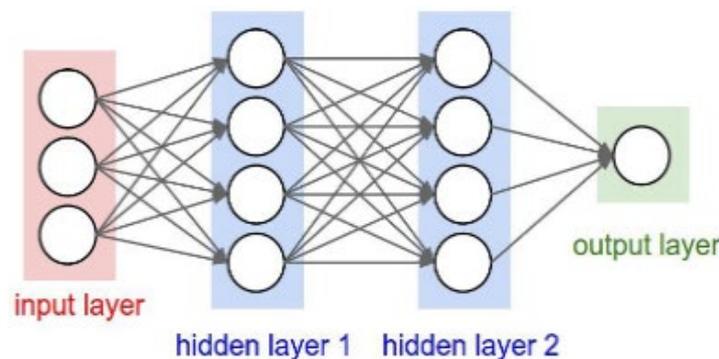


Figure 63. Représentation schématique d'un réseau de neurones artificiels (Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Aditya Srivastava et coll.) [60]

e.2. La diversité des réseaux de neurones

Les différents types de réseaux de neurones se distinguent donc selon les aspects suivants [338] :

- L'architecture du réseau en lui-même, c'est-à-dire le nombre de couches, le nombre de neurones et la topologie des synapses ;
- La fonction d'agrégation (ex. : somme pondérée), c'est-à-dire la fonction qui calcule la valeur résultant des entrées pour chaque neurone ;
- La fonction d'activation (ex. : fonction de Heaviside, fonction sigmoïde, fonction d'unité linéaire rectifiée, etc.), chargée de contrôler l'activation du neurone suite au calcul de la fonction d'agrégation et d'émettre la valeur de sortie. Usuellement, la fonction d'activation choisie est la même pour l'ensemble des neurones des couches cachées. Il est cependant commun de choisir une fonction d'activation différente pour la couche de sortie (*figure 64*).

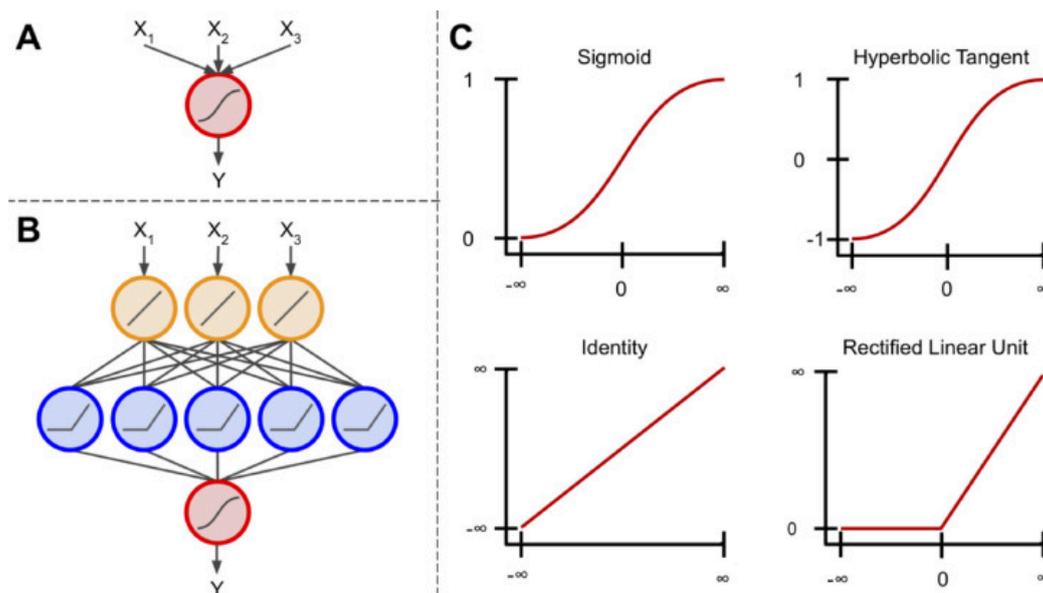


Figure 64. Composants d'un réseau de neurones ; (A) perceptron dont la fonction d'activation est sigmoïde ; (B) réseau de neurones artificiels ayant une fonction d'activation spécifique pour chaque couche ; (C) exemples de fonctions d'activation : sigmoïde (mise à l'échelle des entrées entre 0 et 1), tangente hyperbolique (mise à l'échelle des entrées entre -1 et 1), identité (linéarité) et unité linéaire rectifiée (linéarité avec annulation des entrées négatives). (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Rene Y. Choi et coll.) [45]

Cela a naturellement donné naissance à une multitude de réseaux de neurones aux structures diverses et variées, chacune étant adaptée à des cas d'usage plus ou moins spécifiques (*figure 65*) [285].

A mostly complete chart of Neural Networks

©2019 Fjodor van Veen & Stefan Leijnen asimovinstitute.org

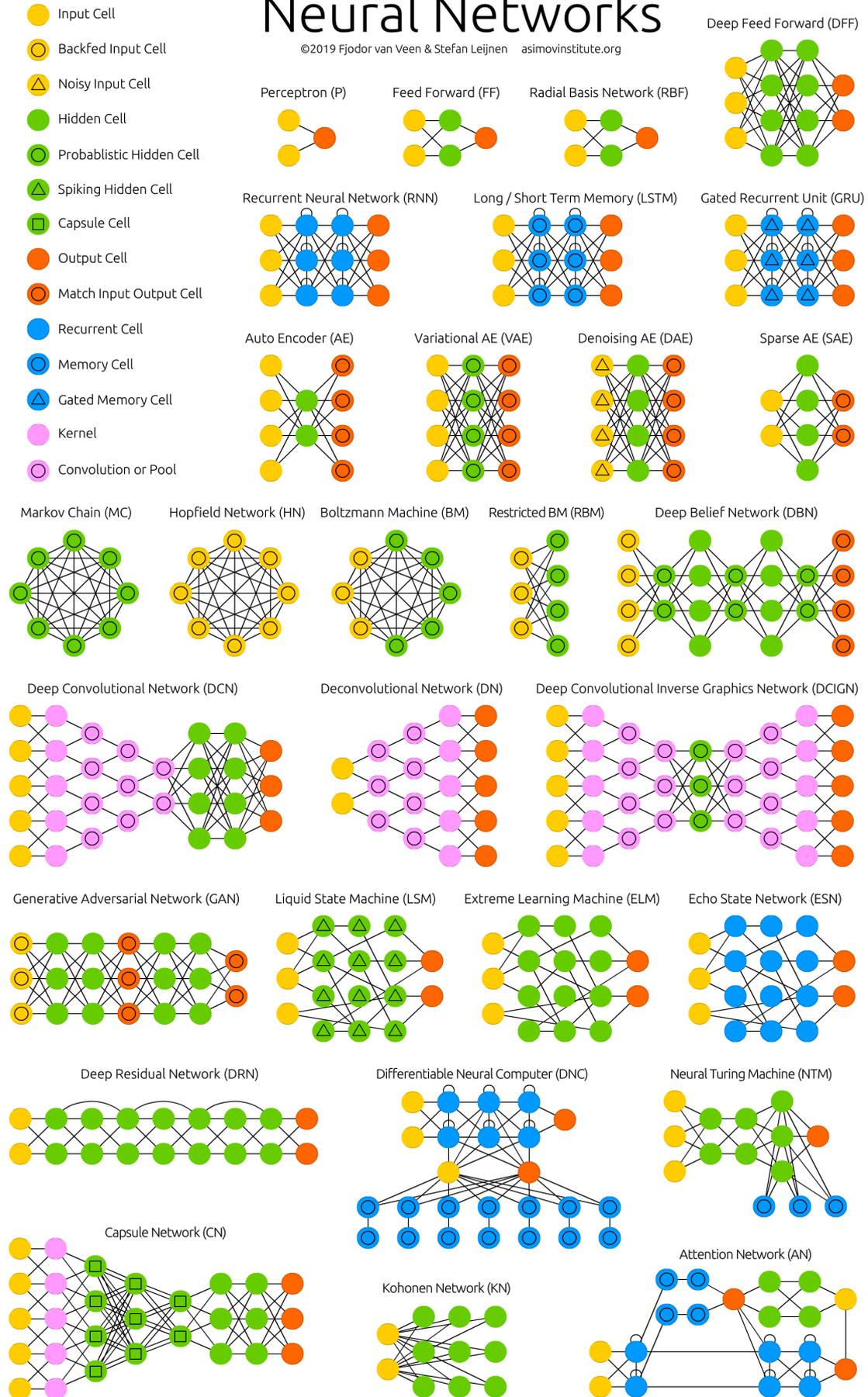


Figure 65. Variété d'architectures de réseaux de neurones artificiels (Reproduit avec la permission de l'auteur, Fjodor van Veen) [341]

Parmi les plus populaires, nous pouvons par exemple citer :

- Les réseaux de neurones convolutifs qui disposent de couches de convolution capables d'isoler les caractéristiques d'une observation en décomposant celle-ci sous forme de filtres (*figure 66*). Cette opération est non supervisée, autrement dit les caractéristiques ne sont pas prédéfinies et sont identifiées lors de l'apprentissage. Les réseaux de neurones convolutifs comportent également des couches de regroupement (*pooling*) dont le rôle est de réduire la taille des observations en privilégiant leurs caractéristiques les plus importantes. La fonction d'activation des réseaux de neurones convolutifs est classiquement la fonction d'unité linéaire rectifiée (ReLU). Les réseaux de neurones convolutifs sont largement employés pour la reconnaissance visuelle, les systèmes de recommandation et le traitement du langage naturel [338,342].

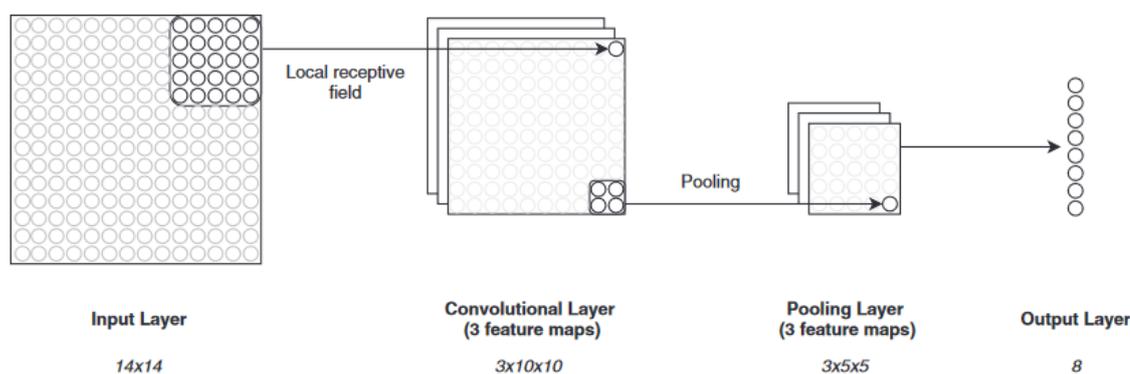


Figure 66. Représentation schématique d'une architecture de réseau de neurones profond convolutif : couche d'entrée, couches de convolution, couches de regroupement, couche de sortie
(Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Olivier Niel et Paul Bastard) [343]

- Les réseaux de neurones récurrents caractérisés par les connexions récurrentes qu'ils comportent. Il s'agit de connexions capables de rediriger la valeur de sortie d'une couche de neurones vers l'entrée de cette même couche ou d'une couche antérieure (*figure 67*). Cela permet aux réseaux de neurones récurrents de disposer en quelque sorte d'une mémoire à court terme. Qui-plus-est, les réseaux de neurones récurrents sont capables de traiter des signaux de taille très variable grâce à leur principe de « fenêtre glissante » qui consiste à analyser un signal de façon séquentielle. Les fonctions d'activations les plus communément utilisées pour les réseaux de neurones récurrents sont la fonction sigmoïde, la fonction tangente hyperbolique (Tanh) et la fonction d'unité linéaire rectifiée. Ce type de réseaux de neurones est particulièrement efficace pour l'analyse de séries temporelles, la traduction automatique ou encore la reconnaissance de forme et du langage (ex. : écriture manuscrite) [338,344].

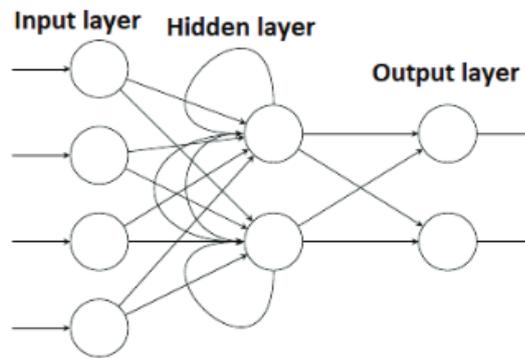


Figure 67. Représentation schématique d'une architecture de réseau de neurones récurrents
(Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Nitin Kumar Chauhan et Krishna Singh) [345]

- Les réseaux de croyance profonde, dont les couches cachées peuvent être comparées à un empilement de machines de Boltzmann restreintes. Une machine de Boltzmann restreinte est un bloc composé de deux couches de neurones fonctionnant de façon stochastique. Elles confèrent ainsi la capacité de construire une distribution de probabilité à partir des caractéristiques des données en entrée. Les réseaux de croyance profonde sont des modèles dits génératifs, c'est-à-dire capables de générer de nouveaux exemples à partir des données d'entraînement. Ce sont des modèles flexibles, pouvant être entraînés de façon supervisée ou non supervisée. En général, la fonction d'activation utilisée pour les réseaux de croyance profonde est la fonction d'unité linéaire rectifiée mais des études mettent en évidence que la fonction tangente hyperbolique convient également. Les réseaux de croyance profonde sont particulièrement utilisés dans la reconnaissance visuelle [338,339].

Ces trois exemples font partie de la famille des réseaux de neurones comportant au minimum deux couches cachées, appelés « réseaux de neurones profonds » [345]. Ceci étant dit, la construction d'un réseau de neurones n'est pas suffisante car, en l'état, ceux-ci effectueraient des prédictions éloignées de leur performance optimale. Pour atteindre leur plein potentiel, les réseaux de neurones ont besoin d'être entraînés. En l'occurrence, lorsque l'entraînement porte sur un réseau de neurones profond, nous pouvons employer les termes « apprentissage profond ».

e.3. L'apprentissage profond

Comme nous l'avons précédemment énoncé, les poids associés aux synapses d'entrée d'un neurone artificiel jouent un grand rôle dans son fonctionnement en influençant son activation et en conditionnant sa valeur de sortie. Par extension, l'ensemble des poids au sein d'un réseau de neurones détermine le fonctionnement de celui-ci ainsi que les valeurs présentées par la couche de sortie et par conséquent la qualité des prédictions. L'apprentissage appliqué aux réseaux de neurones consiste donc à ajuster ces poids de façon à ce que ceux-ci permettent au réseau d'aboutir à de meilleures prédictions. La méthode d'apprentissage typiquement employée de nos jours sur les réseaux de neurones est la rétropropagation du gradient, c'est pourquoi nous présenterons celle-ci pour illustrer l'apprentissage profond. A titre informatif, nous pouvons citer les exemples d'alternatives suivants : algorithmes évolutionnistes, *difference target propagation*, goulot de HSIC (*Hilbert-Schmidt independence criterion*), minimisation alternée en ligne avec variables auxiliaires, interfaces neuronales découplées à l'aide de gradients synthétiques, signaux d'erreur locale [159,346,347,348,349,350,351].

La rétropropagation du gradient est un algorithme d'apprentissage supervisé qui ajuste les poids selon l'erreur mesurée, c'est-à-dire le décalage entre le label des observations et la prédiction réalisée par le modèle pour ces mêmes observations. Cet ajustement se fait itérativement, à chaque passage des données d'entraînement dans le réseau. Il existe de nombreux algorithmes d'optimisation (appelés aussi « optimiseurs ») pouvant être associés à la rétropropagation du gradient, nous pouvons notamment citer la célèbre descente de gradient stochastique, l'optimiseur *Adagrad* ou encore le très populaire optimiseur *Adam*. Ces différents algorithmes s'appuient tous sur le principe de convergence vers un état optimal qui ne peut s'appliquer que sur des fonctions dérivables, ce qui est en l'occurrence le cas pour les réseaux de neurones [159].

L'algorithme de rétropropagation du gradient procède comme suit [352] :

- 1) Les poids du réseau sont initialisés, généralement de façon aléatoire.
- 2) Les données d'entraînement sont présentées à la couche d'entrée. Puisqu'il s'agit d'un apprentissage supervisé, ces données sont composées d'observations labellisées (ex. : photographies de tumeurs cutanées diagnostiquées). La couche d'entrée contient autant de neurones qu'il y a de variables à considérer pour chaque observation (ex. : les valeurs colorimétriques des pixels d'une image de tumeur cutanée).

- 3) A tour de rôle, chaque observation se propage en avant du réseau, traversant successivement les couches cachées. Selon l'observation propagée, les poids associés aux synapses à l'instant t , les fonctions d'activation et d'agrégation, cette propagation en avant va activer ou inactiver des neurones similairement à un effet domino.
- 4) La couche de sortie émet les valeurs résultant de la propagation en avant. S'il s'agit d'une classification, la couche de sortie contient autant de neurones qu'il y a de classes possibles pour la prédiction (ex. : deux neurones pour « tumeur bénigne » et « tumeur maligne », chacun émettant respectivement la probabilité associée à son diagnostic). S'il s'agit d'une régression, celle-ci contient un seul neurone ayant pour rôle d'émettre la valeur numérique prédite.
- 5) La prédiction effectuée pour chaque observation des données d'entraînement est alors comparée à son label (ex. : comparaison du diagnostic prédit par le réseau de neurones avec le diagnostic réel). L'erreur est alors calculée par une fonction d'erreur sur l'ensemble des données d'entraînement. Pour une régression, la fonction d'erreur peut par exemple être le calcul de l'erreur quadratique moyenne ou de l'erreur absolue moyenne. Pour une classification, elle peut être le calcul de l'entropie croisée binaire ou catégorique.
- 6) Enfin, l'optimiseur propage l'erreur calculée en arrière afin d'ajuster les poids des synapses et ainsi réduire la probabilité d'erreur lors des prochaines prédictions.
- 7) Le processus décrit à l'instant, que l'on nomme *epoch*, est ensuite itéré en repartant de la deuxième étape, autant de fois que l'utilisateur l'indique à l'algorithme. En théorie, le choix du nombre d'*epochs* doit être fait en vue d'obtenir des poids permettant une prédiction optimale sur des observations que le réseau de neurones n'a jamais rencontrées.

L'utilisateur dispose de trois leviers principaux pour améliorer la performance d'une rétropropagation du gradient. Tout d'abord, l'architecture et la profondeur du réseau [353]. Les réseaux de neurones profonds ont démontré leur supériorité face aux premiers réseaux de neurones comportant peu de couches, c'est l'une des raisons pour lesquelles l'apprentissage profond a le vent en poupe depuis les années 2000 [345]. La seconde raison, qui est d'ailleurs également le second levier, est la quantité de données disponibles. Plus les données utilisées pour l'apprentissage sont nombreuses, de qualité et correctement labellisées, plus les prédictions du réseau de neurones à l'issue de l'apprentissage auront tendance à être justes et précises [354].

Le troisième levier essentiel est le choix des hyperparamètres qui sont les suivants [342] :

- Les fonctions d'activation, d'agrégation et d'erreur que nous avons abordées précédemment.
- L'optimiseur et ses hyperparamètres propres.
- Le nombre d'*epochs* lors de l'apprentissage.
- Le décrochage (ou *dropout*) qui consiste à inhiber temporairement et aléatoirement, durant l'entraînement, une proportion de neurones différents à chaque *epoch*. Le but est de pousser le réseau de neurones à s'adapter et ainsi renforcer sa capacité de généralisation (figure 68). La proportion de neurones à inhiber est à indiquer par l'utilisateur [355].

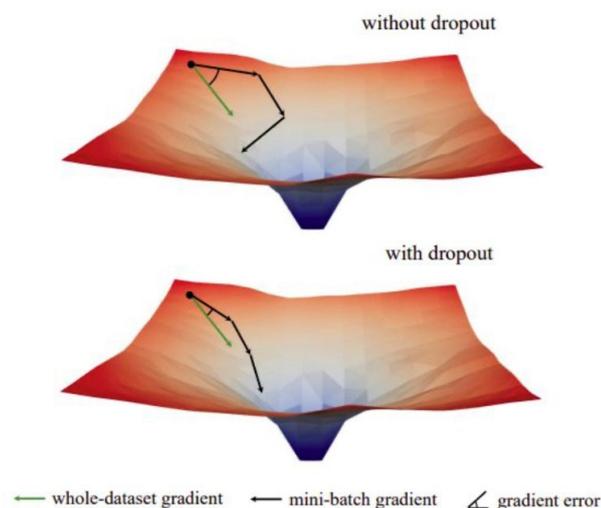


Figure 68. Effet du décrochage sur l'entraînement : le décrochage favorise la convergence de l'erreur vers un minimum global (Reproduit avec la permission de l'auteur principal. Auteurs : Zhuang Liu et coll.) [355]

- Le taux d'apprentissage (ou *learning rate*) qui contrôle le pas de la descente de gradient. Celui-ci ne doit être ni trop élevé ni trop faible pour que l'erreur (la fonction de coût) puisse converger vers un état minimal (figure 69). S'il est trop élevé, le risque est que l'algorithme d'optimisation ne parvienne pas à identifier l'intervalle de valeurs vers lequel la convergence amènerait un état optimal (minimum global de l'erreur). A l'inverse, s'il est trop faible, l'algorithme d'optimisation risque de se maintenir à un minimum local de l'erreur et de ne plus pouvoir s'en défaire.

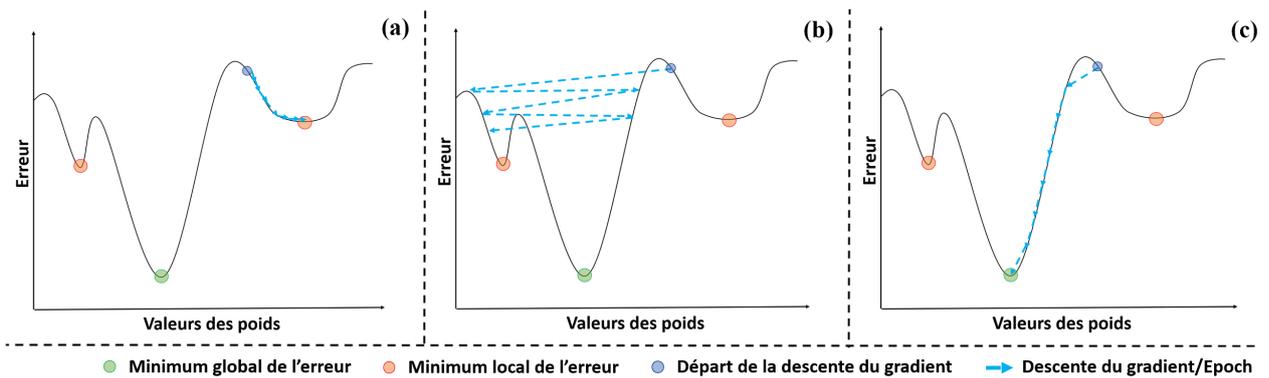


Figure 69. (a) Descente de gradient avec un taux d'apprentissage trop faible, entraînant une stagnation de l'erreur au niveau d'un minimum local ; (b) descente de gradient avec un taux d'apprentissage trop élevé, empêchant la convergence de l'erreur vers le minimum global ; (c) descente de gradient avec un taux d'apprentissage permettant la convergence de l'erreur vers le minimum global.

- Le momentum qui permet au gradient de conserver plus ou moins sa cinétique lors de sa descente. D'un état à un instant t à l'état de l'instant $t+1$ (dont la distance dépend du taux d'apprentissage), le gradient évolue. Grâce au momentum, le gradient peut conserver à $t+1$ une fraction de la cinétique qu'il possédait à l'instant t . Le momentum est en fait le pourcentage de la cinétique conservée entre les états, sa valeur est donc comprise entre 0 et 1. L'intérêt du momentum est d'assouplir la descente de gradient et d'éviter que l'erreur se maintienne à un minimum local.
- La taille du lot (ou *batch size*) qui désigne le nombre d'observations propagées à travers le réseau de neurones à chaque *epoch*. L'un des intérêts majeurs de propager un échantillon d'observations plutôt que l'ensemble du jeu de données est de mobiliser moins de puissance de calcul et ainsi de réduire le temps nécessaire à l'entraînement. Il convient toutefois de constituer des lots suffisamment importants en taille pour que l'échantillon soit représentatif.

Les modèles d'apprentissage profond sont certes complexes et leur explicabilité est limitée, néanmoins ceux-ci sont particulièrement puissants. La grande majorité des avancées réalisées en intelligence artificielle durant ces dernières années leur sont dues et c'est pourquoi l'amalgame entre intelligence artificielle et réseaux de neurones est courant de nos jours.

e.4. Les principaux champs d'application

À défaut d'être comparables à de véritables cerveaux biologiques, les réseaux de neurones parviennent néanmoins à accomplir des tâches antérieurement réalisables uniquement par des êtres vivants. Et bien qu'il s'agisse d'une sous-catégorie de l'apprentissage automatique, certains considèrent que l'apprentissage profond devrait faire l'objet d'une discipline à part entière tant les applications permises par celle-ci sont spécifiques [356].

L'un des champs d'application les plus explorés en apprentissage profond est la vision par ordinateur. Cette branche célèbre de l'intelligence artificielle porte sur la capacité des programmes à recevoir, analyser, catégoriser et générer des stimuli visuels, en particulier des images et des vidéos. L'apprentissage profond appliqué à la vision par ordinateur tente de doter des réseaux de neurones de diverses capacités, telles que par exemple la reconnaissance d'objets, la détection d'événements, le suivi vidéo, la création ou la restauration d'images et de vidéos. La santé fait partie des domaines d'application majeurs notamment pour le diagnostic mais pas uniquement (ex. : détection de tumeurs malignes sur des photographies ou des radiographies, exploration cardiovasculaire, amélioration de la qualité des radiographies, évaluation thérapeutique, détection de symptômes ophtalmologiques, etc.) [357,358,359,360,361,362].

La reconnaissance des sons est également un domaine particulièrement exploré en apprentissage profond. Sur le même principe que la vision par ordinateur, la reconnaissance acoustique étudie la capacité des programmes à recevoir, analyser, catégoriser et générer des stimuli cette fois-ci sonores. Les réseaux de neurones entraînés à la reconnaissance acoustique sont capables d'identifier des motifs au sein d'un flux sonore et d'effectuer de la classification ou, pour les modèles génératifs, de synthétiser ou restaurer des pistes sonores. En santé, la reconnaissance acoustique peut par exemple être employée pour détecter la maladie de Parkinson à partir de l'élocution du/de la patient(e) [363,364].

Enfin, nous citerons le traitement automatique du langage naturel qui se subdivise en deux catégories : la compréhension et la génération du langage naturel. Les applications en traitement automatique du langage naturel sont extrêmement variées. Son principe général est de décrypter, traduire, corriger et générer du langage naturel sous forme manuscrite, numérique ou orale. Pour ce faire, le langage naturel est prétraité à l'aide de différentes techniques : tokenisation (division du texte en plus petites unités), normalisation (standardisation des mots

suivant des règles spécifiques), suppression des *stopwords* (réduction du bruit par suppression des mots n'apportant aucune valeur sémantique), lemmatisation (réduction des mots à leur forme de base), *stemming* (réduction des mots à leur racine). Les modèles d'apprentissage profond les plus employés en traitement du langage naturel sont les transformeurs (ou modèles auto-attentifs). Les agents conversationnels par exemple s'appuient de plus en plus sur cette technologie. L'extraction et la réorganisation des données de santé textuelles non structurées serait l'une des utilisations possibles en santé [365,366].

Ces différents champs d'application s'intersectionnent souvent pour s'appliquer à des domaines très variés. Les applications sont vastes et s'enrichissent continuellement, dans le cadre de ce manuscrit nous nous focaliserons sur les cas d'usage en diagnostic et en thérapeutique.

f. Forces et faiblesses de l'apprentissage automatique

L'apprentissage automatique tire son épingle du jeu de par la notion d'apprentissage sur laquelle il se fonde. Contrairement aux systèmes experts, les modèles d'apprentissage automatique sont en mesure de généraliser et par conséquent de s'adapter au contexte à condition de leur fournir des données fiables et suffisamment nombreuses. En identifiant des motifs généralisables, ceux-ci peuvent également mettre en évidence les cas particuliers par contraste. Ces modèles peuvent traiter des signaux complexes et des données non structurées de façon autonome, ce qui leur confère dans une certaine mesure la faculté d'assimiler et exploiter des connaissances implicites contenues dans les données d'apprentissage. Qui-plus-est, la grande diversité des algorithmes et types de modèles permet un éventail d'applications extrêmement large. Les réseaux de neurones convolutifs sont par exemple particulièrement adaptés à la perception et à la classification d'images ou de sons [367].

Ces différents algorithmes ont chacun leurs forces et leurs faiblesses, et leur puissance se troque souvent contre leur explicabilité (*figure 70*). Typiquement, les réseaux de neurones profonds sur-représentent l'apprentissage automatique actuellement car leur potentiel est difficile à délimiter tant leurs limites opérationnelles ne cessent d'être repoussées. En revanche, bien que leur fonctionnement mathématique soit dument étudié, celui-ci reste à ce jour partiellement opaque. Ce problème de boîte noire impacte l'intérêt pratique des modèles qui en souffrent lorsque la prise de décision doit être précisément décortiquée (ex. : les domaines de la santé, de la justice, de la sécurité ou encore de la finance) [368].

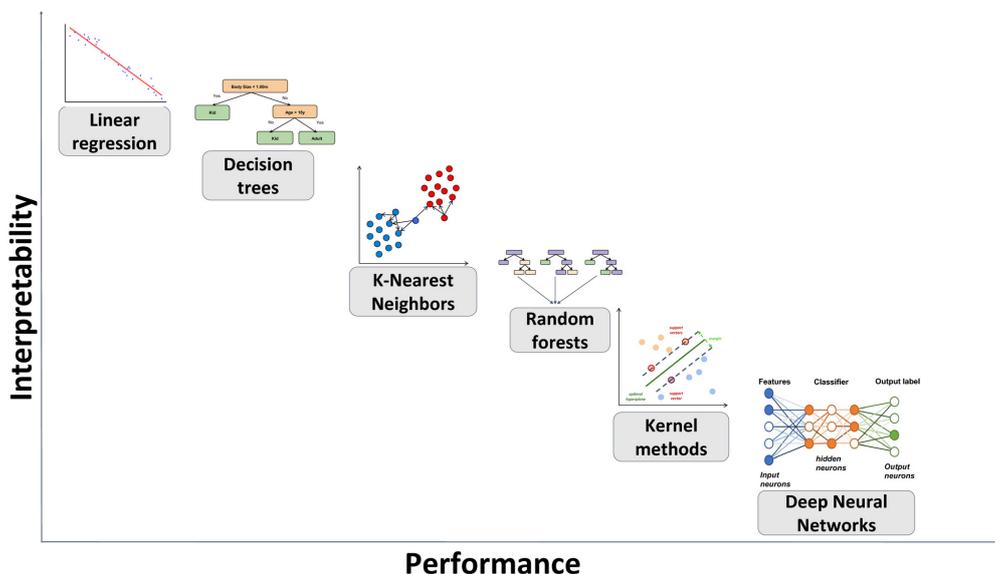


Figure 70. Vue d'ensemble de l'explicabilité et de la performance des algorithmes d'apprentissage automatique (Reproduit avec la permission de l'éditeur John Wiley and Sons. Auteurs : S. Badillo et coll.) [958]

Un autre problème majeur à soulever pour l'apprentissage automatique est ce que nous pourrions nommer le revers de la généralisation. L'apprentissage automatique est en effet incapable de conceptualiser, celui-ci se fonde sur une approche intégralement probabiliste et basée sur les données d'apprentissage. Un entraînement réalisé avec des données trop peu nombreuses ou non représentatives est fatalement exposé au risque de produire un modèle aux prédictions biaisées et inconsistantes [369]. Pour être performant, l'apprentissage automatique a donc nécessairement besoin de grandes quantités de données, ce qui le place en porte-à-faux lorsque la problématique à traiter ne dispose que de peu de données (ex. : maladies rares). De plus, le principe de généralisation expose également l'apprentissage automatique au phénomène du cygne noir, autrement dit les modèles sont très mauvais, voire parfaitement inopérants, lorsqu'il s'agit de prédire un évènement très improbable [370]. L'apprentissage automatique n'est donc pas non plus adapté à des problématiques nécessitant d'anticiper des événements rares et dont la logique est trop complexe (ex. : sécurité nationale ou certaines spécialités médicales telles que la réanimation).

3. L'intelligence artificielle neuro-symbolique : le meilleur des deux mondes ?

Bien que les deux points de vue de l'intelligence artificielle développés précédemment soient différents, ceux-ci ne sont pas antinomiques. De plus en plus sont envisagés des systèmes combinant la logique formelle du symbolisme et la capacité d'apprentissage du connexionnisme, c'est ce que l'on nomme l'intelligence artificielle neuro-symbolique (ou hybride) [371].

L'intérêt majeur de l'intelligence artificielle neuro-symbolique est de s'affranchir des compromis entre la puissance et l'explicabilité ainsi qu'entre la formalisation de la logique et l'apport de données (figure 71). En théorie, une intelligence artificielle neuro-symbolique devrait réunir les qualités des systèmes experts et de l'apprentissage automatique tout en limitant leurs défauts. Cela permettrait au système de bénéficier de la puissance et de la possibilité d'analyser des données non structurées offerte par l'apprentissage profond ainsi que du raisonnement structuré et de l'explicabilité caractérisant les systèmes experts. Le système peut également faire usage des données nouvelles et des méthodes statistiques propres à l'apprentissage automatique pour faire évoluer sa base de connaissances et ses règles de décision. Enfin, la quantité de données nécessaire pour permettre des prédictions de qualité est moins importante comparée à l'entraînement d'un réseau de neurones profond [74].

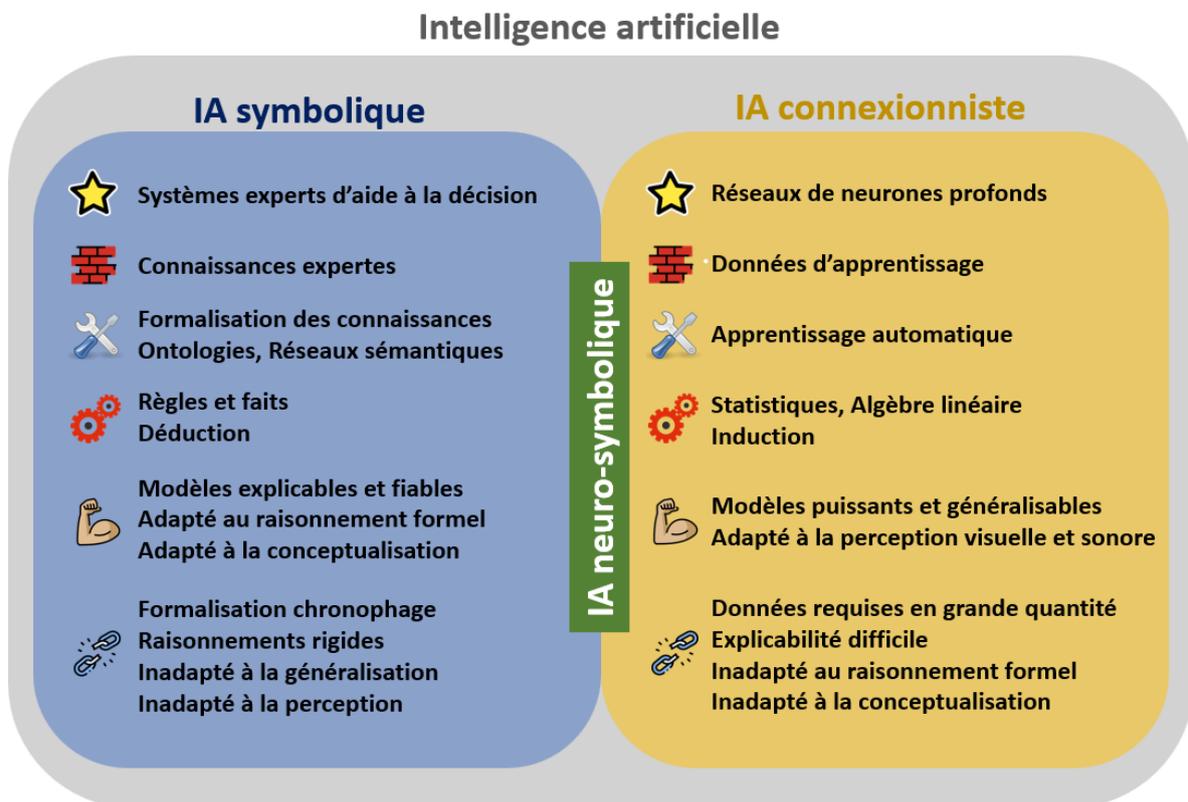


Figure 71. Comparaison des courants symbolique et connexionniste de l'intelligence artificielle. Le courant neuro-symbolique se situe à l'interstice de ces deux approches.

Selon plusieurs experts en la matière, les architectures neuro-symboliques seraient à ce jour la voie la plus prometteuse pour l'intelligence artificielle [372]. De par leur potentiel, celles-ci permettraient de repousser significativement ses capacités opérationnelles et potentiellement de tendre vers l'intelligence artificielle générale. C'est pourquoi, le neuro-symbolisme représente un enjeu majeur de la recherche dans ce domaine [373].

PARTIE 2 : Intelligences artificielles appliquées au diagnostic et à la thérapeutique

PARTIE 2 : Intelligences artificielles appliquées au diagnostic et à la thérapeutique

Comme nous l'avons exposé au cours du premier volet de ce manuscrit, les intelligences artificielles permettent des applications variées et dont l'étendue du potentiel reste à ce jour inconnue. Il en va naturellement de même pour les applications en santé qui s'étendent d'ailleurs au-delà du diagnostic et de la thérapeutique (ex. : prédiction et gestion d'une pandémie) [374]. Nous nous attacherons cependant à présenter des cas d'usage uniquement en diagnostic et en thérapeutique, ce qui représente déjà un défi tant il est difficile de sélectionner les plus pertinents à commenter parmi la multitude de ceux existants [375]. De fait, cette démarche n'a pas pour but d'être exhaustive, et ne pourrait définitivement pas l'être au vu de l'évolution quotidienne de l'état de l'art. Nous nous limiterons donc, au plus, à un cas d'usage en diagnostic et un cas d'usage en thérapeutique pour chaque spécialité abordée.

I. Cas d'usage

1. Définitions

Avant de présenter des cas d'usage en diagnostic et thérapeutique, il convient tout d'abord de définir brièvement ces termes.

L'Académie française propose les définitions du diagnostic suivantes : « *Identification d'une maladie chez un patient d'après l'étude des symptômes et des signes, les résultats de divers examens, etc.* » ou « *Art d'identifier une maladie.* » [376].

La thérapeutique est quant à elle définie, toujours par l'Académie française, comme suit : « *Partie de la médecine qui a pour objet la manière de traiter, de soigner et de guérir les maladies.* » [377].

Enfin, un cas d'usage (ou cas d'utilisation) désigne une manière d'employer un système ayant une valeur ou une utilité pour les acteurs impliqués [378]. Dans notre contexte, cela fait donc référence à l'application d'une technique d'intelligence artificielle dans un but diagnostic ou thérapeutique [379].

2. Addictologie

a. Cas d'usage en diagnostic

Wajid Mumtaz et son équipe présentent en 2016 un modèle d'intelligence artificielle visant à détecter les troubles liés à la consommation d'alcool [380].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de régression logistique. Les données d'entraînement sont constituées de caractéristiques extraites d'électroencéphalographies au repos labellisées selon que le/la patient(e) présente (30) ou non (15) des troubles liés à la consommation d'alcool.

Il s'agit donc d'un modèle de classification prenant en entrée des caractéristiques extraites d'une électroencéphalographie et indiquant en sortie la classe prédite : présence de troubles liés à la consommation d'alcool *vs.* absence de troubles liés à la consommation d'alcool. Ce modèle parvient à atteindre des performances élevées^{xxvi} : exactitude = 89,3 % ; sensibilité = 88,5 % ; spécificité = 91 % ; F-score = 90%.

Le diagnostic des troubles liés à la consommation d'alcool est souvent compliqué en raison de la dimension subjective du processus [381]. La fiabilité du diagnostic serait grandement favorisée par des méthodes objectives et robustes. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, les électroencéphalographies pourraient potentiellement être considérés en tant qu'éléments objectifs en mesure de renforcer la justesse du diagnostic des troubles liés à la consommation d'alcool, de façon automatique, précise et non invasive [380].

b. Cas d'usage en thérapeutique

Laura Acion et son équipe présentent en 2017 des modèles d'intelligence artificielle visant à prédire la réussite d'un traitement contre les troubles liés à la consommation de substances [382].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : régression logistique, régression pénalisée, forêt aléatoire, apprentissage profond, super apprentissage. Les données d'entraînement comportent 28 variables socio-démographiques et cliniques issues d'une base de données collectées auprès de 79210 individus. Ces données ont été labellisées selon la réussite ou l'échec du traitement.

Il s'agit donc de modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite ainsi que sa

^{xxvi} Pour plus de détail concernant les métriques d'évaluation, se référer à la section III.B.2..

probabilité associée : réussite du traitement *vs.* échec du traitement. Le modèle de super apprentissage est celui présentant les meilleures performances : aire sous la courbe de ROC = 82 %.

Le parcours de soins des patients souffrant de troubles liés à la consommation de substances fait partie des cadres nécessitant une médecine de précision [383]. La possibilité de prédire la réussite ou l'échec d'un traitement contre les troubles liés à la consommation de substances permettrait aux soignants de prescrire les traitements les plus adaptés à chaque patient(e) et ainsi améliorer leur taux de réussite. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire la réussite ou l'échec d'un traitement contre les troubles liés à la consommation de substances sur un(e) patient(e) donné(e), de façon automatique et non invasive [382].

3. Anatomie pathologique (cas d'usage en diagnostic)

Babak Ehteshami Bejnordi et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à détecter les stromata entourant un cancer invasif [384].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif. Les données d'entraînement sont constituées d'images numérisées de coupes tissulaires colorées (à l'hématoxyline et à l'éosine) de biopsies mammaires collectées auprès de 552 patientes âgées de 40 à 65 ans. Ces images ont été labellisées selon que la biopsie indique ou non la présence d'un cancer invasif.

Il s'agit donc d'un modèle de classification prenant en entrée une image de coupe tissulaire colorée de biopsie mammaire et indiquant en sortie la classe prédite : stroma entourant un cancer invasif *vs.* stroma n'entourant pas un cancer invasif. Ce modèle parvient à atteindre des performances élevées : aire sous la courbe de ROC = 96,2 %.

Le cancer du sein est fortement caractérisé par son microenvironnement stromal qui conditionne son développement, sa croissance et ses métastases [385]. Les pathologistes sont en mesure de détecter ces variations morphologiques par microscopie optique, cependant la classification visuelle de ces changements est subjective et imprécise. La fiabilité du dépistage du cancer du sein à partir d'un stroma serait donc grandement favorisée par des méthodes objectives et robustes. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, les images de coupe tissulaire de biopsie mammaire pourraient potentiellement être considérées en tant qu'éléments objectifs en mesure de renforcer la justesse du diagnostic des cancers du sein, de façon automatique, rapide et précise [384].

4. Andrologie

a. Cas d'usage en diagnostic

Guanjin Wang et son équipe présentent en 2018 des modèles d'intelligence artificielle visant à détecter les cancers de la prostate [386].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : machine à vecteurs de support, forêt aléatoire, apprentissage profond. Les données d'entraînement comportent des informations pré-biopsie (âge, examen rectal, échographie transrectale, volume de la prostate) ainsi que des taux de PSA issus d'une base de données de biopsies transrectales échoguidées. Celles-ci ont été labellisées après avoir été collectées auprès de 828 patients présentant un cancer de la prostate avéré, 539 patients présentant un cancer de la prostate non avéré (non significatif) et 258 patients contrôle en bonne santé.

Il s'agit donc de modèles de classification prenant en entrée les informations pré-biopsie susmentionnées ainsi qu'un taux de PSA et indiquant en sortie la classe prédite : absence de cancer de la prostate ou présence d'un cancer de la prostate non avéré *vs.* présence d'un cancer de la prostate avéré. Le modèle de réseau de neurones est celui présentant les meilleures performances : exactitude = 95,27 % ; sensibilité = 99,96 % ; spécificité = 90,35 % ; F-score = 95,58% ; aire sous la courbe de ROC = 97,55 %.

L'un des leviers importants dans le traitement du cancer de la prostate est son diagnostic précoce [387]. Malgré la fiabilité de la biopsie, il s'agit d'un examen invasif sujet à des effets secondaires et des risques. Une méthode de diagnostic moins invasive permettrait d'améliorer le rapport bénéfice/risque, plus particulièrement chez les patients les moins susceptibles de développer un cancer. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, des informations pré-biopsie et le taux de PSA pourraient potentiellement être considérés en tant qu'éléments suffisants pour exclure la présence d'un cancer de la prostate, de façon automatique, précise et non invasive [386].

b. Cas d'usage en thérapeutique

Alexandru Nicolae et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à proposer des programmes de curiethérapie à faible dose en traitement du cancer de la prostate [389].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé. Les données d'entraînement sont constituées de caractéristiques extraites de fichiers DICOM (*Digital*

Imaging and Communications in Medicine) collectées auprès de 100 patients ayant un cancer de la prostate et traités par curiethérapie à faible dose. Ces fichiers comportent notamment des informations pré-opératoires : contours anatomiques en trois dimensions, modèles de la source de radiation et de l'aiguille, le placement des patrons, les propriétés des radionucléides. Les caractéristiques extraites sont également soumises à une optimisation stochastique et à des critères de qualité clinique (*figure 72*).

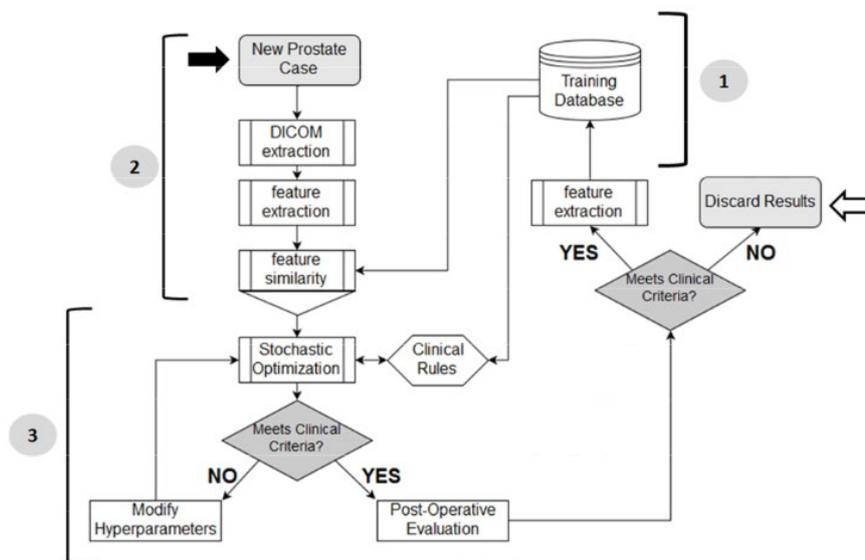


Figure 72. Processus d'apprentissage automatique. La flèche noire indique le point de départ de la boucle, la flèche blanche indique sa fin ; (1) Prétraitement des données et apprentissage ; (2) Mesure de similarité ; (3) Optimisation stochastique et adaptation. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Alexandru Nicolae et coll.) [389]

Il s'agit d'un modèle prenant en entrée les caractéristiques extraites d'un fichier DICOM et indiquant en sortie un programme de curiethérapie à faible dose en traitement du cancer de la prostate. Les comparaisons statistiques menées entre les programmes établis par le modèle et ceux établis par des curiethérapeutes expérimentés n'indiquent pas de différence significative en termes de pertinence clinique. Celles-ci montrent en revanche une différence significative en termes de vitesse d'exécution en faveur du modèle d'apprentissage automatique (50 secondes en moyenne).

La forte incidence de cancer de la prostate, le second cancer le plus fréquent chez l'homme mondialement, représente un challenge pour les services de curiethérapie manquant de ressources humaines ou de personnel expérimenté [390]. Des méthodes fiables, rapides et automatiques de conception de programmes de curiethérapie s'avèreraient donc bénéfiques pour les curiethérapeutes et leurs patients. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait en mesure de fournir aux soignants des programmes fiables de curiethérapie à faible dose, de façon automatique, rapide et non invasive [389].

5. Anesthésie-réanimation

a. Cas d'usage en diagnostic

Gabriel Louis Cuendet et son équipe présentent en 2016 un modèle d'intelligence artificielle visant à détecter les intubations endotrachéales difficiles chez des patients anesthésiés [391]. Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de forêt aléatoire. Les données d'entraînement sont constituées de caractéristiques anatomiques et morphologiques extraites de photos et vidéos de visages collectées auprès de 496 patients s'appêtant à recevoir une anesthésie générale nécessitant une intubation trachéale (*figure 73*). Ces données ont été labellisées selon que le/la patient(e) ait été sujet(te) à une intubation endotrachéale facile (448) ou difficile (48).



Figure 73. Exemple de caractéristiques (en vert) anatomiques et morphologiques constituant les données d'entraînement (*Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Gabriel Louis Cuendet et coll.*) [391]

Il s'agit donc d'un modèle de classification prenant en entrée des caractéristiques extraites d'une photographie de visage de face et indiquant en sortie la classe prédite : intubation endotrachéale facile vs. intubation endotrachéale difficile. Ce modèle parvient à atteindre des performances comparables à l'état de l'art : aire sous la courbe de ROC = 81 %. Le temps d'exécution moyen du modèle est de l'ordre de la seconde.

L'intubation trachéale difficile fait partie des causes majeures de complications liées à l'anesthésie générale [392]. La possibilité d'évaluer en préopératoire le risque d'intubation difficile serait donc fortement bénéfique pour la sécurité des patients. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de détecter un risque d'intubation difficile chez un(e) patient(e) de façon automatique, rapide et non invasive [391].

b. Cas d'usage en thérapeutique

Mehran Pesteie et son équipe présentent en 2015 un modèle d'intelligence artificielle visant à proposer des protocoles optimaux pour les infiltrations épidurales et facettaires [393].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones. Les données d'entraînement sont constituées de caractéristiques extraites d'échographies spinales en trois dimensions collectées auprès de 13 volontaires en bonne santé, âgés de 20 à 35 ans et sans antécédents de pathologie ou chirurgie spinales (*figure 74*).

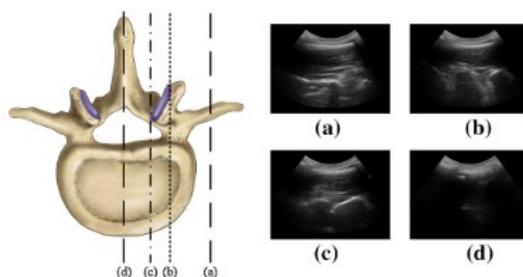


Figure 74. Plans de coupe échographique d'une apophyse transverse (a), d'une articulation facettaire (b), d'une lame vertébrale (c) et d'une apophyse épineuse (d). Les plans paramédians ciblés pour les infiltrations facettaires et épidurales sont respectivement (b) et (c). (Reproduit avec la permission de l'éditeur Springer Nature. Auteurs : Mehran Pesteie et coll.) [393]

Il s'agit d'un modèle de classification prenant en entrée une échographie spinale en trois dimensions et indiquant en sortie les sites d'injection pouvant être ciblés. Ce modèle parvient à atteindre des performances élevées : exactitude = 95 % ; sensibilité = 93 % ; spécificité = 97 % pour les sites épiduraux / exactitude = 94 % ; sensibilité = 96 % ; spécificité = 93 % pour les articulations facettaires. Le temps d'exécution moyen du modèle est de 20 millisecondes.

Les infiltrations épidurales et facettaires ont de multiples indications [394]. Pratiquer cet acte implique l'insertion percutanée d'une aiguille entre ou autour des vertèbres. Lorsque cet acte est réalisé à l'aveugle ou par palpation, celui-ci présente des risques importants en raison du manque de précision de la méthode [395]. L'acte peut également être guidé à l'aide d'une fluoroscopie, qui comporte les risques propres à l'exposition aux rayonnements ionisants [396]. Il en est de même pour la tomodensitométrie qui expose le/la patient(e) aux risques propres aux rayons X [397]. C'est pourquoi, les échographies spinales sont de plus en plus employées pour identifier des sites d'injection [398]. Néanmoins, l'anatomie complexe de la colonne vertébrale rend l'identification à l'œil nu difficile à partir d'une échographie [399]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à identifier des sites d'injection à partir d'échographies spinales, de façon automatique, rapide, précise et non invasive [393].

6. Angiologie

a. Cas d'usage en diagnostic

Jelmer M. Wolterink et son équipe présentent en 2016 un modèle d'intelligence artificielle capable d'estimer un score calcique coronaire [359,400].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif. Les données d'entraînement comportent des coronarographies collectées auprès de 250 patients. Les pixels représentant des calcifications coronaires ont été annotés numériquement sur celles-ci.

Il s'agit donc d'un modèle de segmentation prenant en entrée une coronarographie et indiquant en sortie une estimation de la localisation des calcifications coronaires ainsi que du score calcique coronaire (figure 75). Ce modèle parvient à atteindre les performances suivantes lorsqu'il estime la localisation des calcifications coronaires : sensibilité = 71 % ; nombre de faux positifs (pixels) par coronarographie = 48%. Concernant l'estimation du score calcique, le modèle attribue 83% des patients dans la même catégorie de risque cardiovasculaire que l'examen de référence (tomodensitométrie). Le temps d'exécution moyen du modèle est de l'ordre de la minute.

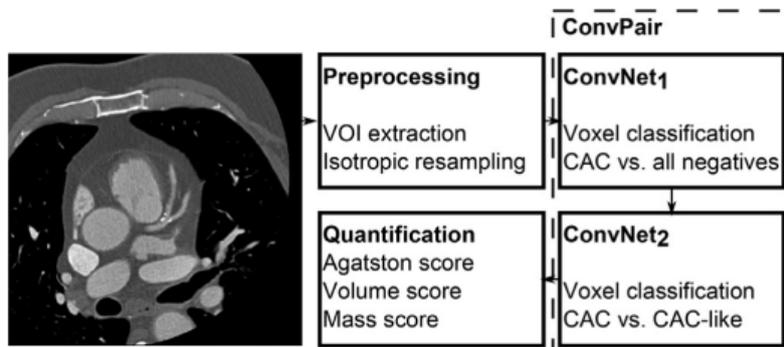


Figure 75. Représentation schématique de l'architecture présentée. L'image d'entrée est prétraitée, les voxels (groupes de pixels) indiquant potentiellement la présence d'une calcification coronaire sont extraits puis classifiés par une paire de réseaux de neurones convolutifs (*ConvNet₁* et *ConvNet₂*). Enfin, l'architecture détermine un score calcique pour chaque voxel présentant une calcification coronaire. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Jelmer M. Wolterink et coll.) [359]

Le score calcique est un prédicteur robuste d'évènements cardiovasculaires. Celui-ci est généralement estimé à partir d'une tomodensitométrie avant de poursuivre l'exploration par d'autres examens cliniques [401]. La possibilité d'estimer le score calcique à partir d'une coronarographie permettrait de se passer de la tomodensitométrie et ainsi réduire la durée d'exposition du/de la patient(e) aux rayons X [397]. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, les coronarographies pourraient potentiellement permettre d'estimer le score calcique de façon automatique, rapide et non invasive [359].

b. Cas d'usage en thérapeutique

Yingjing Feng et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à proposer un guidage durant une ablation par cathéter [402].

Ce modèle est obtenu à l'issue d'un apprentissage automatique actif (semi-supervisé) suivant un processus gaussien. Les données d'entraînement sont constituées de points de cartographies cardiaques en trois dimensions ainsi que des valeurs de voltage associées à ces points (*figure 76*). Ces données ont été collectées à partir de 25 cartographies de ventricule droit réalisées en vue d'une ablation cardiaque sur des patients atteints d'une tétralogie de Fallot.

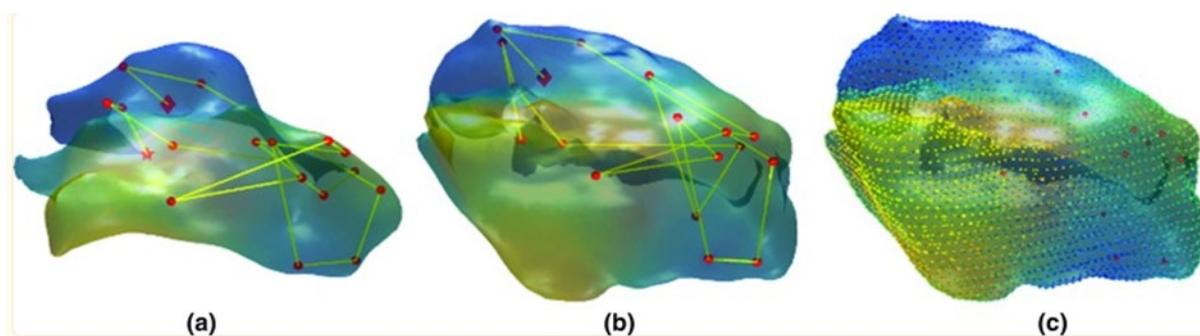


Figure 76. Points (rouges) de cartographie issus des données d'entraînement. Les couleurs en surface représentent les valeurs de tension estimées par un système de cartographie cardiaque. La séquence de cartographie et les valeurs de tension du maillage préopératoire en (a) ont toutes deux été projetées sur le maillage préopératoire en (b) à l'aide d'un k-NN. En (c), les valeurs de tension estimées (couleurs des sommets) sont similaires aux valeurs de tension exportées du système de cartographie cardiaque (couleurs de surface). (*Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Yingjing Feng et coll.*) [402]

Il s'agit d'un modèle de régression prenant en entrée des points de cartographie cardiaque en trois dimensions et indiquant en sortie les voltages correspondant à ces points. Ce modèle parvient à atteindre de meilleures performances, en termes de précision et de temps d'exécution, qu'une méthode rivale basée sur la géométrie.

La précision d'une cartographie cardiaque et la localisation des points gâchettes sont des éléments essentiels qui conditionnent la qualité d'une ablation par cathéter [403]. Cependant, le caractère chronophage de cette démarche peut exposer les patients au risque d'évènement cardiaque avant que celle-ci soit achevée [404]. De plus, l'identification des points gâchettes peut s'avérer complexe lorsque le/la patient(e) est atteint(e) d'une malformation cardiaque [405]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de proposer un guidage efficace durant une ablation par cathéter, manuelle ou robotique, en établissant une cartographie cardiaque de façon automatique, rapide, précise et non invasive [402].

7. Biologie médicale (cas d'usage en diagnostic)

Edward K. Chang et son équipe présentent en 2016 un modèle d'intelligence artificielle visant à identifier des patients atteints de cirrhose [406].

Ce modèle est obtenu à l'issue d'un apprentissage automatique réalisé par un algorithme de traitement du langage naturel. Les données d'entraînement sont constituées de données cliniques, biologiques et administratives issues de dossiers de santé numériques. Ces derniers étaient affiliés à des patients atteints de pathologies diverses.

Il s'agit d'un modèle de classification prenant en entrée des données sous forme de texte d'un ou plusieurs patients (ex. : bilans biologiques numérisés) et indiquant en sortie les patients pour lesquels le modèle prédit la présence d'une cirrhose. Ce modèle parvient à atteindre des performances élevées : valeur prédictive positive = 91,78 % ; valeur prédictive négative = 96,84 % ; sensibilité = 95,71 % ; spécificité = 93,88 %.

L'identification de certaines pathologies non diagnostiquées est possible au sein d'une population de patients en se basant uniquement sur leurs données de santé numérisées [407]. Cela représente notamment un intérêt lorsque le pronostic de la pathologie en question est favorisé par un diagnostic précoce. Cependant, passer manuellement en revue ces données est chronophage et fastidieux. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'identifier les patients touchés par une pathologie ciblée à partir de données de santé sous forme de texte (ex. : bilans biologiques numérisés), de façon automatique et précise [406].

8. Cardiologie

a. Cas d'usage en diagnostic

Lejla Divović Mustafić et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à détecter les sténoses aortiques sévères [408,409].

Ce modèle est obtenu à l'issue d'un apprentissage supervisé réalisé sur un réseau de neurones. Les données d'entraînement comportent 12 variables cliniques, anatomiques et socio-démographiques. Ces données ont été labellisées après avoir été collectées auprès de 45 patients présentant une sténose aortique et de 24 patients contrôle en bonne santé.

Il s'agit donc d'un modèle de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : présence d'une sténose aortique *vs.* absence d'une sténose aortique. Ce modèle parvient à atteindre d'excellentes performances : exactitude = 100 %.

La sténose aortique fait partie des pathologies cardiovasculaires les plus courantes et le traitement de sa forme sévère nécessite une intervention chirurgicale [410]. En 2018, les statistiques disponibles indiquaient que la sténose aortique était sous-diagnostiquée [411]. Pourtant, sa prise en charge tardive impacte négativement le pronostic vital des patients [412]. Cependant, plusieurs facteurs peuvent rendre difficile ou compromettre le diagnostic [413]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer les sténoses aortiques, de façon automatique, précise et non invasive [408,409].

b. Cas d'usage en thérapeutique

Dennis Medved et son équipe présentent en 2018 une architecture d'intelligence artificielle visant à prédire la survie à court et à long terme suite à une transplantation cardiaque [414].

Cette architecture est obtenue à l'issue d'un apprentissage automatique profond réalisé sur un ensemble de réseaux de neurones. Les données d'entraînement comportent 32 variables relatives aux receveurs d'organe ainsi que 11 variables relatives aux donneurs d'organe. Ces données ont été collectées auprès de 27705 patients donneurs ou receveurs d'organe et ont été labellisées selon que le/la receveur(se) survive ou décède à court (une année après la transplantation) et à long terme.

Il s'agit donc d'une architecture composée de modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite à court et à long terme : survie du/de la patient(e) *vs.* décès du/de la patient(e). Celle-ci parvient à atteindre des performances modestes mais néanmoins supérieures à l'état de l'art : aire sous la courbe de ROC = 65,4 % ; C-index = 62,7 %.

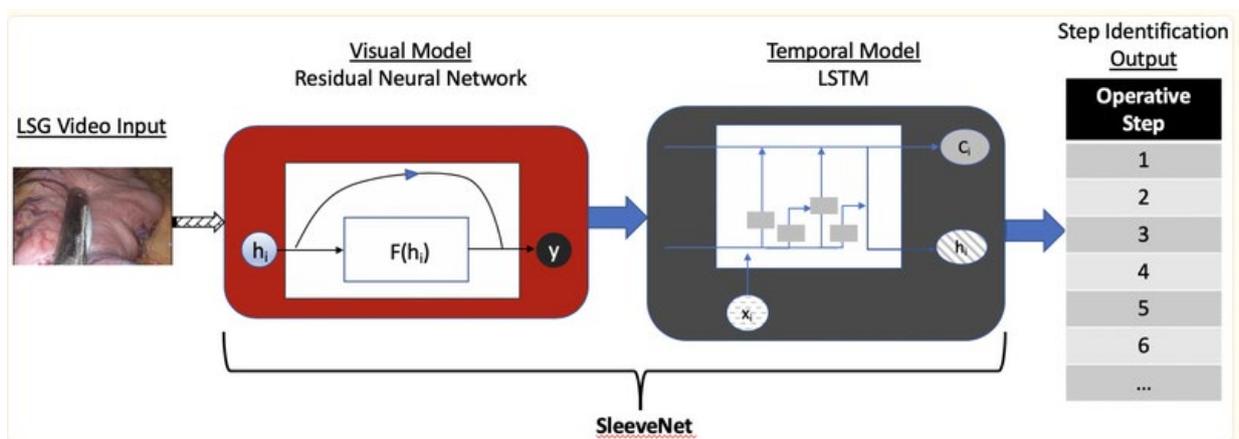
Les transplantations cardiaques permettent de prolonger la vie des patients souffrant d'une pathologie cardiaque en phase terminale [415]. Cependant, la quantité limitée de greffons disponibles restreint le nombre de patients pouvant bénéficier d'une transplantation [416]. Qui-plus-est, le succès de cette dernière est soumis à plusieurs facteurs [417]. La possibilité de prédire la réussite de la transplantation, pour un donneur et un receveur donnés, permettrait d'optimiser l'allocation des greffons et ainsi garantir de meilleures chances de survie pour les patients transplantés. Aussi, cette prédiction aiderait les soignants à anticiper les dysfonctions du greffon et à améliorer leur prise en charge. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire la survie à court et à long terme après une transplantation cardiaque, de façon automatique et non invasive [414].

9. Chirurgie (cas d'usage en thérapeutique)

Daniel A. Hashimoto et son équipe présentent en 2020 une architecture d'intelligence artificielle capable de détecter les différentes étapes d'une sleeve gastrectomie par laparoscopie [418].

Cette architecture est obtenue à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisés sur des réseaux de neurones résiduels (*ResNet-18* et mémoire longue à court terme). Les données d'entraînement sont constituées de vidéos montrant l'entièreté d'une sleeve gastrectomie par laparoscopie filmées lors de 88 opérations. Ces vidéos ont été annotées par des chirurgiens bariatriques de façon à ce que chaque segment de vidéo soit labélisé selon l'étape de l'opération lui correspondant.

Il s'agit donc d'une architecture de classification prenant en entrée un signal vidéo et indiquant en sortie l'étape identifiée : placement du port-à-cath, rétractation du foie, biopsie du foie, dissection du ligament gastro-colique, agrafage de l'estomac, extraction de l'estomac dans un sac, inspection finale de la ligne d'agrafage (*figure 77*). Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 85.6 %.



L'annotation de vidéos montrant une opération chirurgicale est un processus chronophage et fastidieux. La possibilité d'annoter automatiquement ces vidéos permettrait aux chirurgiens, notamment peu expérimentés, d'avoir rapidement et facilement accès à des passages spécifiques et ainsi consolider leurs compétences techniques. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'annoter les différentes étapes d'une sleeve gastrectomie par laparoscopie sur une vidéo, de façon automatique et précise [418].

10. Dermatologie

a. Cas d'usage en diagnostic

Andre Esteva et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter les cancers cutanés [358].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*Inception v3*) (figure 78). Les données d'entraînement sont constituées de 129450 images photographiques et dermatoscopiques de lésions cutanées. Chaque image a été labellisée par un dermatologue selon une taxonomie de 2032 pathologies. Cette taxonomie a été par la suite réduite à 757 classes.

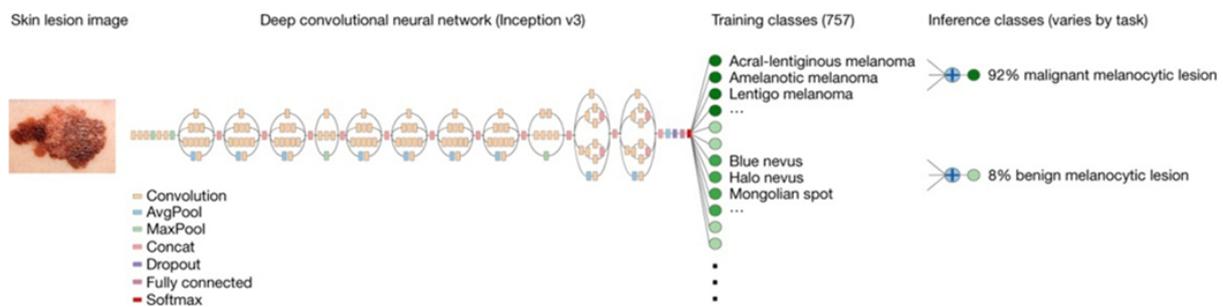


Figure 78. Représentation schématique du modèle présenté avec un exemple de prédiction de lésion mélanocytaire maligne (92%) (Reproduit avec la permission de l'éditeur Springer Nature. Auteurs : Andre Esteva et coll.) [358]

Il s'agit donc d'un modèle de classification prenant en entrée une image photographique ou dermatoscopique et indiquant en sortie la classe prédite : présence d'une tumeur mélanocytaire maligne vs. absence d'une tumeur mélanocytaire maligne (figure 78). Ce modèle parvient à atteindre des performances supérieures à l'état de l'art : aire sous la courbe de ROC = 91 %.

Selon les chiffres avancés par l'Organisation mondiale de la Santé en 2017, le nombre de cancers cutanés enregistrés est de deux à trois millions par an et son incidence est en hausse [419]. Parmi ses différentes formes, les mélanomes sont statistiquement minoritaires, ceux-ci sont néanmoins les formes les plus mortelles [420]. La prise en charge précoce d'un mélanome améliore drastiquement le pronostic vital du/de la patient(e) [421]. Les analyses histologiques permettent un diagnostic précis cependant celles-ci nécessitent de pratiquer une biopsie [422]. La possibilité de détecter les mélanomes de façon non invasive mais néanmoins précise pourrait faciliter leur diagnostic et ainsi favoriser leur prise en charge précoce. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux dermatologues de diagnostiquer précocement les tumeurs cutanées malignes uniquement à partir de photographies de lésions cutanées, de façon automatique, précise, non invasive et moins coûteuse [358].

b. Cas d'usage en thérapeutique

Simone Cazzaniga et son équipe présentent en 2009 deux modèles d'intelligence artificielle, l'un visant à prédire la réussite d'un traitement au laser excimer contre le vitiligo, l'autre visant à prédire la durée nécessaire à la repigmentation totale [423].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés sur deux réseaux de neurones. Les données d'entraînement sont constituées de variables socio-démographiques, hygiéno-diététiques, phénotypiques, des antécédents médicaux, des caractéristiques des vitiligos et des modalités de traitement collectés auprès de 120 patients. Ces données comportent un total de 325 cas de vitiligos traités par laser excimer et pour lesquels les données relatives à la repigmentation sont disponibles. Celles-ci ont été labellisées selon la réussite ou l'échec du traitement ainsi que la durée entre le début du traitement et la repigmentation totale.

Il s'agit donc d'un modèle de classification et d'un modèle de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées. Le modèle de classification indique en sortie la classe prédite : réussite du traitement au laser excimer vs. échec du traitement au laser excimer. Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 85,85 % ; sensibilité = 88,94 % ; spécificité = 80,95 % ; aire sous la courbe de ROC = 93,19 %. Le modèle de régression, quant à lui, indique en sortie une prédiction de la durée nécessaire à la repigmentation totale. Ce modèle parvient à atteindre les performances suivantes : exactitude = 69,85 % ; erreur absolue moyenne = 9,5468 ; racine de l'erreur quadratique moyenne = 11,9158 ; coefficient de détermination = 66,84 %.

Les vitiligos, notamment lorsque ceux-ci sont situés au niveau du visage, peuvent affecter la santé psychologique et sociale des personnes atteintes [424]. Les modalités de traitement des vitiligos sont diverses et leur efficacité peut grandement varier d'un(e) patient(e) à l'autre [425]. Le traitement au laser excimer en fait partie et a démontré son efficacité au cours d'un essai clinique [426]. Cependant, le traitement au laser excimer est chronophage et demande un investissement personnel important de la part du/de la patient(e) [427]. La possibilité de prédire la réussite et la durée du traitement permettrait de mieux cibler les patients les plus réceptifs. De plus, cela permettrait également d'optimiser les modalités du traitement en fonction du/de la patient(e) (ex. : nombre de sessions nécessaires, doses, association d'un corticostéroïde à appliquer sur la peau). Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux soignants de prédire la réussite et la durée d'un traitement au laser excimer contre le vitiligo [423].

11. Endocrinologie

a. Cas d'usage en diagnostic

Chin Lin et son équipe présentent en 2021 un modèle d'intelligence artificielle visant à détecter les paralysies périodiques thyrotoxiques [428].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones. Les données d'entraînement sont constituées de 39 électrocardiographies collectées auprès de 31 patients présentant une paralysie périodique thyrotoxique et de 502 électrocardiographies collectées auprès de 414 patients présentant une hypokaliémie. Ces électrocardiographies ont été labellisées selon qu'elles soient issues d'un(e) patient(e) atteint(e) de paralysie périodique thyrotoxique ou d'hypokaliémie.

Il s'agit donc d'un modèle de classification prenant en entrée une électrocardiographie et indiquant en sortie la classe prédite : présence d'une paralysie périodique thyrotoxique vs. absence d'une paralysie périodique thyrotoxique. Ce modèle parvient à atteindre des performances supérieures à l'état de l'art : sensibilité = 88,9 % ; spécificité = 69,2 % ; aire sous la courbe de ROC = 81,31 % ; F-score = 76,2 % (figure 79). Le temps d'exécution moyen du modèle est de cinq minutes. En associant la prédiction effectuée par le modèle avec des données biologiques (taux de filtration glomérulaire et concentration de chlorure sérique), les performances diagnostiques sont améliorées : sensibilité = 77,8 % ; spécificité = 100 % ; aire sous la courbe de ROC = 98,6% ; F-score = 87,5 % (figure 79).

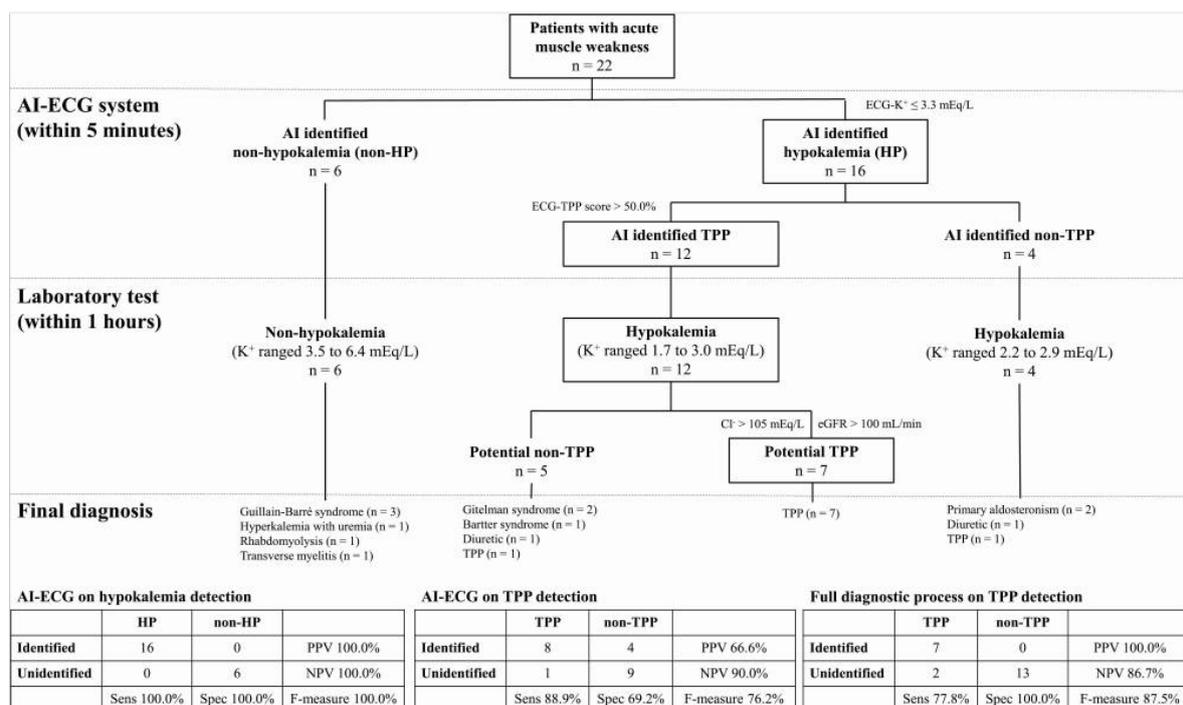


Figure 79. Méthode d'évaluation et performances des différents modèles présentés (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Chin Lin et coll.) [428]

Les paralysies périodiques thyrotoxiques entraînent des symptômes aigus et urgents à traiter lorsque ceux-ci surviennent [429]. Ses symptômes avant-coureurs sont en revanche discrets, rendant difficile son diagnostic précoce [430]. La possibilité de diagnostiquer précocement les paralysies périodiques thyrotoxiques permettrait d'anticiper les symptômes aigus et ainsi d'améliorer leur pronostic. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les paralysies périodiques thyrotoxiques, de façon automatique, rapide, précise et non invasive [428].

b. Cas d'usage en thérapeutique

Giacomo Cappon et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à optimiser les calculs de bolus d'insuline [431].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé par un algorithme de descente de gradient (*RMSprop*) sur un réseau de neurones (*figure 80*). Les données d'entraînement sont générées à partir de 80 patients virtuels simulés à partir d'un simulateur *UVa/Padova T1D*. Ces patients virtuels sont mis en situation postprandiale et caractérisés par différents paramètres : glycémie actuelle (G_c), taux de variation de la glycémie préprandiale (ROC), insuline active (IOB), ratio insuline/glucides (CR), facteur de correction (CF), valeur cible de glycémie (G_T), taux de perfusion basale (I_b), poids (BW), sensibilité à l'insuline (V_c), quantité de glucides ingérée (CHO). L'ensemble de ces paramètres (X_{pm}) aboutissent à une valeur numérique correspondant au bolus d'insuline optimal (Y_{pm}).

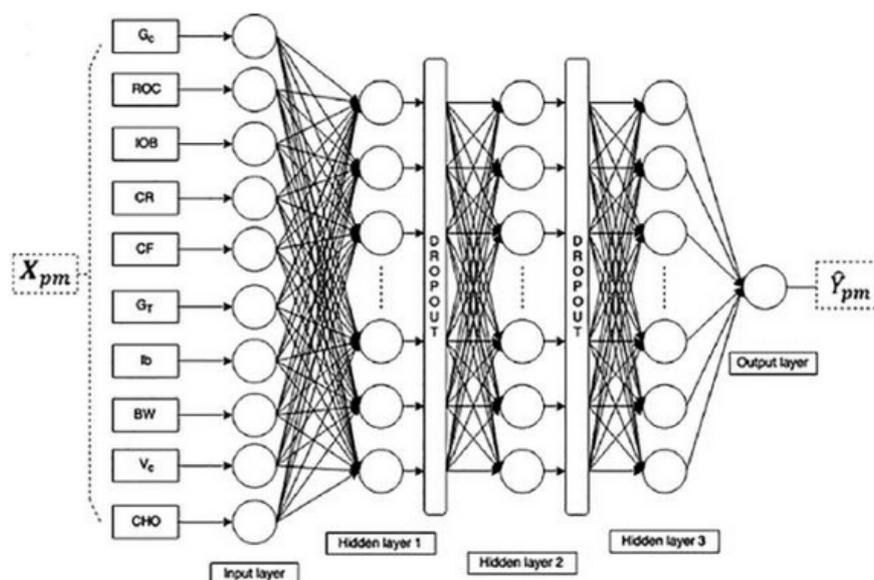


Figure 80. Représentation schématique du modèle présenté (Reproduit avec la permission de l'éditeur SAGE Publications. Auteurs : Giacomo Cappon et coll.) [431]

Il s'agit donc d'un modèle de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie le bolus d'insuline optimal prédit (*figure 80*). Les comparaisons statistiques menées indiquent une réduction significative ($p < 0,001$) par notre modèle de l'indice de risque glycémique comparativement à des méthodes de calcul préexistantes : réduction de 0,37 pour la formule standard ; 0,23 pour la méthode de Scheiner ; 0,20 pour la méthode de Pettus et Edelman.

Le traitement du diabète de type 1 implique notamment l'administration d'un bolus d'insuline après les repas [432]. Cependant, la détermination de la quantité d'insuline à administrer est un challenge pour les patients car la quantité idéale dépend de plusieurs facteurs [433]. C'est pourquoi des méthodes de calcul du bolus d'insuline sont développées afin de faciliter la gestion du diabète de type 1 par les patients [434]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'améliorer la précision du calcul du bolus d'insuline à administrer chez des patients atteints du diabète de type 1, de façon automatique [431].

12. Gastro-entérologie

a. Cas d'usage en diagnostic

Rintaro Hashimoto et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à détecter les néoplasies précoces de l'œsophage chez des patients présentant un œsophage de Barrett [435].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*Inception-ResNet-v2*). Les données d'entraînement sont constituées d'images endoscopiques collectées auprès de patients atteints d'un œsophage de Barret. Ces images ont été labellisées selon que celles-ci indiquent la présence (960) ou l'absence (990) d'une néoplasie précoce de l'œsophage.

Il s'agit donc d'un modèle de classification prenant en entrée un signal vidéo et indiquant en sortie les localisations détectées comme présentant des tissus dysplasiques. Ce modèle parvient à atteindre des performances élevées : exactitude = 95,4 % ; sensibilité = 96,4 % ; spécificité = 94,2 %. Le temps d'exécution moyen du modèle est de l'ordre de la seconde.

Le cancer de l'œsophage est le huitième cancer le plus courant et le sixième cancer le plus mortel mondialement. Son incidence était estimée à 570 000 nouveaux cas en 2018. Son pronostic est fortement favorisé s'il est diagnostiqué précocement, pourtant le diagnostic du cancer de l'œsophage est réalisé après que celui-ci ait métastasé dans plus de 40% des cas. L'un de ses facteurs de risque connus est l'œsophage de Barrett, c'est pourquoi une surveillance

endoscopique est conseillée chez les patients présentant cette pathologie [436]. Cependant, pour être efficace, cette surveillance requiert une biopsie effectuée à partir d'échantillons de tissu œsophagien prélevés tous les un à deux centimètres. La possibilité d'effectuer cette surveillance efficacement sans nécessiter de prélèvement favoriserait grandement la pratique de celle-ci. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux soignants de contrôler immédiatement durant l'endoscopie la présence de dysplasies chez des patients présentant un œsophage de Barret, de façon automatique, rapide, précise et moins invasive [435].

b. Cas d'usage en thérapeutique

Iolanda Valentina Popa et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à prédire la réussite d'un traitement inhibiteur du facteur de nécrose tumorale α contre les rectocolites hémorragiques [437,438].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones. Les données d'entraînement comportent des variables pré-traitement cliniques (scores de Mayo endoscopiques) et biologiques (neutrophiles, largeur de distribution plaquettaire, protéine C réactive et alpha globulines) collectées auprès de 40 patients atteints de rectocolite hémorragique. Les scores de Mayo endoscopiques ont aussi été déterminés après un an de traitement afin de définir pour chaque patient(e) si il ou elle présente une rémission ou une rectocolite hémorragique active. Les données ont été labellisées sur la base de cette information.

Il s'agit donc d'un modèle de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : rémission après un an de traitement *vs.* pathologie active après un an de traitement. Ce modèle parvient à atteindre des performances élevées : exactitude = 90 % ; sensibilité = 100 % ; spécificité = 75 % ; valeur prédictive positive = 86% ; valeur prédictive négative = 100% ; aire sous la courbe de ROC = 92 %.

Les inhibiteurs du facteur de nécrose tumorale α ont démontré leur intérêt thérapeutique dans le cadre de la rectocolite hémorragique en cas d'échec ou de contre-indication aux thérapies conventionnelles [439]. Cependant, il s'agit de traitements engageant des coûts importants et pouvant provoquer des effets indésirables graves d'ordre immunologique ou hépatique [440,441,442]. La possibilité d'estimer la réussite du traitement permettrait aux soignants de mieux cibler les patients les plus réceptifs, d'adapter les posologies et d'anticiper les rechutes. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire la réussite d'un traitement inhibiteur du

facteur de nécrose tumorale α dans le cadre d'une rectocolite hémorragique, de façon automatique et précise [437].

13. Gériatrie (cas d'usage en diagnostic)

Elizabeth Ford et son équipe présentent en 2021 des modèles d'intelligence artificielle visant à détecter des patients atteints de démence [443].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : régression logistique, forêt aléatoire, classification naïve bayésienne. Les données d'entraînement sont constituées de données cliniques, biologiques, pharmaceutiques et hygiéno-diététiques sous forme de texte issues de 93426 dossiers de santé numériques. Ces dossiers sont affiliés à des patients âgés de 65 ans ou plus, ayant reçu un diagnostic de démence ou non. Chaque patient(e) ayant reçu un diagnostic de démence a été apparié(e) à un(e) patient(e) témoin, n'ayant pas reçu de diagnostic de démence, mais ayant le même âge et le même sexe. Les informations disponibles dans leurs dossiers de santé devaient couvrir au moins trois années avant le premier diagnostic de démence le cas échéant ou, pour les patients témoins, trois années avant la date correspondant au diagnostic de démence du/de la patient(e) leur étant apparié(e).

Il s'agit de modèles de classification prenant en entrée des données sous forme de texte d'un ou plusieurs patients et indiquant en sortie les patients pour lesquels le modèle prédit la présence d'une démence. Le modèle de régression logistique est celui présentant les meilleures performances : sensibilité = 84 % ; spécificité = 93 % ; aire sous la courbe de ROC = 94% ; valeur prédictive positive = 51 %.

L'identification de certaines pathologies non diagnostiquées est possible au sein d'une population de patients en se basant uniquement sur leurs données de santé numérisées [407]. Cependant, passer manuellement en revue ces données est chronophage et fastidieux. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement être en mesure d'identifier les patients atteints de démence à partir de données de santé sous forme de texte, de façon automatique et précise [443].

14. Génétique (cas d'usage en diagnostic)

Francisco M. De La Vega et son équipe présentent en 2021 une architecture d'intelligence artificielle visant à identifier les gènes en cause dans une pathologie génétique [444].

Cette architecture, commercialisée sous le nom de *Fabric GEM*, intègre dans son analyse des ontologies et des facteurs de Bayes. Celle-ci prend en entrée des variants génétiques ainsi que des métadonnées (ex. : phénotype du ou de la patiente au format *Human Phenotype Ontology*)

et indique en sortie un classement détaillé des variants les plus probables d'être en cause dans une pathologie. *Fabric GEM* parvient à atteindre des performances élevées lorsqu'il s'agit d'identifier des gènes en cause dans les maladies rares : exactitude = 90 %. Le temps d'exécution moyen du modèle est d'environ 15 minutes.

Le diagnostic des maladies rares fait de plus en plus intervenir l'interprétation des variants génétiques [445]. Cependant, dans le cadre des maladies rares, l'interprétation de variants est une pratique complexe, chronophage et très coûteuse [446]. De fait, les patients atteints d'une maladie rare vivent fréquemment une errance diagnostique pouvant parfois durer de nombreuses années [447]. La possibilité d'interpréter automatiquement des variants génétiques pourrait permettre un diagnostic plus rapide des maladies rares et accélérer également l'étude étiologique de ces pathologies. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'identifier les gènes causaux d'une maladie rare, de façon automatique, rapide, précise, moins coûteuse, et ainsi réduire l'errance diagnostique dont souffrent les patients atteints de ces pathologies [444].

15. Gynécologie-obstétrique

a. Cas d'usage en diagnostic

E. Salamalekis et son équipe présentent en 2002 un modèle d'intelligence artificielle visant à détecter les hypoxies fœtales intrapartum [448,449].

Ce modèle est obtenu à l'issue d'un apprentissage automatique non supervisé réalisé sur un réseau de neurones (*réseau de Kohonen*, aussi nommé carte autoadaptative ou *self-organizing map*) (*figure 81*). Un second apprentissage non supervisé a été effectué par un algorithme de k-moyennes afin de définir des niveaux de risque pour les clusters formés. Les données d'entraînement sont constituées d'enregistrements fœtaux de fréquence cardiaque et de saturation pulsée en oxygène collectés durant les accouchements de 61 femmes à plus de 37 semaines de grossesse. Ces enregistrements furent segmentés en portions de 10 minutes et débarrassés des artefacts puis chaque segment a été labellisé selon le niveau de risque associé de présence d'une hypoxie fœtale (*figure 81*).

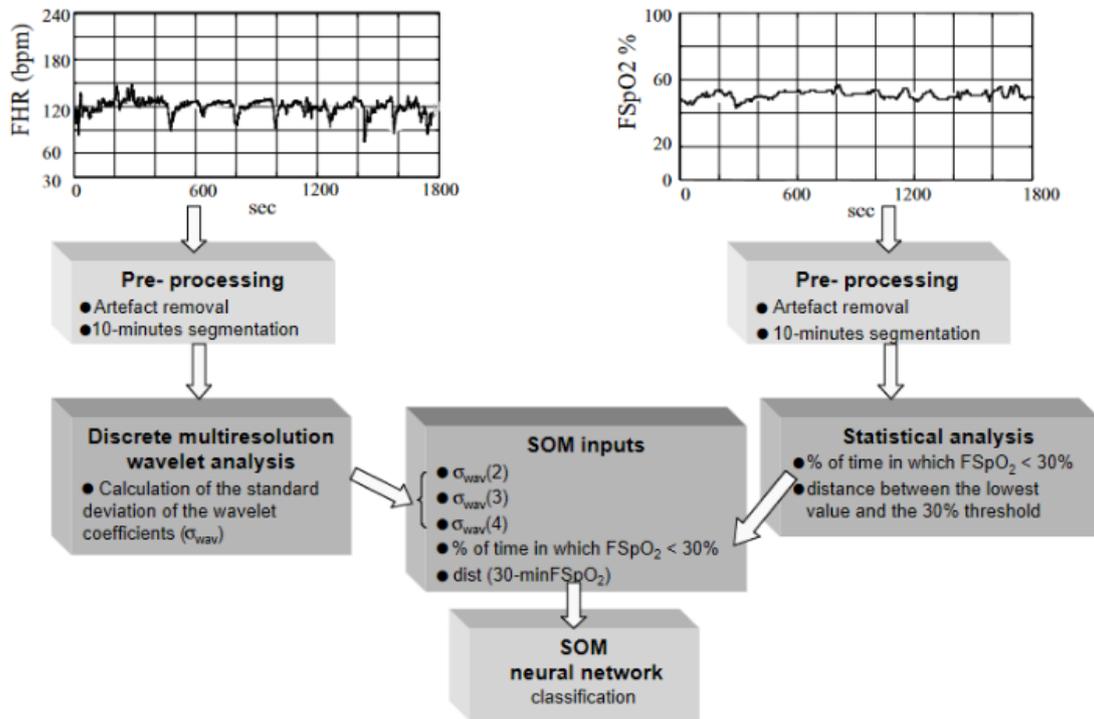


Figure 81. Logigramme du parcours des données d'entraînement. *SOM : self-organizing map (carte autoadaptative).*
 (Reproduit avec la permission de l'éditeur John Wiley and Sons. Auteurs : E. Salamalekis et coll.) [448]

Il s'agit donc d'un modèle de classification prenant en entrée un enregistrement de fréquence cardiaque fœtale ainsi qu'un enregistrement de la saturation pulsée en oxygène concomitante et indiquant en sortie la classe prédite : présence très probable d'une hypoxie fœtale (cluster 3) vs. présence probable d'une hypoxie fœtale (cluster 2) vs. absence d'une hypoxie fœtale (cluster 1). Ce modèle parvient à atteindre des performances élevées : sensibilité = 83,3 % ; spécificité = 97,9 %.

La mesure de la fréquence cardiaque fœtale est communément réalisée avant et durant les accouchements pour s'assurer de la bonne santé du fœtus [450]. Détecter à temps une fréquence cardiaque fœtale anormale permet aux soignants d'agir de sorte à limiter les complications (ex. : dégâts neurologiques, décès) pouvant en résulter. Cependant, l'interprétation de ce marqueur seul manque de spécificité et peut donc dans certains cas conduire à des interventions inutiles (ex. : césarienne en cas de faux positif) [451]. L'oxymétrie de pouls par réflectance est une technique non invasive pouvant être réalisée en continu et associée à la mesure de la fréquence cardiaque fœtale afin de renforcer la précision de la surveillance [452]. La possibilité d'analyser ces deux signaux en continu et de façon automatique permettrait aux soignants de pouvoir être avertis au plus vite d'un risque d'hypoxie fœtale, favorisant ainsi son pronostic. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'analyser en continu ces signaux et de détecter la présence d'une hypoxie fœtale intrapartum, de façon automatique et précise [448].

b. Cas d'usage en thérapeutique

Moshe Wald et son équipe présentent en 2005 un modèle d'intelligence artificielle visant à prédire la réussite d'une fécondation in-vitro par injection intra-cytoplasmique de spermatozoïde après prélèvement chirurgical dans le cadre d'une infertilité masculine [449,453]. Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones. Les données d'entraînement comportent des variables relatives à 83 fécondations in-vitro par injection intra-cytoplasmique de spermatozoïde après prélèvement chirurgical : âge maternel, technique de prélèvement chirurgical, type de spermatozoïdes utilisés, type d'infertilité masculine. Les patients auprès desquels ces données ont été collectées étaient âgés de 22 à 43 ans. Chaque fécondation in-vitro du jeu de données a été labellisée selon son issue (survenue ou absence de grossesse).

Il s'agit donc d'un modèle de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : grossesse à l'issue de la fécondation in-vitro vs. pas de grossesse à l'issue de la fécondation in-vitro (*figure 82*). Ce modèle parvient à atteindre des performances relativement élevées : aire sous la courbe de ROC = 78,3 %.

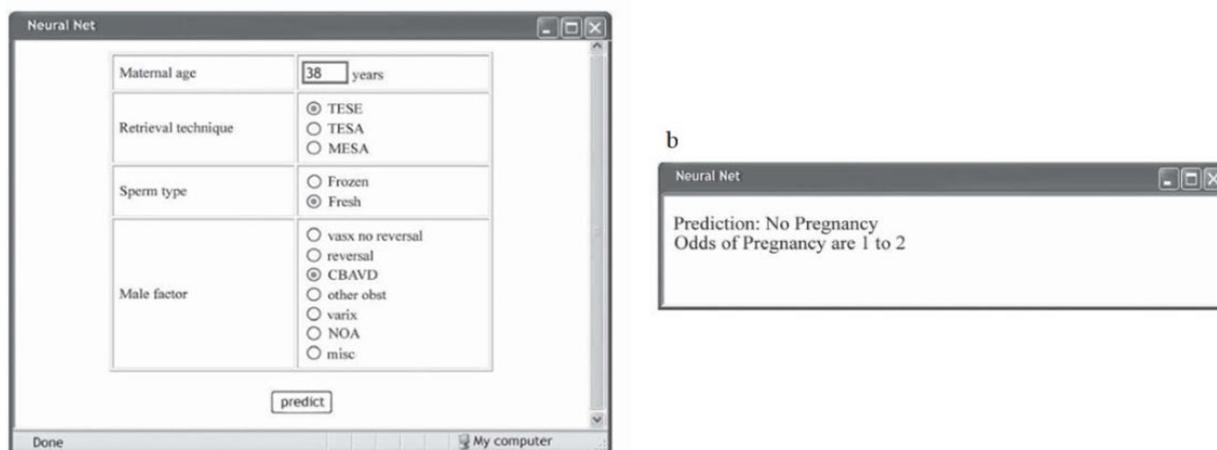


Figure 82. Interface web du modèle : (a) sélection des données d'entrée ; (b) prédiction affichée à l'issue de l'inférence. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : M. Wald et coll.) [453]

La fécondation in-vitro par injection intra-cytoplasmique de spermatozoïde après prélèvement chirurgical s'est imposée comme une alternative de procréation prometteuse en cas d'infertilité masculine sévère [454]. Cependant, les facteurs favorisant sa réussite sont encore incertains actuellement. En outre, la fécondation in-vitro est une expérience pouvant être éprouvante physiquement et psychologiquement pour les personnes qui la vivent [455]. La possibilité de prédire la réussite de celle-ci permettrait de mieux cibler les patients ayant le plus de chance de réussite et d'optimiser les modalités pouvant l'être. Les résultats obtenus révèlent qu'un modèle

d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire la réussite d'une fécondation in-vitro par injection intra-cytoplasmique de spermatozoïde après prélèvement chirurgical dans le cadre d'une infertilité masculine, de façon automatique, précise et non invasive [453].

16. Hématologie

a. Cas d'usage en diagnostic

Sophie Cai et son équipe présentent en 2021 un modèle d'intelligence artificielle visant à détecter la présence de néovaisseaux en sea fan chez des patients atteints de drépanocytose [456,457].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*Inception v4*). Les données d'entraînement sont constituées de photographies larges et en couleur de fond d'œil collectées auprès de 190 patients atteints de drépanocytose. Aucun(e) patient(e) n'était diabétique ni n'avait subi de chirurgie ou de traitement au laser contre la rétinopathie. Chaque photographie a été labellisée selon la présence (57) ou l'absence (133) de néovaisseaux en sea fan.

Il s'agit d'un modèle de classification prenant en entrée une photographie large et en couleur de fond d'œil et indiquant en sortie la classe prédite : présence de néovaisseaux en sea fan *vs.* absence de néovaisseaux en sea fan. Ce modèle parvient à atteindre des performances élevées : sensibilité = 97,4 % ; spécificité = 97 % ; aire sous la courbe de ROC = 98,8 %.

Les patients atteints de drépanocytose sont sujets au risque de rétinopathie drépanocytaire pouvant conduire, si celle-ci n'est pas prise en charge, à une perte sévère et parfois définitive des fonctions visuelles [458]. C'est pourquoi les examens ophtalmologiques sont recommandés dès le plus jeune âge chez les patients atteints de drépanocytose [459]. L'un des signes précoces de complication à surveiller est l'apparition de néovaisseaux en sea fan visibles sur un fond d'œil [460]. La possibilité de détecter automatiquement et avec précision la présence de néovaisseaux en sea fan chez des patients atteints de drépanocytose permettrait aux soignants de pouvoir rapidement évaluer le risque de complication et proposer un traitement prophylactique (ex. : photocoagulation au laser) si nécessaire. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à détecter la présence de néovaisseaux en sea fan chez des patients atteints de drépanocytose, de façon automatique, précise et non invasive [456].

b. Cas d'usage en thérapeutique

Toshiaki Ohara et son équipe présentent en 2021 une architecture d'intelligence artificielle visant à prévenir les anémies chez des patients recevant une hémodialyse d'entretien [461].

Cette architecture est obtenue à l'issue d'un apprentissage profond supervisé réalisé sur deux réseaux de neurones, le premier étant un réseau de neurones dense destiné à prédire la procédure optimale quant à l'administration d'agents stimulant l'érythropoïèse, le second étant un réseau de neurones récurrents destiné à prédire la procédure optimale quant à l'administration d'une supplémentation en fer. Les données d'entraînement comportent des variables dont les valeurs ont été mesurées avant 6080 hémodialyses d'entretien : concentration en hémoglobine (Hb), volume globulaire moyen (MCV), concentration de ferritine, saturation de la transferrine (TSAT), antécédent des dosages d'agents stimulant l'érythropoïèse (ESA) et de supplémentation en fer (IS). Ces données ont été collectées auprès de 130 patients ne présentant ni inflammation, ni pathologie infectieuse, ni cancer au moment de leurs hémodialyses. Les décisions prises par les soignants quant à l'augmentation, la diminution ou le maintien des dosages d'agents stimulant l'érythropoïèse et de supplémentation en fer ont permis de labelliser chaque hémodialyse du jeu de données.

Il s'agit donc d'une architecture de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie les classes prédites : augmentation du dosage *vs.* maintien du dosage *vs.* diminution du dosage pour les agents stimulant l'érythropoïèse / augmentation du dosage *vs.* maintien du dosage pour la supplémentation en fer (*figure 83*). Ce modèle parvient à atteindre des performances élevées : exactitude = 97 % pour les agents stimulant l'érythropoïèse ; exactitude = 98 % pour la supplémentation en fer.

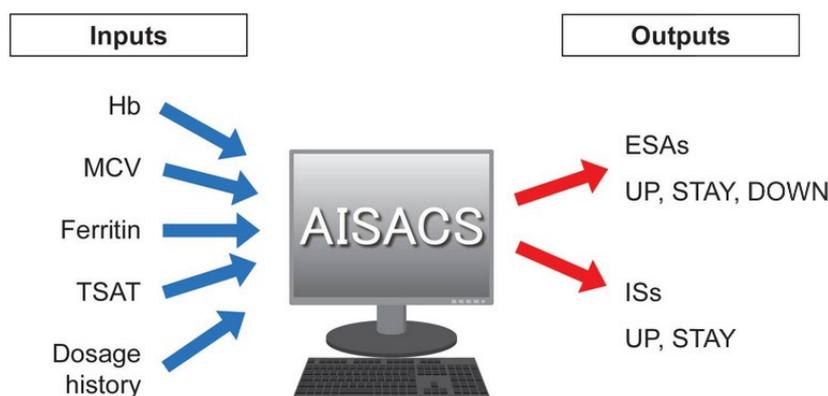


Figure 83. Données en entrée et classes prédites en sortie. AISACS : artificial intelligence supported anemia control system (système de contrôle de l'anémie assisté par intelligence artificielle). (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Toshiaki Ohara et coll.) [461]

L'anémie fait partie des complications courantes chez les patients hémodialysés chroniques. La mesure préventive standard de l'anémie chez ces patients est l'administration d'agents stimulant l'érythropoïèse et d'une supplémentation en fer [462]. Cependant, le nombre de soignants en mesure d'assurer la supervision de ce traitement augmente moins rapidement que le nombre de patients nécessitant une hémodialyse. Qui plus est, les agents stimulant l'érythropoïèse sont des traitements coûteux [463]. La possibilité d'estimer automatiquement la marche à suivre optimale permettrait aux soignants de superviser cette prise en charge avec moins de contrainte de temps et éventuellement de réduire ses coûts. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à superviser l'administration des traitements visant à prévenir l'anémie chez des patients hémodialysés, de façon automatique, précise et moins coûteuse [461].

17. Hépatologie

a. Cas d'usage en diagnostic

Mindaugas Marozas et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter les hypertensions portales [464,465].

Ce modèle est obtenu à l'issue d'un méta-apprentissage automatique supervisé réalisé par un ensemble d'algorithmes : classification naïve bayésienne, régression logistique, *K Star*, table de décision, forêt aléatoire. Les données d'entraînement comportent des variables démographiques, cliniques, biologiques, hématologiques, des mesures instrumentales spléniques ainsi que des élastographies transitoires hépatiques et spléniques. Ces données ont été collectées auprès de 107 patients atteints de pathologie hépatique chronique puis labellisées selon une mesure du gradient de pression hépatique inférieure ou supérieure à 10 mmHg.

Il s'agit donc d'un modèle de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : présence d'une hypertension portale *vs.* absence d'une hypertension portale. Ce modèle parvient à atteindre des performances élevées : exactitude = 89,72 % ; sensibilité = 83 % ; spécificité = 92 % ; aire sous la courbe de ROC = 96 %.

La survenue d'une hypertension portale est un facteur d'aggravation majeur pour plusieurs pathologies hépatiques chroniques [466]. Son diagnostic s'effectue classiquement par le biais d'une mesure du gradient de pression hépatique, une mesure invasive et coûteuse [467]. L'élastographie transitoire est une autre méthode permettant le diagnostic de l'hypertension portale, moins invasive mais également moins précise [468]. La possibilité de détecter les hypertensions portales de façon précise mais néanmoins non invasive favoriserait leur

dépistage. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux soignants de diagnostiquer les hypertensions portales de façon automatique, précise, non invasive et moins coûteuse [464].

b. Cas d'usage en thérapeutique

Haixin Ai et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à prédire l'hépatotoxicité d'une molécule [469].

Ce modèle est obtenu à l'issue d'un méta-apprentissage automatique supervisé réalisé par un ensemble d'algorithmes (*figure 84*) : machine à vecteurs de support, forêt aléatoire, amplification de gradient extrême (*XG Boost*). Les données d'entraînement comportent 683 molécules hépatotoxiques et 558 molécules non hépatotoxiques au format SMILES (*simplified molecular-input line-entry system*). A l'aide du logiciel *PaDEL-Descriptor*, ces molécules ont été synthétisées en 12 motifs moléculaires avant d'être fournies à l'algorithme (*figure 84*).

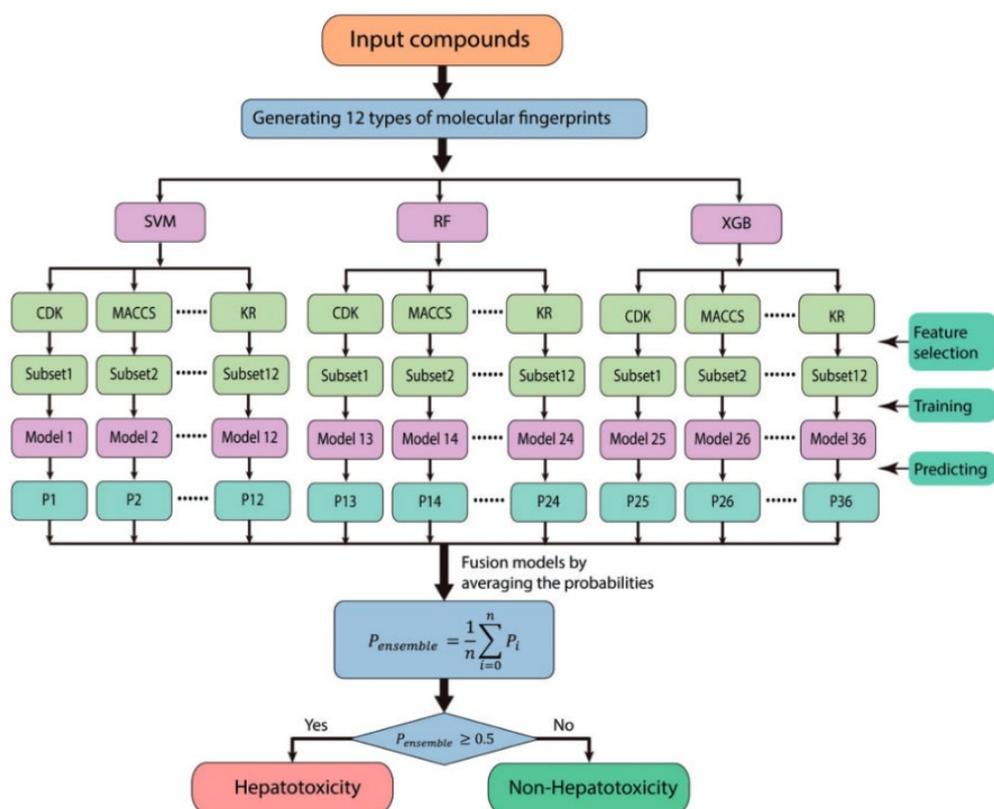


Figure 84. Représentation schématique du méta-apprentissage à l'origine du modèle présenté (Reproduit avec la permission de l'éditeur Oxford University Press. Auteurs : Haixin Ai et coll.) [469]

Il s'agit d'un modèle de classification prenant en entrée une molécule au format SMILES et indiquant en sortie la classe prédite : hépatotoxique vs. non hépatotoxique (*figure 84*). Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 84,3 % ; sensibilité = 86,9 % ; spécificité = 75,4 % ; aire sous la courbe de ROC = 90,4 %.

L'hépatotoxicité fait partie des causes majeures de retrait du marché d'un traitement ou d'échec d'une molécule soumise à un essai clinique [470]. Diverses approches *in vivo* et *in vitro* visant à évaluer l'hépatotoxicité iatrogène existent, cependant celles-ci sont chronophages, coûteuses et souvent non conformes aux résultats obtenus chez des sujets humains. Les approches *in silico*, plus rapides et moins coûteuses, gagnent en popularité à mesure que celles-ci démontrent leur performance [471]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des chercheurs à évaluer l'hépatotoxicité *in silico* de façon automatique et précise [469].

18. Immunologie

a. Cas d'usage en diagnostic

Giovanni Melioli et son équipe présentent en 2014 un modèle d'intelligence artificielle visant notamment à identifier des allergies [472].

Il s'agit d'un système expert nommé *Allergenius* qui intègre dans son analyse plus de 700 règles hiérarchisées selon leurs répercussions ainsi qu'une base de connaissances construite à partir de la littérature disponible et de bases de données spécialisées : *Allergome*, *AllergenOnline*, *RCSB (Research Collaboratory for Structural Bioinformatics)*, *Uniprot*, *WHO/IUIS Allergen Nomenclature*.

Allergenius prend en entrée des données personnelles et cliniques d'un(e) patient(e) (résultats de prick test, dosage des IgE, résultats fournis par des biopuces à allergènes) et indique en sortie un rapport détaillé présentant notamment les allergènes susceptibles de provoquer une réaction allergique chez ce/cette patient(e). *Allergenius* parvient à atteindre des performances élevées lorsqu'il s'agit d'identifier des allergènes, atteignant jusqu'à 100% de précision pour certains d'entre eux.

L'identification de la source d'une allergie chez un(e) patient(e) représente souvent un challenge en raison du nombre important d'allergènes potentiels et de leur hétérogénéité [473]. La possibilité d'identifier et catégoriser automatiquement certains de ces allergènes (notamment à partir d'une biopuce à allergènes) pourrait faciliter la démarche diagnostique des soignants, en particulier chez des patients pour lesquels le prick test n'est pas concluant [474]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer des allergies et à identifier les allergènes responsables, de façon automatique, précise et non invasive [472].

b. Cas d'usage en thérapeutique

Basem Ameen Moghram et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à prédire la structure d'épitopes en vue de concevoir des vaccins basés sur ceux-ci [475,476].

Ce modèle est obtenu à l'issue d'un apprentissage automatique élitiste^{xxvii} réalisé par un algorithme génétique (figure 85). Les données d'entraînement sont constituées de séquences d'épitopes.

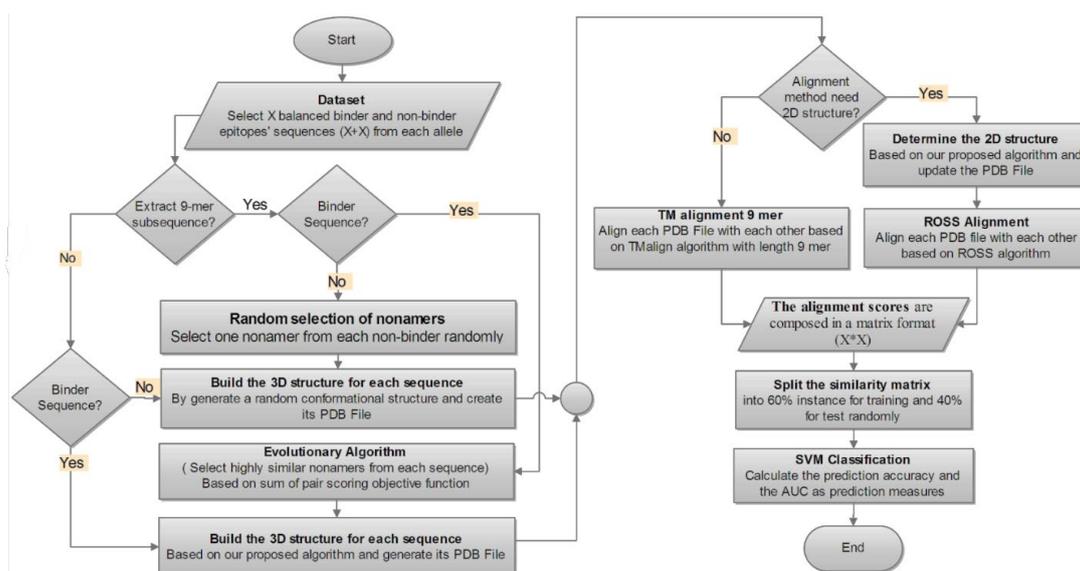


Figure 85. Processus de prédiction de structures d'épitopes par le modèle présenté (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Basem Ameen Moghram et coll.) [475]

Le modèle parvient à construire des structures d'épitopes en trois dimensions de façon performante : exactitude = 95,13 % ; aire sous la courbe de ROC = 98,7 %.

La conception de vaccins basée sur les épitopes est notamment un terrain prometteur mais challengeant [478]. L'étude des structures tertiaires d'épitopes est une piste à poursuivre en recherche thérapeutique car de nombreuses pathologies résultent d'un dysfonctionnement d'une ou plusieurs protéines [479]. De fait, l'identification de la structure des épitopes restreints au complexe majeur d'histocompatibilité de classe II est une étape importante vers la conception de vaccins à base d'épitopes [480]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des chercheurs à identifier des structures d'épitopes pouvant être exploitées dans la conception de nouveaux vaccins inactivés, de façon automatique et précise [475].

^{xxvii} Un apprentissage automatique élitiste procède en favorisant la sélection des meilleurs individus à chaque itération ou génération du processus d'apprentissage. Dans le cadre d'un algorithme évolutionniste, un « individu » désigne une solution candidate dans l'espace de recherche (ex. : configuration spécifique de paramètres ou de gènes). [477]

19. Infectiologie

a. Cas d'usage en diagnostic

Runwen Hu et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à détecter les infections à SARS-CoV-2 [481].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*ShuffleNet V2*) (figure 86). Les données d'entraînement sont constituées d'images de tomodensitométrie thoracique. Celles-ci ont été labellisées selon que le/la patient(e) présente (313) ou non (242) une infection à SARS-CoV-2.

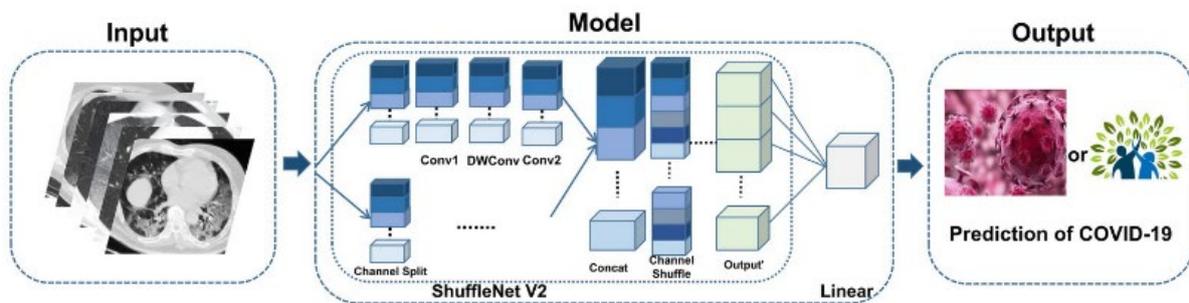


Figure 86. Représentation schématique du modèle présenté (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Runwen Hu et coll.) [481]

Il s'agit donc d'un modèle de classification prenant en entrée une image de tomodensitométrie thoracique et indiquant en sortie la classe prédite : présence d'une infection à SARS-CoV-2 vs. absence d'une infection à SARS-CoV-2 (figure 86). Ce modèle parvient à atteindre des performances élevées : exactitude = 91,21 % ; sensibilité = 90,52 % ; spécificité = 91,58 % ; aire sous la courbe de ROC = 96,89 %.

En 2020, le monde entier s'est vu plongé dans une crise sanitaire brutale suite à l'émergence d'un nouveau virus, le SARS-CoV-2 [482]. La pandémie provoquée par ce virus a émergé de Chine à la fin du mois de décembre 2019 puis s'est propagée de façon exponentielle pour aboutir à des centaines de millions de cas et des millions de décès de par le monde [483]. L'urgence de la situation nécessitait des moyens de diagnostic efficaces et rapides permettant de limiter la propagation et la létalité du virus [484]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle était en mesure de renforcer la capacité des soignants à diagnostiquer les infections à SARS-CoV-2 au commencement de la pandémie, de façon automatique, précise et non invasive [481].

b. Cas d'usage en thérapeutique

Ying Shen et son équipe présentent en 2018 un modèle d'intelligence artificielle visant notamment à optimiser les choix d'antibiothérapies [485].

Il s'agit d'un système d'aide à la décision nommé IDDAP (*infectious diseases data analysis program*) qui intègre dans son analyse une ontologie fondée sur le diagnostic de pathologies infectieuses et les antibiothérapies (*figure 87*). Celle-ci a été construite à partir d'ontologies existantes comprenant un total de 1267000 classes, 1266993 sous-classes et 7608725 axiomes portant sur les pathologies infectieuses, syndromes infectieux, bactéries et traitements médicamenteux antibactériens. IDDAP comprend 507 pathologies infectieuses et leurs traitements, 332 sites d'infection, 936 symptômes, 371 types de complications, 838407 types de bactéries, 341 types d'antibiotiques, 1504 taux de réaction antibiotique-bactérie, 431 interactions médicamenteuses et 86 contre-indications spécifiques aux populations.

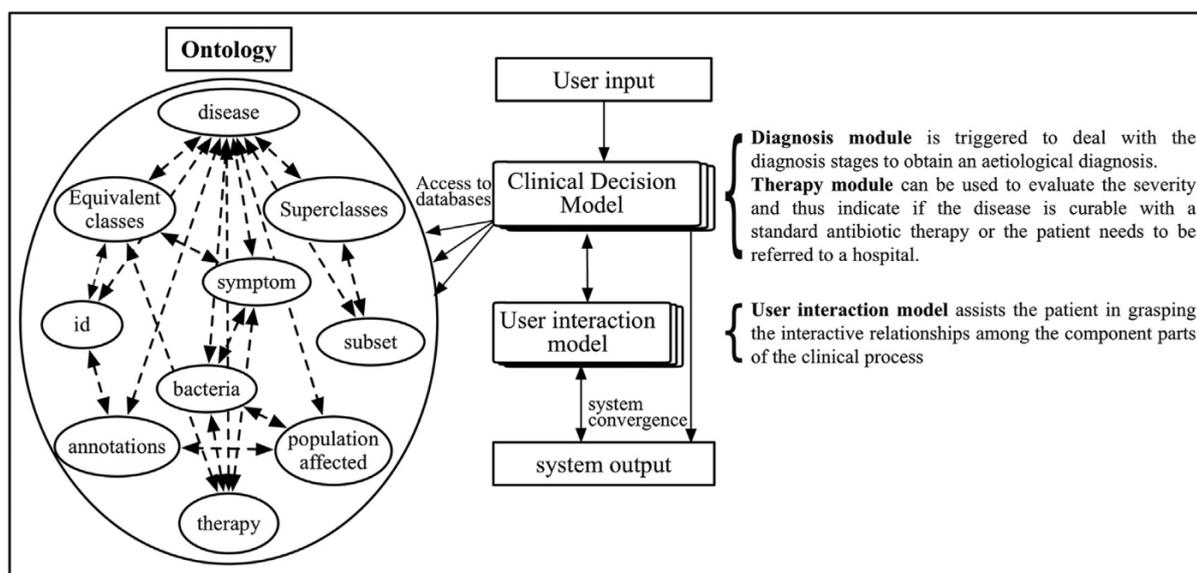


Figure 87. Représentation schématique de l'architecture présentée (IDDAP) (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Ying Shen et coll.) [485]

IDDAP prend successivement en entrée les données cliniques et thérapeutiques d'un(e) patient(e) (pathologies connues chez le ou la patiente, température corporelle, site d'infection, symptômes, complications relatives à la pathologie infectieuse en cours, traitements pris actuellement, contraindications médicamenteuses connues) et indique en sortie un rapport présentant notamment l'antibiothérapie optimale conseillée. IDDAP parvient à atteindre des performances élevées : aire sous la courbe de ROC = 89,91%.

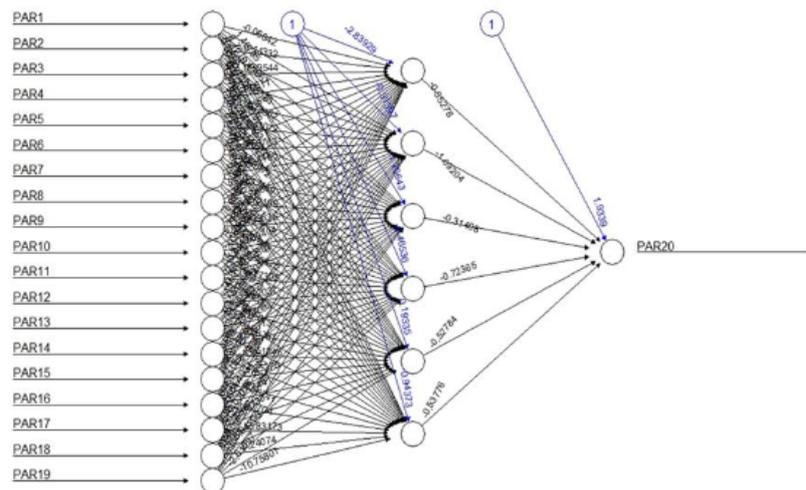
Une administration d'antibiotiques requiert en principe l'expertise d'un professionnel de santé afin que celle-ci soit ajustée au mieux selon le/la patient(e) et que les risques d'antibiorésistance engendrée soient minimisés. Malgré ces recommandations et les campagnes de sensibilisation menées, la prise d'antibiotiques en automédication est courante mondialement, en particulier dans les pays ne disposant pas de couverture santé universelle [486]. Ce phénomène accroît l'émergence d'antibiorésistances, l'une des causes de décès les plus significatives mondialement [487]. Selon les auteurs (Ying Shen et coll.), une atténuation de ces conséquences pourrait être envisagée si les patients avaient accès à un système capable d'évaluer la nécessité d'une antibiothérapie et d'identifier le traitement optimal le cas échéant. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de recommander le traitement le plus adéquat contre une infection, de façon automatique, précise et non invasive [485].

20. Médecine d'urgence

a. Cas d'usage en diagnostic

Vida Abedi et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter les accidents vasculaires cérébraux [488].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones (*figure 88*). Les données d'entraînement comportent des variables socio-démographiques, hygiéno-diététiques, cliniques, biologiques ainsi que des antécédents médicaux collectés auprès de 260 patients présentant des symptômes évocateurs d'un accident vasculaire cérébral (*figure 88*). Ces données ont été labellisées selon que le/la patient(e) soit effectivement atteint(e) d'un accident vasculaire cérébral (130) ou ne l'était pas malgré les symptômes évocateurs (130). Les variables ont été mesurées au cours des quatre heures et demie suivant l'apparition des symptômes.



1. Age
2. Race
3. Sex
4. History of hypertension
5. History of diabetes
6. History of hyperlipidemia
7. History of stroke
8. History of atrial fibrillation
9. History of psychiatry illness
10. History of end stage renal disease
11. History of seizure
12. History of tobacco use
13. Fist recorded systolic blood pressure
14. Fist recorded diastolic blood pressure
15. Fist recorded plasma glucose level
16. Low-density lipoprotein (LDL) – previously charted
17. Hemoglobin A1c (HbA1c) – previously charted
18. National Institute of Health Stroke Scale (NIHSS)
19. Facial droop on initial presentation

Figure 88. **Représentation schématique du modèle présenté et liste des 19 variables d'entraînement**
(Reproduit avec la permission de l'éditeur Wolters Kluwer Health. Auteurs : Vida Abedi et coll.) [488]

Il s'agit donc d'un modèle de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : présence d'un accident vasculaire cérébral vs. absence d'un accident vasculaire cérébral. Ce modèle parvient à atteindre des performances relativement élevées : sensibilité = 80 % ; spécificité = 86,2 % ; valeur prédictive positive = 92%.

La justesse du diagnostic d'accident vasculaire cérébral lors de la survenue de symptômes évocateurs est capitale en raison de la morbidité importante induite par cet événement [489]. Des séquelles irréversibles peuvent se manifester très rapidement après les premiers symptômes, c'est pourquoi le délai de diagnostic et d'intervention est limité [490]. Cependant, les symptômes évocateurs d'accident vasculaire cérébral peuvent résulter de pathologies autres, rendant le diagnostic difficile dans certains cas [491,492]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer les accidents vasculaires cérébraux, de façon automatique et précise [488].

b. Cas d'usage en thérapeutique

Zhongheng Zhang et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à prédire la compliance pulmonaire au cours d'une ventilation mécanique chez des patients atteints d'un syndrome de détresse respiratoire aiguë [493,494].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme d'arbre décisionnel intégrant des régressions linéaires (*M5P*). Les données d'entraînement sont constituées de séries temporelles indiquant la pression respiratoire collectées auprès de 18 patients ventilés mécaniquement dans le cadre d'un syndrome de détresse respiratoire aiguë.

Il s'agit d'un modèle de régression prenant en entrée une série de pressions expiratoires positives de fin d'expiration et indiquant en sortie une prédiction de l'évolution de la compliance pulmonaire ainsi qu'une recommandation quant à l'ajustement de la pression pulmonaire. Ce modèle parvient à atteindre des performances élevées avec un taux de réussite s'élevant à 93,2 %.

La prise en charge des syndromes de détresse respiratoire aiguë nécessite une assistance ventilatoire invasive pouvant entraîner des lésions pulmonaires [495]. Ces dernières peuvent être évitées en ajustant la pression expiratoire positive de fin d'expiration de façon à ce que la compliance pulmonaire se maintienne à un niveau élevé [496]. Cependant, l'influence de la variation de la pression expiratoire sur la compliance pulmonaire est obtenue après un délai conséquent. La possibilité de prédire l'évolution de la compliance pulmonaire selon une modification de la pression expiratoire positive permettrait aux soignants d'ajuster cette dernière avec précision et ainsi limiter les lésions pulmonaires provoquées par l'assistance ventilatoire. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à estimer l'évolution de la compliance pulmonaire au cours d'une ventilation mécanique, de façon automatique et précise [493].

21. Médecine nucléaire

a. Cas d'usage en diagnostic

Junshen Xu et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à générer des images tomographiques de qualité à partir d'images tomographiques en faible dose [360].

Ce modèle est obtenu à l'issue d'un apprentissage profond non supervisé réalisé sur un réseau de neurones convolutif de type auto-encodeur (*figure 89*). Les données d'entraînement sont constituées d'images tomographiques et par résonance magnétique (PET–MRI) cérébrales sur plusieurs couches. Ces examens d'imagerie ont été effectués sur neuf patients atteints d'un glioblastome, après administration d'une dose standard de 18F-fluorodeoxyglucose.

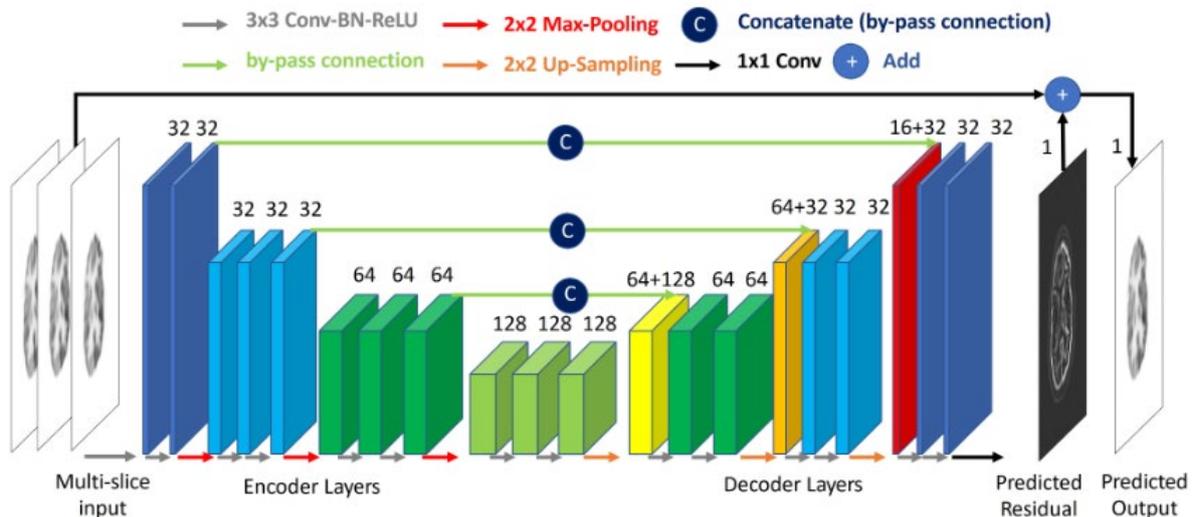


Figure 89. Représentation schématique du modèle présenté (*Reproduit avec la permission de l'auteur principal. Auteurs : Junshen Xu et coll.*) [360]

Il s'agit d'un modèle prenant en entrée une image tomographique en faible dose et générant en sortie une image tomographique similaire mais de qualité comparable à une tomographie en dose standard (*figure 89*). Ce modèle parvient à reconstruire une image tomographique de qualité en administrant au/à la patient(e) seulement 1/200 de la dose standard de 18F-fluorodeoxyglucose.

La tomographie par émission de positons est communément employée en imagerie médicale, notamment dans le diagnostic de cancers. Cet examen nécessite l'injection d'un traceur dont les propriétés radioactives exposent le/la patient(e) à des risques [497]. La possibilité de réduire les doses de traceur injectées sans altérer la qualité des images tomographiques permettrait de réduire ces risques tout en préservant la justesse du diagnostic. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'améliorer automatiquement la qualité des images tomographiques en faible dose afin de rendre celles-ci exploitables par les soignants [360].

b. Cas d'usage en thérapeutique

Wei Mu et son équipe présentent en 2020 un modèle d'intelligence artificielle d'aide à la décision quant au traitement du cancer du poumon non à petites cellules [361].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*figure 90*). Les données d'entraînement sont constituées d'images tomographiques-tomodensitométriques (^{18}F -PET-CT) collectées auprès de 429 patients présentant un cancer du poumon non à petites cellules. Ces images ont été labellisées selon le statut de mutation de l'EGFR (*Epithelial Growth Factor Receptor*) déterminé par biopsie pour chacun des patients.

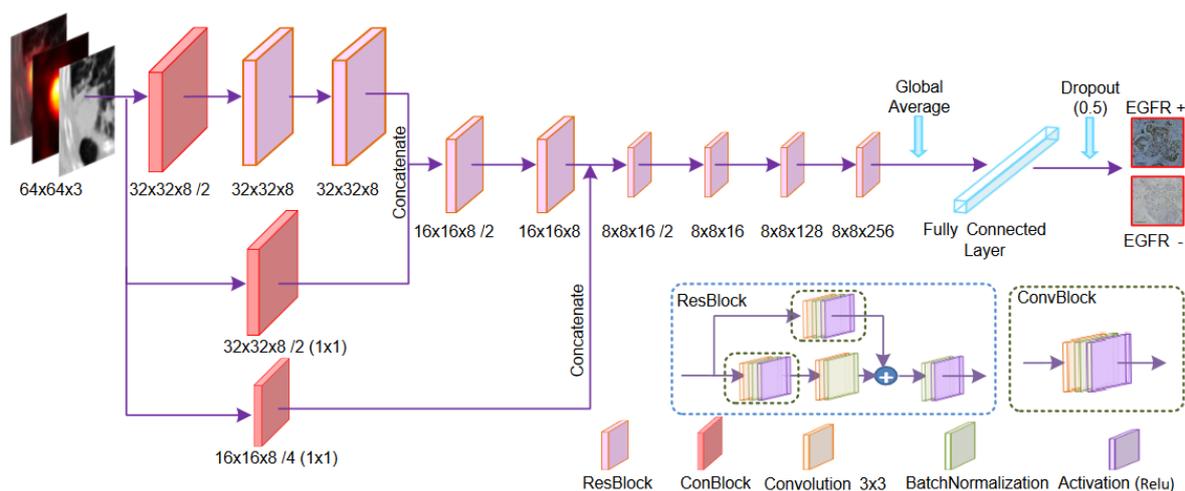


Figure 90. Représentation schématique du modèle présenté (SRResCNN) (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Wei Mu et coll.) [361]

Il s'agit donc d'un modèle de classification prenant en entrée une image tomographique-tomodensitométrique et indiquant en sortie la classe prédite : présence d'une mutation de l'EGFR vs. absence d'une mutation de l'EGFR (*figure 90*). Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 78,5 % ; aire sous la courbe de ROC = 81 %. Le choix de chimiothérapie contre le cancer du poumon non à petites cellules à un stade avancé (inhibiteur de la tyrosine kinase, inhibiteur de point de contrôle immunitaire) se fonde essentiellement sur la présence ou non d'une mutation de l'EGFR au sein des cellules tumorales [498]. Cependant, l'identification des mutations de l'EGFR nécessite une biopsie invasive associée à une morbidité importante [499]. Le test diagnostique est coûteux, chronophage et sa précision dépend de la quantité et de la qualité, difficile à évaluer durant la biopsie, des tissus prélevés [500]. De plus, le statut de mutation de l'EGFR peut évoluer avec le temps, nécessitant idéalement sa réévaluation régulière durant la chimiothérapie [501]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'estimer le statut de mutation de l'EGFR de façon automatique et non invasive, et pourrait ainsi renforcer la capacité des soignants à prescrire le traitement le plus adapté à chaque patient(e) contre le cancer du poumon non à petites cellules tout au long du parcours thérapeutique [361].

22. Médecine nutritionnelle

a. Cas d'usage en diagnostic

Liangyu Yin et son équipe présentent en 2021 des modèles d'intelligence artificielle visant à détecter les malnutritions chez des patients atteints de cancer [502].

Les données d'entraînement comportent des variables socio-démographiques, hygiéno-diététiques, cliniques et biologiques, des comorbidités, des antécédents familiaux, des variations pondérales, des scores nutritionnels, des indices de santé générale et des données relatives à la qualité de vie collectés auprès de 14134 patients atteints de cancer. En premier lieu, un apprentissage non supervisé a été effectué par un algorithme de k-moyennes afin de définir des niveaux de malnutrition (*figure 91*). Quatre clusters distincts ont été obtenus et qualifiés comme suit : 8193 patients n'étant pas en malnutrition, 2195 patients en malnutrition légère, 2491 patients en malnutrition modérée, 1255 patients en malnutrition sévère. Après validation de la pertinence clinique de ces clusters, des modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : régression logistique multiple, arbre de décision, forêt aléatoire, machine à vecteurs de support, apprentissage profond (*figure 91*).

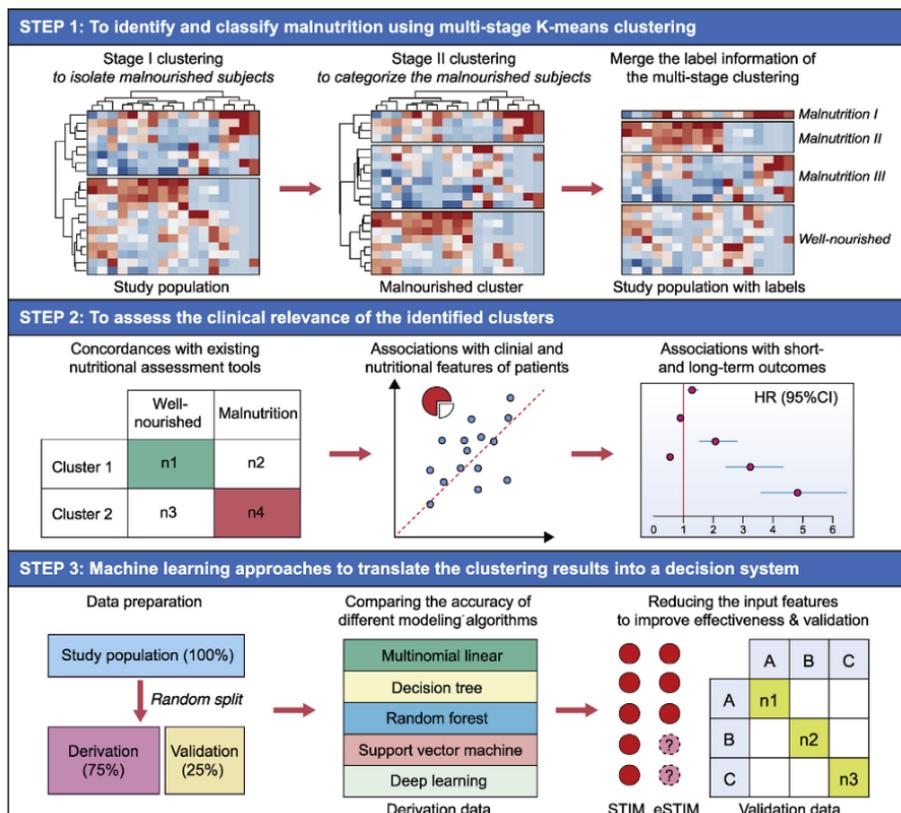


Figure 91. Illustration des différentes étapes de la démarche : identification de niveaux de malnutrition à l'aide d'un algorithme de partitionnement, validation de la pertinence clinique des niveaux définis, entraînement et validation des modèles de classification. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Liangyu Yin et coll.) [502]

Deux types de modèles ont été conçus : des modèles (nommés *STIM*) prenant en entrée l'ensemble des variables d'entraînement sus-mentionnées et des modèles (nommés *eSTIM*) prenant en entrée uniquement les variables ne nécessitant aucun acte invasif. Il s'agit dans tous les cas de modèles de classification indiquant en sortie la classe prédite : absence de malnutrition *vs.* présence d'une malnutrition légère *vs.* présence d'une malnutrition modérée *vs.* présence d'une malnutrition sévère. Les modèles de régression logistique multiple sont ceux présentant les meilleures performances : exactitude = 99,9 % ; kappa de Cohen = 99,9 % ; aire sous la courbe de ROC = 100 % pour *STIM* / exactitude = 94,3 % ; kappa de Cohen = 90,5 % ; aire sous la courbe de ROC = 94,1 % pour *eSTIM*.

Les patients atteints de cancer présentent un risque important de malnutrition en raison notamment des troubles métaboliques pouvant être induits par leur(s) tumeur(s) [503]. Cette malnutrition peut être à l'origine ou accentuer une cachexie, entraîner des complications, impacter l'efficacité des traitements, altérer la qualité de vie et globalement affecter négativement le pronostic du/de la patient(e) [504]. L'identification des malnutritions est par conséquent primordiale dans la prise en charge thérapeutique de ces patients. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait en mesure de renforcer la capacité des soignants à diagnostiquer les malnutritions chez des patients atteints de cancer, de façon automatique, précise et non invasive [502].

b. Cas d'usage en thérapeutique

Ya Lu et son équipe présentent en 2020 une architecture d'intelligence artificielle visant à estimer l'apport nutritionnel des repas chez des patients hospitalisés [505].

Le modèle qu'embarque cette architecture est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif de type auto-encodeur. Les données d'entraînement sont constituées d'images RGB-D^{xxviii} de 232 plateaux repas servis à des patients hospitalisés. Les pixels représentant de la nourriture ou des couverts ont été annotés sur les images du jeu de données.

Il s'agit donc d'un modèle de segmentation prenant en entrée deux photographies de plateau repas, l'une avant et l'autre après consommation par le/la patient(e), et indiquant en sortie les segments (groupes de pixels identifiés comme partageant des caractéristiques communes) correspondant à des aliments ou à des couverts. Cette segmentation est ensuite affinée à l'aide d'un modèle statistique (champ aléatoire conditionnel), puis le volume et la nature de nourriture

^{xxviii} Les cameras RGB-D (*Red Green Blue - Depth*) sont des périphériques d'acquisition vidéo ayant la capacité de détecter la profondeur. Celles-ci produisent des images en couleur dont chaque pixel est associé à une valeur numérique correspondant à la distance entre l'objectif de la caméra et la matière photographiée [506].

consommée sont estimés en comparant les segments identifiés sur les photographies avant et après le repas. Enfin, à partir des volumes estimés, l'apport nutritionnel est calculé (figure 92). Cette architecture parvient à estimer l'apport nutritionnel avec une erreur quadratique moyenne de 15 %.

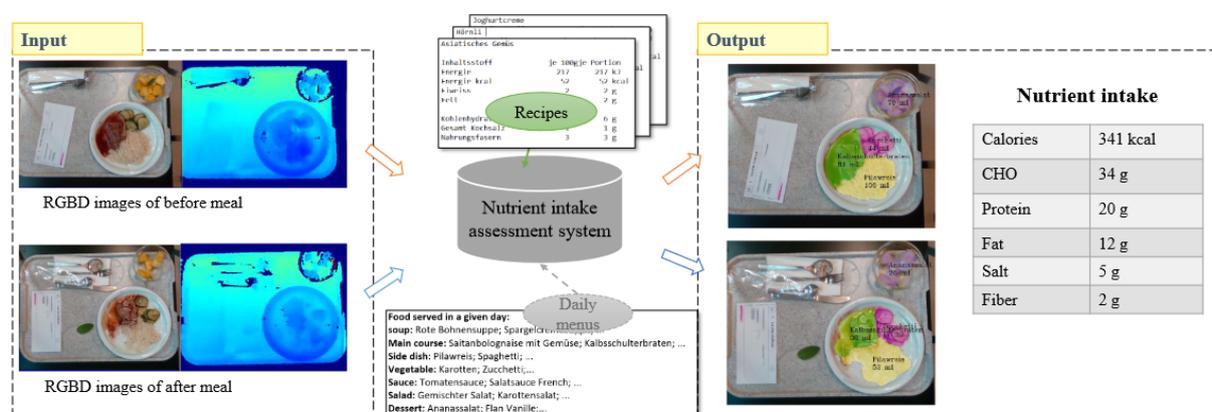


Figure 92. Représentation schématique de l'architecture présentée (Reproduit avec la permission de l'éditeur *Institute of Electrical and Electronics Engineers. Auteurs : Ya Lu et coll.*) [505]

Chez les patients hospitalisés, la malnutrition est fréquente, tend à prolonger leurs séjours hospitaliers et est associée à une augmentation de la mortalité et de la morbidité [507]. La malnutrition en contexte hospitalier est essentiellement due à un manque de suivi des apports nutritionnels individuels [508]. La possibilité d'évaluer ces apports nutritionnels de façon automatique faciliterait le suivi de la part des soignants et permettrait d'adapter au mieux les repas et les soins administrés. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement être en mesure d'estimer les apports nutritionnels des repas de façon automatique et précise, renforçant ainsi la capacité des soignants à suivre et adapter les régimes des patients hospitalisés [505].

23. Médecine palliative (cas d'usage en thérapeutique)

Anand Avati et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à prédire les besoins en soins palliatifs [509].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones. Les données d'entraînement sont constituées de caractéristiques extraites de dossiers médicaux informatisés relatifs à 164424 patients en vie et 12587 patients décédés.

Il s'agit d'un modèle de classification prenant en entrée des informations issues d'un dossier médical informatisé et indiquant en sortie la survie prédite au cours des 12 mois à venir. Le modèle est conçu pour analyser chaque jour les dossiers médicaux informatisés des patients admis au sein d'un service et indiquer aux soignants la liste des patients pour lesquels la survie prédite est inférieure à 12 mois ainsi que les facteurs pouvant expliquer cette prédiction. Ce

modèle parvient à atteindre les performances suivantes : valeur prédictive positive = 69 % ; sensibilité = 90 % ; aire sous la courbe de ROC = 93 %.

Les soins palliatifs représentent une dimension essentielle dans la prise en charge des patients atteints de maladie grave ou en phase terminale. Il subsiste pourtant un décalage important entre les besoins en soins palliatifs de ces patients et leur prise en charge factuelle [510]. Les principaux obstacles à l'amélioration de l'accès aux soins palliatifs sont la surestimation du pronostic et le manque de personnel soignant dédié [511]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait en mesure d'estimer automatiquement le pronostic vital des patients et pourrait ainsi renforcer la capacité des soignants à répondre plus précisément et proactivement aux besoins individuels en soins palliatifs [509].

24. Médecine physique et de réadaptation (cas d'usage en thérapeutique)

Anett Seeland et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à améliorer la performance des rééducations sensorimotrices faisant intervenir un exosquelette à neuro-interface [512,513].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un perceptron de type passif-agressif. Les données d'entraînement sont constituées de caractéristiques extraites d'électroencéphalographies et d'électromyographies (effectuées à partir des biceps) collectées auprès d'un patient au cours de 14 sessions de rééducation sensorimotrice via exosquelette à neuro-interface (*figure 93*). Chaque session comportait deux séries d'au moins 40 mouvements de chaque bras. Les données d'entraînement ont été segmentées et chaque segment a été labellisé selon le mouvement effectué par le patient : vers la droite, vers la gauche ou immobilité (*figure 94*).

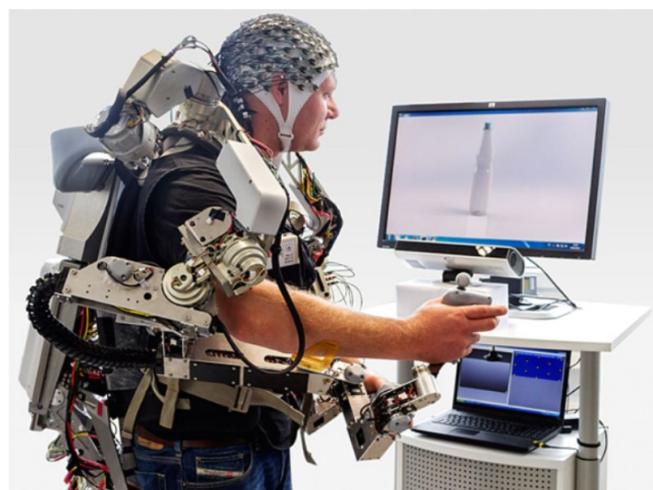


Figure 93. Rééducation du haut du corps assistée par un exosquelette actif. Le soutien du bras (droit ou gauche) par l'exosquelette est déclenché par l'activité de l'EEG et de l'EMG ainsi que par les données de suivi oculaire. (Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Anett Seeland et coll.) [512]

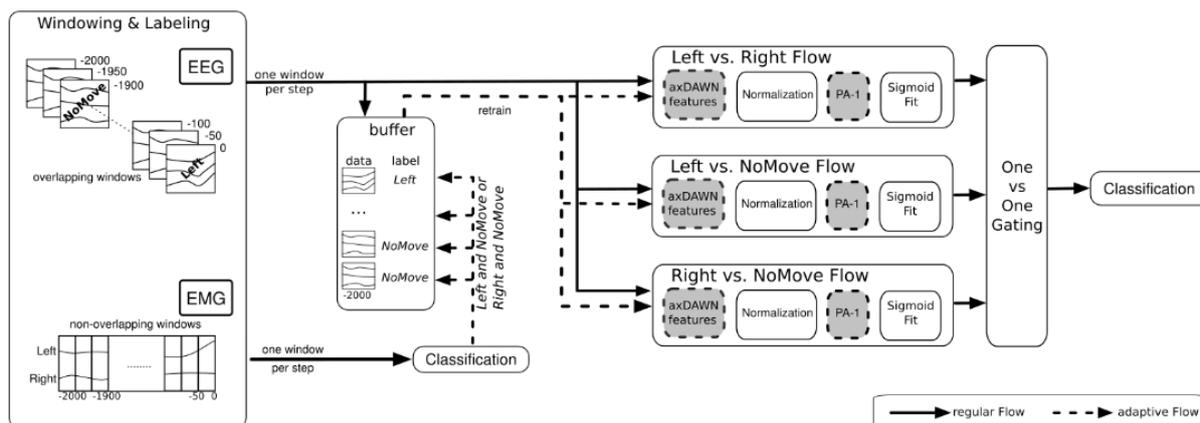


Figure 94. Collecte, pré-traitement et labellisation des données d'entraînement (Reproduit avec la permission de l'éditeur *Institute of Electrical and Electronics Engineers. Auteurs : Anett Seeland et coll.*) [512]

Il s'agit donc d'un modèle de classification prenant en entrée des caractéristiques extraites d'une électroencéphalographie et d'une électromyographie effectuée à partir d'un biceps et indiquant en sortie la classe prédite : bras en mouvement vers la droite vs. bras en mouvement vers la gauche vs. bras immobile (figure 94). Ce modèle parvient à atteindre des performances élevées : exactitude = 97,35%. Le temps d'exécution moyen du modèle est de 50 millisecondes, permettant une analyse en continu des mouvements réalisés par le patient.

Les exosquelettes suscitent de plus en plus l'intérêt en tant qu'outil thérapeutique en rééducation, notamment pour la prise en charge des séquelles sensorimotrices subséquentes à un accident vasculaire cérébral [514]. La possibilité de doter ces exosquelettes d'une intelligence artificielle et d'entraîner celle-ci à partir des données cérébrales et musculaires du patient permettrait de mieux personnaliser sa rééducation (ex. : rythme, exercices à effectuer). Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle embarqué dans un exosquelette serait en mesure d'identifier les mouvements du patient, de façon automatique, continue et précise, permettant ainsi une personnalisation de sa rééducation [512].

25. Néonatalogie

a. Cas d'usage en diagnostic

Henrik Verder et son équipe présentent en 2021 un modèle d'intelligence artificielle visant à détecter les dysplasies bronchopulmonaires chez des nouveau-nés prématurés [515].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de machine à vecteurs de support. Les données d'entraînement sont constituées des poids à la naissance, des âges gestationnels, des agents tensioactifs administrés et de caractéristiques issues de spectroscopies infrarouges obtenues à partir des liquides gastriques prélevés à la naissance (nombres d'onde identifiés comme significatifs). Ces données ont été labellisées après avoir été collectées auprès de 26 nouveau-nés prématurés présentant une dysplasie bronchopulmonaire et de 35 nouveau-nés prématurés en bonne santé. Aucun de ces nouveau-nés ne présentait de sepsis ou de prédispositions génétiques connues à la dysplasie bronchopulmonaire.

Il s'agit donc d'un modèle de classification prenant en entrée les variables d'entraînement susmentionnées et indiquant en sortie la classe prédite : présence d'une dysplasie bronchopulmonaire *vs.* absence d'une dysplasie bronchopulmonaire. Ce modèle parvient à atteindre des performances élevées : sensibilité = 88 % ; spécificité = 91 %.

La dysplasie bronchopulmonaire est une pathologie multifactorielle courante chez le nouveau-né prématuré, associée à un risque élevé de mortalité et de morbidité [516]. Sa prise en charge immédiate permet de limiter fortement ces risques [517]. Cependant, malgré l'existence de biomarqueurs corrélés à la présence de dysplasie bronchopulmonaire, il n'existe à ce jour aucun test diagnostique permettant son dépistage [518]. La possibilité d'identifier automatiquement les dysplasies bronchopulmonaires permettrait de traiter celles-ci précocement et de limiter les risques de complications chez les nouveau-nés atteints. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les dysplasies bronchopulmonaires chez les nouveau-nés prématurés, de façon automatique et précise [515].

b. Cas d'usage en thérapeutique

Yifei Hu et son équipe présentent en 2018 des modèles d'intelligence artificielle d'aide à la décision quant au traitement du sepsis à début tardif chez les nouveau-nés [519].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés : régression logistique, machine à vecteurs de support, forêt aléatoire, arbre de décision à renforcement du gradient, réseau de neurones. Les données d'entraînement sont constituées de signes vitaux (fréquence cardiaque, fréquence respiratoire et saturation en oxygène mesurées à chaque minute) sous forme de séries temporelles. Ces données ont été collectées auprès de nouveau-nés n'ayant présenté aucune anomalie et de nouveau-nés ayant présenté des symptômes justifiant l'administration d'antibiotiques. Les séries temporelles ont été fragmentées par segments de 60 minutes puis ces segments ont été labellisés selon que ceux-ci soient concomitants à un état de santé normal (2416) ou justifiant l'administration d'antibiotiques (637).

Il s'agit donc d'un modèle de classification prenant en entrée une série temporelle comportant les signes vitaux sus-mentionnés mesurés chez un nouveau-né et indiquant en sortie la classe prédite : administration d'antibiotiques non recommandée *vs.* administration d'antibiotiques recommandée. Le modèle d'arbre de décision à renforcement du gradient est celui présentant les meilleures performances : valeur prédictive positive = 72 % ; sensibilité = 72 % ; F-score pondéré = 64 % ; aire sous la courbe de ROC = 75 %.

Le sepsis est une pathologie commune chez les nouveau-nés, particulièrement chez les prématurés en raison de l'immaturation de leur système immunitaire [520]. Il s'agit d'une pathologie associée à une mortalité importante, en l'occurrence la mortalité du sepsis à début tardif (débutant 48 heures après la naissance) s'élève à 15,1% [521]. Le sepsis à début tardif échappe davantage à la vigilance des soignants que son homologue à début précoce car la surveillance de ce dernier s'inscrit généralement dans une routine clinique établie [522]. La possibilité de prédire précocement l'apparition du sepsis à début tardif permettrait aux soignants d'intervenir en amont des premiers symptômes alarmants et d'éviter les complications graves pour le nouveau-né. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de détecter les signes vitaux annonciateurs d'un sepsis à début tardif chez le nouveau-né, de façon automatique et non invasive, renforçant ainsi la capacité des soignants à traiter précocement cette pathologie [519].

26. Néphrologie

a. Cas d'usage en diagnostic

Hang Liu et son équipe présentent en 2022 un modèle d'intelligence artificielle visant à détecter des complications faisant suite à une transplantation rénale [523].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif. Les données d'entraînement sont constituées d'images obtenues par résonance magnétique auprès de 89 patients ayant reçu une transplantation rénale deux semaines auparavant. Parmi ces patients, 39 ont présenté une obstruction entre l'uretère et l'anastomose de la vessie, 13 ont enduré un rejet de la greffe, 10 ont présenté un hématome périrénal, 5 ont présenté un infarctus rénal, et 22 n'ont présenté aucune complication.

Il s'agit d'un modèle de segmentation prenant en entrée une image obtenue par résonance magnétique et indiquant en sortie la localisation et la nature des signes de complication. Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 80,6 % ; sensibilité = 67,8 % ; spécificité = 86,7 %.

Parmi les transplantations pratiquées, la transplantation rénale est la plus courante. Celle-ci représente le traitement offrant le meilleur pronostic aux patients atteints d'insuffisance rénale au stade terminal [524]. Néanmoins, l'incidence d'insuffisance rénale post-transplantation demeure élevée, en raison notamment de ses complications subséquentes. L'une des clés majeures pour réduire ce fardeau réside dans la précocité du diagnostic et du traitement de ces complications [525]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les complications faisant suite à une transplantation rénale, de façon automatique, précise et non invasive, pouvant ainsi favoriser la réussite à long terme de la transplantation [523].

b. Cas d'usage en thérapeutique

Ahmed I. Akl et son équipe présentent en 2001 des modèles d'intelligence artificielle visant à estimer un SRI (*solute removal index*) durant une hémodialyse et à prédire le temps nécessaire pour atteindre un SRI cible [343,525].

Ces modèles sont obtenus à l'issue d'un apprentissage automatique supervisé réalisé sur deux réseaux de neurones (*figure 95*). Les données d'entraînement sont constituées de concentrations sanguines d'azote uréique, des poids des patients, des durées de dialyse et des SRI (élimination totale de l'azote uréique mesurée par quantification directe du dialysat) collectés auprès de 15 patients au cours d'une hémodialyse. Ces mesures ont été effectuées toutes les 30 minutes durant les hémodialyses.

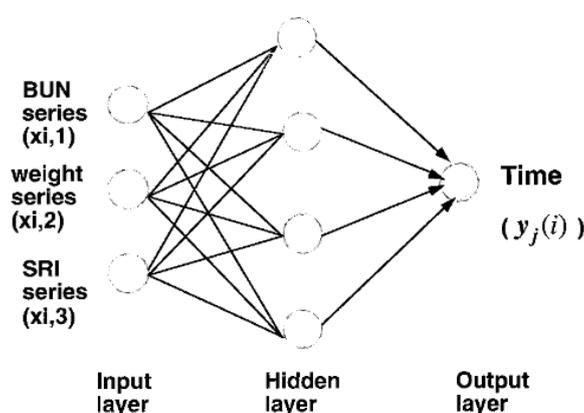


Figure 95. Représentation schématique du second modèle présenté (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Ahmed I. Akl et coll.) [525]

Il s'agit de modèles de régression. Le premier modèle prend en entrée le poids du/de la patient(e), sa concentration sanguine d'azote uréique (BUN), la durée de dialyse associée et indique en sortie le SRI prédit. Son erreur moyenne de prédiction s'élève à 10,9 %. Le second modèle prend en entrée le poids du/de la patient(e), sa concentration sanguine d'azote uréique (BUN), le SRI cible et indique en sortie la durée de dialyse optimale pour atteindre ce SRI (*figure 95*). Son erreur moyenne de prédiction s'élève à 8,3 %.

L'estimation des effets d'une hémodialyse sur un(e) patient(e) est généralement réalisée à partir de modèles mathématiques formels mais impliquant de nombreuses suppositions [526]. Cette contrainte est source d'imprécision, c'est pourquoi la modélisation mathématique formelle de l'urémie au cours d'une hémodialyse est contestée. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de prédire les effets d'une hémodialyse de façon automatique, précise et non invasive, permettant ainsi aux soignants d'optimiser ses paramètres tels que la durée notamment [525].

27. Neurologie

a. Cas d'usage en diagnostic

Taha Khan et son équipe présentent en 2014 des modèles d'intelligence artificielle visant à détecter la maladie de Parkinson et à estimer son stade de sévérité [527,528].

Ces modèles sont obtenus à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de machine à vecteurs de support. Les données d'entraînement sont constituées de caractéristiques extraites d'enregistrements vidéo montrant des tests de tapotement rapide des doigts (*figure 96*). Les caractéristiques extraites étaient relatives à la vitesse, l'amplitude, le rythme et la fatigue caractérisant les mouvements enregistrés. Parmi ces enregistrements, 387 proviennent de 13 patients présentant une maladie de Parkinson avancée et 84 proviennent de six patients contrôle en bonne santé. Sur la base de l'échelle UPDRS (*Unified Parkinson's Disease Rating Scale*), des stades de sévérité ont été attribués en guise de labels à chaque enregistrement provenant des patients atteints de la maladie de Parkinson.

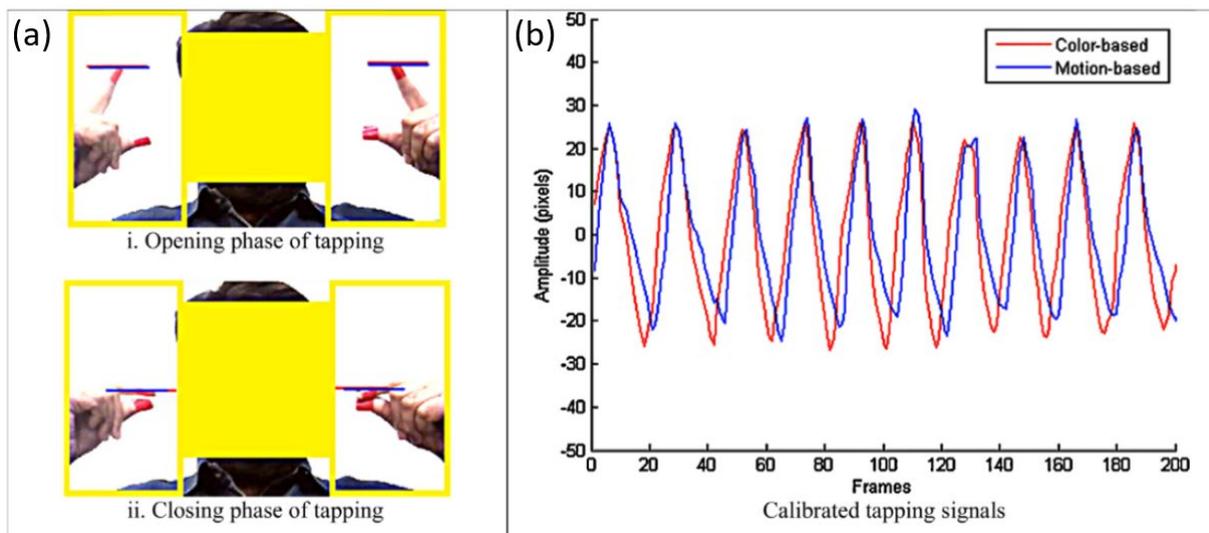


Figure 96. Test de tapotement rapide des doigts. (a) Phases d'ouverture et de fermeture du tapotement avec des marqueurs de position de l'index en mouvement. (b) Amplitude des tapotements en fonction du temps. (*Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Taha Khan et coll.*) [527]

Il s'agit donc de modèles de classification prenant en entrée des caractéristiques extraites d'un enregistrement vidéo montrant un test de tapotement rapide des doigts et indiquant en sortie les classes prédites : présence d'une maladie de Parkinson vs. absence d'une maladie de Parkinson / stade de sévérité des symptômes pouvant aller de 0 (pour normal) à 3 (pour sévère). Ces modèles parviennent à atteindre des performances élevées : exactitude = 95 % / exactitude = 88 %.

Le test de tapotement rapide des doigts permet d'évaluer les mouvements rapides et ainsi de déceler des troubles moteurs [529]. Il peut notamment en être fait usage pour appuyer un diagnostic de la maladie de Parkinson et identifier son stade de sévérité. Cependant, l'évaluation à l'œil nu du tapotement rapide des doigts comporte une part de subjectivité et induit par conséquent des résultats imprécis [530]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure d'évaluer le tapotement rapide des doigts de façon automatique, précise et non invasive, pouvant ainsi renforcer la capacité des soignants à diagnostiquer la maladie de Parkinson et son stade de sévérité [527].

b. Cas d'usage en thérapeutique

Jiansong Fang et son équipe présentent en 2022 un modèle d'intelligence artificielle visant à identifier de nouvelles cibles thérapeutiques en vue de traiter la maladie d'Alzheimer par repositionnement médicamenteux [531].

Ce modèle est obtenu à l'issue d'un apprentissage automatique non supervisé réalisé par un algorithme de réseau bayésien. Les données d'entraînement comportent des résultats d'études d'association pangénomique, des données multiomiques issues d'échantillons cérébraux prélevés chez des patients atteints de la maladie d'Alzheimer ou chez des animaux transgéniques, des réseaux d'interactions médicament-cible, des interactomes protéine-protéine humains et des mécanismes de cellules microgliales humaines observés *in vitro*.

Il s'agit d'un modèle bayésien prenant en entrée les données d'entraînement sus-mentionnées et indiquant en sortie une prédiction des gènes causaux et des cibles thérapeutiques potentielles dans le traitement de la maladie d'Alzheimer (*figure 97*). Ce modèle est parvenu à identifier correctement 103 gènes causaux de la maladie d'Alzheimer ainsi que trois molécules médicamenteuses (pioglitazone, febuxostat et aténolol) pouvant prétendre à un repositionnement.

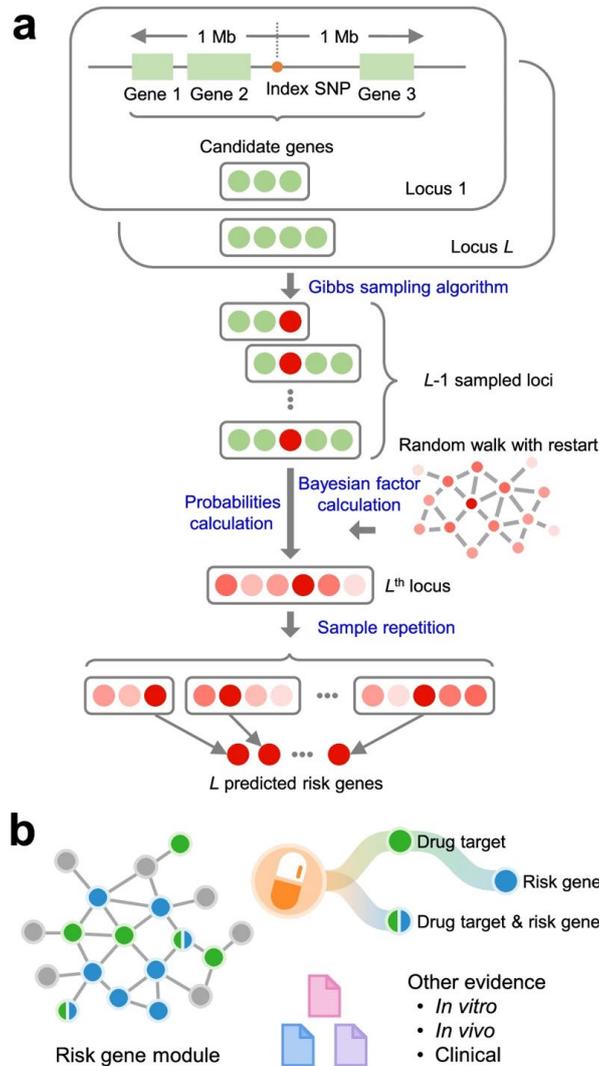


Figure 97. Méthodologie de réseau basée sur les études d'association pangénomique portant sur la maladie d'Alzheimer. (a) Algorithme bayésien, intégrant des données multiomiques et des réseaux de gènes, appliqué à l'identification de locus associés au développement de la maladie d'Alzheimer. (b) Repositionnement médicamenteux basé sur les gènes identifiés comme à risque et le réseau d'interactomes humains. (Reproduit selon les termes de la licence CC BY 4.0. Auteurs Jiansong Fang et coll.) [531]

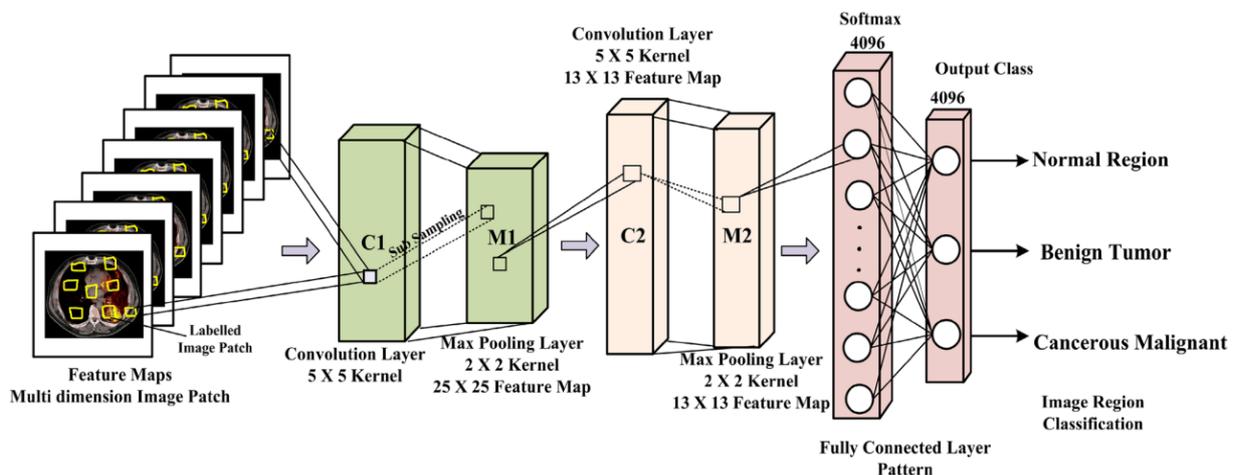
Les démences constituent un fardeau dont la prévalence est en hausse en raison du vieillissement de la population [532]. De nos jours, nous dénombrons 50 millions de personnes atteintes de démence dans le monde et la maladie d'Alzheimer est estimée comme en étant l'origine dans 60 à 70 % des cas [533]. Même si d'importants progrès ont été réalisés dans le traitement de la maladie d'Alzheimer, l'amélioration des stratégies thérapeutiques subsiste à ce jour comme étant un enjeu majeur. Les études d'association pangénomique ont permis d'identifier de nombreux locus relatifs à la maladie d'Alzheimer, cependant ces découvertes peinent à mener à l'identification de gènes causaux et de cibles thérapeutiques [534]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait être en mesure d'exploiter ces locus pour identifier automatiquement des gènes causaux, renforçant ainsi la capacité des chercheurs à identifier de nouvelles cibles thérapeutiques en vue de traiter la maladie d'Alzheimer notamment par repositionnement médicamenteux [531].

28. Odontologie

a. Cas d'usage en diagnostic

Pandia Rajan Jeyaraj et son équipe présentent en 2019 des modèles d'intelligence artificielle visant à détecter précocement les cancers de la cavité buccale [535,536].

Ces modèles sont obtenus à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur des réseaux de neurones convolutifs (*Inception v3*) (figure 98). Les données d'entraînement sont constituées d'images hyperspectrales de cavités buccales labellisées selon que celles-ci indiquent la présence d'une tumeur maligne, d'une tumeur bénigne ou une absence de tumeur. Ces images ont été collectées via des entrepôts de données en libre accès : *BioGPS, The Cancer Imaging Archive, Genomic Data Commons*.



Il s'agit donc de deux modèles de classification prenant en entrée une image hyperspectrale de cavité buccale. Le premier modèle indique en sortie la classe prédite : présence d'une tumeur maligne vs. présence d'une tumeur bénigne (figure 98). Ce modèle parvient à atteindre des performances élevées : exactitude = 91,4 % ; sensibilité = 94 % ; spécificité = 91 %. Le second modèle indique en sortie la classe prédite : présence d'une tumeur maligne vs. absence de tumeur (figure 98). Ce modèle parvient également à atteindre des performances élevées : exactitude = 94,5 % ; sensibilité = 94 % ; spécificité = 98 %.

Les cancers de la cavité buccale font partie des cancers les plus courants et les plus mortels. En 2020, l'Organisation mondiale de la Santé (OMS) estimait mondialement le nombre de nouveaux cas à 377 713 et le nombre de décès à 177 757 [537]. « Si son pronostic est défavorable cinq ans après le diagnostic (taux de survie s'élevant seulement à 40%), son diagnostic précoce permet en revanche de hausser le taux de survie à 80 % [538]. De fait, les recommandations

suggèrent d'effectuer une biopsie dès lors qu'une lésion de la cavité buccale est suspecte [539]. La possibilité d'identifier précocement les tumeurs de la cavité buccale à partir d'une imagerie permettrait d'éviter l'invasivité d'une biopsie. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, l'imagerie hyperspectrale pourrait potentiellement être considérée en tant qu'élément suffisant pour effectuer un diagnostic de cancer de la cavité buccale de façon automatique, précise et non invasive [535].

b. Cas d'usage en thérapeutique

Peilin Li et son équipe présentent en 2019 des modèles d'intelligence artificielle d'aide à la décision dans le cadre d'une extraction dentaire [540].

Ces modèles sont obtenus à l'issue d'un apprentissage automatique supervisé réalisé sur des réseaux de neurones (perceptrons à trois couches) (figure 99). Les données d'entraînement comportent 24 variables démographiques, céphalométriques, dentaires et tissulaires (figure 100). Ces données ont été collectées à partir des dossiers médicaux de 182 patients s'appêtant à recevoir un traitement orthodontique.

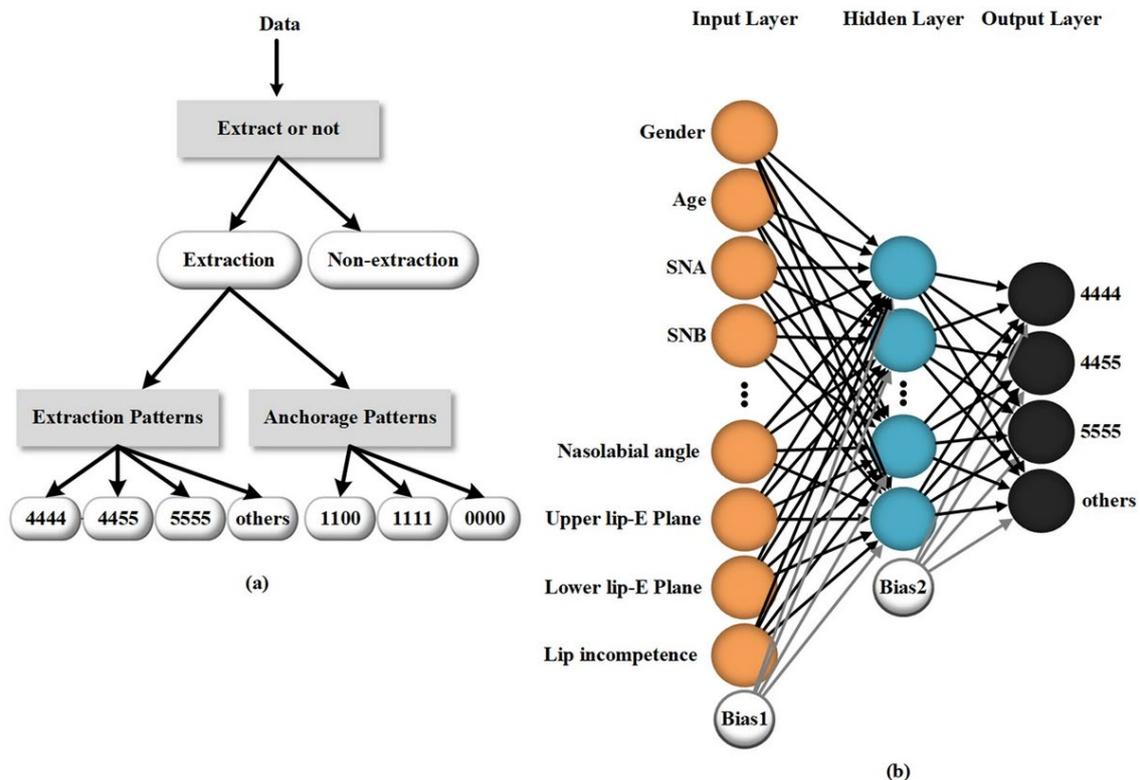


Figure 99. (a) Arbre faisant figurer les différentes modalités organisées selon leur ordre d'intervention dans le processus de planification d'extraction dentaire. (b) Représentation schématique du second modèle. (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Peilin Li et coll.) [540]

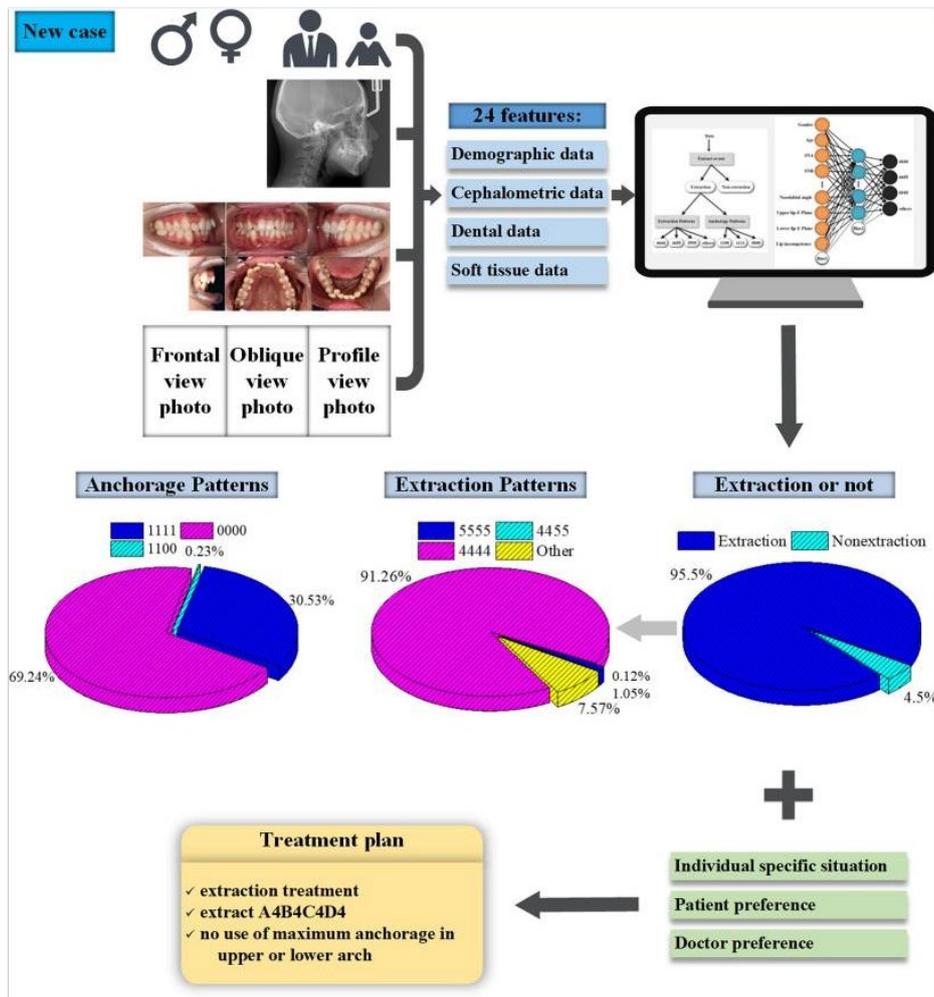


Figure 100. Vue d'ensemble de l'entraînement des modèles et de leur application en contexte clinique
(Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Peilin Li et coll.) [540]

Il s'agit de trois modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées. Le premier modèle indique en sortie la prédiction relative au besoin d'extraction : extraction dentaire nécessaire vs. extraction dentaire non nécessaire (*figure 100*). Ce modèle parvient à atteindre des performances élevées : exactitude = 94 % ; sensibilité = 94,6 % ; spécificité = 93,8 % ; aire sous la courbe de ROC = 98,2 %. Le second modèle indique en sortie la prédiction du type d'extraction nécessaire : extraction de la première prémolaire maxillaire et mandibulaire (4444) vs. extraction de la première prémolaire maxillaire et de la deuxième prémolaire mandibulaire (4455) vs. extraction de la deuxième prémolaire maxillaire et mandibulaire (5555) vs. autre type d'extraction nécessaire (*figures 99 et 100*). Ce modèle parvient à atteindre des performances relativement élevées : exactitude = 84,2 %. Le troisième modèle indique en sortie la prédiction relative à la nécessité d'utiliser un ancrage maximal au cours de l'extraction : ancrage maxillaire maximal (1100) vs. ancrages maxillaire et mandibulaire maximaux (1111) vs. ancrage maximal non nécessaire (0000) (*figure 100*). Ce modèle parvient à atteindre des performances élevées : exactitude = 92,8 %.

L'efficacité et l'efficience des traitements orthodontiques dépendent en grande partie de leur planification [541]. Celle-ci repose sur l'analyse au cas par cas de nombreux facteurs dont certains ne peuvent faire l'objet d'une évaluation objective [542]. La possibilité d'établir automatiquement une planification optimale fondée sur les données spécifiques du/de la patient(e) permettrait aux orthodontistes, notamment les moins expérimentés, d'envisager plus aisément et plus précisément leurs interventions. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux orthodontistes d'établir leurs planifications d'extraction dentaire de façon automatique, précise et non invasive [540].

29. Oncologie

a. Cas d'usage en diagnostic

Alejandro Rodriguez-Ruiz et son équipe présentent en 2019 un modèle d'intelligence artificielle visant à détecter les cancers du sein [543,544].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif. Les données d'entraînement sont constituées de mammographies numérisées présentant ou non une tumeur maligne. L'ensemble de ces mammographies ont été labellisées selon le degré de suspicion de présence d'une tumeur maligne (allant de 1 à 10).

Il s'agit donc d'un modèle de classification prenant en entrée une mammographie numérisée et indiquant en sortie le degré de suspicion prédit pouvant aller de 1 (absence de tumeur) à 10 (présence très probable d'une ou plusieurs tumeurs malignes). Ce modèle parvient à atteindre des performances comparables à l'état de l'art : aire sous la courbe de ROC = 84 %. Les comparaisons statistiques menées indiquent une différence non significative entre les performances observées de la part de ce modèle (IC de l'AUC = [82% ; 86%]) et celles observées au sein d'un groupe de 101 radiologues (IC de l'AUC = [78,7% ; 84,1%]).

Le cancer du sein est actuellement le cancer le plus fréquent et le plus mortel chez les femmes. En 2020, l'OMS estimait mondialement le nombre de nouveaux cas à 2 261 419 et le nombre de décès à 684 996 [537]. La mammographie, promue par les campagnes de prévention, a démontré son efficacité en tant qu'outil de dépistage précoce en réduisant la mortalité des cancers du sein [545]. Cependant, le nombre croissant de mammographies à effectuer et à analyser allié à la pratique de la deuxième lecture aboutissent à une charge travail en forte croissance. Cette croissance est plus rapide que celle de l'effectif médical qualifié et disponible [546]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à dépister les cancers du sein, de façon automatique, précise et non invasive [543].

b. Cas d'usage en thérapeutique

Yoosup Chang et son équipe présentent en 2018 une architecture d'intelligence artificielle visant à prédire l'efficacité de médicaments administrés en tant que traitement anti-cancéreux [547,548].

Cette architecture, nommée CDRscan (*Cancer Drug Response profile scan*), est composée de cinq modèles obtenus à l'issue d'un apprentissage profond supervisé réalisé sur cinq réseaux de neurones convolutifs différents. Les données d'entraînement comportent 787 lignées cellulaires correspondant à 25 types de cancers, des informations relatives à 28 328 mutations de 567 gènes, ainsi que des concentrations inhibitrices médianes correspondant à 244 médicaments (*figure 101*). Ces données ont été collectées via les bases de données CCLP (*COSMIC cell line project*) et GDSC (*Genomics in Drug Sensitivity in Cancer*) (*figure 101*).

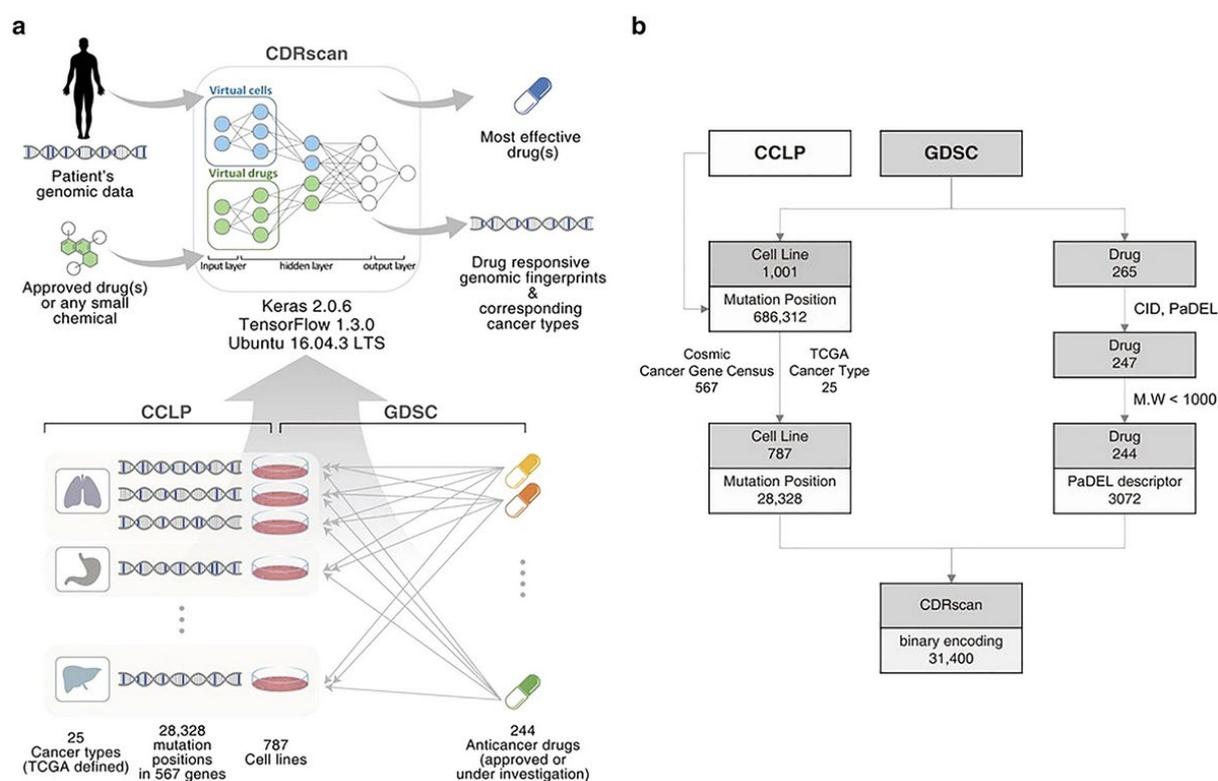


Figure 101. Vue d'ensemble de l'architecture présentée (CDRscan). (a) Représentation schématique de CDRscan et de son entraînement. (b) Collecte et traitement des données d'entraînement. (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Yoosup Chang et coll.) [547]

Il s'agit de modèles de régression prenant en entrée des empreintes génomiques tumorales ainsi que les motifs moléculaires d'un ou plusieurs traitements médicamenteux et estimant en sortie des concentrations inhibitrices médianes. CDRscan calcule ensuite la moyenne des valeurs numériques estimées pour chaque traitement par les cinq modèles et indique en sortie la concentration inhibitrice médiane associée à chaque traitement. CDRscan parvient à atteindre des performances élevées : coefficient de détermination = 0,843 ; racine de l'erreur quadratique

moyenne = 0,98. En outre, parmi 1487 molécules médicamenteuses, CDRscan a pu identifier 37 molécules pouvant potentiellement faire l'objet d'un repositionnement en tant que traitement anti-cancéreux.

Les progrès réalisés en génomique permettent aujourd'hui d'envisager une adaptation des traitements anti-cancéreux selon le profil génomique de la tumeur cible [549]. Le caractère abondant des données génomiques relatives aux cancers permet à celles-ci d'être exploitées par des méthodes d'apprentissage automatique, au service de la médecine de précision [550]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait en mesure d'estimer l'efficacité de médicaments administrés en tant que traitement anti-cancéreux en se fondant sur les données génomiques de la tumeur cible, de façon automatique et précise [547].

30. Ophtalmologie

a. Cas d'usage en diagnostic

Zhixi Li et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à détecter les neuropathies optiques glaucomateuses [362,551].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*Inception v3*). Les données d'entraînement sont constituées de 31 745 photographies de fond d'œil en couleur collectées à partir d'un jeu de données en libre accès (*LabelMe*). L'ensemble des photographies ont été labellisées selon que celles-ci présentent une neuropathie optique glaucomateuse ou non.

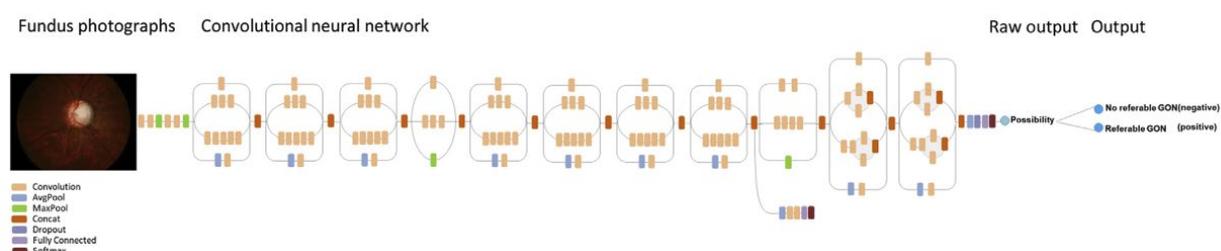


Figure 102. Représentation schématique du modèle présenté (*Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Zhixi Li et coll.*) [362]

Il s'agit donc d'un modèle de classification prenant en entrée une photographie de fond d'œil en couleur et indiquant en sortie la classe prédite : présence d'une neuropathie optique glaucomateuse vs. absence de neuropathie optique glaucomateuse (*figure 102*). Ce modèle parvient à atteindre des performances élevées : exactitude = 92,9 % ; sensibilité = 95,6 % ; spécificité = 92 % ; aire sous la courbe de ROC = 98,6 %.

Selon l'Institut national de la santé et de la recherche médicale (Inserm), le glaucome concerne 1 à 2 % de la population ayant plus de 40 ans et environ 10% de la population ayant plus de 70 ans [552]. En raison de la croissance démographique et du vieillissement de la population, le nombre de personnes atteintes mondialement en 2040 serait estimé à 112 millions [553]. Lorsque les glaucomes ne font pas l'objet d'un traitement, ceux-ci peuvent engendrer des complications oculaires graves pouvant aller jusqu'à une cécité irréversible [554]. Son diagnostic précoce permet de grandement favoriser le pronostic visuel, il s'agit cependant d'une pathologie généralement asymptomatique [555]. De fait, plus de la moitié des personnes présentant un glaucome seraient non diagnostiquées avant l'apparition des premières complications [556]. Le diagnostic du glaucome s'appuie sur plusieurs examens dont celui du nerf optique par le fond d'œil [557]. Il s'agit d'un examen dont l'analyse repose sur l'appréciation visuelle de l'ophtalmologiste, celui-ci comporte donc une part de subjectivité [558]. Effectuer l'ensemble des examens recommandés est, en revanche, fastidieux et coûteux. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, les photographies de fond d'œil pourraient potentiellement être considérées comme étant suffisantes pour réaliser un diagnostic de glaucome, de façon automatique, précise, moins coûteuse et non invasive [362].

b. Cas d'usage en thérapeutique

Reza Rasti et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à prédire la réussite d'un traitement anti-VEGF (*Vascular Endothelial Growth Factor*) dans le cadre d'un œdème maculaire diabétique [558,559].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*figure 103*). Les données d'entraînement sont constituées d'images tomographiques par cohérence optique de rétines. Ces images ont été collectées auprès de 127 patients après qu'ils aient reçu une injection d'agent anti-VEGF (pré-traitement) et après qu'ils aient reçu trois injections d'agent anti-VEGF (post-traitement). Chaque paire d'images (pré-traitement et post-traitement) a été labellisée par la différence d'épaisseur rétinienne associée afin d'entraîner le modèle à estimer cette valeur à partir d'une image pré-traitement et prédire si un traitement anti-VEGF serait efficace (réduction d'au moins 10% de l'épaisseur rétinienne) ou non. Parmi les patients inclus, 80 d'entre eux ont présenté une réduction de l'épaisseur rétinienne d'au moins 10% tandis que les 47 restants ont présenté une réduction de l'épaisseur rétinienne inférieure à 10%.

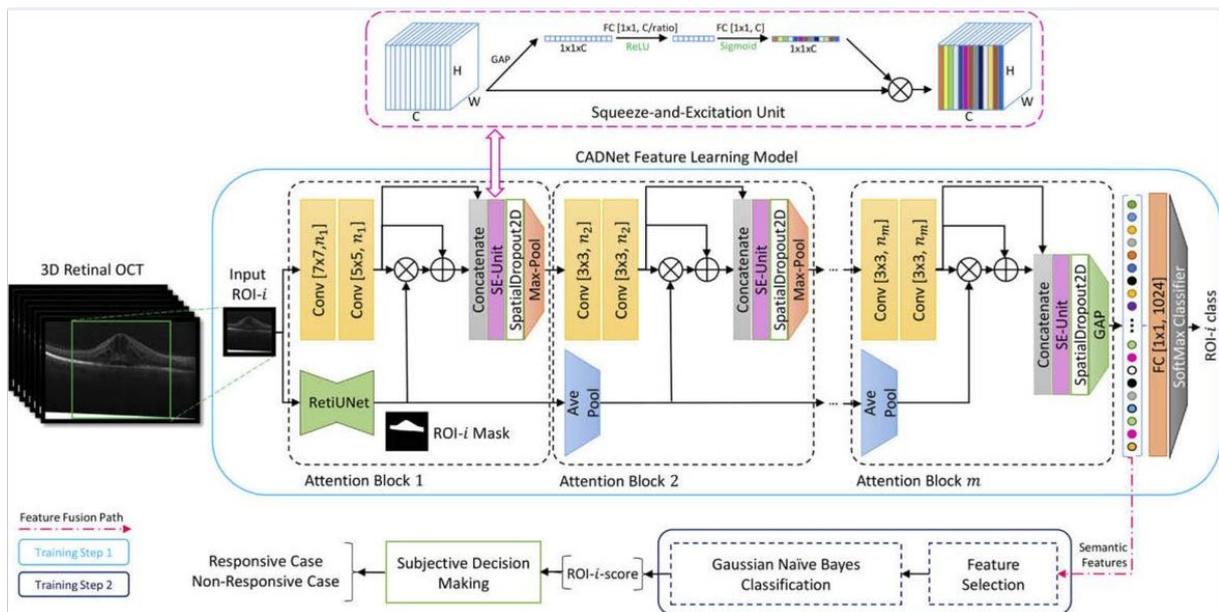


Figure 103. Représentation schématique du modèle présenté (CADNet) (Reproduit selon les termes de la licence OSA Open Access. Auteurs : Reza Rasti et coll.) [558]

Il s'agit donc d'un modèle de classification prenant en entrée une image tomographique par cohérence optique de rétine après une injection d'anti-VEGF et indiquant en sortie la classe prédite : réussite de la poursuite du traitement anti-VEGF vs. échec de la poursuite du traitement anti-VEGF (figure 103). Ce modèle parvient à atteindre des performances relativement élevées : sensibilité = 80,1 % ; spécificité = 85 % ; valeur prédictive positive = 85,5 % ; aire sous la courbe de ROC = 86,6 %. Le temps d'exécution moyen du modèle est d'environ 137 millisecondes.

Chez les patients atteints de rétinopathie diabétique, l'œdème maculaire diabétique non traité est l'une des causes majeures de perte de la vision centrale [560]. Le traitement d'un œdème maculaire diabétique se décline en différentes options plus ou moins efficaces selon les patients [561]. La possibilité de prédire l'efficacité de ces traitements chez un(e) patient(e) donné(e) permettrait d'optimiser la prise de décision thérapeutique. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle pourrait potentiellement permettre aux soignants de prédire la réussite d'un traitement anti-VEGF dans le cadre d'un œdème maculaire diabétique, de façon automatique, rapide, précise et non invasive [558].

31. Orthopédie

a. Cas d'usage en diagnostic

Jakub Olczak et son équipe présentent en 2017 des modèles d'intelligence artificielle visant à détecter les fractures [562,563].

Ces modèles sont obtenus à l'issue d'apprentissages par transfert puis d'apprentissages profonds supervisés réalisés sur cinq réseaux de neurones convolutifs (*CaffeNet*, *VGG-S*, *VGG-16*, *VGG-19*, *Network-in-network*). Les données d'entraînement comportent 256 458 radiographies de mains, de poignets et de chevilles collectées au format DICOM (*Digital Imaging and Communications in Medicine*) et comportant les métadonnées suivantes : latéralité (membre droit ou gauche), type de vue (antéropostérieure, latérale, oblique, etc.), partie du corps (main, poignet, cheville). Ces radiographies ont également été labellisées selon que celles-ci présentent une fracture ou non. Parmi les radiographies utilisées, 56% d'entre elles présentaient des fractures.

Il s'agit donc d'un modèle de classification prenant en entrée une radiographie de main, de poignet ou de cheville et indiquant en sortie les classes prédites correspondant aux quatre variables sus-mentionnées : présence d'une fracture *vs.* absence de fracture / membre droit *vs.* membre gauche / vue antéropostérieure *vs.* vue latérale *vs.* vue oblique *vs.* vue proximale *vs.* vue distale *vs.* vue ulnaire *vs.* vue radiale / main *vs.* poignet *vs.* cheville. Le réseau de neurones entraîné à partir de *VGG-16* est celui présentant les meilleures performances : exactitude = 83 % ; kappa de Cohen = 76 %.

L'interprétation des radiographies nécessite des années d'expérience et comporte par nature une part de subjectivité [564]. La fiabilité du diagnostic de fracture à partir d'une radiographie serait donc grandement favorisée par des méthodes d'interprétation objectives. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer les fractures à partir de radiographies, de façon automatique, précise et non invasive [562].

b. Cas d'usage en thérapeutique

Jun S. Kim et son équipe présentent en 2018 des modèles d'intelligence artificielle visant à prédire les complications résultant d'une arthrodèse lombaire postérieure [565].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés, le premier étant réalisé par un algorithme de régression logistique, le second sur un réseau de neurones. Les données d'entraînement comportent des variables sociodémographiques (genre, âge, ethnie), hygiéno-diététiques (tabagisme), cliniques (indice de masse corporelle, score ASA) ainsi que des données relatives à l'état de santé du/de la patient(e) (état fonctionnel, diabète, coagulopathie, comorbidités pulmonaires ou cardiaques) et aux traitements pris par celle-ci/celui-ci (stéroïdes). Ces données ont été collectées auprès de 22 629 patients ayant subi une arthrodèse lombaire postérieure.

Il s'agit de modèles de classification prenant en entrée des valeurs correspondant aux 15 variables sus-mentionnées et indiquant en sortie les complications post-opératoires prédites : survenue de complications cardiaques *vs.* absence de complications cardiaques / survenue d'une maladie thromboembolique veineuse *vs.* absence d'une maladie thromboembolique veineuse / survenue de complications de plaie *vs.* absence de complications de plaie / décès du patient *vs.* survie du patient. Les modèles parvenant à atteindre les meilleures performances pour chaque type de complications sont les suivants : réseau de neurones pour les complications cardiaques (aire sous la courbe de ROC = 71 %) / régression logistique pour les maladies thromboemboliques veineuses (aire sous la courbe de ROC = 58,8 %), les complications de plaie (aire sous la courbe de ROC = 61,3 %) et la mortalité (aire sous la courbe de ROC = 70,3 %). Les deux modèles sont significativement plus précis que le score ASA pour prédire chaque type de complications énoncé.

L'arthrodèse lombaire postérieure est une stratégie thérapeutique efficace pour traiter diverses affections dégénératives de la colonne lombaire [566,567]. Cependant, il s'agit d'une opération susceptible d'entraîner des complications sévères pouvant aller jusqu'au décès du/de la patient(e) [568]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire les complications résultant d'une arthrodèse lombaire postérieure, de façon automatique, non invasive et plus précise que les méthodes standards [565].

32. Oto-rhino-laryngologie

a. Cas d'usage en diagnostic

Shih-Hau Fang et son équipe présentent en 2019 des modèles d'intelligence artificielle visant à détecter les dysphonies [569,570].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés : machine à vecteurs de support, mélange gaussien, réseau de neurones. Les données d'entraînement sont constituées de caractéristiques extraites d'enregistrements vocaux collectés auprès de 462 patients. Ces caractéristiques ont été labellisées selon que le/la patient(e) présente (402) ou non (60) une dysphonie.

Il s'agit donc de modèles de classification prenant en entrée des caractéristiques extraites d'un enregistrement vocal et indiquant en sortie la classe prédite : présence d'une dysphonie vs. absence de dysphonie (figure 104). Le modèle de réseau de neurones est celui présentant les meilleures performances : exactitude = 94,26 % pour les hommes, 90,52% pour les femmes.

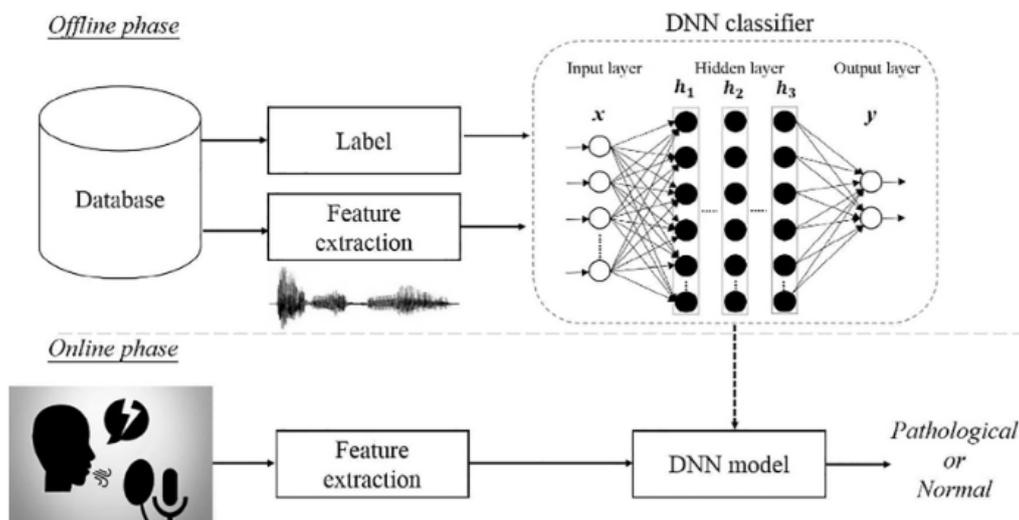


Figure 104. Représentation schématique de l'architecture présentée. L'entraînement du modèle (hors ligne) et ses prédictions (en ligne) sont respectivement schématisés dans les parties supérieure et inférieure de la figure. (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Shih-Hau Fang et coll.) [569]

Les dysphonies peuvent considérablement altérer la qualité de vie et les performances professionnelles des personnes atteintes [571]. De plus, une dysphonie peut être le signe de la présence d'une tumeur [572]. La possibilité de détecter les dysphonies, automatiquement et objectivement à partir d'un enregistrement vocal, permettrait d'accélérer et faciliter leur diagnostic, notamment en téléconsultation. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à télédiagnostiquer les dysphonies, de façon automatique et précise [569].

b. Cas d'usage en thérapeutique

Hosung Kim et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à prédire l'efficacité d'un implant cochléaire chez des patients atteints de surdité post-linguale (apparaissant après l'apprentissage du langage) [570,573].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de forêt aléatoire. Les données d'entraînement comportent les variables suivantes : âge au moment de l'opération, durée de la surdité, durée d'utilisation des appareils auditifs, seuil d'audition préopératoire, score de reconnaissance de phrases préopératoire. Ces données ont été collectées auprès de 120 patients atteints de surdité post-linguale et ayant subi une implantation cochléaire. Des scores de reconnaissance de mots (*word recognition score*) postopératoires ont été mesurés et attribués à chaque patient(e) en guise de labels.

Il s'agit donc d'un modèle de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie une valeur numérique correspondant au score de reconnaissance de mots postopératoire prédit. Ce modèle parvient à atteindre des performances élevées : erreur absolue moyenne = 9,6 ; coefficient de corrélation = 0,91.

La presbyacousie est un problème de santé qui concerne pratiquement tous les adultes de plus de 50 ans [574]. Les pertes sévères de l'audition sont particulièrement handicapantes et, en raison notamment du vieillissement de la population, leur prévalence est en hausse [575]. Des appareils auditifs ne nécessitant aucune opération peuvent généralement suffire à corriger une perte auditive modérée à modérément sévère. Les pertes auditives sévères quant à elles peuvent dans la plupart des cas être corrigées grâce à un implant cochléaire [576]. Cependant, l'efficacité des implants cochléaires est difficile à prédire car celle-ci varie selon plusieurs facteurs [577]. De plus, l'installation d'un implant cochléaire implique une opération chirurgicale présentant des risques [578]. La possibilité de prédire l'efficacité d'un implant cochléaire pour un/une patient(e) donné(e) permettrait de mieux évaluer le rapport bénéfice/risque. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire l'efficacité d'un implant cochléaire chez un/une patient(e) donné(e), de façon automatique, précise et non invasive [573].

33. Pédiatrie

a. Cas d'usage en diagnostic

Xiyang Liu et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter les cataractes chez l'enfant [579,580].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*figure 105*). Les données d'entraînement sont constituées de photographies oculaires prises chez des enfants à partir d'une lampe à fente. Parmi ces photographies, 410 présentaient une cataracte tandis que 476 présentaient un œil sain. Chaque photographie du jeu de données a été labellisée selon que celle-ci indique la présence ou l'absence d'une cataracte.

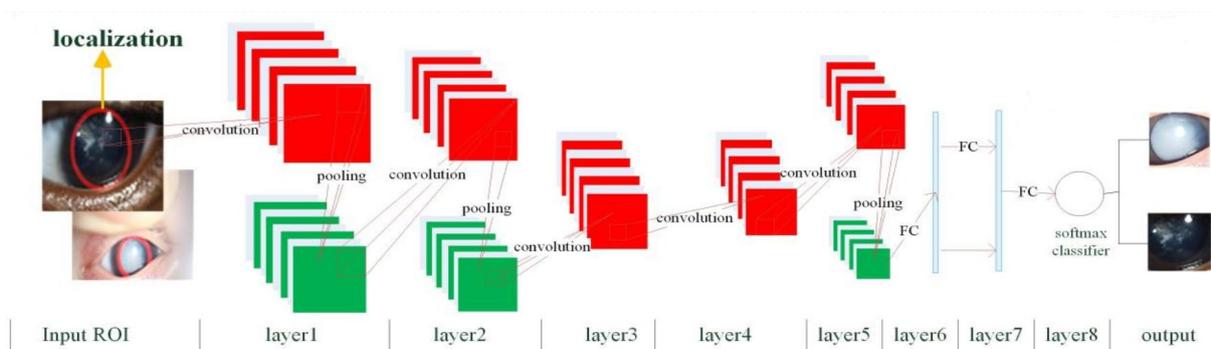


Figure 105. Représentation schématique du modèle présenté (*Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Xiyang Liu et coll.*) [579]

Il s'agit donc d'un modèle de classification prenant en entrée une photographie prise à partir d'une lampe à fente et indiquant en sortie la classe prédite : présence d'une cataracte *vs.* absence de cataracte (*figure 105*). Ce modèle parvient à atteindre des performances élevées : exactitude = 97,07 % ; sensibilité = 97,28 % ; spécificité = 96,83 %.

La cataracte est une pathologie ophtalmique courante pouvant être à l'origine d'une perte irréversible de la vision [581]. Chez les enfants, il s'agit d'une des causes principales de cécité [582]. Cependant, constater une cataracte chez l'enfant en bas âge peut être compliqué en raison de sa progression asymptomatique et, lorsque celle-ci devient sévère, la prévention des troubles de la vision s'avère complexe voire impossible [583,584]. En outre, le diagnostic de la cataracte est généralement effectué visuellement à partir d'une lampe à fente, un examen chronophage et comportant une dimension subjective [585,586]. La possibilité de détecter automatiquement et avec précision les cataractes pourrait permettre une économie de temps pour les ophtalmologues, une facilitation des diagnostics précoces et de fait une meilleure prévention de leurs progressions. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait

potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les cataractes chez l'enfant, de façon automatique, précise et non invasive [579].

b. Cas d'usage en thérapeutique

Olivier Niel et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à déterminer le poids sec chez des enfants hémodialysés chroniques [587].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones (perceptron multi-couches). Les données d'entraînement sont constituées des variables suivantes : statut d'hydratation mesuré par bio-impédancemétrie, volume sanguin relatif, pression artérielle. Ces données ont été générées *in silico* au cours d'hémodialyses virtuelles administrées sur des patients virtuels atteints d'insuffisance rénale chronique terminale. Après entraînement et validation, le modèle a ensuite été évalué sur 14 patients réels hémodialysés chroniques âgés de 7 à 17 ans et dont les poids variaient entre 20 et 60,2 kg.

Il s'agit d'un modèle de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie une valeur numérique correspondant à la prédiction du poids sec idéal à atteindre. Ce modèle parvient à atteindre des performances élevées : coefficient de corrélation = 0,9997. L'écart moyen mesuré entre les poids secs prédits par le modèle et ceux estimés par les néphrologues participant à l'étude est de 0,497 kg. Chez les patients participant à l'étude, le modèle a permis une diminution significative de la pression artérielle post-dialytique ($p = 0,022$), une diminution ou un arrêt des traitements antihypertenseurs dans 28,7% des cas et une diminution des symptômes liés à une sous-estimation du poids sec (hypotension artérielle intra-dialytique chez un patient, asthénie inter-dialytique chez deux patients).

Entre 2003 et 2016, la prévalence mondiale d'insuffisances rénales au stade terminal a connu une croissance rapide [588]. L'hémodialyse représente la thérapie de remplacement rénale la plus commune dans la plupart des pays [589]. Hélas, les patients hémodialysés chroniques sont exposés à des risques de morbi-mortalité importants, notamment des complications cardiovasculaires [590]. La tolérance cardiovasculaire lors d'une hémodialyse dépend essentiellement de la gestion des paramètres de celle-ci, en particulier le poids sec chez l'enfant [591,592]. Cependant, l'absence de recommandations précises quant à la gestion du poids sec entraîne les néphrologues à estimer celui-ci empiriquement et selon les technologies à disposition. Cette approche imprécise s'avère globalement source de morbidité pour les patients hémodialysés [593]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à déterminer le poids sec idéal à atteindre chez un enfant hémodialysé, de façon automatique, précise et non invasive [587].

34. Pharmacologie (cas d'usage en thérapeutique)

Izhar Wallach et son équipe présentent en 2015 un modèle d'intelligence artificielle visant à prédire les affinités de liaison d'une molécule à visée thérapeutique [594,595].

Ce modèle, nommé *AtomNet*, est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif. Les données d'entraînement ont été constituées à partir de bases de données en libre accès : DUD-E (*Directory of Useful Decoys, Enhanced*) et *ChEMBL-20*. DUD-E comporte 22 886 principes actifs, 102 cibles et 50 molécules leures pour chaque principe actif. Le jeu de données construit à partir de *ChEMBL-20* comporte 78 904 principes actifs, 290 cibles et 2 367 120 molécules leures. Une version alternative de ce jeu de données a été construite au sein de laquelle les 2 367 120 molécules leures ont été remplacées par 363 187 molécules dont l'inactivité est vérifiée expérimentalement. Avant d'être fournies à *AtomNet*, les molécules à l'étude et les cibles sont annotées à l'aide de maillages cubiques en trois dimensions. Ces derniers sont placés au niveau des sites de liaison des co-complexes. Les contenus des maillages sont ensuite convertis en vecteurs.

AtomNet est un modèle de classification prenant en entrée un vecteur tel que décrit ci-dessus et indiquant en sortie les probabilités prédites associées à chacune des classes suivantes : activité vs. inactivité. Celui-ci parvient à atteindre des performances supérieures aux outils concurrents (ex. : *Smina*) : aire sous la courbe de ROC = 89,5 % pour DUD-E / 78,1 % pour *ChEMBL-20* / 74,5% pour la version alternative de *ChEMBL-20*. En intégrant dans son analyse la géométrie structurelle des molécules à l'étude et des cibles, *AtomNet* est capable de mettre en évidence des affinités de liaison inconnues jusqu'alors.

L'étude et l'identification des affinités de liaison constituent l'un des fondements de l'innovation thérapeutique [596]. Malgré les efforts déployés dans le développement d'outils informatiques appliqués à la recherche thérapeutique, la précision des outils actuels en matière d'identification des liaisons ligand-récepteur reste insatisfaisante. De fait, les méthodes les plus sollicitées s'appuient majoritairement sur une expérimentation chimique pouvant être coûteuse et chronophage. La possibilité de pouvoir mener ces recherches *in silico* mais avec davantage de précision pourrait réduire les coûts et le temps nécessaire à la découverte de nouveaux traitements et faciliter également l'identification des molécules toxiques avant les études pré-cliniques [597]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des chercheurs à identifier de nouvelles affinités de liaison entre des molécules à l'étude et des cibles, de façon automatique, précise, rapide et moins coûteuse [594].

35. Radiologie (cas d'usage en diagnostic)

William Gale et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter les fractures de la hanche [598,599].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif de type *DenseNet*. Les données d'entraînement comportent 45 492 radiographies frontales du bassin dont environ 12 % présentant une fracture de la hanche. L'ensemble de ces radiographies ont été labellisées selon que celles-ci présentent une fracture de la hanche ou non.

Il s'agit donc d'un modèle de classification prenant en entrée une radiographie frontale du bassin et indiquant en sortie la classe prédite : présence d'une fracture de la hanche *vs.* absence d'une fracture de la hanche. Ce modèle parvient à atteindre des performances comparables à celles des radiologues participant à l'étude : exactitude = 97 % ; sensibilité = 95 % ; valeur prédictive positive = 99 % ; F-score = 97 % ; aire sous la courbe de ROC = 99,4%.

Chez les personnes âgées, les fractures de la hanche font partie des causes majeures d'hospitalisation [600]. Celles-ci sont mortelles dans 20% des cas et entraînent une invalidité permanente dans 50% des cas [601]. La fracture de la hanche est généralement diagnostiquée par radiographie, cependant des examens d'imagerie complémentaires peuvent être nécessaires lorsque la radiographie ne révèle pas de signe de fracture [602]. L'inaccessibilité à ces techniques d'imagerie (ex. : pays en voie de développement) peut être à l'origine d'une erreur ou d'un retard de diagnostic, entraînant une hausse de la morbidité et de la mortalité chez ces patients [603]. En outre, l'interprétation des examens d'imagerie nécessite des années d'expérience et comporte par nature une part de subjectivité [604]. La possibilité de détecter automatiquement, précisément et objectivement les fractures de la hanche à partir d'une radiographie pourrait favoriser leur diagnostic en rendant celui-ci plus rapide, accessible et reproductible. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, la radiographie pourrait potentiellement être suffisante pour effectuer un diagnostic de fracture de la hanche, de façon automatique, précise, rapide, moins coûteuse et non invasive [598].

36. Pneumologie

a. Cas d'usage en diagnostic

Rajkumar Palaniappan et son équipe présentent en 2014 des modèles d'intelligence artificielle visant à détecter des pathologies respiratoires [605].

Ces modèles sont obtenus à l'issue d'un apprentissage automatique supervisé réalisé par deux algorithmes différents : machine à vecteurs de support, méthode des k plus proches voisins. Les

données d'entraînement sont constituées de caractéristiques extraites de 68 signaux acoustiques pulmonaires provenant d'une base de données en accès payant nommée R.A.L.E. (*Respiration Acoustic Laboratory Environment*). Ceux-ci ont été enregistrés au niveau de la paroi thoracique à l'aide d'un capteur de vibration (*Siemens EMT25C*). Ces enregistrements ont été labellisés selon que ceux-ci indiquent une obstruction pathologique des voies respiratoires (26), une pathologie parenchymateuse (25) ou l'absence de pathologie respiratoire (17) (*figure 106*).

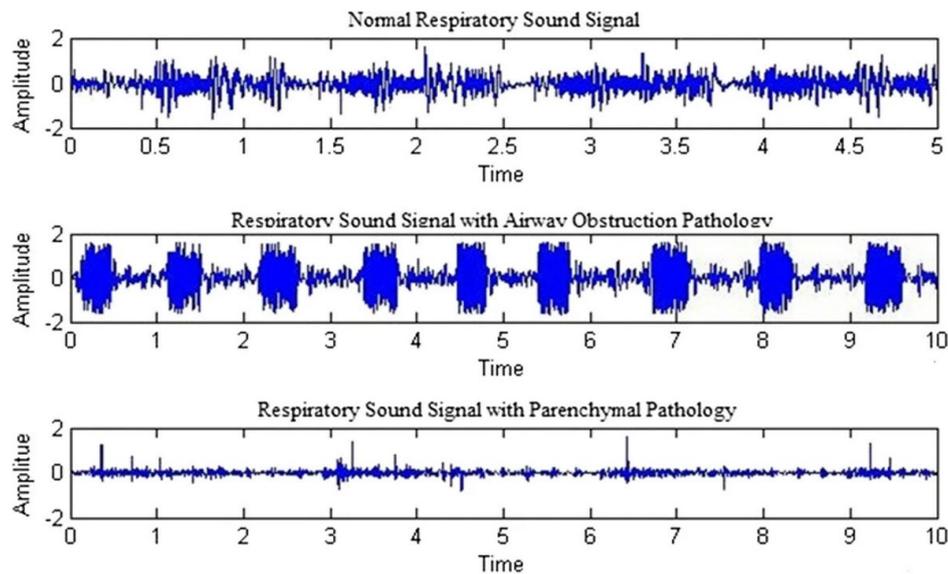


Figure 106. Exemples de signaux acoustiques pulmonaires associés à une obstruction pathologique des voies respiratoires, une pathologie parenchymateuse ou l'absence de pathologie respiratoire (*Reproduit selon les termes de la licence CC BY 2.0. Auteurs : Laszlo Balkanyi et Ronald Cornet.*) [605]

Il s'agit donc d'un modèle de classification prenant en entrée des caractéristiques extraites d'un signal acoustique pulmonaire et indiquant en sortie la classe prédite : présence d'une obstruction pathologique des voies respiratoires *vs.* présence d'une pathologie parenchymateuse *vs.* absence de pathologie respiratoire. Le modèle des *k* plus proches voisins est celui présentant les meilleures performances : exactitude = 98,26 %.

Les signaux acoustiques pulmonaires véhiculent des informations pouvant être exploitées dans la détection de pathologies respiratoires. L'analyse automatisée de ces enregistrements pourrait notamment assister les pneumologues durant leurs diagnostics différentiels et constituer un outil d'apprentissage pour les pneumologues inexpérimentés ou les étudiants. En rendant cet outil accessible publiquement, celui-ci pourrait également permettre aux patients d'effectuer une autosurveillance de leur santé respiratoire. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait en mesure d'identifier la présence de pathologies respiratoires à partir d'un signal acoustique pulmonaire, de façon automatique, précise et non invasive [605].

b. Cas d'usage en thérapeutique

Guy Amit et son équipe présentent en 2015 un modèle d'intelligence artificielle visant à optimiser l'orientation des faisceaux lors d'une radiothérapie par modulation d'intensité en traitement d'une tumeur thoracique [606,607].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de forêt aléatoire. Les données d'entraînement sont constituées de caractéristiques anatomiques relatives aux volumes cibles. Celles-ci sont extraites de 122 planifications de radiothérapies par modulation d'intensité visant à traiter des tumeurs thoraciques. L'ensemble des planifications utilisées pour entraîner le modèle ont été validées cliniquement et associées à des scores de faisceau. Le score de faisceau correspond à la probabilité qu'un faisceau doive être administré avec un angle de portique donné (*figure 107*).

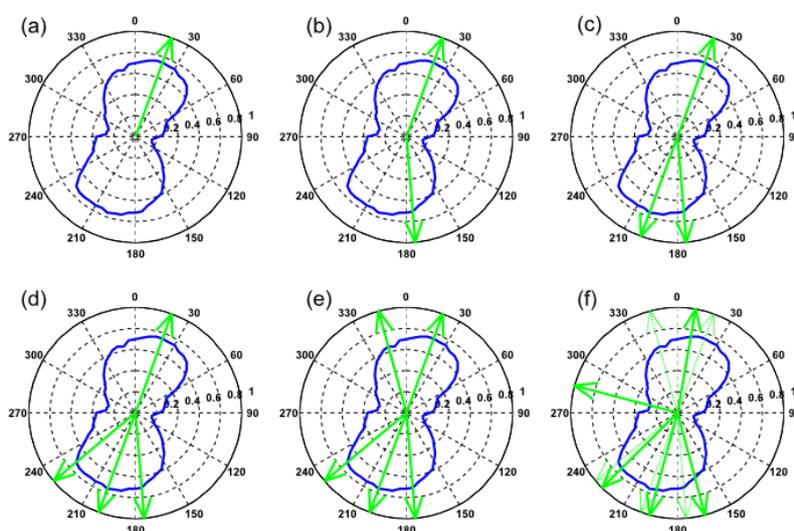


Figure 107. Sélection et ajustement des faisceaux. Les angles 0° , 90° , 180° et 270° correspondant respectivement à l'avant, la gauche, l'arrière, et la droite du patient. (a)-(e) Cinq faisceaux sont sélectionnés un à un, selon un schéma propre aux algorithmes gloutons^{xxix}. (f) Les angles de faisceau sont ensuite ajustés selon un score de faisceau composite. Le score de faisceau pour un point spécifique dépend de la distance entre ce point et le point central. (*Reproduit avec la permission de l'éditeur John Wiley and Sons. Auteurs : Vladimir Pekar et coll.*) [606]

Il s'agit d'un modèle de régression prenant en entrée les caractéristiques anatomiques relatives au volume cible spécifique d'un(e) patient(e) et indiquant en sortie les scores de faisceau prédits. L'orientation optimale des faisceaux est ensuite déduite à partir des scores de faisceau (*figure 107*). Ce modèle parvient à atteindre des performances relativement élevées : coefficient de corrélation = 0,75 ; distance angulaire moyenne (entre les orientations prédites et les orientations de référence) = $16,8^\circ$; pourcentage d'orientations prédites validées (par trois radiothérapeutes) = 93 %.

^{xxix} Un algorithme glouton est un algorithme qui suit le principe de réaliser, étape par étape, un choix optimum local, afin d'obtenir un résultat optimum global (ex. : *algorithme de Lloyd*) [608].

La planification d'une radiothérapie par modulation d'intensité est un processus chronophage durant lequel les paramètres les plus adaptés au/à la patient(e) sont établis [609]. Le choix de l'orientation des faisceaux est notamment primordial en vue de couvrir efficacement le volume cible tout en préservant au mieux les tissus sains [610]. Lorsqu'il n'existe pas de planification standard, comme c'est le cas pour le traitement des tumeurs thoraciques, le radiothérapeute établit celle-ci lui-même à partir de son expérience. Pouvoir estimer automatiquement l'orientation optimale des faisceaux permettrait de réduire la durée globale de planification. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à optimiser l'orientation des faisceaux lors d'une radiothérapie par modulation d'intensité en traitement d'une tumeur thoracique, de façon automatique, rapide et non invasive [606].

37. Psychiatrie

a. Cas d'usage en diagnostic

Bartosz Bohaterewicz et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à détecter les risques suicidaires chez des patients atteints de schizophrénie [611].

Ce modèle est obtenu à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : *gradient boosting*, LASSO (*Least Absolute Shrinkage and Selection Operator*), régression logistique, forêt aléatoire, machine à vecteurs de support. Les données d'entraînement sont constituées de caractéristiques extraites (par segmentation) d'images par résonance magnétique fonctionnelle de repos. Ces caractéristiques correspondent aux indices statiques et dynamiques de l'amplitude des fluctuations de basse fréquence (ALFF), de la fraction d'amplitude des fluctuations de basse fréquence (fALFF), de l'homogénéité régionale (ReHo) et de la connectivité fonctionnelle (FC) (*figure 108*). Après avoir été collectées auprès de 19 patients atteints de schizophrénie, ces caractéristiques ont été labellisées selon que le/la patient(e) présente un risque suicidaire ou non. Le risque suicidaire a été estimé à l'aide du questionnaire sur les comportements suicidaires révisé (SBQ-R, *Suicide Behaviors Questionnaire-Revised*).

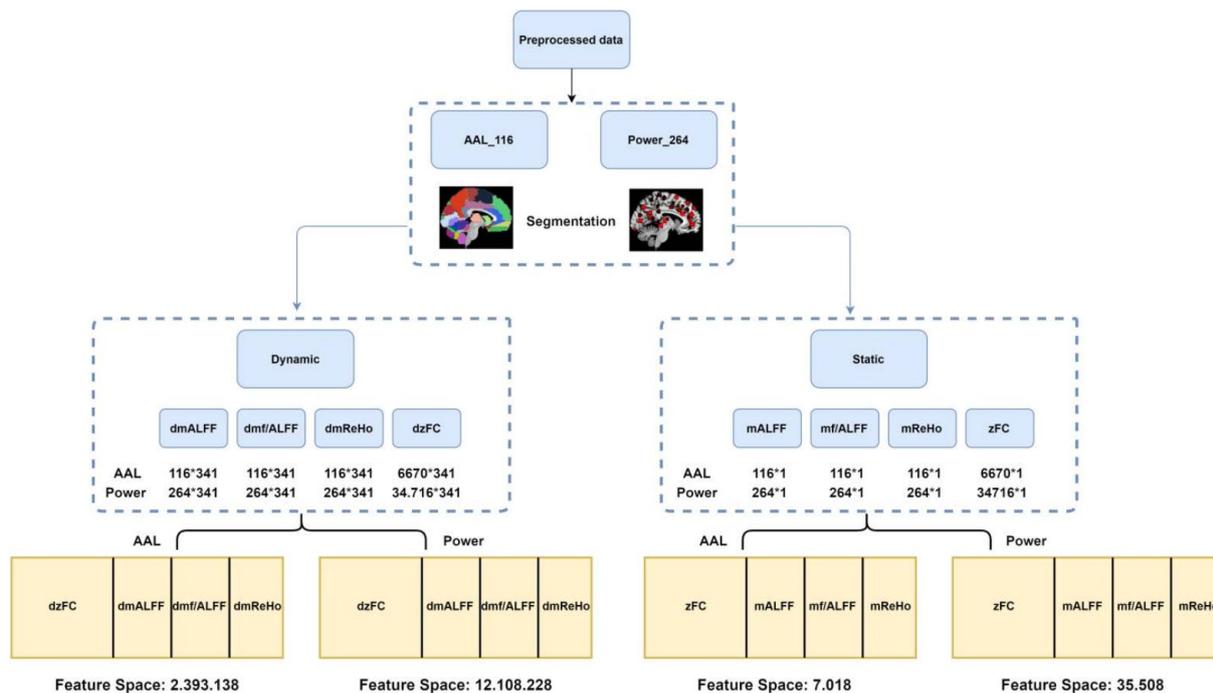


Figure 108. Schéma des extractions de caractéristiques réalisées par segmentation sur des images par résonance magnétique fonctionnelle de repos (Reproduit selon les termes de la licence CC BY 2.0. Auteurs : Bartosz Bohaterewicz et coll.) [611]

Il s'agit donc de modèles de classification prenant en entrée des caractéristiques extraites d'une image par résonance magnétique fonctionnelle de repos et indiquant en sortie la classe prédite : présence d'un risque suicidaire vs. absence de risque suicidaire. Le modèle de LASSO prenant en entrée la connectivité fonctionnelle est celui présentant les meilleures performances : exactitude = 70 % ; aire sous la courbe de ROC = 76%.

L'évaluation du risque suicidaire est l'une des préoccupations majeures en psychiatrie car les suicides constituent la seconde cause de mortalité chez les jeunes adultes [612]. Plus spécifiquement chez les patients atteints de schizophrénie, le suicide serait la cause de 40% des décès [613]. En outre, 25 à 50% des personnes atteintes de schizophrénie commettraient au moins une tentative de suicide au cours de leur vie [614,615]. Ce risque semble difficile à identifier par le dialogue, les prédictions effectuées par les soignants et par les patients eux-mêmes quant au risque de passage à l'acte se sont avérées très imprécises [616]. Parmi les patients dont le décès est dû à un suicide, environ 80% n'avaient pas des pensées suicidaires lors de leur dernier échange avec un professionnel de la santé mentale [617]. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, l'imagerie par résonance magnétique fonctionnelle de repos pourrait potentiellement renforcer la capacité des soignants à détecter le risque suicidaire chez les patients atteints de schizophrénie, de façon automatique et non invasive [611].

b. Cas d'usage en thérapeutique

Buru Chang et son équipe présentent en 2019 un modèle d'intelligence artificielle visant à prédire l'efficacité d'un traitement antidépresseur [618,619].

Ce modèle, nommé ARPNet (*Antidepressant Response Prediction Network*), est obtenu à l'issue d'un apprentissage automatique supervisé réalisé sur un réseau de neurones (*figure 109*). Les données d'entraînement sont constituées d'informations relatives aux consultations et prescriptions, de variables démographiques, génétiques et de caractéristiques extraites d'images par résonance magnétique cérébrale (*figure 109*). Ces données ont été collectées auprès de 121 patients atteints de trouble dépressif majeur au cours de leurs consultations (consultation initiale, après une semaine, après quatre semaines, après huit semaines, après six mois). Chaque consultation a été labellisée selon le degré de sévérité des symptômes observés. Le degré de sévérité a été mesuré à l'aide de l'échelle de dépression de Hamilton (HAM-D, *Hamilton Depression Rating Scale*).

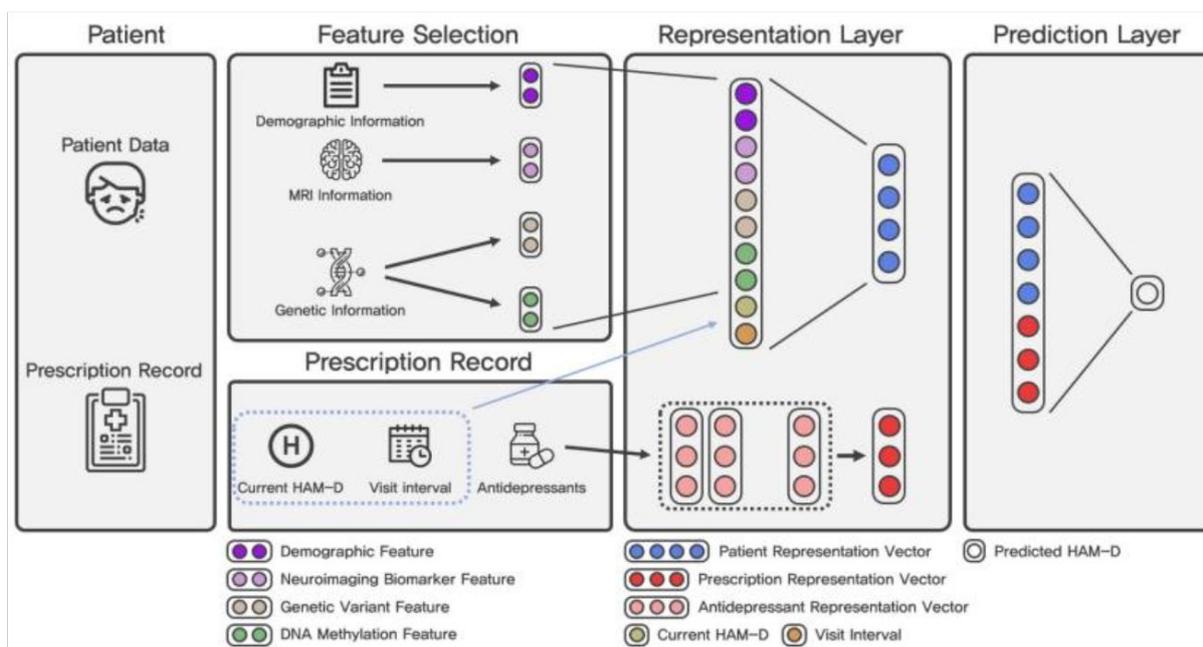


Figure 109. Représentation schématique du modèle présenté (ARPNet) et de son entraînement (*Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Buru Chang et coll.*) [618]

Il s'agit donc d'un modèle de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie une prédiction de l'échelle de dépression de Hamilton après x semaines de traitement (*figure 109*). ARPNet parvient à atteindre des performances relativement élevées : écart quadratique moyen = 3,3022 ; coefficient de détermination = 0,5523 ; exactitude = 84,62 % ; sensibilité = 80 % ; spécificité = 87,50 % ; valeur prédictive positive = 80 % ; F-score = 80 %.

Le traitement des troubles dépressifs majeurs est souvent complexe car il n'est possible d'apprécier l'efficacité d'un antidépresseur donné sur un(e) patient(e) donné(e) qu'après plusieurs semaines de médication. Selon plusieurs études d'efficacité, seuls 11 à 30% des patients parviendraient à guérir de leur trouble dépressif majeur au cours de leur traitement initial [620]. Cette difficulté à identifier le traitement optimal selon les patients prolonge la durée de traitement pour la majorité d'entre eux, entraînant par conséquent une exposition prolongée aux risques de morbidité et de mortalité propres aux troubles dépressifs ainsi qu'aux effets indésirables des antidépresseurs administrés [621,622]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire l'efficacité d'un antidépresseur donné chez un(e) patient(e) donné(e), de façon automatique, précise et non invasive, et ainsi à prescrire l'antidépresseur le plus efficace pour ce/cette patient(e) [618].

38. Rhumatologie

a. Cas d'usage en diagnostic

Beth G. Ashinsky et son équipe présentent en 2017 un modèle d'intelligence artificielle visant à détecter précocement les gonarthroses [623].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme nommé WND-CHARM (*Weighted Neighbor Distance using Compound Hierarchy of Algorithms Representing Morphology*). Les données d'entraînement sont constituées d'images par résonance magnétique réalisées au niveau du condyle fémoral médial. Celles-ci ont été collectées auprès de 68 patients présentant ou non une gonarthrose. Ces examens d'imagerie ont été labellisés selon l'état du genou des patients 36 mois plus tard. Quarante d'entre eux ont présenté une progression de leur gonarthrose (score de WOMAC > 10) tandis que 28 d'entre eux n'ont pas présenté de gonarthrose ou de progression de leur gonarthrose (score de WOMAC ≤ 10).

Il s'agit donc d'un modèle de classification prenant en entrée une image par résonance magnétique et indiquant en sortie la classe prédite : présence d'une gonarthrose tendant à s'aggraver vs. absence de gonarthrose tendant à s'aggraver. WND-CHARM parvient à atteindre les performances suivantes : exactitude = 75 % ; sensibilité = 74 % ; spécificité = 76 %.

L'arthrose est une pathologie répandue et dont l'incidence augmente avec l'âge et la surcharge pondérale [624]. Celle-ci se caractérise notamment par une dégénérescence du cartilage articulaire [625]. Selon des études récentes, des mesures hygiéno-diététiques telles qu'une activité physique régulière et un régime équilibré pourraient diminuer le risque de progression

des arthrites [626,627]. Un diagnostic précoce de l'arthrose pourrait donc dans certains cas permettre de freiner la dégénérescence articulaire et ainsi préserver la qualité de vie des patients atteints. Cependant, les signes d'une arthrose précoce sont généralement difficiles à observer [628]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les gonarthroses, de façon automatique et non invasive [623].

b. Cas d'usage en thérapeutique

André Wirries et son équipe présentent en 2021 des modèles d'intelligence artificielle visant à prédire l'efficacité d'un traitement de la hernie discale lombaire [629].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés : régression linéaire, *elastic net*, méthode des k plus proches voisins, arbres décisionnels, gradient stochastique, machine à vecteurs de support, réseau de neurones. Les données d'entraînement comportent des variables socio-démographiques, cliniques et des informations relatives au traitement administré (chirurgie ou traitement conservateur). Ces données ont été collectées auprès de 54 patients présentant une hernie discale lombaire et ont été labellisées selon l'échelle d'incapacité d'Oswestry (ODI, *Oswestry Disability Index*) mesurée six mois après l'opération ou le début du traitement conservateur.

Il s'agit donc de modèles de régression prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie une prédiction de l'échelle d'incapacité d'Oswestry six mois après l'opération ou le début du traitement conservateur. Le réseau de neurones est le modèle parvenant à atteindre les meilleures performances : erreur absolue moyenne = 1,5 % ; différence maximale entre l'échelle d'incapacité d'Oswestry prédite et celle observée après six mois = 3,4%.

Le traitement de la hernie discale lombaire se décline en plusieurs options pouvant être regroupées selon deux catégories, les traitements chirurgicaux et les traitements conservateurs. L'efficacité d'un traitement conservateur et la nécessité d'une chirurgie sont difficiles à évaluer chez un(e) patient(e) donné(e), rendant complexe cette prise de décision par les soignants [630]. Un traitement conservateur est généralement privilégié en première intention chez les patients ne souffrant pas de déficit neurologique, la chirurgie est ensuite envisagée si les symptômes se maintiennent ou s'aggravent après six mois de traitement conservateur [631]. Chez les patients dont la guérison nécessite une chirurgie, ce délai prolonge la douleur, l'invalidité professionnelle et l'exposition au risque de complications [632]. La possibilité de prédire précocement l'efficacité d'un traitement pour un(e) patient(e) donné(e) permettrait d'optimiser le choix du traitement en première intention et ainsi réduire l'errance thérapeutique. Les

résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prescrire le traitement le plus adapté à chaque patient(e) atteint(e) d'une hernie discale lombaire, de façon automatique, précise et non invasive [629].

39. Soins intensifs

a. Cas d'usage en diagnostic

Nam K. Tran et son équipe présentent en 2019 des modèles d'intelligence artificielle visant à détecter la présence de lésions rénales aiguës chez des patients gravement brûlés [633,634].

Ces modèles sont obtenus à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de k plus proches voisins. Les données d'entraînement comportent des variables cliniques (débit urinaire) et biologiques (NGAL, créatinine, NT-proBNP). Celles-ci ont été collectées auprès de 50 patients majeurs présentant des brûlures sur au moins 20% de leur surface corporelle totale, au cours des 24 heures suivant leur admission en soins intensifs. Les données ont été labellisées selon que le/la patient(e) ait présenté (25) ou non (25) une lésion rénale aiguë durant la semaine suivant son admission.

Il s'agit donc de modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : présence d'une lésion rénale aiguë *vs.* absence de lésion rénale aiguë (*figure 110*). Parmi les modèles construits, deux se distinguent par leurs performances. Le premier, prenant en entrée l'ensemble des variables d'entraînement, parvient à atteindre une exactitude de 98 %. Le second, prenant en entrée le débit urinaire (UOP), la créatinine et le NT-proBNP, parvient à atteindre une exactitude de 88%. En outre, le temps post-admission moyen d'identification d'une lésion rénale aiguë à l'aide d'un des modèles mentionnés était de 18,8 heures contre 42,7 heures en s'appuyant uniquement sur les mesures de débit urinaire et créatinine.

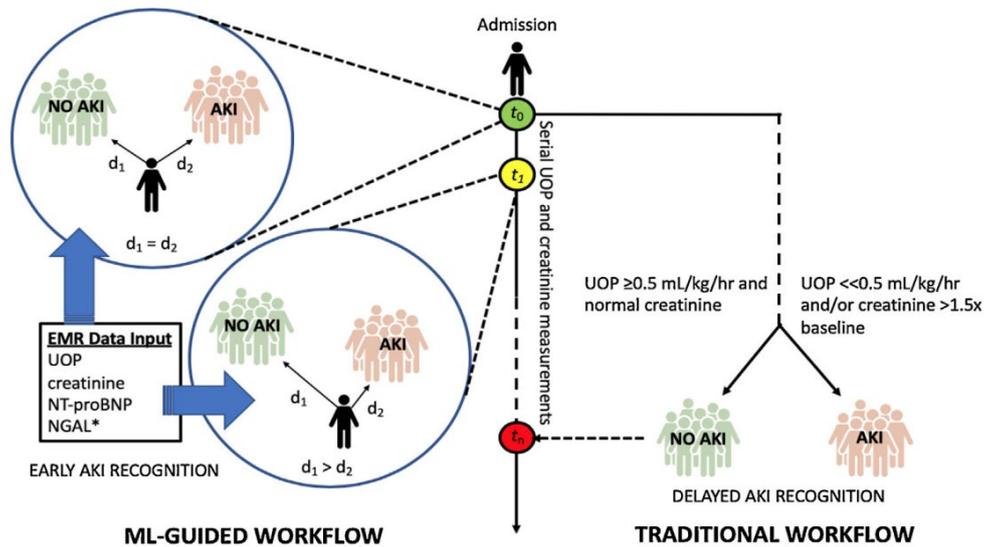


Figure 110. Comparaison d'un flux de travail traditionnel et d'un flux de travail intégrant le modèle présenté, dans le cadre d'une admission d'un(e) patient(e) gravement brûlé(e). AKI : acute kidney injury (lésion rénale aigüe), UOP : urine output (débit urinaire), NGAL : neutrophil gelatinase-associated lipocalin, NT-proBNP : N-Terminal pro-Brain Natriuretic Peptide). (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Nam K. Tran et coll.) [633]

Les lésions rénales aigües concernent plus de la moitié des patients gravement brûlés reçus en soins intensifs [635]. Malgré cette prévalence élevée, le diagnostic précoce des lésions rénales aigües est difficile car celui-ci repose généralement sur une mesure du débit urinaire et de la créatinine, des biomarqueurs ne garantissant pas à eux seuls une performance diagnostique satisfaisante (figure 110) [636,637,638,639]. Des études récentes ont permis d'identifier le NT-proBNP et le NGAL comme étant des biomarqueurs prometteurs pouvant contribuer à améliorer la précision du diagnostic des lésions rénales aigües [640]. Les résultats obtenus révèlent qu'à l'aide d'un modèle d'intelligence artificielle, ces biomarqueurs seraient potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer précocement les lésions rénales aigües chez les patients gravement brûlés, de façon automatique et précise [633].

b. Cas d'usage en thérapeutique

N.L. Loo et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à surveiller la survenue d'asynchronies patient-ventilateur durant une ventilation mécanique [634,641].

Ce modèle est obtenu à l'issue d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (figure 111). Les données d'entraînement sont constituées de 2000 aires sous la courbe de la pression inspiratoire au cours d'un cycle respiratoire (figure 112). Celles-ci ont été collectées auprès de patients sous ventilation mécanique et labellisées selon qu'elles indiquent la survenue d'une asynchronie patient-ventilateur (1000) ou non (1000).

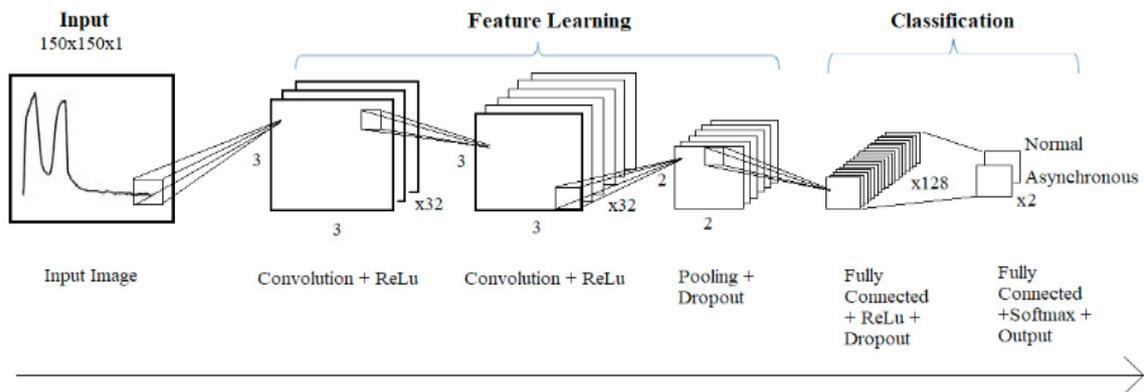


Figure 111. Représentation schématique du modèle présenté (Reproduit avec la permission de l'International Federation of Automatic Control. Auteurs : Nien Loong Loo et coll.) [641]

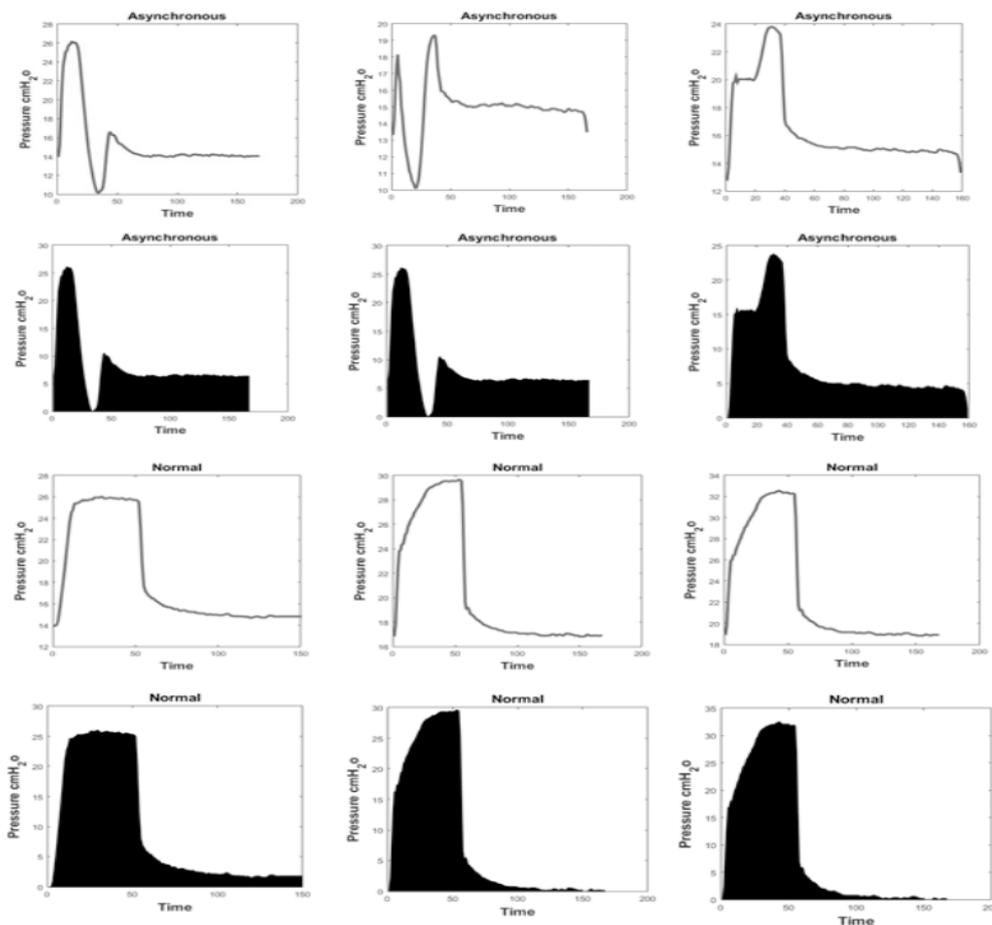


Figure 112. Exemples de graphiques représentant la pression inspiratoire au cours d'un cycle respiratoire associées à leurs aires sous la courbe respectives. (Reproduit avec la permission de l'International Federation of Automatic Control. Auteurs : Nien Loong Loo et coll.) [641]

Il s'agit donc d'un modèle de classification prenant en entrée une aire sous la courbe représentant la pression inspiratoire au cours d'un cycle respiratoire et indiquant en sortie la classe prédite : présence d'une asynchronie patient-ventilateur vs. absence d'asynchronie patient-ventilateur (figure 111). Ce modèle parvient à atteindre des performances élevées : exactitude = 96,7 % ; sensibilité = 98,5 % ; spécificité = 89,4 %. Le temps d'exécution moyen du modèle est de l'ordre de la milliseconde.

La survenue répétée d'asynchronies patient-ventilateur peut provoquer une dyspnée chez le/la patient(e), prolonger la durée de ventilation mécanique, accroître la quantité de sédatifs administrés et globalement augmenter le taux de mortalité [642,643,644,645]. L'identification des asynchronies patient-ventilateur repose sur la détection de motifs anormaux au sein du cycle respiratoire du/de la patient(e) [646]. La tâche peut être difficile car ces motifs varient d'un(e) patient(e) à l'autre [647]. De plus, il s'agit d'un processus chronophage et fastidieux nécessitant une surveillance en continue des cycles respiratoires du/de la patient(e) intubé(e) [648]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à surveiller en continue les asynchronies patient-ventilateur, de façon automatique et précise [641].

40. Télémédecine

a. Cas d'usage en diagnostic

Miles F. Greenwald et son équipe présentent en 2020 un usage en télémédecine d'un modèle d'intelligence artificielle visant à détecter les rétinopathies du prématuré [649].

L'étude incluait 613 examens ophtalmologiques effectués chez 81 patients ayant pour âge post-menstruel^{xxx} moyen 29,2 semaines. Les examens ophtalmologiques consistaient en des fonds d'œil durant lesquels des infirmières praticiennes spécialisées en néonatalogie ont acquis des images rétiniennes à l'aide d'une caméra grand angle (*RetCam*). Ces images étaient ensuite télétransmises à des ophtalmologues en charge du diagnostic de rétinopathie.

En parallèle, les images étaient analysées par le modèle à l'étude (nommé *i-ROP DL*) chargé d'attribuer à chaque image un score de sévérité vasculaire. Ce score de sévérité vasculaire, pouvant aller de 1 à 9, correspond au degré de certitude quant à la présence d'une rétinopathie. Il s'agit donc d'un modèle de classification prenant en entrée une photographie de fond d'œil et indiquant en sortie une prédiction du score de sévérité vasculaire. L'aire sous la courbe de ROC mesurée était de 99 %. En positionnant le seuil du score de sévérité vasculaire à 3, le modèle aurait atteint une sensibilité de 100 % et une spécificité de 90 %.

La prévalence de rétinopathie du prématuré est en forte augmentation mondialement, s'expliquant en partie par l'amélioration du taux de survie des nouveau-nés prématurés [651]. Le télédiagnostic semble être une solution partielle à cette problématique en permettant aux ophtalmologues d'examiner davantage de fonds d'œil [652]. Cependant, le diagnostic de la rétinopathie et de sa sévérité peut dans certains cas être long en raison notamment de la dimension subjective de cet exercice [653]. La possibilité de détecter automatiquement et

^{xxx} L'âge post-menstruel désigne l'âge gestationnel auquel s'ajoute l'âge chronologique. [650]

précisément les rétinopathies permettrait aux ophtalmologues de fournir rapidement un diagnostic aux services de néonatalogie demandeurs et ainsi d'accroître l'efficacité permise par la télémédecine. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à télédiagnostiquer les rétinopathies du prématuré, de façon automatique, rapide et précise [649].

b. Cas d'usage en thérapeutique

Robert LeMoine et son équipe présentent en 2015 un modèle d'intelligence artificielle visant à surveiller à distance la progression d'une rééducation de la cheville [654].

Ce modèle est obtenu à l'issue d'un apprentissage automatique supervisé réalisé par un algorithme de machine à vecteurs de support. Les données d'entraînement sont constituées de caractéristiques extraites de signaux gyroscopiques mesurés durant une dorsiflexion : extremums des vitesses de rotation mesurées lors de la dorsiflexion et intervalles de temps entre les extremums mesurés (*figure 113-b,c*). Ces signaux gyroscopiques ont été collectés à l'aide d'un système composé d'un dorsiflexeur et d'un gyroscope, auprès d'un(e) patient(e) atteint(e) d'une hémiplégié chronique (*figure 113-a*). A partir de ce système, le/la patient(e) a effectué 30 dorsiflexions avec la cheville atteinte par l'hémiplégié et 30 dorsiflexions avec la cheville non atteinte par l'hémiplégié. Les signaux collectés ont été labellisés selon que ceux-ci aient été mesurés à partir de l'une ou l'autre des chevilles.

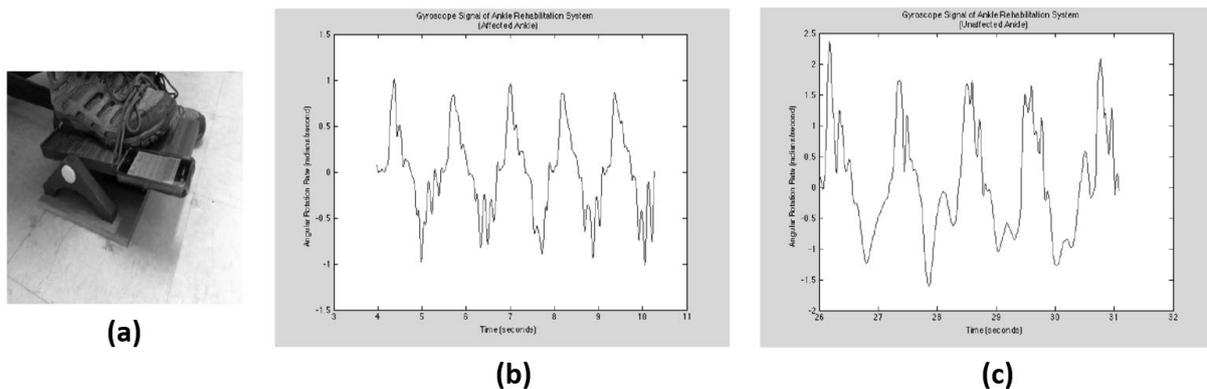


Figure 113. (a) système de rééducation de la cheville avec le pied du sujet monté sur repose-pied et plateforme gyroscopique ; (b) signal émis par le gyroscope durant la dorsiflexion d'une cheville atteinte par une hémiplégié (cinq répétitions) ; (c) signal émis par le gyroscope durant la dorsiflexion d'une cheville saine (cinq répétitions). (*Reproduit avec la permission de l'éditeur Institute of Electrical and Electronics Engineers. Auteurs : Robert LeMoine et coll.*) [654]

Il s'agit donc d'un modèle de classification prenant en entrée les caractéristiques extraites d'un signal gyroscopique et indiquant en sortie la classe prédite : cheville affectée par une hémiplégié vs. cheville non affectée par une hémiplégié. Ce modèle parvient à atteindre des performances élevées : exactitude = 97 %.

Une hémiparésie peut être causée par divers événements (accident vasculaire cérébral, tumeur, traumatisme) et se manifester à travers diverses formes [655]. Lorsque la paralysie engendrée concerne un membre inférieur et affecte notamment la cheville, l'un des symptômes observables est le pied tombant [656]. Le pied tombant désigne une incapacité totale ou partielle d'un(e) patient(e) à effectuer une dorsiflexion, affectant ainsi ses fonctions motrices et accroissant son risque de chutes [657]. Les capacités motrices du/de la patient(e) peuvent être retrouvées grâce à une rééducation adaptée [658]. Son efficacité peut être estimée au fur et à mesure en évaluant la capacité du/de la patient(e) à effectuer une dorsiflexion [659]. La possibilité pour les soignants d'assurer ce suivi automatiquement et à distance permettrait de réduire les déplacements à effectuer pour le/la patient(e). Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à surveiller à distance la progression d'une rééducation de la cheville, de façon automatique et précise [654].

41. Toxicologie

a. Cas d'usage en diagnostic

Michael Chary et son équipe présentent en 2021 un modèle d'intelligence artificielle visant à identifier un toxidrome [660].

Il s'agit d'un système expert nommé *Tak* construit à partir d'un réseau logique probabiliste. Celui-ci intègre dans son analyse une représentation des connaissances organisée autour de 34 règles logiques probabilistes. Ces règles ont été déterminées à partir du consensus de trois toxicologues ainsi que des prévalences indiquées par le rapport annuel de l'*American Association of Poison Control Centers*.

Tak prend en entrée un cas clinique (symptômes observés) et indique en sortie une prédiction du toxidrome correspondant. *Tak* parvient à atteindre les performances suivantes : kappa de Cohen = 0,8432 pour les cas les plus évidents ; 0,4396 pour les cas modérément complexes ; 0,3331 pour les cas les plus complexes. Ces performances sont comparables à celles des toxicologues ayant participé à l'étude pour les cas les plus simples et modérément complexes. En revanche, les performances des toxicologues étaient supérieures pour les cas les plus complexes.

La prise en charge des intoxications a pour étape initiale la caractérisation du risque de mort imminente afin de traiter en urgence les patients à risque. Ces situations critiques ne permettent pas aux toxicologues de recourir à des tests biologiques en raison du délai requis pour effectuer ceux-ci. A défaut, les toxicologues effectuent leurs diagnostics en se référant aux informations

disponibles, aux symptômes observables ainsi qu'à l'interrogatoire lorsque l'état mental du/de la patient(e) le permet [661]. La possibilité d'identifier automatiquement un toxidrome chez un(e) patient(e) donné(e) pourrait accélérer la démarche de diagnostic, permettre une prise en charge plus rapide et ainsi améliorer le pronostic vital des patients les plus à risque. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à identifier rapidement les toxidromes simples et modérément complexes, de façon automatique et non invasive [660].

b. Cas d'usage en thérapeutique

Wei-Hsuan Lo-Ciganic et son équipe présentent en 2019 des modèles d'intelligence artificielle visant à prédire le risque de survenue à court terme d'une surdose d'opioïdes [662].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : régression logistique multivariée, LASSO, forêt aléatoire, amplification de gradient, apprentissage profond. Les données d'entraînement comportent des variables sociodémographiques, géographiques, des antécédents médicaux et addictologiques ainsi que des informations relatives à l'état de santé des patients, leurs prescriptions d'opioïdes et leurs prescripteurs. Ces données ont été collectées auprès de 186 686 patients ayant reçu au moins une prescription d'opioïde. Les données collectées ont ensuite été fragmentées par tranche de trois mois puis chaque période de trois mois a été labellisée selon que le/la patient(e) correspondant(e) ait subi ou non une surdose au cours de la période de trois mois suivante (*figure 114*).

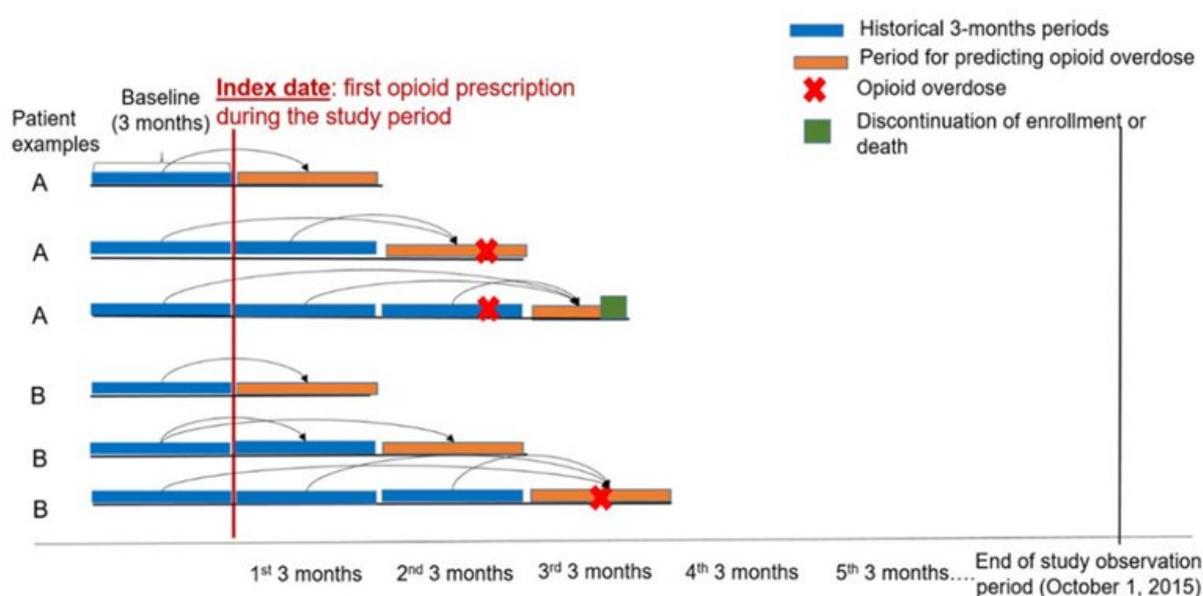


Figure 114. Méthodologie de collecte, segmentation et labellisation des données d'entraînement (*Reproduit selon les termes de la licence CC BY 2.0. Auteurs : Wei-Hsuan Lo-Ciganic et coll.*) [662]

Il s'agit de modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : risque faible de survenue d'une surdose au cours des trois mois suivant la prédiction vs. risque modéré de survenue d'une surdose au cours des trois mois suivant la prédiction vs. risque élevé de survenue d'une surdose au cours des trois mois suivant la prédiction. Le réseau de neurones est le modèle présentant les meilleures performances : sensibilité = 92,3 % ; spécificité = 75,7 % ; aire sous la courbe de ROC = 91 % ; valeur prédictive positive = 0,18 % ; valeur prédictive négative = 99,9 %.

Les analgésiques opioïdes sont couramment prescrits en traitement des douleurs aiguës ou chroniques [663]. Ces traitements présentent comme effet indésirable potentiel la dépendance, notamment suite à une administration prolongée [664]. Lorsque celle-ci évolue en addiction, le/la patient(e) est exposé(e) à des risques importants de mésusages et de surdosages pouvant entraîner la mort [665]. L'addiction n'est néanmoins pas systématique et, même chez les patients touchés par l'addiction, le risque de survenue d'une surdose dépend de plusieurs facteurs relatifs au/à la patient(e) et à la prescription [666]. La possibilité de prédire ce risque permettrait de mieux adapter les modalités de la thérapie selon le/la patient(e). Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire le risque de survenue à court terme d'une surdose d'opioïdes chez un(e) patient(e) donné(e), de façon automatique, précise et non invasive [662].

42. Transplantations

a. Cas d'usage en diagnostic

Sara Moccia et son équipe présentent en 2018 des modèles d'intelligence artificielle visant à détecter une stéatose au sein d'un greffon [667].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés et semi-supervisés réalisés par différents algorithmes : machine à vecteurs de support, forêt aléatoire, apprentissage multi-instances. Les données d'entraînement sont constituées de caractéristiques extraites de photographies de foies transplantés ou non (*figure 115*). Celles-ci ont été collectées à partir de 40 greffons et ont été labellisées selon que la biopsie effectuée sur le greffon indique la présence (20) ou l'absence (20) d'une stéatose.



Figure 115. Exemples de prédictions effectuées sur des foies transplantés et non transplantés. Les cadres verts et rouges désignent respectivement les prédictions justes et erronées. (Reproduit avec la permission de l'éditeur Springer Nature. Auteurs : Sara Moccia et coll.) [667]

Il s'agit donc de modèles de classification prenant en entrée des caractéristiques extraites d'une photographie de foie et indiquant en sortie la classe prédite : présence d'une stéatose hépatique vs. absence de stéatose hépatique. Le modèle issu de l'apprentissage multi-instances est celui présentant les meilleures performances : exactitude = 88 % ; sensibilité = 95 % ; spécificité = 81 % (figure 115).

La transplantation hépatique est le seul traitement pouvant améliorer le pronostic vital des patients souffrant d'une insuffisance hépatique au stade terminal [668]. La pénurie en organes disponibles s'accroît en raison de l'accroissement du besoin, entraînant un élargissement des critères de sélection des donneurs [669]. Cet assouplissement des critères exige cependant de redoubler d'efforts quant à l'évaluation de la viabilité des greffons. La stéatose hépatique notamment, dont la prévalence est estimée à environ un quart au sein de la population générale, est source de morbidité et de mortalité chez les patients receveurs [670]. L'analyse histopathologique est la méthode la plus précise à ce jour pour identifier une stéatose hépatique, celle-ci est néanmoins invasive et chronophage [671]. En raison de la brièveté du délai dont les soignants disposent entre le prélèvement du greffon et la transplantation (six heures en moyenne pour un foie), le diagnostic de stéatose hépatique s'effectue généralement en se basant sur les données cliniques du donneur et l'apparence extérieure du greffon [672,673]. Cette méthode de diagnostic comporte donc une part de subjectivité. Une méthode de diagnostic automatique et objective pourrait accélérer et affiner la détection des stéatoses hépatiques, et ainsi favoriser le pronostic du/de la patient(e) après transplantation [674]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer les stéatoses hépatiques au sein d'un greffon, de façon automatique, précise et non invasive [667].

b. Cas d'usage en thérapeutique

Javier Briceño et son équipe présentent en 2014 des modèles d'intelligence artificielle visant à prédire la réussite à court terme d'une transplantation hépatique [675].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques réalisés par un algorithme évolutionniste (*Neural Net Evolutionary Programming*) sur deux réseaux de neurones (*figure 116*). Les données d'entraînement comportent des variables sociodémographiques, cliniques, biologiques, des antécédents médicaux ainsi que des informations relatives au greffon. Ces données ont été collectées auprès de 1003 paires donneur-receveur et ont été labellisées selon la survie ou le décès du receveur ainsi que selon la survie ou non du greffon au cours des trois mois suivant la transplantation.

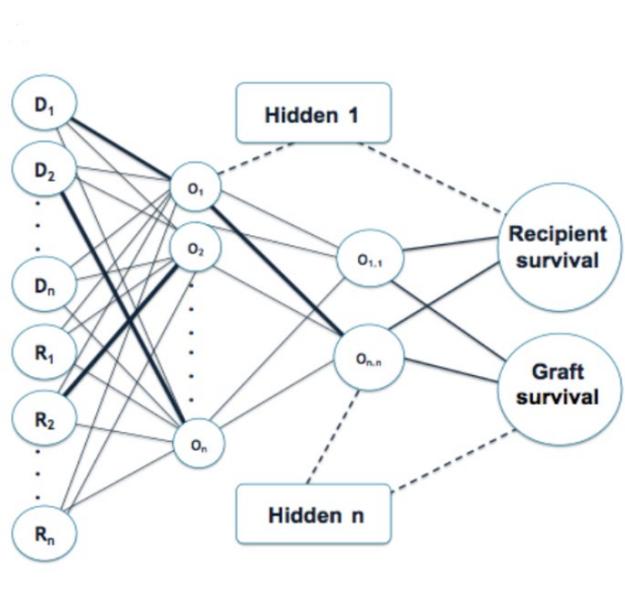


Figure 116. Représentation schématique des modèles présentés (*Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Javier Briceño et coll.*) [675]

Il s'agit donc de deux modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées. Chacun des modèles indique en sortie la classe prédite ainsi que la probabilité associée à celle-ci : survie du/de la patient(e) au cours des trois mois après la transplantation *vs.* décès du/de la patient(e) au cours des trois mois après la transplantation / survie du greffon au cours des trois mois après la transplantation *vs.* défaillance du greffon au cours des trois mois après la transplantation (*figure 116*). Ces modèles parviennent à atteindre les performances suivantes : exactitude = 90,79 % ; aire sous la courbe de ROC = 80,6 % pour la survie du receveur / exactitude = 71,42 % ; aire sous la courbe de ROC = 82,15 % pour la survie du greffon. Dans le cadre de cette étude, les performances des modèles ont été significativement supérieures ($p < 0,001$) à celles des méthodes proposées antérieurement (MELD, DRI, D-MELD, SOFT, P-SOFT and BAR) [676,677,678,679]. L'écart

croissant entre le nombre de patients ayant besoin d'une transplantation hépatique et le nombre de foies disponibles sous-tend une problématique de priorisation [680]. Celle-ci repose notamment sur l'urgence d'une transplantation pour un(e) patient(e) donné(e), calculée à partir des scores MELD et PELD [681]. Plusieurs variables sont également considérées (compatibilité receveur-donneur, probabilité de réussite, contraintes techniques liées au prélèvement, au transport et au maintien de la viabilité du greffon), induisant une complexité du système d'attribution et un risque d'erreur non négligeable en contexte d'urgence [682]. La possibilité de prédire automatiquement et objectivement les probabilités de survie suite à une transplantation hépatique, pour un donneur et un receveur spécifiques, pourrait affiner et accélérer la démarche de sélection des receveurs et ainsi réduire le nombre de greffons perdus en raison d'une attribution inadéquate ou trop tardive. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à prédire les probabilités de survie à court terme suite à une transplantation hépatique, de façon automatique et précise [675].

43. Traumatologie

a. Cas d'usage en diagnostic

Seok Won Chung et son équipe présentent en 2018 un modèle d'intelligence artificielle visant à détecter et classifier les fractures de l'humérus proximal [683,684].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*ResNet-152*). Les données d'entraînement sont constituées de radiographies antéro-postérieures de l'épaule (*figure 117*). Celles-ci ont été collectées auprès de 1891 patients, à raison d'une radiographie par patient(e), et ont été labellisées selon que l'humérus proximal présente ou non une fracture et selon le type de fracture : absence de fracture (515), fracture du trochiter (346), fracture du col chirurgical (514), fracture à trois fragments (269), fracture à quatre fragments (247).



Figure 117. Exemples de données d'entraînement : (A) absence de fracture, (B) fracture du trochiter, (C) fracture du col chirurgical, (D) fracture à trois fragments et (E) fracture à quatre fragments. Chaque radiographie antéropostérieure a été recadrée de façon à ce que les têtes et cols huméraux représentent environ 50 % de l'image. Les images ont ensuite été redimensionnées à 256 × 256 pixels. (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Seok Won Chung et coll.) [683]

Il s'agit donc d'un modèle de classification prenant en entrée une radiographie antéro-postérieure de l'épaule et indiquant en sortie la classe prédite : absence de fracture vs. présence d'une fracture du trochiter vs. présence d'une fracture du col chirurgical vs. présence d'une fracture à trois fragments vs. présence d'une fracture à quatre fragments. Ce modèle parvient à atteindre des performances élevées : exactitude = 96 % ; sensibilité = 99 % ; spécificité = 97 % ; aire sous la courbe de ROC = 100 % ; indice de Youden = 97 % lorsqu'il s'agit de détecter la présence d'une fracture au niveau de l'humérus proximal / exactitude = 65-86 % ; sensibilité = 88-97 % ; spécificité = 83-94 % ; aire sous la courbe de ROC = 90-98 % ; indice de Youden = 71-90 % lorsqu'il s'agit d'identifier le type de fracture. Les performances globales du modèle étaient significativement supérieures ($p < 0,001$) à celles des médecins ayant participé à l'étude. Les fractures de l'humérus proximal ainsi que le type de fracture sont généralement diagnostiqués à partir de radiographies [685]. Cependant, l'interprétation des radiographies nécessite des années d'expérience et comporte par nature une part de subjectivité. De fait, les diagnostics inexacts quant à la présence ou au type de fracture de l'humérus sont relativement courantes, en particulier de la part des chirurgiens orthopédistes les moins expérimentés [686,687]. La fiabilité du diagnostic pourrait être favorisée par des méthodes d'interprétation automatiques et objectives. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à diagnostiquer les fractures de l'humérus proximal, de façon automatique, précise et non invasive [683].

b. Cas d'usage en thérapeutique

Alireza Borjali et son équipe présentent en 2020 un modèle d'intelligence artificielle visant à identifier le modèle de prothèse totale de hanche portée par un(e) patient(e) [688,689].

Ce modèle est obtenu à l'issue d'un apprentissage par transfert puis d'un apprentissage profond supervisé réalisé sur un réseau de neurones convolutif (*DenseNet-201*). Les données d'entraînement sont constituées de 198 radiographies antéro-postérieures de la hanche collectées auprès de patients ayant subi une implantation de prothèse totale de hanche (*figure 118*). Ces radiographies ont été labellisées selon le modèle de prothèse de hanche implantée (*Accolade II, Corail, S-ROM*).

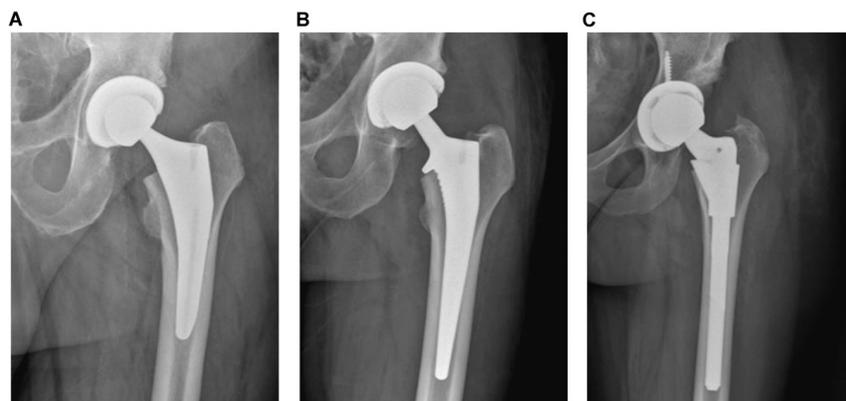


Figure 118. Exemples de données d'entraînement. Ces radiographies antéro-postérieures de la hanche présentent trois modèles de prothèses communément implantées : (A) *Accolade II*, (B) *Corail*, (C) *S-ROM*. (Reproduit avec la permission de l'éditeur John Wiley and Sons. Auteurs : Kartik M. Varadarajan et coll.) [688]

Il s'agit donc d'un modèle de classification prenant en entrée une radiographie antéro-postérieure de la hanche et indiquant en sortie la classe prédite : modèle *Accolade II* vs. modèle *Corail* vs. modèle *S-ROM*. En considérant uniquement les modèles de prothèse sus-cités, le modèle parvient à atteindre des performances élevées : exactitude = 100 %. Le temps d'exécution moyen du modèle est de l'ordre de la seconde.

Les révisions des prothèses totales de hanche exposent les patients à une morbidité plus importante que les arthroplasties de première intention [690]. Le risque de complications lié à cette intervention peut être atténué en identifiant précisément le modèle de la prothèse portée par le/la patient(e) avant l'opération. Dans le cas contraire, le temps et la complexité de la chirurgie peuvent être augmentés, exposant le/la patient(e) à des risques plus élevés d'hémorragie, de perte osseuse et de prolongement du temps de rétablissement [691]. Plusieurs éléments peuvent rendre difficile l'identification du modèle de prothèse porté par le/la patient(e), dont notamment l'incomplétude des informations conservées dans le dossier médical du/de la patient(e) [692]. De fait, les soignants s'appuient le plus souvent sur l'interprétation de radiographies pour identifier le modèle de prothèse portée et les composants à l'origine de l'échec du traitement. Cependant, cette méthode est chronophage, nécessite des années d'expérience et comporte par nature une part de subjectivité [692]. La fiabilité de cette démarche pourrait être favorisée par des méthodes d'interprétation automatiques et objectives, augmentant ainsi le taux de réussite de la révision et réduisant le risque de complications. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à identifier le modèle de prothèse totale de hanche portée par un(e) patient(e), de façon automatique, précise et non invasive [688].

44. Urologie

a. Cas d'usage en diagnostic

Ilker Ali Ozkan et son équipe présentent en 2018 des modèles d'intelligence artificielle visant à différencier les cystites des urétrites [693,694].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés : arbre de décision, forêt aléatoire, machine à vecteurs de support, réseau de neurones (*figure 119*). Les données d'entraînement comportent 15 variables socio-démographiques, cliniques et biologiques collectées auprès de 48 patients atteints d'infection urinaire. Ces données ont été labellisées selon que le/la patient(e) présente une cystite ou une urétrite (déterminé à partir d'échographies rénales et vésicales).

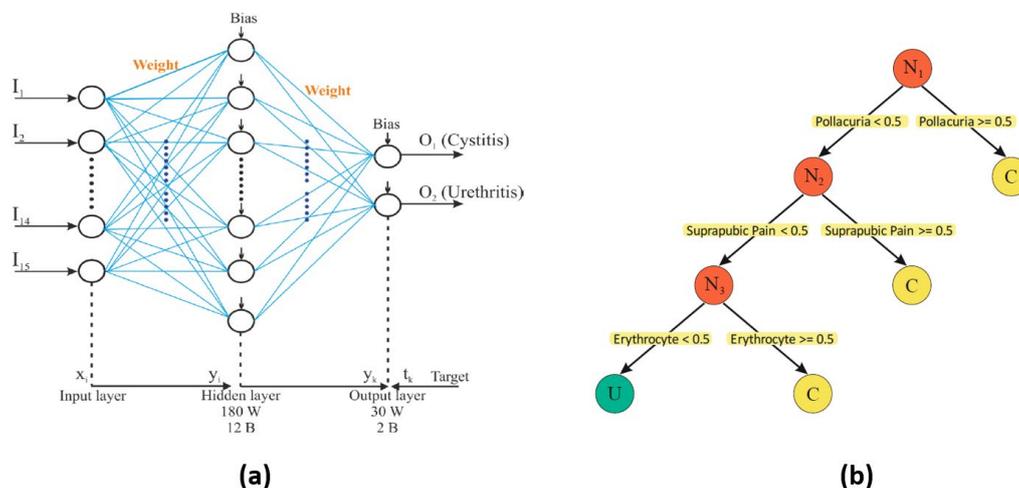


Figure 119. Représentations schématiques de deux modèles parmi ceux présentés : (a) réseau de neurones à propagation avant ; (b) arbre de décision (N : nœud, C : cystite, U : urétrite). (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : Ilker Ali Ozkan et coll.) [693]

Il s'agit donc de modèles de classification prenant en entrée des valeurs correspondant aux variables d'entraînement sus-mentionnées et indiquant en sortie la classe prédite : présence d'une cystite vs. présence d'une urétrite non-spécifique (*figure 119*). Le réseau de neurones est le modèle présentant les meilleures performances : exactitude = 98,30 % ; sensibilité = 97,77 % ; spécificité = 100 % ; valeur prédictive positive = 100 % ; valeur prédictive négative = 93,33 %.

L'infection urinaire fait partie des maladies infectieuses les plus courantes et celle-ci surviendrait chez plus de la moitié de la population générale au cours de leur vie. Cette dénomination regroupe différents cas de figure se différenciant par la localisation de l'infection mais dont les symptômes peuvent être similaires [695]. De fait, le diagnostic de la localisation est difficile en routine. Pourtant, ce diagnostic est primordial car, si l'infection n'est pas traitée adéquatement, celle-ci peut provoquer des complications graves et engendrer des

antibiorésistances bactériennes [696,697]. Identifier la localisation de l'infection peut chez certains patients nécessiter l'emploi de méthodes invasives de prélèvement (cystoscopie) à partir des uretères, exposant le/la patient(e) à un risque supplémentaire de complications [698]. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à différencier les cystites des urétrites en s'appuyant uniquement sur des symptômes et résultats biologiques, de façon automatique, rapide, précise, non invasive et moins coûteuse [693].

b. Cas d'usage en thérapeutique

Andrew J. Hung et son équipe présentent en 2019 des modèles d'intelligence artificielle visant à prédire la durée de récupération de la continence urinaire suite à une prostatectomie totale robot-assistée [699].

Ces modèles sont obtenus à l'issue d'apprentissages automatiques supervisés réalisés par différents algorithmes : régression de Cox, forêt aléatoire, apprentissage profond (*DeepSurv*). Les données d'entraînement comportent des variables clinicopathologiques et des métriques de performance automatiquement calculées durant les opérations. Ces données ont été collectées auprès de 60 patients ayant subi une prostatectomie totale robot-assistée et ont été labellisées pour chaque patient selon la durée de récupération de sa continence urinaire.

Il s'agit d'un modèle de régression prenant en entrée des valeurs correspondant aux variables sus-mentionnées et indiquant en sortie une valeur numérique correspondant à la prédiction du temps de récupération de la continence urinaire. Le réseau de neurones (*DeepSurv*) est le modèle présentant les meilleures performances : erreur absolue moyenne = 85,9 ; C-index = 59,9 %.

La prostatectomie totale constitue la cause majeure d'incontinence urinaire chez l'homme [700]. De nombreux facteurs de risque, relatifs à la technique de chirurgie et au patient (ex. : âge, indice de masse corporelle, longueur de l'urètre membraneux, insuffisance sphinctérienne), ont été identifiés [701,702,703]. Malgré les progrès réalisés par les techniques de chirurgie, la durée nécessaire pour retrouver une continence urinaire après l'opération reste très variable d'un patient à l'autre [704]. La possibilité d'estimer cette durée avant l'opération permettrait d'ajuster au mieux la technique employée selon le patient. En post-opératoire, cela permettrait d'anticiper la durée et la nature du traitement visant à favoriser la récupération de la continence. Les résultats obtenus révèlent qu'un modèle d'intelligence artificielle serait potentiellement en mesure de renforcer la capacité des soignants à estimer la durée de récupération de la continence urinaire suite à une prostatectomie totale robot-assistée, de façon automatique et non invasive [699].

II. Intérêts et perspectives

1. Synthèse des bénéfices pour les patients et les professionnels de santé

Les progrès spectaculaires réalisés par l'intelligence artificielle lui valent souvent d'être considérée comme étant précurseur d'un changement de paradigme pour le champ médical [705,706,707]. Lorsque celle-ci est bien entraînée, les bénéfices sont significatifs pour les patients et les professionnels de santé.

Tout d'abord, le caractère automatique et la vitesse d'exécution des modèles permet aux soignants de déléguer à ces technologies les tâches répétitives, chronophages et fastidieuses [708]. Le temps et l'énergie économisés peuvent ainsi être réinvestis sur des tâches plus complexes et/ou requérant des compétences humaines telles que l'empathie (ex. : annonce de fin de vie) ou la négociation (ex. : explication du rapport bénéfice/risque, éducation thérapeutique en cas de mauvaise observance d'un traitement) [709,710]. Cela permet aussi d'accorder plus de temps à chaque patient(e) et/ou de prendre en charge davantage de patients [711].

De plus, les modèles développés présentent des performances de plus en plus élevées, allant jusqu'à dépasser dans certains cas celles des professionnels de santé expérimentés (ex. : diagnostic à partir d'un examen d'imagerie) [712]. Une autre approche consiste à, non pas comparer indépendamment les performances d'un modèle avec celles des professionnels de santé, mais plutôt combiner celles-ci. Il s'avère que, dans certains cas, l'association des capacités humaines et des capacités propres à l'intelligence artificielle offre des performances optimales [713].

La télémédecine se voit également renforcée grâce à des modèles d'intelligence artificielle capables d'améliorer la précision du télédiagnostic ou du suivi à distance d'un traitement au moyen/long court [714].

Bien que le développement, l'évaluation et la mise en place de ces outils impliquent des coûts importants, les intelligences artificielles ont la capacité de réduire les coûts liés à la santé sur le moyen et long terme : nombre réduit d'examen cliniques nécessaires, optimisation des soins, diminution de l'errance diagnostique et thérapeutique, etc [715]. De fait, il se pourrait que les intelligences artificielles appliquées au diagnostic et à la thérapeutique contribuent à préserver et étendre les couvertures sanitaires universelles [716]. Cela dépendra cependant des coûts de production et de maintenance de ces architectures, ainsi que de leurs modèles économiques [717].

Plus spécifiquement en diagnostic, l'intelligence artificielle est en mesure de réduire le nombre, la durée ou la dose des examens nécessaires (ex. : diagnostic de fracture uniquement à partir d'une radiographie, reconstruction d'images tomographiques de qualité à partir d'images tomographiques en faible dose), permettant ainsi de s'affranchir de certains examens invasifs ou coûteux tout en préservant la précision du diagnostic [360,598,599]. Cela permet notamment de réduire les risques auxquels sont exposés les patients et pourrait constituer un levier d'amélioration de l'accès au diagnostic pour les territoires les plus démunis en matériel médical (pays en voie de développement, déserts médicaux) [718].

En lui fournissant des données collectées antérieurement à l'apparition des premiers symptômes visibles d'une pathologie, un modèle peut également être entraîné à détecter précocement celle-ci [579,623,633]. Sous réserve que des solutions thérapeutiques ou préventives existent et démontrent leur efficacité, cette utilisation pourrait grandement améliorer le pronostic des patients présentant une pathologie dégénérative au stade asymptomatique ou caractérisée par des symptômes équivoques.

D'un point de vue épidémiologique, cette approche peut être d'autant plus bénéfique qu'il est possible d'entraîner un modèle à identifier des patients à risque d'une ou plusieurs pathologies à partir de leur dossier médical [406,443].

L'autodiagnostic, par exemple à partir d'applications mobiles embarquant une intelligence artificielle (ex. : détection d'un mélanome à partir d'une photographie), pourrait également constituer un levier de diagnostic précoce à grande échelle [719].

Les fabricants devront cependant veiller à ce que la spécificité (ou « taux de vrais négatifs ») de ces différents outils diagnostics soit significativement élevée, sans quoi la surmédicalisation (en particulier le surdiagnostic et le surtraitement) qui en résulterait pourrait représenter un risque important pour les patients et engendrer des dépenses de santé inutiles [720].

Quant à la thérapeutique, les progrès de l'intelligence artificielle offrent des opportunités prometteuses en médecine de précision. Pour un(e) patient(e) donné(e), il est possible de prédire l'efficacité d'un traitement, ses effets indésirables, la durée de traitement nécessaire ainsi que le pronostic suite à l'administration [382,414,423,565,662,699]. Cela s'applique également aux transplantations pour lesquelles une intelligence artificielle est en mesure d'évaluer le rapport bénéfique/risque pour un receveur et un donneur spécifiques, permettant ainsi d'optimiser l'attribution des greffons [675].

L'intelligence artificielle peut aussi être employée pour surveiller ou suivre un(e) patient(e) au cours de son traitement [461]. En anesthésie-réanimation, un modèle peut être entraîné pour surveiller en continu et avec précision les constantes d'un(e) patient(e) et prédire les

événements nécessitant une intervention (ex. : survenue d'asynchronies patient-ventilateur durant une ventilation mécanique) [641]. Dans le cadre d'un traitement au moyen/long court, la surveillance et le suivi peuvent être assurés par un matériel médical embarquant un modèle d'intelligence artificielle. Ainsi, le/la patient(e) peut être guidé(e) dans son autosurveillance et/ou le soignant peut effectuer le suivi à distance si le matériel médical est connecté [654]. Cette approche peut, dans certains cas, améliorer significativement la qualité de vie des patients (ex. : en limitant les déplacements dans le cas d'un handicap moteur), de surcroît lorsque ceux-ci résident dans un désert médical [562].

L'intelligence artificielle s'applique également aux problématiques de la recherche. Un modèle peut par exemple être entraîné à effectuer des génotypages (ex. : identification de gènes causaux d'une maladie rare) [444]. Les progrès réalisés en apprentissage profond sont également très prometteurs en recherche thérapeutique *in silico*. En facilitant et en accélérant l'identification de nouvelles cibles thérapeutiques, l'intelligence artificielle ouvre la voie vers de nouvelles perspectives thérapeutiques, notamment via le repositionnement médicamenteux [469,531,594].

2. Vers une numérisation du patient ?

La numérisation des données est une notion récurrente et essentielle que nous retrouvons à chaque cas d'usage présenté dans ce manuscrit. En effet, les données d'entraînement doivent nécessairement être converties sous forme numérique avant d'être fournies aux modèles à entraîner. De façon plus générale, la tendance à numériser les données de santé s'affirme par souci de traçabilité, de conservabilité, d'interopérabilité, de facilité de stockage et de transmission [721]. L'extension de cette démarche pourrait mener à terme vers la démocratisation des jumeaux numériques en santé (*figure 120*).

Un jumeau numérique désigne une réplique virtuelle d'un objet, processus ou système [723]. En s'appuyant sur l'apprentissage automatique, les jumeaux numériques peuvent simuler la dynamique de fonctionnement de ce qui est répliqué tout au long de son cycle de vie. Appliquée au contexte clinique, cette technologie serait en mesure de calquer les processus physiologiques d'un(e) patient(e) à partir de l'ensemble des données numériques disponibles relatives à la santé de ce/cette patient(e) (imagerie, pathologie numérique, génomique, etc.). L'intérêt majeur de cette technologie résiderait dans la capacité d'automatisation holistique que celle-ci offre pour un individu, facilitant notamment les diagnostics précoces, les prises en charge en urgence ou en soins intensifs ou encore l'optimisation des parcours de soins (*figure 120*) [724,725].

Cependant, cela impliquerait de centraliser les données de santé d'un(e) patient(e), que celles-ci soient nombreuses et se rapportent à différents organes [722]. Cette application des jumeaux numériques pourrait potentiellement devenir, dans un futur hypothétique, l'aboutissement de la démarche initiée par le dossier médical informatisé/partagé [726].

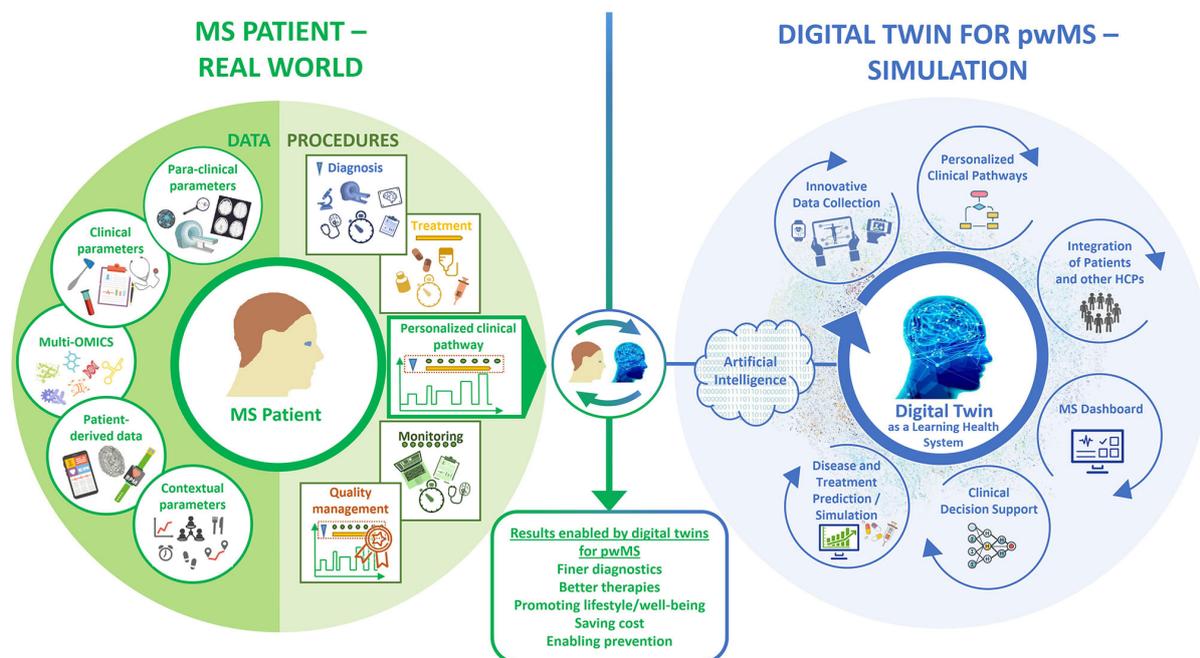


Figure 120. Application du concept de jumeau numérique aux patients atteints de sclérose en plaques. *MS : multiple sclerosis (sclérose en plaques) ; pwMS : people with multiple sclerosis (personne atteinte de sclérose en plaques) ; HCPs : healthcare professionals (professionnels de santé). (Reproduit selon les termes de la licence CC BY 4.0. Auteurs : Isabel Voigt et coll.) [722]*

III. Limites, défis et appréhensions

Nonobstant le potentiel et les bénéfices énumérés, l'intelligence artificielle n'est pas une solution universelle et incontestable. Certaines problématiques propres à la santé challengent ces outils et mettent en lumière leurs limites actuelles. De plus, leur utilisation dans le cadre clinique soulève de nombreuses questions d'ordres éthique et juridique [727]. Ce volet a pour but de présenter succinctement ces limites et défis majeurs ainsi que les appréhensions suscitées par l'application des intelligences artificielles au diagnostic et à la thérapeutique.

1. L'applicabilité et la performance en contexte clinique

La première question que nous pouvons nous poser est celle de l'applicabilité. Celle-ci est variable selon la spécialité médicale, la pathologie à diagnostiquer ou à traiter, ainsi que la technique d'intelligence artificielle sollicitée.

La conception des systèmes experts, dont le principe est de calquer le raisonnement d'un expert, s'appuient sur trois postulats peu voire non applicables au corps humain [259] :

- L'ensemble du fonctionnement de l'objet d'étude (organe, corps humain) est connu ;
- L'ensemble des problèmes potentiels à résoudre (pathologies) ainsi que leurs solutions (traitements) sont connus ;
- La formulation des éléments observables et imputables au problème (symptômes) ne peut pas prêter à l'équivoque.

Lorsqu'il s'agit d'appliquer les systèmes experts au diagnostic et à la thérapeutique, ceux-ci sont confrontés à la limite tracée par l'étendue des connaissances. De plus, il se pose un problème d'ontologie en raison de la subjectivité du langage employé pour désigner les symptômes (ex. : le terme « malaise »). Certains symptômes sont d'ailleurs par nature subjectifs car soumis à l'appréciation du/de la patient(e) (ex. : échelle de douleur). Par ailleurs, les recommandations quant à la pratique des professionnels de santé sont fondées sur un consensus évoluant au fil de l'émergence des connaissances et leur remise en cause. Cette pratique est, dans les faits, le résultat de facteurs individuels plus nombreux encore (ex. : ressources matérielles et humaines à disposition, connaissances et expériences individuelles, atypisme relatif des patients, dimension psycho-sociale, etc.). Ces variations sont trop nombreuses et spécifiques pour être toutes intégrées au sein d'un système expert [261].

L'apprentissage automatique, de par sa nature généralisatrice, ne se prête pas non plus à cette variabilité. Qui-plus-est, l'apprentissage automatique est également soumis par nature à l'état des connaissances puisqu'il s'appuie sur des données d'entraînement. S'ajoute à cela le problème de fiabilité des données qui conditionne la qualité de l'entraînement [728].

De fait, à l'heure actuelle, l'intelligence artificielle appliquée au diagnostic est davantage à rapprocher d'un test diagnostique que d'un assistant capable de prendre en charge la démarche diagnostique dans sa globalité. Il en va de même lorsque l'intelligence artificielle est appliquée à la thérapeutique. Celles-ci n'en demeurent pas moins efficaces sur des tâches spécifiques et bien délimitées (*voir section P2.I.*). Cependant, l'évaluation d'un modèle d'intelligence

artificielle à partir d'un jeu de données est-elle suffisante pour préjuger de son efficacité et de sa sécurité une fois intégrée au contexte clinique [391] ? Nous pouvons supposer que non et qu'il convient d'évaluer ce modèle avec les mêmes précautions et méthodologies que les tests diagnostiques et protocoles thérapeutiques, c'est-à-dire en condition expérimentale et clinique [729].

C'est pourquoi, lorsqu'un modèle vise à être intégré dans une démarche diagnostique ou thérapeutique, les modalités de son implémentation dans les usages médicaux devraient être pensées dès les premières phases de conception [730]. Cette réflexion devrait être menée par l'équipe de développement de concert avec un échantillon d'utilisateurs finaux. Le pilotage de cette démarche, ainsi que ses aspects périphériques (interfaçage, adhérence de la part des utilisateurs cibles, pérennité du modèle), devraient dans l'idéal être confiés à du personnel spécialisé en sciences de l'implémentation [729]. En outre, encourager la mise en place de comités et congrès sur le sujet pourrait favoriser le développement d'une vision commune au sein d'une spécialité [731].

Une autre réflexion importante à mener dès la conception du projet concerne la disponibilité des données d'entrée lors de l'utilisation du modèle par l'utilisateur. Les données d'entrée nécessaires à l'inférence devraient être, dans l'idéal, limitées en nombre et disponibles au moment prévu de l'utilisation du modèle (ex. : au sein d'un service de radiologie, une radiographie pour prédire la présence d'une tumeur cancéreuse) [730,732]. Si un modèle est performant en théorie mais que ses données d'entrée sont trop difficiles à obtenir pour l'utilisateur (ex. : données génétiques en soins courants) ou trop nombreuses pour la prédiction souhaitée (ex. : 100 informations à renseigner pour estimer la présence d'une angine bactérienne), ce modèle est inapplicable et/ou peu utile concrètement. Cela constitue une raison supplémentaire pour laquelle l'implication des acteurs du terrain dans la conception d'un modèle d'intelligence artificielle est essentielle. Dans le cas où la disponibilité des données d'entrée serait variable selon les contextes d'utilisation, une solution alternative consiste à prévoir différents cas de figure pour lesquels l'architecture serait en mesure d'effectuer une prédiction. Chacun de ces cas de figure correspondrait à des données d'entrée spécifiques, permettant ainsi une souplesse mais impliquant une variabilité de la performance [730].

2. L'accessibilité pour les patients et les professionnels de santé

Le second sujet que nous pouvons aborder est celui de l'accessibilité pour les professionnels de santé et les patients.

Celle-ci dépend en premier lieu des coûts et des capacités d'investissement, notamment de la part des établissements de santé ou professionnels de santé demandeurs, des institutions publiques, des universités et des entreprises du secteur de la technologie et de la santé [730]. Les coûts de production d'une intelligence artificielle sont généralement élevés en raison du niveau de qualification que cela exige et les coûts liés à l'implémentation ne sont pas non plus à négliger [733]. S'ajoutent à cela des coûts relatifs au traitement des données, tel que la labellisation qui exige généralement l'expertise d'un ou plusieurs spécialistes [734]. Ces dépenses peuvent et devraient être étudiées relativement à l'impact sur la santé publique que le projet est en mesure de garantir (ex. : évaluation d'impact sur la santé) [735]. Les autorités de santé peuvent également influencer l'accessibilité selon la prise en charge ou non, par les régimes de sécurité sociale, des dispositifs médicaux embarquant un modèle d'intelligence artificielle [736]. Outre leur potentiel d'investissement dans la recherche et le développement des intelligences artificielles, les universités endossent également la charge d'investir dans la formation des futurs professionnels de santé à employer ces technologies, comprendre leur fonctionnement et être acteurs de leur évolution [737].

En effet, l'accessibilité dépend aussi et surtout de la capacité des professionnels de santé et des patients à s'approprier ces outils. L'un des leviers fondamentaux est la conception d'interfaces intuitives et simples d'utilisation pour l'utilisateur final cible (*UI/UX design*). L'interopérabilité de cette interface et du modèle qu'il embarque devrait également faire l'objet d'une attention particulière. L'architecture devrait être pensée pour fonctionner sur le plus de plateformes possibles et être adaptable à des formats variés de données (de surcroît si celle-ci est conçue pour analyser des dossiers médicaux). Les déserts médicaux et les pays en voie de développement sont particulièrement concernés par l'interopérabilité en raison de leur accès limité à du matériel récent [730].

3. Les données de santé

Comme l'indique chacun des cas d'usage que nous avons présenté, les données de santé sont la matière première indispensable à la création d'intelligences artificielles appliquées au diagnostic ou à la thérapeutique. Cela est particulièrement vrai en apprentissage automatique, qui est la tendance majoritaire à l'heure actuelle.

Les données de santé employées en apprentissage automatique devraient donc respecter des standards de qualité car la qualité des données d'entraînement conditionne la qualité de l'entraînement et par conséquent la qualité des prédictions du modèle produit. Les données utilisées pour évaluer le modèle doivent également être de qualité sans quoi la sélection du modèle le plus performant est biaisée [738]. Schématiquement, la qualité des données peut être décomposée selon sept aspects majeurs : exactitude, fiabilité, complétude, pertinence, validité, actualité, uniformité [739]. Lorsque l'apprentissage est supervisé, ce qui est souvent le cas en diagnostic et en thérapeutique, l'un des enjeux majeurs réside dans la qualité de la labellisation car celle-ci influence directement la fiabilité de la prédiction [740]. La labellisation devrait donc s'appuyer exclusivement sur la méthode contemporaine la plus précise pour mesurer l'information permettant la déduction du label. Il existe des modèles capables de se charger de la labellisation, leur fiabilité exige cependant d'avoir à disposition des exemples en nombre suffisant et dont les labels sont fiables [741].

La qualité n'est cependant pas suffisante puisque des données d'entraînement en faible quantité induisent également des biais potentiels [408]. De façon générale, il est recommandé d'entraîner et évaluer un modèle d'apprentissage automatique avec une quantité importante de données, et ce pour les raisons suivantes :

- Plus les données d'entraînement sont nombreuses, plus il est aisé pour un modèle d'identifier des tendances et motifs généralisables. Par exemple en apprentissage supervisé, plus le modèle dispose d'exemples labellisés, plus il sera aisé pour celui-ci d'associer ces tendances et motifs aux différents labels et de généraliser ces connaissances à de nouvelles données.
- Plus les données d'entraînement sont nombreuses, plus il est aisé pour un modèle de s'affranchir des biais et imprécisions caractérisant ces données. Cela peut être particulièrement utile lorsque les données sont très variables ou imparfaites (ex. : reconnaissance visuelle ou acoustique).
- Plus le jeu de données destiné à l'évaluation est riche, plus celle-ci est précise et fiable.

Parmi les problématiques induites par un déficit en données, celle du déséquilibre de classes fait partie des plus connues. Comme son nom l'indique, ce problème se produit en classification lorsque le nombre d'exemples est significativement différent pour une ou plusieurs classes au sein d'un jeu de données (ex. : 10000 photographies de tumeurs bénignes + 100 photographies de tumeurs malignes). Cela a pour conséquence d'altérer la capacité du modèle à détecter les exemples rares de la classe minoritaire [391]. Plus spécifiquement lorsque ce problème concerne un modèle entraîné au diagnostic binaire (malade vs. non malade), la spécificité se voit favorisée au détriment de la sensibilité si le jeu de données comporte moins de cas malade que de non malade, l'inverse se produit si le jeu de données comporte plus de cas malade que de non malade [742]. Il existe différentes méthodes permettant d'atténuer un déséquilibre de classes (ex. : *Synthetic Minority Oversampling Technique*), cependant l'utilisation d'une quantité de données importante et équilibrée subsiste comme étant la solution offrant les meilleures perspectives de performance [712].

C'est pourquoi les initiatives de partage (ex. : *MIMIC-III*) ou de centralisation (ex. : *Health Data Hub*) des données sont précieuses et primordiales pour le développement des intelligences artificielles appliquées au diagnostic et à la thérapeutique [730].

La représentativité des données d'entraînement est également essentielle pour garantir de bonnes performances au modèle. Les prédictions émises par un modèle entraîné uniquement ou majoritairement sur des données issues d'un groupe spécifique présente le risque d'être significative uniquement pour les patients représentés par ce groupe [743,744]. De fait, la disparité des moyens de collecte et de conservation des données induit un risque d'inéquité quant à la représentativité des modèles. L'entraînement des modèles devrait donc prendre en considération cette problématique en intégrant dans son jeu de données des exemples issus de populations variées : différentes zones géographiques, différents systèmes de santé, différentes situations socio-économiques, patients présentant des comorbidités variées, etc. [46]. Dans le même esprit, l'évaluation des modèles devrait comporter une analyse des potentiels biais discriminatoires [730,745].

Un autre défi majeur posé par les données de santé concerne leur diversité. Celles-ci peuvent être de différentes natures (ex. : cliniques, biologiques, socio-démographiques) et se présenter selon des formats divers (ex. : tableurs, photographies, textes structurés ou non) [272,273]. En l'occurrence, environ 80% des données patients sont non structurées et sous forme de texte (ex. : comptes rendus d'examen ou d'hospitalisation) [746]. Rendre exploitables ces données en

vue d'un apprentissage automatique implique une démarche d'extraction et de structuration qui est très chronophage lorsque celle-ci est manuelle [747].

Des méthodes faisant intervenir du traitement du langage naturel peuvent être employées pour automatiser l'extraction des informations souhaitées, cependant celles-ci présentent à l'heure actuelle des performances en deçà des exigences [748]. Les algorithmes d'apprentissage non supervisé sont notamment prometteurs pour cette tâche spécifique [749]. Parallèlement, une réflexion collaborative devrait être menée quant à la standardisation des dossiers médicaux informatisés [730].

D'autre part, cette diversité des données de santé est en partie le reflet des démarches diagnostique et thérapeutique qui s'appuient souvent sur des éléments de différentes natures. De fait, lorsque cela est pertinent, les modèles devraient être entraînés de façon à prendre en entrée autant de formats de données que nécessaire (apprentissage multimodal) [750].

4. L'explicabilité

L'explicabilité est un sujet très populaire en intelligence artificielle. Celle-ci s'intéresse aux causes et au cheminement ayant amené un modèle à fournir une prédiction. L'explicabilité s'appuie notamment sur l'interprétabilité qui désigne la capacité d'une intelligence artificielle à fournir des justifications pour ses prédictions, autrement dit à mettre en évidence les données et caractéristiques qui en sont à l'origine [751].

L'explicabilité est une problématique primordiale pour de nombreuses raisons. Tout d'abord, en permettant de saisir le fonctionnement d'un modèle, l'explicabilité contribue à améliorer et affiner ce modèle. De plus, l'explicabilité permet de mieux gérer les risques liés à l'utilisation d'un modèle, en facilitant l'identification de ses limites et des cas spécifiques pour lesquels son utilisation pourrait ne pas être appropriée. Enfin, l'explicabilité est également importante pour des raisons éthiques puisque celle-ci permet aux utilisateurs de mieux comprendre comment les décisions prises par un modèle sont influencées par les données et l'algorithme qui le sous-tendent [752]. En outre, des modèles à l'interprétabilité démontrée pourraient être sollicités par les chercheurs pour tenter d'expliquer des phénomènes mal compris jusqu'alors.

Si l'explicabilité est généralement satisfaisante pour les systèmes experts et les algorithmes d'apprentissage automatique les plus simples (ex. : arbres décisionnels, régression linéaire), celle-ci devient de plus en plus insaisissable à mesure que les algorithmes se complexifient

[302]. Notamment, les modèles entraînés par un algorithme d'apprentissage profond sont communément qualifiés de « boîtes noires » en raison de leur faible explicabilité [753].

En l'occurrence, ce sont les réseaux de neurones profonds qui, à l'heure actuelle, garantissent les meilleures performances lorsqu'il s'agit d'appliquer l'intelligence artificielle au diagnostic et à la thérapeutique. Il s'agit cependant de deux disciplines pour lesquels l'explicabilité est fondamentalement recherchée [754]. Pour autant, nous n'avons actuellement pas connaissance de l'ensemble des processus relatifs aux pathologies et aux mécanismes d'actions des traitements [755]. Le crédit accordé à la sémiologie médicale ainsi qu'aux protocoles diagnostiques et thérapeutiques se fonde en grande partie sur des démonstrations statistiques [756]. La question de l'explicabilité en diagnostic et en thérapeutique induit donc une autre question, toute aussi complexe et irrésolue : sommes-nous disposés à accorder ce même crédit à une intelligence artificielle si sa performance est statistiquement et cliniquement démontrée ? Cette question se pose particulièrement pour les décisions relatives à des pathologies ou des traitements mal connus, lorsque la véracité de la prédiction ne peut pas être vérifiée ou lorsque celle-ci contredit le résultat obtenu par d'autres méthodes [75].

Plusieurs pistes sont explorées en vue d'améliorer l'interprétabilité des intelligences artificielles [757]. Concernant les systèmes experts, les efforts se poursuivent quant à la description des raisonnements amenant aux décisions suggérées [758]. En apprentissage automatique, la mise à disposition du niveau de confiance des prédictions est encouragée. Cela se traduit par l'indication des probabilités de certitude assorties à chaque valeur possible en sortie (ex. : 96% malade / 4% non malade) [730]. Les méthodes de visualisation, telles que les cartes thermiques, gagnent également en popularité. En vision par ordinateur, les cartes thermiques peuvent être générées sous forme d'une visualisation d'activation de classe ou d'une cartographie de la saillance. Les cartes de saillance fournissent une représentation chromatique des zones d'intérêt considérées par le modèle pour effectuer sa prédiction (*figure 121*) [358]. Sur le même principe, l'activation de classe permet de visualiser les zones d'une image impliquées dans l'identification d'une classe lors d'une classification [759]. Générer des cartes de saillance ou d'activation de classe sur différentes couches d'un réseau de neurones permet de décliner cette visualisation à la globalité du modèle pour ainsi disposer d'une vue panoramique de l'analyse effectuée par celui-ci (*figure 122*) [760]. En outre, l'approche neuro-symbolique semble être une piste prometteuse permettant d'associer la puissance de l'apprentissage profond avec le niveau d'interprétabilité des systèmes experts (*voir section P1.III.3.*) [372].

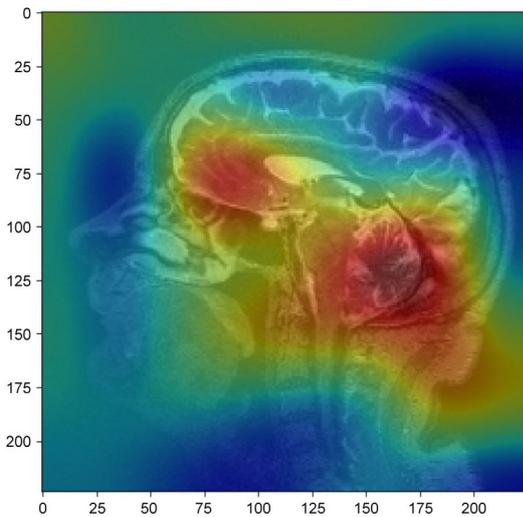


Figure 122. Exemple de carte de saillance réalisée lors d'une prédiction de démence à partir d'une image par résonance magnétique (Reproduit avec la permission de la revue *The Journal of Nuclear Medicine*. Auteurs : Felix Nensa et coll.) [75]



Figure 121. Exemple de cartes de saillance générées sur différentes couches d'un réseau de neurones lors d'une prédiction de trouble de la conduite à partir d'une image par résonance magnétique. Les couleurs représentent des scores de contribution à la classification. (Reproduit avec la permission de l'éditeur Springer Nature. Auteurs : Jianing Zhang et coll.) [760]

Au-delà des exemples présentés ci-dessus, l'interprétabilité des intelligences artificielles fait l'objet de recherches intensives. Il est cependant difficile d'envisager si ces efforts parviendront à suivre la cadence toujours plus rapide des progrès réalisés en intelligence artificielle [761,762].

5. Les questions d'ordre éthique

De nombreuses questions éthiques se posent quant à l'application des intelligences artificielles au secteur de la santé. Les opportunités offertes par l'intelligence artificielle ne doivent pas faire perdre de vue les dangers potentiels qui leur sont inhérents, notamment en cas d'utilisation abusive [763]. La pratique des professionnels de santé est caractérisée par des codes éthiques hérités d'une longue tradition, telle que décrite par les serments d'Hippocrate et de Galien, et enrichis par des codes plus modernes (ex. : *Déclaration de Genève*) [764,765]. Ces valeurs ont su être préservées au cours de l'évolution des pratiques, cela ne devrait pas être différent avec l'implémentation des intelligences artificielles au sein de celles-ci [766]. Pour ce faire, et afin de gérer au mieux les défis et les risques, il est important de constamment considérer ces questions éthiques et de poursuivre la réflexion autour de celles-ci. Dans ce volet, nous

proposons une synthèse des enjeux éthiques majeurs relatifs aux intelligences artificielles appliquées au diagnostic et à la thérapeutique.

- L'équité [767] : comment s'assurer que l'intelligence artificielle appliquée au diagnostic et à la thérapeutique soit accessible à tous, que les patients de tous les milieux sociaux et économiques aient accès aux bénéfices qu'elle peut apporter ? (*voir section P2.III.1.*)
- La relation soignant-patient : comment implémenter l'intelligence artificielle sans altérer la relation soignant-patient ?

Appliquée à la robotique, l'intelligence artificielle permet d'attribuer à un robot des capacités de perception, de décision et d'action [768]. Dans le domaine de la santé, cela permet par exemple d'assister une chirurgie, et ainsi d'améliorer la précision des gestes chirurgicaux ou d'opérer à distance [769]. Par ailleurs, les perspectives d'automatisation des opérations chirurgicales semblent de moins en moins hypothétiques (ex. : *robot chirurgien STAR*) [770]. Les prothèses intelligentes ou encore les robots d'assistance aux personnes sont d'autres exemples d'application réputées [226,771]. En 2017, une équipe chinoise met au point le premier robot-médecin (nommé *Xiaoyi*, signifiant « petit docteur » en chinois) entraîné à partir d'un million d'images, deux millions de dossiers de santé électroniques, 53 livres et 400 000 rapports médicaux. En 2018, celui-ci s'avérera en mesure de réussir l'examen national chinois de médecine avec un score de 456 points sur 600 [772,773]. Bien que l'utilisation actuelle de *Xiaoyi* se limite à établir des prédiagnostics et orienter les patients vers les bons spécialistes, cette initiative pose la question suivante : souhaitons-nous des programmes qui remplacent l'homme ou des programmes qui assistent et augmentent les capacités humaines ? L'idée d'un robot médecin autonome divise car si certains y voient l'opportunité de solutionner la pénurie de soignants, d'autres poussent la sonnette d'alarme en rappelant que la relation soignant-patient ne peut se résumer à des actes techniques [774,775]. Dans leur ouvrage *Santé et intelligence artificielle*, Bernard Nordlinger et Cédric Villani soulignent quatre qualités essentielles des soignants dont les intelligences artificielles sont dépourvues : la compassion, la compréhension, l'esprit critique et la conscience professionnelle [9]. La dimension empathique propre à la relation soignant-patient est un aspect fondamental de la démarche diagnostique ou thérapeutique qui ne peut, et ne pourra peut-être jamais, être égalée par un programme informatique [776]. En effet, l'empathie permet au soignant de considérer les préoccupations émotionnelles et les perspectives du/de la patient(e), et ainsi de mieux saisir les facteurs sociaux et psychologiques qui influencent sa santé. Qui plus est, cette démarche favorise l'établissement d'une relation de confiance favorisant la mise

en place d'un environnement de soins plus positif pour le/la patient(e) [777]. En plaçant l'intérêt des patients au centre des priorités, l'intelligence artificielle telle que nous la connaissons aujourd'hui semble se prêter davantage à assister les professionnels de santé sur des tâches spécifiques et délimitées [766,778].

- L'autonomie et la décision médicale : comment intégrer l'intelligence artificielle dans le processus de prise de décision médicale tout en préservant l'autonomie des professionnels de santé ?

Les performances atteintes par les modèles d'intelligence artificielle peuvent exposer leurs utilisateurs au biais d'automatisation [730]. Ce dernier pourrait amener les professionnels de santé à accorder excessivement leur confiance aux résultats fournis par les modèles, endossant ainsi une décision prise à l'aveugle. Pour éviter cela, les professionnels de santé devraient être informés des bénéfices et surtout des limites spécifiques aux modèles dont ils font l'usage. Ces informations pourraient être communiquées par le biais d'une documentation à l'attention des utilisateurs. La rédaction de celle-ci, afin d'être accessible aux utilisateurs cibles, devraient être le fruit d'une collaboration entre les développeurs et les professionnels de santé. Cette documentation devrait intégrer, dans la mesure du possible, des éléments relatifs à l'explicabilité. Globalement, l'explicabilité est l'élément central de cette question, ce qui nous amène à la question suivante [779].

- L'explicabilité des modèles [754] : comment permettre aux modèles d'intelligence artificielle d'être plus transparents et compréhensibles, afin que les professionnels de santé et les patients puissent comprendre leur fonctionnement et comment ils génèrent leurs décisions ? (*voir section P2.III.4.*)
- L'évaluation de l'efficacité et de la sécurité [729] : comment évaluer l'efficacité et la sécurité des modèles employés en diagnostic et en thérapeutique ? Comment s'assurer qu'ils respectent les standards de rigueur ? (*voir section P3.III.2.d.*)
- Les biais potentiels [745] : comment s'assurer que les modèles employés en diagnostic et en thérapeutique ne soient pas biaisés, et ne discriminent pas certaines populations ? (*voir section P2.III.3.*)
- La responsabilité juridique [780] : qui devrait être tenu responsable en cas d'erreur ou de décision inappropriée prise par un modèle employé en diagnostic ou en thérapeutique ? (*voir section P3.III.4.*)

- La confidentialité des données [781] : comment protéger les données de santé et gérer les risques de fuites de données ou de piratage ? (*voir section P3.I.*)
- Les conséquences sur le marché du travail : faut-il s'attendre à ce que l'intelligence artificielle appliquée au diagnostic et à la thérapeutique ait des répercussions sur le marché du travail ? Auquel cas, comment s'assurer que les personnes perdant leur emploi soient adéquatement formées et réintégrées professionnellement ?

L'une des préoccupations majeures de ce début du 21^{ème} siècle concerne l'évolution du marché du travail due à l'automatisation permise par l'intelligence artificielle. Les économistes avancent que, à l'instar de la révolution industrielle, nous pourrions potentiellement assister à un transfert d'emplois en masse plutôt qu'à une forte hausse du chômage [782]. Cette refonte présupposée des pratiques professionnelles pourrait particulièrement s'appliquer au secteur de la santé. La tendance démographique actuelle (croissance démographique associée à la hausse de l'espérance de vie et au vieillissement de la population) laisse entrevoir une hausse des besoins en personnels soignants [783,784]. L'intelligence artificielle pourrait donc se présenter comme un appui permettant aux soignants de prendre en charge davantage de patients (*voir section P2.II.1.*) [711].

Ces questions n'explorent pas exhaustivement l'ensemble des conséquences potentielles qui, pour certaines, relèvent de la futurologie et/ou sont imprévisibles [785]. Les réflexions qu'elles engagent n'en demeurent pas moins importantes à mener car, du point de vue de la santé publique, les bénéfices et opportunités offertes par l'intelligence artificielle sont résolument à l'ordre du jour [375].

6. Le cadre réglementaire et juridique

Toutes ces questions éthiques amènent naturellement à considérer la nécessité d'un cadre réglementaire et juridique [786]. Il s'agit de l'un des défis majeurs relatifs à l'intelligence artificielle appliquée au diagnostic et à la thérapeutique car ce cadre doit être apte à limiter au possible les dangers et mésusages sans entraver outre mesure l'évolution des pratiques médicales pouvant bénéficier aux patients. Celle-ci devrait donc reposer sur une évaluation scrupuleuse du rapport bénéfice/risque ainsi que sur une surveillance continue. Le volet suivant s'attachera à présenter ces aspects tels qu'ils sont appliqués actuellement et à proposer des pistes d'amélioration.

PARTIE 3 : Évaluation et réglementation des intelligences artificielles appliquées au diagnostic et à la thérapeutique

PARTIE 3 : Évaluation et réglementation des intelligences artificielles appliquées au diagnostic et à la thérapeutique

Qu'il s'agisse des lois visant à protéger les données personnelles ou du cadre réglementaire concernant la mise sur le marché, la réglementation des intelligences artificielles appliquées à la santé est variable d'un pays à l'autre. Il en va de même pour l'évaluation des modèles d'intelligence artificielle dont les pratiques ne sont pas standardisées [787]. Au cours de ce volet, nous tenterons de présenter une synthèse des exigences réglementaires au sein de l'Espace économique européen (EEE) et des méthodes à privilégier lors de l'évaluation d'une intelligence artificielle appliquée au diagnostic ou à la thérapeutique.

I. Réglementation relative aux données de santé

1. Pourquoi est-il important de réglementer les données de santé ?

La question de la réglementation des données est fondamentale en intelligence artificielle puisque celles-ci sont omniprésentes tout au long du cycle de vie d'un modèle, en particulier en apprentissage automatique (*figure 123*). En effet, la production d'un modèle nécessite a minima de collecter, conserver et traiter des données afin d'entraîner, d'évaluer et d'utiliser ce modèle [46,788]. Leur encadrement est une préoccupation majeure lorsqu'il s'agit de données de santé, et ce pour plusieurs raisons.

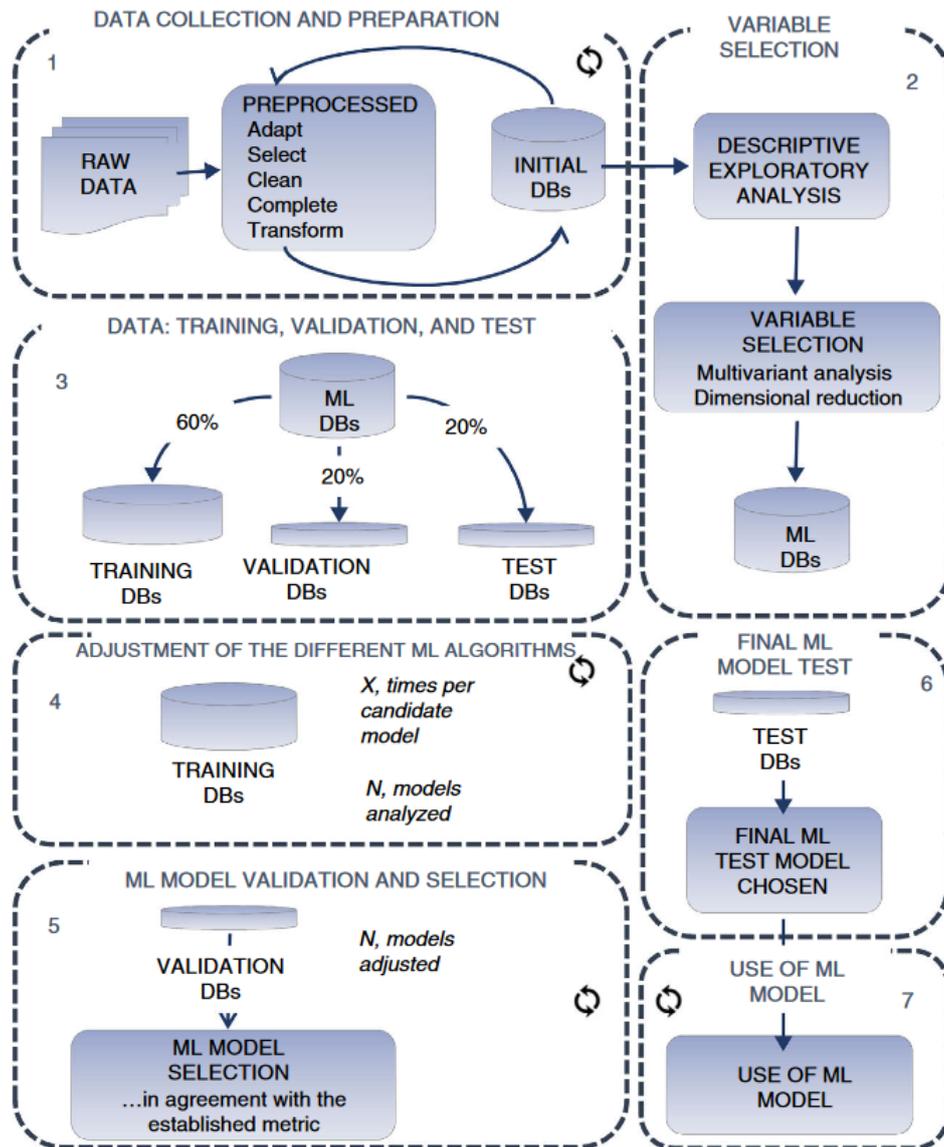


Figure 123. Cycle de vie d'un modèle d'apprentissage automatique.

DB : database (base de données) ; ML : machine learning (apprentissage automatique). (Reproduit avec la permission de l'éditeur Elsevier. Auteurs : P. Ignacio Dorado-Díaz, et coll.) [46]

Tout d'abord, il s'agit de données à caractère personnel dont la divulgation ou l'absence d'anonymisation peut porter atteinte à la vie privée du/de la patient(e). Il est à noter que les données de santé sont sujettes au risque de réidentification indirecte, c'est-à-dire à l'identification de la personne par recoupement, par exemple à partir de la date à laquelle se produit un évènement sanitaire ou un acte de soins spécifique. L'anonymisation doit donc tenir compte de cela et s'appuyer sur une méthodologie robuste [789].

Veiller à ce qu'aucune atteinte ne soit portée à la vie privée des patients permet notamment de prévenir les usages délétères ou pernicieux. Les données de santé sont qualifiées de données sensibles car celles-ci peuvent informer au sujet de l'état de santé d'une personne physique [790]. En cas de fuite accidentelle, vol ou partage non autorisé à des tiers, ces informations peuvent être exploitées à des fins commerciales (ex. : systèmes de recommandation), politiques

et militaires (ex. : stratégies selon l'état de santé des dirigeants), criminelles (ex. : hameçonnage) ou par des organismes à but lucratif (ex. : discriminations par les assurances ou par les banques) [781]. Ces utilisations peuvent donc exposer les patients, et dans une plus large mesure les populations, à des risques considérables.

Un usage éthique et strictement encadré par la loi est, de fait, nécessaire à la préservation de la confiance accordée par les patients à la recherche et aux innovations telles que les intelligences artificielles. La réglementation des données de santé peut également favoriser leur qualité en imposant des normes de collecte, de stockage et de partage. Ainsi, ce cadre réglementaire est en mesure d'accroître la quantité et la qualité des données collectées, et par ce biais de promouvoir l'innovation.

Pour ces différentes raisons, les données de santé doivent faire l'objet d'une réglementation visant à assurer leur confidentialité, leur sécurité, et à circonscrire leur utilisation et leur partage.

2. Les réglementations en vigueur

Si certains états ne disposent, à l'heure actuelle, d'aucune réglementation concernant la protection des données, d'autres ont mis en place des législations plus ou moins exigeantes et applicables à l'international. Qu'en est-il des états les plus investis dans la production d'intelligences artificielles ?

Le journal *Tortoise* propose un classement positionnant 62 pays selon un indice nommé *Global AI Index*. Leur analyse se fonde sur 143 indicateurs relatifs à l'investissement, l'innovation et l'implémentation des intelligences artificielles au sein de chacun des pays inclus. Selon le *Global AI Index* (avril 2023), les 20 pays les plus avancés en matière d'intelligence artificielle sont les suivants (*tableau 1*) [791] :

Tableau 1. Classement des 20 pays les plus avancés en matière d'IA selon le *Global AI Index* (avril 2023)

Rang	Pays	Autorité de protection des données	Réglementation
1	États-Unis	Département de la Santé et des Services sociaux (<i>United States Department of Health and Human Services</i>) [792]	Loi sur la transférabilité et la responsabilité de l'assurance maladie (<i>Health Insurance Portability and Accountability Act</i>) [793]
2	Chine	Administration du cyberspace de Chine (<i>中华人民共和国国家互联网信息办公室</i>) [794]	Loi sur la protection des informations personnelles (<i>中华人民共和国个人信息保护法</i>) [795]
3	Royaume-Uni	Bureau du Commissaire à l'information (<i>Information Commissioner's Office</i>) [796]	Règlement général sur la protection des données au Royaume-Uni (<i>United Kingdom General Data Protection Regulation</i>) [797]
4	Canada	Commissariat à la protection de la vie privée du Canada (<i>Office of the Privacy Commissioner of Canada</i>) [798]	Loi sur la protection des renseignements personnels et les documents électroniques (<i>Personal Information Protection and Electronic Documents Act</i>) [799]
5	Israël	Autorité de protection de la vie privée (<i>הרשות להגנת הפרטיות</i>) [800]	Loi sur la protection de la vie privée (<i>חוק הגנת הפרטיות</i>) [801]
6	Singapour	Commission de protection des données personnelles (<i>Personal Data Protection Commission</i>) [802]	Loi sur la protection des données personnelles (<i>Personal Data Protection Act</i>) [803]
7	Corée du Sud	Commission de protection des informations personnelles (<i>국립개인정보위원회</i>) [804]	Loi sur la protection des informations personnelles (<i>개인정보보호법</i>) [805]
8	Pays-Bas	Autorité de protection des données (<i>Autoriteit Persoonsgegevens</i>) [806]	Règlement Général sur la Protection des Données [807]
9	Allemagne	Commission fédérale de la protection des données (<i>Bundesbeauftragte für den Datenschutz und die Informationsfreiheit</i>) [808]	Règlement Général sur la Protection des Données [807]
10	France	Commission nationale de l'informatique et des libertés [809]	Règlement Général sur la Protection des Données [807]
11	Australie	Bureau du commissaire australien à l'information (<i>Office of the Australian Information Commissioner</i>) [810]	Loi sur la confidentialité 1988 (<i>Privacy Act 1988</i>) [811]
12	Irlande	Commission de contrôle des données (<i>Data Protection Commission</i>) [811]	Règlement Général sur la Protection des Données [807]
13	Finlande	Office de protection des données (<i>Tietosuojavaltuutustointo</i>) [812]	Règlement Général sur la Protection des Données [807]
14	Danemark	Agence de protection des données (<i>Datatilsynet</i>) [813]	Règlement Général sur la Protection des Données [807]
15	Luxembourg	Commission nationale pour la protection des données [814]	Règlement Général sur la Protection des Données [807]
16	Japon	Commission de protection des informations personnelles (<i>個人情報保護委員会</i>) [815]	Loi sur la protection des informations personnelles (<i>個人情報保護に関する法律</i>) [816]
17	Inde		Loi sur les technologies de l'information, 2000 (<i>Information Technology Act, 2000</i>) [817]
18	Suisse	Préposé fédéral à la protection des données et à la transparence [818]	Loi fédérale sur la protection des données [819]
19	Suède	Autorité de protection des données (<i>Integritetsskyddsmyndigheten</i>) [820]	Règlement Général sur la Protection des Données [807]
20	Hong Kong	Bureau du Commissaire à la protection des données personnelles (<i>個人資料私隱專員公署</i>) [821]	Ordonnance sur les données personnelles (vie privée) (<i>個人資料(私隱)條例</i>) [822]

La réglementation la plus représentée au sein de ce classement est celle du *Règlement Général sur la Protection des Données* (RGPD), celui-ci concerne l'ensemble de l'EEE depuis le 25 mai 2018. Le RGPD s'applique à toute organisation publique ou privée, établie sur l'EEE ou ayant une activité ciblant ses résidents, dès lors que cette organisation traite des données personnelles [807]. Le RGPD doit également s'appliquer aux sous-traitants tels que les services de stockage et de traitement en ligne (*cloud*) de plus en plus communément employés dans le cadre de l'intelligence artificielle en raison des capacités de stockage et puissances de calcul que ces services offrent [823].

Le RGPD définit les données de santé comme suit : « *Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne.* ». Cette définition large englobe non seulement les données de santé par nature (ex. : antécédents médicaux, résultats d'examens), les données employées à des fins médicales, mais également les données à partir desquelles il est possible de déduire une information sur l'état de santé d'un individu par croisement (ex. : les apports caloriques d'un individu dont le poids serait connu) [824]. Vis-à-vis des données de santé, le RGPD impose un certain nombre d'exigences que nous pouvons synthétiser comme suit (*figure 124*) :

- Les finalités de la collecte, du traitement et/ou du transfert de ces données doivent être spécifiques et légitimes (ex. : but diagnostic ou thérapeutique, recherche jugée éthique, santé publique). La quantité et la nature des données collectées doivent être minimisées à ce qui est nécessaire selon l'objectif défini [825].
- Les personnes dont les données de santé ont été recueillies doivent en être informées et donner leur consentement explicite pour leur collecte, leur traitement et/ou leur transfert, excepté dans certains cas de figure spécifiques (ex. : traitements nécessaires à la sauvegarde de la vie humaine). Le consentement exprimé doit être libre, spécifique, éclairé et univoque. Les personnes concernées peuvent à tout moment revenir sur leur consentement [826].
- Les personnes concernées disposent d'un droit d'accès, de portabilité, de rectification, d'effacement, de limitation ou d'opposition au traitement de leurs données de santé [827].
- Des mesures de sécurité appropriées doivent être mises en place afin de protéger ces données contre la perte, le vol ou l'accès non autorisé. Ces mesures doivent être adaptées aux risques inhérents au traitement : contrôle d'accès des utilisateurs, traçabilité, sauvegarde et chiffrement des données, etc. De plus, les organisations sont tenues

d'informer les personnes concernées de tout incident de sécurité impliquant leurs données. [828]

- Un(e) délégué(e) à la protection des données doit être nommé(e) par chaque organisation collectant, traitant ou transférant des données. Il/Elle est chargé(e) de s'assurer que les dispositions du RGPD sont respectées par l'organisation au sein de laquelle il/elle exerce [829].
- Une analyse d'impact relative à la protection des données doit être effectuée lorsque des données de santé sont collectées, traitées et/ou transférées. Celle-ci doit comprendre une description détaillée du traitement mis en œuvre, une évaluation de la nécessité et de la proportionnalité^{xxx} concernant les principes et droits fondamentaux (volet juridique) ainsi qu'une étude des risques sur la sécurité des données et leurs impacts potentiels sur la vie privée (volet technique) [831].
- Toute organisation traitant des données doit tenir un registre recensant l'ensemble des traitements mis en œuvre. Celui-ci doit préciser le nom et les coordonnées du responsable du traitement, les finalités du traitement, les catégories d'individus concernés, les catégories de données personnelles, les catégories des parties prenantes, les transferts de données, la durée de conservation des données, ainsi qu'une description des mesures de sécurité mises en place [832].

En revanche, le RGPD n'impose pas l'anonymisation des données. L'anonymisation désigne l'action de rendre des données parfaitement anonymes, c'est-à-dire sans aucune possibilité d'identification de la personne même indirectement et ce de façon irréversible. En ce sens, l'anonymisation est à distinguer de la pseudonymisation. Le RGPD ne s'applique donc pas lorsque les données sont anonymisées, car l'anonymisation garantit par essence le respect de la vie privée des personnes concernées [833]. De même, le RGPD ne s'applique pas lorsque le traitement concerne des données de santé à l'usage exclusif de la personne concernée (ex. : stockage et utilisation de données de santé par une application en local, sans partage de ces données) [824].

^{xxx} Dans le contexte d'une analyse d'impact, la proportionnalité désigne le principe selon lequel les mesures doivent être adaptées et équilibrées en fonction de l'ampleur du projet, des avantages attendus ainsi que des coûts ou des inconvénients potentiels qu'elles pourraient engendrer. [830]

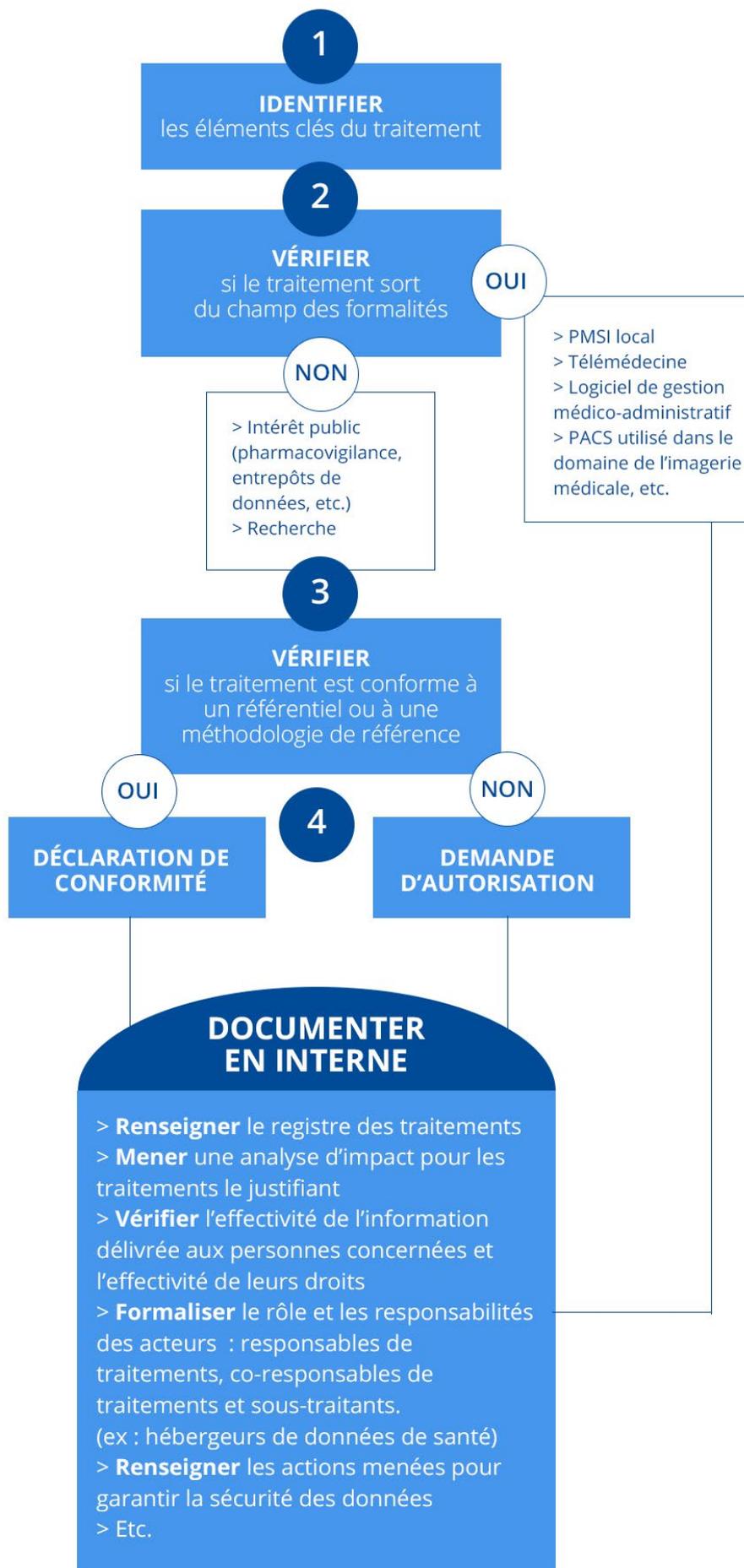


Figure 124. Synthèse des questions à se poser pour chaque nouveau traitement de données de santé (CNIL)

[826]

L'ensemble de ces exigences s'applique de la même façon à l'emploi des données de santé en intelligence artificielle, avec quelques subtilités à prendre en compte [834] :

- La phase d'apprentissage du modèle et la phase de production (déploiement du modèle) doivent être considérées séparément au regard du RGPD. En effet, un assouplissement vis-à-vis du volume et de la diversité des données peut être envisagé durant la phase d'apprentissage dans la mesure où celle-ci est d'autant plus favorisée que la quantité de données est importante. De plus, la phase d'apprentissage comporte une dimension expérimentale, la nature des données pertinentes à être exploitées peut être déterminée lors de cette phase. A l'inverse, la phase de production doit être soumise à des contraintes plus strictes quant au traitement en limitant celui-ci aux données avérées indispensables à l'issue de la phase d'apprentissage.
- Des mesures de sécurité spécifiques aux intelligences artificielles doivent être mises en œuvre, notamment contre les méthodes d'extraction d'information à partir des modèles (ex. : attaques par inférence d'appartenance, par exfiltration de modèle ou par inversion de modèle). La capacité de détecter ces attaques doit également être assurée afin de pouvoir, le cas échéant, procéder au retrait du modèle en question et notifier la violation de données aux autorités compétentes ainsi qu'aux personnes concernées.
- Un protocole d'évaluation rigoureux doit être élaboré et appliqué. Cette évaluation doit porter sur le modèle en lui-même, sur les risques de dérive de celui-ci au cours du temps (ex. : biais discriminatoires) et sur ses performances une fois déployé.

En somme, le RGPD vise à renforcer les droits des personnes, à responsabiliser les acteurs traitant des données et à crédibiliser la régulation en renforçant la coopération transnationale. Chaque pays membre de l'EEE pour lequel s'applique le RGPD dispose d'une autorité nationale de protection des données, dite « cheffe de file », chargée de s'assurer de la conformité des organisations vis-à-vis de ces exigences (*voir tableau 1*). En harmonisant la réglementation au niveau européen, le RGPD permet aux autorités nationales cheffes de file d'émettre des décisions valables dans l'ensemble de l'EEE. En pratique, si une organisation est établie dans un pays membre de l'EEE et souhaite collecter, traiter ou transférer des données concernant des résidents de l'EEE, celle-ci doit soumettre son projet à son autorité nationale cheffe de file qui doit émettre une décision quant à l'autorisation du projet. Dans le cas où le projet est transnational, l'autorité cheffe de file référente se concerta avec l'ensemble des autorités cheffe de file des pays concernés avant d'émettre sa décision. Ce système de guichet unique simplifie les démarches des organisations qui, lorsque leur projet concerne un ou plusieurs pays membres de l'EEE, peuvent désormais s'adresser directement à l'autorité cheffe

de file de leur pays. De même, les personnes concernées peuvent soumettre leurs plaintes à l'autorité cheffe de file de leur pays même si celles-ci concernent une organisation établie dans un pays étranger, au sein de l'EEE ou non car le RGPD s'applique également aux organisations présentes hors de l'EEE souhaitant collecter, traiter ou transférer des données concernant des résidents de l'EEE [835].

De fait, le RGPD est perçu comme étant un modèle en tant que réglementation internationale de protection des données à caractère personnel. Les organisations de la société civile et les experts en protection des données ont notamment salué son approche globale et ses exigences élevées en matière de transparence, de responsabilité et de sécurité des données [836]. Le RGPD fait d'ailleurs office de référence pour de nombreux pays qui s'en sont inspirés pour améliorer leur réglementation nationale (ex. : Qatar, Brésil, Nouvelle-Zélande) [837].

En se plaçant du point de vue du RGPD, la Commission européenne considère six niveaux de protection des données distinguant les différents pays du monde (*figure 125*) [838] :

- Pays membres de l'EEE, c'est-à-dire les 27 États membres de l'Union européenne auxquels s'ajoutent la Norvège, le Liechtenstein et l'Islande. Il s'agit des pays pour lesquels la protection des données est encadrée par le RGPD.
- Pays reconnus comme adéquats par la Commission européenne, tels que le Royaume-Uni, Israël, la Corée du Sud, le Japon ou encore la Suisse. Il s'agit des pays tiers (hors EEE) dont la législation ou/et les engagements internationaux offrent un niveau de protection des données comparable à celui du RGPD. Les transferts de données à caractère personnel entre les pays adéquats et les pays membres de l'EEE peuvent être effectués sans nécessiter d'autorisations ou d'encadrements supplémentaires [839,840].
- Pays en adéquation partielle, tels que le Canada. Il s'agit des pays tiers reconnus comme adéquats par la Commission européenne uniquement pour certains traitements spécifiques (ex. : les traitements réalisés dans le cadre d'activités commerciales entre le Canada et l'EEE ne nécessitent pas d'encadrement spécifique). Les traitements ou transferts sortant de ce périmètre spécifique doivent faire l'objet d'autorisations ou d'encadrements supplémentaires lorsqu'ils sont opérés au sein d'un pays en adéquation partielle et qu'ils concernent des données à caractère personnel provenant de l'EEE.
- Pays disposant d'une autorité indépendante et de lois, tels que les États-Unis, l'Australie ou Hong Kong. Il s'agit des pays tiers non reconnus comme adéquats par la Commission européenne mais disposant néanmoins d'une législation nationale en matière de protection des données personnelles et d'une autorité de protection des données

reconnues par la *Global Privacy Assembly*. Les traitements ou transferts de données à caractère personnel provenant de l'EEE doivent faire l'objet d'autorisations ou d'encadrements supplémentaires lorsqu'ils sont opérés au sein de ces pays.

- Pays disposant d'une législation, tels que la Chine, Singapour ou l'Inde. Il s'agit des pays tiers non reconnus comme adéquats par la Commission européenne mais disposant néanmoins d'une législation générale sur la protection des données personnelles ou de dispositions spécifiques. Les traitements ou transferts de données à caractère personnel provenant de l'EEE doivent faire l'objet d'autorisations ou d'encadrements supplémentaires lorsqu'ils sont opérés au sein de ces pays.
- Pays ne disposant d'aucune législation, tels que la Lybie, le Venezuela ou le Cambodge. Il s'agit des pays tiers non reconnus comme adéquats par la Commission européenne et ne disposant pas d'une législation ou de dispositions spécifiques concernant la protection des données personnelles. Les traitements ou transferts de données à caractère personnel provenant de l'EEE doivent faire l'objet d'autorisations ou d'encadrements spécifiques lorsqu'ils sont opérés au sein de ces pays.

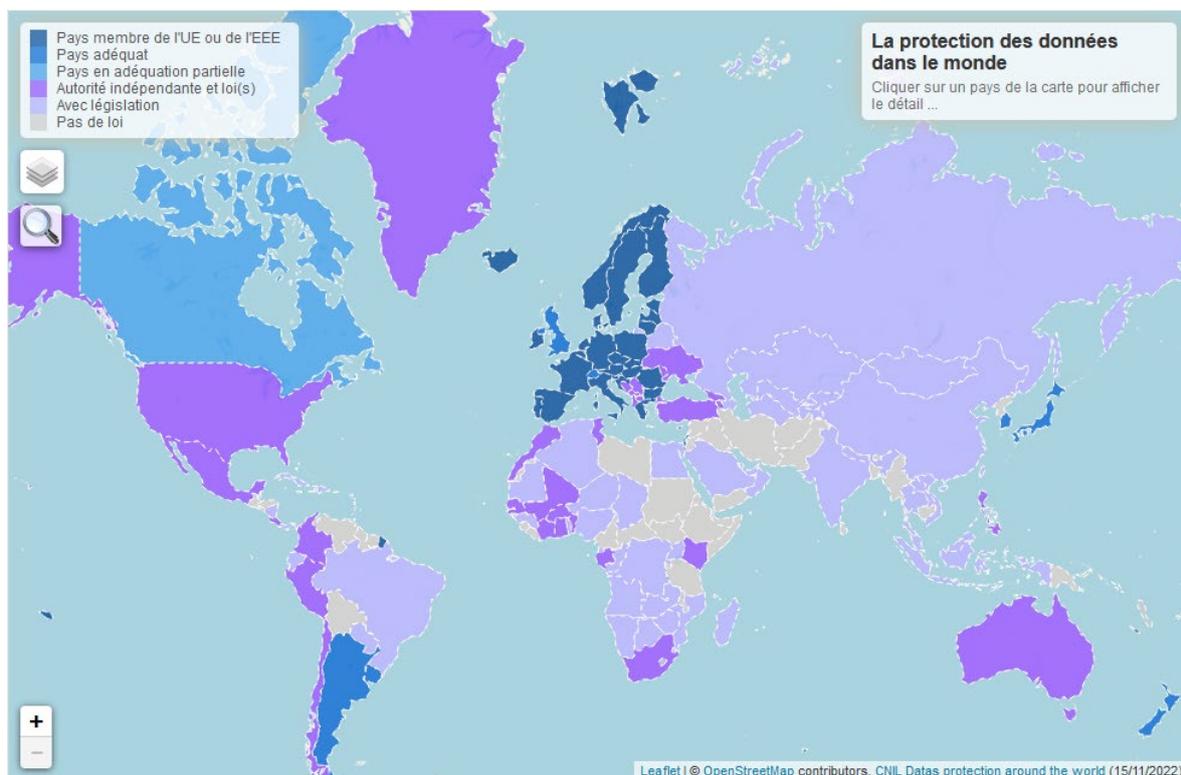


Figure 125. La protection des données dans le monde (CNIL) [838]

Le modèle international de réglementation quant à la protection des données personnelles, dont la forme la plus aboutie actuellement est celle du RGPD, est l'une des pierres angulaires de l'intelligence artificielle éthique. En favorisant la coopération et le transit de données, cette approche réglementaire permet d'élargir leur collecte à une population plus large en termes de

quantité mais également de représentativité. Par extension, cela est susceptible d'accroître la qualité des modèles produits en réduisant notamment les biais discriminatoires. De surcroît, des standards juridiques élevés en matière de transparence, sécurité et responsabilité assurent que cette performance n'est pas obtenue au détriment de la vie privée des personnes concernées, de leur droit à l'information ou de leur consentement. Néanmoins, au vu de l'absence de réglementation vis-à-vis de la protection des données personnelles dans de nombreux pays ainsi que des discordances entre certaines réglementations (ex. : article 48 du RGPD et *CLOUD Act*), des défis persistent [841]. Les efforts de globalisation et de conciliation des réglementations sont donc à poursuivre.

II. Évaluation des modèles

L'évaluation occupe une place prépondérante durant le cycle de vie d'un modèle d'intelligence artificielle. Il est capital de soigner l'évaluation car la qualité de son exécution conditionne l'efficacité de l'intelligence artificielle produite. En s'appuyant sur des métriques adaptées à la finalité, évaluer les modèles permet d'ajuster au mieux les conditions d'apprentissage et ainsi d'optimiser les performances des modèles retenus en dernier lieu. Cette démarche peut se décomposer en trois étapes inter-communicantes : choix et calcul des métriques, ajustement des conditions d'apprentissage, sélection du ou des modèles à retenir [46]. Au cours de ce volet, nous traiterons ces trois étapes en nous focalisant sur le cas des modèles obtenus à l'issue d'un apprentissage supervisé car il s'agit du type d'apprentissage le plus largement employé en diagnostic et en thérapeutique.

1. Qu'est-ce qu'un modèle idéal ?

Il convient tout d'abord de définir ce qui est considéré comme étant un bon ou un mauvais modèle. La qualité d'un modèle est communément définie par son ajustement vis-à-vis de ses données d'entraînement. Schématiquement, un modèle peut être caractérisé par trois états que nous allons développer : le sous-apprentissage, le sur-apprentissage ou l'ajustement optimal [842].

a. Le sous-apprentissage

Employer le terme « sous-apprentissage » (ou sous-ajustement) pour désigner un modèle prédictif signifie que celui-ci s'ajuste insuffisamment aux données d'entraînement (*figure 126*).

Autrement dit, le modèle en question peine à identifier des corrélations entre les variables prédictives d'entrée (X) et la variable cible à prédire (Y). Cela a pour conséquence d'affecter la capacité du modèle à généraliser et donc à fournir des prédictions justes, que ce soit à partir de données qu'il connaît ou de données auxquelles il n'a pas été exposé durant son entraînement. Plus un modèle sous-apprend, plus ses prédictions sont susceptibles d'être biaisées (erreurs systématiques et indépendantes du nombre d'observations) [302].

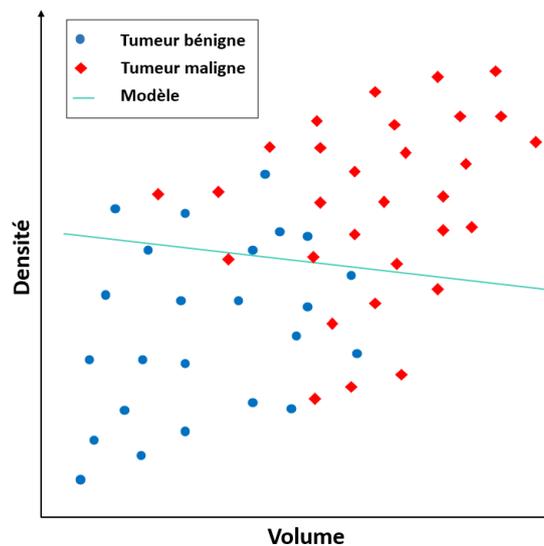


Figure 126. Exemple de modèle de classification ayant sous-appris

Le sous-apprentissage s'évalue avec la qualité des prédictions du modèle vis-à-vis de ses données d'entraînement. En général, un modèle est considéré comme ayant sous-appris lorsque celui-ci est peu performant pour prédire la variable cible de ses données d'entraînement. La notion de sous-apprentissage est cependant très subjective, les performances mesurées doivent être comparées à une valeur de référence (ex. : modèles similaires préexistants ou performances atteintes par un test diagnostic) [843].

Les causes possibles du sous-apprentissage sont les suivantes [843] :

- Le modèle est trop simple et considère excessivement les déviations comme étant du bruit ;
- Les variables d'entraînement ne sont pas suffisamment nombreuses ;
- La qualité des données d'entraînement n'est pas suffisante (ex. : attrition ; sous-effectifs trop faibles).

Plusieurs solutions peuvent être envisagées pour atténuer le sous-apprentissage [843] :

- Entraîner davantage le modèle ;

- Utiliser un algorithme d'apprentissage et/ou des hyperparamètres à même de produire un modèle plus complexe (ex. : réduction de la régularisation) ;
- Augmenter le nombre de variables d'entraînement ;
- Améliorer la qualité des données d'entraînement.

Remarque : la complexité d'un modèle fait référence aux paramètres considérés par le modèle pour effectuer sa prédiction. Plus les paramètres d'un modèle sont nombreux et étendus, plus celui-ci est jugé complexe [844]. Les paramètres d'un modèle ne sont pas à confondre avec les hyperparamètres. Les seconds concernent les algorithmes d'apprentissage et peuvent être sélectionnés avant l'entraînement (ex. : le nombre de voisins pour un algorithme de k-NN), tandis que les premiers sont intrinsèques au modèle en lui-même et se définissent au cours de l'entraînement. Nous pourrions presque affirmer que les paramètres d'un modèle sont le modèle. Ils désignent par exemple les coefficients caractérisant une régression linéaire ou encore les poids d'un réseau de neurones [845].

b. Le sur-apprentissage

Employer le terme « sur-apprentissage » (ou sur-ajustement) pour désigner un modèle prédictif signifie que celui-ci s'ajuste excessivement aux données d'entraînement (*figure 127*). Autrement dit, le modèle prend en compte à la fois les corrélations généralisables mais également les fluctuations aléatoires que comportent les données d'entraînement. Cela a pour conséquence d'affecter la capacité du modèle à généraliser et donc à fournir des prédictions précises à partir de données auxquelles il n'a pas été exposé durant son entraînement. Plus un modèle sur-apprend, plus ses prédictions sont susceptibles d'être sujettes à une forte variance (erreurs aléatoires et d'autant plus importantes que le nombre d'observations est faible) [302].

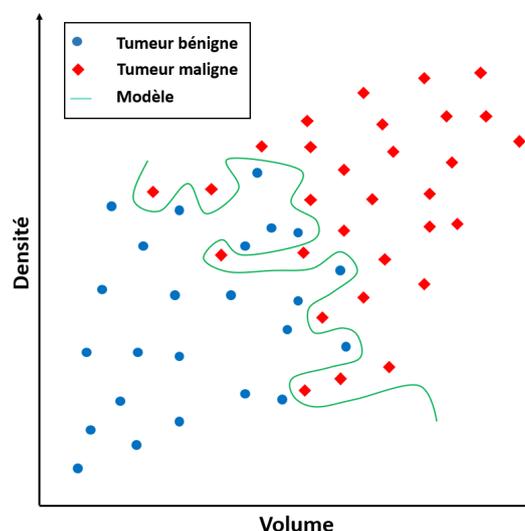


Figure 127. Exemple de modèle de classification ayant sur-appris

Le sur-apprentissage est une notion beaucoup plus objective. Celui-ci s'évalue à la différence de qualité selon que les prédictions du modèle concernent des données de validation (que le modèle ne connaît pas) ou ses données d'entraînement. En pratique, un modèle est considéré comme ayant sur-appris lorsque ses performances de prédiction avec ses données d'entraînement sont éloignées de celles obtenues avec des données de validation [843].

Les causes possibles du sur-apprentissage sont les suivantes [843] :

- Le modèle est trop complexe (nombre trop important de paramètres) et n'ignore pas suffisamment les déviations ;
- Les données d'entraînement comportent trop de bruit ;
- Les variables d'entraînement sont trop nombreuses ;
- La quantité des données d'entraînement est insuffisante ;
- La qualité des données d'entraînement est insuffisante (ex. : attrition ; sous-effectifs trop faibles).

Plusieurs solutions peuvent être envisagées pour atténuer le sur-apprentissage [843] :

- Ecourter l'entraînement du modèle (*early stopping*) ;
- Utiliser un algorithme d'apprentissage et/ou des hyperparamètres à même de produire un modèle moins complexe (ex. : régularisation, *pruning*) ;
- Diminuer le nombre de variables d'entraînement ;
- Augmenter la quantité des données d'entraînement par le nombre d'observations ou artificiellement (ex. : augmentation de données) ;
- Améliorer la qualité des données d'entraînement ;
- Adopter une approche ensembliste (ex. : *bagging*, *boosting*).

c. L'ajustement optimal

Au cours d'un apprentissage automatique, les paramètres du modèle entraîné s'ajustent petit à petit aux données d'entraînement. Durant ce processus, le biais des prédictions du modèle diminue tandis que leur variance augmente. Cela a pour effet, dans un premier temps, de diminuer l'erreur du modèle à la fois vis-à-vis des données d'entraînement mais aussi de données inconnues du modèle. Jusque-là, le modèle est qualifié de sous-ajusté (*figure 128*) [315].

Dans un second temps, après le point de complexité optimale, l'erreur relative aux données d'entraînement poursuit sa baisse tandis que l'erreur relative à des données inconnues entame sa croissance. A partir de cette inflexion, plus le modèle sera entraîné plus celui-ci sera sur-ajusté (*figure 128*) [315].

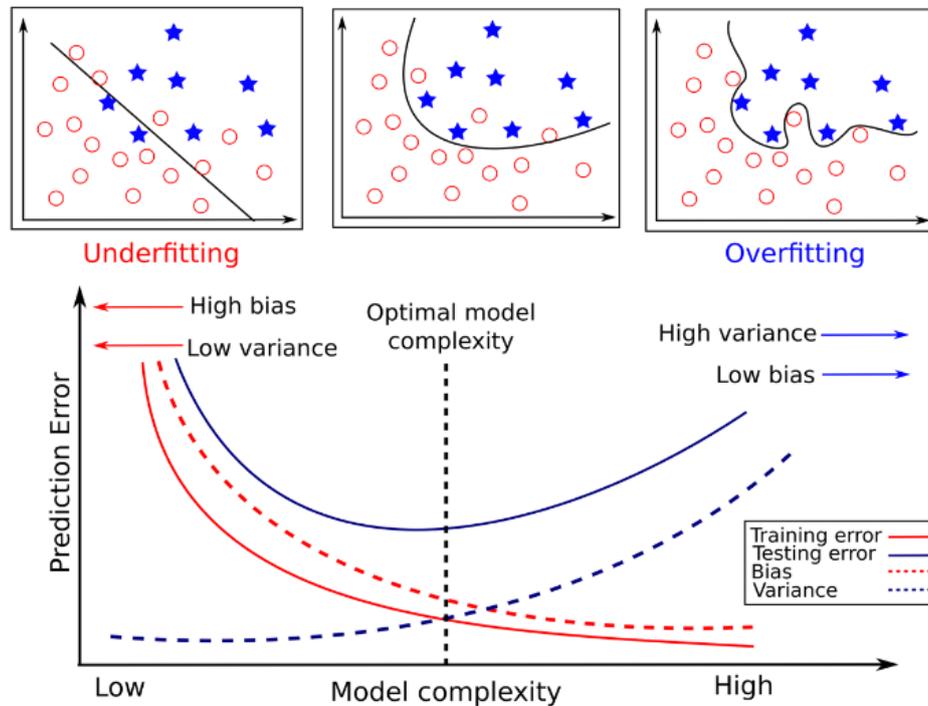


Figure 128. Illustration de l'évolution de l'ajustement au cours de l'entraînement d'un modèle (*Reproduit avec la permission de l'éditeur Springer Nature. Auteur : Alaa Tharwat*) [315]

Entre les deux états décrits ci-dessus se trouve l'ajustement optimal (*figure 128*). Celui-ci correspond à un modèle ne souffrant ni de sous-apprentissage ni de sur-apprentissage, dont le biais et la variance sont donc minimisés. Le challenge majeur en apprentissage automatique est de parvenir à générer des modèles se rapprochant le plus possible de cet état d'ajustement optimal [315].

2. Métriques usuellement employées en apprentissage supervisé

La démarche d'optimisation de l'ajustement ne peut être menée sans mesurer l'erreur du modèle entraîné et cette mesure de l'erreur repose sur le calcul de métriques. Les métriques à employer diffèrent selon notamment qu'il s'agisse d'un problème de classification ou de régression, ou selon la finalité du modèle.

a. Métriques employées en classification

Les métriques les plus couramment employées pour évaluer un modèle de classification sont les suivantes :

- L'exactitude (ou *accuracy*), qui désigne le taux de prédictions correctes parmi l'ensemble des prédictions effectuées par le modèle. L'exactitude se calcule comme ceci [618,662] :

$$\text{Exactitude} = \frac{\text{Vrais positifs} + \text{Vrais négatifs}}{\text{Vrais positifs} + \text{Vrais négatifs} + \text{Faux positifs} + \text{Faux négatifs}}$$

Celle-ci est d'autant plus informative que les différentes classes sont équilibrées.

- La sensibilité (ou *recall*), qui désigne le taux d'identifications positives correctement prédites par le modèle parmi l'ensemble des cas réellement positifs. D'un point de vue médical, il s'agit de la probabilité qu'un résultat de test soit positif lorsque la pathologie recherchée est réellement présente chez le/la patient(e). Dans le cadre d'une classification binaire, la sensibilité se calcule comme ceci [558,618,662] :

$$\text{Sensibilité} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}}$$

En pratique, il est pertinent de privilégier la sensibilité lorsque les coûts associés aux faux négatifs sont très élevés par rapport aux coûts associés aux faux positifs (ex. : dépistage de masse d'une pathologie infectieuse ayant un taux de reproduction important) [846].

- La valeur prédictive positive (ou *precision*), qui désigne le taux d'identifications positives correctes parmi l'ensemble des identifications positives effectuées par le modèle. D'un point de vue médical, il s'agit de la probabilité qu'un résultat de test positif indique la présence réelle chez le/la patient(e) de la pathologie recherchée. Dans le cadre d'une classification binaire, la valeur prédictive positive se calcule comme ceci [558,618,662] :

$$\text{Valeur préd.pos.} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}} = \frac{\text{Sensibilité} \times \text{Prévalence}}{\text{Sensibilité} \times \text{Prévalence} + (1 - \text{Spécificité})(1 - \text{Prévalence})}$$

Sa valeur dépend non seulement de la qualité du test évalué mais également de la prévalence de la pathologie recherchée.

- Le *F-score* (ou *F-measure*), qui correspond à la moyenne harmonique de la sensibilité et de la valeur prédictive positive [386]. Celle-ci permet d'exprimer une synthèse de ces deux métriques sous la forme d'un rapport moyen pouvant éventuellement être pondéré en faveur de la sensibilité ou de la valeur prédictive positive. Dans le cadre d'une

classification binaire, le calcul du *F-score* se décline en deux formules différentes selon qu'une pondération (β) soit appliquée ou non [618,662] :

$$F1 - score = \frac{2 \times \text{Vrais positifs}}{2 \times \text{Vrais positifs} + \text{Faux positifs} + \text{Faux négatifs}} = 2 \times \frac{\text{sensibilité} \times \text{valeur préd. pos.}}{\text{sensibilité} + \text{valeur préd. pos.}}$$

$$F\beta - score = \frac{(1 + \beta^2) \times \text{Vrais positifs}}{(1 + \beta^2) \times \text{Vrais positifs} + \text{Faux positifs} + \beta^2 \times \text{Faux négatifs}} = (1 + \beta^2) \times \frac{\text{sensibilité} \times \text{valeur préd. pos.}}{\text{sensibilité} + (\beta^2 \times \text{valeur préd. pos.})}$$

En utilisant une valeur x pour β , cela attribue x fois plus de poids à la sensibilité au détriment de la valeur prédictive positive. L'emploi des moyennes harmoniques, telles que le *F-score*, est très répandu en apprentissage automatique.

Lorsque l'évaluation porte plus spécifiquement sur des modèles à vocation diagnostique ou thérapeutique, il est recommandé de faire intervenir les métriques suivantes [46] :

- La spécificité, qui désigne le taux d'identifications négatives correctement prédites par le modèle parmi l'ensemble des cas réellement négatifs. D'un point de vue médical, il s'agit de la probabilité qu'un résultat de test soit négatif lorsque la pathologie recherchée est réellement absente chez le/la patient(e). Dans le cadre d'une classification binaire, la spécificité se calcule comme ceci [558,618,662] :

$$\text{Spécificité} = \frac{\text{Vrais négatifs}}{\text{Vrais négatifs} + \text{Faux positifs}}$$

L'emploi de la spécificité comme métrique d'évaluation est très répandu en recherche clinique, en association avec la sensibilité. En pratique, il est pertinent de privilégier la spécificité lorsque les coûts associés aux faux positifs sont très élevés par rapport aux coûts associés aux faux négatifs (ex. : confirmation d'un diagnostic incertain) [846].

- La statistique C, qui désigne la capacité d'un modèle à distinguer différentes classes. Dans le cadre d'une classification binaire, celle-ci correspond à l'aire sous la courbe de ROC (*Receiver Operating Characteristic*) et se calcule à partir de l'intégrale de la fonction ROC ayant pour abscisse le taux de faux positifs ($1 - \text{spécificité}$) et pour ordonnée le taux de vrais positifs (sensibilité) [662]. L'aire sous la courbe de ROC permet d'évaluer l'aptitude d'un test diagnostique à distinguer les individus malades des individus sains, en tenant compte à la fois de la sensibilité et de la spécificité. En pratique, plus l'aire sous la courbe de ROC est proche de 1, plus le test évalué est jugé performant (*figure 129*) [386].

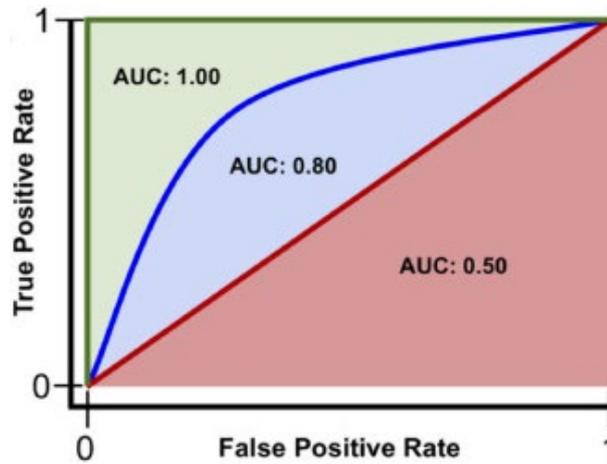


Figure 129. Exemples de courbe de ROC associées à leur aire sous la courbe. Les courbes de ROC se trouvant entre la droite rouge et le demi-carré vert correspondent aux modèles dont les performances de prédiction sont meilleures que le hasard. (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Rene Y. Choi et coll.) [45]

- L'indice de Youden (ou *Youden's J statistic*), qui réunit la sensibilité et la spécificité dans son calcul : $J = \text{sensibilité} + \text{spécificité} - 1$. En se référant à la représentation graphique de la courbe de ROC, la diagonale passant par l'origine représente les tests n'ayant aucune valeur diagnostique. A l'inverse, le point de la courbe de ROC (valeur seuil) le plus éloigné de cette droite correspond au test caractérisé par des valeurs optimales de sensibilité et spécificité [846]. Il est possible d'identifier cette valeur seuil optimale par le calcul des indices de Youden correspondant à chaque valeur seuil, la valeur seuil optimale est celle qui est associée à l'indice de Youden le plus élevé (*figure 130*) [847].

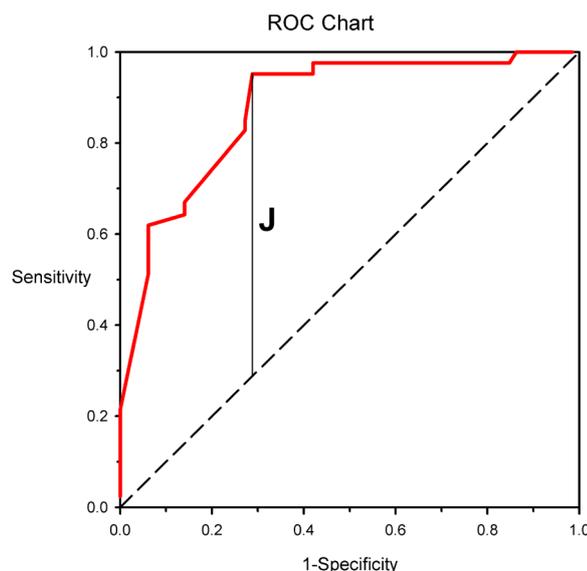


Figure 130. Exemple d'indice de Youden situé sur une courbe de ROC (Reproduit selon les termes de la licence CC BY-SA 4.0, Auteur : Kognos. https://commons.wikimedia.org/wiki/File:ROC_Curve_Youden_J.png)

- La valeur prédictive négative, qui désigne le taux d'identifications négatives correctes parmi l'ensemble des identifications négatives effectuées par le modèle. D'un point de vue médical, il s'agit de la probabilité qu'un résultat de test négatif indique l'absence réelle chez le/la patient(e) de la pathologie recherchée. Dans le cadre d'une classification binaire, la valeur prédictive négative se calcule comme ceci [662] :

$$\text{Valeur préd. nég.} = \frac{\text{Vrais négatifs}}{\text{Vrais négatifs} + \text{Faux négatifs}} = \frac{\text{Spécificité} \times (1 - \text{Prévalence})}{(\text{Spécificité} (1 - \text{Prévalence}) + (1 - \text{Sensibilité}) \times \text{Prévalence}}$$

Tout comme la valeur prédictive positive, sa valeur dépend non seulement de la qualité du test évalué mais également de la prévalence de la pathologie recherchée.

- Les rapports de vraisemblance, qui désignent le rapport entre la probabilité d'avoir un test positif (ou négatif) chez les sujets malades et celle d'avoir un test positif (ou négatif) chez les sujets sains [846]. Les rapports de vraisemblance permettent d'exprimer la valeur diagnostique d'un test médical en considérant à la fois la sensibilité et la spécificité. Toutefois, contrairement aux valeurs prédictives, les rapports de vraisemblance sont indépendants de la prévalence. Ils se calculent comme ceci [662,848] :

$$\text{Rapport de vraisemblance positif} = \frac{\text{Sensibilité}}{1 - \text{Spécificité}}$$

$$\text{Rapport de vraisemblance négatif} = \frac{1 - \text{Sensibilité}}{\text{Spécificité}}$$

En pratique, un test diagnostique est considéré comme fiable lorsque son rapport de vraisemblance positif est supérieur à 10 et/ou que son rapport de vraisemblance négatif est inférieur à 0,1 [849].

Nous pourrions également considérer l'indice de concordance (ou *Harrell's C-index*) ou encore le Kappa de Cohen, qui permettent de mesurer la capacité discriminante d'un modèle de classification [850,851].

Les 11 métriques que nous avons citées peuvent être employées seules mais il est recommandé de baser l'évaluation d'un modèle sur plusieurs d'entre-elles indépendamment ou/et réunies sous la forme de moyennes harmoniques [852]. Classiquement, la sensibilité et la spécificité sont souvent considérées conjointement pour évaluer la qualité d'un test diagnostic (*figure 131*). Il est à noter que la sensibilité et la spécificité varient en sens opposé pour un test donné, l'un augmente lorsque l'autre diminue [853]. En théorie, un test serait parfait si celui-ci était caractérisé par une sensibilité et une spécificité toutes deux égales à 1, mais cela est en pratique rare ou impossible selon les cas [846]. Dans l'idéal, il convient d'optimiser simultanément la sensibilité et la spécificité, en privilégiant éventuellement l'une des deux métriques si cela est justifié [854].

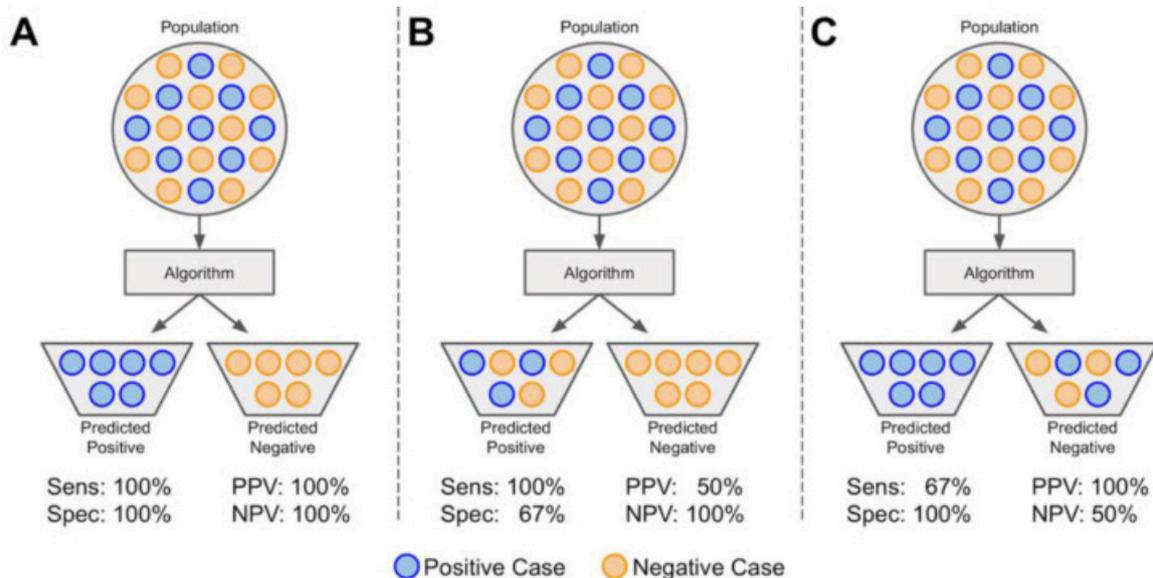


Figure 131. Exemple schématique d'évaluation d'un modèle de classification binaire. Trois cas de figure sont illustrés. (A) Le modèle a correctement classifié l'ensemble des observations. Les quatre métriques calculées s'élèvent à 100%. (B) Le modèle a correctement classifié l'ensemble des cas positifs. En revanche, certains cas négatifs ont été classifiés comme positifs. Les calculs de métriques indiquent donc une sensibilité plus élevée que la spécificité. (C) Le modèle a correctement classifié l'ensemble des cas négatifs. En revanche, certains cas positifs ont été classifiés comme négatifs. Les calculs de métriques indiquent donc une spécificité plus élevée que la sensibilité. (Reproduit selon les termes de la licence CC BY-NC-ND 4.0. Auteurs : Rene Y. Choi et coll.) [45]

En outre, précisons que l'applicabilité de ces métriques ne se limite pas aux classifications binaires, celles-ci peuvent également être mobilisées dans le cadre d'une classification multiclassée. Pour ce faire, il s'agit de calculer la métrique pour chaque classe par rapport à l'ensemble des autres classes regroupées, puis de calculer une moyenne des métriques calculées pour chacune des classes. Le calcul de la moyenne peut éventuellement être pondéré par la fréquence des classes [842].

Notons enfin que l'évaluation d'un modèle de segmentation (classification des pixels ou groupe de pixels) peut également s'appuyer sur ces métriques, telles quelles ou sous des formes déclinées (ex. : indice de Sørensen-Dice, indice de Jaccard) [855].

b. Métriques employées en régression

Du fait que leurs prédictions concernent une variable quantitative (ex. : score clinique, dose de médicament à administrer), les modèles de régression doivent être évalués à partir de métriques spécifiques à ce type de variables :

- L'erreur quadratique moyenne (ou MSE, *mean squared error*), qui désigne la moyenne arithmétique des carrés des écarts entre les valeurs réelles observées (y_i) et les valeurs

prédites par le modèle (\hat{y}_i). L'erreur quadratique moyenne se calcule comme ceci (n étant le nombre d'observations) [618] :

$$MSE = \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

L'une des déclinaisons de l'erreur quadratique moyenne est sa racine carrée (ou RMSE, *root mean squared error*) qui présente l'avantage de pouvoir être exprimée dans la même unité que la variable à prédire, facilitant ainsi son interprétation. L'erreur quadratique moyenne permet de quantifier les erreurs réalisées par le modèle, avec pour principal avantage d'être dérivable, facilement optimisable et donc de pouvoir être utilisée en tant que fonction d'erreur. En revanche, puisque son calcul comporte une mise au carré des erreurs, l'erreur quadratique moyenne est d'autant plus forte que les erreurs sont grandes. Cela rend cette métrique très sensible aux valeurs aberrantes, ce qui constitue son inconvénient majeur. En pratique, la minimisation de l'erreur quadratique moyenne favorise les performances du modèle. Remarque : la MSE et la RMSE sont respectivement assimilables à la variance et à l'écart-type, celles-ci seraient identiques pour un modèle dont toutes les valeurs prédites seraient égales à la moyenne.

- L'erreur absolue moyenne (ou MAE, *mean absolute error*), qui désigne la moyenne arithmétique des valeurs absolues des écarts entre les valeurs réelles observées (y_i) et les valeurs prédites par le modèle (\hat{y}_i). L'erreur absolue moyenne se calcule comme ceci (n étant le nombre d'observations) [856] :

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n}$$

Celle-ci permet de quantifier les erreurs réalisées par le modèle et s'exprime dans la même unité que la variable à prédire, ce qui en fait une métrique simple à interpréter. A l'instar de l'erreur quadratique moyenne, l'erreur absolue moyenne est très sensible aux valeurs aberrantes. En pratique, la minimisation de l'erreur absolue moyenne favorise les performances du modèle.

- L'erreur absolue moyenne en pourcentage (ou MAPE, *mean absolute percentage error*), qui désigne le pourcentage moyen d'écart entre les valeurs réelles observées (y_i) et les valeurs prédites par le modèle (\hat{y}_i). L'erreur absolue moyenne en pourcentage se calcule comme ceci (n étant le nombre d'observations) [857] :

$$MAPE = \frac{\sum \frac{|y_i - \hat{y}_i|}{|y_i|}}{n}$$

Celle-ci permet de quantifier les erreurs réalisées par le modèle et s'exprime en pourcentage, ce qui en fait une métrique simple à interpréter. Cette déclinaison de

l'erreur absolue moyenne présente l'avantage d'être moins sensible à l'échelle des valeurs et aux valeurs aberrantes, excepté lorsque les valeurs sont proches de 0. En outre, l'erreur absolue moyenne en pourcentage se prête moins à l'optimisation car elle n'est pas dérivable en 0. En pratique, la minimisation de l'erreur absolue moyenne en pourcentage favorise les performances du modèle.

- Le coefficient de détermination, noté R^2 , qui désigne le rapport de la variance expliquée par la régression (SSE) sur la variance totale (SST)^{xxxii}. Il se calcule comme ceci [618,858] :

$$R^2 = \frac{SSE}{SST} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST}$$

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Le coefficient de détermination permet de mesurer la qualité d'ajustement d'un modèle vis-à-vis de ses données d'entraînement, et par extension sa capacité à effectuer des prédictions justes. Cependant, il s'agit d'une métrique délicate à interpréter car celle-ci augmente mécaniquement avec le nombre de variables d'entraînement. Une utilisation pertinente du coefficient de détermination serait, par exemple, de comparer des modèles ayant le même nombre de variables prédictives. Une autre alternative consiste à calculer le coefficient de détermination ajusté [858]. En pratique, le modèle évalué est jugé d'autant mieux ajusté que son coefficient de détermination est proche de 1.

Nous pouvons également citer les métriques suivantes qui peuvent faire office d'alternatives et/ou s'avérer plus adaptées à certains types de données d'entraînement :

- La variance expliquée [859] ;
- L'erreur maximale [860] ;
- L'écart médian absolu (ou MedAE, *median absolute error*) [861] ;
- L'erreur logarithmique quadratique moyenne (ou MSLE, *mean squared logarithmic error*) [862] ;
- La racine carrée de l'erreur logarithmique quadratique moyenne (ou RMSLE, *root mean squared logarithmic error*) [862] ;
- Le critère d'information d'Akaike (ou AIC, *Akaike information criterion*) [863] ;
- Le critère d'information bayésien (ou BIC, *bayesian information criterion*) [864].

^{xxxii} SSE : *sum of squares error* (variance expliquée par la régression) ; SST : *sum of squares total* (variance totale) ; SSR : *sum of squared residuals* (variance résiduelle).

Tout comme en classification, les 12 métriques que nous avons mentionnées ci-dessus peuvent être employées seules mais il est recommandé de les considérer concurremment [852].

3. Validation des modèles

Une fois que les données ont été préparées, que les métriques ont été désignées, que l'algorithme d'apprentissage et ses hyperparamètres ont été sélectionnés, l'entraînement peut avoir lieu. Seulement, il est rare d'obtenir un modèle idéal à l'issue de la première tentative. C'est en réponse à cette problématique qu'intervient l'étape de validation.

a. Le principe de la validation

La finalité de la phase de validation est de tendre autant que possible vers un modèle idéal. Il s'agit en pratique d'ajuster au mieux les conditions d'apprentissage (hyperparamètres, algorithme d'apprentissage, quantité et qualité des données, etc.) afin que le modèle obtenu soit le plus proche possible de l'ajustement optimal [865].

La validation repose sur une boucle dont la première étape consiste à mesurer l'erreur des prédictions effectuées par le modèle entraîné. Cette mesure de l'erreur doit être effectuée sur les données d'entraînement, dont le modèle s'est servi pour apprendre, ainsi que sur des données de validation auxquelles le modèle n'a jamais été exposé. L'intérêt d'effectuer cette mesure à partir de ces deux jeux de données distincts est de pouvoir constater à fois le degré de sous-apprentissage et celui de sur-apprentissage. La boucle se poursuit par une tentative d'ajustement des conditions d'apprentissage (ou *fine-tuning*) selon les résultats obtenus. Le modèle est ensuite entraîné cette fois-ci avec les conditions d'apprentissage ajustées, puis évalué à nouveau à partir des données d'entraînement et des données de validation, et ainsi de suite jusqu'à obtenir le ou les modèles jugés optimaux (*figure 132*). Concrètement, ce qui est recherché est l'amélioration des métriques (minimiser le sous-apprentissage) ainsi qu'un écart minimal de ces métriques entre les données d'entraînement et les données de validation (minimiser le sur-apprentissage) [865]. Il est à noter qu'il existe des outils permettant d'automatiser ce processus (ex. : *GridSearchCV*, *RandomizedSearchCV*) [866,867].

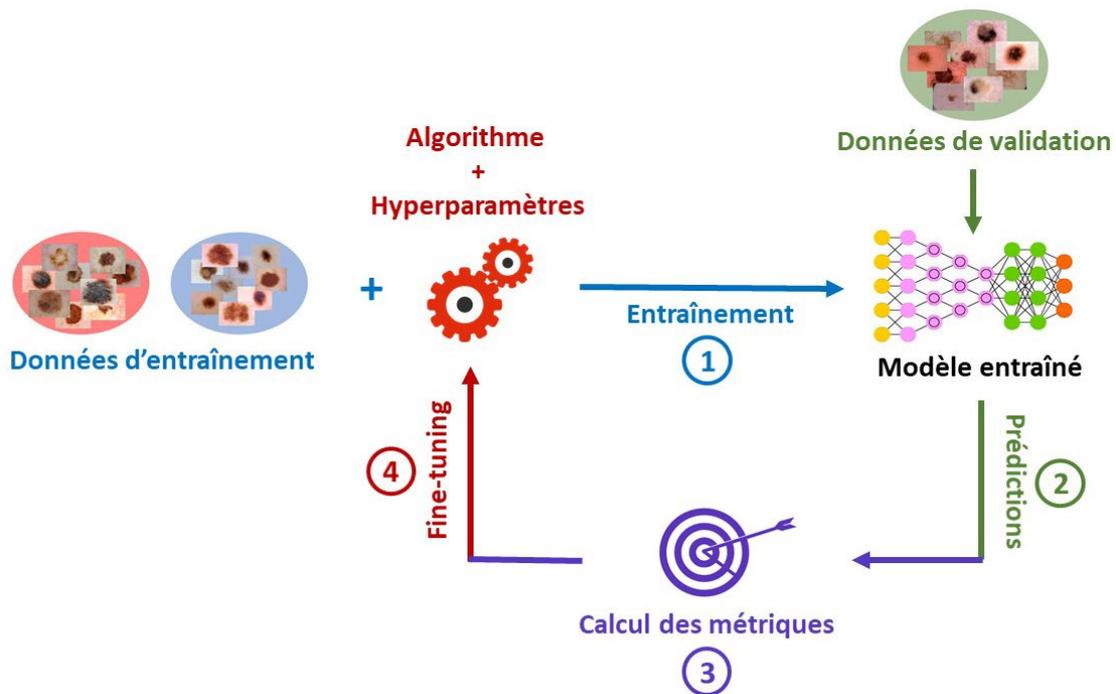


Figure 132. Illustration schématique du processus de validation d'un modèle d'apprentissage automatique

b. Les méthodes d'échantillonnage

Il est tout à fait possible de réaliser ce processus à partir de jeux de données d'entraînement et de validation fixes, cependant cela peut fortement exposer le modèle au biais de sélection. Ce dernier peut être important lorsque le fractionnement des données (entraînement-validation) n'est pas réalisé de façon aléatoire ou encore lorsque certaines classes sont sur-représentées ou sous-représentées [868]. Il existe différentes méthodes d'échantillonnage en mesure d'atténuer ce biais [869].

Parmi ces méthodes, la plus populaire est la validation croisée. Celle-ci se décline sous plusieurs formes :

- La validation croisée à k blocs (ou *k-fold cross-validation*), qui consiste à diviser les données d'entraînement en k sous-échantillons de tailles similaires, à entraîner le modèle sur $k-1$ sous-échantillons puis à l'évaluer à partir du sous-échantillon restant (*figure 133*). Ce processus est itéré k fois en interchangeant le rôle des sous-échantillons de façon à ce que chacun de ces sous-ensembles endosse à la fois le rôle de données d'entraînement et de données de validation. La performance du modèle est alors évaluée en calculant la moyenne des métriques obtenues sur chacune des k itérations [865].

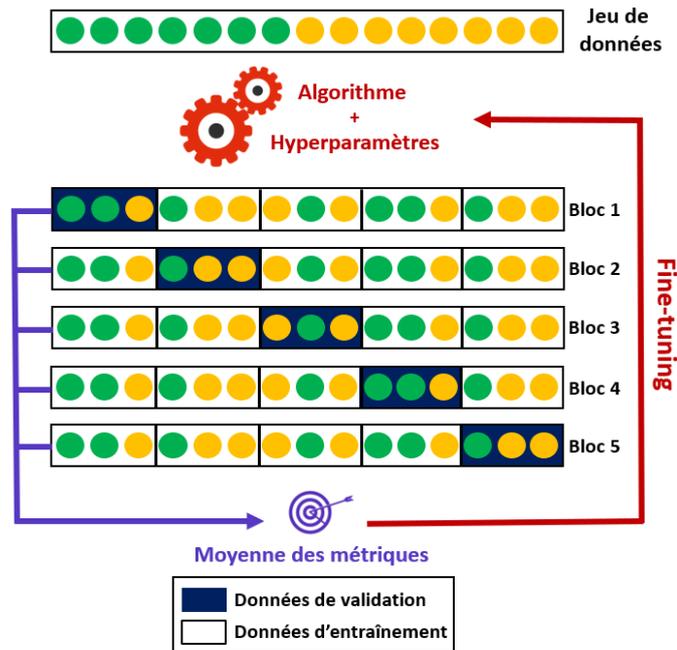


Figure 133. Exemple de validation croisée à cinq blocs appliquée à un jeu de données comportant 15 observations

- La validation croisée d'un contre tous (ou *leave-one-out cross-validation*), dont le principe est identique à celui de la validation croisée à k blocs, à la différence qu'un bloc correspond à une observation ($k = n$) (figure 134). Chaque itération attribue à $n - 1$ observations le rôle de données d'apprentissage et à l'observation restante le rôle de donnée de validation [865].

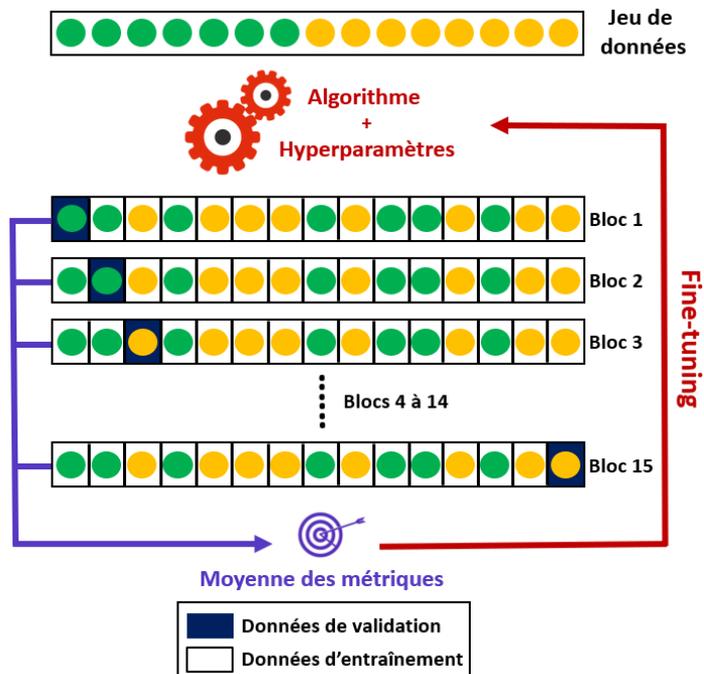


Figure 134. Exemple de validation croisée d'un contre tous appliquées à un jeu de données comportant 15 observations

- La validation croisée d'un groupe contre tous (ou *leave-one-group-out cross-validation*), qui est une variante de la validation croisée à k blocs pour laquelle sont contrôlées une ou plusieurs variables de façon à ce que chaque sous-échantillon corresponde à une classe ou à un sous-groupe (ex. : tranches d'âge) [865].
- La validation croisée stratifiée (ou *stratified cross-validation*), qui est également similaire à la validation croisée à k blocs mais qui s'en différencie par le maintien de la distribution des classes au sein des sous-échantillons (*figure 135*) [865].

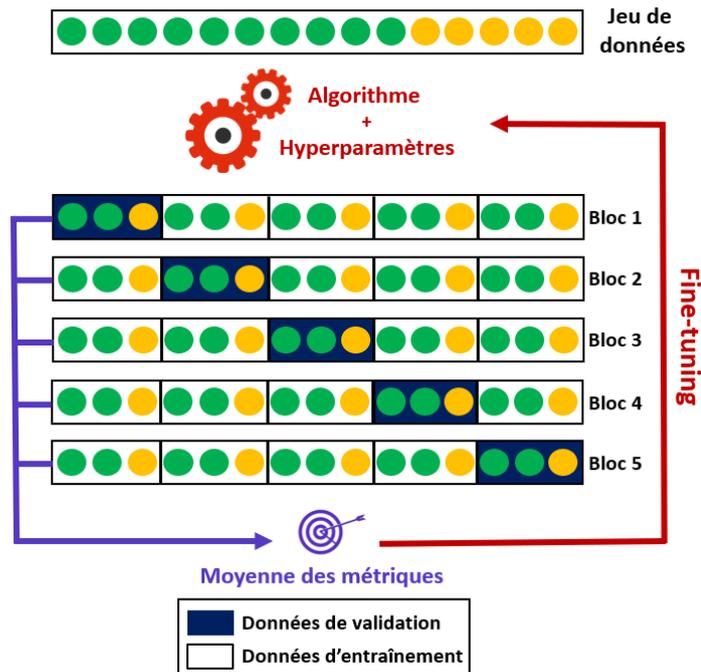


Figure 135. Exemple de validation croisée stratifiée appliquée à un jeu de données comportant 15 observations

Nous pouvons également mentionner le *bootstrapping* qui est une méthode de composition aléatoire des sous-échantillons. Ce procédé se fait avec remplacement, c'est-à-dire que certaines observations peuvent être sélectionnées plusieurs fois pour former un sous-échantillon, tandis que d'autres peuvent ne jamais être sélectionnées (*figure 136*) [870].

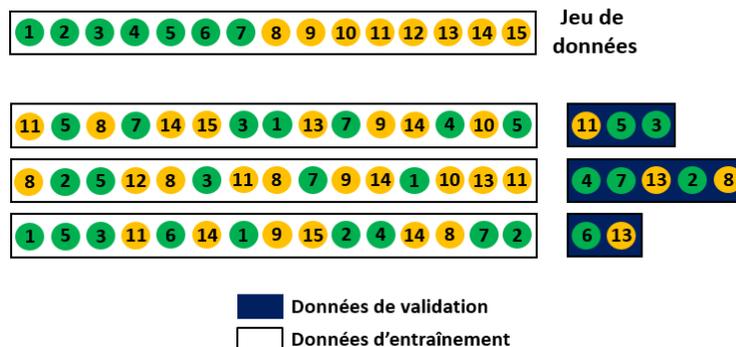


Figure 136. Exemple de *bootstrapping* appliqué à un jeu de données comportant 15 observations

Le choix de la méthode employée doit tenir compte du coût computationnel, de la nature des données, de leur quantité ainsi que de la distribution des variables. Faire intervenir ces méthodes d'échantillonnage a pour effet d'accroître la fiabilité de l'évaluation, ce qui peut notamment favoriser la détection d'un sur-apprentissage [865].

4. Évaluation finale

a. La loi de Goodhart appliquée à l'apprentissage automatique

La loi de Goodhart, formulée par l'économiste Charles Goodhart en 1975, nous dit ceci : « *Toute régularité statistique observée aura tendance à s'effondrer dès lors qu'une pression lui sera exercée à des fins de contrôle.* » [871]. Autrement dit, la fiabilité d'un indicateur statistique est à remettre en question lorsque maximiser ou minimiser celui-ci devient un objectif.

Appliquée à l'apprentissage automatique, cette loi se traduit par le fait que les résultats obtenus lors de la phase de validation ne reflètent pas nécessairement la qualité réelle d'un modèle car l'objectif d'optimisation des conditions d'apprentissage est poursuivi au regard des métriques calculées sur un échantillon d'observations donné [872]. Les bonnes pratiques suggèrent donc d'effectuer une évaluation finale du ou des modèles retenus à partir d'un échantillon attiré (nommées données de test ou *hold-out set*), indépendant de l'entraînement et de la validation (figure 137).

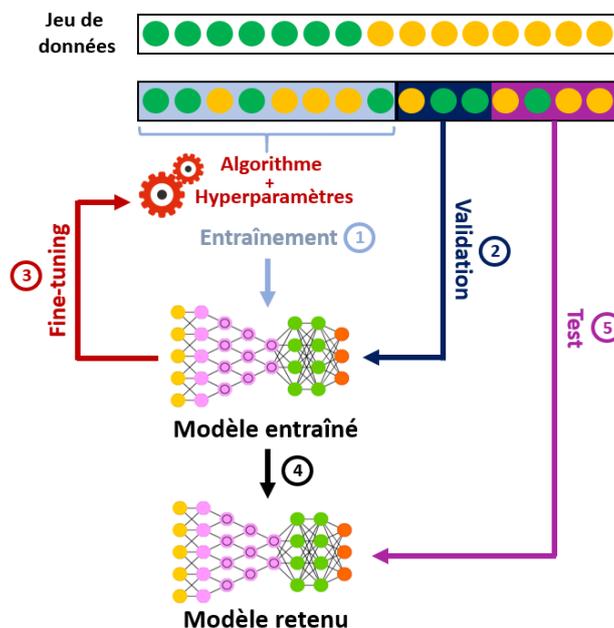


Figure 137. Exemple de *hold-out* appliqué à un jeu de données comportant 15 observations

Cette bonne pratique, parfois appelée technique du *hold-out*, est compatible avec l'emploi des méthodes d'échantillonnage telles que la validation croisée [865].

b. La validation croisée imbriquée

La validation croisée imbriquée (ou *nested cross-validation*) est une variante de validation croisée dans laquelle est imbriquée la technique du *hold-out*. Celle-ci parvient donc à réunir en un seul procédé la phase de validation et l'évaluation finale, limitant ainsi le biais décrit par la loi de Goodhart [865].

Pour ce faire, cette méthode comporte deux types de boucles opérant à des niveaux différents. Le premier type de boucles, celles d'optimisation des conditions d'apprentissage, est interne. Le second type de boucle, ayant pour rôle d'évaluer la performance des modèles retenus, est externe. Le type de validation croisée intervenant au sein des boucles est à la discrétion de l'utilisateur (ex. : une validation croisée imbriquée faisant intervenir une validation croisée à k -blocs pour ses boucles internes et une validation croisée d'un groupe contre tous pour sa boucle externe) [865].

En appliquant la validation croisée à k blocs pour la boucle externe et à k' blocs pour la boucle interne, le fonctionnement de la validation croisée imbriquée est le suivant (*figure 138*) :

- 1) Les données d'entraînement sont divisées en k blocs externes de tailles similaires.
- 2) Parmi ces k blocs externes, $k-1$ d'entre eux sont utilisés pour constituer un échantillon destiné à la boucle interne, celui-ci est divisé en k' blocs internes.
- 3) Le modèle est entraîné sur $k'-1$ des blocs internes puis validé sur le bloc interne restant.
- 4) Le processus est itéré k' fois en interchangeant le rôle des blocs internes de façon à ce que chacun de ces blocs endosse à la fois le rôle de données d'entraînement et de données de validation, il s'agit de la boucle interne.
- 5) À l'issue de cette boucle interne, k' modèles ont été entraînés puis évalués. La performance de ces k' modèles peut alors être estimée en calculant la moyenne des métriques obtenues avec chacun d'eux. L'intérêt de la boucle interne réside dans la possibilité de reproduire celle-ci en faisant varier les hyperparamètres, pour ainsi satisfaire deux objectifs à la fois : sélectionner le modèle présentant les meilleures métriques et identifier les hyperparamètres optimaux à privilégier.
- 6) Le processus de boucle interne est itéré k fois en interchangeant le rôle des blocs externes de façon à ce que chacun de ces blocs endosse à la fois le rôle de données d'entraînement-validation et de données de test, il s'agit de la boucle externe.
- 7) À l'issue de cette boucle externe, le modèle présentant les meilleures performances a été retenu au sein de chaque boucle interne. Chacun de ces k modèles est évalué avec les données de test correspondant à sa boucle interne. Enfin, la performance globale de

ces k modèles peut alors être estimée en calculant la moyenne et la variance des métriques obtenues avec chacun d'eux.

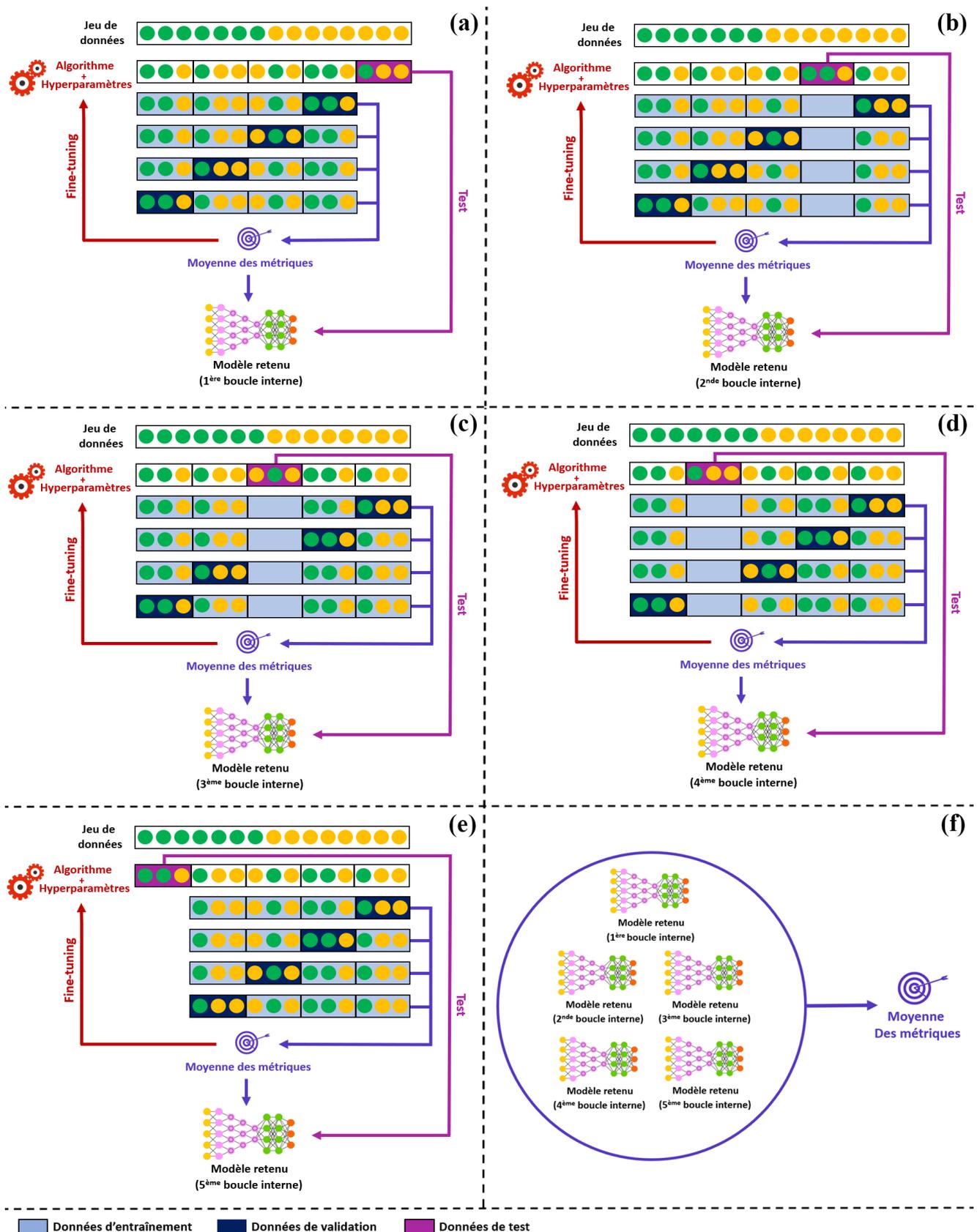


Figure 138. Exemple de validation croisée imbriquée à cinq blocs pour la validation externe et quatre blocs pour la validation interne. Chacune des cinq itérations de la boucle externe est représentée (a-e). Ces itérations se concluent par le calcul de la moyenne des métriques à partir des modèles retenus à l'issue de chaque boucle interne (f).

III. Mise sur le marché (Espace économique européen)

Une fois notre modèle (ou nos modèles) abouti et jugé apte à un usage médical, l'étape suivante est de le déployer auprès des professionnels de santé et/ou des patients. Pour ce faire, le modèle est généralement intégré dans une architecture (application, logiciel, équipement médical, etc.) adaptée à la finalité et à l'utilisateur final. Avant que cette architecture puisse être rendue accessible à ses utilisateurs cibles, celle-ci doit répondre à un certain nombre d'exigences.

1. La qualification des intelligences artificielles appliquées au diagnostic et à la thérapeutique

La mise sur le marché ou mise en service des produits ayant une finalité médicale est soumise à une réglementation qui varie selon leur qualification. Il convient donc, avant d'aborder les exigences réglementaires, de préciser la qualification désignant les intelligences artificielles appliquées au diagnostic et à la thérapeutique.

A ce jour, le sujet des intelligences artificielles n'est pas explicitement abordé dans les textes de loi européens relatifs à la réglementation des produits de santé. Néanmoins, le *Règlement 2017/745 du parlement européen et du conseil* stipule la chose suivante : « *On entend par « dispositif médical », tout instrument, appareil, équipement, logiciel, implant, réactif, matière ou autre article, destiné par le fabricant à être utilisé, seul ou en association, chez l'homme pour l'une ou plusieurs des fins médicales précises suivantes :*

- *diagnostic, prévention, contrôle, prédiction, pronostic, traitement ou atténuation d'une maladie,*
- *diagnostic, contrôle, traitement, atténuation d'une blessure ou d'un handicap ou compensation de ceux-ci,*
- *investigation, remplacement ou modification d'une structure ou fonction anatomique ou d'un processus ou état physiologique ou pathologique,*

[...]

et dont l'action principale voulue dans ou sur le corps humain n'est pas obtenue par des moyens pharmacologiques ou immunologiques ni par métabolisme, mais dont la fonction peut être assistée par de tels moyens. » [873].

Cette définition met en avant deux aspects fondamentaux qui conditionnent la qualification en dispositif médical : la nature du produit et sa finalité.

Pour être qualifié de dispositif médical, un produit doit tout d'abord effectuer son action autrement que par des moyens pharmacologiques, immunologiques ou par métabolisme. Concernant la forme que peut prendre ce produit, cette définition se veut large puisqu'elle énonce divers substantifs, ainsi que le complément « *ou autre article* ». Que le modèle soit intégré dans une architecture matérielle ou logicielle, le volet concernant la nature du produit couvre donc les modèles d'intelligence artificielle que nous traitons dans le cadre de ce manuscrit.

En revanche, les finalités pour lesquelles cette définition est applicable sont plus spécifiquement énoncées. Selon cette définition, toute intelligence artificielle (ou architecture embarquant une intelligence artificielle) n'étant pas destinée à être utilisée chez l'humain ou ne présentant aucune des finalités médicales citées n'est pas considérée comme étant un dispositif médical, et ce même si celle-ci contribue in fine à l'amélioration ou au développement des pratiques curatives ou de diagnostic. Les intelligences artificielles conçues pour être utilisées uniquement dans le cadre de la recherche préclinique ne pourraient par exemple pas être qualifiées de dispositifs médicaux.

Le *Medical Device Coordination Group* propose, dans son guide de qualification et de classification des logiciels, un arbre de décision sur lequel les fabricants peuvent s'appuyer pour déterminer si leur logiciel peut être qualifié ou non de dispositif médical (*figure 139*). Cet arbre peut s'employer de la même façon pour les modèles d'intelligence artificielle [874].

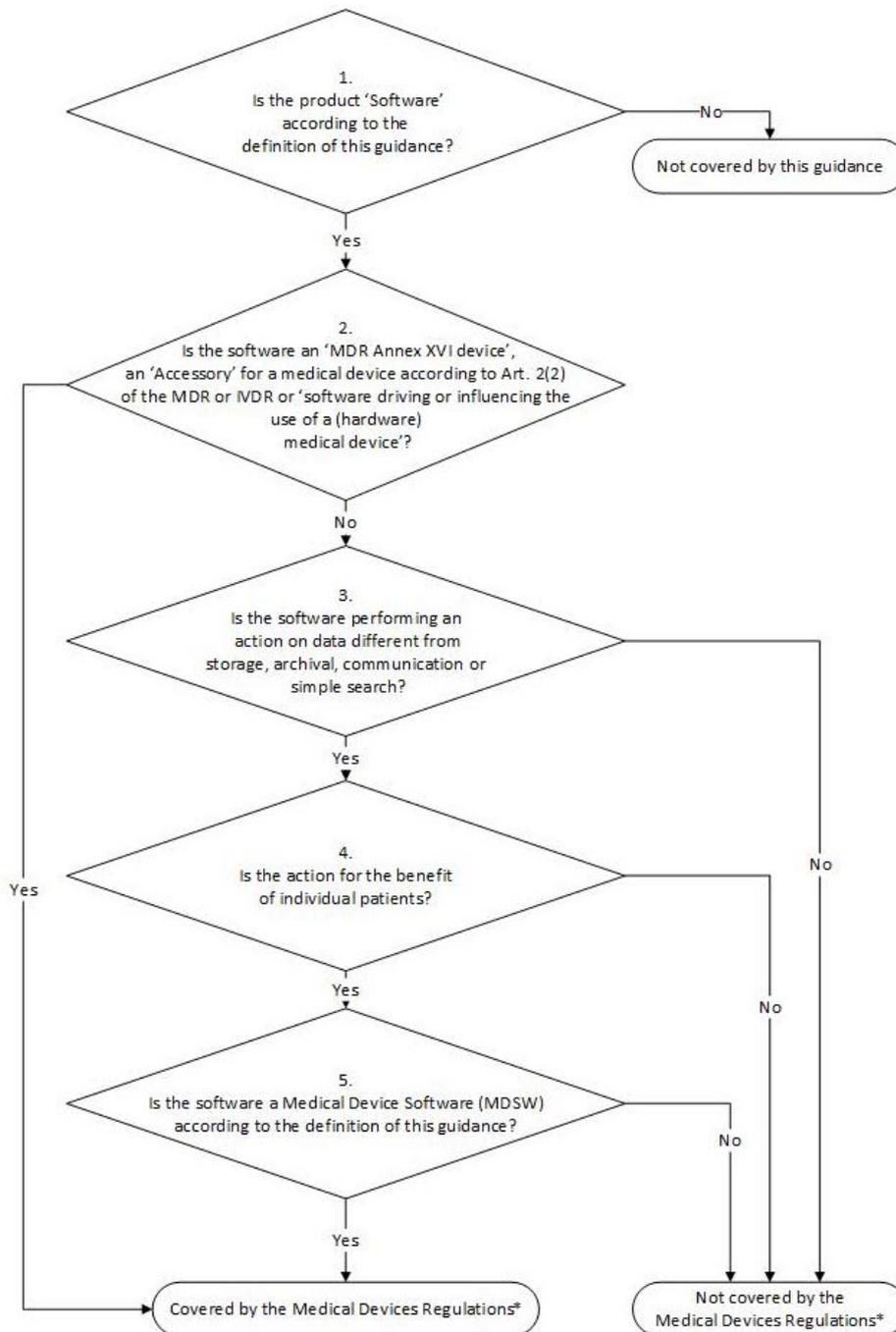


Figure 139. Arbre de décision destiné à éclairer les fabricants quant à la qualification de leur logiciel en dispositif médical (Medical Device Coordination Group) [874]

L'annexe VIII du *Règlement 2017/745* précise les différentes classes de dispositifs médicaux. Celles-ci expriment le niveau de risque associé aux dispositifs médicaux qu'elles qualifient : classe I (risque faible), classe IIa (risque potentiel modéré), classe IIb (risque potentiel élevé/important), classe III (risque élevé). Concernant la classification des logiciels qualifiés de dispositif médical, le *Règlement 2017/745* stipule ceci :

- « *Le logiciel commandant un dispositif ou agissant sur son utilisation relève de la même classe que le dispositif.* » ;
- « *Si le logiciel est indépendant de tout autre dispositif, il est classé en tant que tel.* » ;

- « Les logiciels destinés à fournir des informations utilisées pour prendre des décisions à des fins thérapeutiques ou diagnostiques relèvent de la classe IIa, sauf si ces décisions ont une incidence susceptible de causer la mort ou une détérioration irréversible de l'état de santé d'une personne, auxquels cas ils relèvent de la classe III, ou une grave détérioration de l'état de santé d'une personne ou une intervention chirurgicale, auxquels cas ils relèvent de la classe IIb. Les logiciels destinés à contrôler des processus physiologiques relèvent de la classe IIa, sauf s'ils sont destinés à contrôler des paramètres physiologiques vitaux, lorsque des variations de certains de ces paramètres peuvent présenter un danger immédiat pour la vie du patient, auxquels cas ils relèvent de la classe IIb. » [873].

Enfin, un modèle d'intelligence artificielle qualifié comme étant un dispositif médical est classifié comme « *dispositif actif* », ce qui est désigné par le Règlement 2017/745 comme « *tout dispositif dont le fonctionnement dépend d'une source d'énergie autre que celle générée par le corps humain à cette fin ou par la pesanteur et agissant par modification de la densité de cette énergie ou par conversion de celle-ci. [...] Les logiciels sont aussi réputés être des dispositifs actifs.* » [873].

2. L'autorisation de mise sur le marché

a. Le marquage CE médical

Sur le territoire de l'EEE, un marquage CE (conformité européenne) est obligatoire pour tous les produits couverts par des directives ou règlements européens qui le prévoient explicitement. C'est notamment le cas des dispositifs médicaux dont la mise sur le marché ou la mise en service nécessitent préalablement l'apposition d'un marquage CE dit « *médical* ». Il est important de noter que celui-ci est spécifique aux produits qualifiables en dispositifs médicaux, les modèles d'intelligence artificielle dont les caractéristiques divergent de cette qualification ne sont donc pas concernés par le marquage CE médical [873].

Un marquage CE, lorsqu'il est apposé sur un dispositif médical, indique sa conformité quant au bénéfice clinique et à la sécurité des patients. Pour qu'un fabricant puisse apposer un marquage CE sur un dispositif médical, il doit donc s'assurer que les exigences détaillées dans le Règlement 2017/745 soient satisfaites : « *Un dispositif ne peut être mis sur le marché ou mis en service que s'il est conforme au présent règlement au moment où il est dûment fourni et dès lors qu'il est correctement installé, entretenu et utilisé conformément à sa destination.* » [873].

b. Les exigences réglementaires essentielles

La démonstration de la conformité d'un dispositif médical vis-à-vis des directives européennes repose sur l'évaluation du produit concerné mais également sur la capacité de l'entreprise à assurer cette conformité. Les exigences de sécurité et de performance auxquelles doivent répondre les dispositifs médicaux marqués CE sont énoncées dans l'annexe I du *Règlement 2017/745* : « *Un dispositif est conforme aux exigences générales en matière de sécurité et de performances énoncées à l'annexe I qui lui sont applicables, compte tenu de sa destination.* ». Celles-ci portent sur la conception, la fabrication ainsi que le respect des normes et référentiels pré-cliniques et cliniques [873].

Il est à la charge du fabricant de constituer une documentation regroupant les éléments de sécurité et de performance pertinents à communiquer : « *Chaque dispositif est accompagné des informations nécessaires à l'identification de celui-ci et de son fabricant, ainsi que de toute information relative à la sécurité et aux performances utile à l'utilisateur ou à toute autre personne, le cas échéant. Ces informations peuvent figurer sur le dispositif lui-même, sur le conditionnement ou dans la notice d'utilisation et, si le fabricant dispose d'un site internet, sont mises à disposition et mises à jour sur le site internet [...].* ». Le fabricant doit également tenir une documentation technique récapitulant toutes les informations relatives au dispositif médical tout au long de son cycle de vie. Remarque : le fabricant doit prendre en considération les exigences linguistiques spécifiques à chaque état membre de l'EEE dans lequel il souhaite rendre disponible son dispositif médical [873].

Lorsque le dispositif médical inclut ou est lui-même un modèle d'intelligence artificielle, le fabricant doit s'assurer que son environnement technique soit maîtrisé et conforme aux normes en vigueur : « *Pour les dispositifs qui comprennent des logiciels ou pour les logiciels qui sont des dispositifs à part entière, ces logiciels sont développés et fabriqués conformément à l'état de l'art, compte tenu des principes du cycle de développement, de gestion des risques, y compris la sécurité de l'information, de vérification et de validation.* ». Le fabricant doit veiller à ce que son dispositif médical soit en mesure de fonctionner pleinement sur tout matériel pour lequel il est destiné (ex. : responsivité d'une application mobile) : « *Les logiciels visés à la présente section qui sont destinés à être utilisés en combinaison avec des plateformes informatiques mobiles sont conçus et fabriqués en tenant compte des caractéristiques spécifiques de la plateforme mobile (par exemple, taille et rapport de contraste de l'écran) et des facteurs externes liés à leur utilisation (variation du niveau sonore ou de la luminosité dans l'environnement).* » [873].

Les dispositifs médicaux doivent être sécurisés sur le plan mécanique, électronique et informatique. Un point d'honneur doit notamment être mis sur la cybersécurité : « *Les fabricants énoncent les exigences minimales concernant le matériel informatique, les caractéristiques des réseaux informatiques et les mesures de sécurité informatique, y compris la protection contre l'accès non autorisé, qui sont nécessaires pour faire fonctionner le logiciel comme prévu.* ». Le fabricant doit mettre en place des procédures et systèmes aptes à gérer les incidents déclarés, les actions de sécurité ou la diffusion des mises à jour vers les utilisateurs. En outre, il doit fonder sa gestion des risques sur des textes de référence telles que les normes harmonisées suivantes [875] :

- NF EN ISO 14971 (gestion des risques relatifs aux dispositifs médicaux) [876] ;
- NF EN 62304 (gestion des risques spécifiques aux logiciels de dispositifs médicaux) [877] ;
- NF EN 60601-1-4 (sécurité des appareils et systèmes électromédicaux comprenant des sous-systèmes électroniques programmables) [878].

S'il choisit de ne pas se conformer à ces normes, le fabricant doit le justifier et appliquer des référentiels équivalents [873].

Aussi, le fabricant doit mettre en place un système de gestion de la qualité en mesure d'assurer la reproductibilité de la conformité aux exigences. Celui-ci peut par exemple s'appuyer sur la norme NF EN ISO 13485 (exigences relatives au système de management de la qualité d'un fabricant de dispositifs médicaux) [879].

Dernier point essentiel, mais non des moindres, le *Règlement 2017/745* impose désormais aux fabricants une évaluation clinique de chacun de leurs dispositifs médicaux avant et après leur mise sur le marché [873].

c. L'évaluation clinique

L'évaluation clinique d'un dispositif médical est définie par le *Règlement 2017/745* comme étant « *un processus systématique et planifié visant à produire, collecter, analyser et évaluer en continu les données cliniques relatives à un dispositif afin de vérifier la sécurité et les performances, y compris les bénéfices cliniques, de celui-ci lorsqu'il est utilisé conformément à la destination prévue par le fabricant* ». Celle-ci constitue le pilier majeur de la démonstration de la conformité aux exigences générales de performance et de sécurité, comme indiqué dans l'article 61 dudit règlement: « *La confirmation de la conformité aux exigences générales pertinentes en matière de sécurité et de performances énoncées à l'annexe I dans des conditions*

normales d'utilisation d'un dispositif, ainsi que l'évaluation des effets secondaires indésirables et du caractère acceptable du rapport bénéfice/risque visé à l'annexe I, sections 1 et 8, sont fondées sur des données cliniques apportant une preuve clinique suffisante, [...]. » Le chapitre VI dudit règlement détaille, de l'article 61 à l'article 82, le cadre réglementaire relatif à l'évaluation clinique et aux investigations cliniques [873].

L'objectif de l'évaluation clinique est de définir précisément les bénéfices cliniques offerts par le produit évalué ainsi que son rapport bénéfice/risque. Ce dernier doit être mis en parallèle avec le rapport bénéfice/risque des solutions diagnostiques ou thérapeutiques comparables. Il appartient au fabricant de déterminer et justifier le niveau de preuve clinique à privilégier : « *Le fabricant précise et justifie le niveau de preuve clinique nécessaire pour démontrer la conformité aux exigences générales pertinentes en matière de sécurité et de performances. Ce niveau de preuve clinique est approprié, eu égard aux caractéristiques du dispositif et à sa destination.* » [873].

L'évaluation clinique d'un dispositif médical est la combinaison d'un ou plusieurs des moyens suivants :

- L'évaluation par la littérature, qui consiste à dresser un état de l'art de la solution proposée par le fabricant.
- L'évaluation par équivalence, dont le principe est de s'appuyer sur des résultats cliniques obtenus à partir d'un produit équivalent déjà présent sur le marché. Pour que deux produits puissent être admis comme étant équivalents, ceux-ci doivent être similaires techniquement, biologiquement et cliniquement.
- L'évaluation à partir d'investigations cliniques, définies par le *Règlement 2017/745* comme « *toute investigation systématique impliquant un ou plusieurs participants humains destinée à évaluer la sécurité ou les performances d'un dispositif* ». Les investigations cliniques doivent évaluer l'emploi du dispositif dans le cadre de la pratique médicale revendiquée par le fabricant [873].

A l'instar de sa qualification, c'est, non pas le modèle seul, mais l'ensemble de l'architecture dans laquelle ce modèle est intégré qui doit faire l'objet d'une évaluation clinique. Pour obtenir et conserver le marquage CE pour l'un de ses produits, le fabricant doit démontrer que celui-ci présente un rapport bénéfice/risque favorable tout au long de sa commercialisation [873].

d. L'évaluation de la conformité au marquage CE

Pour obtenir le droit d'apposer un marquage CE sur l'un de ses produits, le fabricant doit constituer un dossier de demande d'autorisation de mise sur le marché (AMM) regroupant l'ensemble des éléments que nous avons mentionné. Ces éléments doivent être en mesure de démontrer que le produit en question répond aux exigences de sécurité et de performance telles que définies dans le *Règlement 2017/745* [873].

Le fabricant doit ensuite adresser ce dossier à un organisme indépendant dit « notifié », qui doit être habilité à évaluer la conformité des dispositifs médicaux. Les organismes notifiés sont désignés par les autorités sanitaires compétentes (ex. : en France, il s'agit de l'Agence Nationale de Sécurité du Médicament et des produits de santé). La liste des organismes notifiés habilités varie selon le pays et la classe du dispositif médical aspirant au marquage CE. À la réception du dossier, l'organisme notifié est chargé d'évaluer à la fois la conformité du dispositif médical mais aussi celles du plan de surveillance post-AMM et du système qualité mis en place par le fabricant. L'organisme notifié doit également s'assurer que les personnes ayant réalisé l'évaluation clinique aient les compétences requises [879].

Remarque : les dispositifs médicaux de classe I font exception à la règle. L'évaluation de leur conformité ne nécessite pas l'intervention d'un organisme notifié, le fabricant effectue son évaluation de conformité lui-même et peut apposer un marquage CE sur son produit à l'issue de celle-ci [879].

Le fabricant est libre de choisir l'organisme notifié auquel il adresse son dossier de demande d'AMM, dès lors que cet organisme notifié est habilité pour la classe du dispositif médical à évaluer [880]. De même, c'est au fabricant de déterminer la classe du dispositif médical concerné par sa demande d'AMM, selon les caractéristiques du produit, la finalité revendiquée et les règles spécifiques prévues par les directives européennes. Les exigences quant à la performance et à la sécurité sont identiques quelle que soit la classe du dispositif médical, cependant les modalités de démonstration de la conformité à ces exigences sont d'autant plus exigeantes que la classe attribuée au dispositif est élevée (ex. : l'évaluation par investigation clinique est obligatoire pour les dispositifs médicaux de classe III, sauf cas particuliers) [881].

Si l'organisme notifié juge conformes à la réglementation le dispositif médical, le plan de surveillance post-AMM et le système qualité présentés, il délivre un certificat autorisant le fabricant à apposer le marquage CE sur son produit [881].

e. La mise sur le marché

Le certificat autorisant l'apposition du marquage CE est valable seulement pour le produit dont la conformité a été évaluée et sur une durée maximale de cinq ans après avoir été délivré [881]. Celui-ci confère, au cours de sa validité, le droit de commercialiser ou mettre à disposition le produit concerné dans l'ensemble de l'EEE sans qu'aucune formalité, norme de sécurité ou étude supplémentaire ne puisse être réclamée. En effet, le *Règlement 2017/745* stipule ceci : « *D'une manière générale, le marquage CE devrait être apposé sur les dispositifs pour indiquer leur conformité avec le présent règlement afin qu'ils puissent circuler librement dans l'Union et être mis en service conformément à leur destination. Les États membres devraient s'abstenir de créer des entraves à la mise sur le marché ou à la mise en service de dispositifs conformes aux exigences du présent règlement. Néanmoins, les États membres devraient être autorisés à décider s'il y a lieu de restreindre l'utilisation de tout type particulier de dispositif en ce qui concerne des aspects ne relevant pas du présent règlement.* » [873].

Le système du marquage CE médical présente deux avantages essentiels. De concert avec l'harmonisation de la réglementation sur le territoire de l'EEE, le marquage CE simplifie et favorise la libre circulation des dispositifs médicaux sur ce territoire. De plus, cela permet à quiconque de vérifier rapidement et simplement la conformité d'un produit au regard de la réglementation.

Le fabricant peut en outre, s'il le souhaite, procéder à une demande de prise en charge de son dispositif médical. En France, un produit doit être inscrit sur la liste des produits et prestations remboursables (LPPR) pour être pris en charge par l'assurance maladie. Pour ce faire, le fabricant doit soumettre un dossier de demande de prise en charge dérogatoire à la Haute Autorité de Santé (HAS). Les éléments constitutifs du dossier sont détaillés dans le guide de la HAS dédié aux produits innovants. Ce dossier doit notamment comporter un argumentaire de conformité aux critères d'éligibilité ainsi que les données cliniques et/ou médico-économiques relatives au produit [881].

Dans l'annexe 4 du guide sus-mentionné, la HAS propose une grille regroupant les informations supplémentaires à transmettre lorsque le dispositif médical est (ou inclut) un modèle d'apprentissage automatique. Ces informations doivent décrire les finalités d'usage, les données d'apprentissage et d'entrée, la méthode d'entraînement et d'évaluation, la surveillance vis-à-vis de la robustesse et la résilience du système, ainsi que les éléments d'explicabilité et d'interprétabilité [882]. En parallèle, le Comité économique des produits de santé (CEPS) est chargé de fixer avec les fabricants un tarif de remboursement ou un prix limite de vente. Si le dispositif médical n'est pas à usage individuel, c'est à l'Union nationale des caisses d'assurance

maladie (UNCAM) d'établir le prix et le tarif de remboursement de l'acte médical correspondant [736].

Suite à l'obtention d'un avis favorable de la part de la HAS et à l'établissement des modalités de remboursement, c'est au ministère chargé de la Santé que revient la décision d'inscrire la prestation médicale ou le dispositif médical sur la LPPR [736,883].

Tout dispositif médical mis en service pour la première fois sur un territoire national doit être déclaré auprès de son autorité compétente (ex. : en France, l'Agence nationale de sécurité du médicament et des produits de santé) [881].

3. La surveillance post-AMM

Dès lors que débute la distribution d'un dispositif médical, son fabricant doit mettre en œuvre son plan de surveillance post-AMM et ce tout au long de son cycle de vie (*figure 140*).

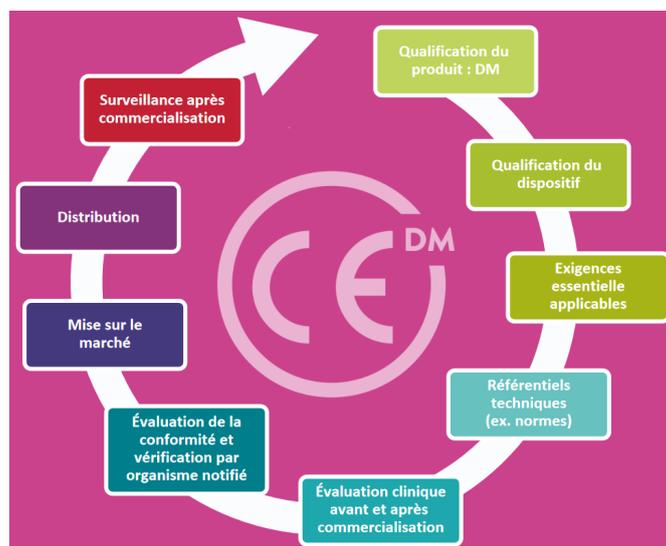


Figure 140. Cycle des exigences relatives à la mise sur le marché d'un dispositif médical (*SNITEM*) [884]

Ce système doit permettre, dans un premier temps, la collecte (via la littérature, les alertes et données en vie réelle) des données relatives aux performances et à la sécurité du produit. Dans un second temps, le système doit prévoir et assurer [873] :

- Une analyse de la performance et de la sécurité fondée sur les données collectées.
- Une gestion efficace du risque à travers des mesures préventives ou correctives. La survenue d'incidents doit notamment faire l'objet d'une surveillance continue (via le site du fabricant et les retours de l'autorité compétente) et d'une rétroaction le cas échéant.
- La notification des autorités compétentes en cas d'événements indésirables graves ou de défectuosité du dispositif.

- Une mise à jour continue du rapport bénéfice/risque et de l'évaluation clinique. Cette mise à jour s'appuie notamment sur le suivi clinique après commercialisation réalisé à partir des données cliniques en vie réelle.

L'organisme notifié ayant accordé le marquage CE est pour sa part chargé d'effectuer un suivi vis-à-vis du fabricant. Ce suivi s'opère via des audits sur site qui sont en principe systématiques et annuels, cependant ceux-ci peuvent également être réalisés inopinément. Lors de modifications ou du renouvellement du marquage CE, l'audit effectué est généralement plus approfondi. Au cours de ces audits, l'organisme notifié examine notamment la pertinence de la documentation technique, la bonne exécution du plan de surveillance post-AMM et vérifie que le rapport bénéfice/risque soit toujours favorable [885].

Les autorités compétentes des pays de l'EEE sont quant à elles chargées d'assurer la matériovigilance sur leurs territoires respectifs. C'est donc l'Agence nationale de sécurité du médicament et des produits de santé (ANSM) qui est responsable de la matériovigilance en France, qu'elle accomplit à travers le recueil et le traitement des notifications d'incidents et effets indésirables ainsi qu'en effectuant des inspections sur sites. Les fabricants et les professionnels de santé sont tenus de déclarer à l'ANSM tout incident ou risque d'incident de toute nature lié à l'utilisation d'un dispositif médical. Les patients, bien qu'ils n'en aient pas l'obligation, peuvent également signaler à l'ANSM tout effet indésirable consécutif à l'utilisation d'un dispositif médical [886]. En cas d'écart vis-à-vis de la réglementation ou d'identification d'un risque pour la sécurité des patients, et selon le degré d'écart ou de risque, l'ANSM peut prendre la décision d'émettre des recommandations, de suspendre temporairement l'AMM ou de retirer définitivement du marché le dispositif médical en cause [887].

4. La question de la responsabilité juridique

Malgré toutes les précautions pouvant être prises par le fabricant et les autorités compétentes, les dispositifs médicaux ne sont pas exempts de risque pour la santé des patients. A plus forte raison lorsque ceux-ci intègrent un modèle d'apprentissage automatique, qui comme nous l'avons vu est sujet aux biais algorithmiques. Du fait de la complexité de son explicabilité, l'intelligence artificielle appliquée au diagnostic ou à la prise de décision thérapeutique induit des problématiques juridiques toutes aussi complexes à traiter. L'une des problématiques majeures est celle de la responsabilité juridique en cas d'erreur induite par le modèle [780].

Il n'existe, à l'heure actuelle, aucune réglementation à ce sujet qui soit spécifique à l'utilisation des intelligences artificielles de diagnostic et d'aide à la décision thérapeutique. L'approche actuelle consiste à leur appliquer la législation propre aux dispositifs médicaux [873].

Dans la plupart des États, la législation en matière de responsabilité médicale s'applique aux décisions prises par les professionnels de santé. Indépendamment des moyens employés pour justifier ces décisions, l'engagement de la responsabilité du professionnel de santé dépend du respect ou non des recommandations médicales en vigueur. Cela signifie que la responsabilité d'une erreur de diagnostic ou une mauvaise décision thérapeutique pourraient être imputées avant tout au professionnel de la santé si cette erreur est causée par l'utilisation non prévue par les recommandations d'un modèle d'intelligence artificielle [888].

Si en revanche cette utilisation est prévue par les recommandations et que l'erreur est imputable au dispositif médical en lui-même, c'est la responsabilité du fabricant qui est engagée et c'est en principe aux autorités sanitaires de statuer sur la mesure corrective imposée au fabricant. Cependant, la démonstration de cette imputabilité peut s'avérer difficile voire impossible selon l'explicitabilité du modèle utilisé [888].

De surcroît, il faut considérer l'émergence des nouvelles méthodes d'apprentissage automatique. Qu'en est-il par exemple de l'apprentissage incrémental, un type d'apprentissage permettant à un modèle de poursuivre son apprentissage après sa mise sur le marché [889] ? Faut-il encadrer ce type d'apprentissage ou en interdire radicalement l'usage dans le contexte clinique ?

Au vu de ces défis, l'intelligence artificielle apparaît comme un dispositif atypique et aux multiples facettes, un atypisme et une diversité qui semblent justifier une législation spécifique [890]. Ce constat fait l'objet d'une réflexion de la part de nombreux législateurs et professionnels de santé [891,892,893].

Différentes approches sont explorées en vue de définir précisément les exigences et la responsabilité juridique de chaque acteur selon les différents cas d'usage (contexte d'utilisation, degré d'intervention) [730]. Les questions relatives aux exigences portent essentiellement sur la transparence, la documentation des décisions prises par le modèle et la rigueur attribuée aux protocoles de validation des modèles. Ces réflexions sont primordiales pour garantir une utilisation éthique de l'intelligence artificielle. C'est pourquoi, malgré l'apparente inextricabilité de certains de ces sujets, cet élan de questionnement et d'échange doit être préservé en encourageant toujours davantage les collaborations et l'intégration des connaissances entre les champs du droit, des technologies et de la santé.

PARTIE 4 : Mise en pratique - Méla-Safe

PARTIE 4 : Mise en pratique - Méla-Safe

Dans l'optique de mettre en pratique les notions abordées au cours de ce manuscrit, nous avons souhaité développer une architecture d'intelligence artificielle appliquée au diagnostic. Nous avons choisi d'axer notre projet sur la détection des mélanomes.

I. Justification du projet

1. Contexte

En 2020, le nombre de nouveaux cas de cancers de la peau (carcinomes et mélanomes) était estimé à 1,5 millions [894]. Les mélanomes, plus rares mais plus dangereux que les carcinomes, représentent le 17ème cancer le plus fréquent mondialement [895]. L'incidence mondiale des mélanomes est globalement en hausse des années 50 à nos jours, atteignant en 2020 les 325 000 nouveaux cas par an (*figure 141*) [894]. Ceux-ci sont actuellement responsables du décès de près de 57 000 personnes chaque année.

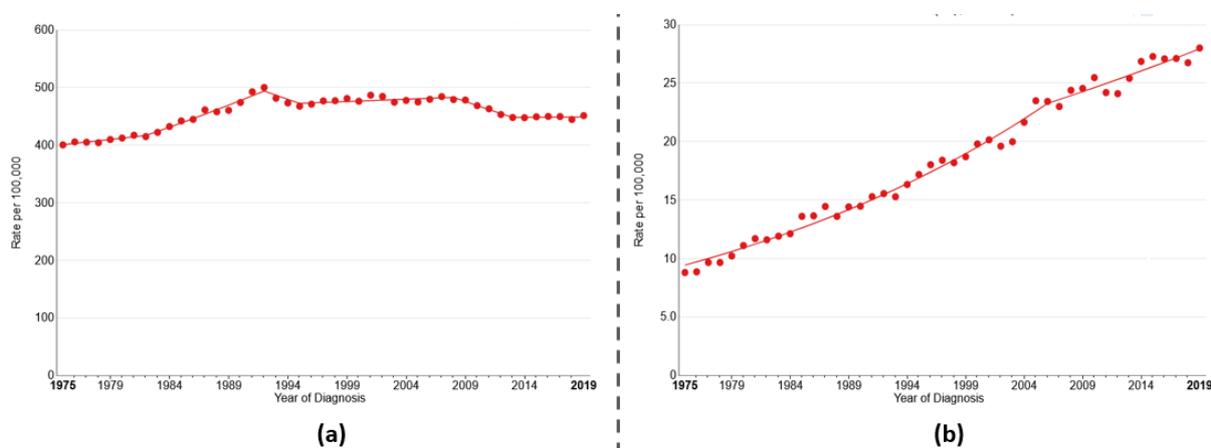


Figure 141. Incidence de cancer au États-Unis entre 1975 et 2019, tous cancers confondus (a) vs. mélanomes uniquement (b) (*National Cancer Institute*) [896,897]

Une poursuite de la hausse de ces chiffres est à prévoir au cours des années à venir selon les estimations du Centre international de recherche sur le cancer. En 2040, le nombre de nouveaux cas annuels devrait ainsi atteindre 510 000 tandis que le nombre de décès annuels serait de 96 000 [894]. Ce cancer mortel et ayant un fort potentiel métastatique (en particulier les mélanomes nodulaires) présente pourtant un taux de guérison élevé lorsque celui-ci est pris en charge précocement [898]. Ce constat met en évidence l'intérêt et l'urgence de développer des moyens favorisant le diagnostic précoce des mélanomes, et ces moyens peuvent notamment s'appuyer sur l'intelligence artificielle.

2. Réponse au besoin

Notre projet, que nous avons choisi de nommer *Méla-Safe*, est une application permettant à ses utilisateurs de prédire la malignité ou la bénignité d'une tumeur cutanée. En téléchargeant une photographie de tumeur cutanée, l'utilisateur obtient instantanément une prédiction quant au caractère malin de la tumeur photographiée, ainsi que les recommandations adéquates. Si la prédiction indique la présence d'une tumeur maligne, l'application suggère à l'utilisateur de s'orienter vers un dermatologue afin d'établir un diagnostic médical fiable et pouvoir prendre en charge une éventuelle tumeur maligne dans les plus brefs délais. Dans le cas où l'application prédit la présence d'une tumeur bénigne, celle-ci suggère malgré tout à l'utilisateur d'être attentif aux signes évocateurs d'une tumeur cutanée maligne (augmentation de taille, changement d'aspect, forme non circulaire, bords irréguliers, couleur non homogène) et de consulter un dermatologue en cas de doute.

Bien que *Méla-Safe* n'ait pas vocation à proposer un diagnostic fiable et définitif, la prédiction générée permet à l'utilisateur de prendre des dispositions pouvant à terme permettre un diagnostic en bonne et due forme. En concourant à la réduction du délai moyen entre l'apparition d'un symptôme évocateur d'un cancer cutané et le diagnostic de celui-ci, la finalité de *Méla-Safe* est de contribuer à réduire les complications et la létalité liées à ce type de cancers (*figure 142*).

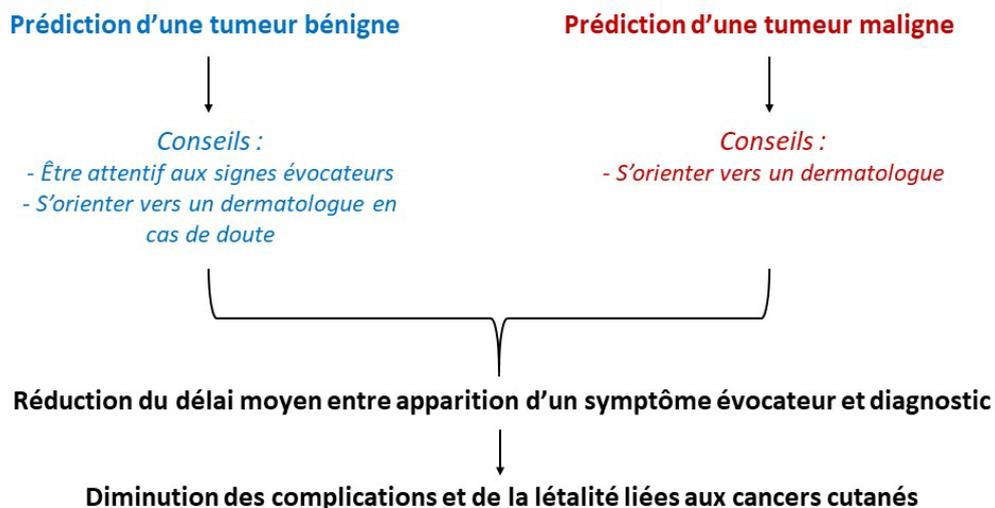


Figure 142. Finalité du projet *Méla-Safe*

II. Réalisation

La concrétisation de ce projet exigeait un certain nombre de prérequis et d'opérations que nous pouvons regrouper en trois étapes majeures. Tout d'abord, nous devions constituer un jeu de données adapté à notre objectif. Ensuite, nous devions construire, entraîner et évaluer des modèles à partir du jeu de données constitué. Enfin, le modèle retenu devait être intégré au sein d'une architecture adaptée et accessible à l'utilisateur final.

1. Constitution d'un jeu de données

a. Collecte des données

La première étape de notre projet consistait à nous procurer les données que nous allions utiliser pour entraîner et évaluer nos modèles, c'est-à-dire des photographies de tumeurs cutanées labellisées (bénigne/maligne), en nombre et qualité suffisants.

Par commodité, nous avons choisi d'employer des données mises à disposition publiquement et librement réutilisables. En menant nos recherches, nous avons pu identifier deux jeux de données correspondant à nos besoins, tous deux téléchargeables sur la plateforme Kaggle :

- *Skin Cancer: Malignant vs. Benign*, provenant des archives de l'*International Skin Imaging Collaboration (ISIC)* [899] ;
- *Melanoma Detection Dataset*, mis à disposition par une équipe de chercheur (Noel C.F. Codella et al.) à l'issue de l'*International symposium on biomedical imaging* organisé par l'ISIC en 2018 [900].

b. Description des données

En explorant les deux jeux de données sus-cités, nous avons pu constater que ceux-ci n'étaient pas partagés selon les mêmes classes (*figure 143*). Aussi, nous avons remarqué que les résolutions et dimensions des photographies variaient.

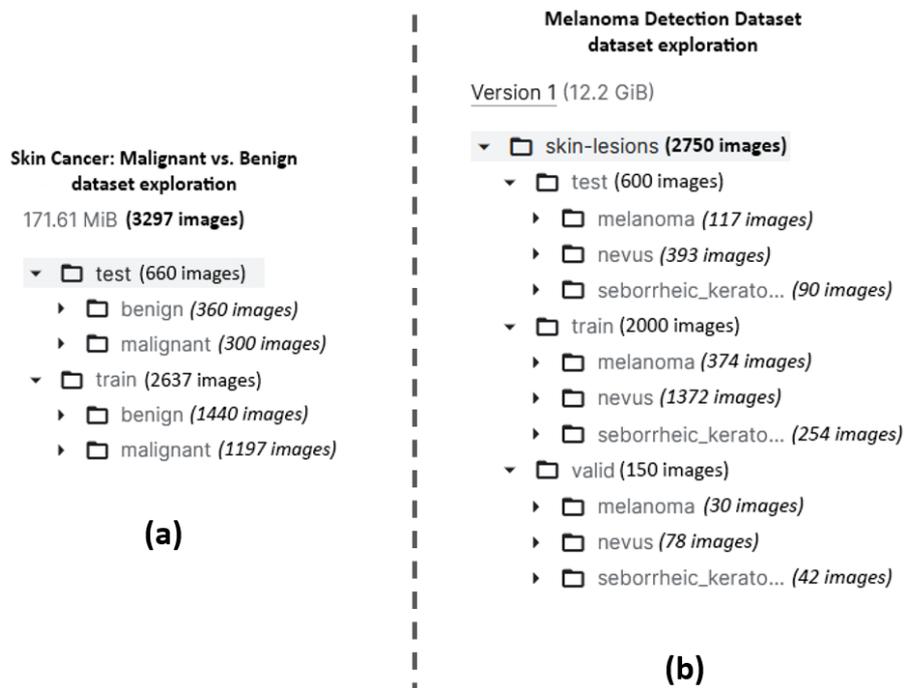


Figure 143. Arborences des jeux de données retenus (Kaggle) [899,900]

En l'occurrence, le jeu de données *Skin Cancer: Malignant vs. Benign* présentait les classes « bénigne » et « maligne », divisées entre un jeu d'entraînement (*training set*) et un jeu de test (*testing set*) (figure 143-a). L'ensemble des photographies avaient une résolution de 224 pixels sur 224 pixels, celles-ci étaient donc carrées.

Le jeu de données *Melanoma Detection Dataset*, quant à lui, présentait trois classes : « mélanome », « névus » et « kératoses séborrhéiques ». Ce jeu de données était divisé entre un jeu d'entraînement (*training set*), un jeu de validation (*validation set*) et un jeu de test (*testing set*) (figure 143-b). L'ensemble des photographies avaient une résolution de 4288 pixels sur 2848 pixels, celles-ci étaient donc rectangulaires. Nous pouvons d'ailleurs constater la haute résolution des photographies par leur poids élevé (12,2 gb) (figure 143-b).

Nous pouvons également noter que ces deux jeux de données comportent plus de photographies de tumeurs bénignes que de tumeurs malignes (figure 143).

c. Répartition des données

Afin d'optimiser la quantité et l'hétérogénéité des observations, nous avons souhaité bâtir un unique jeu de données en fusionnant les deux jeux présentés ci-dessus. Cela impliquait donc d'uniformiser la répartition des données.

Notre application doit être en mesure de distinguer les tumeurs cutanées bénignes et malignes. Nous avons donc opté pour les deux classes déjà présentes dans le jeu de données *Skin Cancer: Malignant vs. Benign* : « tumeur bénigne » et « tumeur maligne ». De façon à pouvoir évaluer notre modèle dans le respect des bonnes pratiques, nous avons en revanche choisi de conserver la répartition adoptée par le jeu de données *Melanoma Detection Dataset* : jeu d’entraînement, jeu de validation, jeu de test (voir section P3.II.4.a.) [872].

Pour ce faire, nous avons tout d’abord réuni les classes “névus” et “kératoses séborrhéiques” afin de constituer une classe “tumeur bénigne” dans les jeux d’entraînement, de validation et de test du *Melanoma Detection Dataset*. Nous avons ensuite réuni les jeux d’entraînement issus des deux jeux de données de façon à constituer un unique jeu d’entraînement divisé en deux classes. Pour finir, nous avons divisé en deux le jeu de test issu de *Skin Cancer: Malignant vs. Benign*, puis réuni chaque moitié avec la classe correspondante des jeux de validation et de test du *Melanoma Detection Dataset*, ce qui aboutit sur un unique jeu de validation et un unique jeu de test, tous deux divisés en deux classes (figure 144).

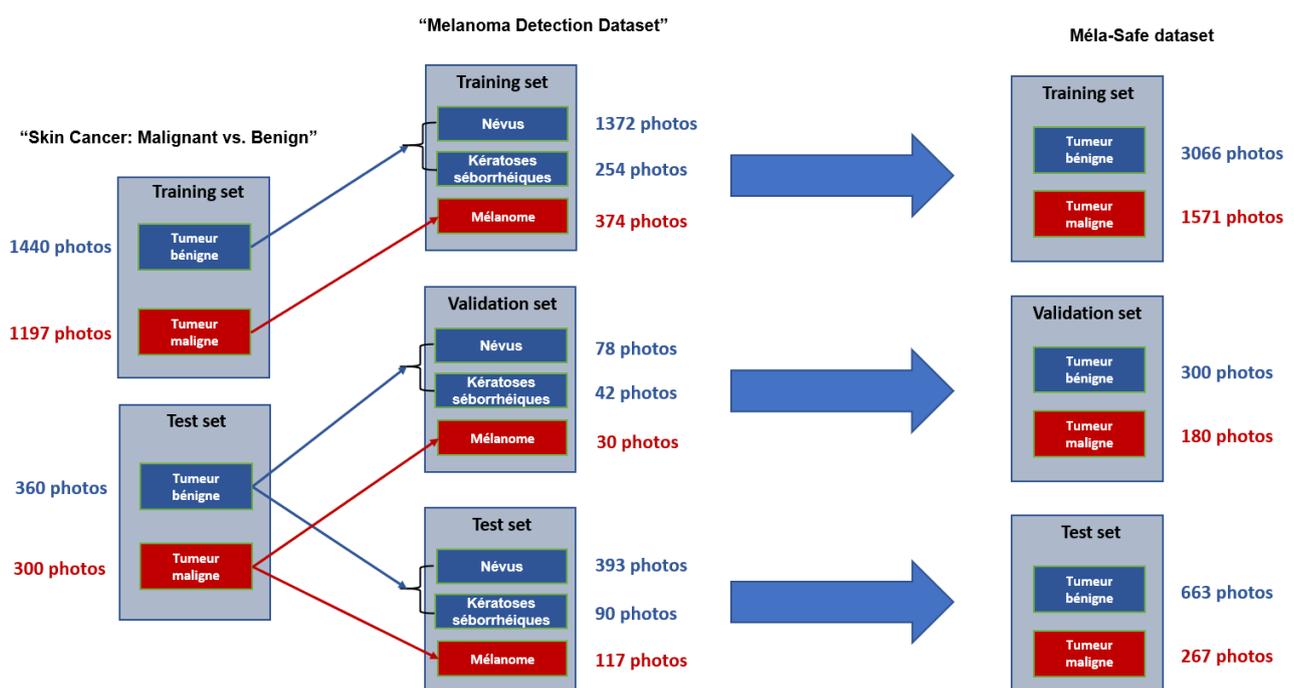


Figure 144. Constitution et répartition du jeu de données final

Nous obtenons donc un jeu de données final comportant un total de 6047 photographies distribuées de la façon suivante :

- 4637 photographies dans le jeu d’entraînement (environ 77% des données), dont 3066 de tumeurs bénignes et 1571 de tumeurs malignes ;
- 480 photographies dans le jeu de validation (environ 8% des données), dont 300 de tumeurs bénignes et 180 de tumeurs malignes ;

- 930 photographies dans le jeu de test (environ 15% des données), dont 663 de tumeurs bénignes et 267 de tumeurs malignes.

d. Autres prétraitements des données

Nous avons appliqué des prétraitements supplémentaires aux photographies de notre jeu de données, nécessaires pour certains, préférables pour d'autres. L'ensemble de ces prétraitements ont été réalisés en *Python*.

Nous avons choisi de redimensionner les photographies de haute qualité dans une résolution plus basse. Cela présente l'avantage de réduire le temps nécessaire à l'entraînement des modèles ainsi qu'à leurs prédictions. A l'aide de la librairie *Pillow* (version 8.4.0), les photographies du jeu de données ont été ajustées au format 224 pixels sur 224 pixels (*figure 145*) [901]. Nous avons opté pour un redimensionnement sans rognage donc permettant une altération des dimensions, cela présente l'avantage de rendre notre modèle plus adaptable aux différents formats de photographies qu'il reçoit.



Figure 145. Redimensionnement des photographies. (a) Boucle de redimensionnement (codée en *Python*). (b) Exemple de redimensionnement réalisé sur une photographie de tumeur cutanée bénigne. (Photographie provenant des archives de l'International Skin Imaging Collaboration)

Nous avons également opéré de l'augmentation de données sur notre jeu d'entraînement afin de réduire le sur-apprentissage (*figure 147*). En vision par ordinateur, l'augmentation de données consiste à appliquer aux images un ensemble de modifications plus ou moins aléatoires telles que des rotations ou des inclinaisons. L'objectif de ce traitement est d'accroître l'efficacité de l'entraînement en exposant les modèles entraînés à des variantes des images d'origine. Ainsi, les modèles acquièrent une meilleure adaptabilité vis-à-vis des différentes prises de vue [902]. A titre d'exemple, en appliquant une modification de type rotation et inclinaison, nous pouvons obtenir le type de résultat visible à droite du schéma ci-dessous (*figure 146*).

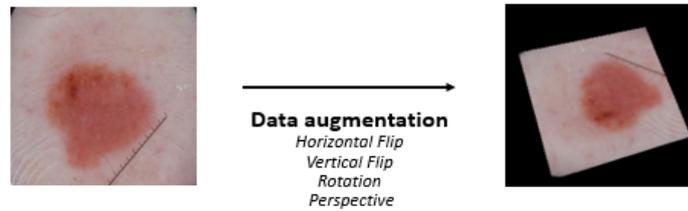


Figure 146. Exemple d'augmentation de données réalisé sur une photographie de tumeur cutanée bénigne (Photographie provenant des archives de l'International Skin Imaging Collaboration)

Aussi, nous avons converti ces photographies en tenseurs et nous leur avons appliqué une normalisation (figure 147). Notre normalisation est identique à celle appliquée aux données ayant entraîné les réseaux de neurones à partir desquels nous réaliserons par la suite des apprentissages par transfert (voir section P4.II.2.a.). Selon la documentation de *PyTorch*, la normalisation doit être effectuée selon les valeurs suivantes : moyenne = [0.485, 0.456, 0.406], écart-type = [0.229, 0.224, 0.225] [903].

```

train_set = torchvision.datasets.ImageFolder('data2/data/train/',
    transform=torchvision.transforms.Compose([
        transforms.RandomHorizontalFlip(),
        transforms.RandomVerticalFlip(),
        transforms.RandomRotation(degrees=(0, 180)),
        transforms.RandomPerspective(distortion_scale=0.6, p=1.0),
        transforms.ToTensor(),
        transforms.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225))
    ]))

```

Figure 147. Importation, augmentation, transformation en tenseurs et normalisation (codées en *Python*) des données d'entraînement

La transformation en tenseur et la normalisation ont également été appliquées aux jeux de validation et de test (figure 148). En revanche, ces deux jeux de données n'ont pas été soumis à l'augmentation de données puisque ceux-ci interviennent uniquement dans l'évaluation des modèles.

```

validation_set = torchvision.datasets.ImageFolder('data2/data/val/',
    transform=torchvision.transforms.Compose([
        torchvision.transforms.ToTensor(),
        transforms.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225))
    ]))

test_set = torchvision.datasets.ImageFolder('data2/data/test/',
    transform=torchvision.transforms.Compose([
        torchvision.transforms.ToTensor(),
        transforms.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225))
    ]))

```

Figure 148. Importation, transformation en tenseurs et normalisation (codées en *Python*) des données de validation et de test

La transformation en tenseur et la normalisation devaient aussi être appliquées aux photographies destinées à être interprétées par notre modèle (données d'entrée utilisateur) (*figure 149*). Nous avons également choisi d'appliquer un redimensionnement en 224x224 pixels, de façon à ce que la photographie importée soit, avant l'inférence, remise aux mêmes résolution et dimension que les photographies ayant entraîné notre modèle (*figure 149*).

```
imported_normalized = torchvision.datasets.ImageFolder('imported/',
    transform=torchvision.transforms.Compose([
        transforms.Resize((224,224)),
        torchvision.transforms.ToTensor(),
        transforms.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225))
    ]))
```

Figure 149. Importation, redimensionnement, transformation en tenseurs et normalisation (codés en *Python*) des données d'entrée utilisateur

2. Construction du modèle d'intelligence artificielle

a. Méthode d'apprentissage

Les prédictions réalisées par notre modèle devaient s'appuyer sur de la reconnaissance d'image, une sous-catégorie de la vision par ordinateur [904]. Nos recherches nous ont amenés à considérer l'apprentissage par transfert comme approche à privilégier. Notre méthode consiste donc à importer un réseau de neurones préalablement entraîné pour des tâches plus ou moins spécifiques, puis de l'entraîner à nouveau sur les 4637 images labellisées de notre jeu d'entraînement afin que les poids s'ajustent à notre problématique (*figure 35*). Cette méthode de plus en plus populaire est réputée comme étant particulièrement efficace dans le cadre de la classification d'image [905]. En l'occurrence, les réseaux de neurones pré-entraînés qui promettent les meilleures performances en vision par ordinateur sont les réseaux convolutifs du *Visual Geometry Group* (VGG) et les réseaux neuronaux résiduels (ResNet, *residual neural network*) [906,907,908,909].

Avant leur réentraînement par rétropropagation du gradient de l'erreur, nous avons doté tous ces réseaux de neurones d'une couche de sortie composée de deux neurones, chacun correspondant à l'une des classes (tumeur bénigne/tumeur maligne). En revanche, le nombre, la composition, la structure et la nature des couches d'entrée et des couches intermédiaires variaient selon le réseau de neurones importé.

b. Choix des métriques

Concernant les métriques à employer, nous avons opté pour la sensibilité et la spécificité. Il s'agit des métriques usuelles lorsque l'évaluation porte sur la performance d'un test diagnostique (voir section P3.II.2.a.). En outre, nous avons choisi de privilégier la spécificité en nous basant sur les bonnes pratiques. En effet, il convient de privilégier la spécificité lorsque le but de la démarche est d'affirmer ou de confirmer un diagnostic [910]. Aussi, nous avons tenu compte des risques encourus dans le cas où la sensibilité serait privilégiée, c'est-à-dire un nombre potentiellement conséquent de faux positifs pouvant provoquer une sur-consultation et donc des dépenses publiques non nécessaires ainsi qu'un accroissement des difficultés d'accès aux soins en dermatologie [911].

Malgré notre choix de privilégier la spécificité, nous voulions construire un modèle présentant une sensibilité acceptable. De fait, il nous fallait employer une métrique d'arbitrage entre la sensibilité et la spécificité. En l'occurrence, l'exactitude (*accuracy*) et l'aire sous la courbe de ROC sont classiquement de bonnes métriques d'arbitrage si le jeu d'entraînement est équilibré, c'est-à-dire si ses différentes classes présentent un nombre identique (ou proche) d'observations. Dans le cas contraire, l'arbitrage se voit être influencé par le déséquilibre de classes [912]. Si le nombre d'images de tumeurs bénignes est plus important, la spécificité aura tendance à être favorisée par l'arbitrage. De même, si le nombre d'images de tumeurs malignes est plus important, c'est la sensibilité qui aura tendance à être favorisée par l'arbitrage [742]. Dans notre cas, l'exactitude se présente comme étant une métrique pertinente puisque notre jeu de données comporte davantage d'images de tumeurs bénignes et que nous souhaitons privilégier la spécificité.

Nous avons également exploré une autre alternative faisant intervenir une métrique permettant de s'affranchir de ce déséquilibre, à savoir une moyenne harmonique entre la sensibilité et la spécificité :

$$\text{Moyenne harmonique} = \frac{2 \times \text{Sensibilité} \times \text{Spécificité}}{\text{Sensibilité} + \text{Spécificité}}$$

c. Entraînement et validation des modèles

Une fois notre méthode d'apprentissage et nos métriques déterminées, nous pouvions dès lors entraîner nos modèles. C'est en *Python*, en l'occurrence à l'aide de la librairie *Pytorch* (version 1.8.1), qu'ont été menées l'ensemble des démarches d'apprentissage (importation des réseaux de neurones et entraînements) et d'évaluation (validation et évaluation finale) [913].

La phase de validation que nous avons mis en œuvre est classique. Celle-ci consistait à entraîner, à partir du jeu d'entraînement et en faisant varier divers hyperparamètres, plusieurs réseaux de neurones appartenant à la famille des VGG et des ResNet. En évaluant ces modèles à partir du jeu de validation, nous avons pu ajuster ces hyperparamètres et identifier les réseaux de neurones les plus performants [914].

Les hyperparamètres que nous avons fait varier sont les suivants (*voir section P1.III.2.e.e.3.*) [342] :

- Le taux d'apprentissage ;
- La fonction d'erreur ;
- Le nombre d'*epochs* ;
- Le décrochage (*dropout*) [355] ;
- L'optimiseur ;
- La normalisation des lots (*batch normalization*) ;
- Le nombre d'observations par lot (*batch size*) ;
- Le pas du planificateur de taux d'apprentissage (*scheduler step size*) [915] ;
- Le facteur de multiplication du planificateur de taux d'apprentissage (*scheduler gamma*) [915] ;
- La métrique privilégiée par l'optimiseur.

Pour certains de ces hyperparamètres, nous avons pu identifier des valeurs idéales quel que soit le réseau de neurones importé :

- Optimiseur : *Adaptive Moment Estimation (Adam)* ;
- Fonction d'erreur : entropie croisée (*cross-entropy loss*) ;
- Décrochage : 0,5 ;
- Pas du planificateur : 1 ;
- Facteur de multiplication du planificateur : 0,5.

En faisant varier le réseau de neurones importé et les autres hyperparamètres, les performances que nous avons pu atteindre sont diverses (*tableau 2*).

Tableau 2. Résultats les plus significatifs obtenus à partir du jeu de validation

Modèle chargé	Norm. des lots	Taux d'appr.	Taille des lots	Nombre d'epochs	Métrique privilégiée	Exactitude	MH	Sensibilité	Spécificité	AUC
<i>ResNet-18</i>	Non	0,001	32	15	Exactitude	84,58%	82,15%	75,56%	90,00%	91,89%
<i>ResNet-18</i>	Non	0,001	32	15	MH	80,42%	81,09%	84,44%	78,00%	88,48%
<i>ResNet-34</i>	Non	0,001	32	15	Spécificité	69,37%	58,62%	20,56%	98,67%	91,21%
<i>ResNet-34</i>	Non	0,001	32	25	Exactitude	81,67%	79,89%	77,78%	84,00%	88,84%
<i>ResNet-50</i>	Non	0,001	32	25	Exactitude	81,46%	78,73%	72,78%	86,67%	88,80%
<i>ResNet-101</i>	Non	0,001	32	25	Exactitude	82,29%	79,28%	72,22%	88,33%	90,14%
<i>ResNet-152</i>	Non	0,001	16	25	Exactitude	78,75%	77,58%	73,89%	81,67%	85,02%
<i>VGG-11</i>	Non	0,001	32	25	Exactitude	75,00%	72,53%	66,11%	80,33%	81,83%
<i>VGG-11</i>	Oui	0,001	16	25	Exactitude	78,75%	73,91%	63,89%	87,67%	85,56%
<i>VGG-13</i>	Non	0,001	32	25	Exactitude	74,38%	74,79%	76,67%	73,00%	81,10%
<i>VGG-13</i>	Oui	0,001	16	25	Exactitude	84,38%	83,64%	81,11%	86,33%	91,12%
<i>VGG-16</i>	Non	0,001	32	25	Exactitude	77,50%	77,76%	78,89%	76,67%	85,70%
<i>VGG-16</i>	Oui	0,001	16	25	Exactitude	80,83%	76,35%	66,67%	89,33%	89,96%
<i>VGG-19</i>	Non	0,0001	32	25	Exactitude	81,46%	80,32%	76,67%	84,33%	89,41%
<i>VGG-19</i>	Oui	0,001	16	25	Exactitude	79,17%	78,19%	75,00%	81,67%	87,22%

Nous pouvons constater que lorsque l'optimiseur privilégie la spécificité, la sensibilité est drastiquement affectée. C'est par exemple le cas du modèle *ResNet-34* (troisième ligne du tableau 2) qui dispose de la meilleure spécificité mais de la sensibilité la plus faible. En outre, lorsque l'optimiseur privilégie la moyenne harmonique (MH), la sensibilité est modérément plus élevée que la spécificité. De fait, les résultats les plus satisfaisants sont obtenus lorsque l'exactitude est privilégiée.

En définitive, nous avons retenu les deux modèles surlignés, qui présentent une bonne spécificité ainsi qu'une sensibilité acceptable : *ResNet-18* (spécificité = 90%, sensibilité = 75,56%) et *VGG-13* (spécificité = 86,33%, sensibilité = 81,11%) (tableau 2).

d. Evaluation finale

En tenant compte de la loi de Goodhart, nous avons intégré la méthode du *hold-out* à notre évaluation (voir section P3.II.4.a.). Il convenait donc d'évaluer à nouveau les deux modèles retenus à partir d'un jeu de données n'ayant pris part ni à leur entraînement ni à leur validation, en l'occurrence notre jeu de test (tableau 3) [872].

Tableau 3. Résultats obtenus à partir du jeu de test

Modèle chargé	Norm. des lots	Taux d'appr.	Taille des lots	Nombre d'epochs	Métrique privilégiée	Exactitude	MH	Sensibilité	Spécificité	AUC
<i>ResNet-18</i>	Non	0,001	32	15	Exactitude	80,22%	70,76%	58,80%	88,84%	82,92%
<i>VGG-13</i>	Oui	0,001	16	25	Exactitude	77,96%	71,18%	61,42%	84,62%	82,53%

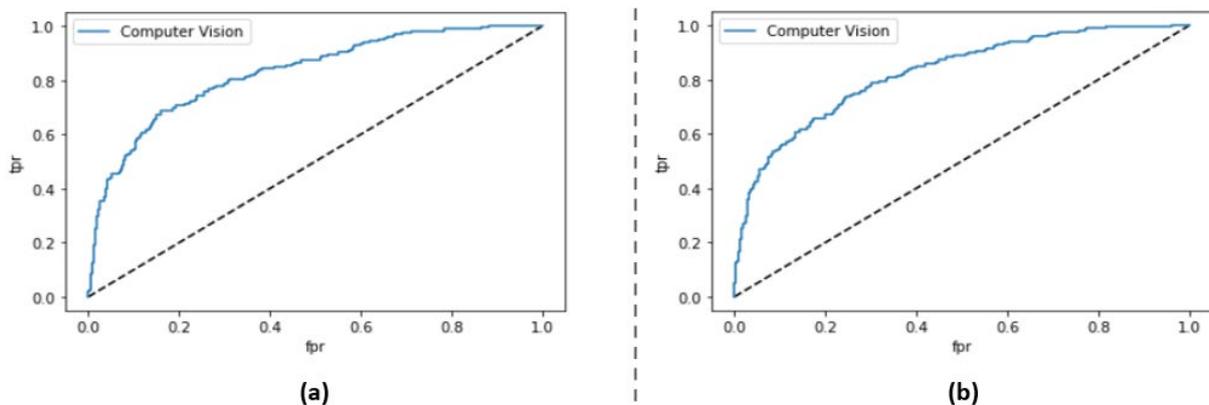


Figure 150. Courbes de ROC des deux modèles retenus à l'issue de la phase de validation (a : *ResNet-18*, b : *VGG-13*)

Les deux modèles évalués présentent des performances semblables, ce qui est illustré par la similarité de leurs courbes de ROC (figure 150). Nous avons finalement sélectionné le modèle *ResNet-18* car celui-ci dispose d'une spécificité plus élevée. Le modèle retenu présente donc une spécificité et une sensibilité s'élevant respectivement à 88,84% et 58,80% (tableau 3).

3. Intégration du modèle au sein d'une architecture

a. Infrastructure de développement

Afin de rendre son utilisation possible pour l'utilisateur, nous devons intégrer notre modèle au sein d'une architecture. Celle-ci devait comporter une interface utilisateur (développement frontal ou *front-end*) et permettre la mise en œuvre des fonctionnalités nécessaires au fonctionnement de notre application (développement dorsal ou *back-end*) [916]. Parmi les nombreuses infrastructures permettant le développement d'applications (*framework*), notre choix s'est porté sur *Dash* (version 2.5.1) car celui-ci présente les avantages suivants [917] :

- *Dash* permet la création d'interfaces interactives entièrement en *Python*, facilitant ainsi son intégration à notre environnement de développement.
- *Dash* permet le développement d'applications web. Celles-ci peuvent être utilisées sans nécessiter aucune installation et sont accessibles depuis n'importe quel terminal disposant d'un navigateur web et d'une connexion à internet.
- *Dash* permet l'emploi de composants *Bootstrap*. Ceux-ci assurent la responsivité de l'application développée, c'est-à-dire sa capacité à s'adapter à la taille de l'écran du terminal y accédant.

b. Fonctionnalités majeures

Notre application devait être en mesure de proposer une prédiction s'appuyant sur le modèle retenu et sur une photographie importée. Nous devons donc implémenter, tout au moins, des fonctionnalités permettant à l'application d'avoir accès à la photographie puis de fournir celle-ci au modèle. Pour ce faire, l'application devait être en capacité de charger une image soit via une photographie prise à partir d'un périphérique d'acquisition d'image soit via un fichier présent sur le périphérique de stockage (*figure 151*).

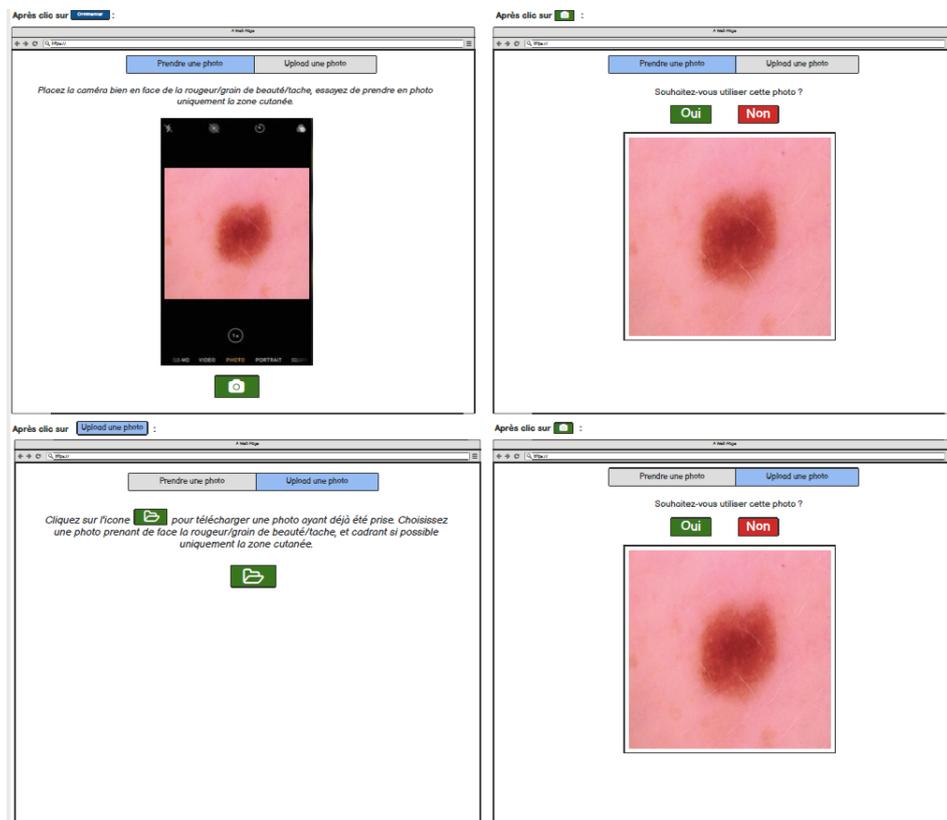


Figure 151. Extrait de la maquette de *Méla-Safe* réalisée avant son développement : importation d'une photographie (*Balsamiq*) (Photographie provenant des archives de l'*International Skin Imaging Collaboration*)

La solution que nous avons retenue afin d'implémenter cette fonctionnalité est le composant *Dash* « *dcc.Upload* ». Ce composant permet à l'application d'importer un fichier quelconque sous la forme d'une série de caractères. Cette série de caractères correspond à un encodage en Base64 du contenu du fichier importé (*figure 152*) [918].

Figure 152. Exemple de codage d'un fichier jpeg en Base64 exécuté par le composant « *dcc.Upload* ».
Remarque : l'encart affiché ici présente uniquement les premiers caractères de la série. (*Photographie provenant des archives de l'International Skin Imaging Collaboration*)

En l'état, cette série de caractères ne pouvait pas être prise en charge par notre modèle. Il était nécessaire de supprimer le début de cette série de caractères (dans le cas d'un fichier *jpeg*, « *data:image/jpeg;base64,* ») et de convertir le reste de cette série en un fichier image exploitable par notre modèle. Nous avons créé la fonction « *get_image_and_write_it* » chargée de réaliser ces actions (*figure 153*). Cette fonction est conçue pour prendre en entrée l'objet « *contents* » (correspondant à la série de caractères Base64 mentionnée ci-dessus) et renvoyer en sortie un fichier image en son format d'origine. Chaque format de fichier à prendre en charge devait donc être prévu au sein de notre fonction. Pour le moment, celle-ci prend uniquement en charge les fichiers *jpeg* et *png*.

```
def get_image_and_write_it(contents) :
    if 'data:image/jpeg;base64,' in contents :
        contents = contents[len('data:image/jpeg;base64,'):]
        imgdata = base64.b64decode(contents)

    if 'data:image/png;base64,' in contents :
        contents = contents[len('data:image/png;base64,'):]
        imgdata = base64.b64decode(contents)

    filename = 'imported/New_photo/photo.jpg'
    with open(filename, 'wb') as f:
        f.write(imgdata)
```

Figure 153. Fonction de décodage d'une série de caractère en Base64 vers un fichier *jpeg* ou *png* (codée en Python)

Une fois la photographie reconstruite et prête à l'emploi, notre application devait être en mesure de faire appel à notre modèle et d'exécuter ses inférences. Nous avons programmé ces actions par le biais de deux fonctions distinctes :

- La fonction « *load_model_and_predict* » ayant pour rôle de charger notre modèle et d'amorcer l'inférence (figure 154). Cette fonction renvoie en sortie le résultat de la fonction « *get_prediction* ».

```
def load_model_and_predict(path):
    ResNet18_Comp = models.resnet18(pretrained=True)
    ResNet18_Comp.fc = Classifier()
    device = torch.device('cpu')
    model = ResNet18_Comp
    model.load_state_dict(torch.load("chosen_model.pth", map_location=device))
    model.eval()
    return get_prediction(model)
```

Figure 154. Fonction de chargement du modèle et d'initiation de l'inférence (codée en Python)

- La fonction « *get_prediction* » ayant pour rôle d'appliquer à la photographie les transformations nécessaires (redimensionnement, transformation en tenseur, normalisation) puis d'effectuer l'inférence à partir de celle-ci (figure 155). Cette fonction prend en entrée le modèle chargé et renvoie en sortie la prédiction effectuée par celui-ci.

```
def get_prediction(model):
    imported = torchvision.datasets.ImageFolder('imported/',
        transform=torchvision.transforms.Compose([
            torchvision.transforms.Resize((224,224)),
            torchvision.transforms.ToTensor(),
            torchvision.transforms.Normalize(mean=(0.485, 0.456, 0.406), std=(0.229, 0.224, 0.225))
        ]))
    x, y = observation_extraction(imported, 1)
    model.eval()
    prediction = model(x.view(1,3,224,224))
    return int(prediction.argmax(1)[0])
```

Figure 155. Fonction de préparation de la donnée d'entrée et d'inférence (codée en Python)

Enfin, nous avons programmé une fonction de rappel (ou *callback*) chargée de faire le lien entre l'objet « *contents* », notre modèle et les trois fonctions présentées ci-dessus (figure 156).

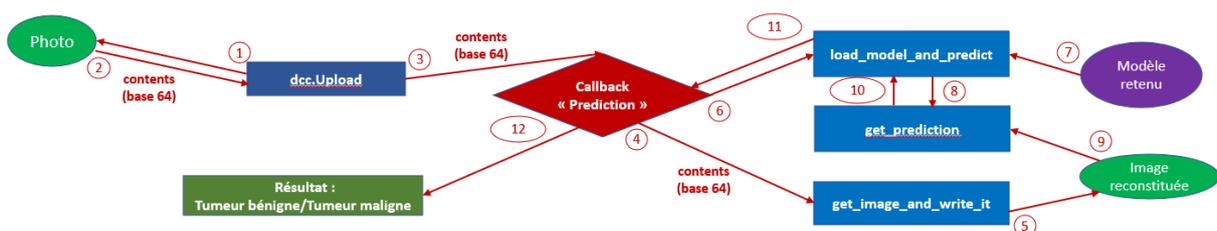


Figure 156. Fonction de rappel dédiée à la prédiction

c. Interface et expérience utilisateur

Pour qu'un utilisateur puisse interagir avec notre application sans que cela ne nécessite des connaissances spécifiques en informatique, ces fonctionnalités doivent être associées à une interface graphique. Afin d'optimiser son esthétique et son ergonomie, nous avons élaboré l'interface de notre application en tenant compte des recommandations propres à l'UI (*User Interface*) et l'UX (*User Experience*) design. Nous nous sommes notamment appuyés sur des lois communément appliquées à l'*UX design* : loi de Miller, loi de Jakob, loi de Hick, loi de Fitts [919]. Nous avons également adopté une approche *design thinking* que nous pourrions résumer en cinq principes essentiels : l'empathie vis-à-vis de l'utilisateur, la définition du problème, la recherche de solution, son prototypage et son expérimentation [920]. Nous espérons qu'ainsi notre interface permette un parcours utilisateur accommodant et intuitif.

Nous avons choisi de subdiviser notre interface en trois pages distinctes. Nous souhaitons dédier la première page à la présentation de l'application. La page suivante devait permettre l'importation d'une photographie et l'exécution de la prédiction. Enfin, la troisième et dernière page devait présenter le résultat obtenu à l'issue de l'inférence du modèle ainsi que les recommandations adaptées. Pour concrétiser cela, nous avons besoin d'intégrer les composants *Bootstrap* suivants :

- Des onglets en guise de pages (*dbc.Tabs, dbc.Tab*) [921] ;
- Une structure de mise en page (*dbc.Col, dbc.Row*) [922] ;
- Des zones de texte (*html.H1, html.H5, html.P*) [923] ;
- Des boutons cliquables (*dbc.Button*) [924] ;
- Un menu déroulant permettant un choix parmi plusieurs possibilités (*dbc.DropdownMenu, dbc.DropdownMenuItem*) [925] ;
- Une alerte qui s'active lorsque l'utilisateur tente de lancer la prédiction avant d'importer un fichier conforme (*dbc.Alert*) [926] ;
- Une zone d'affichage d'image (*dbc.Card, dbc.CardImg*) [927] ;
- Un indicateur de chargement qui s'affiche pour indiquer une action en cours de traitement (*dbc.Spinner*) [928].

Nous avons également programmé des fonctions de rappel (*app.callback*) (*figure 157*) [929] :

- La fonction de rappel « Home » faisant le lien entre le bouton cliquable « Revenir à l'accueil » et la page d'accueil.
- La fonction de rappel « Start » faisant le lien entre le bouton cliquable « Lancer l'application » et la page de prédiction.

- La fonction de rappel « Prediction » dont nous avons déjà décrit le rôle (voir section P4.II.3.b.). Celle-ci s'active lorsque l'utilisateur clique sur le bouton « Obtenir ma prédiction » et mène, en dernier lieu, à la page de résultat.

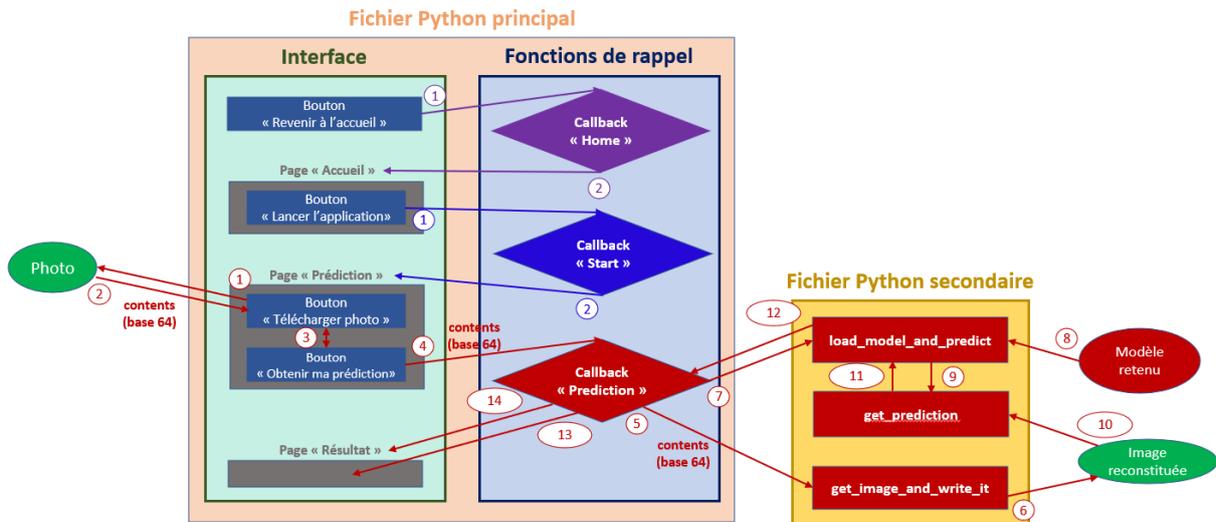


Figure 157. Vue d'ensemble de l'architecture de *Méla-Safe*. Le fichier *Python* principal comporte les fonctions de rappel ainsi que l'ensemble des composants nécessaires à l'interface tandis que le fichier *Python* secondaire comporte les fonctions ordinaires mobilisées par la fonction de rappel « Prediction ».

- La fonction de rappel « Language » qui permet à l'utilisateur de changer le contenu des zones de texte (en français/en anglais) en cliquant sur les items « Français » ou « English » du menu déroulant « Change the language » (figure 158).

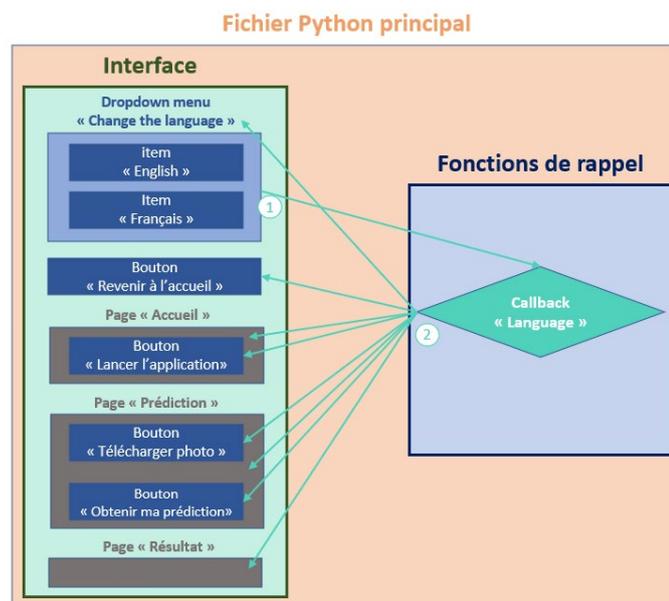


Figure 158. Fonction de rappel dédiée au changement de langue

L'emploi des composants *Bootstrap* permet à notre interface d'être responsive. Celle-ci est donc fonctionnelle aussi bien sur un navigateur pour ordinateur que pour smartphone et quelles que soient la taille et les dimensions de l'écran ou de la fenêtre de navigation.

d. Parcours utilisateur

Lorsque l'utilisateur accède à l'application, celle-ci se lance sur la page d'accueil (*figure 159*). Cette page contient le nom de l'application, une présentation succincte ainsi qu'un bouton « Lancer l'application ».

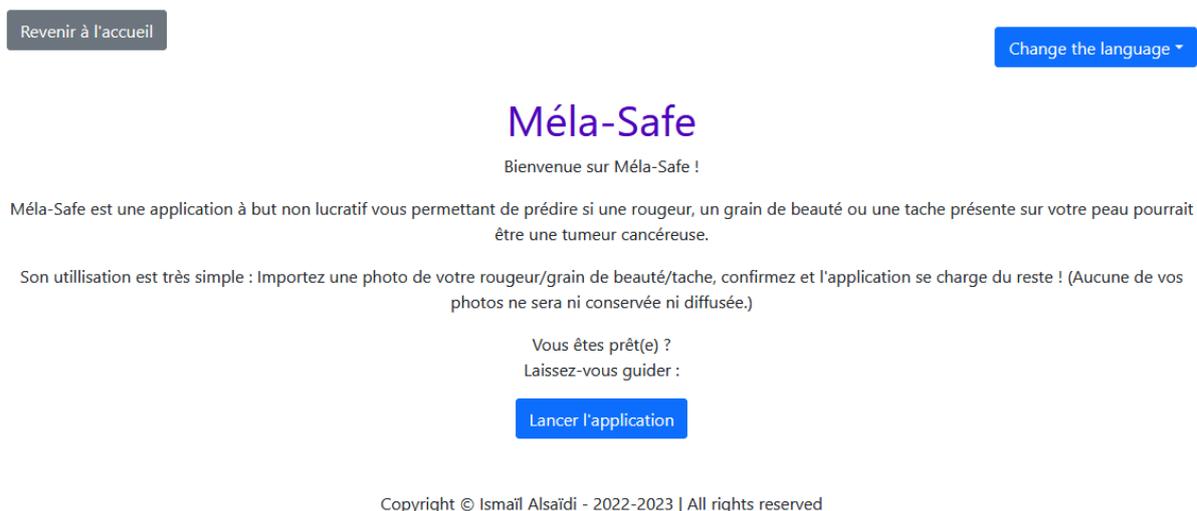


Figure 159. Page d'accueil (*Méla-Safe*)

En cliquant sur le bouton « Lancer l'application », l'application bascule sur la page de prédiction (*figure 160*). Celle-ci comporte deux courtes phrases d'instruction invitant l'utilisateur à uploader une photographie. Cette page affiche les boutons « Télécharger photo » et « Obtenir ma prédiction ».



Figure 160. Page de prédiction (*Méla-Safe*)

En cliquant sur “Télécharger photo” (composant *dbc.Button* niché dans un composant *dcc.Upload*), l'utilisateur peut sélectionner la photographie qu'il souhaite soumettre au modèle. A partir d'un navigateur pour ordinateur, seuls les fichiers présents sur un périphérique de stockage peuvent être uploadés (*figure 161*).

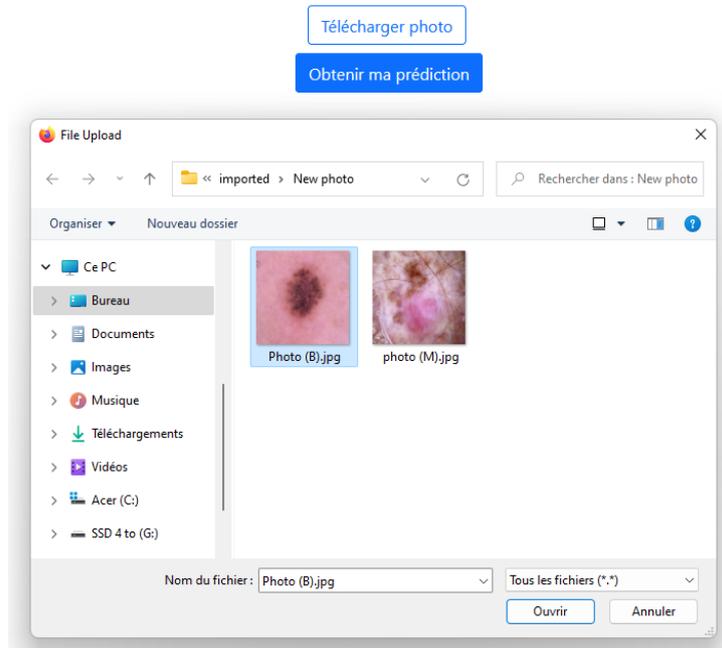


Figure 161. Importation d'un fichier à partir d'un périphérique de stockage (navigateur pour ordinateur)
(Photographie provenant des archives de l'International Skin Imaging Collaboration)

A partir d'un navigateur pour smartphone, l'utilisateur a la possibilité d'uploader un fichier présent sur l'espace de stockage ou de capturer la photographie à partir de la caméra (*figure 162*).

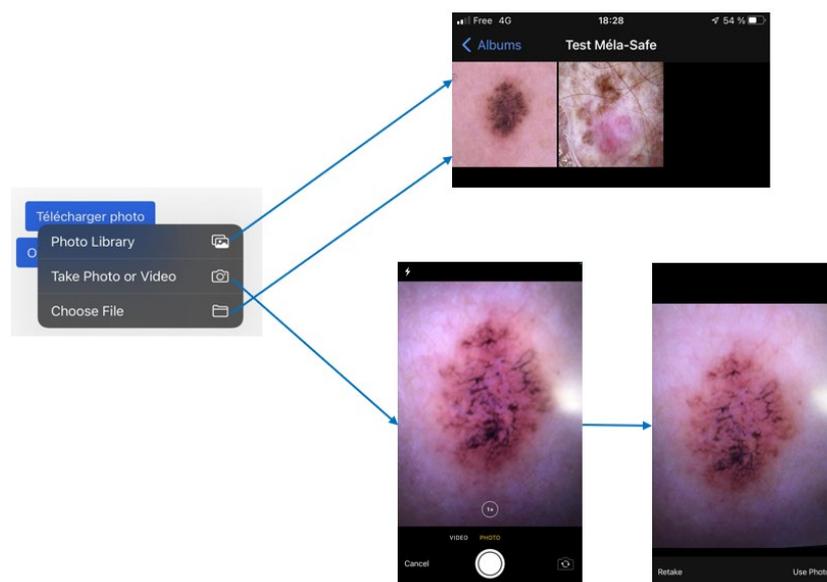


Figure 162. Menu déroulant d'importation de fichier à partir d'un périphérique de capture ou de stockage (navigateur pour smartphone)
(Photographie provenant des archives de l'International Skin Imaging Collaboration)

Dans le cas où l'utilisateur upload un fichier dans un format non pris en charge (autre que *jpeg* ou *png*) ou qu'il clique sur "Obtenir ma prédiction" sans avoir préalablement uploadé un fichier, nous avons prévu l'affichage d'un message d'alerte (composant *dbc.Alert*) (figure 163).

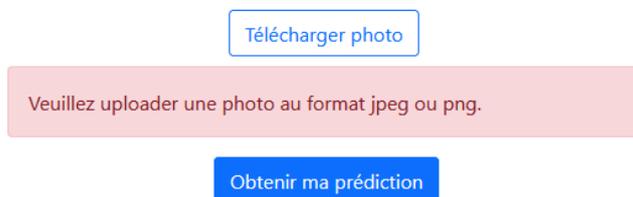


Figure 163. Message d'alerte consécutif à la tentative de prédiction sans upload préalable d'un fichier conforme

À l'inverse, si un fichier est bien uploadé et que son format est pris en charge, l'utilisateur peut dès lors déclencher la fonction de rappel « Prediction » en cliquant sur le bouton « Obtenir prédiction ». L'application reconstitue alors l'image uploadée à partir de sa version Base64, charge le modèle et procède à la prédiction. Durant ces opérations, la page affiche un indicateur de chargement de type *spinner* ainsi qu'une zone de texte indiquant le chargement en cours (figure 164).



Figure 164. Page de prédiction affichant un indicateur de chargement (*Méla-Safe*)

Les actions effectuées par la fonction de rappel « Prediction » s'achèvent avec le basculement vers la page de résultat. Celle-ci affiche le résultat fourni par le modèle ainsi que des recommandations selon que la tumeur soit prédite comme étant bénigne ou maligne (figure 165). Aussi, la photographie fournie par l'utilisateur est affichée à droite du résultat.

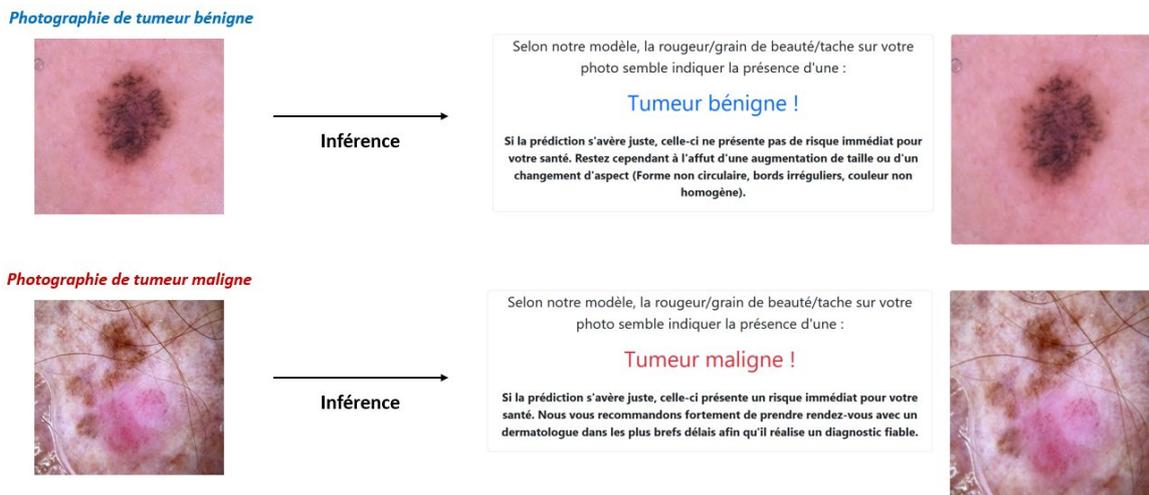


Figure 165. Résultats affichés selon que la photographie ait donné lieu à une prédiction de tumeur bénigne ou de tumeur maligne (Photographie provenant des archives de l'International Skin Imaging Collaboration)

Quel que soit le résultat obtenu, la page de résultat affiche un message d'avertissement visant à mettre en garde l'utilisateur concernant la fiabilité des prédictions émises par notre modèle et incitant à se fier uniquement à un diagnostic émis par un médecin (figure 166).



Malgré les efforts déployés et les progrès réalisés dans le domaine des intelligences artificielles, les prédictions réalisées par notre modèle restent très imparfaites. Les prédictions obtenues ne sont pas en mesure de se substituer à un diagnostic médical. En cas de doute ou si l'application prédit la présence d'une tumeur maligne, nous vous recommandons vivement de consulter un médecin pour obtenir un diagnostic fiable.

Figure 166. Message d'alerte relatif à la fiabilité des prédictions

L'utilisateur peut, à tout moment du parcours, revenir à la page d'accueil en cliquant sur le bouton « Revenir à l'accueil » ou modifier la langue affichée par l'application via le menu déroulant « Change the language » (figure 167).

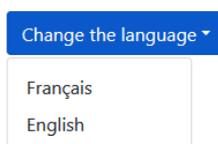


Figure 167. Menu déroulant de changement de langue

e. Gestion des vulnérabilités

Le déploiement d'une application web expose ses utilisateurs à un certain nombre de risques. Ces risques sont inhérents aux vulnérabilités de l'application pouvant être la cible d'attaques de différentes natures [930]. Nous devons faire preuve de vigilance vis-à-vis de ces risques et prendre les dispositions nécessaires. Notre gestion des vulnérabilités s'articule autour de six dimensions majeures.

La première d'entre elles est la confidentialité. Notre application est conçue pour que l'accès à une prédiction ne soit autorisé qu'à partir d'interactions spécifiques (bouton « Obtenir ma prédiction »). De plus, un utilisateur ne peut avoir accès qu'aux prédictions relatives aux images qu'il soumet à l'application. Aussi, l'accès à notre application est sécurisé par un protocole HTTPS (*Hypertext Transfer Protocol Secure*), garantissant ainsi un cryptage des données échangées entre le navigateur et le serveur. Cela protège les utilisateurs des attaques de type « homme du milieu » [931].

Le second aspect auquel nous avons prêté attention est l'intégrité. Si un acteur malveillant parvenait à détériorer le modèle ou toute autre fonctionnalité intervenant dans la prédiction, la disponibilité ou la qualité des inférences pourraient être altérées. Nous devons donc protéger l'application contre les modifications non autorisées. Pour assurer son intégrité, il est capital d'héberger notre application sur une plateforme fiable, mettant à jour régulièrement ses serveurs et ses mesures de protection contre les attaques [932].

Nous avons également axé nos efforts sur la fiabilité de notre application. Pour nous en assurer, nous avons mis au point des tests unitaires afin de contrôler le bon fonctionnement des différentes fonctionnalités au cours du développement. Le choix d'un hébergeur sûr et performant joue également un rôle dans la fiabilité, la plateforme d'hébergement doit notamment être en mesure de protéger ses serveurs contre les attaques par déni de service (DDoS, *Distributed Denial of Service*) [933,934].

Le quatrième aspect que nous avons observé est la maintenabilité. En cas de vulnérabilité décelée au sein de notre code, notre application doit pouvoir aisément faire l'objet d'une correction. Pour ce faire, nous avons suivi les bonnes pratiques relatives à l'écriture du code telles que la modularité, la lisibilité, la documentation et le versionnage [935].

La traçabilité a également fait partie de nos préoccupations majeures. Nous avons donc choisi un hébergeur mettant à disposition les historiques (*logs*) des événements survenus sur le serveur et sur notre application. De plus, nous avons mis en place une surveillance de l'activité de l'application via un service de monitoring en continu [936].

Enfin, nous nous efforçons d'assurer une veille de sécurité, en particulier vis-à-vis des composants de notre application. Les composants tiers peuvent être une source importante de vulnérabilités. Maintenir à jour ces composants permet de disposer des patches de sécurité les plus récents [937].

4. Bilan et axes d'amélioration

Nous avons pour objectif de développer une application capable de détecter les tumeurs cutanées malignes. Nous avons pu mener à bien ce projet par le biais d'outils efficaces et bien documentés tels que *PyTorch*, *Dash* et *Bootstrap*. En entraînant notre modèle à partir d'un jeu de données riche et hétérogène, nous avons pu atteindre une spécificité de 88,84% sans excessivement compromettre la sensibilité. Ainsi, notre application est en mesure de proposer à ses utilisateurs des prédictions relativement fiables. Nous ne pouvons pas affirmer avec certitude que *Méla-Safe* est en mesure de contribuer significativement à diminuer les complications liées aux cancers cutanés, mais nous pensons que celle-ci réunit les qualités nécessaires pour y prétendre. Quoiqu'il en soit, notre rendu final répond aux standards que nous nous étions fixés en proposant des prédictions instantanées concernant la malignité d'une tumeur cutanée à partir d'une simple photographie, à tout moment et quel que soit le terminal accédant à l'application dès lors que celui-ci dispose d'un navigateur web et d'une connexion internet.

Nous envisageons d'ores et déjà des améliorations, notamment des fonctionnalités que nous aimerions implémenter à l'avenir :

- La possibilité de rogner la photographie, ce qui serait utile dans le cas où la surface cutanée photographiée est mal cadrée. Une alternative serait d'implémenter un détournement automatique effectué par suivi de mouvement (ou *video tracking*) [938].
- La possibilité d'effectuer plusieurs prédictions simultanément en soumettant au modèle un lot d'images.
- La possibilité pour l'utilisateur de sauvegarder ses prédictions afin de pouvoir y accéder ultérieurement, ainsi que la possibilité de confirmer ou infirmer la prédiction suite à un diagnostic médical et d'autoriser ou non l'utilisation du résultat pour améliorer le modèle (*figure 168*). Cela impliquerait d'implémenter la création de compte ainsi qu'une authentification sécurisée par mot de passe.

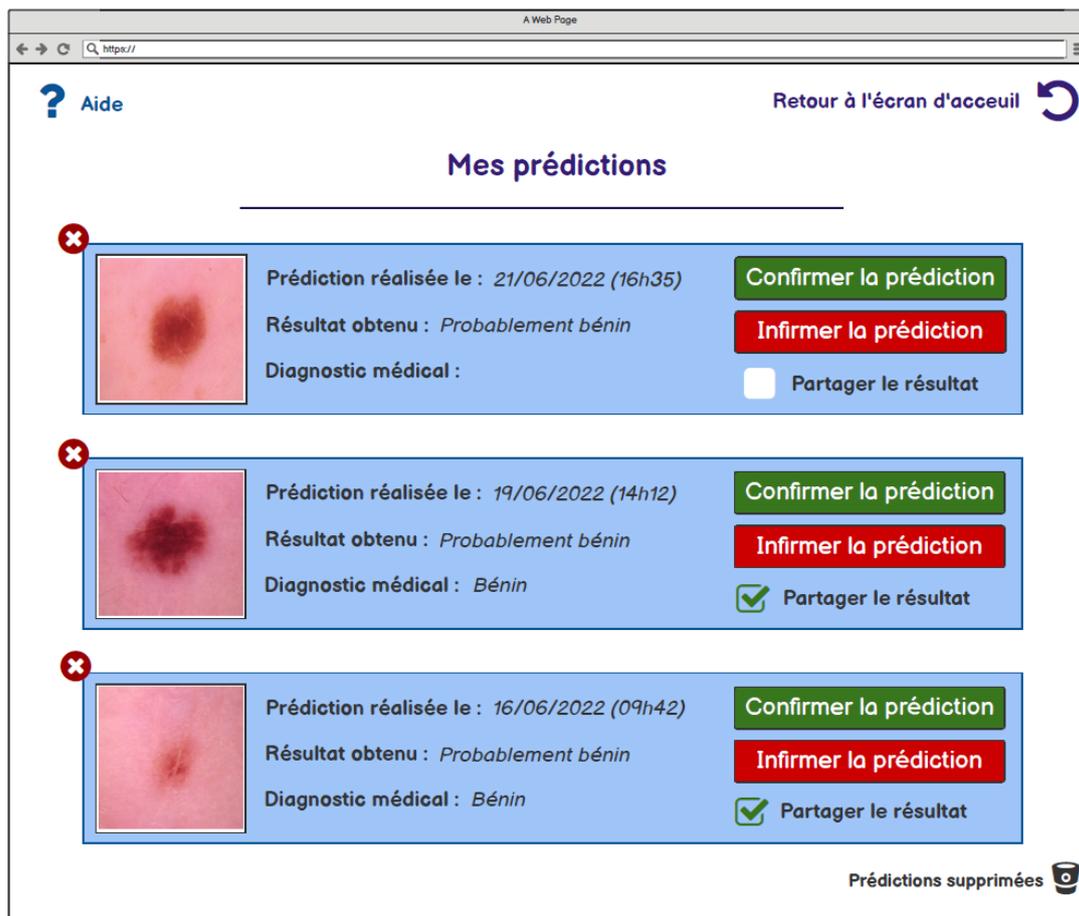


Figure 168. Extrait de la maquette de *Méla-Safe* réalisée avant son développement : page d'accès aux prédictions sauvegardées (*Balsamiq*) (Photographie provenant des archives de l'International Skin Imaging Collaboration)

Nous aimerions clarifier et assouplir l'expérience utilisateur en ajoutant les éléments suivants :

- Un grisage du bouton « Obtenir ma prédiction » avant l'achèvement de l'importation d'une photographie valide.
- L'ajout d'un bouton « Effectuer une autre prédiction » en page de résultat. Celui-ci afficherait la page de prédiction et éviterait ainsi à l'utilisateur de revenir à la page d'accueil pour chaque prédiction supplémentaire.
- L'ajout d'une limite de poids pour le fichier importé, afin d'éviter une saturation de la bande passante par un fichier trop lourd. Le poids maximal serait indiqué dans les instructions présentes en page de prédiction et ferait également l'objet d'une alerte destinée à l'utilisateur si celui-ci tente d'uploader un fichier donc le poids excède cette limite.
- Des éléments d'explicabilité (voir section P2.III.4.).

Nous pourrions également déployer notre application sous forme native, c'est-à-dire dans un format pouvant être installé et utilisé localement sur un système d'exploitation (ex. : .exe pour *Windows*, .app pour *macOS*, .apk pour *Android*).

Concernant le modèle que notre application embarque, nous pensons pouvoir l'améliorer en appliquant les changements suivants :

- Elargir notre jeu d'entraînement, en privilégiant les images de tumeurs cutanées malignes [939].
- Approfondir l'augmentation de données, en expérimentant davantage de transformations et de combinaisons [902]. Nous pourrions inclure ces variantes dans notre processus de validation.
- Elargir l'éventail de modèles que nous entraînons lors de la validation, en y intégrant par exemple les modèles de la famille *Inception* ou *EfficientNet* [940,941].
- Attribuer à notre modèle la capacité de détecter si une photographie correspond à ce qui est attendu ou non (*out-of-distribution detection*), de façon à ce que notre application puisse identifier et refuser une photographie ne comportant aucune zone cutanée [942].

Nous espérons pouvoir effectuer ces ajouts dans un avenir proche et proposer ainsi un outil d'intérêt public dont les performances et l'ergonomie sont optimisées.

III. Aspects réglementaires (Espace économique européen)

1. Réglementation des données

La fonctionnalité majeure de notre application est la prédiction, celle-ci fait intervenir des transformations et une inférence s'appliquant aux photographies importées par les utilisateurs. Ces opérations constituent un traitement de données [943]. Puisqu'il est prévu que ces photographies comportent des tumeurs cutanées, celles-ci constitueraient des données de santé le cas échéant. Il en va de même pour les prédictions qui leur sont associées [824].

Cependant, notre application est conçue pour ne stocker ni partager aucune de ces données, et pour supprimer instantanément celles-ci dès que l'utilisateur quitte ou redémarre l'application. Avant leur suppression, ces données sont anonymisées par design puisque leurs métadonnées sont instantanément effacées lors de l'importation et qu'aucune information concernant l'utilisateur n'est demandée. Ainsi, ni les utilisateurs ni les photographies importées ne sont identifiables. Concernant le modèle en lui-même, ses données d'entraînement sont également anonymisées. Le RGPD ne s'applique pas aux traitements de données de santé à l'usage exclusif de la personne dont sont issues ces données et/ou lorsque ces données sont anonymisées [833]. En outre, les jeux de données que nous avons utilisés sont soumis à une licence *Creative Commons* (CC0, BY, ou BY-NC), autorisant notamment la copie, la transformation et la

création à partir du matériel [900,944,945]. En l'état, notre application est donc conforme aux exigences réglementaires relatives aux données.

L'ajout de la possibilité de sauvegarder et/ou de transmettre le résultat (tel que décrit dans la section IV.B.4.) nécessitera une réévaluation de la conformité au RGPD. Ces fonctionnalités impliquent la conservation des photographies, des prédictions et du diagnostic médical indiqué par l'utilisateur. Or, comme nous l'avons mentionné, il s'agit de données de santé [824]. La conservation, si celle-ci s'effectue via une application web, ainsi que le partage de ces données sont soumis aux exigences du RGPD : information et recueil du consentement, possibilité de retirer son consentement à tout moment, désignation d'un(e) délégué(e) à la protection des données, analyse d'impact, etc. (*voir section P3.I.*). En l'occurrence, notre application est d'ores et déjà en conformité avec le RGPD concernant les points suivants :

- Traitement des données minimisé, spécifique et légitime [825] ;
- Mise en place de mesures de sécurité (*voir section P4.II.3.e.*) [828].

2. Mise sur le marché

Notre application est un logiciel destiné à être utilisé chez l'humain pour obtenir une prédiction d'ordre médical. Celle-ci est destinée à fournir des informations ne présentant pas un danger immédiat pour la santé de l'utilisateur. En outre, son fonctionnement dépend d'une source d'énergie autre que celle générée par le corps humain. Notre application est donc qualifiable en dispositif médical actif de classe IIa [946].

Sa mise sur le marché est conditionnée par l'apposition d'un marquage CE médical. Cela implique d'être en conformité avec les exigences énoncées par le *Règlement 2017/745* : sécurité, performance, système qualité, documentation technique, plan de surveillance post-AMM, etc. (*voir section P3.III.2.*). La conformité aux exigences de sécurité et de performance doit notamment être démontrée par le biais d'une évaluation clinique [873]. Un dossier de demande d'AMM regroupant les éléments de preuve quant à cette conformité doit ensuite être constitué et transmis à un organisme notifié [880].

En cas d'obtention de l'AMM, le plan de surveillance post-AMM validé par l'organisme notifié doit être mis en œuvre (*voir section P3.III.3.*) [879]. L'application doit également être déclarée auprès d'une autorité compétente lors de la première mise en service. Si la durée d'AMM souhaitée excède celle conférée, une réévaluation doit être prévue à l'issue de la durée de validité du certificat fourni par l'organisme notifié [881].

Conclusion

Conclusion

En dressant un état de l'art de l'intelligence artificielle appliquée au diagnostic et à la thérapeutique, nous avons mis en évidence les avancées significatives que celle-ci est en mesure d'engendrer dans ces deux domaines [947]. À ce jour, l'usage qui prédomine est celui de l'aide à la décision. Néanmoins, les applications s'étendent de plus en plus et offrent des perspectives nouvelles : médecine prédictive, médecine de précision, chirurgie assistée, rééducation assistée, suivi thérapeutique automatisé, recherche thérapeutique accélérée, etc. Les résultats issus des expérimentations indiquent que ces différents usages représentent une opportunité majeure d'améliorer l'efficacité des soins de santé et de la recherche biomédicale [375]. Sous réserve de qualité et de mutualisation, le volume sans cesse croissant de données de santé alimenté depuis de nombreuses années par les soignants pourrait de surcroît constituer le terreau fertile de ces innovations [948].

Ces applications de l'intelligence artificielle sont cependant source de préoccupations en matière d'éthique, de confidentialité et de responsabilité. Les défis majeurs qu'elles soulèvent sont typiques en apprentissage automatique [949] : biais algorithmiques, explicabilité, confidentialité et sécurité des données, responsabilité légale, etc. Ces problématiques sont d'autant plus exacerbées que les exigences relatives à la sécurité et à la performance sont élevées. C'est notamment le cas sur le territoire de l'Espace économique européen au sein duquel la réglementation qui s'applique aux intelligences artificielles médicales est celle qui régit les dispositifs médicaux [950]. Bien que ce règlement (2017/745) soit communément considéré comme exemplaire de par son degré élevé de rigueur, celui-ci ne mentionne pas explicitement les intelligences artificielles et leurs spécificités [873]. En l'occurrence, la transposition dudit règlement à une architecture embarquant de l'intelligence artificielle peut dans certains cas s'avérer complexe (ex. : apprentissage incrémental) [889]. Qui plus est, la tendance semble indiquer que la pertinence de cette transposition sera à remettre en question à mesure que les modèles d'intelligence artificielle se complexifieront [951]. De fait, nous souhaitons souligner l'importance de poursuivre les efforts de réflexion visant à établir des normes et approches spécifiques dans l'encadrement et l'évaluation des intelligences artificielles appliquées au diagnostic ou à la thérapeutique. Nous pensons que l'évolution continue des intelligences artificielles devrait être suivie de près par une évolution réglementaire également continue. Cette dernière devrait tenir compte des risques et problématiques propres aux modèles d'apprentissage automatique, notamment aux phases

d'apprentissage et de validation qui occupent une place fondamentale dans leur développement. C'est en tout cas la démarche engagée par la Commission Européenne qui prévoit d'étendre le RGPD via son projet de réglementation sur l'intelligence artificielle, l'*Artificial Intelligence Act* [952,953].

Par ailleurs, l'évolution de la réglementation doit également tenir compte des fonctions que les intelligences artificielles seraient amenées à endosser [954]. Même si leurs performances parviennent parfois à atteindre l'excellence (ex. : vision par ordinateur appliquée à l'imagerie médicale), cela n'est pas généralisable à l'ensemble des cas d'usage [955]. De plus, la dimension interrelationnelle du soin (empathie, négociation, etc.) subsiste comme étant une compétence propre à l'humain [709]. Actuellement, l'applicabilité des intelligences artificielles en contexte clinique se restreint donc à des tâches spécifiques et délimitées. Malgré que l'intégration de ces outils dans les flux opérationnels suggère une redéfinition des pratiques médicales, la perspective d'un remplacement des soignants n'est résolument pas à l'ordre du jour. Au contraire, le cadre d'intégration qui est communément revendiqué par les fabricants est celui d'une approche holistique des soins de santé [956]. De plus, afin de favoriser un usage éthique et responsable de ces technologies, il est crucial que les professionnels de santé soient acteurs de cette évolution en prenant part à la conception des modèles, à leur évaluation, à l'élaboration du cadre réglementaire et en assurant la collecte et la labellisation des données [957]. En mettant l'accent sur la synergie entre les compétences des professionnels de santé et les atouts propres à l'intelligence artificielle, l'intention générale est d'optimiser la précision et l'exactitude des diagnostics et des traitements ainsi que le temps consacré aux patients [713]. Dans un contexte démographique induisant des problématiques d'accès aux soins (médecine à flux tendu et déserts médicaux), l'intelligence artificielle est donc appelée à jouer un rôle majeur dans les pratiques diagnostiques et thérapeutiques de demain [783]. Les enjeux de la santé et de l'intelligence artificielle sont complexes et multiples, aussi nous espérons que l'interdisciplinarité de cette thèse préfigurera un phénomène plus large et salubre de convergence des domaines de la santé et de l'intelligence artificielle.

Bibliographie

Bibliographie

- [1] MEDLINE : Overview. *National Library of Medicine*. Consulté le 17 avril 2023. https://www.nlm.nih.gov/medline/medline_overview.html
- [2] Balkanyi L, Cornet R. The Interplay of Knowledge Representation with Various Fields of Artificial Intelligence in Medicine. *Yearb Med Inform*. 2019 Aug;28(1):27-34.
- [3] Samuel AL. Artificial Intelligence: A Frontier of Automation. *The ANNALS of the American Academy of Political and Social Science*. 1962 Mar;340(1):10-20.
- [4] Hof R. Interview: Inside Google Brain Founder Andrew Ng's Plans To Transform Baidu. *Forbes*. 28 août 2014.
- [5] Islam MR, Ahmed MU, Barua S, Begum S. A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks. *Applied Sciences*. 2022; 12(3):1353.
- [6] Lemke N, Trein P, Varone F. *Defining Artificial Intelligence as a Political Problem: A Discourse Network Analysis from Germany*. University of Lausanne ; 2022.
- [7] Philippe E. *Lettre de mission adressée à Cédric Villani*. 8 septembre 2017. <https://parlement-et-citoyens.fr/media/default/0001/01/81e6a65e224b3de490ecfbd00b6e40c79bcb3a43.pdf>
- [8] Villani C, Schoenauer M, Bonnet Y, et al. *Donner un sens à l'intelligence artificielle*. AI for HUMANITY ; 2018.
- [9] Nordlinger B, Villani C. *Santé et Intelligence artificielle*. CNRS Editions ; 2018.
- [10] Mou X. Artificial Intelligence: Investment Trends and Selected Industry Uses. *International Finance Corporation*. 2019.
- [11] Alami H, Rivard L, Lehoux P, et al. Artificial intelligence in health care: laying the Foundation for Responsible, sustainable, and inclusive innovation in low- and middle-income countries. *Global Health*. 2020 Jun 24;16(1):52.
- [12] Santoro E. Artificial intelligence in medicine: limits and obstacles. *Recenti Prog Med*. 2017 Dec;108(12):500-502.
- [13] Erdélyi OJ, Goldsmith J. Regulating Artificial Intelligence: Proposal for a Global Solution. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AI/ES '18)*. arXiv:2005.11072. 2018 Dec:95-101.
- [14] Borana J. Applications of Artificial Intelligence & Associated Technologies. *Proceeding of International Conference on Emerging Technologies in Engineering, Biomedical, Management and Science*. 5-6 mars 2016.
- [15] Howard J. Artificial intelligence: Implications for the future of work. *Am J Ind Med*. 2019; 62: 917– 926.
- [16] Zhang Y. A Historical Interaction between Artificial Intelligence and Philosophy. *ArXiv, abs/2208.04148*. 2022.
- [17] Thomason RH. *Philosophical Logic and Artificial Intelligence*. Springer Dordrecht ; 1989.
- [18] Mosk C. *Safety First in a Hostile Environment.: The Historical Origins of Artificial Intelligence*. 2021.
- [19] Macpherson T, Churchland A, Sejnowski T, et al. Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research. *Neural Netw*. 2021 Dec;144:603-613.
- [20] Demongeot J, Herve T, Rialle V, et al. *Artificial intelligence and cognitive sciences*. Manchester university press ; 1988.

- [21] Johri P, Khatri SK, Al-Taani AT, Sabharwal M, Suvanov S, Kumar A. Natural Language Processing: History, Evolution, Application, and Future Work. *Proceedings of 3rd International Conference on Computing Informatics and Networks*. Springer Singapore. 23 avril 2020.
- [22] Elkind E, Leyton-Brown K. Algorithmic Game Theory and Artificial Intelligence. *AI Magazine*, 2010 Sep;31(4):9-12.
- [23] Tricou F. Chapitre 1. Au cœur du débat : la théorie économique de la décision. Bénédicte Vidaillet éd., *La décision: Une approche pluridisciplinaire des processus de choix*. De Boeck Supérieur ; 2005:13-42.
- [24] Loève M. *Probability Theory I*. Springer New York ; 1977.
- [25] Owen G. *Game Theory*. Emerald Group Publishing Limited ; 1995.
- [26] Puterman ML. Chapter 8 Markov decision processes. *Handbooks in Operations Research and Management Science Volume 2*. 1990:331-434.
- [27] Zhu L, Zheng WJ. Informatics, Data Science, and Artificial Intelligence. *JAMA*. 2018 Sep 18;320(11):1103-1104.
- [28] Macki J, Strauss A. *Introduction to Optimal Control Theory*. Springer Science & Business Media ; 2012.
- [29] Waterson P. The Cybernetic Return in Human Factors/Ergonomics (HFE). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 20 novembre 2019:894-898.
- [30] Bierlaire M. *Introduction à l'optimisation différentiable*. PPUR presses polytechniques ; 2006.
- [31] Holland O, Husbands P. The origins of British cybernetics: the Ratio Club. *Kybernetes*. 2011;40(1/2):110-123.
- [32] CYBERNÉTIQUE. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 17 avril 2023. <https://www.dictionnaire-academie.fr/article/A9C5393>
- [33] ARTIFICIEL, ARTIFICIELLE. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 17 avril 2023. <https://www.dictionnaire-academie.fr/article/A9A2706>
- [34] intelligence artificielle. *Encyclopédie Larousse en ligne*. Consulté le 17 avril 2023. https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257
- [35] INTELLIGENCE. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 17 avril 2023. <https://www.dictionnaire-academie.fr/article/A9I1608>
- [36] Mays JD. *Physics: Modeling Nature*. Novare Science & Math; 2015.
- [37] INFORMATIQUE. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 17 avril 2023. <https://www.dictionnaire-academie.fr/article/A9I1219>
- [38] Zhao XG, Zhou K, Xing B, et al. JAMIP: an artificial-intelligence aided data-driven infrastructure for computational materials informatics. *Science Bulletin*. 2021 Oct;66(19):1973-1985.
- [39] ALGORITHME. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 17 avril 2023. <https://www.dictionnaire-academie.fr/article/A9A1163>
- [40] HEURISTIQUE, EURISTIQUE. *Centre National de Ressources Textuelles et Lexicales*. Consulté le 17 avril 2023. <https://www.cnrtl.fr/definition/heuristique>
- [41] Pastre D. *L'INTELLIGENCE ARTIFICIELLE DEFINITION - GENERALITES - HISTORIQUE - DOMAINES*. Université Paris 5 ; 2000.
- [42] Stankov L. Complexity, Metacognition, and Fluid Intelligence. *Intelligence*. 2000 May;28(2):121-143.

- [43] Jakhar D, Kaur I. Artificial intelligence, machine learning and deep learning: definitions and differences. *Clinical and Experimental Dermatology*. 2020 Jan;45(1):131-132.
- [44] extensibilité. *Journal officiel de la République française du 27/02/2003*. Consulté le 17 avril 2023. <https://www.culture.fr/franceterme/terme/INFO618>
- [45] Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Translational Vision Science & Technology*. 2020 Feb; 9(2):14.
- [46] Dorado-Díaz PI, Sampedro-Gómez J, Vicente-Palacios V, Sánchez PL. Applications of Artificial Intelligence in Cardiology. The Future is Already Here. *Revista Española de Cardiología (English Edition)*. 2019 Dec;72(12):1065-1075.
- [47] Cielen D, Meysman A. *Introducing Data Science: Big data, machine learning, and more, using Python tools*. Simon and Schuster ; 2016.
- [48] About us. *International Organization for Standardization*. Consulté le 17 avril 2023. <https://www.iso.org/about-us.html>
- [49] ISO/IEC 2382-1:1993, Technologies de l'information — Vocabulaire — Partie 1: Termes fondamentaux. *International Organization for Standardization*. Consulté le 17 avril 2023. <https://www.iso.org/fr/standard/7229.html>
- [50] ISO/IEC 2382-28:1995, Technologies de l'information — Vocabulaire — Partie 28: Intelligence artificielle — Notions fondamentales et systèmes experts. *International Organization for Standardization*. Consulté le 17 avril 2023. <https://www.iso.org/fr/standard/7272.html>
- [51] ISO/IEC 2382:2015, Technologies de l'information — Vocabulaire. *International Organization for Standardization*. Consulté le 17 avril 2023. <https://www.iso.org/fr/standard/63598.html>
- [52] À propos. *Comité économique et social européen*. Consulté le 17 avril 2023. <https://www.eesc.europa.eu/fr/propos>
- [53] Muller C. *L'intelligence artificielle - Les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société (avis d'initiative)*. Comité économique et social européen. 31 août 2017 ; OJ C 288, p. 1.
- [54] Arrêté du 27 juin 1989 relatif à l'enrichissement du vocabulaire de l'informatique, Annexes I à III. 16 septembre 1989. https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000006537939
- [55] Joshi A, Mishra G. Artificial intelligence. *ICWET '10: Proceedings of the International Conference and Workshop on Emerging Trends in Technology*. Février 2010:1023.
- [56] Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. Pearson ; 2016.
- [57] Georges B. Intelligence artificielle : de quoi parle-t-on ?. *Constructif*. 2019;54:5-10.
- [58] Bouret J, Hoarau J, Mauléon F. Chapitre 5. L'IA, nouvelle grille d'analyse des compétences. *Soft Skills: Développez vos compétences comportementales, un enjeu pour votre carrière*. Dunod. 2018:59-67.
- [59] Tack C. Artificial intelligence and machine learning | applications in musculoskeletal physiotherapy. *Musculoskeletal Science and Practice*. 2019 Feb;39:164-169.
- [60] Srivastava A, Saini S, Gupta D. Comparison of Various Machine Learning Techniques and Its Uses in Different Fields. *3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2019;81-86.

- [61] Ng GW, Leung WC. Strong Artificial Intelligence and Consciousness. *Journal of Artificial Intelligence and Consciousness*. 2020;7(1):63-72.
- [62] domotique. *Dictionnaire Larousse en ligne*. Consulté le 17 avril 2023. <https://www.larousse.fr/dictionnaires/francais-monolingue>
- [63] Liu K, Hu X, Wang P. Do Machines Need Education? A Review on the Education of General Artificial Intelligence and Educational Innovation. 2018.
- [64] SAPIENCE. *Centre National de Ressources Textuelles et Lexicales*. Consulté le 17 avril 2023. <https://www.cnrtl.fr/definition/sapience>
- [65] Andrés PG. The Future of Artificial Intelligence. *Fert Batxillerat*. 2019.
- [66] Pavel MA. Direction for Artificial Intelligence to Achieve Sapiency Inspired by Homo Sapiens. *International Journal of Advanced Research in Artificial Intelligence*. 2016;5(9):17-21.
- [67] Hudson AD, Finn E, Wylie R. What can science fiction tell us about the future of artificial intelligence policy?. *AI & SOCIETY*. 2021 Sep;38:197-211.
- [68] Miaillhe N, Hodes C. The third age of artificial intelligence. *Field Actions Science Reports*. 2017;17:6-11.
- [69] Fjelland R. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*. 2020 Jun;7(1):1-9.
- [70] Moteur de règles Drools. *IBM*. Consulté le 17 avril 2023. <https://www.ibm.com/docs/fr/sig-and-i/5.2.3?topic=engine-drools-rules>
- [71] Saibene A, Assale M, Giltri M. Expert systems: Definitions, advantages and issues in medical field applications. *Expert Systems with Applications*. 2021 Sep;177.
- [72] Havel IM. Artificial intelligence and connectionism: Some philosophical implications. *Advanced Topics in Artificial Intelligence: International Summer School Prague*. 1992 Jul:25-41.
- [73] Alekseeva E. The opposition of symbolism and connectionism in the history of artificial intelligence development. *ISTORIYA*. 2020 Nov;11(97).
- [74] Hitzler P. *Neuro-symbolic artificial intelligence: The state of the art*. IOS Press. 2022.
- [75] Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. *Journal of Nuclear Medicine*. 2019 Sep;60(2):29S-37S.
- [76] Eysenck H. *Intelligence: A New Look*. Routledge ; 2018.
- [77] Russell S, Norvig P. Chapitre 1, Introduction. *Intelligence artificielle* (p. 2). Pearson Education France ; 2010.
- [78] Sheikh H, Prins C, Schrijvers E. Artificial Intelligence: Definition and Background. *Mission AI: The New System Technology* (pp. 15-41). Springer International Publishing ; 2023.
- [79] Idel M. *Golem: Jewish magical and mystical traditions on the artificial anthropoid*. State University of New York Press ; 1990.
- [80] Sturluson S. *L'Edda. Récits de mythologie nordique*. Traduit par François-Xavier Dillmann. L'Aube des peuples, Gallimard; 1991.
- [81] Tseu L. *Le vrai classique du vide parfait*. Traduit par Benedyct Grynypas. Gallimard; 2011.
- [82] Bareau A, Eugène D. La Lokapannatti et les idées cosmologiques du bouddhisme ancien. *Bulletin de l'Ecole française d'Extrême-Orient, Tome 66*. 1979.

- [83] El-Zein A. *Islam, Arabs, and the Intelligent World of the Jinn (Contemporary Issues in the Middle East Series)*. Syracuse University Press ; 2009.
- [84] SYLLOGISME. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 18 avril 2023. <https://www.dictionnaire-academie.fr/article/A8S2078>
- [85] Belna JP. *Histoire de la logique*. Ellipses ; 2014.
- [86] Kelsen H. On the Practical Syllogism. *Essays in Legal and Moral Philosophy. Synthese Library, vol 57*. Springer ; 1973.
- [87] Toussaint G. *The Euclidean Algorithm Generates Traditional Musical Rhythms*. Renaissance Banff: Mathematics, Music, Art, Culture ; 2005.
- [88] Mehri B. From Al-Khwarizmi to Algorithm. *Olympiads in Informatics. Sharif University of Technology*. 2017;11:71-74.
- [89] Wear A, French RK, Lonie IM. *The medical Renaissance of the sixteenth century*. Cambridge University Press ; 1985.
- [90] Smith DE. *Medicine and Mathematics in the Sixteenth Century*. Annals of Medical History ; 1917.
- [91] Rommevaux-Tani S. L'édition électronique des manuscrits de Thomas Harriot (1560-1621). *Centre national de la recherche scientifique*. Consulté le 18 avril 2023. <https://images.math.cnrs.fr/L-edition-electronique-des-manuscrits-de-Thomas-Harriot-1560-1621.html>
- [92] Larrère C. "L'animal-machine. Élevage industriel et biotechnologies". *Cahiers d'histoire. Revue d'histoire critique*. 2022 Aug;153.
- [93] Tietz S. Gottfried Wilhelm Leibniz (1646-1716). *Wiley*. 2014 Mar.
- [94] Strickland L, Lewis HR. *Leibniz on Binary: The Invention of Computer Arithmetic*. MIT Press ; 2022.
- [95] automate. *Dictionnaire Larousse en ligne*. Consulté le 18 avril 2023. <https://www.larousse.fr/dictionnaires/francais/automate/6746>
- [96] de Vaucanson J. *Le mécanisme du flûteur automate*. Jacques Guerin ; 1738.
- [97] Leducfayette D. *La Mettrie et Descartes*. Europe ; 1978.
- [98] Alice R. WOLFGANG VON KEMPELEN AND THE ILLUSIONISTIC IMITATIONS OF THE BODY. 2020.
- [99] Barbaud P. Musique algorithmique. *Esprit*. 1960 Jan;280(1):92-97.
- [100] Réunion du 31 Janvier 1964 (Compte-rendu par M. LEPP). 11 Février 1964. http://www.lam.jussieu.fr/Membres/Castellengo/publications/1964a_GAM2-CompositionMachine.pdf
- [101] Van Heijenoort J. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press ; 2002.
- [102] Leclercq B. Rhétorique de l'idéographie. *Nouveaux Actes Sémiotiques*. CeReS, Université de Limoges ; 2011.
- [103] Lerma R. El autómatas ajedrecista de Leonardo Torres Quevedo, precursor de la inteligencia artificial. *Técnica industrial*. 2020;325:92-94.
- [104] Čejková J. Robots: The century past and the century ahead, an Introduction to the 2021 ALIFE conference. *Proceedings of the ALIFE 2022: The 2022 Conference on Artificial Life*. MIT Press. 2022 Jul:5.

- [105] Reilly K. From Automata to Automation: The Birth of the Robot in R.U.R. (Rossum's Universal Robots). *Automata and Mimesis on the Stage of Theatre History*. Palgrave Macmillan ; 2011.
- [106] Hoare G. A survey of mathematical logic, part II: Post-1931. *The Mathematical Gazette*. 1996 Jul;80(488):286-297.
- [107] Weiss E. Konrad zuse obituary. *IEEE Annals of the History of Computing*. 1996 Aug;18(2):3.
- [108] Kaminka GA, Spokoini-Stern R, Amir Y, Agmon N, Bachelet I. Molecular Robots Obeying Asimov's Three Laws of Robotics. *Artif Life*. 2017 Aug;23(3):343-350.
- [109] Heims SJ. *The Cybernetics Group*. MIT Press ; 1991.
- [110] Palm G. Warren McCulloch and Walter Pitts: A Logical. *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory*. Springer ; 1986.
- [111] Kline RR. *The cybernetics moment: Or why we call our age the information age*. JHU Press ; 2015.
- [112] Abraham TH. The Macy Conferences on Cybernetics: Reinstantiating the Mind. *Oxford Research Encyclopedia of Psychology*. 2020 Apr.
- [113] Dupuy JP. *Aux origines des sciences cognitives*. La Découverte ; 1994.
- [114] Dugas É, Collard L, Nakas R, Hébert T. Le coronavirus à l'épreuve des stratégies interactionnelles. Entre dilemmes et enjeux sociétaux. *Recherches & éducations*. 2020.
- [115] McCartney S. *ENIAC: The triumphs and tragedies of the world's first computer*. Walker & Company ; 1999.
- [116] Campbell-Kelly M. Origin of computing. *Scientific American*. 2009 Sep;301(3):62-69.
- [117] Shannon CE. A mathematical theory of communication. *The Bell system technical journal*. 1948 Jul;27(3):379-423.
- [118] Shannon CE. XXII. Programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1949 Nov;41(314):256-275.
- [119] Soni J, Goodman R. *A mind at play: how Claude Shannon invented the information age*. Simon and Schuster ; 2017.
- [120] De Carli A, Franchini F. Trends in Industrial Automation: An Approach to Training and Education. *IFAC Proceedings Volumes*. 1997 Jun;30(7):399-403.
- [121] Haggan GL. *A history of computer numerical control*. Education Ressources Information Center ; 1988.
- [122] Turing AM. Computing machinery and intelligence-AM Turing. *Mind*. 1950 Oct;59(236):433-460.
- [123] Shieber SM. *The Turing test: verbal behavior as the hallmark of intelligence*. MIT Press ; 2004.
- [124] Ashrafian H, Darzi A, Athanasiou T. A novel modification of the Turing test for artificial intelligence and robotics in healthcare. *The International Journal of Medical Robotics and Computer Assisted Surgery*. 2015 Mar;11(1):38-43.
- [125] Geman D, Geman S, Hallonquist N, Younes L. Visual turing test for computer vision systems. *Proceedings of the National Academy of Sciences*. 2015 Mar;112(12):3618-3623.
- [126] Marcus G. What Comes After the Turing Test?. *The New Yorker*. 9 juin 2014.
- [127] Crevier D. *AI: the tumultuous history of the search for artificial intelligence*. Basic Books ; 1993.
- [128] Schaeffer J. Didn't Samuel Solve That Game?. *One Jump Ahead: Computer Perfection at Checkers*. 2009:1-11.

- [129] The IBM 700 Series. *IBM*. Consulté le 18 avril 2023. <https://www.ibm.com/ibm/history/ibm100/us/en/icons/ibm700series/impacts/>
- [130] Schaeffer J, Burch N, Bjornsson Y, et al. Checkers is solved. *Science*. 2007 Sep;317(5844):1518-1522.
- [131] Schaeffer J, Lake R. Solving the game of checkers. *Games of no chance*. Cambridge University Press. 1998 Nov;29:119-133.
- [132] Newell A. The chess machine: an example of dealing with a complex task by adaptation. *Proceedings of the March 1-3, 1955, western joint computer conference*. 1955 Mar:101-108.
- [133] Data Structures. *Clojure*. Consulté le 18 avril 2023. https://clojure.org/reference/data_structures#Data%20Structures-Symbols
- [134] Newell A, Shaw JC. Programming the logic theory machine. *Papers presented at the February 26-28, 1957, western joint computer conference: Techniques for reliability*. 1957 Feb:230-240.
- [135] McCarthy J. History of LISP. In *History of programming languages*. 1978:173-185.
- [136] Newell A, Shaw JC, Simon HA. Empirical explorations of the logic theory machine: a case study in heuristic. *Papers presented at the February 26-28, 1957, western joint computer conference: Techniques for reliability*. 1957 Feb:218-230.
- [137] Alexis Nasr. Licence Info : Théorie des langages. *Université d'Aix-Marseille*. Consulté le 18 avril 2023. <http://www.pageperso.lif.univ-mrs.fr/~alexis.nasr/Ens/IntroLing/logique.pdf>
- [138] Whitehead AN, Russell B. *Principia Mathematica to *56*. Cambridge University Press ; 1997.
- [139] Gugerty L. Newell and Simon's logic theorist: Historical background and impact on cognitive modeling. *Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications. 2006 Oct;50(9):880-884.
- [140] Kline R. Cybernetics, automata studies, and the Dartmouth conference on artificial intelligence. *IEEE Annals of the History of Computing*. 2010 Jun;33(4):5-16.
- [141] McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the dartmouth summer research project on artificial intelligence. *AI magazine*. 1955 Aug;27(4):12.
- [142] Larson M. Optimizing chess: philology and algorithmic culture. *Diacritics*. 2018;46(1):30-53.
- [143] Hoekenga BC. *Mind over machine: what Deep Blue taught us about chess, artificial intelligence, and the human spirit*. Massachusetts Institute of Technology ; 2007.
- [144] Rosenblatt F. Perceptron simulation experiments. *Proceedings of the IRE*. 1960 Mar;48(3):301-309.
- [145] Luhn HP. A business intelligence system. *IBM Journal of research and development*. 1958 Oct;2(4):314-319.
- [146] Newell A, Shaw JC. A variety of intelligent learning in a general problem solver. *RAND Report*. 1959 Jul.
- [147] Gelernter HL. Realization of a geometry theorem proving machine. *IFIP congress*. 1959 Jun:273-281.
- [148] Green Jr BF, Wolf AK, Chomsky C, Laughery K. Baseball: an automatic question-answerer. *Papers presented at the May 9-11, 1961, western joint IRE-AIEE-ACM computer conference*. 1961 May:219-224.
- [149] Zadeh LA. Fuzzy sets. *Information and control*. 1965 Jun;8(3):338-353.
- [150] Lindsay RK, Buchanan BG, Feigenbaum EA, Lederberg J. DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence*. 1993 Jun;61(2):209-261.

- [151] Weizenbaum J. ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*. 1966 Jan;9(1):36-45.
- [152] Neff G. Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*. 2016;10:4915-4931.
- [153] Greenblatt RD, Eastlake DE, & Crocker SD. The Greenblatt chess program. *Proceedings of the November 14-16, 1967, fall joint computer conference*. 1967 Nov:801-810.
- [154] Proceedings of the First International Joint Conference on Artificial Intelligence, IJCAI-69 Contents. *International Joint Conferences on Artificial Intelligence Organization*. Consulté le 18 avril 2023. <https://www.ijcai.org/proceedings/1969>
- [155] Information about previous IJCAI Conferences and Proceedings. *International Joint Conferences on Artificial Intelligence Organization*. Consulté le 18 avril 2023. https://www.ijcai.org/past_conferences
- [156] Codd EF. A relational model of data for large shared data banks. *Communications of the ACM*. 1970 Jun;13(6):377-387.
- [157] Linnainmaa S. *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*. Univ. Helsinki ; 1970.
- [158] Werbos PJ. *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. Harvard University ; 1974.
- [159] Wythoff BJ. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*. 1993 Feb;18(2):115-155.
- [160] Durkin J. Expert systems: a view of the field. *IEEE Intelligent Systems*. 1996 Apr;11(02):56-63.
- [161] Sowa JF. Semantic networks. *Encyclopedia of artificial intelligence*. 1992;2:1493-1511.
- [162] Quillian MR. Semantic memory. *Semantic information processing*. 1968.
- [163] Minsky M. A framework for representing knowledge. *The Psychology of Computer Vision*. 1974.
- [164] Schank RC, Abelson RP. Scripts, plans, and knowledge. *IJCAI*. 1975 Sep;75:151-157.
- [165] Al-Fedaghi S. Three levels of modeling: static (structure/trajectories of flow), dynamic (events) and behavioral (chronology of events). *ArXiv, abs/2005.00149*. 2020.
- [166] SHRDLU. *Stanford HCI Group*. Consulté le 18 avril 2023. <http://hci.stanford.edu/~winograd/shrdlu/>
- [167] Winograd T. Procedures as a representation for data in a computer program for understanding natural language. *MASSACHUSETTS INST OF TECH CAMBRIDGE PROJECT MAC*. 1971.
- [168] Simon HA. Artificial Intelligence Systems That Understand. *IJCAI*. 1977 Aug:1059-1073.
- [169] Colmerauer A, Roussel P. The birth of Prolog. *History of programming languages---II*. 1996 Jan:331-367.
- [170] Goldberg A, Kay A. Teaching smalltalk. *Xerox Parc SSL*. 1977 Aug;77.
- [171] Klein M, Frana P. Encyclopedia of artificial intelligence: The past, present, and future of AI. *ABC-CLIO*. 2021.
- [172] Moravec H. *The Role of Raw Power in Intelligence*. Stanford University. 1976.
- [173] Dreyfus HL. Alchemy and artificial intelligence. *RAND CORP SANTA MONICA CA*. 1965.
- [174] Marvin M, Seymour AP. Perceptrons. *Cambridge, MA: MIT Press*. 1969. <http://134.208.26.59/math/AI/AI.pdf>

- [175] Lighthill J. Lighthill Report: Artificial Intelligence: a paper symposium. *Science Research Council*. 1973.
- [176] Weizenbaum J. Computer power and human reason: From judgment to calculation. *Plymouth, MI: WH Freeman and Company*. 1976.
- [177] Searle JR. Minds, brains, and programs. *Behavioral and brain sciences*. 1980 Sep;3(3):417-424.
- [178] Yao X, Zhou J, Zhang J, Boër CR. From intelligent manufacturing to smart manufacturing for industry 4.0 driven by next generation artificial intelligence and further on. *2017 5th international conference on enterprise systems. IEEE*. 2017 Sep:311-318.
- [179] National Research Council. *Funding a revolution: Government support for computing research*. National Academies Press ; 1999.
- [180] Garvey C. Artificial intelligence and Japan's fifth generation: the information society, neoliberalism, and alternative modernities. *Pacific Historical Review*. 2019 Nov;88(4):619-658.
- [181] Frontain M. Microelectronics and Computer Technology Corporation [MCC]. *Handbook of Texas Online. Texas State Historical Association*. 2020.
- [182] Roland A, Shiman P. *Strategic computing: DARPA and the quest for machine intelligence*. MIT Press ; 2002.
- [183] Appendix 4: Historic Government policy on artificial intelligence in the United Kingdom. *UK Parliament*. Consulté le 18 avril 2023. <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10018.htm>
- [184] Goldstein I, Papert S. Artificial intelligence, language, and the study of knowledge. *Cognitive science*. 1977 Jan;1(1):84-123.
- [185] Bonnist E. The Future of Artificial Intelligence in the Healthcare Industry. *Honors Theses*. 2021.
- [186] Sviokla JJ. An examination of the impact of expert systems on the firm: the case of XCON. *MIS Quarterly*. 1990 Jun;14(2):127-140.
- [187] Erman LD, Lesser VR. Hearsay-II. Tutorial Introduction and Retrospective View. *CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE*. 1978.
- [188] Hart PE, Duda RO, Einaudi MT. PROSPECTOR—a computer-based consultation system for mineral exploration. *Journal of the International Association for Mathematical Geology*. 1978 Oct;10(5):589-610.
- [189] Lenat DB. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*. 1995 Nov;38(11):33-38.
- [190] Mascardi V, Cordi V, Rosso P. A Comparison of Upper Ontologies. *Woa*. 2007 Sep;2007:55-64.
- [191] asanchez75/opencyc. *GitHub*. Consulté le 18 avril 2023. <https://github.com/asanchez75/opencyc>
- [192] Paschke A, Schroeder M. *Inductive logic programming for bioinformatics in Prova*. VLDB DBM ; 2007.
- [193] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*. 1982 Apr;79(8):2554-2558.
- [194] Parker DB. *Learning-logic: Casting the cortex of the human brain in silicon*. Center for Computational Research in Economics and Management Science ; 1985.
- [195] Le Cun Y. Learning process in an asymmetric threshold network. *Disordered systems and biological organization. Springer Berlin Heidelberg*. 1986:233-240.
- [196] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986 Oct;323(6088):533-536.

- [197] Pearl J. *Probabilistic Reasoning in Intelligent Systems*. Elsevier ; 1988.
- [198] Golberg DE. Genetic algorithms in search, optimization, and machine learning. *Addion wesley*. 1989 Aug;1989(102):36.
- [199] McCorduck P, Cfe C. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. CRC Press ; 2004.
- [200] Brooks RA. Intelligence without representation. *Artificial intelligence*. 1991 Jan;47(1-3),139-159.
- [201] Moravec H. *Mind children: The future of robot and human intelligence*. Harvard University Press ; 1988.
- [202] Allam Z. Big data, artificial intelligence and the rise of autonomous smart cities. *The Rise of Autonomous Smart Cities: Technology, Economic Performance and Climate Resilience*. Palgrave Macmillan ; 2020:7-30.
- [203] Gil Y, Deelman E, Blythe J, Kesselman C, Tangmunarunkit H. Artificial intelligence and grids: Workflow planning and beyond. *IEEE Intelligent Systems*. 2004 Jan-Feb;19(1):26-33.
- [204] Hassaballah M, Hosny KM. Recent advances in computer vision. *Springer Cham*. 2019 Jan;804:1-84.
- [205] Barceló P, Monet M, Pérez J, Subercaseaux B. Model interpretability through the lens of computational complexity. *Advances in neural information processing systems*. 2020;33:15487-15498.
- [206] Markoff J. Behind artificial intelligence, a squadron of bright real people. *The New York Times*. 14 octobre 2005.
- [207] Fourmentraux JP. A work in common: the dynamics of credits in coproduced works of digital art. *SOCIOLOGIE DU TRAVAIL*. 2007 Apr-Jun;49(2):162-179.
- [208] Pomerleau DA. Alvin: An autonomous land vehicle in a neural network. *Advances in neural information processing systems*. 1988.
- [209] Computer Beats Champ Again - This Time in Othello. *The New York Times*. 9 août 1997.
- [210] Kitano H. *RoboCup-97: robot soccer world cup I (Vol. 1395)*. Springer Science & Business Media ; 1998.
- [211] Past RoboCup events. *RoboCup*. Consulté le 18 avril 2023. https://www.robocup.org/past_robocup_events
- [212] Noda I, Suzuki SJ, Matsubara H, Asada M, Kitano H. RoboCup-97: The first robot world cup soccer games and conferences. *AI magazine*. 1998 Sep;19(3):49.
- [213] Miller GA. *WordNet: An electronic lexical database*. MIT press ; 1998.
- [214] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The unified medical language system: an informatics research collaboration. *Journal of the American Medical Informatics Association*. 1998 Jan;5(1):1-11.
- [215] Jacksi K, Abass SM. Development history of the world wide web. *Int. J. Sci. Technol. Res*. 2019 Sep;8(9):75-79.
- [216] Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*. 1998 Apr;30(1-7):107-117.
- [217] Nayak P, Kurien J, Dorais G, et al. Validating the ds-1 remote agent experiment. *Artificial intelligence, robotics and automation in space*. 1999 Aug;440:349.
- [218] Pransky J. AIBO—the No. 1 selling service robot. *Industrial robot: An international journal*. 2001 Feb;28(1):24-26.
- [219] Tang KY, Chang CY, Hwang GJ. Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998–2019). *Interactive Learning Environments*. 2021 Jan:1-19.

- [220] Verma P, Sharma S. Artificial Intelligence based Recommendation System. *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE. 2020 Dec:669-673.
- [221] Hou YT, Chang Y, Chen T, Laih CS, Chen CM. Malicious web content detection by machine learning. *expert systems with applications*. 2010 Jan;37(1):55-60.
- [222] Von Ahn L, Blum M, Langford, J. Telling humans and computers apart automatically. *Communications of the ACM*. 2004 Feb;47(2),56-60.
- [223] Srihari SN, Cha SH, Arora H, Lee S. Individuality of handwriting. *Journal of forensic sciences*. 2002 Jul;47(4),856-872.
- [224] Holden C. At the head of his class. *Science*. 2003 Feb;299(5611):1311.
- [225] David Hanson CV. *Hanson Robotics*. Consulté le 18 avril 2023. https://www.hansonrobotics.com/wp-content/uploads/2018/12/DavidHansonPhD-CV_2018-12-05.pdf
- [226] Shiotani S, Tomonaka T, Kemmotsu, K, Asano S, Oonishi K, Hiura R. World's first full-fledged communication robot" Wakamaru" capable of living with family and supporting persons. *Mitsubishi Juko Giho*. 2006 Jan;43(1):44-45.
- [227] Langen M, Heinrich S. Humanoid Robots: Use Cases as AI-Lab Companion: Can an empathic and collaborative digital companion motivate innovation?. *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*. 2019 Jun: 1-6.
- [228] Newborn M, Newborn M. 2003: Deep Junior Confounds Kasparov, Drawing 3–3 in New York. *Beyond Deep Blue: Chess in the Stratosphere*. 2011 Feb:71-81.
- [229] Markoff J. Computer wins on 'Jeopardy!': Trivial, it's not. *The New York Times*. 16 février 2011.
- [230] Chao X, Kou G, Li T, Peng Y. Jie Ke versus AlphaGo: A ranking approach using decision making method for large-scale data with incomplete information. *European Journal of Operational Research*. 2018 Feb;265(1),239-247.
- [231] VITAL POSITIONING SYSTEM. *Le site officiel du gouvernement du Canada*. Consulté le 18 avril 2023. <https://ised-isde.canada.ca/opic/recherche-marques/1196643?lang=fra>
- [232] Thrun S, Montemerlo M, Dahlkamp H, et al. Stanley: The robot that won the DARPA Grand Challenge. *Journal of field Robotics*. 2006 Sep;23(9),661-692.
- [233] Rouff C. *Experience from the DARPA urban challenge*. M. Hinchey (Ed.). Springer-Verlag London Limited ; 2012.
- [234] Che D, Safran M, Peng Z. From big data to big data mining: challenges, issues, and opportunities. *Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18*. Springer Berlin Heidelberg. 2013:1-15.
- [235] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. *Communications of the ACM*. 2008 Jan;51(1):107-113.
- [236] HDFS Architecture Guide. *The Apache Software Foundation*. Consulté le 18 avril 2023. https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html
- [237] O'malley O. Terabyte sort on apache hadoop. *Yahoo!*. 2008 May.

- [238] Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Stoica I. Spark: Cluster computing with working sets. *HotCloud*. 2010;10(10):95.
- [239] About us. *scikit-learn*. Consulté le 18 avril 2023. <https://scikit-learn.org/stable/about.html>
- [240] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;12:2825-2830.
- [241] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural computation*. 2006 Jul;18(7):1527-1554.
- [242] Russakovsky O, Deng J, Su H. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015 Apr;115:211-252.
- [243] Pinto N, Stone Z, Zickler T, Cox D. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. *CVPR 2011 WORKSHOPS. IEEE*. 2011 Jun: 35-42.
- [244] Hinds J, Williams EJ, Joinson AN. "It wouldn't happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal. *International Journal of Human-Computer Studies*. 2020.
- [245] Reis A, Paulino D, Paredes H, Barroso J. Using intelligent personal assistants to strengthen the elderlies' social bonds: A preliminary evaluation of amazon alexa, google assistant, microsoft cortana, and apple siri. *Universal Access in Human-Computer Interaction. Human and Technological Environments: 11th International Conference, UAHCI 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part III 11*. Springer International Publishing. 2017:593-602.
- [246] Warwick K, Shah H. Can machines think? A report on Turing test experiments at the Royal Society. *Journal of experimental & Theoretical artificial Intelligence*. 2016;28(6):989-1007.
- [247] No, A 'Supercomputer' Did NOT Pass The Turing Test For The First Time And Everyone Should Know Better. *techdirt*. Consulté le 18 avril 2023. <https://www.techdirt.com/2014/06/09/no-supercomputer-did-not-pass-turing-test-first-time-everyone-should-know-better/>
- [248] Zemčik T. Failure of chatbot Tay was evil, ugliness and uselessness in its nature or do we judge it through cognitive shortcuts and biases?. *AI & SOCIETY*. 2021 Mar;36:361-367.
- [249] Retto J. *Sophia, first citizen robot of the world*. Universidad Nacional Mayor de San Marcos ; 2017.
- [250] Borji A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. *ArXiv, abs/2210.00586*. 2022.
- [251] Liebrezn M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *The Lancet Digital Health*. 2023 Feb;5(3):105-106.
- [252] Seshia SA, Sadigh D, Sastry SS. Towards verified artificial intelligence. *ArXiv, abs/1606.08514*. 2016.
- [253] Maschler B, Weyrich M. Deep transfer learning for industrial automation: a review and discussion of new techniques for data-driven machine learning. *IEEE Industrial Electronics Magazine*. 2021 Jan;15(2):65-75.
- [254] Duarte JC, Cavalcanti MCR, de Souza Costa I, Esteves D. An interoperable service for the provenance of machine learning experiments. *Proceedings of the International Conference on Web Intelligence*. 2017 Aug;132-138.
- [255] Wang YE, Wei GY, Brooks D. Benchmarking TPU, GPU, and CPU platforms for deep learning. *ArXiv, abs/1907.10701*. 2019.

- [256] Baji T. Evolution of the GPU Device widely used in AI and Massive Parallel Processing. *2018 IEEE 2nd Electron Devices Technology and Manufacturing Conference (EDTM)*. 2018 Mar:7-9.
- [257] Li Y, Gu J, Wang L. Research on artificial intelligence ethics in the field of art design. *Journal of Physics: Conference Series*. 2020 Nov;1673(1):012052.
- [258] Mohsin K. Necessity of Artificial Intelligence Law. *SSRN*. 2019 Jan.
- [259] Buchanan BG, Smith RG. Fundamentals of expert systems. *Annual review of computer science*. 1988 Jun;3(1),23-58.
- [260] Nikolopoulos C. *Expert systems: introduction to first and second generation and hybrid knowledge based systems*. CRC Press ; 1997.
- [261] Singla J, Grover D, Bhandari A. Medical expert systems for diagnosis of various diseases. *International Journal of Computer Applications*. 2014 May;93(7):36-43.
- [262] Schreiber G. Knowledge engineering. *Foundations of Artificial Intelligence, Chapter 25*. 2008;3:929-946.
- [263] Abdullah MS, Benest I, Evans A, Kimble C. Knowledge modelling techniques for developing knowledge management systems. *Third European Conference on Knowledge Management: Trinity College Dublin, Ireland*. 2002 Sep.
- [264] Kendal SL, Creen M. *An introduction to knowledge engineering*. Springer London ; 2007:1-25.
- [265] Wilson M. Knowledge engineering. *Proceedings of SOFSEM*. 1993 Dec;93:357-373.
- [266] Chinniah P, Muttan DS. ICD 10 Based Medical Expert System Using Fuzzy Temporal Logic. *ArXiv, abs/1001.1979*. 2010.
- [267] Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). *Drug safety*. 1999 Feb;20(2):109-117.
- [268] McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry*. 2003 Apr;49(4):624-633.
- [269] ONTOLOGIE. *Centre National de Ressources Textuelles et Lexicales*. Consulté le 18 avril 2023. <https://www.cnrtl.fr/lexicographie/ontologie>
- [270] Breitman K, Casanova MA, Truszkowski W. Ontology in Computer Science. *Semantic Web: Concepts, Technologies and Applications*. Springer Science & Business Media. 2007:17-34.
- [271] Coron E, Vanbiervliet G. Intelligence artificielle et maladies digestives. *Le Grand Métier*. 2021.
- [272] Li J, Liu C, Liu B, et al. Diversity-aware retrieval of medical records. *Computers in Industry*. 2015 May;69:81-91.
- [273] Sindhu CS, Hegde NP. A framework to handle data heterogeneity contextual to medical big data. *2015 IEEE international conference on computational intelligence and computing research (ICCCIC)*. 2015 Dec:1-7.
- [274] Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. *BMJ Health & Care Informatics*. 2015 Apr; 22(2).
- [275] Presence Ontology. *BioPortal*. Consulté le 18 avril 2023. <https://bioportal.bioontology.org/ontologies/PREO>
- [276] Noy NF, Shah NH, Whetzel PL. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*. 2009 Jul;37(2):170-173.
- [277] Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. *Journal of the American Medical Informatics Association*. 2014 Feb;21(1):11-19.

- [278] Bertaud Gounot V, Donfack V, Lasbleiz J, Bourde A, Duvauferrier R. Creating an ontology driven rules base for an expert system for medical diagnosis. *User Centred Networked Health Care*. 2011;169:714-718.
- [279] Grieu J, Lecroq F, Galinho T, Boukachour H. Environnements Industriels Virtualisés et Processus d'Apprentissage. *Séminaire: Virtualia 2016, la réalité virtuelle au service de la recherche*. 2018.
- [280] Galopin A, Bouaud J, Pereira S, Seroussi B. An ontology-based clinical Decision support system for the management of patients with multiple chronic disorders. *MedInfo*. 2015 Jan;275-279.
- [281] Nonaka I, Toyama R. The knowledge-creating theory revisited: knowledge creation as a synthesizing process. *Knowledge management research & practice*. 2003 Jan;1(1),2-10.
- [282] Yen CC, Tang HL. Inside an expert system: strengths, weaknesses, and trends. *Journal of Computer Information Systems*. 1989;30(1),34-39.
- [283] Ivanović M, Budimac Z. An overview of ontologies and data resources in medical domains. *Expert Systems with Applications*. 2014 Sep;41(11),5158-5166.
- [284] Menzies T. Cost benefits of ontologies. *intelligence*. 1999;10(3):26-32.
- [285] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electronic Markets*. 2021 Apr;31(3):685-695.
- [286] Bosshard C. Les données sont l'or noir du 21e siècle. *Bulletin des médecins suisses*. 2016 Mar;97(1213):447.
- [287] Taylor P. Amount of data created, consumed, and stored 2010-2020, with forecasts to 2025. *Statista*. 8 septembre 2022.
- [288] Baviskar MR, Nagargoje PN, Deshmukh PA, Baviskar RR. A survey of data science techniques and available tools. *International Research Journal of Engineering and Technology (IRJET)*. 2021 Apr; 8(04):4258-4263.
- [289] Mitchell TM. *Machine learning (Chapitre 1 - Introduction, 1.1 WELL-POSED LEARNING PROBLEMS)*. McGraw-Hill Science/Engineering/Math ; 1997 Mar.
- [290] Hady MFA, Schwenker F. Semi-supervised learning. *Handbook on Neural Information Processing* (pp. 215-239). Springer ; 2013.
- [291] Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, Tang J. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*. 2021 Jun;35(1),857-876.
- [292] Mahesh B. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. 2020 Jan;9(1):381-386.
- [293] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *Journal of Big data*. 2016 May;3(1),1-40.
- [294] Stulp F, Sigaud O. Many regression algorithms, one unified model: A review. *Neural Networks*. 2015 Sep;69:60-79.
- [295] Harchaoui Z, Bach F. Image classification with segmentation graph kernels. *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007 Jun:1-8.
- [296] Harper PR. A review and comparison of classification algorithms for medical decision making. *Health policy*. 2005 Mar;71(3),315-331.
- [297] Rodriguez MZ, Comin CH, Casanova D, et al. Clustering algorithms: A comparative approach. *PloS one*. 2019 Jan;14(1):0210236.

- [298] Karamizadeh S, Abdullah SM, Manaf AA, Zamani M, Hooman A. An overview of principal component analysis. *Journal of Signal and Information Processing*. 2013 May;4(3B):173-175.
- [299] LinearRegression. *scikit-learn*. Consulté le 18 avril 2023. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- [301] Maulud D, Abdulazeez AM. A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*. 2020 Dec;1(4),140-147.
- [302] Muller AC, Guido S. *Introduction to Machine Learning with Python*. Oreilly ; 2022.
- [303] LogisticRegression. *scikit-learn*. Consulté le 19 avril 2023. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- [304] Schober P, Vetter TR. Logistic regression in medical research. *Anesthesia and analgesia*. 2021 Jan;132(2):365.
- [305] Bisong E. *Building machine learning and deep learning models on Google cloud platform*. Berkeley ; 2019.
- [306] Naive Bayes. *scikit-learn*. Consulté le 19 avril 2023. https://scikit-learn.org/stable/modules/naive_bayes.html
- [307] Berrar D. Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (pp. 403-412). 2018.
- [308] Zhang H. *The Optimality of Naive Bayes*. Faculty of Computer Science, University of New Brunswick ; 2004.
- [309] Kurniawan YI, Razi F, Nofiyati N, Wijayanto B, Hidayat ML. Naive Bayes modification for intrusion detection system classification with zero probability. *Bulletin of Electrical Engineering and Informatics*. 2021 Oct;10(5):2751-2758.
- [310] SVR. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>
- [311] SVC. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
- [312] Amer M, Goldstein M, Abdennadher S. Enhancing one-class support vector machines for unsupervised anomaly detection. *Proceedings of the ACM SIGKDD workshop on outlier detection and description*. 2013 Aug:8-15.
- [313] Noble WS. What is a support vector machine?. *Nature biotechnology*. 2006 Dec;24(12),1565-1567.
- [314] Tharwat A, Hassanien AE, Elnaghi BE. A BA-based algorithm for parameter optimization of support vector machine. *Pattern recognition letters*. 2017 Jul;93:13-22.
- [315] Tharwat A. Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems*. 2019 Feb;61:1269-1302.
- [316] Chen PH, Lin CJ, Schölkopf B. A tutorial on v-support vector machines. *Applied Stochastic Models in Business and Industry*. 2005 Mar;21(2):111-136.
- [317] Schölkopf B, Smola AJ, Bach F. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press ; 2002.
- [318] KNeighborsClassifier. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

- [319] KNeighborsRegressor. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html>
- [320] Kramer O. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors* (pp. 13-23). Springer ; 2013.
- [321] Kotsiantis SB. Decision trees: a recent overview. *Artificial Intelligence Review*. 2013 Apr;39:261-283.
- [322] DecisionTreeClassifier. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>
- [323] DecisionTreeRegressor. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html>
- [324] RandomForestClassifier. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- [325] RandomForestRegressor. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>
- [326] Biau G, Scornet E. A random forest guided tour. *Test*. 2016 Apr;25:197-227.
- [327] Chuprunov A, Fazekas I. Strong laws of large numbers for random forests. *Acta Math Hung*. 2009 Jul;124:59–71.
- [328] Oshiro TM, Perez PS, Baranauskas JA. How many trees in a random forest?. *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8* (pp. 154-168). Springer Berlin Heidelberg; 2012.
- [329] Blakely L, Reno MJ, Broderick RJ. *Evaluation and Comparison of Machine Learning Techniques for Rapid QSTS Simulations*. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States) ; 2018.
- [330] KMeans. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- [331] Yoder J, Priebe CE. Semi-supervised k-means++. *Journal of Statistical Computation and Simulation*. 2017;87(13),2597-2608.
- [332] Ahmed M, Seraj R, Islam SMS. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*. 2019 Aug;9(8):1295.
- [333] Finley T, Joachims T. *Supervised k-means clustering*. 2008.
- [334] PCA. *scikit-learn*. Consulté le 19 avril 2023. <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
- [335] Ghojogh B, Crowley M. Unsupervised and supervised principal component analysis: Tutorial. *ArXiv, abs/1906.03148*. 2019.
- [336] Kim M, Yun J, Cho Y, et al. Deep learning in medical imaging. *Neurospine*. 2019 Dec;16(4):657.
- [337] Hasiewicz Z. Modular neural networks for non-linearity recovering by the Haar approximation. *Neural Networks*. 2000 Dec;13(10):1107-1133.
- [338] Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press ; 2016.
- [339] Lau MM, Lim KH. Investigation of activation functions in deep belief network. *2017 2nd international conference on control and robotics engineering (ICCRE)* (pp. 201-206). IEEE. 2017 Apr.

- [340] Singh J, Banerjee R. A study on single and multi-layer perceptron neural network. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 35-40). IEEE. 2019 Mar.
- [341] Van Veen F, Leijnen S. The Neural Network Zoo. *The Asimov Institute*. 22 avril 2019.
- [342] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*. 2022 Dec;32(12):6999-7019.
- [343] Niel O, Bastard P. Artificial intelligence in nephrology: core concepts, clinical applications, and perspectives. *American Journal of Kidney Diseases*. 2019 Dec;74(6):803-810.
- [344] Medsker L, Jain LC. *Recurrent neural networks: design and applications*. CRC press ; 1999.
- [345] Chauhan N., Singh K. A review on conventional machine learning vs deep learning. *2018 International conference on computing, power and communication technologies (GUCON)* (pp. 347-352). IEEE. 2018 Sep.
- [346] Such FP, Madhavan V, Conti E, Lehman J, Stanley KO, Clune J. Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. *ArXiv, abs/1712.06567*. 2017.
- [347] Lee DH, Zhang S, Fischer A, Bengio Y. Difference target propagation. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I 15* (pp. 498-515). Springer International Publishing ; 2015.
- [348] Ma WDK, Lewis JP, Kleijn WB. The HSIC bottleneck: Deep learning without back-propagation. *Proceedings of the AAAI conference on artificial intelligence*. 2020 Apr;34(4):5085-5092.
- [349] Choromanska A, Cowen B, Kumaravel S, et al. Beyond backprop: Online alternating minimization with auxiliary variables. *International Conference on Machine Learning* (pp. 1193-1202). PMLR ; 2019 May.
- [350] Jaderberg M, Czarnecki, WM, Osindero S, et al. Decoupled neural interfaces using synthetic gradients. *International conference on machine learning* (pp. 1627-1635). PMLR ; 2017 Jul.
- [351] Nøkland A, Eidnes LH. Training neural networks with local error signals. *International conference on machine learning* (pp. 4839-4850). PMLR ; 2019 May.
- [352] Rojas R. The backpropagation algorithm. *Neural networks: a systematic introduction* (pp. 149-182). Springer ; 1996.
- [353] Sekeroglu B, Dimililer K. Review and analysis of hidden neuron number effect of shallow backpropagation neural networks. *Neural Network World*. 2020;30(2):97-112.
- [354] Clair DCS, Peterson GE, Aylward S, Bond WE. Design techniques for the control of errors in backpropagation neural networks. *Science of Artificial Neural Networks II* (Vol. 1966, pp. 372-383). SPIE ; 1993 Aug.
- [355] Liu Z, Xu Z, Jin J, Shen Z, Darrell T. Dropout Reduces Underfitting. *ArXiv, abs/2303.01500*. 2023.
- [356] Qiao L. Teaching design of online ideological and political course based on deep learning model evaluation. *Scientific Programming*. 2022 Jul;2022:1-8.
- [357] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*. 2018.
- [358] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb;542(7639):115-118.

- [359] Wolterink JM, Leiner T, de Vos BD, van Hamersvelt RW, Viergever MA, Išgum I. Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks. *Medical image analysis*. 2016 Dec; 34:123-136.
- [360] Xu J, Gong E, Pauly J, Zaharchuk G. 200x low-dose PET reconstruction using deep learning. *ArXiv, abs/1712.04119*. 2017.
- [361] Mu W, Jiang L, Zhang J, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nature communications*. 2020 Oct;11(1):5228.
- [362] Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology*. 2018 Aug;125(8):1199-1206.
- [363] Purwins H, Li B, Virtanen T, Schlüter J, Chang SY, Sainath T. Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*. 2019 May;13(2):206-219.
- [364] Jeancolas L, Petrovska-Delacrétaz D, Mangone G, et al. X-vectors: New quantitative biomarkers for early Parkinson's disease detection from speech. *Frontiers in Neuroinformatics*. 2021 Feb;15:578369.
- [365] Wu C, Li X, Guo Y, Wang J, Ren Z, Wang M, Yang Z. Natural language processing for smart construction: Current status and future directions. *Automation in Construction*. 2022 Feb;134:104059.
- [366] Houssein EH, Mohamed RE, Ali AA. Machine learning techniques for biomedical natural language processing: a comprehensive review. *IEEE Access*. 2021;9:140628-140653.
- [367] Khanzode KCA, Sarode RD. Advantages and disadvantages of artificial intelligence and machine learning: A literature review. *International Journal of Library & Information Science (IJLIS)*. 2020 Jan-Apr;9(1):3.
- [368] Joshi G, Walambe R, Kotecha K. A review on explainability in multimodal deep neural nets. *IEEE Access*. 2021;9:59800-59821.
- [369] Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*. 2022 Jul;54(6):1-35.
- [370] Taleb NN. *The Black Swan: Why Don't We Learn that We Don't Learn?*. NY: Random House. 2005:1145.
- [371] Sarker MK, Zhou L, Eberhart A, Hitzler P. Neuro-symbolic artificial intelligence. *AI Communications*. 2021;34(3):197-209.
- [372] Feinman R, Lake BM. Learning task-general representations with generative neuro-symbolic modeling. *ArXiv, abs/2006.14448*. 2020.
- [373] Hitzler P, Eberhart A, Ebrahimi M, Sarker MK, Zhou L. Neuro-symbolic approaches in artificial intelligence. *National Science Review*. 2022 Jun;9(6):nwac035.
- [374] Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics*. 2020;8(6):890.
- [375] Bhagat PM. Artificial Intelligence in Healthcare. *International Journal of Scientific Research & Engineering Trends*. 2021 Mar-Apr;7(2):796-800.
- [376] DIAGNOSTIC. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 19 avril 2023. <https://www.dictionnaire-academie.fr/article/A9D2326>
- [377] THÉRAPEUTIQUE. *Dictionnaire de l'Académie française, 9e édition*. Consulté le 19 avril 2023. <https://www.dictionnaire-academie.fr/article/A8T0546>

- [378] Définition des cas d'utilisation. *IBM*. Consulté le 19 avril 2023. <https://www.ibm.com/docs/fr/elms/elm/6.0.5?topic=requirements-defining-use-cases>
- [379] cas d'usage. *Office québécois de la langue française*. Consulté le 19 avril 2023. https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8364726/cas-dusage?utm_campaign=Redirection%20des%20anciens%20outils&utm_content=id_fiche%3D8364726&utm_source=GDT
- [380] Mumtaz W, Vuong PL, Xia L, Malik AS, Rashid RBA. An EEG-based machine learning method to screen alcohol use disorder. *Cognitive neurodynamics*. 2017 Apr;11:161-171.
- [381] Watson CG, Tilleskior C, Hoodecheck-Schow EA, Pucel J, Jacobs L. Do alcoholics give valid self-reports?. *Journal of Studies on Alcohol*. 1984 Jul;45(4):344-348.
- [382] Acion L, Kelmansky D, van der Laan M, Sahker E, Jones D, Arndt S. Use of a machine learning framework to predict substance use disorder treatment success. *PloS one*. 2017;12(4):e0175383.
- [383] Abijo T, Blum K, Gondré-Lewis MC. Neuropharmacological and neurogenetic correlates of opioid use disorder (OUD) as a function of ethnicity: relevance to precision addiction medicine. *Current Neuropharmacology*. 2020;18(7):578-595.
- [384] Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, et al. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Modern Pathology*. 2018 Oct; 31(10):1502-1512.
- [385] Soysal SD, Tzankov A, Muenst SE. Role of the tumor microenvironment in breast cancer. *Pathobiology*. 2015 Sep;82(3-4):142-152.
- [386] Wang G, Teoh JYC, Choi KS. Diagnosis of prostate cancer in a Chinese population by using machine learning methods. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-4). IEEE ; 2018 Jul.
- [387] Barry MJ, Simmons LH. Prevention of prostate cancer morbidity and mortality: primary prevention and early detection. *Medical Clinics*. 2017 Jul;101(4):787-806.
- [388] Bhanji Y, Allaway MJ, Gorin MA. Recent advances and current role of transperineal prostate biopsy. *Urologic Clinics*. 2021 Feb;48(1):25-33.
- [389] Nicolae A, Morton G, Chung H, et al. Evaluation of a machine-learning algorithm for treatment planning in prostate low-dose-rate brachytherapy. *International Journal of Radiation Oncology* Biology* Physics*. 2017 Mar; 97(4):822-829.
- [390] Gandaglia G, Leni R, Bray F. Epidemiology and prevention of prostate cancer. *European urology oncology*. 2021 Dec;4(6):877-892.
- [391] Cuendet GL, Schoettker P, Yüce A, et al. Facial image analysis for fully automatic prediction of difficult endotracheal intubation. *IEEE Transactions on Biomedical Engineering*. 2016 Feb;63(2):328-339.
- [392] Andrade RGADC, Lima BLS, Lopes DKDO, Couceiro Filho RO, Lima LC, Couceiro TCDM. Difficult laryngoscopy and tracheal intubation: observational study. *Revista Brasileira de Anestesiologia*. 2018 Mar-Apr;68:168-173.
- [393] Pesteie M, Abolmaesumi P, Ashab HAD, et al. Real-time ultrasound image classification for spine anesthesia using local directional Hadamard features. *International journal of computer assisted radiology and surgery*. 2015 Jun;10:901-912.

- [394] Rivera CE. Lumbar epidural steroid injections. *Physical Medicine and Rehabilitation Clinics*. 2018 Feb;29(1):73-92.
- [395] Helm S, Harmon PC, Noe C, et al. Transforaminal epidural steroid injections: a systematic review and meta-analysis of efficacy and safety. *Pain Physician*. 2021;24(S1):S209-S232.
- [396] Stahl CM, Meisinger QC, Andre MP, Kinney TB, Newton IG. Radiation risk to the fluoroscopy operator and staff. *American Journal of Roentgenology*. 2016 Oct;207(4):737-744.
- [397] Buchberger B, Scholl K, Krabbe L, Spiller L, Lux B. Radiation exposure by medical X-ray applications. *GMS German Medical Science*. 2022 Mar;20:Doc06.
- [398] Chi M, Chen AS. Ultrasound for lumbar spinal procedures. *Physical Medicine and Rehabilitation Clinics*. 2018 Feb;29(1):49-60.
- [399] Hurdle MFB. Ultrasound-guided spinal procedures for pain: a review. *Physical Medicine and Rehabilitation Clinics*. 2016 Aug;27(3):673-686.
- [400] Dey D, Slomka PJ, Leeson P. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *Journal of the American College of Cardiology*. 2019 Mar;73(11):1317-1335.
- [401] Azevedo CF, Rochitte CE, Lima JA. Coronary artery calcium score and coronary computed tomographic angiography for cardiovascular risk stratification. *Arq Bras Cardiol*. 2012 Jun;98(6):559-68.
- [402] Feng Y, Guo Z, Dong Z, et al. An efficient cardiac mapping strategy for radiofrequency catheter ablation with active learning. *International journal of computer assisted radiology and surgery*. 2017 Jul;12:1199-1207.
- [403] Kumagai K, Gondo N, Matsumoto N, et al. New technique for simultaneous catheter mapping of pulmonary veins for catheter ablation in focal atrial fibrillation. *Cardiology*. 2001 Apr;94(4):233-238.
- [404] Skadsberg ND, He B, Laske TG, Ramanathan C, Iaizzo PA. Cardiac mapping technology. *Handbook of Cardiac Anatomy, Physiology, and Devices* (pp. 599-614). Springer ; 2015.
- [405] Sim I, Bishop M, O'Neill M, Williams SE. Left atrial voltage mapping: defining and targeting the atrial fibrillation substrate. *Journal of Interventional Cardiac Electrophysiology*. 2019 Dec;56:213-227.
- [406] Chang EK, Yu CY, Clarke R, et al. Defining a Patient Population With Cirrhosis: An Automated Algorithm With Natural Language Processing. *J Clin Gastroenterol*. 2016 Nov/Dec;50(10):889-894.
- [407] Rubbo B, Fitzpatrick NK, Denaxas S, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations. *Int J Cardiol*. 2015 May;187:705-711.
- [408] Mustafić LD, Gurbeta L, Badnjevic-Cengic A, et al. Diagnosis of severe aortic stenosis using implemented expert system. *CMBEBIH 2019: Proceedings of the International Conference on Medical and Biological Engineering, 16–18 May 2019, Banja Luka, Bosnia and Herzegovina* (pp. 149-153). Springer International Publishing ; 2020.
- [409] Šećkanović A, Šehovac M, Spahić L, et al. Review of artificial intelligence application in cardiology. *2020 9th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1-5). IEEE ; 2020 Jun.
- [410] Marquis-Gravel G, Redfors B, Leon MB, Généreux P. Medical Treatment of Aortic Stenosis. *Circulation*. 2016 Nov;134(22):1766-1784.
- [411] Durko AP, Osnabrugge RL, Van Mieghem NM, et al. Annual number of candidates for transcatheter aortic valve implantation per country: current estimates and future projections. *Eur Heart J*. 2018 Jul;39(28):2635-2642.

- [412] Thoenes M, Bramlage P, Zamorano P, et al. Patient screening for early detection of aortic stenosis (AS)-review of current practice and future perspectives. *J Thorac Dis*. 2018 Sep;10(9):5584-5594.
- [413] Brennan JM. Under-treatment of Aortic Stenosis in the United States. *TVT 2019, Chicago*. 2019 Jun.
- [414] Medved D, Ohlsson M, Höglund P, Andersson B, Nugues P, Nilsson J. Improving prediction of heart transplantation outcome using deep learning techniques. *Sci Rep*. 2018 Feb;8(1):3613.
- [415] Awad MA, Shah A, Griffith BP. Current status and outcomes in heart transplantation: a narrative review. *Rev Cardiovasc Med*. 2022 Jan;23(1):11.
- [416] Quader M, Toldo S, Chen Q, Hundley G, Kasirajan V. Heart transplantation from donation after circulatory death donors: Present and future. *J Card Surg*. 2020 Apr;35(4):875-885.
- [417] Schulze PC, Jiang J, Yang J, et al. Preoperative assessment of high-risk candidates to predict survival after heart transplantation. *Circ Heart Fail*. 2013 May;6(3):527-34.
- [418] Hashimoto DA, Rosman G, Witkowski ER, et al. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann Surg*. 2019 Sep;270(3):414-421.
- [419] Rayonnement ultraviolet (UV) et cancer de la peau. *Organisation mondiale de la Santé*. Consulté le 19 avril 2023. [https://www.who.int/fr/news-room/questions-and-answers/item/ultraviolet-\(uv\)-radiation-and-skin-cancer](https://www.who.int/fr/news-room/questions-and-answers/item/ultraviolet-(uv)-radiation-and-skin-cancer)
- [420] Saginala K, Barsouk A, Aluru JS, Rawla P, Barsouk A. Epidemiology of Melanoma. *Med Sci (Basel)*. 2021 Oct;9(4):63.
- [421] Tripp MK, Watson M, Balk SJ, Swetter SM, Gershenwald JE. State of the science on prevention and screening to reduce melanoma incidence and mortality: The time is now. *CA Cancer J Clin*. 2016 Nov;66(6):460-480.
- [422] Yasui Y, Kato H, Oda T, Nakamura M, Morita A. Complications and risk factors of punch biopsy: A retrospective large-scale study. *J Dermatol*. 2023 Jan;50(1):98-101.
- [423] Cazzaniga S, Sassi F, Mercuri SR, Naldi L. Prediction of clinical response to excimer laser treatment in vitiligo by using neural network models. *Dermatology*. 2009;219(2):133-7.
- [424] Kussainova A, Kassym L, Akhmetova A, et al. Vitiligo and anxiety: A systematic review and meta-analysis. *PLoS One*. 2020 Nov;15(11):e0241445.
- [425] Speeckaert R, van Geel N. Vitiligo: An Update on Pathophysiology and Treatment Options. *Am J Clin Dermatol*. 2017;18(6):733-744.
- [426] Sassi F, Cazzaniga S, Tessari G, et al. Randomized controlled trial comparing the effectiveness of 308-nm excimer laser alone or in combination with topical hydrocortisone 17-butyrate cream in the treatment of vitiligo of the face and neck. *Br J Dermatol*. 2008;159(5):1186-1191.
- [427] Post NF, Ezekwe N, Narayan VS, et al. The use of lasers in vitiligo, an overview. *J Eur Acad Dermatol Venereol*. 2022;36(6):779-789.
- [428] Lin C, Lin CS, Lee DJ, et al. Artificial Intelligence-Assisted Electrocardiography for Early Diagnosis of Thyrotoxic Periodic Paralysis. *J Endocr Soc*. 2021;5(9):bvab120.
- [429] Pinzon RT, Wijaya BM, Japanto ET. Sudden weakness in a 44 year old male: Thyrotoxic Periodic Paralysis (Case Report). *Journal of the Medical Sciences (Berkala Ilmu Kedokteran)*. 2022;54(1).

- [430] Chakraborty J, Chakraborty S, Moitra R. Thyrotoxic periodic paralysis: an update. *Journal of Endocrinology and Metabolism*. 2020;10(3-4),60-62.
- [431] Cappon G, Vettoretti M, Marturano F, Facchinetti A, Sparacino G. A Neural-Network-Based Approach to Personalize Insulin Bolus Calculation Using Continuous Glucose Monitoring. *J Diabetes Sci Technol*. 2018;12(2):265-272.
- [432] Slattery D, Amiel SA, Choudhary P. Optimal prandial timing of bolus insulin in diabetes management: a review. *Diabet Med*. 2018;35(3):306-316.
- [433] Wolpert HA, Atakov-Castillo A, Smith SA, Steil GM. Dietary fat acutely increases glucose concentrations and insulin requirements in patients with type 1 diabetes: implications for carbohydrate-based bolus dose calculation and intensive diabetes management. *Diabetes Care*. 2013;36(4):810-816.
- [434] Noaro G, Cappon G, Vettoretti M, Sparacino G, Favero SD, Facchinetti A. Machine-Learning Based Model to Improve Insulin Bolus Calculation in Type 1 Diabetes Therapy. *IEEE Trans Biomed Eng*. 2021;68(1):247-255.
- [435] Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest Endosc*. 2020;91(6):1264-1271.e1.
- [436] Uhlenhopp DJ, Then EO, Sunkara T, Gaduputi V. Epidemiology of esophageal cancer: update in global trends, etiology and risk factors. *Clin J Gastroenterol*. 2020;13(6):1010-1021.
- [437] Popa IV, Burlacu A, Mihai C, Prelipcean CC. A Machine Learning Model Accurately Predicts Ulcerative Colitis Activity at One Year in Patients Treated with Anti-Tumour Necrosis Factor α Agents. *Medicina (Kaunas)*. 2020;56(11):628.
- [438] Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: Emerging technologies and future directions. *World J Gastroenterol*. 2021;27(17):1920-1935.
- [439] Yamamoto T, Shimoyama T, Umegae S, Matsumoto K. Tacrolimus vs. anti-tumour necrosis factor agents for moderately to severely active ulcerative colitis: a retrospective observational study. *Aliment Pharmacol Ther*. 2016;43(6):705-716.
- [440] Adalimumab. VIDAL. Consulté le 19 avril 2023. https://www.vidal.fr/medicaments/substances/adalimumab-22520.html#substance_mecanisme
- [441] Infliximab. VIDAL. Consulté le 19 avril 2023. https://www.vidal.fr/medicaments/substances/infliximab-18833.html#substance_mecanisme
- [442] Êtanercept. VIDAL. Consulté le 19 avril 2023. https://www.vidal.fr/medicaments/substances/etanercept-20006.html#substance_mecanisme
- [443] Ford E, Sheppard J, Oliver S, Rooney P, Banerjee S, Cassell JA. Automated detection of patients with dementia whose symptoms have been identified in primary care but have no formal diagnosis: a retrospective case-control study using electronic primary care records. *BMJ Open*. 2021;11(1):e039248.
- [444] De La Vega FM, Chowdhury S, Moore B, et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Med*. 2021;13(1):153.
- [445] Pijuan J, Rodríguez-Sanz M, Natera-de Benito D, et al. Translational Diagnostics: An In-House Pipeline to Validate Genetic Variants in Children with Undiagnosed and Rare Diseases. *J Mol Diagn*. 2021;23(1):71-90.

- [446] Nordgren A, Lindstrand A, Wu HY, Fossum M. Precision medicine and rare diseases in pediatric urology [published online ahead of print, 2023 Mar 11]. *J Pediatr Urol.* 2023;S1477-5131(23)00093-1.
- [447] Dong D, Chung RY, Chan RHW, Gong S, Xu RH. Why is misdiagnosis more likely among some people with rare diseases than others? Insights from a population-based cross-sectional study in China. *Orphanet J Rare Dis.* 2020;15(1):307.
- [448] Salamalekis E, Thomopoulos P, Giannaris D, et al. Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks. *BJOG.* 2002;109(10):1137-1142.
- [449] Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorgiou AT, Jouannic JM. Contributions of Artificial Intelligence Reported in Obstetrics and Gynecology Journals: Systematic Review. *J Med Internet Res.* 2022;24(4):e35465.
- [450] Martis R, Emilia O, Nurdianti DS, Brown J. Intermittent auscultation (IA) of fetal heart rate in labour for fetal well-being. *Cochrane Database Syst Rev.* 2017;2(2):CD008680.
- [451] Wretler S, Holzmann M, Graner S, Lindqvist P, Falck S, Nordström L. Fetal heart rate monitoring of short term variation (STV): a methodological observational study. *BMC Pregnancy Childbirth.* 2016;16:55.
- [452] Fong DD, Knoesen A, Motamedi M, O'Neill T, Ghiasi S. Recovering the fetal signal in transabdominal fetal pulse oximetry. *Smart Health.* 2018;9:23-36.
- [453] Wald M, Sparks A, Sandlow J, Van-Voorhis B, Syrop CH, Niederberger CS. Computational models for prediction of IVF/ICSI outcomes with surgically retrieved spermatozoa. *Reprod Biomed Online.* 2005;11(3):325-331
- [454] Esteves SC, Coimbra I, Hallak J. Surgically retrieved spermatozoa for ICSI cycles in non-azoospermic males with high sperm DNA fragmentation in semen. *Andrology.* 2023;10.1111/andr.13405.
- [455] Malina A, Pooley JA. Psychological consequences of IVF fertilization - Review of research. *Ann Agric Environ Med.* 2017;24(4):554-558.
- [456] Cai S, Parker F, Urias MG, Goldberg MF, Hager GD, Scott AW. Deep Learning Detection of Sea Fan Neovascularization From Ultra-Widefield Color Fundus Photographs of Patients With Sickle Cell Hemoglobinopathy. *JAMA Ophthalmol.* 2021;139(2):206-213.
- [457] Cai S, Han IC, Scott AW. Artificial intelligence for improving sickle cell retinopathy diagnosis and management. *Eye (Lond).* 2021;35(10):2675-2684.
- [458] Oluleye TS, Babalola YO, Majekodunmi OI, Ijaluola MA. Sickle cell retinopathy: Patient awareness, mode of presentation, and treatment modalities in Ibadan, South-West Nigeria. *Nigerian Journal of Medicine.* 2021;30(5):481-486.
- [459] Pahl DA, Green NS, Bhatia M, Chen RWS. New Ways to Detect Pediatric Sickle Cell Retinopathy: A Comprehensive Review. *J Pediatr Hematol Oncol.* 2017;39(8):618-625.
- [460] Scott AW. Ophthalmic Manifestations of Sickle Cell Disease. *South Med J.* 2016;109(9):542-548.
- [461] Ohara T, Ikeda H, Sugitani Y, et al. Artificial intelligence supported anemia control system (AISACS) to prevent anemia in maintenance hemodialysis patients. *Int J Med Sci.* 2021;18(8):1831-1839.
- [462] Cazzola M. How and when to use erythropoietin. *Current Opinion in Hematology.* 1998;5(2):103-108.
- [463] Goodkin DA, Zhao J, Cases A, Nangaku M, Karaboyas A. Resistance to Erythropoiesis-Stimulating Agents among Patients on Hemodialysis Is Typically Transient. *Am J Nephrol.* 2022;53(5):333-342.

- [464] Marozas M, Zykus R, Sakalauskas A, Kupčinskas L, Lukoševičius A. Noninvasive Evaluation of Portal Hypertension Using a Supervised Learning Technique. *J Healthc Eng.* 2017;2017:6183714.
- [465] Ahn JC, Connell A, Simonetto DA, Hughes C, Shah VH. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology.* 2021;73(6):2546-2563.
- [466] Turco L, Garcia-Tsao G. Portal Hypertension: Pathogenesis and Diagnosis. *Clin Liver Dis.* 2019;23(4):573-587.
- [467] La Mura V, Nicolini A, Tosetti G, Primignani M. Cirrhosis and portal hypertension: The importance of risk stratification, the role of hepatic venous pressure gradient measurement. *World J Hepatol.* 2015;7(4):688-695.
- [468] Castera L, Pinzani M, Bosch J. Non invasive evaluation of portal hypertension using transient elastography. *J Hepatol.* 2012;56(3):696-703.
- [469] Ai H, Chen W, Zhang L, et al. Predicting Drug-Induced Liver Injury Using Ensemble Learning Methods and Molecular Fingerprints. *Toxicol Sci.* 2018;165(1):100-107.
- [470] Lin H, Ewing LE, Koturbash I, Gurley BJ, Miousse IR. MicroRNAs as biomarkers for liver injury: Current knowledge, challenges and future prospects. *Food Chem Toxicol.* 2017;110:229-239.
- [471] Lagarde N. *Méthodes de criblage virtuel in silico: importance de l'évaluation et application à la recherche de nouveaux inhibiteurs de l'interleukine 6.* Paris, CNAM ; 2014.
- [472] Melioli G, Spenser C, Reggiardo G, et al. Allergenius, an expert system for the interpretation of allergen microarray results. *World Allergy Organ J.* 2014;7(1):15.
- [473] Siegel PD, Law BF, Warshaw EM. Chemical Identification and Confirmation of Contact Allergens. *Dermatitis.* 2020;31(2):99-105.
- [474] Kalli M, Blok A, Jiang L, Starr N, Alcocer MJC, Falcone FH. Development of a protein microarray-based diagnostic chip mimicking the skin prick test for allergy diagnosis. *Sci Rep.* 2020;10(1):18208.
- [475] Moghram BA, Nabil E, Badr A. Ab-initio conformational epitope structure prediction using genetic algorithm and SVM for vaccine design. *Comput Methods Programs Biomed.* 2018;153:161-170.
- [476] Xu Z, Wang X, Zeng S, Ren X, Yan Y, Gong Z. Applying artificial intelligence for cancer immunotherapy. *Acta Pharm Sin B.* 2021;11(11):3393-3405.
- [477] Alexander V, Annamalai P. An elitist genetic algorithm based extreme learning machine. *Computational Intelligence, Cyber Security and Computational Models: Proceedings of ICC3 2015* (pp. 301-309). Springer Singapore ; 2016.
- [478] Floudas CA. Computational methods in protein structure prediction. *Biotechnol Bioeng.* 2007;97(2):207-213.
- [479] Hecker M, Wagner AH. Role of protein carbonylation in diabetes. *J Inherit Metab Dis.* 2018;41(1):29-38.
- [480] Singh SP, Mishra BN. Major histocompatibility complex linked databases and prediction tools for designing vaccines. *Hum Immunol.* 2016;77(3):295-306.
- [481] Hu R, Ruan G, Xiang S, Huang M, Liang Q, Li J. Automated diagnosis of covid-19 using deep learning and data augmentation on chest ct. *Medrxiv.* 2020.
- [482] Alsaïdi I, De Sousa Santos F, Plard B, et al. Factors associated with SARS-CoV2 infection and care pathways among the most vulnerable populations living in Marseille: a case control study. *BMC Public Health.* 2021;21(1):1704.

- [483] Larker M, Martin SS. COVID-19 Epidemiology and Differences in Incidence and Mortality Between Countries. *Cardiovascular Complications of COVID-19: Acute and Long-Term Impacts* (pp. 19-29). Cham: Springer International Publishing ; 2023.
- [484] Barakat C, Aach M, Schuppert A, Brynjólfsson S, Fritsch S, Riedel M. Analysis of Chest X-ray for COVID-19 Diagnosis as a Use Case for an HPC-Enabled Data Analysis and Machine Learning Platform for Medical Diagnosis Support. *Diagnostics (Basel)*. 2023;13(3):391.
- [485] Shen Y, Yuan K, Chen D, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif Intell Med*. 2018;86:20-32.
- [486] Rehman M, Ahmed S, Ahmed U, Tamanna K, Sabir MS, Niaz Z. An overview of self-medication: A major cause of antibiotic resistance and a threat to global public health. *J Pak Med Assoc*. 2021;71(3):943-949.
- [487] Noor AUZ, Kabir H, Chowdhury MAA, Ather MF, Kamrul Hasan M. An emerging route to antibiotic resistance in South Asia: a correspondence. *Ann Med Surg (Lond)*. 2023;85(2):335-336.
- [488] Abedi V, Goyal N, Tsivgoulis G, et al. Novel Screening Tool for Stroke Using Artificial Neural Network. *Stroke*. 2017;48(6):1678-1681.
- [489] Smit S, Hagemester DT, Van Rooyen C. Clinical review of stroke care at National District Hospital, Bloemfontein. *S Afr Fam Pract (2004)*. 2023;65(1):e1-e7.
- [490] Yu Y, Wen X, Lin JG, et al. Identification of three potential novel biomarkers for early diagnosis of acute ischemic stroke via plasma lipidomics. *Metabolomics*. 2023;19(4):32.
- [491] Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. *Diagnosis (Berl)*. 2014;1(2):155-166.
- [492] Goyal N, Male S, Al Wafai A, Bellamkonda S, Zand R. Cost burden of stroke mimics and transient ischemic attack after intravenous tissue plasminogen activator treatment. *J Stroke Cerebrovasc Dis*. 2015;24(4):828-833.
- [493] Ganzert S, Kramer S, Guttman J. Predicting the lung compliance of mechanically ventilated patients via statistical modeling. *Physiol Meas*. 2012;33(3):345-359.
- [494] Zhang Z, Navarese EP, Zheng B, et al. Analytics with artificial intelligence to advance the treatment of acute respiratory distress syndrome. *J Evid Based Med*. 2020;13(4):301-312.
- [495] Lionetti V, Recchia FA, Ranieri VM. Overview of ventilator-induced lung injury mechanisms. *Curr Opin Crit Care*. 2005;11(1):82-86.
- [496] Cooper AS. Positive End-Expiratory Pressure Levels in Adult Patients With Acute Lung Injury and Acute Respiratory Distress Syndrome. *Crit Care Nurse*. 2023;43(2):77-79.
- [497] Omer H, Salah H, Tamam N, et al. Assessment of occupational exposure from PET and PET/CT scanning in Saudi Arabia. *Radiation Physics and Chemistry*. 2023;204:110642.
- [498] Chhouri H, Alexandre D, Grumolato L. Mechanisms of Acquired Resistance and Tolerance to EGFR Targeted Therapy in Non-Small Cell Lung Cancer. *Cancers (Basel)*. 2023;15(2):504.
- [499] Li YZ, Kong SN, Liu YP, Yang Y, Zhang HM. Can Liquid Biopsy Based on ctDNA/cfDNA Replace Tissue Biopsy for the Precision Treatment of EGFR-Mutated NSCLC?. *J Clin Med*. 2023;12(4):1438.
- [500] Diep R, MacDonald M, Cooper R, et al. Biopsy Method and Needle Size on Success of Next-Generation Sequencing in NSCLC: A Brief Report. *JTO Clin Res Rep*. 2023;4(4):100497.

- [501] Bai H, Wang Z, Chen K, et al. Influence of chemotherapy on EGFR mutation status among patients with non-small-cell lung cancer. *J Clin Oncol*. 2012;30(25):3077-3083.
- [502] Yin L, Song C, Cui J, et al. A fusion decision system to identify and grade malnutrition in cancer patients: Machine learning reveals feasible workflow from representative real-world data. *Clin Nutr*. 2021;40(8):4958-4970.
- [503] Casirati A, Da Prat V, Cereda E, et al. The Key Role of Patient Empowerment in the Future Management of Cancer-Related Malnutrition. *Nutrients*. 2023;15(1):235.
- [504] Watanabe H, Oshima T. The Latest Treatments for Cancer Cachexia: An Overview. *Anticancer Res*. 2023;43(2):511-521.
- [505] Lu Y, Stathopoulou T, Vasiloglou MF, et al. An Artificial Intelligence-Based System for Nutrient Intake Assessment of Hospitalised Patients. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:5696-5699.
- [506] Peng H, Li B, Xiong W, Hu W, Ji R. RGBD salient object detection: A benchmark and algorithms. *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13* (pp. 92-109). Springer International Publishing ; 2014.
- [507] Katona P, Katona-Apte J. The interaction between nutrition and infection. *Clin Infect Dis*. 2008;46(10):1582-1588.
- [508] Monacelli F, Sartini M, Bassoli V, et al. Validation of the Photography Method for Nutritional Intake Assessment in Hospitalized Elderly Subjects. *J Nutr Health Aging*. 2017;21(6):614-621.
- [509] Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving palliative care with deep learning. *BMC Med Inform Decis Mak*. 2018;18(Suppl 4):122.
- [510] Tietbohl CK, Dafoe A, Jordan SR, et al. Palliative Care across Settings: Perspectives from Inpatient, Primary Care, and Home Health Care Providers and Staff. *Am J Hosp Palliat Care*. 2023;10499091231163156.
- [511] Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ*. 2000;320(7233):469-472.
- [512] Seeland A, Tabie M, Kim SK, Kirchner F, Kirchner EA. Adaptive multimodal biosignal control for exoskeleton supported stroke rehabilitation. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2431-2436). IEEE ; 2017 Oct.
- [513] Vélez-Guerrero MA, Callejas-Cuervo M, Mazzoleni S. Artificial Intelligence-Based Wearable Robotic Exoskeletons for Upper Limb Rehabilitation: A Review. *Sensors (Basel)*. 2021;21(6):2146.
- [514] Leow XRG, Ng SLA, Lau Y. Overground robotic exoskeleton training for patients with stroke on walking-related outcomes: A systematic review and meta-analysis of randomised controlled trials. *Arch Phys Med Rehabil*. 2023;S0003-9993(23)00165-X.
- [515] Verder H, Heiring C, Ramanathan R, et al. Bronchopulmonary dysplasia predicted at birth by artificial intelligence. *Acta Paediatr*. 2021;110(2):503-509.
- [516] Yu Z, Wang L, Wang Y, Zhang M, Xu Y, Liu A. Development and Validation of a Risk Scoring Tool for Bronchopulmonary Dysplasia in Preterm Infants Based on a Systematic Review and Meta-Analysis. *Healthcare (Basel)*. 2023;11(5):778.
- [517] Cui X, Fu J. Early prediction of bronchopulmonary dysplasia: can noninvasive monitoring methods be essential?. *ERJ Open Res*. 2023;9(2):00621-2022.

- [518] Rivera L, Siddaiah R, Oji-Mmuo C, Silveyra GR, Silveyra P. Biomarkers for Bronchopulmonary Dysplasia in the Preterm Infant. *Front Pediatr*. 2016;4:33.
- [519] Hu Y, Lee VC, Tan K. Prediction of clinicians' treatment in preterm infants with suspected late-onset sepsis—An ML approach. *2018 13th IEEE conference on industrial electronics and applications (ICIEA)* (pp. 1177-1182). IEEE ; 2018 May.
- [520] Parra-Llorca A, Pinilla-Gonzalez A, Torrejón-Rodríguez L, et al. Effects of Sepsis on Immune Response, Microbiome and Oxidative Metabolism in Preterm Infants. *Children (Basel)*. 2023;10(3):602.
- [521] Hornik CP, Fort P, Clark RH, et al. Early and late onset sepsis in very-low-birth-weight infants from a large group of neonatal intensive care units. *Early Hum Dev*. 2012;88 Suppl 2(Suppl 2):S69-S74.
- [522] Seven EK, Aydemir C, Tekin IO. Evaluation of Changes in Leukocyte Surface Markers in the Early Diagnosis of Late-Onset Neonatal Sepsis. *Journal of Pediatric Infectious Diseases*. 2023.
- [523] Liu H, Ren L, Fan B, Wang W, Hu X, Zhang X. Artificial Intelligence Algorithm-Based MRI in the Diagnosis of Complications after Renal Transplantation. *Contrast Media Mol Imaging*. 2022;2022:8930584.
- [524] Sigera LSM, Denning DW. Invasive Aspergillosis after Renal Transplantation. *J Fungi (Basel)*. 2023;9(2):255.
- [525] Reyna-Sepúlveda F, Ponce-Escobedo A, Guevara-Charles A, et al. Outcomes and Surgical Complications in Kidney Transplantation. *Int J Organ Transplant Med*. 2017;8(2):78-84.
- [525] Akl AI, Sobh MA, Enab YM, Tattersall J. Artificial intelligence: a new approach for prescription and monitoring of hemodialysis therapy. *Am J Kidney Dis*. 2001;38(6):1277-1283.
- [526] Kovacic V, Roguljic L, Jukic I, Kovacic V. Comparison of methods for hemodialysis dose calculation. *Dialysis & transplantation*. 2003;32(4):170-178.
- [527] Khan T, Nyholm D, Westin J, Dougherty M. A computer vision framework for finger-tapping evaluation in Parkinson's disease. *Artif Intell Med*. 2014;60(1):27-40.
- [528] Belić M, Bobić V, Badža M, Šolaja N, Đurić-Jovičić M, Kostić VS. Artificial intelligence for assisting diagnostics and assessment of Parkinson's disease-A review. *Clin Neurol Neurosurg*. 2019;184:105442.
- [529] Criswell S, Sterling C, Swisher L, Evanoff B, Racette BA. Sensitivity and specificity of the finger tapping task for the detection of psychogenic movement disorders. *Parkinsonism Relat Disord*. 2010;16(3):197-201.
- [530] Jakubowski J, Potulska-Chromik A, Chmielińska J, Nojszewska M, Kostera-Pruszczyk A. Application of imaging techniques to objectify the Finger Tapping test used in the diagnosis of Parkinson's disease. *Bulletin of the Polish Academy of Sciences: Technical Sciences*. 2023;71(2):e144886.
- [531] Fang J, Zhang P, Wang Q, et al. Artificial intelligence framework identifies candidate targets for drug repurposing in Alzheimer's disease. *Alzheimers Res Ther*. 2022;14(1):7.
- [532] Kelle N, Ehrlich U. *Situation of people providing support and care to someone with dementia*. 2023.
- [533] Démence. *Organisation mondiale de la Santé*. Consulté le 19 avril 2023. <https://www.who.int/fr/news-room/fact-sheets/detail/dementia>
- [534] Jung YJ, Kim YH, Bhalla M, Lee SB, Seo J. Genomics: New Light on Alzheimer's Disease Research. *Int J Mol Sci*. 2018;19(12):3771.
- [535] Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J Cancer Res Clin Oncol*. 2019;145(4):829-837.

- [536] García-Pola M, Pons-Fuster E, Suárez-Fernández C, Seoane-Romero J, Romero-Méndez A, López-Jornet P. Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review. *Cancers (Basel)*. 2021;13(18):4600.
- [537] Global Cancer Observatory (GLOBOCAN) 2020. Lip, oral cavity. *Organisation mondiale de la Santé*. Consulté le 19 avril 2023. <https://gco.iarc.fr/today/data/factsheets/cancers/1-Lip-oral-cavity-fact-sheet.pdf>
- [538] Abati S, Bramati C, Bondi S, Lissoni A, Trimarchi M. Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis. *Int J Environ Res Public Health*. 2020;17(24):9160.
- [539] Guide pour le diagnostic clinique différentiel des lésions de la muqueuse buccale. *dentalcare*. Consulté le 19 avril 2023. <https://docplayer.fr/105877461-Guide-pour-le-diagnostic-clinique-differentiel-des-lesions-de-la-muqueuse-buccale.html>
- [540] Li P, Kong D, Tang T, et al. Orthodontic Treatment Planning based on Artificial Neural Networks. *Sci Rep*. 2019;9(1):2037.
- [541] Proffit WR, Fields HW, Larson B, Sarver DM. *Contemporary orthodontics-e-book*. Elsevier Health Sciences ; 2018.
- [542] Proffit WR. The timing of orthodontic treatment: Effectiveness and Efficiency. *Rev Odont Stomat*. 2003;32:171-189.
- [543] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J Natl Cancer Inst*. 2019;111(9):916-922.
- [544] Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci*. 2020;111(5):1452-1460.
- [545] Broeders M, Moss S, Nyström L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen*. 2012;19 Suppl 1:14-25.
- [546] Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*. 2017;359:j4683.
- [547] Chang Y, Park H, Yang HJ, et al. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. *Sci Rep*. 2018;8(1):8857.
- [548] Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13(1):152.
- [549] Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-724.
- [550] Alvarellos M, Sheppard HE, Knarston I, et al. Democratizing clinical-genomic data: How federated platforms can promote benefits sharing in genomics. *Front Genet*. 2023;13:1045450.
- [551] Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103(2):167-175.
- [552] Glaucome. *Inserm*. Consulté le 20 avril 2023. <https://www.inserm.fr/dossier/glaucome/>
- [553] Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090.
- [554] Stevens GA, White RA, Flaxman SR, et al. Global prevalence of vision impairment and blindness: magnitude and temporal trends, 1990-2010. *Ophthalmology*. 2013;120(12):2377-2384.
- [555] Tatham AJ, Medeiros FA, Zangwill LM, Weinreb RN. Strategies to improve early diagnosis in glaucoma. *Prog Brain Res*. 2015;221:103-133.

- [556] Shaikh Y, Yu F, Coleman AL. Burden of undetected and untreated glaucoma in the United States. *Am J Ophthalmol*. 2014;158(6):1121-1129.e1.
- [557] Wagner IV, Stewart MW, Dorairaj SK. Updates on the Diagnosis and Management of Glaucoma. *Mayo Clin Proc Innov Qual Outcomes*. 2022;6(6):618-635.
- [558] Nutt RJ, Dowlut MS, McLoone SF, McLoone E. Epidemiology and long-term outcomes of primary congenital glaucoma: a population-based study. *Eye (Lond)*. 2023;10.1038/s41433-023-02382-6.
- [558] Rasti R, Allingham MJ, Mettu PS, et al. Deep learning-based single-shot prediction of differential effects of anti-VEGF treatment in patients with diabetic macular edema. *Biomed Opt Express*. 2020;11(2):1139-1152.
- [559] Gunasekaran DV, Ting DSW, Tan GSW, Wong TY. Artificial intelligence for diabetic retinopathy screening, prediction and management. *Curr Opin Ophthalmol*. 2020;31(5):357-365.
- [560] Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. 2012;35(3):556-564.
- [561] Chen J, Wang H, Qiu W. Intravitreal anti-vascular endothelial growth factor, laser photocoagulation, or combined therapy for diabetic macular edema: A systematic review and network meta-analysis. *Frontiers in Endocrinology*. 2023;14.
- [562] Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop*. 2017;88(6):581-586.
- [563] Cabitza F, Locoro A, Banfi G. Machine Learning in Orthopedics: A Literature Review. *Front Bioeng Biotechnol*. 2018;6:75.
- [564] Andersen DJ, Blair WF, Steyers CM Jr, Adams BD, el-Khoury GY, Brandser EA. Classification of distal radius fractures: an analysis of interobserver reliability and intraobserver reproducibility. *J Hand Surg Am*. 1996;21(4):574-582.
- [565] Kim JS, Merrill RK, Arvind V, et al. Examining the Ability of Artificial Neural Networks Machine Learning Models to Accurately Predict Complications Following Posterior Lumbar Spine Fusion. *Spine (Phila Pa 1976)*. 2018;43(12):853-860.
- [566] Mobbs RJ, Phan K, Malham G, Seex K, Rao PJ. Lumbar interbody fusion: techniques, indications and comparison of interbody fusion options including PLIF, TLIF, MI-TLIF, OLIF/ATP, LLIF and ALIF. *J Spine Surg*. 2015;1(1):2-18.
- [567] Rompe JD, Eysel P, Hopf C. Clinical efficacy of pedicle instrumentation and posterolateral fusion in the symptomatic degenerative lumbar spine. *Eur Spine J*. 1995;4(4):231-237.
- [568] Lowe TG, Tahernia AD, O'Brien MF, Smith DA. Unilateral transforaminal posterior lumbar interbody fusion (TLIF): indications, technique, and 2-year results. *J Spinal Disord Tech*. 2002;15(1):31-38.
- [569] Fang SH, Tsao Y, Hsiao MJ, et al. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. *J Voice*. 2019;33(5):634-641.
- [570] Tama BA, Kim DH, Kim G, Kim SW, Lee S. Recent Advances in the Application of Artificial Intelligence in Otorhinolaryngology-Head and Neck Surgery. *Clin Exp Otorhinolaryngol*. 2020;13(4):326-339.
- [571] Stemple JC, Roy N, Klaben BK. *Clinical voice pathology: Theory and management*. Plural Publishing ; 2018.
- [572] Cohen SM, Kim J, Roy N, Asche C, Courey M. Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope*. 2012;122(2):343-348.

- [573] Kim H, Kang WS, Park HJ, et al. Cochlear Implantation in Postlingually Deaf Adults is Time-sensitive Towards Positive Outcome: Prediction using Advanced Machine Learning Techniques. *Sci Rep*. 2018;8(1):18004.
- [574] Cunningham LL, Tucci DL. Hearing Loss in Adults. *N Engl J Med*. 2017;377(25):2465-2473.
- [575] Wittorff MG, Lewin G, Burton E. Acquired Combined Vision and Hearing Loss: Awareness and Perceptions of Australian Aged Care Workers. *Journal of Visual Impairment & Blindness*. 2023;117(1):74-86.
- [576] Pacala JT, Yueh B. Hearing deficits in the older patient: "I didn't notice anything". *JAMA*. 2012;307(11):1185-1194.
- [577] Lazard DS, Vincent C, Venail F, et al. Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: a new conceptual model over time. *PLoS One*. 2012;7(11):e48739.
- [578] Vijendren A, Ajith A, Borsetto D, et al. Cochlear Implant Infections and Outcomes: Experience From a Single Large Center. *Otol Neurotol*. 2020;41(9):e1105-e1110.
- [579] Liu X, Jiang J, Zhang K, et al. Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network. *PLoS One*. 2017;12(3):e0168606.
- [580] Reid JE, Eaton E. Artificial intelligence for pediatric ophthalmology. *Curr Opin Ophthalmol*. 2019;30(5):337-346.
- [581] Resnikoff S, Keys TU. Future trends in global blindness. *Indian J Ophthalmol*. 2012;60(5):387-395.
- [582] Lin D, Chen J, Lin Z, et al. 10-Year Overview of the Hospital-Based Prevalence and Treatment of Congenital Cataracts: The CCPMOH Experience. *PLoS One*. 2015;10(11):e0142298.
- [583] Medsinge A, Nischal KK. Pediatric cataract: challenges and future directions. *Clin Ophthalmol*. 2015;9:77-90.
- [584] Lenhart PD, Courtright P, Wilson ME, et al. Global challenges in the management of congenital cataract: proceedings of the 4th International Congenital Cataract Symposium held on March 7, 2014, New York, New York. *J AAPOS*. 2015;19(2):e1-e8.
- [585] Martonyi CL, Bahn CF, Meyer RF. Slit lamp: examination and photography. *Cornea E-Book*. Elsevier Health Sciences ; 2010.
- [586] West SK, Rosenthal F, Newland HS, Taylor HR. Use of photographic techniques to grade nuclear cataracts. *Invest Ophthalmol Vis Sci*. 1988;29(1):73-77.
- [587] Niel O, Bastard P, Boussard C, Hogan J, Kwon T, Deschênes G. Artificial intelligence outperforms experienced nephrologists to assess dry weight in pediatric patients on chronic hemodialysis. *Pediatr Nephrol*. 2018;33(10):1799-1803.
- [588] Thurlow JS, Joshi M, Yan G, et al. Global Epidemiology of End-Stage Kidney Disease and Disparities in Kidney Replacement Therapy. *Am J Nephrol*. 2021;52(2):98-107.
- [589] Pantelias K, Grapsa E. Vascular access today. *World J Nephrol*. 2012;1(3):69-78.
- [590] Neu AM, Frankenfield DL. Clinical outcomes in pediatric hemodialysis patients in the USA: lessons from CMS' ESRD CPM Project. *Pediatr Nephrol*. 2009;24(7):1287-1295.
- [591] Fischbach M, Zaloszc A, Shroff R. The interdialytic weight gain: a simple marker of left ventricular hypertrophy in children on chronic haemodialysis. *Pediatr Nephrol*. 2015;30(6):859-863.

- [592] Gotta V, Marsenic O, Pfister M. Age- and weight-based differences in haemodialysis prescription and delivery in children, adolescents and young adults. *Nephrol Dial Transplant*. 2018;33(9):1649-1660.
- [593] Raimann J, Liu L, Tyagi S, Levin NW, Kotanko P. A fresh look at dry weight. *Hemodial Int*. 2008;12(4):395-405.
- [594] Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *ArXiv, abs/1510.02855*. 2015.
- [595] Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol Sci*. 2019;40(8):592-604.
- [596] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004;3(11):935-949.
- [597] Wang L, Wu Y, Deng Y, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc*. 2015;137(7):2695-2703.
- [598] Gale W, Oakden-Rayner L, Carneiro G, Bradley AP, Palmer LJ. Detecting hip fractures with radiologist-level performance using deep neural networks. *ArXiv, abs/1711.06504*. 2017.
- [599] Syed AB, Zoga AC. Artificial Intelligence in Radiology: Current Technology and Future Directions. *Semin Musculoskelet Radiol*. 2018;22(5):540-545.
- [600] Brauer CA, Coca-Perrillon M, Cutler DM, Rosen AB. Incidence and mortality of hip fractures in the United States. *JAMA*. 2009;302(14):1573-1579.
- [601] Musculoskeletal conditions affect millions. *Organisation mondiale de la santé*. Consulté le 20 avril 2023. <https://www.who.int/news/item/27-10-2003-musculoskeletal-conditions-affect-millions>
- [602] Cannon J, Silvestri S, Munro M. Imaging choices in occult hip fracture. *J Emerg Med*. 2009;37(2):144-152.
- [603] Shiga T, Wajima Z, Ohe Y. Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression. *Can J Anaesth*. 2008;55(3):146-154.
- [604] Collin D, Dunker D, Göthlin JH, Geijer M. Observer variation for radiography, computed tomography, and magnetic resonance imaging of occult hip fractures. *Acta Radiol*. 2011;52(8):871-874.
- [605] Palaniappan R, Sundaraj K, Sundaraj S. A comparative study of the SVM and K-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. *BMC Bioinformatics*. 2014;15:223.
- [606] Amit G, Purdie TG, Levinshtein A, et al. Automatic learning-based beam angle selection for thoracic IMRT. *Med Phys*. 2015;42(4):1992-2005.
- [607] Zhang H, Meng D, Cai S, et al. The application of artificial intelligence in lung cancer: a narrative review. *Transl Cancer Res*. 2021;10(5):2478-2487.
- [608] Barron AR, Cohen A, Dahmen W, DeVore RA. Approximation and learning by greedy algorithms. *Ann. Statist*. 2008;36(1):64-94.
- [609] Tubiana M, Eschwège F. Conformal radiotherapy and intensity-modulated radiotherapy--clinical data. *Acta Oncol*. 2000;39(5):555-567.

- [610] Liu HH, Jauregui M, Zhang X, Wang X, Dong L, Mohan R. Beam angle optimization and reduction for intensity-modulated radiation therapy of non-small-cell lung cancers. *Int J Radiat Oncol Biol Phys*. 2006;65(2):561-572.
- [611] Bohaterewicz B, Sobczak AM, Podolak I, et al. Machine Learning-Based Identification of Suicidal Risk in Patients With Schizophrenia Using Multi-Level Resting-State fMRI Features. *Front Neurosci*. 2021;14:605697.
- [612] WISQARS™ — Web-based Injury Statistics Query and Reporting System. *Centers for Disease Control and Prevention*. Consulté le 20 avril 2023. <https://www.cdc.gov/injury/wisqars/index.html>
- [613] Wildgust HJ, Hodgson R, Beary M. The paradox of premature mortality in schizophrenia: new research questions. *J Psychopharmacol*. 2010;24(4 Suppl):9-15.
- [614] Bohaterewicz B, Jabłońska A, Szulc A. Suicidal thoughts and behaviors and their comorbidity with psychopathological symptoms among patients with schizophrenia hospitalized at long term ward. *Studia Psychologiczne*. 2018;56(1).
- [615] Cassidy RM, Yang F, Kapczinski F, Passos IC. Risk Factors for Suicidality in Patients With Schizophrenia: A Systematic Review, Meta-analysis, and Meta-regression of 96 Studies. *Schizophr Bull*. 2018;44(4):787-797.
- [616] Glenn CR, Nock MK. Improving the short-term prediction of suicidal behavior. *Am J Prev Med*. 2014;47(3 Suppl 2):S176-S180.
- [617] Busch KA, Fawcett J, Jacobs DG. Clinical correlates of inpatient suicide. *J Clin Psychiatry*. 2003;64(1):14-19.
- [618] Chang B, Choi Y, Jeon M, et al. ARPNet: Antidepressant Response Prediction Network for Major Depressive Disorder. *Genes (Basel)*. 2019;10(11):907.
- [619] Lin E, Lin CH, Lane HY. Precision Psychiatry Applications with Pharmacogenomics: Artificial Intelligence and Machine Learning Approaches. *Int J Mol Sci*. 2020;21(3):969.
- [620] Rush AJ, Wisniewski SR, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry*. 2008;65(8):870-880.
- [621] Cartwright C, Gibson K, Read J, Cowan O, Dehar T. Long-term antidepressant use: patient perspectives of benefits and adverse effects. *Patient Prefer Adherence*. 2016;10:1401-1407.
- [622] Fawcett J. The morbidity and mortality of clinical depression. *Int Clin Psychopharmacol*. 1993;8(4):217-220.
- [623] Ashinsky BG, Bouhrara M, Coletta CE, et al. Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *J Orthop Res*. 2017;35(10):2243-2250.
- [624] Davis MA. Epidemiology of osteoarthritis. *Clin Geriatr Med*. 1988;4(2):241-255.
- [625] Swedberg JA, Steinbauer JR. Osteoarthritis. *Am Fam Physician*. 1992;45(2):557-568.
- [626] Hunter DJ, Eckstein F. Exercise and osteoarthritis. *Journal of anatomy*. 2009;214:197-207.
- [627] Morales-Ivorra I, Romera-Baures M, Roman-Viñas B, Serra-Majem L. Osteoarthritis and the Mediterranean Diet: A Systematic Review. *Nutrients*. 2018;10(8):1030.
- [628] Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. *Bone*. 2012;51(2):278-288.

- [629] Wirries A, Geiger F, Hammad A, Oberkircher L, Blümcke I, Jabari S. Artificial intelligence facilitates decision-making in the treatment of lumbar disc herniations. *Eur Spine J.* 2021;30(8):2176-2184.
- [630] Pearson AM, Lurie JD. Surgical versus nonoperative treatment: how do we choose the right approach to lumbar disk herniation?. *Pain Manag.* 2014;4(4):247-249.
- [631] Bailey CS, Rasoulinejad P, Taylor D, et al. Surgery versus Conservative Care for Persistent Sciatica Lasting 4 to 12 Months. *N Engl J Med.* 2020;382(12):1093-1102.
- [632] Quon JA, Sobolev BG, Levy AR, et al. The effect of waiting time on pain intensity after elective surgical lumbar discectomy. *Spine J.* 2013;13(12):1736-1748.
- [633] Tran NK, Sen S, Palmieri TL, et al. Artificial intelligence and machine learning for predicting acute kidney injury in severely burned patients: A proof of concept. *Burns.* 2019;45(6):1350-1358.
- [634] Gutierrez G. Artificial Intelligence in the Intensive Care Unit. *Crit Care.* 2020;24(1):101.
- [635] Palmieri T, Lavrentieva A, Greenhalgh DG. Acute kidney injury in critically ill burn patients. Risk factors, progression and impact on mortality. *Burns.* 2010;36(2):205-211.
- [636] Kellum JA, Lameire N, Aspelin P, et al. Kidney disease: improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury. *Kidney international supplements.* 2012;2(1):1-138.
- [637] Chiou WL, Hsu FH. Pharmacokinetics of creatinine in man and its implications in the monitoring of renal function and in dosage regimen modifications in patients with renal insufficiency. *J Clin Pharmacol.* 1975;15(5-6):427-434.
- [638] Reinhard M, Erlandsen EJ, Randers E. Biological variation of cystatin C and creatinine. *Scand J Clin Lab Invest.* 2009;69(8):831-836.
- [639] Legrand M, Payen D. Understanding urine output in critically ill patients. *Ann Intensive Care.* 2011;1(1):13.
- [640] Howell E, Sen S, Palmieri T, et al. Point-of-care B-type natriuretic peptide and neutrophil gelatinase-associated lipocalin measurements for acute resuscitation: a pilot study. *J Burn Care Res.* 2015;36(2):e26-e33.
- [641] Loo NL, Chiew YS, Tan CP, et al. A machine learning model for real-time asynchronous breathing monitoring. *IFAC-PapersOnLine.* 2018;51(27):378-383.
- [642] Sassoon CS, Foster GT. Patient-ventilator asynchrony. *Curr Opin Crit Care.* 2001;7(1):28-33.
- [643] Blanch L, Villagra A, Sales B, et al. Asynchronies during mechanical ventilation are associated with mortality. *Intensive Care Med.* 2015;41(4):633-641.
- [644] de Wit M, Miller KB, Green DA, Ostman HE, Gennings C, Epstein SK. Ineffective triggering predicts increased duration of mechanical ventilation. *Crit Care Med.* 2009;37(10):2740-2745.
- [645] Dodd-Sullivan R, Quirin J, Newhart J. Ventilator autotriggering: a caution in brain death diagnosis. *Prog Transplant.* 2011;21(2):152-155.
- [646] Dres M, Rittayamai N, Brochard L. Monitoring patient-ventilator asynchrony. *Curr Opin Crit Care.* 2016;22(3):246-253.
- [647] Holanda MA, Vasconcelos RDS, Ferreira JC, Pinheiro BV. Patient-ventilator asynchrony. *J Bras Pneumol.* 2018;44(4):321-333.

- [648] Qiao X, Li H, Wang B, Xiong F, Yan Y, Wang L. Research on Classification of Patient-ventilator Asynchrony Using Permutation Disalignment Index. *2021 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (pp. 1-5). IEEE ; 2021.
- [649] Greenwald MF, Danford ID, Shahrawat M, et al. Evaluation of artificial intelligence-based telemedicine screening for retinopathy of prematurity. *J AAPOS*. 2020;24(3):160-162.
- [650] Sisman J, Campbell DE, Brion LP. Amplitude-integrated EEG in preterm infants: maturation of background pattern and amplitude voltage with postmenstrual age and gestational age. *J Perinatol*. 2005;25(6):391-396.
- [651] Gergely K, Gerinec A. Retinopathy of prematurity--epidemics, incidence, prevalence, blindness. *Bratisl Lek Listy*. 2010;111(9):514-517.
- [652] Ells AL, Holmes JM, Astle WF, et al. Telemedicine approach to screening for severe retinopathy of prematurity: a pilot study. *Ophthalmology*. 2003;110(11):2113-2117.
- [653] Trese MT. Subjectivity in retinopathy of prematurity screening. *Am J Ophthalmol*. 2015;160(3):406-407.
- [654] LeMoyné R, Mastroianni T, Hessel A, Nishikawa K. Ankle rehabilitation system with feedback from a smartphone wireless gyroscope platform and machine learning classification. *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 406-409). IEEE ; 2015.
- [655] Hurwitz LJ, Adams GF. Rehabilitation of hemiplegia: indices of assessment and prognosis. *Br Med J*. 1972;1(5792):94-98.
- [656] Everaert DG, Stein RB, Abrams GM, et al. Effect of a foot-drop stimulator and ankle-foot orthosis on walking performance after stroke: a multicenter randomized controlled trial. *Neurorehabil Neural Repair*. 2013;27(7):579-591.
- [657] Mills PM, Barrett RS, Morrison S. Toe clearance variability during walking in young and elderly men. *Gait Posture*. 2008;28(1):101-107.
- [658] Zhang M, Cao J, Xie SQ. A preliminary study on robot-assisted ankle rehabilitation for the treatment of drop foot. *Journal of Intelligent & Robotic Systems*. 2018;91:207-215.
- [659] Terada M, Pietrosimone BG, Gribble PA. Therapeutic interventions for increasing ankle dorsiflexion after ankle sprain: a systematic review. *J Athl Train*. 2013;48(5):696-709.
- [660] Chary M, Boyer EW, Burns MM. Diagnosis of Acute Poisoning using explainable artificial intelligence. *Comput Biol Med*. 2021;134:104469.
- [661] Van Hoving DJ, Veale DJH, Müller GF. Clinical Review: Emergency management of acute poisoning. *African Journal of Emergency Medicine*. 2011;1(2):69-78.
- [662] Lo-Ciganic WH, Huang JL, Zhang HH, et al. Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. *JAMA Netw Open*. 2019;2(3):e190968.
- [663] Pasricha SV, Tadrous M, Khuu W, et al. Clinical indications associated with opioid initiation for pain management in Ontario, Canada: a population-based cohort study. *Pain*. 2018;159(8):1562-1568.
- [664] MORPHINE (CHLORHYDRATE) LAVOISIER. VIDAL. Consulté le 20 avril 2023. <https://www.vidal.fr/medicaments/gammes/morphine-chlorhydrate-lavoisier-26775.html>
- [665] Hornberger J, Chhatwal J. Opioid Misuse: A Global Crisis. *Value Health*. 2021;24(2):145-146.

- [666] Volkow ND, Jones EB, Einstein EB, Wargo EM. Prevention and Treatment of Opioid Misuse and Addiction: A Review. *JAMA Psychiatry*. 2019;76(2):208-216.
- [667] Moccia S, Mattos LS, Patrini I, et al. Computer-assisted liver graft steatosis assessment via learning-based texture analysis. *Int J Comput Assist Radiol Surg*. 2018;13(9):1357-1367.
- [668] Chen CL, Fan ST, Lee SG, Makuuchi M, Tanaka K. Living-donor liver transplantation: 12 years of experience in Asia. *Transplantation*. 2003;75(3 Suppl):S6-S11.
- [669] Adam R, Hoti E. Liver transplantation: the current situation. *Semin Liver Dis*. 2009;29(1):3-18.
- [670] Armstrong MJ, Houlihan DD, Bentham L, et al. Presence and severity of non-alcoholic fatty liver disease in a large prospective primary care cohort. *J Hepatol*. 2012;56(1):234-240.
- [671] Karanjia RN, Crossey MM, Cox IJ, et al. Hepatic steatosis and fibrosis: Non-invasive assessment. *World J Gastroenterol*. 2016;22(45):9880-9897.
- [672] Mor E, Klintmalm GB, Gonwa TA, et al. The use of marginal donors for liver transplantation. A retrospective study of 365 liver donors. *Transplantation*. 1992;53(2):383-386.
- [673] Yersiz H, Lee C, Kaldas FM, et al. Assessment of hepatic steatosis by transplant surgeon and expert pathologist: a prospective, double-blind evaluation of 201 donor livers. *Liver Transpl*. 2013;19(4):437-449.
- [674] Koneru B, Dikdan G. Hepatic steatosis and liver transplantation current clinical and experimental perspectives. *Transplantation*. 2002;73(3):325-330.
- [675] Briceño J, Cruz-Ramírez M, Prieto M, et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. *J Hepatol*. 2014;61(5):1020-1028.
- [676] Feng S, Goodrich NP, Bragg-Gresham JL, et al. Characteristics associated with liver graft failure: the concept of a donor risk index [published correction appears in *Am J Transplant*. 2018 Dec;18(12):3085]. *Am J Transplant*. 2006;6(4):783-790.
- [677] Kamath PS, Kim WR; Advanced Liver Disease Study Group. The model for end-stage liver disease (MELD). *Hepatology*. 2007;45(3):797-805.
- [678] Rana A, Hardy MA, Halazun KJ, et al. Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation. *Am J Transplant*. 2008;8(12):2537-2546.
- [679] Dutkowski P, Oberkofler CE, Slankamenac K, et al. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. *Ann Surg*. 2011;254(5):745-753.
- [680] Merion RM, Sharma P, Mathur AK, Schaubel DE. Evidence-based development of liver allocation: a review. *Transpl Int*. 2011;24(10):965-972.
- [681] Wiesner RH, McDiarmid SV, Kamath PS, et al. MELD and PELD: application of survival models to liver allocation. *Liver Transpl*. 2001;7(7):567-580.
- [682] Wood NL, VanDerwerken DN, Segev DL, Gentry SE. Logistical burden of offers and allocation inefficiency in circle-based liver allocation. *Liver Transpl*. 2023;29(1):26-33.
- [683] Chung SW, Han SS, Lee JW, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. 2018;89(4):468-473.

- [684] Kalmet PHS, Sanduleanu S, Primakov S, et al. Deep learning in fracture detection: a narrative review. *Acta Orthop*. 2020;91(2):215-220.
- [685] Basti JJ, Dionysian E, Sherman PW, Bigliani LU. Management of proximal humeral fractures. *J Hand Ther*. 1994;7(2):111-121.
- [686] Mora Guix JM, Pedrós JS, Serrano AC. Updated classification system for proximal humeral fractures. *Clin Med Res*. 2009;7(1-2):32-44.
- [687] Foroohar A, Tosti R, Richmond JM, Gaughan JP, Ilyas AM. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. *J Orthop Surg Res*. 2011;6:38.
- [688] Borjali A, Chen AF, Muratoglu OK, Morid MA, Varadarajan KM. Detecting total hip replacement prosthesis design on plain radiographs using deep convolutional neural network. *J Orthop Res*. 2020;38(7):1465-1471.
- [689] Beyaz S. A brief history of artificial intelligence and robotic surgery in orthopedics & traumatology and future expectations. *Jt Dis Relat Surg*. 2020;31(3):653-655.
- [690] Laude F. Les révisions de prothèse totale de hanche par voie antérieure : intérêt, indications et limites. *Maîtrise Orthopédique*. 2014;238:20-5.
- [691] Wilson N, Broatch J, Jehn M, Davis C 3rd. National projections of time, cost and failure in implantable device identification: Consideration of unique device identification use. *Healthc (Amst)*. 2015;3(4):196-201.
- [692] Wilson NA, Jehn M, York S, Davis CM 3rd. Revision total hip and knee arthroplasty implant identification: implications for use of Unique Device Identification 2012 AAHKS member survey results. *J Arthroplasty*. 2014;29(2):251-255.
- [693] Ozkan IA, Koklu M, Sert IU. Diagnosis of urinary tract infection based on artificial intelligence methods. *Comput Methods Programs Biomed*. 2018;166:51-59.
- [694] Checcucci E, De Cillis S, Granato S, et al. Applications of neural networks in urology: a systematic review. *Curr Opin Urol*. 2020;30(6):788-807.
- [695] Geerlings SE. Clinical Presentations and Epidemiology of Urinary Tract Infections. *Microbiol Spectr*. 2016;4(5):10.1128/microbiolspec.UTI-0002-2012.
- [696] Nicolle LE. Urinary tract infection. *Crit Care Clin*. 2013;29(3):699-715.
- [697] Waller TA, Pantin SAL, Yenior AL, Pujalte GGA. Urinary Tract Infection Antibiotic Resistance in the United States. *Prim Care*. 2018;45(3):455-466.
- [698] Messer JS, Chew DJ, McLoughlin MA. Cystoscopy: techniques and clinical applications. *Clin Tech Small Anim Pract*. 2005;20(1):52-64.
- [699] Hung AJ, Chen J, Ghodoussipour S, et al. A deep-learning model using automated performance metrics and clinical features to predict urinary continence recovery after robot-assisted radical prostatectomy. *BJU Int*. 2019;124(3):487-495.
- [700] Pacik D, Fedorko M. Literature review of factors affecting continence after radical prostatectomy. *Saudi Med J*. 2017;38(1):9-17.
- [701] Ficarra V, Novara G, Rosen RC, et al. Systematic review and meta-analysis of studies reporting urinary continence recovery after robot-assisted radical prostatectomy. *Eur Urol*. 2012;62(3):405-417.

- [702] Herranz-Amo F. Radical retropubic prostatectomy: Preservation of urinary continence. La prostatectomía radical retropública: preservación de la continencia urinaria. *Actas Urol Esp (Engl Ed)*. 2020;44(10):674-681.
- [703] Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442.
- [704] Mungovan SF, Carlsson SV, Gass GC, et al. Preoperative exercise interventions to optimize continence outcomes following radical prostatectomy. *Nat Rev Urol*. 2021;18(5):259-281.
- [705] Sahu M, Gupta R, Ambasta RK, Kumar P. Artificial intelligence and machine learning in precision medicine: A paradigm shift in big data analysis. *Prog Mol Biol Transl Sci*. 2022;190(1):57-100.
- [706] Barrett M, Boyne J, Brandts J, et al. Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care. *EPMA J*. 2019;10(4):445-464.
- [707] Rashid MBMA. Artificial Intelligence Effecting a Paradigm Shift in Drug Development. *SLAS Technol*. 2021;26(1):3-15.
- [708] He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30-36.
- [709] Kerasidou A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bull World Health Organ*. 2020;98(4):245-250.
- [710] Haine N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The Bright, Artificial Intelligence-Augmented Future of Neuroimaging Reading. *Front Neurol*. 2017;8:489.
- [711] Kostić EJ, Pavlović DA, Živković MD. Applications of artificial intelligence in medicine and pharmacy: ethical aspects. *Acta Medica Medianae*. 2019;58(3):128-137.
- [712] Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230-243.
- [713] Jarrahi MH. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business horizons*. 2018;61(4):577-586.
- [714] Pieczynski J, Kuklo P, Grzybowski A. The Role of Telemedicine, In-Home Testing and Artificial Intelligence to Alleviate an Increasingly Burdened Healthcare System: Diabetic Retinopathy. *Ophthalmol Ther*. 2021;10(3):445-464.
- [715] Golding LP, Nicola GN. A Business Case for Artificial Intelligence Tools: The Currency of Improved Quality and Reduced Cost. *J Am Coll Radiol*. 2019;16(9 Pt B):1357-1361.
- [716] Meessen B. The Role of Digital Strategies in Financing Health Care for Universal Health Coverage in Low- and Middle-Income Countries. *Glob Health Sci Pract*. 2018;6(Suppl 1):S29-S40.
- [717] Garbuio M., Lin N. Artificial intelligence as a growth engine for health care startups: Emerging business models. *California Management Review*. 2019;61(2):59-83.
- [718] Gennatas ED, Chen JH. Artificial intelligence in medicine: past, present, and future. *Artificial Intelligence in Medicine* (pp. 3-18). Academic Press ; 2021.
- [719] Aljurayfani M, Alghernas S, Shargabi A. Medical self-diagnostic system using artificial neural networks. *2019 International Conference on Computer and Information Sciences (ICIS)* (pp. 1-5). IEEE ; 2019.
- [720] Hanslik T, Flahault A. La surmédicalisation: quand trop de médecine nuit à la santé [Overmedicalization: When too much medicine harms]. *Rev Med Interne*. 2016;37(3):201-205.

- [721] Willemin B. Numérisation des données de soin: impact sur l'élaboration et les fonctions du dossier médical. *Droit, Santé et Société*. 2018;5(5):33-41.
- [722] Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K, Ziemssen T. Digital Twins for Multiple Sclerosis. *Front Immunol*. 2021;12:669811. Published 2021 May 3.
- [723] Singh M, Fuenmayor E, Hinchey EP, Qiao Y, Murray N, Devine D. Digital twin: Origin to future. *Applied System Innovation*. 2021;4(2):36.
- [724] Lonsdale H, Gray GM, Ahumada LM, Yates HM, Varughese A, Rehman MA. The Perioperative Human Digital Twin. *Anesth Analg*. 2022;134(4):885-892.
- [725] Sinisi S, Alimguzhin V, Mancini T, Tronci E, Mari F, Leeners B. Optimal personalised treatment computation through in silico clinical trials on patient digital twins. *Fundamenta Informaticae*. 2020;174(3-4):283-310.
- [726] Bertezenne S. Le jumeau numérique en santé - Apports organisationnels et limites épistémologiques dans un contexte de crise sanitaire [The digital twin in health: Organizational contributions and epistemological limits in a context of health crisis]. *Med Sci (Paris)*. 2022;38(8-9):663-668.
- [727] Matheny M, Israni ST, Ahmed M, Whicher D. Artificial intelligence in health care: The hope, the hype, the promise, the peril. *Washington, DC: National Academy of Medicine*. 2019.
- [728] Gupta N, Patel H, Afzal S, et al. Data Quality Toolkit: Automatic assessment of data quality and remediation for machine learning datasets. *ArXiv, abs/2108.05935*. 2022.
- [729] Antoniou T, Mamdani M. Evaluation of machine learning solutions in medicine. *CMAJ*. 2021;193(36):E1425-E1429.
- [730] Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect*. 2020;26(5):584-595.
- [731] Intelligence Artificielle & Néphrologie 2023. *Artificial Intelligence + MEDICINE*. Consulté le 20 avril 2023. <https://targeting-ai.com/>
- [732] Derathé A, Reche F, Jannin P, Moreau-Gaudry A, Gibaud B, Voros S. Explaining a model predicting quality of surgical practice: a first presentation to and review by clinical experts. *Int J Comput Assist Radiol Surg*. 2021;16(11):2009-2019.
- [733] Florian E, Sgarbossa F, Zennaro I. Machine learning-based predictive maintenance: A cost-oriented model for implementation. *International Journal of Production Economics*. 2021;236:108114.
- [734] Dimitrakakis C, Savu-Krohn C. Cost-minimising strategies for data labelling: optimal stopping and active learning. *Foundations of Information and Knowledge Systems: 5th International Symposium, FoIKS 2008, Pisa, Italy, February 11-15, 2008. Proceedings 5* (pp. 96-111). Springer Berlin Heidelberg ; 2008.
- [735] Rubeis G, Fang ML, Sixsmith A. Equity in AgeTech for Ageing Well in Technology-Driven Places: The Role of Social Determinants in Designing AI-based Assistive Technologies. *Sci Eng Ethics*. 2022;28(6):49.
- [736] La prise en charge des dispositifs médicaux par l'assurance maladie. *Haute Autorité de Santé*. Consulté le 20 avril 2023. https://www.has-sante.fr/jcms/c_930603/fr/la-prise-en-charge-des-dispositifs-medicaux-par-l-assurance-maladie
- [737] Pucchio A, Eisenhauer EA, Moraes FY. Medical students need artificial intelligence and machine learning training. *Nat Biotechnol*. 2021;39(3):388-389.

- [738] Alves VM, Auerbach SS, Kleinstreuer N, et al. Curated Data In - Trustworthy In Silico Models Out: The Impact of Data Quality on the Reliability of Artificial Intelligence Models as Alternatives to Animal Testing. *Altern Lab Anim*. 2021;49(3):73-82.
- [739] Sidi F, Panahy PHS, Affendey LS, Jabar MA, Ibrahim H, Mustapha A. Data quality: A survey of data quality dimensions. *2012 International Conference on Information Retrieval & Knowledge Management* (pp. 300-304). IEEE ; 2012.
- [740] Chen Z, Lin Z, Wang P, Ding M. Negative-ResNet: noisy ambulatory electrocardiogram signal classification scheme. *Neural Computing and Applications*. 2021;33:8857-8869.
- [741] Wang D, Shang Y. A new active labeling method for deep learning. *2014 International joint conference on neural networks (IJCNN)* (pp. 112-119). IEEE ; 2014.
- [742] Drummond C, Holte RC. C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. *Workshop on learning from imbalanced datasets II*. 2003;11:1-8.
- [743] Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017;356(6334):183-186.
- [744] Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR ; 2018.
- [745] Varona D, Suárez JL. Discrimination, Bias, Fairness, and Trustworthy AI. *Applied Sciences*. 2022;12(12):5826.
- [746] Santé numérique : quand la recherche monte en puissance. *Inserm*. Consulté le 20 avril 2023. <https://www.inserm.fr/actualite/sante-numerique-quand-la-recherche-monte-en-puissance/>
- [747] Adnan K, Akbar R, Khor SW, Ali ABA. Role and challenges of unstructured big data in healthcare. *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019*. 2020;1:301-323.
- [748] Attrey R, Levit A. The promise of natural language processing in healthcare. *University of Western Ontario Medical Journal*. 2018;87(2):21-23.
- [749] Li I, Pan J, Goldwasser J. Neural Natural Language Processing for unstructured data in electronic health records: A review. *Computer Science Review*. 2022;46:100511.
- [750] Ramachandram D, Taylor GW. Deep multimodal learning: A survey on recent advances and trends. *IEEE signal processing magazine*. 2017;34(6):96-108.
- [751] Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE ; 2018.
- [752] Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: A survey on methods and metrics. *Electronics*. 2019;8(8):832.
- [753] Oh SJ, Schiele B, Fritz M. Towards reverse-engineering black-box neural networks. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. 2019:121-144.
- [754] Reddy S. Explainability and artificial intelligence in medicine. *Lancet Digit Health*. 2022;4(4):e214-e215.
- [755] Djulbegovic B, Hozo I, Greenland S. Uncertainty in clinical medicine. *Philosophy of medicine* (pp. 299-356). North-Holland ; 2011.

- [756] Cataldo R, Arancibia M, Stojanova J, Papuzinski C. General concepts in biostatistics and clinical epidemiology: Observational studies with cross-sectional and ecological designs. *Conceptos generales en bioestadística y epidemiología clínica: estudios observacionales con diseños transversal y ecológico. Medwave.* 2019;19(8):e7698.
- [757] Minh D, Wang HX., Li,YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review.* 2022;1-66.
- [758] Mishra P. Model Explainability for Rule-Based Expert Systems. *Practical Explainable AI Using Python: Artificial Intelligence Model Explanations Using Python-based Libraries, Extensions, and Frameworks* (pp. 315-326). Berkeley, CA: Apress ; 2021.
- [759] Muhammad MB, Yeasin M. Eigen-cam: Class activation map using principal components. *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE ; 2020.
- [760] Zhang J, Li X, Li Y, et al. Three dimensional convolutional neural network-based classification of conduct disorder with structural MRI. *Brain Imaging Behav.* 2020;14(6):2333-2340.
- [761] Sellam T, Lin K, Huang I, Yang M, Vondrick C, Wu E. Deepbase: Deep inspection of neural networks. *Proceedings of the 2019 International Conference on Management of Data* (pp. 1117-1134). 2019.
- [762] Gunning D. Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA). and Web.* 2017;2(2):1.
- [763] Matheny M, Israni ST, Ahmed M, Whicher D. Artificial intelligence in health care: The hope, the hype, the promise, the peril. *Washington, DC: National Academy of Medicine.* 2019.
- [764] Wiesing U. The Hippocratic Oath and the Declaration of Geneva: legitimisation attempts of professional conduct. *Med Health Care Philos.* 2020;23(1):81-86.
- [765] de Bienassis K, Slawomirski L, Klazinga NS. The economics of patient safety Part IV: Safety in the workplace: Occupational safety as the bedrock of resilient health systems. 2021.
- [766] Souza Filho EM, Fernandes FA, Pereira NCA, Mesquita CT, Gismondi RA. Ethics, Artificial Intelligence and Cardiology. *Ética, Inteligência Artificial e Cardiologia. Arq Bras Cardiol.* 2020;115(3):579-583.
- [767] Thomasian NM, Eickhoff C, Adashi EY. Advancing health equity with artificial intelligence. *J Public Health Policy.* 2021;42(4):602-611.
- [768] Andreu-Perez J, Deligianni F, Ravi D, Yang GZ. Artificial intelligence and robotics. *ArXiv, abs/1803.10813.* 2018.
- [769] Zhou XY, Guo Y, Shen M, Yang GZ. Application of artificial intelligence in surgery. *Front Med.* 2020;14(4):417-430.
- [770] Saeidi H, Opfermann JD, Kam M, et al. Autonomous robotic laparoscopic surgery for intestinal anastomosis. *Sci Robot.* 2022;7(62):eabj2908.
- [771] Nayak S, Das RK. Application of artificial intelligence (AI) in prosthetic and orthotic rehabilitation. *Service Robotics.* IntechOpen ; 2020.
- [772] Yan A. How a robot passed China's medical licensing exam. *South China Morning Post.* 20 novembre 2017.
- [773] Jalinière H. Xiaoyi, le petit docteur chinois robotisé est testé depuis un an dans plusieurs centres de santé. *Sciences et Avenir.* 11 novembre 2018.
- [774] Arnaud D. La médecine ne peut pas se résumer à de la technique. *Libération.* 9 décembre 2022.

- [775] Formarier M. La relation de soin, concepts et finalités [The caring relationship, concepts and conclusions]. *Rech Soins Infirm.* 2007;(89):33-42.
- [776] Yalcin ÖN, DiPaola S. A computational model of empathy for interactive agents. *Biologically inspired cognitive architectures.* 2018;26:20-25.
- [777] Hojat M, Louis DZ, Maio V, Gonnella JS. Empathy and health care quality. *Am J Med Qual.* 2013;28(1):6-7.
- [778] Kuziemy C, Maeder AJ, John O, et al. Role of Artificial Intelligence within the Telehealth Domain. *Yearb Med Inform.* 2019;28(1):35-40.
- [779] Avis de la CERNA. *Éthique de la recherche en apprentissage machine.* ALLISTENE ; 2017.
- [780] Singh J, Walden I, Crowcroft J, Bacon J. *Responsibility & machine learning: Part of a process.* Available at SSRN 2860048 ; 2016.
- [781] Douville T. Les dangers de la collecte des données de santé par les tiers intéressés (gafam, assureurs...). *Journal du Droit de la Sante et de l'Assurance-Maladie (JDSAM).* 2018;20(3):12-16.
- [782] Su Z, Togay G, Côté AM. Artificial intelligence: a destructive and yet creative force in the skilled labour market. *Human Resource Development International.* 2021;24(3):341-352.
- [783] Harper S. The challenges of twenty-first-century demography. *Challenges of aging: Pensions, retirement and generational justice.* 2015:17-29.
- [784] Kirch DG, Petelle K. Addressing the Physician Shortage: The Peril of Ignoring Demography. *JAMA.* 2017;317(19):1947-1948.
- [785] Christianto V, Smarandache F. *Remark on Artificial Intelligence, humanoid and Terminator scenario: A Neutrosophic way to futurology.* Infinite Study ; 2020.
- [786] Price II, Nicholson W. Artificial intelligence in health care: applications and legal issues. *14 SciTech Lawyer 10. U of Michigan Public Law Research Paper.* 2017;599.
- [787] Freeman L. Test and evaluation for artificial intelligence. *Insight.* 2020;23(1):27-30.
- [788] Polyzotis N, Roy S, Whang SE, Zinkevich M. Data lifecycle challenges in production machine learning: a survey. *ACM SIGMOD Record.* 2018;47(2):17-28.
- [789] Cytermann L. Promesses et risques de l'open et du big data: les réponses du droit. *Informations sociales.* 2015;5,80-90.
- [790] Donnée sensible. *CNIL.* Consulté le 20 avril 2023. <https://www.cnil.fr/fr/definition/donnee-sensible>
- [791] The Global AI Index. *Tortoise.* Consulté le 20 avril 2023. <https://www.tortoisemedia.com/intelligence/global-ai/>
- [792] Laws & Regulations. *United States Department of Health and Human Services.* Consulté le 20 avril 2023. <https://www.hhs.gov/regulations/index.html>
- [793] The HIPAA Privacy Rule. *United States Department of Health and Human Services.* Consulté le 20 avril 2023. <https://www.hhs.gov/hipaa/for-professionals/privacy/index.html>
- [794] Cyberspace Administration of China launches official website. *State Council of the People's Republic of China.* Consulté le 20 avril 2023. http://english.www.gov.cn/news/top_news/2014/12/31/content_281475032291728.htm

- [795] The PRC Personal Information Protection Law (Final): A Full Translation. *China Briefing*. Consulté le 20 avril 2023. <https://www.china-briefing.com/news/the-prc-personal-information-protection-law-final-a-full-translation/>
- [796] Legislation we cover. *Information Commissioner's Office*. Consulté le 20 avril 2023. <https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/>
- [797] UK General Data Protection Regulation. *Information Commissioner's Office*. Consulté le 20 avril 2023. <https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/general-data-protection-regulation/>
- [798] About the OPC. *Office of the Privacy Commissioner of Canada*. Consulté le 20 avril 2023. <https://www.priv.gc.ca/en/about-the-opc/>
- [799] The Personal Information Protection and Electronic Documents Act (PIPEDA). *Office of the Privacy Commissioner of Canada*. Consulté le 20 avril 2023. <https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/>
- [800] The Privacy Protection Authority, About. *gov.il - Government services and information*. Consulté le 20 avril 2023. https://www.gov.il/en/departments/about/about_ppa
- [801] Protection of privacy regulations (data security). *gov.il - Government services and information*. Consulté le 20 avril 2023. https://www.gov.il/en/Departments/legalInfo/data_security_regulation
- [802] pdpc, About Us. *Personal Data Protection Commission*. Consulté le 20 avril 2023. <https://www.pdpc.gov.sg/Who-We-Are/About-Us>
- [803] PDPA Overview. *Personal Data Protection Commission*. Consulté le 20 avril 2023. <https://www.pdpc.gov.sg/Overview-of-PDPA/The-Legislation/Personal-Data-Protection-Act>
- [804] Vision and Mission. *Personal Information Protection Commission*. Consulté le 20 avril 2023. <https://www.pipc.go.kr/eng/user/itc/visionMission.do>
- [805] Law. *Personal Information Protection Commission*. Consulté le 20 avril 2023. <https://www.pipc.go.kr/eng/user/lgp/law/lawsRegulations.do#none>
- [806] Tasks and powers of the Dutch DPA. *Autoriteit Persoonsgegevens*. Consulté le 20 avril 2023. <https://www.autoriteitpersoonsgegevens.nl/en/about-dutch-dpa/tasks-and-powers-dutch-dpa>
- [807] RÈGLEMENT (UE) 2016/679 DU PARLEMENT EUROPÉEN ET DU CONSEIL. *EUR-Lex*. Consulté le 20 avril 2023. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>
- [808] Tasks and powers. *Bundesbeauftragte für den Datenschutz und die Informationsfreiheit*. Consulté le 20 avril 2023. https://www.bfdi.bund.de/EN/DerBfDI/UeberUns/DieBehoerde/diebehoerde_node.html
- [809] Les missions de la CNIL. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/les-missions-de-la-cnil>
- [810] About the OAIC. *Office of the Australian Information Commissioner*. Consulté le 20 avril 2023. <https://www.oaic.gov.au/about-the-OAIC>
- [811] The Privacy Act. *Office of the Australian Information Commissioner*. Consulté le 20 avril 2023. <https://www.oaic.gov.au/privacy/privacy-legislation/the-privacy-act>
- [811] Who We Are. *Data Protection Commission*. Consulté le 20 avril 2023. <https://www.dataprotection.ie/en/who-we-are#>
- [812] Duties of the Data Protection Ombudsman. *Office of the Data Protection Ombudsman*. Consulté le 20 avril 2023. <https://tietosuoja.fi/en/duties>

- [813] What we do. *Datatilsynet*. Consulté le 20 avril 2023. <https://www.datatilsynet.dk/english/about-us/what-we-do>
- [814] CNPD, Duties. *Commission Nationale pour la Protection des Données*. Consulté le 20 avril 2023. <https://cnpd.public.lu/en/commission-nationale/missions.html>
- [815] Roles and Responsibilities. *Personal Information Protection Commission JAPAN*. Consulté le 20 avril 2023. <https://www.ppc.go.jp/en/aboutus/roles/>
- [816] Act on the Protection of Personal Information. *Japanese Law Translation*. Consulté le 20 avril 2023. <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>
- [817] THE INFORMATION TECHNOLOGY ACT, 2000 (No. 21 OF 2000). *MINISTRY OF LAW, JUSTICE AND COMPANY AFFAIRS (Legislative Department)*. Consulté le 20 avril 2023. <https://eprocure.gov.in/cppp/rulesandprocs/kbadqkdlcswfjdelrquehwuxcfmijmuixngudufgbuubgubfugbububjxcgfvsbdihbgfGhdfgFHtyhRtMjk4NzY=>
- [818] Préposé fédéral à la protection des données et à la transparence PFPDT. *Chancellerie fédérale*. Consulté le 20 avril 2023. <https://www.bk.admin.ch/bk/fr/home/chancellerie-federale/organisation-de-la-chancellerie-federale/prepose-federal-a-la-protection-des-donnees-et-a-la-transparence.html>
- [819] Loi fédérale sur la protection des données. *Chancellerie fédérale*. Consulté le 20 avril 2023. https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/fr
- [820] Our mission. *Swedish Authority for Privacy Protection*. Consulté le 20 avril 2023. <https://www.imy.se/en/about-us/swedish-authority-for-privacy-protections-assignment/>
- [821] What We Do. *Office of the Privacy Commissioner for Personal Data*. Consulté le 20 avril 2023. https://www.pcpd.org.hk/english/about_pcpd/our_role/what_we_do.html
- [822] The Personal Data (Privacy) Ordinance. *Office of the Privacy Commissioner for Personal Data*. Consulté le 20 avril 2023. https://www.pcpd.org.hk/english/data_privacy_law/ordinance_at_a_Glance/ordinance.html
- [823] Barati M, Rana O. Tracking GDPR compliance in cloud-based service delivery. *IEEE Transactions on Services Computing*. 2020;15(3):1498-1511.
- [824] Qu'est-ce qu'une donnée de santé ?. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/quest-ce-que-une-donnee-de-sante>
- [825] Définir une finalité. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/definir-une-finalite>
- [826] Quelles formalités pour les traitements de données de santé à caractère personnel ?. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/quelles-formalites-pour-les-traitements-de-donnees-de-sante-caractere-personnel>
- [827] CHAPITRE III - Droits de la personne concernée. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre3>
- [828] Garantir la sécurité des données. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/garantir-la-securite-des-donnees>
- [829] Le délégué à la protection des données (DPO). *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/le-delegue-la-protection-des-donnees-dpo>
- [830] Lemaître V, Khallouki D, Millard F, Lesueur T. *Le principe de proportionnalité dans l'évaluation environnementale*. Commissariat général au développement durable ; 2019.

- [831] Analyse d'impact (AIPD). *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/definition/analyse-dimpact-aid>
- [832] Le registre des activités de traitement. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/RGDP-le-registre-des-activites-de-traitement>
- [833] L'anonymisation de données personnelles. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>
- [834] IA : comment être en conformité avec le RGPD ?. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/intelligence-artificielle/ia-comment-etre-en-conformite-avec-le-rgpd>
- [835] CHAPITRE VII - Coopération et cohérence. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre7>
- [836] Regnaud I. Un an de RGPD : l'heure du premier bilan. *Les Echos*. 21 mai 2019.
- [837] Siggia S. RGPD : source d'inspiration pour la protection des données personnelles. *Pideeco*. 25 janvier 2022.
- [838] La protection des données dans le monde. *CNIL*. Consulté le 20 avril 2023. <https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde>
- [839] Questions & réponses sur la décision d'adéquation concernant le Japon. *Commission Européenne*. Consulté le 20 avril 2023. https://ec.europa.eu/commission/presscorner/detail/fr/MEMO_18_4503
- [840] ARTICLE 29 DATA PROTECTION WORKING PARTY. 6 février 2018. <https://ec.europa.eu/newsroom/article29/redirection/document/57550>
- [841] Fischer C. The CLOUD Act: A dangerous expansion of police snooping on cross-border data. *Electronic Frontier Foundation*. 8 février 2018.
- [842] Amazon Machine Learning: Guide du développeur. *Amazon Web Services*. Consulté le 20 avril 2023. https://docs.aws.amazon.com/fr_fr/machine-learning/latest/dg/machinelearning-dg.pdf#resources
- [843] Géron A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc. ; 2022.
- [844] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*. 2002;64(4):583-639.
- [845] Alridha A, Wahbi FA, Kadhim, MK. Training analysis of optimization models in machine learning. *International Journal of Nonlinear Analysis and Applications*. 2021;12(2):1453-1461.
- [846] Bertrand D, Fluss J, Billard C, Ziegler JC. Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture. *L'Année psychologique*. 2010;110(2):299-320.
- [847] Delacour H, Servonnet A, Perrot A, Vigezzi JF, Ramirez JM. La courbe ROC (receiver operating characteristic): principes et principales applications en biologie clinique [ROC (receiver operating characteristics) curve: principles and application in biology]. *Ann Biol Clin (Paris)*. 2005;63(2):145-154.
- [848] Goeldlin AO, Perrig M. Examen clinique base sur l'évidence. *PRIMARY AND HOSPITAL CARE – MÉDECINE INTERNE GÉNÉRALE*. 2016;16(6):109-112.
- [849] McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002;17(8):646-649.
- [850] Schmid M, Wright MN, Ziegler A. On the use of Harrell's C for clinical risk prediction via random survival forests. *Expert Systems with Applications*. 2016;63:450-459.

- [851] Magboo VPC, Magboo MSA. Machine learning classifiers on breast cancer recurrences. *Procedia Computer Science*. 2021;192:2742-2752.
- [852] Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol Pharm*. 2017;14(12):4462-4475.
- [853] Dab W. 18. La sensibilité, la spécificité et les valeurs prédictives. *Les fondamentaux de l'épidémiologie* (pp. 113-117). Presses de l'EHESP, « Vade-mecum Pro » ; 2021.
- [854] Bertrand D, Fluss J, Billard C, Ziegler JC. Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture. *L'Année psychologique*. 2010;110(2):299-320.
- [855] Soun JE, Chow DS, Nagamine M, et al. Artificial Intelligence and Acute Stroke Imaging. *AJNR Am J Neuroradiol*. 2021;42(1):2-11.
- [856] Wang W, Lu Y. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP conference series: materials science and engineering* (Vol. 324, No. 1, p. 012049). IOP Publishing ; 2018.
- [857] Khair U, Fahmi H, Al Hakim S, Rahim R. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. *journal of physics: conference series* (Vol. 930, No. 1, p. 012002). IOP Publishing ; 2017.
- [858] Liu H, Erdem E, Shi J. Comprehensive evaluation of ARMA–GARCH (-M) approaches for modeling the mean and volatility of wind speed. *Applied Energy*. 2011;88(3):724-732.
- [859] Tony Cai T, Guo Z. Semisupervised inference for explained variance in high dimensional linear regression and its applications. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2020;82(2):391-419.
- [860] Gómez-Ramírez J, Fernández-Blázquez MA, González-Rosa JJ. Prediction of Chronological Age in Healthy Elderly Subjects with Machine Learning from MRI Brain Segmentation and Cortical Parcellation. *Brain Sci*. 2022;12(5):579.
- [861] Cho BH, Kaji D, Cheung ZB, et al. Automated Measurement of Lumbar Lordosis on Radiographs Using Machine Learning and Computer Vision. *Global Spine J*. 2020;10(5):611-618.
- [862] Istaiteh O, Owais T, Al-Madi N, Abu-Soud S. Machine learning approaches for covid-19 forecasting. *2020 international conference on intelligent data science technologies and applications (IDSTA)* (pp. 50-57). IEEE ; 2020.
- [863] Ogink PT, Karhade AV, Thio QCBS, et al. Development of a machine learning algorithm predicting discharge placement after surgery for spondylolisthesis. *Eur Spine J*. 2019;28(8):1775-1782.
- [864] Drton M, Plummer M, Claeskens G, et al. A Bayesian information criterion for singular models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. 2017;323-380.
- [865] Maleki F, Muthukrishnan N, Ovens K, Reinhold C, Forghani R. Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment. *Neuroimaging Clin N Am*. 2020;30(4):433-445.
- [866] Ranjan GSK, Verma AK, Radhika S. K-nearest neighbors and grid search cv based real time fault monitoring system for industries. *2019 IEEE 5th international conference for convergence in technology (I2CT)* (pp. 1-5). IEEE ; 2019.

- [867] Rasid A, Kenedy S. IMPLEMENTATION OF SUPPORT VECTOR MACHINE ALGORITHM WITH HYPER-TUNING RANDOMIZED SEARCH IN STROKE PREDICTION. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*. 2023;6(2):61-65.
- [868] Jacobusse G, Veenman C. On Selection Bias with Imbalanced Classes. *Discovery Science. 19th International Conference, DS 2016, Bari, Italy, October 19–21, 2016, Proceedings* (pp. 325–340). Springer International Publishing ; 2016.
- [869] Kouw WM, Krijthe JH, Loog M. Robust importance-weighted cross-validation under sample selection bias. *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE ; 2019.
- [870] Tsamardinos I, Greasidou E, Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach Learn*. 2018;107(12):1895-1922.
- [871] Chrystal KA, Mizen PD, Mizen PD. Goodhart's Law: its origins, meaning and implications for monetary policy. *Central banking, monetary theory and practice: Essays in honour of Charles Goodhart*. 2003;1:221-243.
- [872] Hennessy C, Goodhart C. Goodhart's Law and Machine Learning: A Structural Perspective. *Available at SSRN 3639508*. 2021.
- [873] RÈGLEMENT (UE) 2017/745 DU PARLEMENT EUROPÉEN ET DU CONSEIL du 5 avril 2017 relatif aux dispositifs médicaux, modifiant la directive 2001/83/CE, le règlement (CE) no 178/2002 et le règlement (CE) no 1223/2009 et abrogeant les directives du Conseil 90/385/CEE et 93/42/CEE. 5 mai 2017. <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32017R0745>
- [874] Medical Device Coordination Group. *Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR*. 2019. https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2019_11_guidance_qualification_classification_software_en_0.pdf
- [875] Agence nationale de sécurité du médicament et des produits de santé. *Cybersécurité des Dispositifs Médicaux Intégrant du Logiciel Au cours de leur Cycle de Vie*. 2022. <https://ansm.sante.fr/uploads/2022/09/23/20220923-recommandations-ansm-cybersecurite-des-dmil.pdf>
- [876] Afnor. *NF EN ISO 14971*. 2019. <https://www.boutique.afnor.org/fr-fr/norme/nf-en-iso-14971/dispositifs-medicaux-application-de-la-gestion-des-risques-aux-dispositifs-/fa190418/1838>
- [877] Afnor. *NF EN 62304*. 2006. <https://www.boutique.afnor.org/fr-fr/norme/nf-en-62304/logiciels-de-dispositifs-medicaux-processus-du-cycle-de-vie-du-logiciel/fa140619/790>
- [878] Afnor. *NF EN 60601-1-4*. 1999. <https://www.boutique.afnor.org/fr-fr/norme/nf-en-6060114/appareils-electromedicaux-partie-1-regles-generales-de-securite-4-norme-col/fa042666/5262>
- [879] Afnor. *NF EN ISO 13485*. 2016. <https://www.boutique.afnor.org/fr-fr/norme/nf-en-iso-13485/dispositifs-medicaux-systemes-de-management-de-la-qualite-exigences-a-des-f/fa161550/1575>
- [879] Mise sur le marché des dispositifs médicaux et des dispositifs médicaux de diagnostic in vitro. *Agence nationale de sécurité du médicament et des produits de santé*. Consulté le 21 avril 2023. <https://ansm.sante.fr/uploads/2022/09/23/20220923-recommandations-ansm-cybersecurite-des-dmil.pdf>

- [880] Agence nationale de sécurité du médicament et des produits de santé. *Certification des dispositifs médicaux : le GMED est désigné comme organisme notifié au titre du nouveau règlement européen*. 2021. <https://ansm.sante.fr/actualites/certification-des-dispositifs-medicaux-le-gmed-est-designe-comme-organisme-notifie-au-titre-du-nouveau-reglement-europeen>
- [881] Haute Autorité de Santé. *Parcours du dispositif médical en France*. 2021. https://www.has-sante.fr/upload/docs/application/pdf/2009-12/guide_pratique_dm.pdf
- [882] Haute Autorité de Santé. *Forfait innovation : dépôt de dossier de demande de prise en charge dérogatoire pour un produit innovant – L.165-1-1 du CSS – Dispositif médical, dispositif médical de diagnostic in vitro ou acte*. 2020. https://has-sante.fr/upload/docs/application/pdf/2020-10/fi_guide_de_depot_2020_10_01.pdf
- [883] Décret n° 2022-1767 du 30 décembre 2022 relatif à la prise en charge et au remboursement des activités de télésurveillance médicale. 31 décembre 2022. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000046849110>
- [884] SNITEM. *Logiciel et intelligence artificielle*. 2021. <https://www.snitem.fr/publications/fiches-et-syntheses/logiciel-et-ia/>
- [885] SNITEM. *Les 4 étapes du marquage CE médical*. 2021. <https://www.snitem.fr/publications/fiches-et-syntheses/les-4-etapes-du-marquage-ce-medical/>
- [886] Agence nationale de sécurité du médicament et des produits de santé. *Déclarer un effet indésirable*. 2023. <https://ansm.sante.fr/documents/referance/declarer-un-effet-indesirable>
- [887] Les suites de l'inspection : mesures administratives. *Agence nationale de sécurité du médicament et des produits de santé*. Consulté le 21 avril 2023. <https://ansm.sante.fr/page/les-suites-de-linspection-mesures-administratives>
- [888] Aredoc. *LA RESPONSABILITÉ MÉDICALE EN CAS DE DOMMAGES IMPLIQUANT UN PRODUIT DE SANTÉ OU UN DISPOSITIF MÉDICAL*. 2013. <https://www.aredoc.com/wp-content/uploads/2017/10/Brochure-La-responsabilite-medicale-en-cas-de-dommages-impliquant-un-produit-de-sante-ou-un-dispositif-medical-2013.pdf>
- [889] Gepperth A, Hammer B. Incremental learning algorithms and applications. *European symposium on artificial neural networks (ESANN)*. 2016.
- [890] Rich BA. Medical custom and medical ethics: rethinking the standard of care. *Camb Q Healthc Ethics*. 2005;14(1):27-39.
- [891] Gruson D, Grass E. Quelle régulation positive éthique de l'intelligence artificielle en santé ?. *Les Tribunes de la santé*. 2020;63:25-33.
- [892] Lequillier C. L'impact de l'IA sur la relation de soin. *Journal du Droit de la Santé et de l'Assurance - Maladie (JDSAM)*. 2020;25:84-91.
- [893] Breton X, Touraine JL. RAPPORT D'INFORMATION DÉPOSÉ en application de l'article 145 du Règlement PAR LA MISSION D'INFORMATION sur la révision de la loi relative à la bioéthique. *ASSEMBLÉE NATIONALE*. 23 janvier 2019.
- [894] Arnold M, Singh D, Laversanne M, et al. Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040. *JAMA Dermatol*. 2022;158(5):495-503.
- [895] Worldwide cancer data. *World Cancer Research Fund International*. Consulté le 21 avril 2023. <https://www.wcrf.org/cancer-trends/worldwide-cancer-data/>

- [896] All Cancer Sites Combined, Long-Term Trends in SEER Age-Adjusted Incidence Rates, 1975-2020. *SEER*Explorer Application*. Consulté le 21 avril 2023. https://seer.cancer.gov/statistics-network/explorer/application.html?site=1&data_type=1&graph_type=1&compareBy=sex&chk_sex_1=1&rate_type=2&race=1&age_range=1&adopt_precision=1&adopt_show_ci=on&hdn_view=0&adopt_show_apc=on&adopt_display=2#resultsRegion0
- [897] Melanoma of the Skin, Long-Term Trends in SEER Age-Adjusted Incidence Rates, 1975-2020. *SEER*Explorer Application*. Consulté le 21 avril 2023. https://seer.cancer.gov/statistics-network/explorer/application.html?site=53&data_type=1&graph_type=1&compareBy=sex&chk_sex_1=1&rate_type=2&race=1&age_range=1&adopt_precision=1&adopt_show_ci=on&hdn_view=0&adopt_show_apc=on&adopt_display=2
- [898] Terushkin V, Halpern AC. Melanoma early detection. *Hematol Oncol Clin North Am*. 2009;23(3):481-viii.
- [899] Skin Cancer: Malignant vs. Benign. *Kaggle*. Consulté le 21 avril 2023. <https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign?datasetId=174469>
- [900] Melanoma Detection Dataset. *Kaggle*. Consulté le 21 avril 2023. <https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection/code>
- [901] Pillow 8.4.0. *Paquets Python*. Consulté le 21 avril 2023. <https://pypi.org/project/Pillow/8.4.0/>
- [902] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *Journal of big data*. 2019;6(1):1-48.
- [903] torchvision.models. *PyTorch*. Consulté le 21 avril 2023. <https://pytorch.org/vision/0.8/models.html>
- [904] Pak M, Kim S. A review of deep learning in image recognition. *2017 4th international conference on computer applications and information processing technology (CAIPT)* (pp. 1-3). IEEE ; 2017.
- [905] Zhuang F, Qi Z, Duan K. A comprehensive survey on transfer learning. *Proceedings of the IEEE*. 2020;109(1):43-76.
- [906] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv, abs/1409.1556*. 2014.
- [907] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770-778.
- [908] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*. 2015:1026-1034.
- [909] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14* (pp. 630-645). Springer International Publishing ; 2016.
- [910] Grenier B. *Justifier les décisions médicales et maîtriser les coûts* (pp. 45-63). Masson ; 2006.
- [911] « La dermatologie est au bord du précipice ». *Le Monde*. 11 mars 2023.
- [912] Mortaz E. Imbalance accuracy metric for model selection in multi-class imbalance classification problems. *Knowledge-Based Systems*. 2020;210:106490.
- [913] PyTorch documentation. *PyTorch*. Consulté le 21 avril 2023. <https://pytorch.org/docs/1.8.1/>

- [914] Maleki F, Muthukrishnan N, Ovens K, Reinhold C, Forghani R. Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment. *Neuroimaging Clin N Am*. 2020;30(4):433-445.
- [915] StepLR. *PyTorch*. Consulté le 21 avril 2023. https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.StepLR.html
- [916] Mardan A. *Full Stack JavaScript*. Berkeley, CA: Apress ; 2015.
- [917] Nokeri TC. Python Web Frameworks and Apps. *Web App Development and Real-Time Web Analytics with Python: Develop and Integrate Machine Learning Algorithms into Web Apps* (pp. 79-85). Berkeley, CA: Apress ; 2021.
- [918] dcc.Upload. *plotly*. Consulté le 21 avril 2023. <https://dash.plotly.com/dash-core-components/upload>
- [919] Yablonski J. *Laws of UX: Using psychology to design better products & services*. O'Reilly Media ; 2020.
- [920] Brown T. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. HarperBusiness ; 2009.
- [921] Tabs. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/tabs/>
- [922] Layout. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/layout/>
- [923] Dash HTML Components. *plotly*. Consulté le 21 avril 2023. <https://dash.plotly.com/dash-html-components>
- [924] Buttons. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/button/>
- [925] DropdownMenu. *Dash Bootstrap Components*. Consulté le 21 avril 2023. https://dash-bootstrap-components.opensource.faculty.ai/docs/components/dropdown_menu/
- [926] Alerts. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/alert/>
- [927] Cards. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/card/>
- [928] Spinner. *Dash Bootstrap Components*. Consulté le 21 avril 2023. <https://dash-bootstrap-components.opensource.faculty.ai/docs/components/spinner/>
- [929] Basic Dash Callbacks. *plotly*. Consulté le 21 avril 2023. <https://dash.plotly.com/basic-callbacks>
- [930] Lai ST. An interface design secure measurement model for improving web App security. *International Conference on Broadband and Wireless Computing, Communication and Applications* (pp. 422-427). IEEE ; 2011.
- [931] Tataryn D, Burachonak I. Hypertext Transfer Protocol Secure, Security Web Services. 2019.
- [932] Albahar M, Alansari D, Jurcut A. An Empirical Comparison of Pen-Testing Tools for Detecting Web App Vulnerabilities. *Electronics*. 2022;11(19):2991.
- [933] Williams G. Testing and Deploying Your Mobile Web App. *Learn HTML5 and JavaScript for Android* (pp. 317-350). Berkeley, CA: Apress ; 2012.
- [934] Mansfield-Devine S. DDoS: threats and mitigation. *Network Security*. 2011;(12):5-12.

- [935] Marshall DA, Burgos-Liz L, IJzerman MJ, et al. Selecting a dynamic simulation modeling method for health care delivery research-part 2: report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force. *Value Health*. 2015;18(2):147-160.
- [936] Majd M, Najafi P, Alhosseini SA, Cheng F, Meinel C. A Comprehensive Review of Anomaly Detection in Web Logs. *2022 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)* (pp. 158-165). IEEE ; 2022.
- [937] Shema M. *Hacking web apps: detecting and preventing web application security problems*. Newnes ; 2012.
- [938] Yao R, Lin G, Xia S, Zhao J, Zhou Y. Video object segmentation and tracking: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2020;11(4):1-47.
- [939] He H, Garcia EA. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*. 2009;21(9):1263-1284.
- [940] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the AAAI conference on artificial intelligence*. 2017;31(1).
- [941] Koonce B. EfficientNet. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization* (pp. 109-123). Berkeley, CA: Apress ; 2021.
- [942] Yang J, Zhou K, Li Y, Liu Z. Generalized out-of-distribution detection: A survey. *ArXiv, abs/2110.11334*. 2021.
- [943] Traitement de données personnelles. *CNIL*. Consulté le 21 avril 2023. <https://www.cnil.fr/fr/definition/traitement-de-donnees-personnelles>
- [944] Terms of Use Agreement and Privacy Policy. *International Skin Imaging Collaboration ("ISIC")*. Consulté le 21 avril 2023. <https://www.isic-archive.com/#!/topWithHeader/tightDarkContentTop/termsOfUse>
- [945] About CC Licenses. *Creative Commons*. Consulté le 21 avril 2023. <https://creativecommons.org/about/ccllicenses/>
- [946] SNITEM. *Mon logiciel est-il un dispositif médical ?*. 2021. <https://www.snitem.fr/publications/fiches-et-syntheses/les-4-etapes-du-marquage-ce-medical/>
- [947] Asada K, Komatsu M, Shimoyama R, et al. Application of Artificial Intelligence in COVID-19 Diagnosis and Therapeutics. *J Pers Med*. 2021;11(9):886.
- [948] Bentayeb M, Benda L, Vlaar T, Combes S. Présentation et avancées du Health Data Hub. *Revue d'Épidémiologie et de Santé Publique*. 2022;70:S7-S8.
- [949] World Health Organization. *Ethics and governance of artificial intelligence for health*. 2021. <https://www.who.int/publications/i/item/9789240029200>
- [950] Zanca F, Brusasco C, Pesapane F, Kwade Z, Beckers R, Avanzo M. Regulatory Aspects of the Use of Artificial Intelligence Medical Software. *Semin Radiat Oncol*. 2022;32(4):432-441.
- [951] Müller H, Holzinger A, Plass M, Brcic L, Stumptner C, Zatloukal K. Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation. *N Biotechnol*. 2022;70:67-72.
- [952] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS. *EUR-Lex*. Consulté le 21 avril 2023. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

- [953] Création d'un service de l'intelligence artificielle à la CNIL et lancement des travaux sur les bases de données d'apprentissage. *CNIL*. Consulté le 21 avril 2023. <https://www.cnil.fr/fr/creation-dun-service-de-lintelligence-artificielle-la-cnil-et-lancement-des-travaux-sur-les-bases-de>
- [954] Justo-Hanani R. The politics of Artificial Intelligence regulation and governance reform in the European Union. *Policy Sciences*. 2022;55(1):137-159.
- [955] Esteva A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med*. 2021;4(1):5.
- [956] Biot J. How will clinical practice be impacted by artificial intelligence?. *Eur J Dermatol*. 2019;29(S1):8-10.
- [957] Grosjean S. L'interopérabilité sociale de l'IA en santé: un enjeu pour le design d'algorithmes situés dans des pratiques. *Revue française des sciences de l'information et de la communication*. 2019;17.
- [958] Badillo S, Banfai B, Birzele F, et al. An Introduction to Machine Learning. *Clin Pharmacol Ther*. 2020;107(4):871-885.



VU, LE PRESIDENT DU JURY

CAEN, LE

VU, LE DIRECTEUR DE LA FACULTE
DES SCIENCES PHARMACEUTIQUES

CAEN, LE

L'université n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses et mémoires. Ces opinions doivent être considérées comme propres à leurs auteurs.

TITRE

RÔLE, ÉVALUATION ET RÉGLEMENTATION DES INTELLIGENCES ARTIFICIELLES APPLIQUÉES AU DIAGNOSTIC ET À LA THÉRAPEUTIQUE

Résumé

Les progrès spectaculaires réalisés récemment par l'intelligence artificielle lui valent souvent d'être considérée comme étant précurseur d'un changement de paradigme pour le champ médical. Ses applications s'étendent de plus en plus et offrent des perspectives nouvelles : diagnostic augmenté, médecine de précision, médecine prédictive, chirurgie assistée, rééducation assistée, suivi thérapeutique automatisé, recherche thérapeutique accélérée, etc. Ces différents usages représentent une opportunité majeure d'améliorer l'efficacité des soins de santé et de la recherche biomédicale mais sont également source de préoccupations en matière d'éthique, de sécurité, de confidentialité et de responsabilité. Au cœur de ces dernières réside l'enjeu de la réglementation, dont la réflexion nécessite un dialogue interdisciplinaire faisant intervenir non seulement des professionnels de santé et de la donnée, mais aussi des juristes et philosophes spécialistes de l'éthique. En naviguant à travers ces thématiques, cet ouvrage dresse un état de l'art de l'intelligence artificielle appliquée au diagnostic et à la thérapeutique, et examine la transposabilité de la réglementation et des méthodes d'évaluation appliquées aux produits de santé. Secondairement, celui-ci propose aux lecteurs une synthèse historique et technique de l'intelligence artificielle, et soulève notamment l'importance de former les professionnels de santé à ces technologies. Ceux-ci sont résolument amenés à devoir jouer un rôle actif dans la convergence de la santé et de l'intelligence artificielle, en étant les garants d'un usage éthique servant au mieux l'intérêt des patients. Enfin, cet ouvrage illustre cette approche interdisciplinaire par la présentation d'une application (détection de mélanome à partir d'une photographie) développée par l'auteur au cours de ses recherches.

TITLE

ROLE, EVALUATION AND REGULATION OF ARTIFICIAL INTELLIGENCE APPLIED TO DIAGNOSIS AND THERAPEUTICS

Summary

The spectacular progress recently achieved by artificial intelligence has often led to it being considered as a precursor of a paradigm shift for the medical field. Its applications are expanding more and more, offering new perspectives: augmented diagnosis, precision medicine, predictive medicine, assisted surgery, assisted rehabilitation, automated therapeutic monitoring, accelerated therapeutic research, etc. These different uses represent a major opportunity to improve the efficiency of health care and biomedical research, but also raise concerns regarding ethics, security, confidentiality and liability. At the heart of these concerns lies the issue of regulation, which requires an interdisciplinary dialogue involving not only healthcare and data professionals, but also legal and philosophical experts in ethics. By navigating through these themes, this thesis provides a state-of-the-art overview of artificial intelligence applied to diagnosis and therapeutics, and examines the transposability of regulation and evaluation methods applied to health products. Secondly, it offers readers a historical and technical synthesis of artificial intelligence, and highlights the importance of training healthcare professionals in these technologies. They are definitely called upon to play an active role in the convergence of health and artificial intelligence, by being the guarantors of an ethical use serving the best interests of patients. Finally, this thesis illustrates this interdisciplinary approach by presenting an application (melanoma detection from a photograph) developed by the author during his research.

Mots-clés

Intelligence artificielle, Apprentissage automatique, Diagnostic, Thérapeutique, Réglementation, Évaluation
