



HAL
open science

L'intelligence artificielle au service des pharmaciens d'officine et biologistes

Assan Sanogo

► **To cite this version:**

Assan Sanogo. L'intelligence artificielle au service des pharmaciens d'officine et biologistes. Sciences du Vivant [q-bio]. 2019. dumas-02477480

HAL Id: dumas-02477480

<https://dumas.ccsd.cnrs.fr/dumas-02477480>

Submitted on 13 Feb 2020

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.



UFR DE PHARMACIE
Université Clermont Auvergne

UNIVERSITÉ CLERMONT AUVERGNE
UFR DE PHARMACIE

Année : 2019

N°

THÈSE D'EXERCICE
pour le
DIPLOME D'ÉTAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement
le 28 octobre 2019
par
Assan SANOGO

L'INTELLIGENCE ARTIFICIELLE AU SERVICE DES PHARMACIENS D'OFFICINE ET BIOLOGISTES

Directeur de thèse : **Mr Jean-Michel CARDOT**

Jury

Président : **Mr Jean-Michel CARDOT**

Professeur,
UFR Pharmacie de Clermont-Ferrand

Membres : **Mme Valérie Weber**

Maître de conférences,
UFR Pharmacie de Clermont-Ferrand

Mme Caroline Decombat

Maître de conférences,
UFR Pharmacie de Clermont-Ferrand

Mr Amir Nakib

Directeur de recherche en
intelligence artificielle,
Cyclope.ai (Groupe Vinci)



UFR DE PHARMACIE
Université Clermont Auvergne

UNIVERSITÉ CLERMONT AUVERGNE
UFR DE PHARMACIE

Année : 2019

N°

THÈSE D'EXERCICE
pour le
DIPLOME D'ÉTAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement
le 28 octobre 2019
par
Assan SANOGO

L'INTELLIGENCE ARTIFICIELLE AU SERVICE DES PHARMACIENS D'OFFICINE ET BIOLOGISTES

Directeur de thèse : **Mr Jean-Michel CARDOT**

Jury

Président : **Mr Jean-Michel CARDOT**

Professeur,
UFR Pharmacie de Clermont-Ferrand

Membres : **Mme Valérie Weber**

Maître de conférences,
UFR Pharmacie de Clermont-Ferrand

Mme Caroline Decombat

Maître de conférences,
UFR Pharmacie de Clermont-Ferrand

Mr Amir Nakib

Directeur de recherche en
Intelligence Artificielle,
Cyclope.ai (Groupe Vinci)

REMERCIEMENTS

Je tiens à remercier mes parents pour avoir cru en ce changement de carrière tardif et ainsi, devenir un professionnel de santé. Je remercie ma petite soeur, qui fut sans doute le catalyseur de ce choix difficile et ma soeur jumelle qui reste un exemple de persévérance. Je tiens aussi à remercier le Docteur Amir Nakib, Directeur de la recherche en intelligence artificielle chez Cyclope.ai et Maître de conférences à l'Université Paris Est Créteil (UPEC) qui a cru en mon potentiel. Il m'a en effet permis de développer au sein de son équipe passionnée, savoir théorique et pratique utiles à la rédaction de cette présente thèse et également nécessaires à l'essor de l'intelligence artificielle au service des sciences pharmaceutiques.

J'exprime ma profonde gratitude envers le président et les membres du jury qui ont su nourrir très tôt cette passion pour la recherche appliquée aux sciences pharmaceutiques.

Je remercie enfin, ma femme et ma fille pour contribuer chaque jour à faire de ma vision du pharmacien une réalité.

TABLE DES MATIERES

TABLE DES MATIERES	3
1. LISTE DES TABLEAUX ET FIGURES	5
2. ABBREVIATIONS	7
3. INTRODUCTION	10
4. GENERALITES	11
4.1. L'intelligence artificielle	11
4.2. L'apprentissage automatique	11
4.3. Les ConvNets ou réseaux de convolutions	12
4.4. L'inspiration biologique:	13
4.5. Les opérations constitutives des réseaux convolutifs	15
4.5.1. L'opération de convolution	15
4.5.2. L'opération d'échantillonnage ou "Pooling"	16
4.5.3. L'opération d'abandon ou "Dropout"	17
4.5.4. Les Fonctions d'activation classiques	18
4.6. L'apprentissage des réseaux convolutifs	19
4.6.1. La descente de gradient	21
4.6.2. La rétro-propagation du gradient	23
4.6.3. L'évanouissement du gradient	24
4.6.4. Le sous-apprentissage et le sur-apprentissage	26
5. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN D'OFFICINE	27
5.1. Généralités	27
5.2. La solution retenue	29
5.3. Le détecteur d'objets	32
5.3.1. Détecteur d'objets "historique"	32
5.3.2. Les détecteurs récents	34
5.4. Méthodologie	37
5.4.1. Entraînement du module de détection	37
5.4.2. Le thésaurus:	41
5.5. Résultats et Limites	43
6. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN CLINICIEN	46
6.1. Jeu de données	46
6.2. Analyse exploratoire des données	48

6.3. Méthodologie	52
6.3.1. Généralités	52
6.3.2. XGBoost	55
6.3.3. Boosting	55
6.3.3.1. Boosting par sous-ensembles	57
6.3.3.2. Boosting probabiliste (bayésien)	59
6.3.4. Le boosting appliqué au gradient	60
6.3.5. Classes non équilibrées et stratification	62
6.3.5.1. Stratification	62
6.3.5.2. Sur-échantillonnage	62
6.4. Paramètres et optimisation	65
6.5. Résultats et limites	66
7. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN BIOLOGISTE	70
7.1. Jeu de données	70
7.1.1. Description	70
7.1.2. Analyse exploratoire des images	73
7.2. Algorithme	74
7.2.1. Classification multi-classes	74
7.3. Résultat et limites	76
7.4. Conclusion	79
8. L'IA POUR LE PHARMACIEN EN PAYS EMERGENTS	80
8.1. Jeu de données	80
8.1.1. Description	80
8.2. Méthodologie	81
8.2.1. Réseau neuronal convolutif très profond	81
8.3. Résultats et limites	83
8.4. Conclusion	85
9. CONCLUSION	87
10. BIBLIOGRAPHIE	90
RESUME	93

1. LISTE DES TABLEAUX ET FIGURES

Figure 1 : Comparaison du neurone biologique avec le neurone artificiel.

Figure 2 : L'opération de convolution pour les CNN.

Figure 3 : L'opération de Max-Pooling est une opération de sous-échantillonnage.

Figure 4 : Le "Drop-out" renforce le caractère prédictif des modèles.

Figure 5 : Les fonctions d'activations classiques.

Figure 6 : Entraînement, validation et test.

Figure 7 : La validation croisée.

Figure 8 : La descente de gradient.

Figure 9 : L'impact du taux d'apprentissage sur la descente de gradient.

Figure 10 : La rétro-propagation du gradient appliquée à un neurone.

Figure 11 : La rétro-propagation du gradient appliquée aux réseaux profonds.

Figure 12 : Rectified linear Unit, Rectified linear Unit.

Figure 13 : Le sur-apprentissage.

Figure 14 : Utilisation de Chatbot dans le coaching du patient.

Figure 15 : Schéma illustratif du prototype d'interaction médicamenteuse.

Figure 16 : Cascade de Haar: filtres de détection en action.

Figure 17 : Détection d'objets par fenêtrage glissant.

Figure 18 : Les particularités d'implémentation de Y.O.L.O.

Figure 19 : Le fenêtrage glissant implémenté par un réseau 100% convolutif.

Figure 20 : Détection d'objets en temps réel.

Figure 21 : Les itérations garantissent la robustesse des algorithmes.

Figure 22 : Détection d'objets appliquée à la santé.

Figure 23 : Etiquetage des données est une étape nécessaire à l'entraînement de modèles.

Figure 24 : Utilisation détournée de Captcha pour annoter des données.

Figure 25 : Thesaurus des interactions médicamenteuses.

Figure 26 : Exemples d'objets détectés après entraînement.

Figure 27 : Technique d'attaque des réseaux de Neurones par "patch" ou étiquette.

Figure 28 : Description du jeu de données "Healthcare Dataset Stroke Data".

Figure 29 : Corrélation en images.

Figure 30 : Corrélation des variables explicatives entre elles, et avec la variable explicative.

Figure 31 : Le machine learning en mode automatique.

Figure 32 : L'intelligence artificielle.

Figure 33 : Des arbres prédictifs aux arbres décisionnels thérapeutiques.

Figure 34 : Boosting & Bagging.

Figure 35 : Boosting et pondération des observations.

Figure 36 : Les remèdes aux jeux de données non équilibrés.

Figure 37 : Technique de sur-échantillonnage synthétique.

Figure 38 : Sensibilité-Spécificité.

Figure 39 : La fonction d'efficacité du récepteur.

Figure 40 : Exemple de courbe de survie.

Figure 41 : Illustration des structures biologiques identifiées à travers différents filtres.

Figure 42 : Déséquilibre de classes.

Figure 43 : Structure schématique du réseau de Neurones.

Figure 44 : Courbes de Précision.

Figure 45 : Plusieurs types de Convolutions.

Figure 46 : Nvidia DGX station: la référence pour la recherche en IA.

Figure 47 : Coloration Giemsa d'hématies de patients infectés par falciparum.

Figure 48 : Bloc caractéristique de ResNet.

Figure 49 : Courbes représentatives (fonction objectif et précision).

Figure 50 : Utilisation d'une solution de Deep Learning en Afrique.

2. ABBREVIATIONS

ADABOOST : Algorithme de Boosting adaptatif (Adaptive boosting in English)

ADASYN : Echantillonnage Synthétique Adaptatif (Adaptive synthetic sampling in English)

AI : Intelligence Artificielle (Artificial Intelligence in English)

ANSM : Agence Nationale de Sécurité du Médicament

ASAQ : Artésunate + Amiodaquine

ASMQ : Artésunate + Méfloquine

AUC : Aire sous la courbe (area under curve in English)

CNIL : Commission Nationale de l'Informatique et des Libertés

CNN : Réseau Neuronal Convolutif (convolutional neural network in English)

COCO : Objets Communs [dans leur] Contexte (Common Objects in context in English)

CONVNET : Réseau neuronal convolutif (Convolutional Neural Network in English)

CSV : Format Valeurs séparées par des virgules (Comma Separated Values in English)

BN : Normalisation de Batch (batch normalization in English)

DNDI : Initiative pour les médicaments pour maladies négligées (Drugs for Neglected Disease Initiative in English).

E.S.I.E.E : Ecole de l'innovation technologique de la Chambre de commerce et d'industrie.

FN: Faux Négatifs (observations de la classe “négative” inexactement prédites)

FP: Faux Positifs (observations de la classe “positive” inexactement prédites)

IA : Intelligence Artificielle.

GAN: modèle génératif antagoniste (Generative adversarial network)

GPU : carte graphique de calcul (Graphical Processing Unit in English).

JSON : format de fichier: javascript object notation.

MAXPOOLING : Echantillonnage Maximum (Pooling Maximum in English).

NIH : National Institute of Health.

PVD: Pays en voie de développement.

RCNN: Réseau de convolution récurrent (Recurrent Convolutional Network in English).

RELU: Unité de rectification linéaire (Rectified Linear Unit in English).

RGPD: Règlement Général sur la Protection des Données.

ROC : Courbe fonction d'efficacité du récepteur (Receiver Operating Characteristic Unit in English).

SMOTE : Sur-échantillonnage Synthétique de la Minorité (Synthetic Minority Over sampling Technique in English).

SMOTEENN : Sur-échantillonnage Synthétique de la Minorité avec Edition des plus proches voisins (Synthetic Minority Oversampling Technique and Edited Nearest Neighbour in English)

VGG : Visual Geometry Group

VN: Vrais Négatifs (observations de la classe "négative" exactement prédites)

VP: Vrais Positifs (observations de la classe "positive" exactement prédites)

XGBOOST : Algorithme extreme gradient boosting

YOLO : Algorithme de "Deep Learning" named "you look only once"

3. INTRODUCTION

Ce travail a pour but d'illustrer comment le pharmacien peut dépasser son utilisation actuelle des outils digitaux classiques et participer plus activement à cette révolution technologique que représente l'essor de l'intelligence artificielle. En effet, ces nouveaux outils vont permettre de personnaliser le suivi du patient, d'en assurer la prévention et plus généralement de repousser les limites des services de santé. Dans une première partie, nous allons introduire les généralités relatives à l'intelligence artificielle. Dans les parties suivantes, nous considérerons des cas d'usage variés pertinents pour le pharmacien.

D'abord, nous nous intéresserons à mini-prototype destiné au pharmacien d'officine. Cet algorithme permet de détecter les types de médicaments présents sur une photo et de statuer quant à une potentielle interaction médicamenteuse. Ensuite, nous nous consacrerons à une application destinée au pharmacien biologiste basée sur un problème de classification de coupes biologiques issues du National institute of Health (N.I.H). Dans une troisième partie, nous décrirons brièvement un modèle prédictif utile au pharmacien clinicien et terminerons cette étude sur l'impact que cette nouvelle technologie peut avoir pour le pharmacien en pays en voie de développement à travers un détecteur de coupes biologiques comprenant des hématies infectées à Plasmodium.

Enfin, la conclusion de ce travail élargit les perspectives de cette étude et questionne notamment les limites des algorithmes actuels extrêmement gourmands en données et remet en question les impacts sociétaux et éthiques d'une telle technologie.

4. GENERALITES

4.1. L'intelligence artificielle

L'intelligence artificielle, encore appelée IA, est la convergence de plusieurs techniques simulant les processus cognitifs humains. Elle existe depuis les années 60 et la recherche s'est accélérée cette dernière décennie et elle se retrouve aujourd'hui dans une multitude d'applications: des voitures autonomes jusqu'au diagnostics médicaux, en passant par les robots industriels. Après 2010, l'essor des cartes dédiées au calcul parallèle a contribué à l'explosion de la puissance de calcul des machines. L'abondance croissante des données, d'autre part, a signé le tournant majeur de l'IA, lui permettant de s'imposer jusque dans les objets du quotidien.

Ses finalités ont suscité depuis toujours de nombreuses inquiétudes ou interprétations tant dans la littérature que dans les médias. Dans une tentative de contrôle qui renforce l'action de la CNIL, il existe depuis 2016, un cadre légal européen grandissant autour de la protection de la donnée défini par le Règlement Général sur la Protection des Données (R.G.P.D) [1]

4.2. L'apprentissage automatique

On distingue deux types d'apprentissage automatique: *supervisé* et *non-supervisé*. L'apprentissage supervisé (supervised learning en anglais) est une tâche d'apprentissage automatique dont le but est l'apprentissage d'une fonction de prédiction à partir d'observations étiquetées.

On catégorise ces problèmes sous les termes de: régression ou de classification. Dans le cas de la prédiction d'une ou plusieurs variable(s) quantitative(s), on parle de régression.

Pour les problèmes liés à une ou plusieurs variable(s) qualitative(s), on parle de classification.

L'apprentissage non supervisé consiste à apprendre sans étiquettes. Le modèle doit extraire des classes ou groupes d'individus présentant des caractéristiques communes (on utilise le terme *clustering* en anglais). La performance d'une méthode de classification est mesurée par sa capacité à trouver des structures sous-jacentes dans les données, alors que les métriques pour évaluer les algorithmes non supervisés sont non triviaux.

4.3. Les ConvNets ou réseaux de convolutions

Ces dernières années, dans le cadre de problèmes liés à l'analyse d'images, les techniques d'apprentissage profond ou "deep learning" se sont imposées comme les solutions de choix. Dans le cas présent nous utiliserons des réseaux de neurones convolutifs profonds. On les retrouve sous le nom de "ConvNet" ou CNN. Ces réseaux sont constitués de plusieurs éléments:

- **Les poids ou valeurs** sont les éléments constitutifs des matrices qui sont modifiés de manière itérative lors de l'entraînement.
- **L'agencement de couches** qui rendent certains réseaux caractéristiques et adaptés à des tâches particulières.
- Les **types de neurones** organisés sous forme matricielle caractérisés par leur fonction de transfert.
- La **Fonction Objectif** qui permet la comparaison des étiquettes avec les prédictions et par suite l'apprentissage itératif des modèles

4.4. L'inspiration biologique:

Un neurone biologique est une cellule nerveuse constituant la base du système nerveux spécialisée dans le traitement des signaux électriques. Le neurone intègre l'information et la propage du corps cellulaire ou soma et emprunte l'axone et est connecté aux autres neurones via les dendrites. Un neurone artificiel est un objet qui va incorporer l'information entrante avant de la transformer et la propager dans un sens (passe avant ou "forward pass" en anglais) à un ou plusieurs neurones. La sommation de l'information entrante tient compte des intensités des connections (pareilles aux synapses biologiques) qui sont définies par les poids. On notera qu'il existe d'autres implémentations des réseaux de neurones (qui incluent les neurones impulsionnels) notamment qui sont plus fidèles aux théories issues de la neuroscience. DeepMind parie sur la convergence de ces 2 disciplines (Machine Learning & Neurosciences) pour permettre le développement de nouvelles architectures.[2]

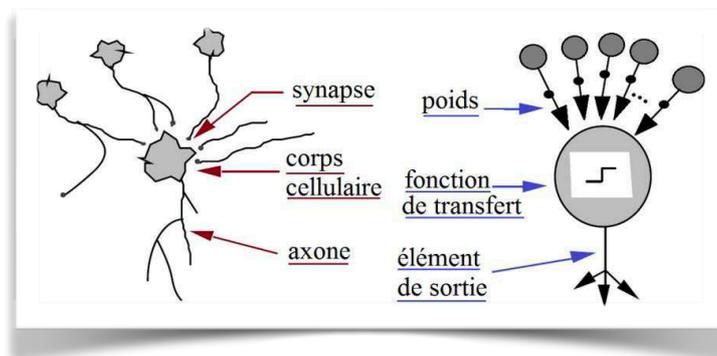


Figure 1: **Comparaison du neurone biologique avec le neurone artificiel.**

Les réseaux de neurones actuels reposent sur l'entité élémentaire: le neurone. Ce dernier a une structure qui s'inspire des neurones biologiques.

Dans les réseaux de neurones convolutifs, chaque couche représente une interprétation de l'information entrante, qui gagne en complexité au fur et à mesure que l'on s'enfonce dans le réseau. La couche finale est une couche de décision. Par exemple dans le cas d'une classification, le réseau produit des probabilités pour chaque classe.

4.5. Les opérations constitutives des réseaux convolutifs

4.5.1. L'opération de convolution

Dans le cadre de l'apprentissage profond, le produit de convolution est le traitement d'une matrice par un filtre appelé encore : noyau ou kernel. Dans l'exemple ci-dessous, il s'agit d'une matrice ou filtre K de taille 3x3. La convolution permet de détecter et extraire des caractères ou propriétés intrinsèques de l'image. L'exemple ci-dessous illustre le calcul en pratique et montre comment l'image initiale est filtrée selon une caractéristique du kernel employé.

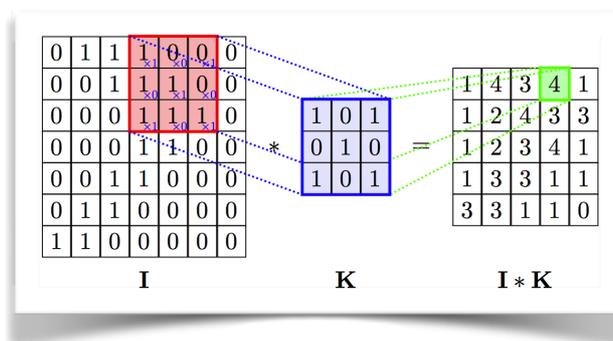


Figure 2: L'opération de convolution pour les CNN

L'opération de convolution peut conduire à une réduction de la taille de l'image. Il est possible de paramétrer cette grandeur afin de conserver la taille de l'image (ajout d'une marge). Ci-dessus, à droite, le produit de convolution réduit la taille de la matrice qui a subi l'opération de convolution.

4.5.2. L'opération d'échantillonnage ou "Pooling"

Elle permet de contrôler la taille (largeur et hauteur) du "volume occupé" par les neurones, tout en prenant en compte la réponse des neurones les plus activés. Ceci permet de conserver l'information tout en réduisant le nombre d'opérations.

Le champ récepteur d'un neurone sensoriel ou d'un neurone sensitif est le volume de l'espace qui modifie la réponse de ce neurone. Dans notre cas, le champ récepteur représente la surface de l'image entrée vue par un neurone des couches de convolution. L'opération d'échantillonnage permet en réalité d'augmenter le champ récepteur des neurones.

Comme décrit ci-dessous, on distingue plusieurs types: "max pooling", "average pooling", "global pooling", etc. L'opération "max-pooling" est sensible aux variations au sein d'une image (gradients d'intensité lumineuse) ce qui permet d'extraire les caractéristiques de l'image comme les bords et les textures. En revanche, le type average pooling va effectuer une sorte de lissage de l'image. Dans le cas de données très bruitées, l'échantillonnage moyen ("average pooling" en anglais) permet d'extraire l'information par rapport au bruit.

Enfin, les deux approches diffèrent par la quantité d'information rejetée. Dans le cas d'une opération d'échantillonnage maximum ("max pooling" en anglais): $\text{MaxPool}(2,2)$, uniquement 1/4 des données entrantes sont conservées. A l'inverse, dans le cas d'un échantillonnage moyen, toutes les données sont utilisées.

Un choix de kernel trop grand peut entraîner une trop forte diminution du signal (perte d'information) et impacter négativement l'apprentissage du réseau de neurones.

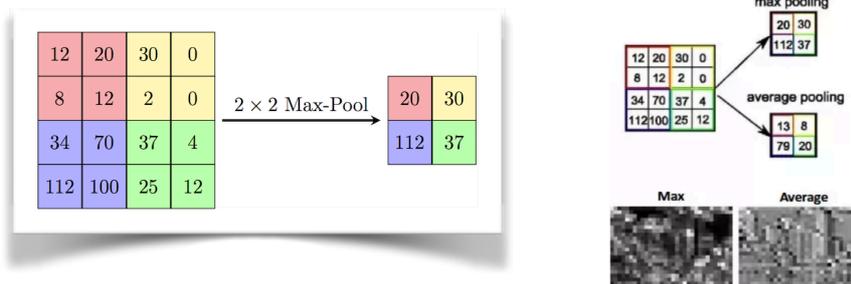


Figure 3: L'opération de "Max Pooling" est une opération de sous-échantillonnage.

Dans le cas de l'opération de sous échantillonnage maximum (communément appelée max pooling) - $MaxPool(2,2)$ uniquement 1/4 des données entrantes sont conservées. Cette opération est nécessaire pour contrôler la taille du réseau et est une façon d'augmenter le champ récepteur au fur et à mesure que l'on s'enfonce dans le réseau convolutif.

4.5.3. L'opération d'abandon ou "Dropout"

La couche d'abandon ou "Dropout" a pour rôle d'éteindre au hasard des neurones avec une probabilité p , uniquement lors de l'entraînement. Autrement dit, seule une partie du réseau va être entraînée à chaque itération. La capacité du réseau de neurones résultant est donc diminuée. Cette opération augmente le temps de convergence, mais permet d'obtenir une meilleure performance. Le choix de la probabilité " p " reste empirique et définie au début de chaque entraînement. Elle est souvent perçue comme un hyperparamètre.

La justification de l'utilité de la couche "Dropout" est justifiée par la complexité des couches entièrement connectées. En effet, ces couches concentrent la majorité des paramètres. Lors de l'apprentissage les neurones développent une co-dépendance

nuisible au pouvoir prédictif individuel de chaque neurone. Le “Dropout” permet donc de réduire les problèmes de sur-apprentissage.

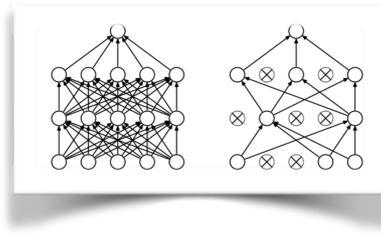


Figure 4: **Le Dropout renforce le caractère prédictif des modèles**

Dans le cas de l’opération de Dropout (à droite) des Neurones sont désactivés lors de l’apprentissage avec une probabilité “p” qui suit une loi de Bernoulli. Il agit comme un paramètre de régularisation qui réduit les phénomènes de sur-apprentissage.

4.5.4. Les Fonctions d’activation classiques

Il existe de nombreuses fonctions d’activation:

Function	Equation	Graphe
Fonction Echelon	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	
Fonction signe	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	
Fonction linéaire	$\phi(z) = z$	
Fonction linéaire	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	
Fonction sigmoïde	$\phi(z) = \frac{1}{1 + e^{-z}}$	
Fonction tangente Hyperbolique	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	

Figure 5: **Les fonctions d’activations classiques.**

On retrouve ici de manière non exhaustive différentes fonctions d’activation traditionnellement utilisées dans la littérature. La fonction sigmoïde a tendance à accentuer le phénomène d’évanouissement du gradient.

4.6. L'apprentissage des réseaux convolutifs

Cet entraînement requiert de diviser le jeu de données en 3 parties de taille différentes: les échantillons d'entraînement, de validation et de test. Lors de l'apprentissage, le modèle améliore son caractère prédictif de manière itérative sur les données d'entraînement et sa capacité à généraliser est évaluée avec les données de validation. Les données de test ne sont jamais utilisées lors de l'apprentissage, mais servent à caractériser la performance du modèle prédictif, à posteriori.

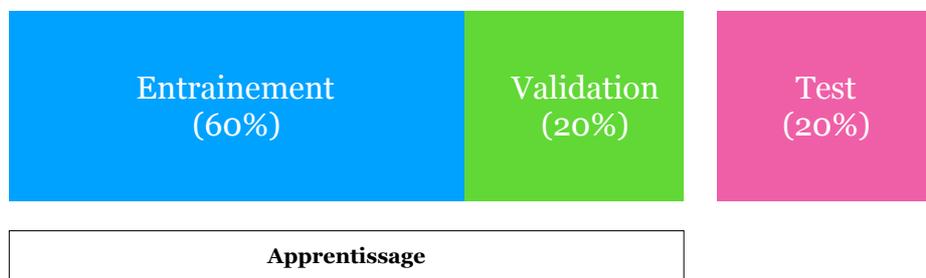


Figure 6: **Entraînement, validation et test.**

Le jeu de données est divisé en 3 parties (entraînement , validation et test).

À chaque passe sur l'entièreté du jeu de données d'entraînement (1 epoch), le modèle est évalué sur un jeu de données de validation. Finalement, à l'issue de l' apprentissage le modèle est testé sur le jeu de données de test.

Parfois, on divise le dataset en 2 parties: le jeu de données d'entraînement et de test. L'entraînement du modèle a lieu sur les données d'entraînement. Le jeu de test est là encore utilisé uniquement à la fin de l'entraînement. On divise ce dernier en k parties égales et disjointes. L'apprentissage est réalisé sur un échantillon formé de k-1 parties tirées au sort. Le taux d'erreur est calculé sur la partie restante (qui fait office de

validation). Cette opération est alors répétée k fois. L'erreur moyenne est appelée erreur de validation et on parle de validation croisée. La fonction d'apprentissage du réseau de convolutions repose sur deux éléments: la descente du gradient et la rétro-propagation de l'erreur au sein du réseau.

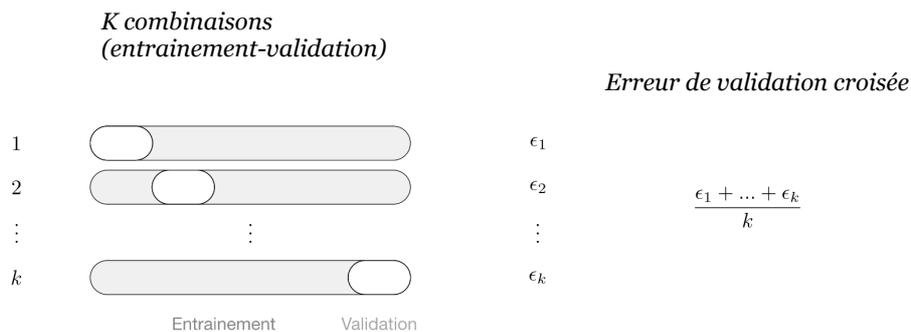


Figure 7: **La validation croisée.**

Le jeu de données est divisé en 2 parties (entraînement et test). Afin d'obtenir le modèle qui généralise le plus, on utilise la validation croisée. On divise le jeu de données d'entraînement en k parties. L'une de ces parties servira de validation. Il s'agit donc de k combinaisons (entraînement-validation). Chaque observation est utilisée k-1 fois lors de l'entraînement et sert exactement 1 fois lors de la validation. On obtient une erreur moyenne de validation.

4.6.1. La descente de gradient

La descente de gradient est un algorithme d'optimisation d'ordre 1 très utilisé pour les entraînements de réseaux de convolution. Le gradient de la fonction objectif en un point est égal au coefficient directeur de la tangente de sa courbe représentative en ce point. Le gradient pointe alors dans la direction de l'augmentation de la fonction objectif.

En d'autres termes : si le gradient est négatif (partie gauche de la courbe) pour se rapprocher de X_{\min} , il faut ajouter à X_1 une valeur proportionnelle opposée au gradient en ce point, pour se rapprocher de X_2 .

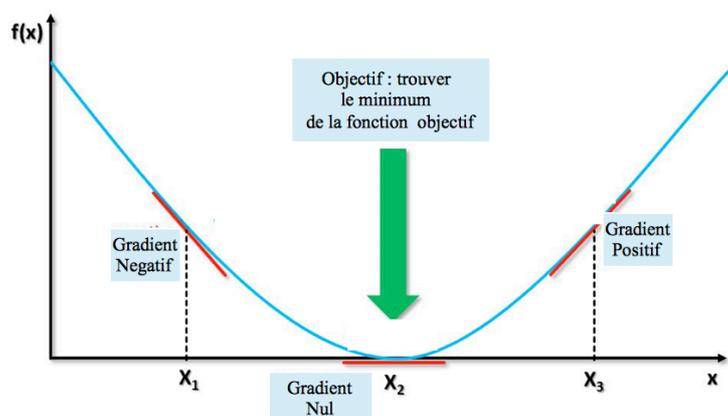


Figure 8: **La descente de gradient.**

On représente schématiquement en bleu la courbe représentative de la fonction objectif.

On retrouve en rouge, les tangentes à la fonction objectif en 3 points (X_1 , X_2 , X_3). En X_1 , la tangente pointe vers le bas (coefficient directeur négatif). En X_2 , la tangente pointe vers le haut (coefficient directeur positif). En X_3 le coefficient directeur est nul. L'objectif est de trouver le plus rapidement possible la valeur X_2 .

Si on se trouve dans la partie droite de la courbe (à droite de X_{\min}), le gradient est positif et il faudrait ajouter à X_3 une valeur proportionnelle et opposée au gradient en ce point pour se rapprocher de X_2 (le minimum)[3].

Devant ce constat, on peut dériver un algorithme d'optimisation qui retranche à l'abscisse du point courant, à chaque itération, une fraction du gradient de la fonction objectif (en ce point). Cette itération est continuée jusqu'à ce que la valeur de la fonction objectif ne change plus.

La fraction du gradient retranchée à chaque itération est communément appelée vitesse d'apprentissage. Comme illustré ci-dessous: si celui-ci est trop élevé, le minimum n'est jamais atteint. S'il est trop faible, le minimum est atteint en très grand nombre d'itérations, ce qui n'est pas désirable car cet algorithme est supposé accélérer la recherche du minimum.

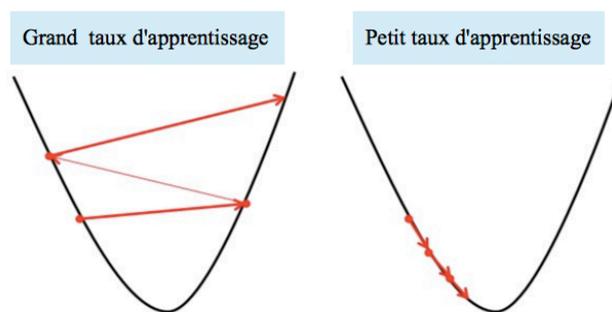


Figure 9: L'impact du taux d'apprentissage sur la descente de gradient.

On présente ici 2 scénarios de descente de Gradient. La courbe convexe représente l'évolution de la fonction objectif. À gauche, le taux d'apprentissage est trop faible, À droite, le taux d'apprentissage est trop grand: le modèle ne parvient pas à converger vers le point minimum.

Ainsi, notre but est de trouver les valeurs optimales des paramètres dont dépend le réseau. Ces termes aussi appelés poids minimisent la fonction objectif et sont obtenus de manière itérative via la formule :

$$\begin{array}{l} \text{Répéter jusqu' à convergence} \{ \\ \theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \\ \} \end{array}$$

4.6.2. La rétro-propagation du gradient

Dans le cadre d'apprentissage supervisé pour les CNN, on cherche à minimiser une fonction objectif. En minimisant cette fonction, les prédictions faites par le modèle deviennent de plus en plus proches des étiquettes pré-annotées. Cette minimisation de la fonction objectif a lieu lorsque le gradient de la fonction objet est nul. Or, nous avons présenté une méthode itérative qui permet de s'approcher du point minimum.[3]

Cette minimisation a lieu grâce à la rétro-propagation du gradient. En réalité, elle n'est que une interprétation de la dérivée en chaîne:

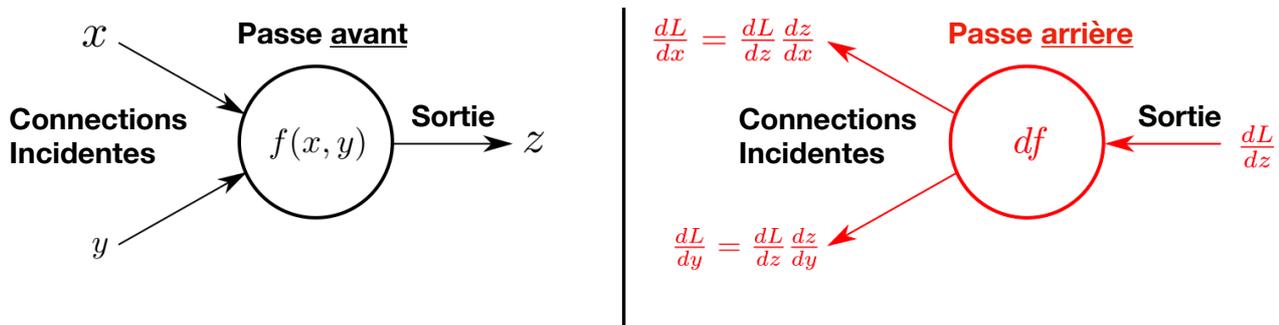


Figure 10: La rétro-propagation du gradient appliquée à un neurone.

L'intuition derrière la rétro-propagation est la suivante: le gradient du coût est noté: dL/dz .

Ce gradient est rétro-propagé de manière proportionnelle à la contribution de x dans

l'erreur z , et de manière proportionnelle à la contribution de y dans l'erreur z . C'est ce qui

constitue le phénomène d'apprentissage des réseaux de neurones actuels.

Cette astuce mathématique permet de calculer les gradients de la fonction objectif par rapport à chacun des paramètres en les décomposant sous forme de produits de gradients connus.

Les éléments qui sont modifiés dans ce phénomène d'apprentissage sont les poids décrits ci-dessus. Les neurones servent "d'activateurs". Dans un premier temps, les entrées sont propagées et comparées à la sortie ou "étiquette". Cette comparaison se fait grâce à la fonction objectif: c'est la passe avant. On mesure dans cette étape : l'erreur entre la prédiction et la valeur cible.

La notion d'apprentissage repose sur la modification des poids de manière proportionnelle à leur contribution dans l'erreur. Il s'agit de modifier les poids positivement ou négativement par rétro-propagation de l'erreur. Cette contribution est communément désignée par le terme "gradients".

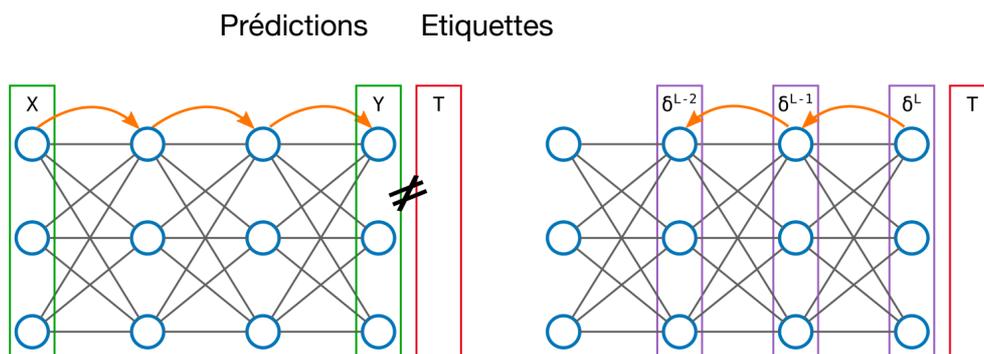


Figure 11: **La rétropropagation appliquée aux réseaux profonds.**

Après chaque itération, l'erreur est rétro-propagée et les poids modifiés par un terme proportionnel à leur contribution dans cette erreur ($-\alpha \cdot (dE/dW)$).

4.6.3. L'évanouissement du gradient

Le phénomène d'évanouissement du gradient est une conséquence de la rétro-propagation de l'erreur. À chaque couche, la rétro-propagation de ce gradient devient de plus en plus petite (redistribution des contributions des erreurs).

Dans la partie négative du graphe, la fonction tend asymptotiquement vers zéro et par conséquent le gradient est nul. Dans la partie positive du graphe, la fonction tend asymptotiquement vers 1 et le gradient s'annule également.

- Ainsi, lorsqu'un gradient faible est rétro-propagé via une fonction d'activation sigmoïde, il résulte une multiplication de termes <1 , qui tend naturellement vers zéro.
- Ainsi, lorsqu'un gradient élevé est rétro-propagé via une fonction d'activation sigmoïde, il résulte une multiplication de termes >1 , qui tend naturellement vers l'infini.

L'utilisation de la fonction de transfert linéaire rectifiée: ReLU (rectified linear unit) répond à ce problème d'évanouissement du gradient. Cette fonction de transfert non linéaire a cependant la limite de créer des neurones morts. En effet, lors de la rétro propagation du gradient via cette fonction de transfert, les signaux négatifs éteignent les neurones. Cet évanouissement des gradients peut être limité grâce à plusieurs évolutions de la fonction de transfert appelée Leaky ReLU.

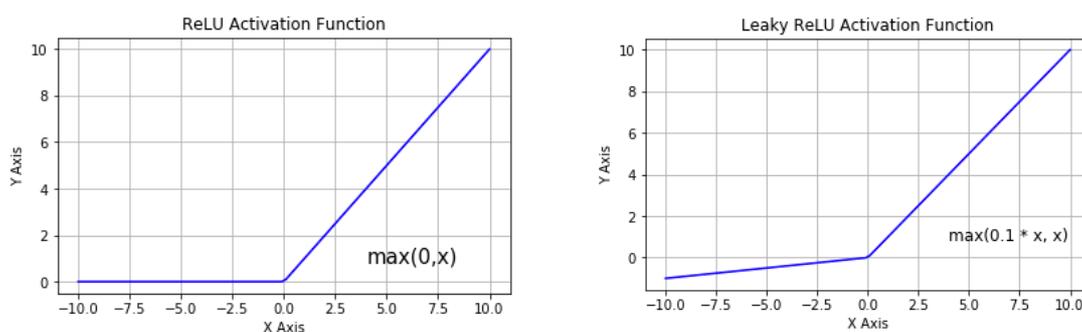


Figure 12: **Rectified linear Unit, Rectified linear Unit .**

Mathématiquement, la fonction de transfert ReLU est définie par $y = \max(0, x)$. La fonction de transfert appelée "leaky ReLU" tente de résoudre les limitations apportées par la fonction de transfert ReLU.

4.6.4. Le sous-apprentissage et le sur-apprentissage

En statistique, le sur-apprentissage est parfois appelé sur-ajustement ou également sur-interprétation (en anglais “overfitting”).

Le modèle prédictif, dans cet état, prédit exactement un ensemble particulier de données rencontrées lors de l’entraînement. Il peut cependant difficilement généraliser à des cas nouveaux (données de test, par exemples) et ne permet pas de prévoir de manière fiable les observations futures. Dans le cas des réseaux convolutifs, un modèle surentraîné est souvent un modèle statistique qui contient un trop grand nombre de paramètres. Il “apprend par coeur” les échantillons rencontrés lors de l'apprentissage et perd ses pouvoirs de prédiction sur de nouveaux échantillons. Un tel modèle est dépourvu de pouvoir prédictif et est donc inutile.

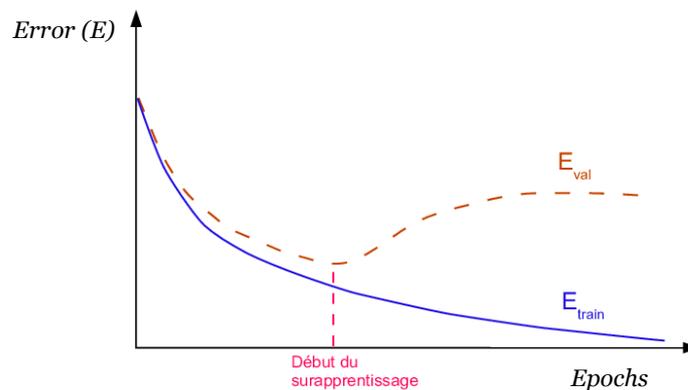


Figure 13: **Le sur-apprentissage.**

En marron: la courbe représente l'évolution de la fonction objectif de validation en fonction des itérations. En bleu: la courbe représente l'évolution de la fonction objectif d'entraînement en fonction des itérations. Lors du sur-apprentissage, le caractère prédictif modèle devient toujours meilleur pour les données d'entraînement, mais se dégrade de manière significative pour les données de validation.

5. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN D'OFFICINE

5.1. Généralités

Le pharmacien d'officine est l'acteur de santé le plus impliqué dans la santé de proximité. Son rôle de conseil doit être exploité et c'est le rôle des pharmaciens de développer des initiatives innovantes pour préserver l'existence du lieu d'échange, de prévention et de santé que représente la pharmacie d'officine.

L'utilisation de techniques innovantes peut par exemple donner naissance à des applications mobiles de prévention de santé recommandées par les pharmaciens d'officine ou à l'inverse être un moyen de formation et/ou de prise de décision.

La connaissance des données de santé du patient associée à des algorithmes prédictifs puissants peuvent être la source d'un système d'aide à la décision du pharmacien. Ceci lui permettra de continuer d'apporter ce degré de proximité, de personnalisation avec encore plus de pertinence sur le plan médical. Le développement de chatbots pour aiguiller les patients ou répondre aux questions fréquentes concernant certaines affections, illustrent cette tendance. Ces assistants conversationnels sont développés pour fournir un service et assurer une relation de proximité avec l'utilisateur. Ils s'utilisent principalement sur des plateformes de messagerie.

Ces applications sont souvent de 2 types. Le premier type peut être qualifié d'intelligent car il repose sur les techniques récentes d'intelligence artificielle. Les chatbots peuvent agir comme des assistants et leur version la plus élaborée se trouve à l'heure actuelle sans doute parmi l'offre grand public représentée par Siri® (Apple), Alexa® (Amazon) ou encore Cortana® (Microsoft).

La seconde forme d'assistants conversationnels est beaucoup plus formatée, elle repose sur des scénarios préétablis ou arbres de décision. Ce type de chatbot est très populaire

pour assister le département marketing ou pour le support après-vente. Ces assistants permettent de rediriger les questions les plus fréquentes vers les solutions appropriées. Dans le cas de scénarios complexes, leur logique est insuffisante pour créer une interaction multi-factorielle.

Si on s'intéresse aux fonctions que pourraient remplir les chatbots, on peut distinguer deux cibles: le pharmacien et le patient. Même si l'adoption est encore timide (dû en particulier à propriété des données et à méfiance quant à leur utilisation), les chatbots permettraient au pharmacien de satisfaire les activités suivantes:

- L'aide à la recherche d'information
- L'amélioration des relations clients avec les fournisseur
- L'assistance à la déclaration d'effets indésirables (Pharmacovigilance)
- L'évaluation de l'efficience de services.

Les assistants conversationnels pourraient aider les patients en pharmacie et hors pharmacie dans les cas ci-dessous :

- L'assistance conversationnelle à la recommandation de produit
- L'aide à l'orientation médicale
- L'accompagnement et le suivi des patients à distance
- Le coaching personnel de bien-être et santé

On citera Ramsay, le Chatbot prévention santé et Vik® l'assistant de WeFight qui accompagne les patients atteints d'un cancer du sein.

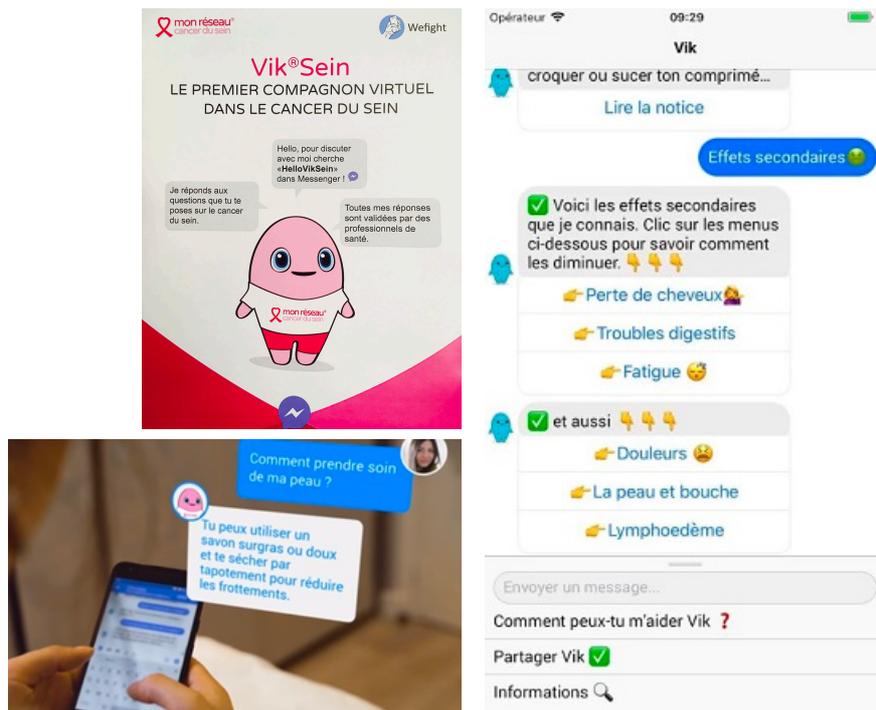


Figure 14: Utilisation de Chatbot dans le coaching du patient.

Vik@, l'assistant de WeFight, permet d'accompagner patients et patientes atteints d'un cancer du sein. Source: <https://www.patientsenreseau.fr/nos-reseaux/viksein/>

5.2. La solution retenue

En tant que pharmacien et data scientist, il m'apparaît nécessaire de faire converger technologie et santé. Dans le cadre de cette thèse d'exercice, nous décrivons une application qui est capable de reconnaître certains médicaments, présents sur une photo et de conclure après identification des médicaments et consultation d'une base de données d'interactions, à une éventuelle interaction médicamenteuse.

Pour que ce prototype soit utilisable sur des données réelles, il faudrait un jeu de photos bien plus divers provenant de plusieurs appareils photos différents, sous différentes conditions d'acquisition. Cette variété de la donnée garantira la robustesse de l'algorithme

final. Cependant, le but de cette étude est d'introduire les techniques à mettre en oeuvre afin de prototyper rapidement.

Le système intelligent sera capable de prévenir une interaction médicamenteuse, une interaction aliments-médicaments. Ce module pourrait en outre, s'intégrer à une structure d'assistance de conversation. Ce système pourrait être étendu à d'autres types d'interaction complément alimentaire – médicament. Il pourrait faire partie d'une solution plus large qui tiendrait compte des alertes sanitaires en cours lorsque le médicament est détecté. Par exemple en 2019, l'ANSM alerte sur les dangers du surdosage du paracétamol [5] en particulier les dégâts hépatiques, ou encore l'utilisation abusive de l'ibuprofène conduisant à des surinfections mortelle [6]

L'aspect général de la solution est donc représenté à travers le diagramme ci-dessous:

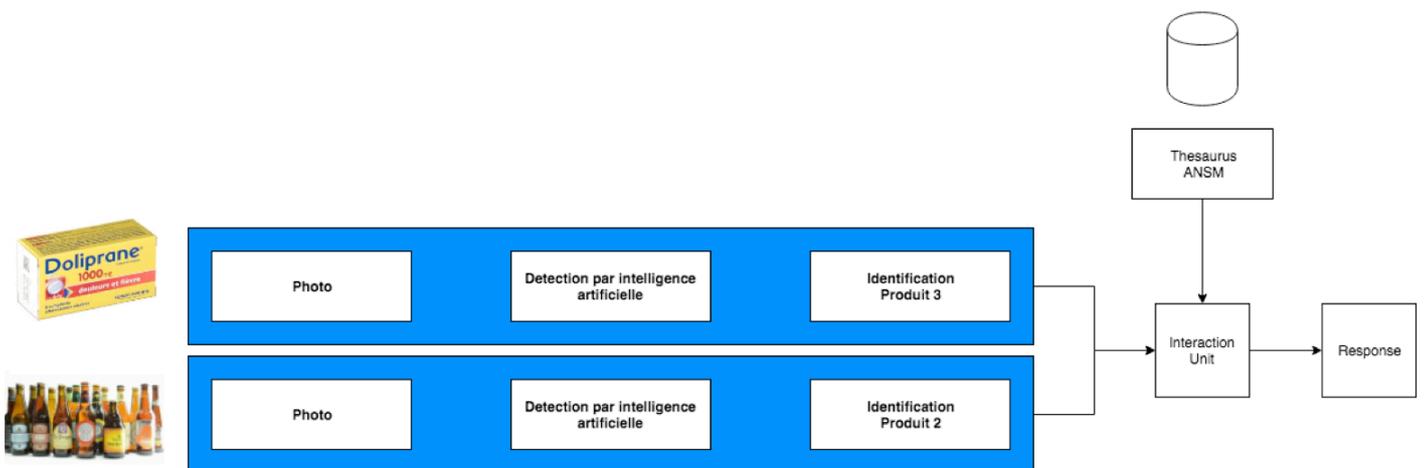


Figure 15: **Schéma illustratif du prototype d'interaction médicamenteuse.**

2 photos sont comparées et identifiées. La base de données issues de L'ANSM qui regroupe les interactions est interrogée, et l'interaction est avérée ou non.

Le développement informatique qui embarque des algorithmes d'intelligence artificielle fonctionne très souvent de manière itérative. Pour montrer l'intérêt d'un projet basé sur le machine learning, nous allons développer un prototype (communément PVM: le produit viable minimum) dont il convient de définir les limites. Dans notre cas, le modèle sera capable d'identifier les médicaments à base de Paracétamol et leur interactions possibles avec d'autres produits. Ce prototype a pour but de montrer comment la connaissance métier alliée à l'intelligence artificielle peut faire émerger des solutions innovantes.

5.3. Le détecteur d'objets

5.3.1. Détecteur d'objets "historique"

La détection d'objets dans les images est devenue de plus en plus commune, en particulier dans le développement de solutions de vision assistées par ordinateur. En outre, la reconnaissance faciale fait partie intégrante des fonctionnalités disponibles sur les réseaux sociaux. Viola et Jones ont eu une contribution significative pour le développement de la détection d'objets notamment via la cascade de Haar [7,8].



Figure 16: **Cascade de Haar: filtres de détection en action.**

Les caractéristiques pseudo Haar sont représentées par une valeur unique obtenue en soustrayant une somme de pixels sous un rectangle blanc et sous un rectangle noir.

Source: https://docs.opencv.org/3.2.0/d7/d8b/tutorial_py_face_detection.html

Cette technique est basée sur l'apparence de l'image. A partir de la valeur des pixels, Viola et Jones ont défini des caractéristiques très simples appelées caractéristiques pseudo-Haar [7]. Ces caractéristiques sont calculées à partir de la différence des sommes de deux ou plusieurs zones rectangulaires adjacentes. Ces techniques encodent une

représentation des zones de l'image qui peut être classifiées plus facilement par un algorithme.

Par un système de fenêtrage de l'image, les zones comprenant l'objet d'intérêt sont classifiées comme positives et la position de l'objet est obtenue. Ces techniques présentent de nombreuses limites: leur sensibilité au bruit présent dans l'image, la difficulté à généraliser la détection d'objets dans des situations nouvelles, l'invariance à la rotation et enfin la sensibilité à l'occlusion. De nouvelles techniques, plus robustes et rapides ont émergé avec le développement du deep learning. YOLO est un algorithme de détection d'objets très rapide et qui fonctionne également sur téléphones mobiles. YOLO est un détecteur d'objets qui n'utilise pas le mécanisme traditionnel de fenêtrage glissant. [9]

En effet, ce fenêtrage permet d'analyser des zones de l'image de manière successive. Cette opération coûteuse peut ralentir le temps d'exécution de manière considérable, lorsque l'image est grande ou lorsque le pas de fenêtrage est petit.

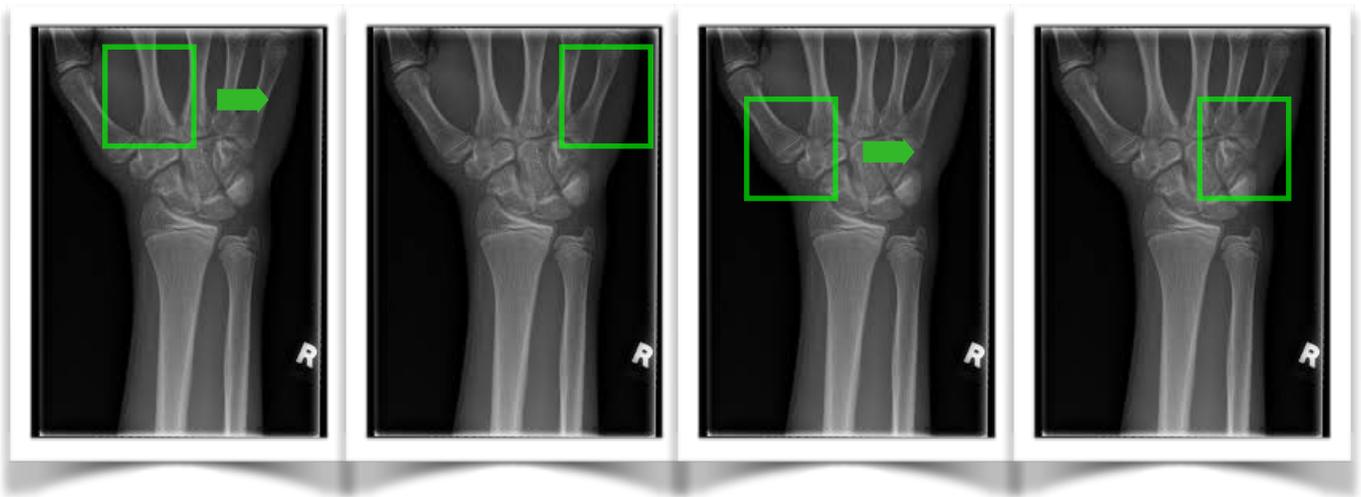


Figure 17: **Détection d'objets par fenêtrage glissant.**

Grâce à la technique de fenêtre glissante: L'image est parcourue progressivement. L'objet d'intérêt est alors recherché de manière séquentielle. Cette technique a été historiquement utilisée dans des travaux de recherche menés par L'E.S.I.E.E sur le traitement d'images

médicales. Cette approche est applicable à la détection de lésion(s) est en revanche très coûteuse en ressource.

Source: <http://xrayhead.com/rsna2016.pdf>

5.3.2. Les détecteurs récents

L'algorithme YOLO s'affranchit de cette approche traditionnelle qui scanne l'image successivement. En effet, ce nouvel algorithme permet une analyse de l'image en une fois. Les opérations de convolution implémentent la notion de fenêtre glissante en une passe, ce qui permet de réduire le temps de calcul de manière significative.

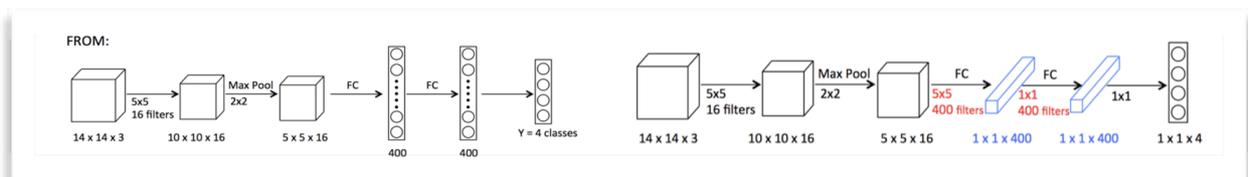


Figure 18: **Les particularités d'implémentation de Y.O.L.O permettent un traitement en temps réel.** Les couches complètement connectées ou dense peuvent être considérées comme des couches de convolution. Cette conversion est nécessaire pour pouvoir implémenter le fenêtrage glissant. C'est une des astuces implémentées par YOLO qui a révolutionné la détection d'objet [11]. Au lieu de réaliser les calculs de manière itérative YOLO permet d'automatiser les calculs en une unique passe.

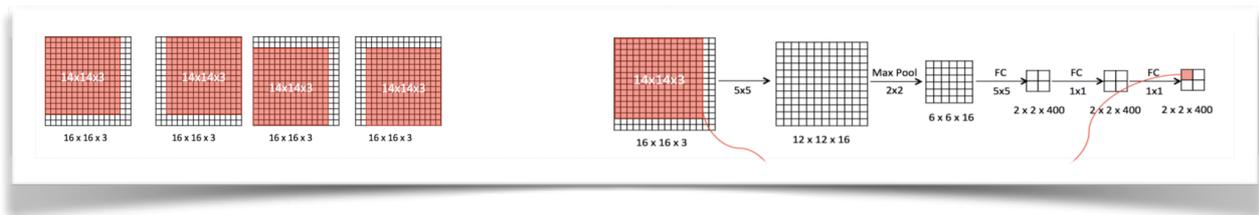


Figure 19: Le fenêtrage glissant implémenté par un réseau 100% convolutif.

À gauche: Le modèle défini ci-dessus a une taille d'entrée de 14×14 pour une sortie de taille $1 \times 1 \times 400$. Si on utilise une image dont la taille est plus grande, ici 16×16 , on obtient en sortie une Matrice de taille $2 \times 2 \times 400$.

À droite: YOLO utilise les propriétés de la convolution pour s'affranchir de la fenêtre glissante. Dans la matrice d'arrivée chaque pixel correspond à une 4 des zones identifiées à gauche.

Supposons que la taille de la fenêtre soit 14×14 . Dans le cas d'un fenêtrage classique, avec un espacement ou pas de 2, on obtiendrait 4 zones successivement analysées comme ci-dessus. En appliquant à cette image (dont la taille est supérieure à 14×14), la succession de "convolutions" et "pooling", on obtient une matrice 2×2 . Chaque élément représente une pièce de taille 14×14 issue de l'image originale. Ce qui correspond à un fenêtrage de taille 14×14 avec un espacement de 2 pixels. Le fenêtrage glissant s'est réalisé en 1 passe.

Les derniers algorithmes de détection sont basés sur les réseaux convolutifs. Un des paramètres cruciaux de la détection d'objets réside dans la proposition de zones d'intérêt (où l'objet pourrait se situer). RCNN, Faster RCNN possèdent une unité qui génère un très grand nombre de zones où l'objet d'intérêt pourrait se situer. Cette suggestion de zones rend l'approche relativement lente, et l'apparition de l'algorithme appelé YOLO a permis de révolutionner la discipline. Le concept derrière YOLO est de regarder l'image une seule fois: l'image est quadrillée, et chaque "patch" est évalué ou classifié selon l'assurance du modèle quant à la présence d'un objet. Le modèle évalue également la probabilité

conditionnelle $P(\text{classe}/\text{objet})$. Enfin, le modèle évalue la position des cadres de délimitations des objets à travers quatre variables :

- x: abscisse du centre de l'objet
- y: ordonnée du centre de l'objet
- w: largeur de l'objet
- h: hauteur de l'objet

L'algorithme YOLO a subi plusieurs améliorations, qui lui permettent à présent de détecter des objets de petite taille.



Figure 20: **Détection d'objets en temps réel.**

La Détection d'objets en temps réel est nécessaire aux solutions intelligentes des villes connectées. L'algorithme non-problème spécifique peut distinguer à gauche: voiture, piétons et bus. Cette approche est facilement extensible au domaine de la grande distribution, au milieu: détection de produits pour bébés, shampoings, produits d'épilation, etc. Et appliqué à la détection de médicaments (à droite).

Source: <https://www.youtube.com/watch?v=EhcpGpFHCrw>

Source : <https://dzone.com/articles/image-processing-with-deep-learning-a-quick-start>

Source: <https://tinyurl.com/IEEE-pills>

5.4. Méthodologie

De manière surprenante, il est plutôt rare de trouver des ressources en ligne des introductions aux modèles prédictifs ayant des applications pertinentes pour le pharmacien. En revanche, L'utilisation du "Deep learning" s'est imposé dans les analyses radiographiques car il est possible d'interpréter et de comprendre les réseaux convolutifs en particulier grâce aux cartes de saillance visuelle. Ces dernières permettent d'identifier les régions de l'image d'entrée qui ont contribué à une classification donnée. Le résultat est une carte de chaleur mettant en sur-brillance la région qui a permis de voir comme un ConvNet et de comprendre les éléments qu'il a pris en compte pour parvenir à une conclusion donnée.

Le but du présent travail est de faciliter la compréhension du processus pour un non-expert. Le langage de développement sera Python 3.6. Nous nous efforcerons d'assurer la compatibilité avec les versions antérieures. Le jeu de données est disponible sur GitHub à l'adresse: https://github.com/assansanogo/AI_for_pharmacists.

5.4.1. Entraînement du module de détection

Nous allons suivre le diagramme présenté ci-dessous afin d'entraîner le module de détection d'objets.

Les étapes de développement nécessitent un premier recueil d'images puis leur annotation manuelle souvent très fastidieuse, et enfin l'entraînement d'un ou plusieurs modèle de détection. On peut ensuite répéter les itérations de manière plus rapide car le modèle servira de pré-annotateur

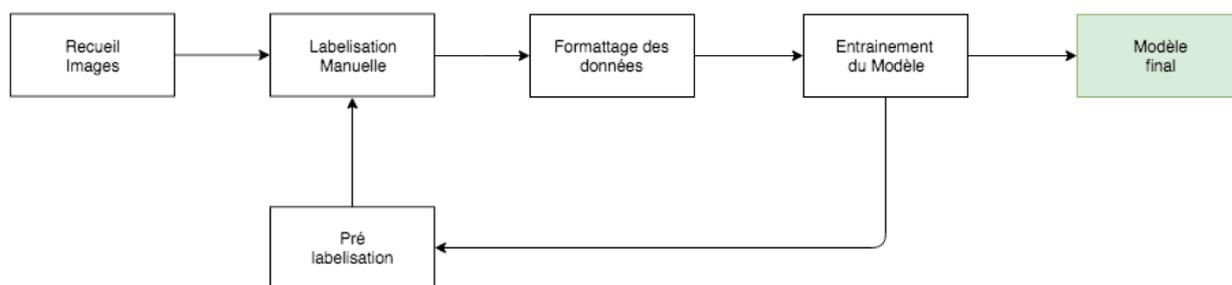


Figure 21: **Les itérations garantissent la robustesse des algorithmes.**

Dans les étapes d'entraînement en Machine Learning, une grande partie des efforts se concentre sur la préparation des données, puis sur l'entraînement itératif du modèle.

On pourra ré-entraîner le modèle sur un volume de données plus élevé et garantir une meilleure généralisation du modèle. On parle d'apprentissage itératif.



Figure 22: **Détection d'objets appliquée à la santé.**

La boîte de Doliprane®/Paracetamol est détectée par l'algorithme dans les mains d'une patiente. Source: <https://www.gettyimages.co.uk/photos/antalgic>

Dans le cadre d'un apprentissage supervisé, l'annotation va permettre au modèle de confronter ses prédictions aux valeurs attendues. Pour ce qui est de l'étiquetage des images, il est possible de développer des outils spécialisés à une tâche particulière, mais nous aurons recours à un outil d'étiquetage développé par le "Oxford Group": VGG Annotator.

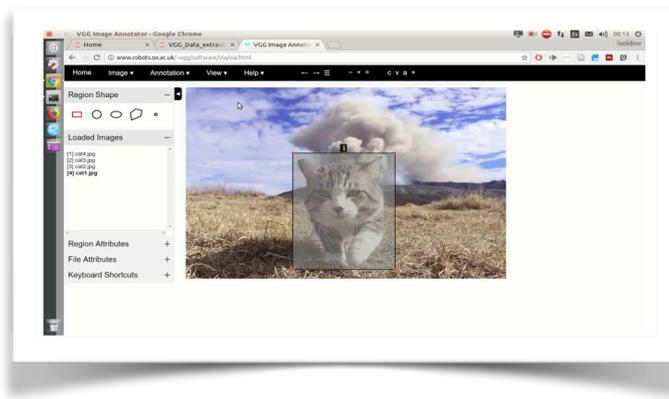


Figure 23: **Etiquetage des données est une étape nécessaire à l’entraînement de modèles.**

VGG annotator est l’outil “open source” de référence pour annoter des objets, c’est à dire dessiner des cadres autour des objets d’intérêt et sauvegarder ces étiquettes sous forme de fichiers exploitables (extension .csv) pour l’entraînement de réseaux neuronaux.

Dans le cadre du Machine Learning, une des difficultés réside dans l’absence de jeu de données pré-annoté. Très souvent les données sont dédiées à un projet particulier ou sont très spécifiques d’une problématique donnée. Le recueil d’images issues du web sera donc la première tâche. Notons, qu’il s’agit là d’une des limites de ces algorithmes. En effet, l’utilisation d’images antérieures non validées, les erreurs d’annotations sont des types d’erreurs qui vont compromettre l’apprentissage des modèles.

Google a été accusée d’utiliser de manière abusive ses internautes, les forçant à résoudre des tests de vérifications les “Captcha” afin de réduire les attaques automatiques par les “bots”.

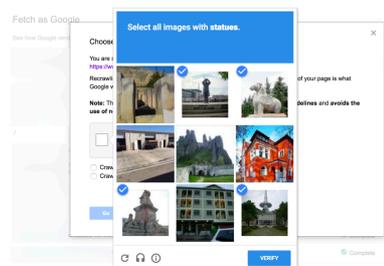




Figure 24: Utilisation détournée de Captcha pour annoter les données.

Les captcha sont utilisés pour empêcher les attaques malveillantes réalisées via des “bots”. Google a été suspecté d'utiliser les internautes à leur insu pour annoter des images ou du texte. Cette utilisation de main d'oeuvre gratuite a pu permettre d'améliorer de manière colossale leur algorithme de détection (Google Vision api) de manière déguisée.

Source : <https://askleo.com/why-captcha/>

Source : <https://medium.com/@thenextcorner/you-are-helping-google-ai-image-recognition-b24d89372b7e>

Les techniques d'apprentissage profond ont cette particularité de nécessiter un nombre très important de données pour converger. Ceci est dû en partie au grand nombre de paramètres (~ Millions de paramètres). En pratique, plutôt que d'entraîner le modèle depuis le début, nous utilisons les modèles pré-entraînés sur des datasets de grande taille tels que ImageNet disponible à (<http://www.image-net.org/>) ou COCO disponibles à (<http://cocodataset.org>). Nous utilisons ces modèles pré-entraînés génériques comme point de départ pour faciliter l'apprentissage sur un dataset particulier et souvent de taille limitée. On parle de réglage fin.

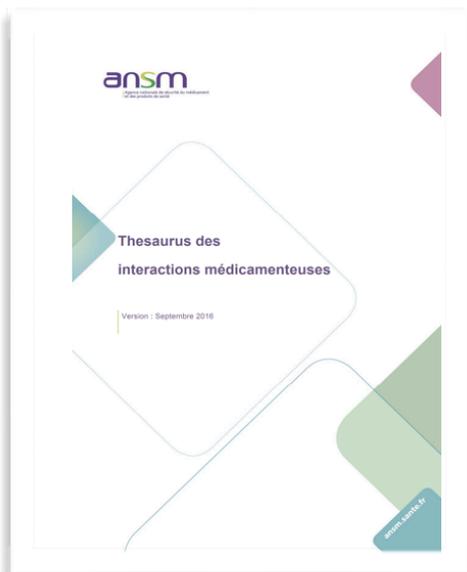
Pour entraîner le modèle de type YOLO, il convient d'avoir quelques milliers d'exemples. Plus le jeu de données initial est petit, plus il est difficile au modèle de généraliser à des données complètement inconnues. Très souvent, l'annotation prend du temps et est coûteuse en personnel. Il existe des plateformes d'annotation telles que Mechanical Turk

créée par Amazon, mais présentent très souvent un coût prohibitif dans le cadre de réalisation de prototypes. Il existe à présent des plateformes qui connecte des populations faiblement qualifiées de bas coût pour réaliser ces tâches répétitives, on peut citer : Figure Eight, iMerit, CloudFactory et la pépite française: Isahit. Certains peuvent voir en ce phénomène une "délocalisation 4.0".

Comme nous ne possédons pas une telle main d'oeuvre pour établir ce prototype, nous allons avoir recours à une technique couramment utilisée en deep learning: "l'augmentation artificielle des données". Pour augmenter artificiellement la taille de notre jeu de données, nous créons des variations des images en notre possession, en appliquant des transformations (flou, bruit, translations et rotations). Ces transformations vont améliorer les capacités de notre modèle à généraliser à des données qui diffèrent du jeu de données initial et palier au manque de données initial. Dans le cas présent, nous augmentons les données d'un facteur 10.

5.4.2. Le thésaurus:

Le bloc logique d'interaction médicamenteuse repose sur une base de données listant les interactions médicamenteuses recensées par l'ANSM. Le thésaurus des interactions médicamenteuses est en libre accès. Comme beaucoup de documents, il n'a pas été réalisé d'une manière optimale pour la data science. Il est donc nécessaire d'en extraire les données et de les restructurer d'une façon consommable par les algorithmes. Dans le jargon de la data science, le processus de réorganisation de la donnée fait passer les données non structurées en données structurées.



		ainsi que pendant et 3 mois suivant l'arrêt du traitement.
ABIRATERONE		
+ ANTICONGULSIVANTS INDUCTEURS ENZYMATIQUES	Diminution notable des concentrations plasmatiques de l'abiraterone, avec risque de moindre efficacité.	Association DECONSEILLÉE
+ FLECAINIDE	Risque d'augmentation des effets indésirables du flecaïnide, par diminution de son métabolisme hépatique par l'abiraterone.	Précaution d'emploi Surveillance clinique. Si besoin, adaptation de la posologie du flecaïnide pendant le traitement par l'abiraterone.
+ METOPROLOL	Chez l'insuffisant cardiaque, risque d'augmentation des effets indésirables du métoprolol, par diminution de son métabolisme hépatique par l'abiraterone.	Précaution d'emploi Surveillance clinique. Si besoin, adaptation de la posologie du métoprolol pendant le traitement par l'abiraterone.
+ PROPAFENONE	Risque d'augmentation des effets indésirables de la propafénone, par diminution de son métabolisme hépatique par l'abiraterone.	Précaution d'emploi Surveillance clinique. Si besoin, adaptation de la posologie de la propafénone pendant le traitement par l'abiraterone.
+ RIFAMPICINE	Diminution notable des concentrations plasmatiques de l'abiraterone, avec risque de moindre efficacité.	Association DECONSEILLÉE

Figure 25: **Thésaurus des interactions médicamenteuses.**

Bien qu'en libre service, le formatage demande une réorganisation substantielle pour lier médicaments et leurs interactions. La molécule princeps est surlignée en bleu, et les interactions en rouge.

Dans la version initiale du prototype, nous ne nous concentrerons uniquement que sur un type d'interaction.

Dans le cadre de notre prototype, nous considérons que l'interaction d'intérêt est :

Paracétamol – Alcool.

5.5. Résultats et Limites



Figure 26: Exemples d'objets détectés après entraînement.

Après entraînement le modèle détecte sur les différentes photo de gauche à droite: la bouteille de bière, la boîte de Doliprane® en comprimés et les boissons non alcoolisées. Advil® n'a pas été fourni lors de l'apprentissage, il est donc invisible pour le modèle. (Détections sur les images)

Le code final est disponible sous forme détaillée et commentée dans le répertoire hébergé sur GitHub. La routine "demo.py" prend en entrée 2 images et est capable de conclure à l'interaction.

Le protocole décrit ci-dessus souligne la simplicité pour développer un prototype. L'idée proposée se veut plus réaliste ou plus proche des cas d'usage réels que les solutions déjà présentes sur le marché. En effet, le patient qui se questionne face au risque d'une intoxication, ou qui préfère se renseigner doit entrer le nom du ou des médicaments concernés.

Cette technologie est cependant très gourmande en données, et requiert une annotation conséquente afin d'avoir un modèle robuste qui ne se limite pas à une unique interaction. D'autres perspectives peuvent être explorées (autres architectures) adaptées aux cas avec peu de données. Il a été montré récemment par ailleurs, qu'il est possible de créer de

manière synthétique des observations qui peuvent perturber ou mettre en échec le modèle d'apprentissage.

Ces perturbations prennent la forme d'étiquettes ou d'images dépourvues de sens pour un oeil humain. Le cerveau humain sait en faire abstraction et classifie correctement une image même en présence de ces moyens d'attaque ce qui prouve la nette infériorité des algorithmes. Certains chercheurs ont également mis au point des simulations d'attaque de type mono pixel (un seul pixel suffit pour déstabiliser certains algorithmes). Ceci nous pousse à renforcer la robustesse des algorithmes actuels ou même à envisager des méthodes d'apprentissage qui ne sont pas vulnérables à ce genre d'attaque.

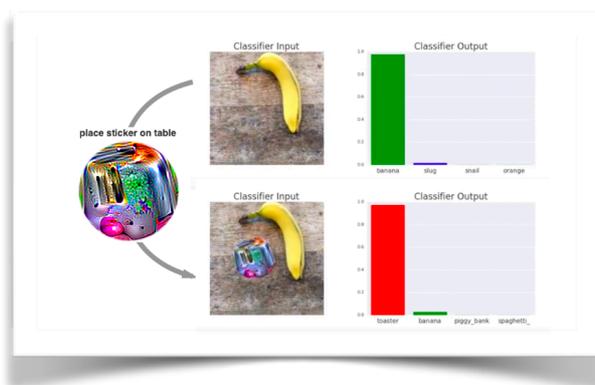


Figure 27: **Perturbation par attaque par “patch” ou étiquette.**

Après entraînement certains modèles sont particulièrement sensibles au patch, faisant basculer l'image originale d'une classe à l'autre : ici l'image d'abord prédite comme une banane devient malheureusement un grille-pain!

Source: <https://mc.ai/know-your-enemy-the-fascinating-implications-of-adversarial-examples/>

Des techniques d'entraînement et de renforcement peuvent limiter l'action de ces attaques sans pour autant garantir à l'heure actuelle d'une protection universelle. Doit-on

condamner les techniques d'apprentissage? Probablement non, mais le machine learning et deep learning tels que nous les connaissons doivent profondément évoluer pour se prémunir de contournement de système de vision assistée par ordinateur.

Jean Emmanuel Bibault développe le point précédent lié aux conséquences de l'utilisation d'une technologie encore faillible et pourtant utilisée sur les patients. En tant que professionnels de santé, nous sommes en droit de nous interroger sur les risques d'une régulation absente, en particulier dans le cas de prédiction d'apparition de pathologies ou de diagnostic erronés.

Aux Etats-Unis, en 2019 la FDA a approuvé une dizaine de solutions médicales s'appuyant sur les techniques d'intelligence artificielle, pourtant seule une d'entre elles a été validée après des essais cliniques. Dans une course à la technologie effrénée où La Chine et les Etats-Unis sont les leaders, les décideurs du domaine de la santé doivent répondre à la question difficile: faut-il une régulation stricte pour garantir la sécurité des médicaments au prix d'un retard dans l'innovation liée à la santé ? [10]

6. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN CLINICIEN

La **pharmacie clinique** est une discipline pharmaceutique qui permet aux pharmaciens : d'optimiser les choix thérapeutiques, la dispensation et l'administration des médicaments au patient. La pharmacie clinique a plus de 30 ans d'existence dans les pays anglo-saxons. J'ai pu constater à l'Université d'Oklahoma que les pharmaciens font partie intégrante des services cliniques et travaillent avec les médecins. Le pharmacien est là au moment de la prescription et l'optimise éventuellement. Il réalise parfois un changement de molécule au sein de la classe thérapeutique. Il échange avec les patients sur leur(s) traitement(s) médicamenteux et diagnostique des problèmes liés aux médicaments (observance ou effets indésirables). L'existence d'outils intelligents peut permettre une meilleure prise en charge (dépistage, évaluation de risques, etc.) et éventuellement une réduction des frais de santé.

6.1. Jeu de données

Le jeu de données utilisé provient du site kaggle (www.kaggle.com) et est disponible dans le répertoire github sus-cité. Il est la propriété de : McKinsey & Company and Analytics Vidhya et est utilisable à des fins non commerciales.

Les variables explicatives sont regroupées dans le tableau ci-dessous.

- **Id:** identification du patient après anonymisation
- **Gender:** genre du patient (Homme/Femme)
- **Hypertension:** variable booléenne (0: pas d' hypertension, 1: hypertension)
- **Heart_disease:** variable booléenne (0: pas de maladie cardiaque, 1: maladie cardiaque)
- **Ever_married:** variable booléenne (0: jamais marié/e, 1: déjà marié/e)

- **Work_type:** variable catégorielle décrivant le type de travail
- **Residence_type:** variable catégorielle prenant 2 valeurs (rural/urbain)
- **Avg_glucose_level:** variable continue décrivant le niveau de glucose après repas - sans indication de la nature du repas.
- **BMI:** variable continue décrivant l'indice de masse corporelle (IMC)
- **smoking_status:** variable catégorielle décrivant la consommation tabagique

La variable cible est:

stroke: variable booléenne (0: n'a pas subi d'accident cardiovasculaire cérébral, 1: a déjà subi une attaque cérébrale)

DATA DICTIONARY	
Variable	Definition
id	Patient ID
gender	Gender of Patient
age	Age of Patient
hypertension	0 - no hypertension, 1 - suffering from hypertension
heart_disease	0 - no heart disease, 1 - suffering from heart disease
ever_married	Yes/No
work_type	Type of occupation
Residence_type	Area type of residence (Urban/ Rural)
avg_glucose_level	Average Glucose level (measured after meal)
bmi	Body mass index
smoking_status	patient's smoking status
stroke	0 - no stroke, 1 - suffered stroke

Figure 28: **Description du jeu de données “Healthcare Dataset Stroke Data”.**

L'origine du jeu de données a été gardée confidentielle. Il appartient cependant à McKinsey & Company and Analytics et Analytics Vidya et fut utilisé dans le cadre d'un hackathon.

6.2. Analyse exploratoire des données

Les variables explicatives sont liées à la variable cible (prédite) et sont liées entre elles.

On définit la corrélation grâce à la grandeur mathématique appelée covariance.

Cette mesure quantifie l'association linéaire ou la relation entre deux variables. Quand 2 variables covarient, un écart à la moyenne d'une variable est accompagné par un écart dans le même sens (corrélation positive) ou dans le sens opposé (corrélation négative) de l'autre variable pour un sujet donné.

Donc, pour une variable donnée, s'il y a corrélation, pour chaque valeur qui s'écarte de la moyenne, on s'attend à trouver un écart à la moyenne pour la seconde variable .

Pearson a défini le coefficient de corrélation:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

avec la moyenne de la variable x telle que: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Ce coefficient de corrélation, représente la covariance standardisée par le produit des écarts-types des 2 variables, varie entre -1 et 1.

Un coefficient égal à 1 indique une corrélation parfaite entre les 2 variables. À l'inverse, un coefficient de -1 indique une corrélation négative parfaite: lorsque la variable v1 augmente, la variable v2 diminue dans la même proportion.

Ci-dessous, les 3 graphiques illustrent les cas décrits ci-dessus : (respectivement corrélation positive, négative et absence de corrélation.) Notons que dans les situations rencontrées en pratique, le degré de corrélation de variables explicatives entre elles est rarement égal à 1 ou -1. Si tel était cas, une des deux variables explicatives totalement corrélée devient inutile. Enfin, si la variable explicative était parfaitement corrélée à la cible, la prédiction deviendrait triviale.

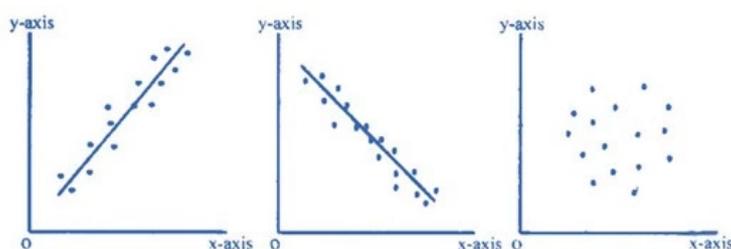


Figure 29: **Corrélation en images**

À gauche : la corrélation entre les 2 variables est positive. Au milieu : la corrélation est négative. À droite : il n’y a pas de corrélation.

Sur la carte de chaleur ci-dessous, en bleu, les variables présentent une corrélation positive prononcée avec l’absence d’accident cardiovasculaires. En orange, les variables présentent une corrélation négative importante avec l’absence d’attaque cérébrale. On voit que les variables concernant l’âge, l’hypertension, les maladies cardiovasculaires, et le niveau de glycémie sont positivement corrélées au risque d’accident cardiovasculaire. En d’autres termes hautes valeurs pour ces variables sont concordants avec un risque accru d’accident cardiovasculaire.

En outre, le statut marital et le type de travail ont également une corrélation (plus faible) avec la survenue d’accident cardiovasculaire.

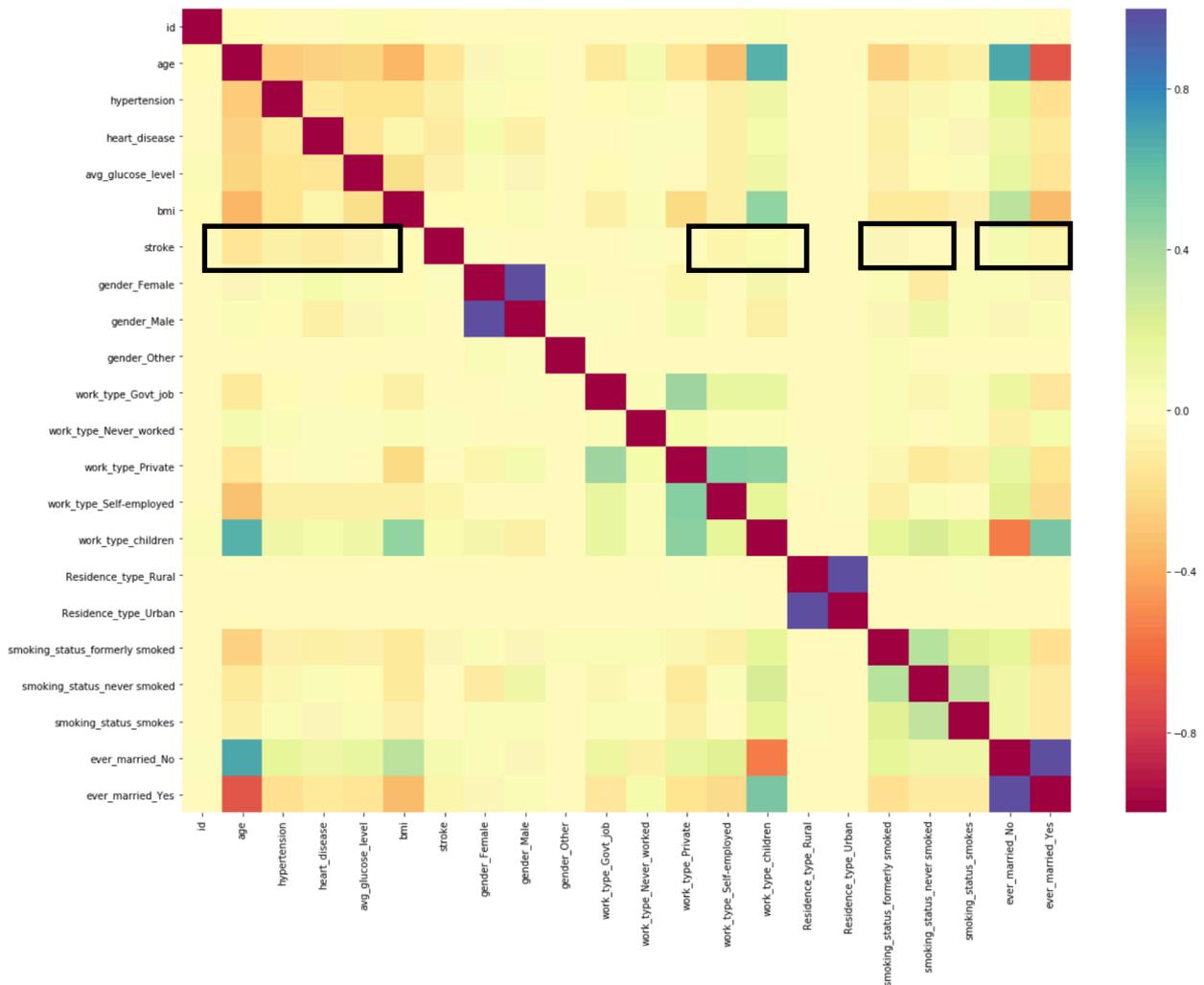


Figure 30: **Corrélation des variables explicatives entre elles et avec la variable explicative.**

En orange, les variables présentent une corrélation négative prononcée. En bleu les variables présentent une corrélation positive importante. On voit que les variables concernant l'âge, l'hypertension, les maladies cardiovasculaires, et le niveau de glycémie sont positivement corrélées au risque d'accident cardiovasculaire. En outre, le statut marital et le type de travail peuvent avoir une corrélation avec la survenue d'accident cardio vasculaire.

De manière surprenante, le statut (fumeur ou non) qui est connu scientifiquement pour être un facteur de risque, ressort ici comme ayant un impact mineur sur la survenue d'un

accident cérébral. Ceci est dû à plusieurs problèmes liés au jeu de données: d'abord les données sont hautement déséquilibrées par rapport à cette variable, ensuite il s'agit de variables agrégées sans information sur le comportement du fumeur (quantité et fréquence) enfin, nous n'avons aucune certitude sur la représentativité du jeu de données. Ceci souligne la dangerosité d'une intelligence artificielle dont les résultats ne sont pas critiqués sous soumis à une analyse rigoureuse.

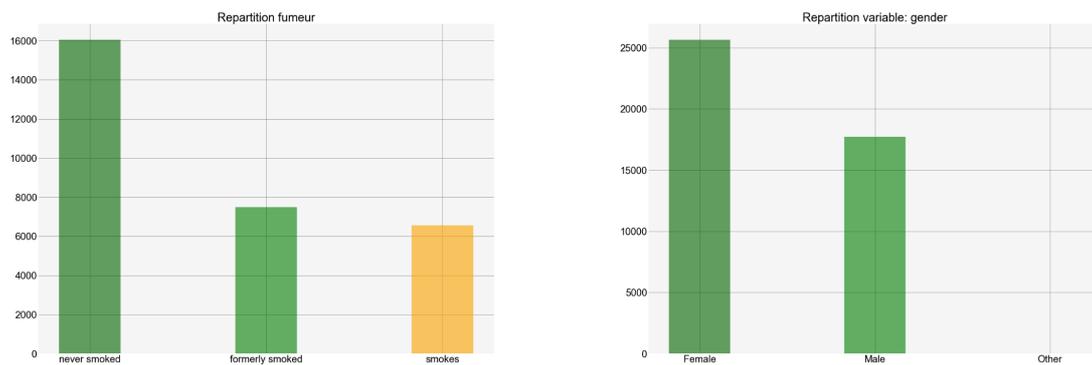


Figure 31: **Dataset non équilibré**

En particulier pour la variable Fumeur. On observe un léger déséquilibre entre le nombre de femmes et d'hommes dans cette étude.

6.3. Méthodologie

6.3.1. Généralités

Nous avons tenté d'explicitier le fonctionnement des réseaux de neurones dans les parties précédentes, Malgré cela, Les réseaux de neurones convolutifs ont souvent été décrits comme des “boîtes noires” et il est vrai que la compréhension de leur mode de fonctionnement est encore parcellaire. Contrairement à d'autres algorithmes dont on a prouvé la convergence, leur maîtrise relève beaucoup de l'empirisme.

La complexité de ces modèles et le désir de rendre l'intelligence artificielle plus accessible possible a naturellement donné lieu au développement de solution qualifiées de machine learning automatique. Google a créé en 2019 Cloud Vision AutoML, H2O.ai a lancé sa solution clé en main “driverless” et d'autres se sont lancé dans cette quête de démocratisation. Bien que louable, cette volonté peut avoir l'effet contraire si les utilisateurs manquent d'éducation statistique et d'esprit critique quant à la pertinence du modèle et les possibles différents biais introduits.

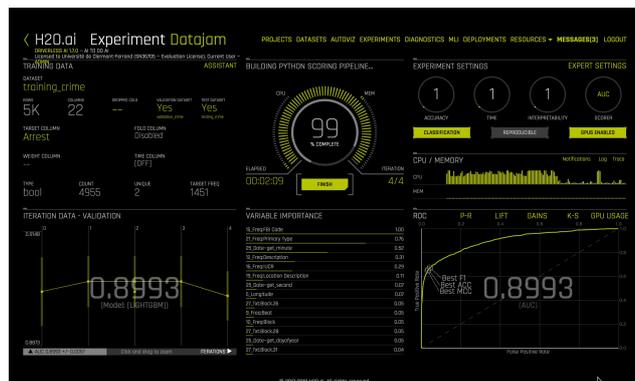


Figure 32: Le machine learning en mode automatique.

L'utilisateur, utilise les modèles comme des boîtes noires et pas une seule ligne de code. Seul certains critères lui sont accessibles : complexité du modèle, temps d'entraînement, exactitude et interprétabilité. Tel un pilote, l'utilisateur (non expert) choisit ses paramètres et obtient un modèle pré-entraîné en quelques heures.

Il est indéniable qu'une bulle s'est créée autour de l'intelligence artificielle, et en particulier du "Deep learning". Or, dans beaucoup de cas, nous avons besoin d'algorithmes puissants, qui gèrent le problème de données manquantes et que l'on peut expliciter.

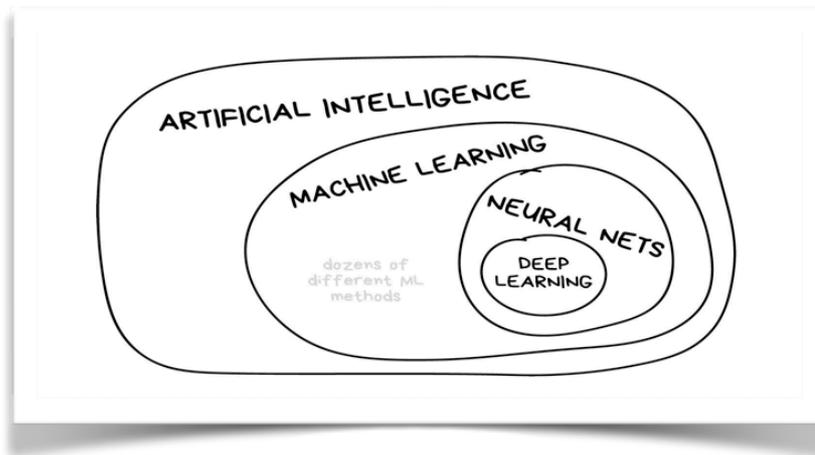


Figure 33: **L'intelligence artificielle.**

Les algorithmes utilisés en intelligence artificielle ne se limitent pas aux réseaux de convolution utilisés en "Deep Learning".

Dans le domaine lié à la santé, la nécessité d'obtenir des modèles interprétables se justifie par le besoin de validation. Cette validation assure la sécurité des patients et permet de garantir le cadre d'utilisation de l'algorithme.

C'est pourquoi les algorithmes basés sur les arbres de décision tels que les forêts d'arbres aléatoires sont encore préférés par la communauté médicale. Ce besoin a donné naissance à un sous-domaine appelé : " explainable machine learning" ou d'apprentissage automatisé interprétable.

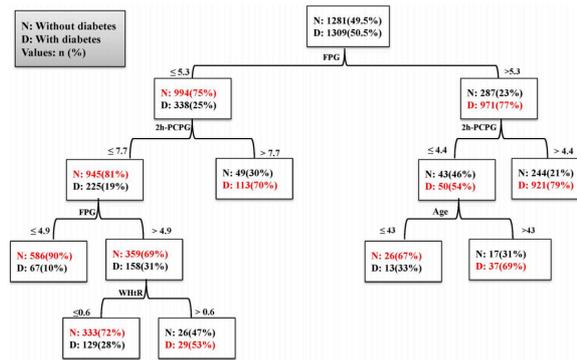
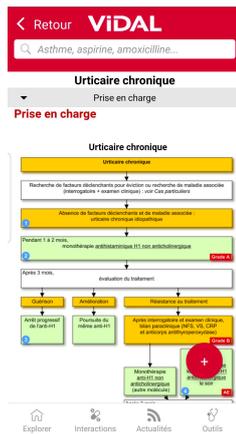


Figure 34: Des arbres prédictifs aux arbres de décision thérapeutique

Dans la prise en charge du patient, on utilise beaucoup les arbres de décision comme sur la photo ci-dessus à gauche. Les algorithmes en forme d'arbres prédictifs (régression ou classification) peuvent donner naissance à des outils d'aide à la décision comme les schémas décisionnels. Comme dans le schéma ci-dessus à droite, à chaque étage de l'arbre une décision est basée sur la valeur d'une variable, qui permet de séparer le mieux possible les patients en 2 groupes (diabétiques, non-diabétiques).

Source : Decision tree-based modelling for identification of potential interactions between type 2 diabetes risk factors: a decade follow-up in a Middle East prospective cohort study

<https://bmjopen.bmj.com/content/bmjopen/6/12/e013336.full.pdf>

6.3.2. XGBoost

XGBoost est l'aboutissement de plusieurs évolutions ou techniques utilisées en Machine learning. Il s'agit d'un modèle performant interprétable, rapide et qui gère les données manquantes. Dans plusieurs compétitions organisées par kaggle, l'organisme de référence des compétitions en Data Sciences, cet algorithme a su dépasser les réseaux de neurones convolutifs gourmands en données. XGBoost est l'acronyme pour extreme gradient boosting. Il convient d'expliquer ce qu'est la technique de Boosting, l'intérêt d'utiliser le gradient et ce qui justifie l'utilisation du terme "extreme" dans cette implémentation.

6.3.3. Boosting

Le boosting n'est pas une technique récente. À la différence des techniques de bagging ou "Bootstrap aggregating", qui agrègent plusieurs classifieurs/régresseurs forts en un méta-classifieur, la technique de boosting s'applique à des méthodes générales capables de produire des décisions très précises à partir d'un ensemble de règles de décision ou classifieurs "faibles". Ces classifieurs "faibles" par un mécanisme de système de vote deviennent un classifieur "fort". On leur donne l'adjectif "faible" car ils sont un peu meilleurs que le hasard. Ces méthodes s'appliquent aussi bien à la classification qu'à la régression. L'entraînement par boosting est un entraînement séquentiel. D'abord, le premier classifieur faible est d'abord entraîné. Ensuite, le boosting ayant lieu au niveau des observations: les exemples mal classés sont "boostés" pour qu'ils aient davantage d'importance lors de l'entraînement du prochain classifieur faible.

Enfin, à l'issue du boosting, les classifieurs sont pondérés par leur qualité de classification. Plus le classifieur classe correctement plus il sera important dans le vote final. Enfin, toute l'approche décrite précédemment est valide dans le cas d'une régression "boostée".

Cette approche possède certaines limites dont on peut citer:

- le risque de sur-apprentissage (surtout en fin d'entraînement ou si le nombre de classifieurs est trop élevé).
- La sensibilité aux valeurs extrêmes et aux valeurs bruitées.

Le critère d'arrêt, le choix du nombre de classifieurs sont des hyperparamètres qui influent sur la performance de l'algorithme qui peuvent être optimisés par optimisation combinatoire, aléatoire ou bayésienne des paramètres.

On distingue deux grands types de boosting simples: boosting par sous-ensembles et boosting probabiliste.

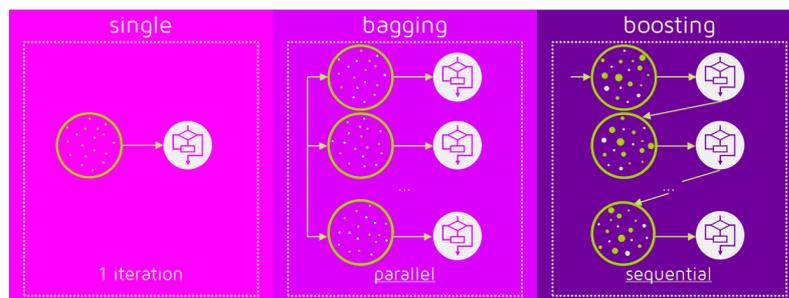


Figure 35: **Boosting et Bagging**

Ces deux approches très différentes permettent de créer des méta-modèles. Dans le cas du bagging, elle repose sur un ensemble de classifieurs “forts” dont on pondère les prédictions. A l’inverse, dans le cas du boosting, le modèle final résulte d’un ensemble de classifieurs faibles. Ces deux techniques permettent d’obtenir des modèles plus complexes mais plus performants qu’un unique algorithme.

Source: <https://quantdare.com/what-is-the-difference-between-bagging-and-boosting/>

6.3.3.1. Boosting par sous-ensembles

L'idée est d'utiliser plusieurs algorithmes d'apprentissage qui peuvent être de nature très diverse (un arbre de décision, une règle bayésienne de classification, une décision dépendant d'un hyperplan, etc.) sur n sous-ensembles d'apprentissage et dont le poids est relié à la capacité à classer certaines observations difficiles.

Quatre idées fondamentales sont à la base des méthodes de boosting sous-ensemblistes:

Plaçons nous dans le cas d'une classification à k classes)

1. On obtient d'abord une première hypothèse h_1 sur un sous-échantillon S_1 d'apprentissage de taille $m_1 < m$ (m étant la taille de l'échantillon d'apprentissage disponible S).
2. On apprend une i -ème hypothèse h_i sur un échantillon S_i de taille m_i choisi dans $S - S_{i-1}$ dont la moitié des exemples sont mal classés par h_{i-1} .
3. On apprend finalement une dernière hypothèse h_{k+1} sur m_{k+1} exemples tirés dans $S - \{S_i \}_{i=1 \dots k}$ pour lesquels $h_i \}_{i=1 \dots k}$ sont en désaccord.
4. L'hypothèse finale est obtenue par un vote majoritaire des $k+1$ hypothèses apprises. $H = \text{vote majoritaire}(h_1, \dots, h_{k+1})$

Dans le cas de la classification: "malade"/ "sain", en suivant le mode opératoire ci-dessus, on obtiendrait alors 3 classifieurs "faibles", dont le vote serait pondéré. C'est cet ensemble de classifieurs qui forme un comité "expert".

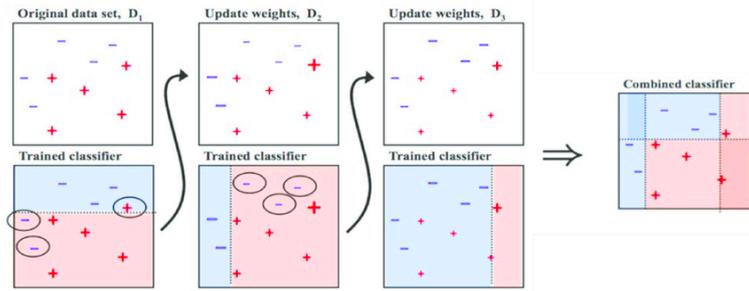


Figure 36: **Boosting et pondération des observations**

Le boosting s'applique à des problèmes de type "supervisé". Un premier classifieur faible h_0 va tenter de partitionner les données en deux groupes (malades et sains). Ce classifieur va réaliser des erreurs. Les exemples mal-classifiés -uniquement- vont être "boostés" (leur poids vont augmenter à l'itération suivante). Après une succession d'entraînements de classeurs faibles, ici $n=3$, la somme pondérée des 3 classifieurs forme un classifieur fort qui partitionne correctement les observations.

6.3.3.2. Boosting probabiliste (bayésien)

Quatre idées fondamentales sont à la base des méthodes de boosting probabiliste:

1. Chaque classifieur "faible" est en réalité un expert pour certaines observations données. Le classifieur faible peut être de type: classifieur bayésien naïf.
2. À la différence du boosting par sous ensembles (disjoints), on utilise une modification de la distribution des exemples disponibles pour entraîner chaque classifieur faible.
3. La pondération adaptative des votes se fait par une technique de mise à jour multiplicative.
4. Enfin, l'utilisation d'un comité de classifieurs faibles spécialisés que l'on fait voter pour atteindre une décision.

L'algorithme le plus connu s'appelle AdaBoost (pour adaptive boosting). L'une des idées principales est de définir à chacune de ses étapes $1 \leq t \leq T$, une nouvelle distribution D_t de probabilités à priori sur les exemples d'apprentissages en fonction des résultats de l'algorithme à l'étape précédente. La distribution s'adapte à chaque itération.

Le poids à l'étape t d'un exemple (x_i, u_i) d'indice i est noté $p_t(i)$. Initialement, tous les exemples sont équiprobables et auront donc un poids identique, puis à chaque étape, les poids des exemples sont recalculés. Le terme "poids" réfère à l'importance attribué à chaque observation dans le cas d'une bonne ou mauvaise classification.

6.3.4. Le boosting appliqué au gradient

XGBoost est un ensemble de modèles dont l'action conjointe permet de classifier ou régresser. Cette méthode est un type de boosting basé sur les résidus. On utilise le terme résidu car les classifieurs/régresseurs faibles vont être entraînés afin de prédire à chaque itération l'erreur provenant de l'itération précédente : $h_m(x) - Y$

- Soit un set d'entraînement (n observations): $\{(x_i, y_i)\}_{i=1}^n$,
- Une fonction objectif différentiable : $L(y, F(x))$,
- Un nombre d'itération : M

Initialisation avec une valeur constante gamma : $F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$.

De $m=1$ à M:

1. Calcul des pseudo-résidus: $r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$ for $i = 1, \dots, n$.

2. Fit a base learner (a tree) to fit the pseudo residuals: $h_m(x)$

3. **1D** Optimization: $\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$.

4. Mise à jour des modèles: $F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$.

Modèle Final: $F_M(x)$.

Les opérations de minimisation se font lorsque la dérivée de l'expression par rapport à la variable est nulle. Résoudre analytiquement ces équations requiert un temps de calcul important. Des méthode itératives ont été alors développées afin de résoudre ce problème d'optimisation.

Le gradient de la fonction objectif en un point est égal au coefficient directeur de la tangente de sa courbe représentative en ce point. Le gradient pointe toujours dans la direction de l'augmentation de la fonction objectif.

Devant ce constat, nous pouvons dériver un algorithme itératif qui retranche à chaque itération une fraction du gradient de la fonction objectif en ce point. Cette itération est continuée jusqu'à ce que la valeur de la fonction objectif ne change plus.

La fraction du gradient retranchée à chaque itération est communément appelée vitesse d'apprentissage. Comme illustré ci-dessous: si celui-ci est trop élevé, le minimum n'est jamais atteint. S'il est trop faible, le minimum est atteint en très grand nombre d'itérations, ce qui n'est pas désirable car cet algorithme est supposé accélérer la recherche du minimum.

La méthode de descente de gradient est une méthode classique d'optimisation du premier ordre, qui reste assez sensible aux minimas locaux. Pour s'affranchir de cette limitation, d'autres algorithmes d'ordre 1 incorporant une pénalisation par des termes d'ordre 2 (approximation de la matrice Hessienne) ont été mis en place.

C'est le cas d'algorithmes tels que Adam ou Adagrad communément préférés aux méthodes d'ordre 1 basées sur la descente de gradient, pour l'entraînement de réseaux de neurones convolutifs.

Il existe cependant d'autres méthodes d'optimisation basée sur le Hessien de la fonction à minimiser (Nelder-Mead est un exemple), et également des méthodes évolutionnistes : (algorithmes génétiques). De plus en plus data scientists questionnent la pertinence d'optimiseurs simples de degré 1 [12].

Enfin, l'implémentation de XGboost repose sur une technique d'optimisation différente des méthodes évoquées et dont le détail sort du cadre de cette étude.

6.3.5. Classes non équilibrées et stratification

Le dataset utilisé est particulièrement déséquilibré. On observe un rapport de taille supérieur à 50 lorsqu'on compare les observations: patients sains contre patients malades. Ceci correspond à ce qu'on attend dans une situation réelle. Afin de construire un modèle prédictif qui ne prédit pas systématiquement "patient sain", il convient de pondérer notre jeu de données afin de l'équilibrer.

6.3.5.1. Stratification

Comme décrit dans les parties précédentes, afin d'entraîner les modèles et quantifier leur performance, nous devons organiser nos données en un jeu de test et un jeu d'entraînement. Le modèle n'apprend uniquement sur le jeu d'entraînement, et le jeu de test représente des données inconnues. C'est ce jeu de test qui sert de validation externe de notre modèle.

Généralement, 80% des données servent à l'entraînement, 20% servent au test.

La stratification permet de s'assurer que le ratio sain/malade est le même dans le jeu de données d'entraînement que dans celui de test. Cette pratique permet de construire un jeu de données équilibré, évitant de biaiser le modèle lors de son apprentissage.

6.3.5.2. Sur-échantillonnage

Nous sommes dans un problème où les données sont déséquilibrées. Il existe plusieurs approches, on peut sous-échantillonner la classe majoritaire, sur-échantillonner la classe minoritaire. Il existe plusieurs méthodes d'échantillonnage, il faut être attentif à ne pas ajouter de biais supplémentaire dans l'équilibrage du jeu de données afin de garantir l'apprentissage du modèle et sa capacité à généraliser sur les données de test.

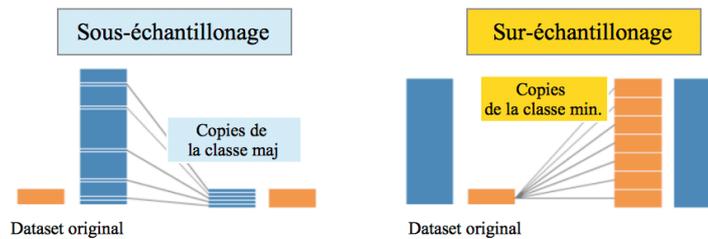


Figure 37: **Les remèdes aux jeux de données non équilibrés.**

Le sous-échantillonnage et le sur-échantillonnage sont 2 approches de base pour ré-équilibrer des jeux de données déséquilibrés.

Source: <https://towardsdatascience.com/having-an-imbalanced-dataset-here-is-how-you-can-solve-it-1640568947eb>

Comme décrit dans la Figure ci-dessus: dans le cas du sous-échantillonnage, n observations de la classe majoritaire sont choisies de telle sorte que l'effectif échantillonné correspond au nombre d'observations de la classe minoritaire. Dans le cas du sur-échantillonnage aléatoire simple ou bootstrapping: n observations de la classe minoritaire sont choisies aléatoirement (répétitions) de telle sorte que l'effectif échantillonné correspondent au nombre d'observations de la classe majoritaire. Dans le premier cas, il faut s'assurer que le jeu de données résultant aie une taille suffisante pour donner un pouvoir prédictif suffisant au modèle.

Dans la population des individus qui ont eu un accident cardiovasculaire les profils ne sont pas nécessairement équiprobables. On peut donc s'intéresser à d'autres techniques d'échantillonnage plus élaborées, qui visent en particulier à créer des observations synthétiques en interpolant à partir de plusieurs voisins, citons par exemple SMOTE, SMOTEENN et ADASYN. [13]

La performance de ces méthodes est souvent limitée par le bruit ou la présence de valeurs extrêmes dans le dataset. En effet, dans le cas de SMOTE, ces valeurs sont

créées à partir des observations, sans garantie d'être des observations plausibles, ce qui peut fausser l'apprentissage du modèle. SMOTEEN offre une solution qui s'affranchit de cette limitation. A la différence de la technique SMOTE qui crée des nouvelles observations par interpolation et introduit parfois du bruit, cette méthode réalise un "nettoyage" des observations synthétiques, par vérification des étiquettes des nouvelles observations par les plus proches voisins. Si elles sont dans le voisinage de données de référence (non artificielles) et que leur label diffère, ces nouvelles observations sont retirées.

Une exploration détaillée des techniques d'échantillonnage sort du cadre de cette étude. Ces techniques sont cependant explorées et décrites dans la partie complémentaire de ce travail à l'adresse suivante: https://github.com/assansanogo/AI_for_pharmacists.

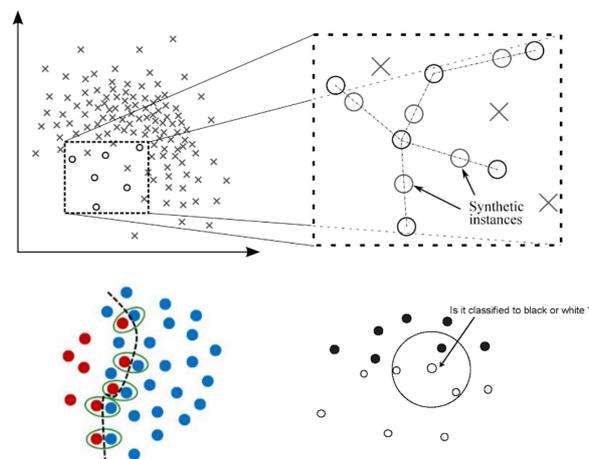


Figure 38: **Technique de sur-échantillonnage synthétique.**

Au dessus: Il s'agit d'interpoler des nouvelles observations à partir d'observations réelles/ initiales. Certaines méthodes fixent le nombre de points à prendre en compte pour définir le voisinage, d'autres incorporent la notion de densité de points.

À gauche: Les points à la frontière pour les 2 classes sont souvent des points complexes.

À droite: Une façon de corriger les incohérences dûes à cette création artificielle de données est l'édition des observations grâce aux plus proches voisins. Si l'étiquette diffère des plus proches voisins, cette observation synthétique est retirée.

6.4. Paramètres et optimisation

La data science est une discipline qui utilise de nombreux cadres applicatifs communément "frameworks". L'objectif d'un "framework" est généralement de simplifier le travail des data scientists, en leur offrant une architecture et des algorithmes "prêts à l'emploi" et qui leur permette de prototyper très vite. XGBoost et la bibliothèque scikit-learn en sont deux exemples. L'implémentation de XGBoost que nous allons utiliser possède de nombreux paramètres qu'il peut être nécessaire d'optimiser. L'exploration détaillée de ces derniers sort du cadre de cette étude. Nous nous limiterons à une description succincte de leur utilisation:

1. **eta** [default=0.3], **gamma** [default=0]

- Ces paramètres influent sur la vitesse d'apprentissage

2. **min_child_weight** [default=1], **max_depth** [default=6], **max_leaf_nodes**, **lambda** [default=1], **alpha** [default=0]

- Ces paramètres permettent de contrôler le sur-apprentissage en agissant sur la forme (profondeur, largeur) des arbres.

3. **max_delta_step** [default=0], **scale_pos_weight** [default=1]

- Ces paramètres sont utilisés dans le cas de dataset déséquilibrés

4. **subsample** [default=1], **colsample_bytree** [default=1], **colsample_bylevel** [default=1]

- Ces paramètres permet de contrôler le sur apprentissage en agissant sur l'échantillonnage des lignes et colonnes (observations, caractéristiques des observations)

L'optimisation des hyperparamètres se fait de manière combinatoire, on parle de l'algorithme GridSearch. Cette optimisation explore un nombre limité de paramètres. La combinaison de paramètres qui optimise la précision de modèle définit le meilleur modèle retenu.

6.5. Résultats et limites

Dans le cas d'un test qui conduit à des résultats

Une courbe ROC (Receiver Operating Characteristic) représente le taux de vrai positifs en fonction du taux de faux positifs.

- La sensibilité de notre modèle est décrite par le taux de vrais positifs.
- La spécificité de notre modèle est décrite par le taux de vrais négatifs.
- 100-Spécificité représente le taux de faux positifs.

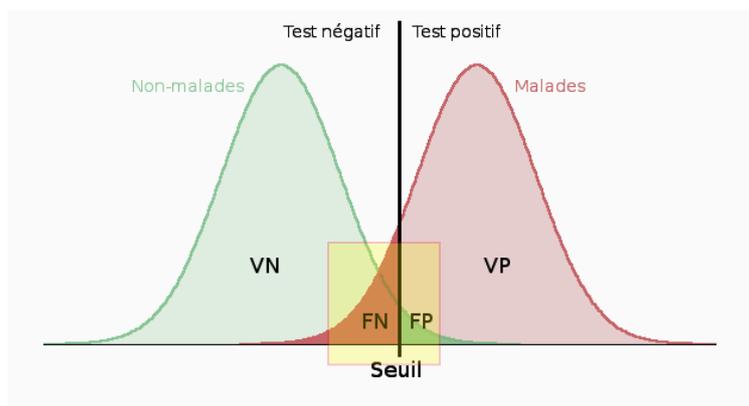


Figure 39: **Sensibilité - Spécificité (exemple).**

La sensibilité est calculée comme : $VP / (VP+FN)$

La spécificité est calculée comme : $VN / (VN+FP)$

$$1\text{-Spécificité} = FP/VN+FP$$

La partie en jaune, schématisent les limites du modèle. Lorsque le seuil utilisé pour prédire si le patient est malade est diminué, le modèle gagne en sensibilité (il inclut plus de faux faux positifs et réduit les faux négatifs) mais perd cependant en spécificité.

La courbe ROC est utilisée pour décrire la performance d'un modèle de classification binaire (2 classes) et représente l'évolution de la sensibilité en fonction de:

100 – Spécificité (%) ou 1 - Spécificité si les valeurs ne sont pas sous forme de pourcentage.

Une courbe optimale aura une aire sous la courbe égale à 1. Pratiquement, elle aura la forme comme ci-dessous c'est à dire que la courbe sera le plus proche possible du coin haut gauche. On dit que l'efficacité est proche de 1.

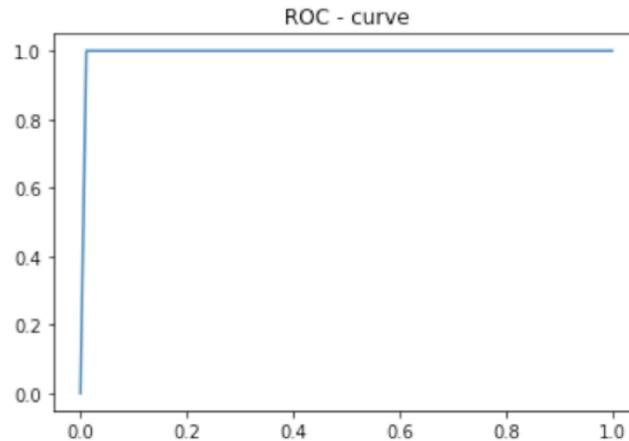


Figure 40: **La fonction d'efficacité du récepteur.**

Plus fréquemment désignée sous le terme « courbe ROC » (de l'anglais *receiver operating characteristic*, pour « caractéristique de fonctionnement du récepteur ») dite aussi *caractéristique de performance (d'un test) ou courbe sensibilité/spécificité*, est une mesure de la performance d'un classificateur binaire, Graphiquement, on représente souvent la mesure ROC sous la forme d'une courbe qui donne le taux de vrais positifs (fraction des positifs qui sont effectivement détectés) en fonction du taux de faux positifs (fraction des négatifs qui sont incorrectement détectés).

Idéalement nous voulons un modèle très sensible (toutes les personnes susceptibles d'avoir une attaque doivent être identifiés) et très spécifique en particulier si les tests de dépistage et traitement sont coûteux. Le modèle retenu présente une sensibilité de 0.989. En d'autres termes, sur un échantillon de 1000 individus qui ont eu un accident cardiovasculaire. 989 auraient pu être identifiés par notre modèle (positifs vrais). 11 individus n'auraient pas pu être identifiés comme à risque (faux négatifs).

Le modèle obtenu est une première étape dans la prévention thérapeutique, mais présente plusieurs limites. Il n'inclue ni la notion de temps et de survie, et ne fait pas de distinction de type de risque entre les patients. On peut imaginer qu'il existe plusieurs niveaux de risque. De ce fait une analyse de survie supplémentaire pourrait dans notre cas être très informative.

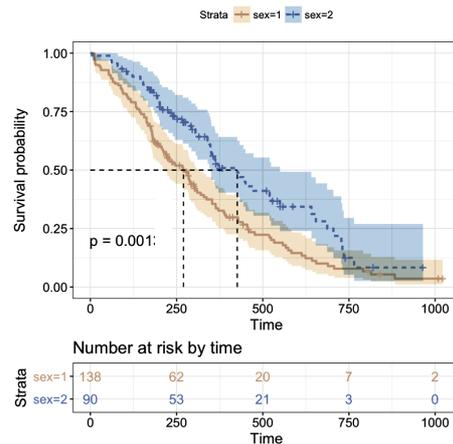


Figure 41: **Exemple de courbe de survie.**

L'analyse de survie cherche à modéliser le temps restant avant la survenue d'un évènement d'intérêt ou le temps restant avant l'échec ou la panne dans les systèmes artificiels, ce que l'on représente graphiquement sous la forme d'une courbe de survie.

Source: <http://www.sthda.com/english/wiki/survival-analysis-basics>

7. L'INTELLIGENCE ARTIFICIELLE POUR LE PHARMACIEN BIOLOGISTE

La vision assistée par ordinateur est une application pertinente pour le pharmacien biologiste, en particulier pour identifier, quantifier les structures d'intérêts. Le pharmacien biologiste est un métier en pleine mutation. D'après [14], le pharmacien biologiste est un expert et son rôle est essentiel pour trier et affiner les règles identifiées par le Machine Learning. « *La combinaison de l'Intelligence Humaine et de l'Intelligence Artificielle est l'approche la plus pertinente pour caractériser finement une pathologie* »

L'atlas protéique humain regroupe un ensemble de clichés pris par microscope confocal. Ce type de cliché est utilisé fréquemment pour la recherche biomédicale. A l'heure actuelle, il existe une grande différence entre la vitesse de production des clichés et leur analyse. L'identification des protéines de manière automatique et fiable et quasi instantanée contribue à l'automatisation des techniques d'imagerie. Ceci permet de faciliter la compréhension des pathologies et d'accélérer la recherche médicale.

7.1. Jeu de données

7.1.1. Description

Le jeu de données est tel qu'il existe 28 différentes étiquettes. Le but de cet algorithme est de prédire l'origine de la structure protéique mise en évidence. Si l'étiquette est "endosome", on doit prédire ce label à partir des 4 images.

Le filtre vert est le filtre qui concentre le plus d'informations par rapport à la structure protéique. Les autres filtres serviront de contexte au modèle.

Ces données sont obtenues à partir d'un microscope confocal. Les cellules humaines présentent une grande variabilité en forme et taille et possèdent de nombreuses structures

cellulaires dont l'aspect varie également. Cette diversité cellulaire rend cette tâche de classification automatique, relativement difficile. Tous les clichés sont organisés en 4 filtres (ici colorisés) de taille 512 pixels*512 pixels. Chaque filtre a pour cible un des 4 éléments constitutifs de la cellule comme décrit ci-dessous:

- filtre bleu (contexte): noyau
- filtre rouge (contexte): microtubules
- filtre jaune (contexte): réticulum endoplasmique
- filtre vert: la structure marquant les protéines

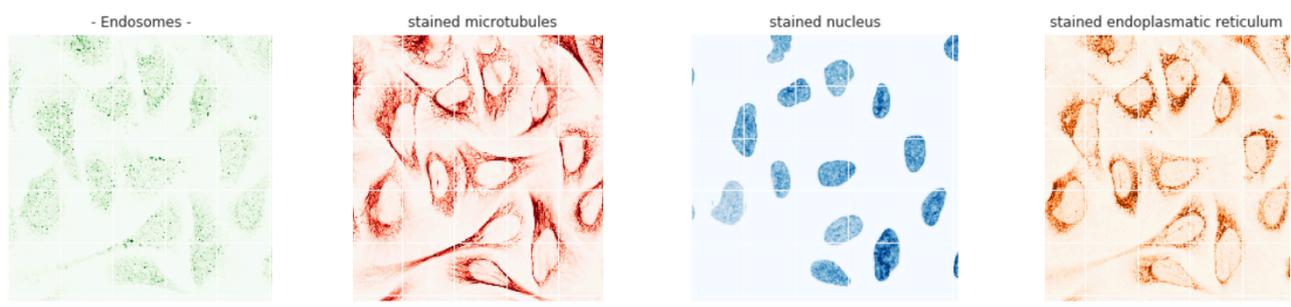


Figure 42: **Illustration des structures biologiques identifiées à travers différents filtres.**

Ici, on identifie sur le filtre vert: la structure protéique marquée (les endosomes). Les autres filtres servent de contexte: le filtre rouge représente toujours les microtubules, le filtre bleu le noyau et en jaune le réticulum endoplasmique.

Le but de cette classification multi-étiquettes est d'identifier la structure protéique d'intérêt. Le filtre vert est donc celui qui contient le plus de poids dans ce problème de classification. Une première approche sera donc de développer un modèle prédictif capable de tirer parti de cette seule source d'information.

Liste des catégories à prédire:

- | | |
|----------------------------------|---------------------------------|
| 0.Nucléoplasme | 15.Terminaisons microtubulaires |
| 1.Membrane cellulaire | 16.Pont cytotinétique |
| 2.Nucléole | 17.Epine mitotique |
| 3.Centre fibrillaire nucléolique | 18.Centre d'organisation |
| 4.Grains nucléaires | Microtubulaire |
| 5.Corps nucléaires | 19.Centrosome |
| 6.Reticulum endoplasmique | 20.Gouttelette lipidique |
| 7.Appareil de Golgi | 21.Membrane plasmique |
| 8.Peroxisomes | 22.Jonction cellulaire |
| 9.Endosomes | 23.Mitochondrie |
| 10.Lysosomes | 24.Aggresome |
| 11.Filaments intermédiaires | 25.Cytosol |
| 12.Filaments d'actine | 26.Corps cytoplasmiques |
| 13.Site d'adhésion focale | 27.Anneaux et cylindres |

$L = \{\lambda_1, \dots, \lambda_k, \dots, \lambda_q\}$ est représenté par la liste des 28 labels:

$D = \{(x_i, y_i) \mid i = 1..|D|\}$ est le jeu de données d'entraînement, c'est à dire les images associées aux étiquettes (annotations).

$S = \{x_i \mid i = 1..|S|\}$ est le jeu de données de test, c'est à dire des images non associées à des étiquettes, qu'il convient de prédire avec le maximum de précision. La précision est calculée en confrontant les prédictions et les étiquettes de test maintenues secrètes pour le modèle prédictif. En revanche, les filtres bleu, jaune et rouge donnent une information sur les autres structures présentes. Préalable à toute analyse de données, il convient d'observer la fréquence relative des étiquettes, d'analyser la présence de données extrêmes ou même d'erreurs lors de de l'étiquetage des données d'entraînement.

7.1.2. Analyse exploratoire des images

D'après le graphe ci-dessous, les clichés sont étiquetés le plus fréquemment avec le nucléoplasme, le cytosol et la membrane cellulaire. Ceci correspond aux éléments cellulaires de ce dataset où l'on retrouve communément des protéines. On peut s'expliquer ce fait car les protéines de traduction/transcription sont retrouvées dans le noyau, les ribosomes se concentrent au niveau du cytosol, et enfin les canaux protéiques au niveau de la membrane cellulaire.

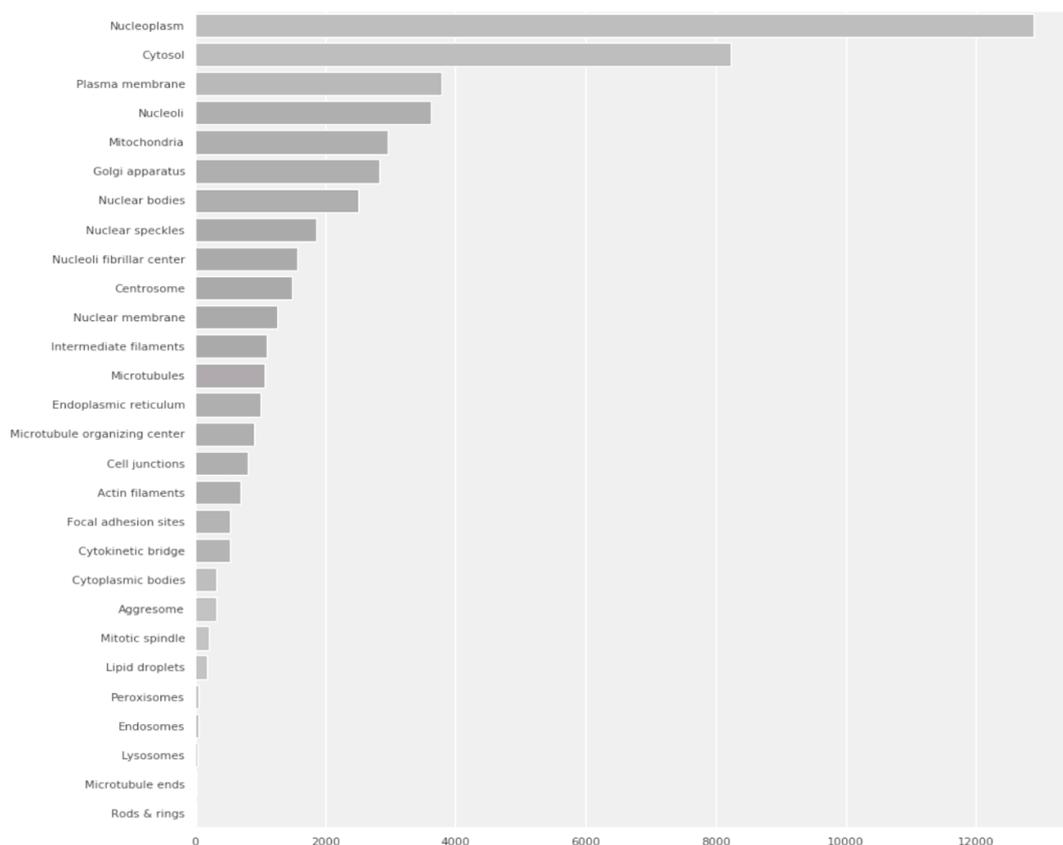


Figure 43: **Déséquilibre de classes.**

Ce déséquilibre entre les différentes étiquettes, ajoute un élément de complexité au problème de classification multi-étiquettes. En effet, sans modification du jeu de données ou sans prise en compte dans l'élaboration du modèle prédictif, le modèle va apprendre à classer les classes les plus fréquentes.

En effet, supposons que le classifieur puisse catégoriser de manière exacte les protéines nucléoplasmiques, cytosoliques et membranaires. L'algorithme pourrait très bien avoir une précision globale proche de 90% alors qu'il serait incapable d'identifier des catégories plus rares comme les enzymes lysosomiques ou encore les anneaux protéiques. Il faut donc tenir compte de ce déséquilibre. Les paramètres de l'entraînement permettent de donner un poids inversement proportionnelle à la fréquence des classes. Une erreur sur une classe sous représentée sera pénalisée de manière beaucoup plus importante que pour une classe sur-représentée.

7.2. Algorithme

7.2.1. Classification multi-classes

La classification des clichés est un problème de classification multi-étiquettes. À la différence d'une classification multi-classes ou chaque échantillon ne peut appartenir qu'à une unique classe. Chaque image peut posséder simultanément plusieurs étiquettes. Prenons l'exemple de deux étiquettes ou classes: **classe 1** (Endosomes) et **classe 2** (Lysosomes). Dans le cadre de la classification multi-étiquettes, les classes ne sont pas mutuellement exclusives. Ainsi, un cliché donné peut être catégorisé simultanément en Endosomes et Lysosomes.

Définition Mathématique du problème:

- Soit F tel que $F = \{f_1, \dots, f_j, \dots, f_m\}$ un espace de m attributs
- Soit L tel que $L = \{\lambda_1, \dots, \lambda_k, \dots, \lambda_q\}$ un espace de q labels
- Soit D tel que $D = \{(x_i, y_i) \mid i = 1..|D|\}$ un jeu de données multi-label.
- Chaque exemple $x_i = \langle x_{1i}, \dots, x_{ji}, \dots, x_{mi} \rangle$ est annoté par un ensemble de labels :

$y_i = \langle y_{1i}, \dots, y_{ki}, \dots, y_{qi} \rangle$ (avec $q \ll n$)

Un classifieur multi-labels h a pour objectif de prédire les labels y_i des exemples non-étiquetés $x_i \in S$ à partir d'un ensemble d'apprentissage $T \subset D$ de taille limitée, où S est un jeu de test de taille relativement petite comparée à celle de T , i.e. $|S| \ll |T|$. Dans le cadre de problèmes liés à l'analyse d'images, les techniques comme l'apprentissage profond ou "deep learning" se sont imposées ces dernières années comme la solution de choix. Dans le cas présent nous utiliserons un réseau de neurones convolutif très profond que nous construirons totalement. Le modèle obtenu est très volumineux et possède environ 58'971'548 paramètres, et assure une très bonne précision. La surparamétrisation reste cependant un problème qui peut conduire au sur-apprentissage. L'optimisation de la structure (réduction du nombre de certaines couches sort du cadre de cette étude mais serait justifiée dans le raffinement final du modèle.

On envisagera une architecture de modèle à 4 entrées, où les 4 filtres sont traités d'abord dans des branches séparées. L'information extraite sera concaténée puis ensuite traitée par un tronc commun (Convolutions, Echantillonnage).

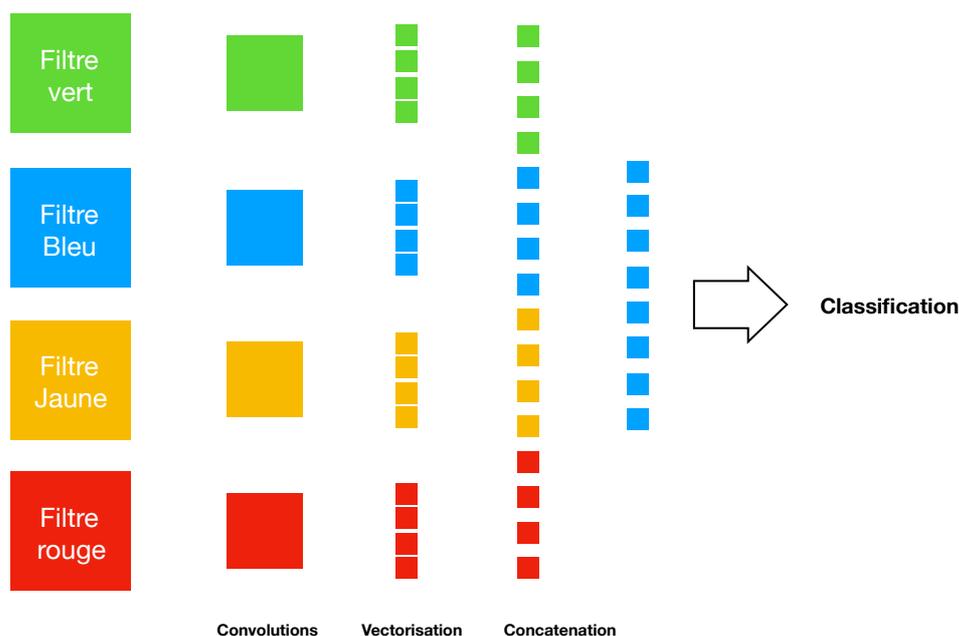


Figure 44: **Structure schématique du réseau de Neurones.**

Chaque filtre va être analysé, via une branche indépendante. Chacune d'entre elles va permettre d'extraire de l'information et de créer des représentations des images d'entrée. Ces représentations des images sont alors jointes et analysées en une branche unique.

7.3. Résultat et limites

La précision obtenue sur les données de validation sans lissage atteint 96%. Cette performance bien que satisfaisante mérite d'être améliorée si ce modèle a pour but d'être utilisé. En effet, le modèle doit être testé sur d'autres jeux de données et devra être ré-entraîné au fil du temps. D'abord parce que plus le modèle apprend sur des données différentes, plus il devient performant et également car il doit pouvoir s'adapter à l'évolution des données au cours du temps (qualité des images microscopiques qui chute avec l'utilisation du matériel).



Figure 45: Courbes de Précision.

Lors de l'entraînement, on obtient 2 graphes: la courbe de gauche donne l'évolution de la précision pour les données d'entraînement, la courbe de gauche donne l'évolution de la précision du modèle pour les données de validation. Les courbes ont été lissées.

La structure du modèle peut et doit être optimisée, et c'est tout le travail du Data scientist. L'optimisation du modèle va dépendre du cas d'utilisation: réduction du temps de calcul pour une analyse en temps réel ou maximisation de la précision. Dans la pratique il faut très souvent faire un compromis entre puissance de calcul et précision.

Par exemple, la représentation de l'information avant concaténation, est paramétrable. La taille des filtres et leur nombre sont autant d'hyperparamètres sur lesquels nous pouvons agir dans le futur pour affiner notre modèle.

Par ailleurs, Nous avons fait l'hypothèse que chaque filtre possédait la même quantité d'information. Or, il apparait que le filtre vert permet à lui seul de faire une classification convenable. Dans le cas des endosomes, les protéines sont identifiées comme des points sur filtre vert. Nous pouvons imaginer un modèle à 2 branches où les filtres de contexte sont analysés dans une branche, et le filtre "protéique" traité à part. Une autre approche possible serait de considérer comme image d'entrée un tenseur composé des 4 images. Chaque observation est un tenseur de dimension 3 qui a donc pour taille: 512*512*4. Les

4 images sont alors considérées comme un unique volume sur lequel les convolutions seraient des opérations volumiques.

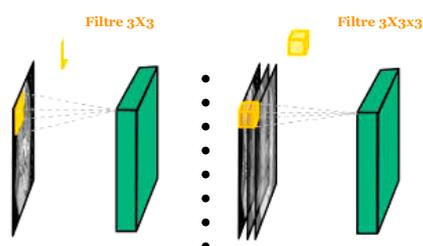


Figure 46: **Plusieurs types de Convolutions.**

Les opérations de convolution peuvent être surfaciques ou volumiques. Une convolution volumique va intégrer les relations spatiales des pixels dans un volume donné, soit le volume défini par les 4 filtres (bleu, rouge, vert et jaune.)

Enfin, le modèle analyse des images dont la taille est de 512 pixels par 512 pixels. Cette taille relativement importante impose une limitation sur la taille du batch utilisé pour l'entraînement. En effet, ces opérations sont contraintes par la puissance de calcul de la carte GPU. Malgré un entraînement réalisé sur une station Nvidia DGX - référence pour la recherche appliquée à l'intelligence artificielle, un compromis est à trouver entre la taille des batchs et celle des images à analyser. Dans ce cas, nous pouvons imaginer un scénario où les images sont réduites de taille et celle du batch augmenté.



Figure 47: **Nvidia DGX station: la référence pour la recherche en IA.**

Cette station permet d'accélérer les entraînements de manière significative. Le prototypage passe de l'ordre de quelque jours à quelques heures. Source: <https://www.bocx.com/systems/deep-learning/dgx-station>

7.4. Conclusion

L'utilisation de cette technique s'inscrit dans l'amélioration du criblage à haut débit. Cette automatisation des tests biologiques requiert des méthodes robustes et rapides.

La classification via les réseaux de neurones permet de localiser des structures cellulaires et accélérer le développement de nouveaux traitements. Par sa simplicité, elle peut être intégrée à des outils d'acquisition d'images ou être même faire partie d'extensions de solutions préexistantes comme la suite logicielle ImageJ.

Afin d'accélérer les découvertes de nouvelles molécules, tout en limitant les risques liés à certains produits ou organismes potentiellement dangereux, des chercheurs ont développé des techniques de criblage virtuel. A l'aide de modèles mathématiques génératifs tels que les réseaux de convolution génératifs antagonistes (GAN), les chercheurs sont capables de générer des nouveaux principes actifs ou protéines qui ne peuvent pas être synthétisées naturellement.

Grâce à des simulations virtuelles d'effets thérapeutiques pour des cibles données sont étudiées. Ceci a également pour but de réduire l'expérimentation animale. En revanche, ces simulations ne peuvent pas complètement garantir l'absence d'effets indésirables sur un organisme.

8. L'I.A POUR LE PHARMACIEN EN PAYS EMERGENTS

8.1. Jeu de données

8.1.1. Description

Le dataset "Malaria" proposé par le NIH a été développé pour les microbiologistes ou pharmaciens dans les régions connues pour leurs ressources limitées, les chercheurs du Lister Hill National Center ont développé une application mobile qui peut être utilisée sur un téléphone mobile conventionnel. Des lames de 151 patients sains et 50 patients infectés préparées à la coloration Giemsa permettent d'identifier une infection à falciparum transmis par les moustiques. Les clichés ont été réalisés au "Chittagong Medical College Hospital" situé au Bangladesh. Les images ont été annotées manuellement par un expert de l'unité de recherche en médecine tropicale de Mahidol-Oxford situé à Bangkok en Thaïlande.

Les hématies ont été déjà pré-segmentées, de telle sorte que chaque image contient une unique hématie. En sus d'un ratio de l'image (hauteur/largeur) très variable, on remarque une variation dans la luminosité significative des images capturées.

En effet, Les images capturées par le téléphone qui correspondent par exemple au patient-ID "C47P8thinOriginal" sont issues de deux modèles de microscopes (Olympus and Motif).

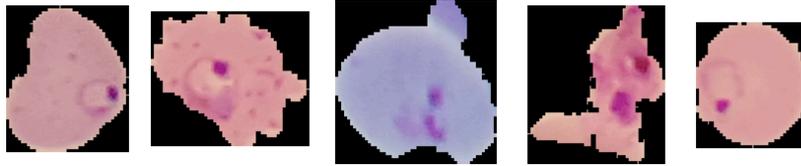


Figure 48: **Coloration Giemsa d'hématies de patients infectés par falciparum.**

Le jeu de données apparaît avec la publication Rajaraman S.(2018) Pre-trained convolutional neural networks as feature extractors toward improved Malaria parasite detection in thin blood smear images PeerJ6:e4568 <https://doi.org/10.7717/peerj.4568>

L'augmentation des données doit être choisie de manière réaliste ou du moins qui simule la variété possible des clichés que l'on peut rencontrer. Les trophozoïtes ont des positions variées au sein des hématies, on peut donc utiliser des opérations de rotation, translation. Le jeu de données n'est pas déséquilibré et par conséquent ne nécessite pas de traitement particulier (i.e: sous échantillonnage ou sur-échantillonnage).

8.2. Méthodologie

8.2.1. Réseau neuronal convolutif très profond

Les réseaux neuronaux utilisés sont considérés comme la référence actuelle en termes d'algorithme de classification basés sur le Deep Learning. Nous utiliserons un modèle très profond appelé ResNet. Ce modèle possède des connexions résiduelles qui permettent de conserver l'information au sein du réseau de neurones. Ceci permet en effet de réduire l'extinction des gradients. Comme décrit précédemment, les modèles appellent par rétro-propagation des gradients. Cependant, l'entraînement de réseaux de neurones très profonds est plus complexe car des portions du gradients doivent être rétro-propagées autant de fois qu'il ya de couches. Ainsi le produit de termes tels que $l_{grad} \cdot J_l < 1$, va

converger vers zero. C'est ce qu'on appelle extinction du gradient. Avec elle, le réseau cesse d'apprendre.

Comme illustré dans la figure ci-dessous: Dans les blocs constitutifs de ResNet, on note 2 éléments distinctifs: un bloc convolutif composé de 3 étapes: expansion, filtration et compression et une connection résiduelle qui empêche l'extinction du gradient.

Dans ce bloc on note la présence d'opérations particulières et une fonction de transfert non linéaire qui limite l'explosion du gradient:

- **BN:** La Normalization de Batch permet de normaliser les activations de la couche activée. Cette opération permet au modèle d'apprendre plus rapidement. En effet, les données étant traitées par batch, on peut observer une grande variabilité d'une observation à une autre au sein d'un même batch.
- **ReLU:** La fonction d'activation non linéaire est utilisée afin de faciliter la rétropropagation du gradient.

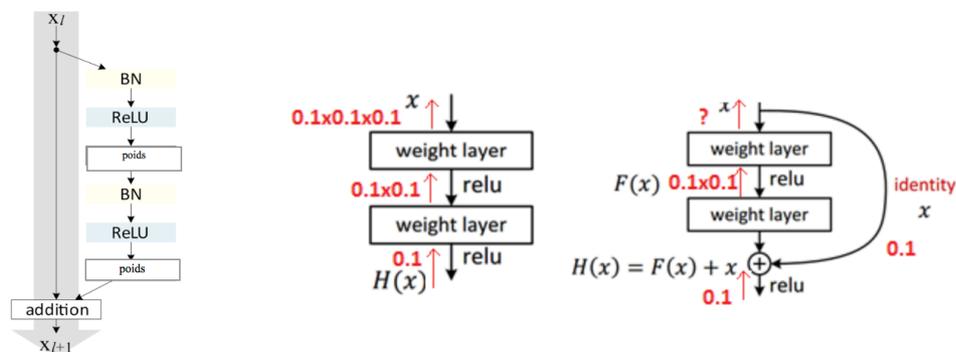


Figure 49: **Bloc caractéristique de ResNet.**

Le Bloc constitutif de ResNet (9) comprend deux branches: la suite de convolutions et la connection "résiduelle". Les informations issues de ces deux branches sont ensuite ajoutées. Ainsi, si le gradient par couche était de 0.1 : dans le cas classique gradient rétropropagé vaudrait 0.1³ alors que dans ResNet50 cette valeur vaut : 0.1 + 0.1³. On évite alors l'évanouissement du gradient en intercalant des blocs résiduels.

Source: <https://arxiv.org/pdf/1603.05027.pdf>

8.3. Résultats et limites

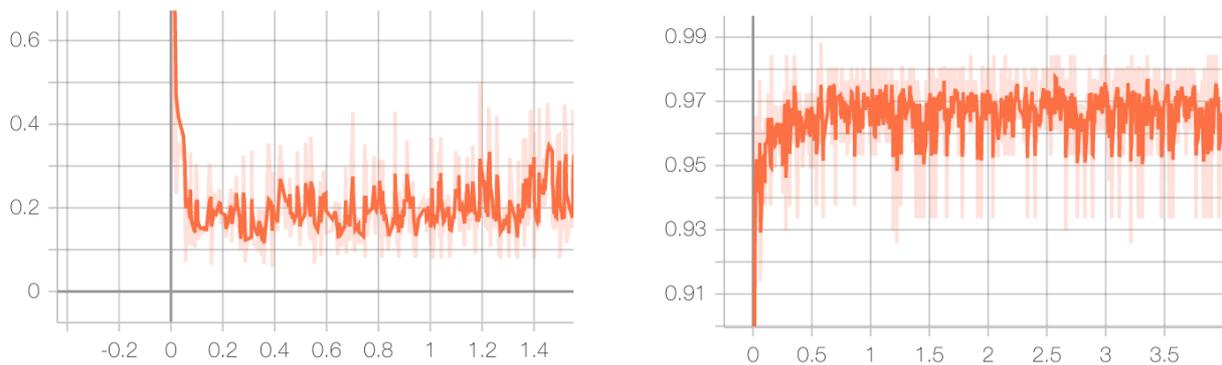


Figure 50: Courbes représentatives (fonction objectif et précision).

À gauche, l'évolution de la fonction objectif avec le temps montre une décroissance vers un plateau. À droite, l'évolution de la courbe montre une croissance vers un plateau. Les courbes sont lissées pour une meilleure lisibilité.

Le modèle entraîné, ResNet50 possède environ 25 Millions de paramètres. Ce modèle a été pré-entraîné sur un dataset très volumineux (>1 million d'images), puis a été affiné à notre cas d'étude. La précision maximale atteinte est de 98.5% (sans lissage). Les versions plus profondes de ce modèle peuvent permettre une meilleure classification au risque d'être prompt au sur-apprentissage. Cette expérimentation sort du cadre de cette étude, mais sera envisagée dans le cadre d'une mise en production. Nous ajouterons, que le modèle entraîné sur ce dataset ou une seule cellule étaient visible nécessite une modification pour être efficace sur des clichés avec de nombreuses hématies. Un pré-traitement de localisation des hématies suivi d'1 classification sera nécessaire. Un entraînement d'un détecteur d'hématies parasitées (algorithme inspiré de YOLO v.3 ou RetinaNet) réalisant en 1 passe: localisation et classification sur des des clichés concentrant plusieurs types d'hématies (infectées et non infectées) est aussi concevable.

L'intérêt pour le pharmacien ayant des ressources limitées est la simplicité d'usage d'une telle solution, notamment grâce à l'usage ubiquitaire des téléphones mobiles. Cet algorithme a été développé pour être porté sur android ou iOS. Cette solution peut aider à terme à désengorger les hopitaux et faciliter la prise en charge. Les techniques d'intelligence artificielle apparaissent comme une solution pour les pays en voie de développement pour réaliser leur révolution technologique. Le calcul haute performance s'est aujourd'hui démocratisé (1) et l'utilisation des GPU n'est plus limité aux seuls centres de recherche.

En effet, certes l'entraînement de ce modèle a été réalisé sur la plateforme de recherche DGX, cependant l'utilisation de la puissance de calcul d'Amazon ne dépasse pas 10 \$/h et permet le développement de prototypes facilement déclinables en solutions par exemple là où la présence d'équipements de santé est limitée. Il existe même une structure en ligne initiée par Google qui permet l'utilisation gratuite de GPU. Les temps d'entraînement seront cependant plus longs qu'avec une station de recherche.

On note le développement de solutions qui détectent automatiquement les problèmes hématologiques comme la drépanocytose, ou comme illustré ci-dessus des systèmes détectant des hématies infectées par les trophozoites. Cette modernisation doit s'accompagner d'un arsenal thérapeutique plus accessible pour les populations à revenu limité. C'est dans ce cadre que les initiatives du DNDI [16] avec la production hors brevets de 2 formules thérapeutiques: "ASAQ": Artésunate + Amiodaquine et "ASMQ": Artésunate + Mefloquine, peuvent contribuer à une évolution durable et favorable face à des maladies qui restent encore très meurtrières.



Figure 51: Utilisation d'une solution de Deep Learning en Afrique.

Le modèle développé ci-dessus permet d'identifier plasmodium sur des échantillons qui sont de petite taille. Ces projets fonctionnent sur des images de haute résolution (contrairement à l'approche ci-dessus). À gauche et au centre: Grâce à un téléphone, la détection de pathologie telle que la malaria devient possible et accessible à tous. Une autre approche classifie et localise les cellules infectées/parasitées.

Source: <http://air.ug/microscopy/>

8.4. Conclusion

Les applications incorporant les réseaux convolutifs se multiplient dans le domaine de l'e-santé. L'algorithme de classification est destiné aux professions de santé et se veut contribuer à l'aide au dépistage des patients. En fixant une valeur seuil suffisamment haute, ce modèle va minimiser les faux négatifs (diagnostiquer les patients malades comme sains) quitte à augmenter légèrement le taux de faux positifs (diagnostiquer les patients sains comme malades).

Cette solution peut aider à désengorger les hôpitaux comme outil de première prise en charge des patients à risques (enfants, personnes âgées et patients immuno-déprimés) et accélérer le traitement des patients réellement affectés par le paludisme. Par sa simplicité elle peut être proposée aux dispensaires et étendue aux zones isolées sans réelle

structure de soins. Elle s'inscrit également dans le combat mené par les ONG à savoir lutter contre les maladies endémiques et sensibiliser l'ensemble de la population sur certains thèmes de santé. Enfin, la miniaturisation des capteurs et leur coût décroissant rend possible le développement de stations mobiles de traitement opérables par un pharmacien.

9. CONCLUSION

L'intelligence artificielle suscite aujourd'hui des inquiétudes ou le rejet de certains et une fascination pour d'autres. Le RGPD est une tentative pour accompagner ces projets de manière réfléchie en Europe. Si certains imaginent un serment d'Hippocrate technologique, il est certain que l'IA doit posséder un cadre éthique stricte pour éviter toute dérive ou mauvaise utilisation des données.

L'intelligence artificielle est capitale car elle permet d'automatiser des compétences cognitives de manière plus efficace et à coût diminué. Ceci représente de nouvelles opportunités dans les domaines où la collecte et l'analyse de données est très importante, en particulier: le domaine de la santé (diagnostic, découverte de nouvelles molécules et suivi de patient). Elle permet également aux populations aux ressources limitées l'accès aux soins grâce à son faible coût lorsqu'employée sur les appareils mobiles [17]

Au delà de ses succès, l'intelligence artificielle peut souffrir de plusieurs limitations. Commençons d'abord par l'aspect spéculatif de son utilisation: Deux cinquièmes des start-up européennes d'intelligence artificielle n'utilisent en fait aucune technologie assimilable à de l'intelligence artificielle, d'après le rapport «The State of Ai» (l'état de l'IA) [18]. Il y a donc un détournement de la technologie pour stimuler l'attrait des startups et faciliter l'investissement[20].

Ensuite, Watson qui a révolutionné le monde de l'intelligence artificielle, peine à transformer les promesses tenues par Watson Health quant à sa capacité à assister les médecins dans le diagnostic et traitement de cancers. Au delà de ces difficultés techniques, s'ajoute le problème de biais dans les données utilisées pour entraîner Watson. Finalement son besoin colossal en données ont rendu son utilisation face à une stratégie d'IBM inadaptée en fait un succès en demi-teinte.

D'autre part, avec le rapport Villani [19] et partie intégrante du plan « Ma Santé 2022 », présenté par Emmanuel Macron le 18 septembre dernier, nous avons assisté à la création du Health Data Hub. A terme, ce guichet unique et sécurisé devra notamment permettre de favoriser les croisements (possibles) entre les données médico-administratives et les données médicales et épidémiologiques existantes. Malgré cette initiative et la notion d'open data grandissante, l'accès à la donnée de santé reste aujourd'hui très difficile en particulier pour une startup. L'acquisition des données va certes se simplifier, mais risque tout de même de nécessiter plusieurs mois. Cette lourdeur administrative reste un frein pour l'éducation et l'avènement des "pharmaciens 4.0".

Enfin, nous assistons à une révolution de la pharmacie en particulier: les associations entre les GAFAs et les grands groupes pharmaceutiques tels que Sanofi ou Novartis. Les industriels comptent donc sur cette intégration d'expertises pour permettre le développement de nouveaux médicaments et de nouvelles solutions de santé connectée toujours plus intelligentes grâce notamment au Big data et au Machine Learning.[21]

Doyen de L'UFR de Pharmacie,
Brigitte Vennat

Le président de Jury,
Jean-Michel CARDOT

10. BIBLIOGRAPHIE

1. CNIL, RGPD de quoi parle t-on ? [Internet] 2019. Disponible sur: <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>
2. DeepMind, Towards integration of deep learning and neurosciences. [Internet] 2016. Disponible sur : <https://deepmind.com/research/publications/towards-integration-deep-learning-and-neuroscience>
3. Wikipédia, Algorithme du gradient [Internet] 2019. Disponible sur: https://fr.wikipedia.org/wiki/Algorithme_du_gradient
4. Sanofi, La digitalisation de l'officine. [Internet] 2019. Disponible sur : <https://www.sanofi.fr/fr/labsante/eclairage-sur-la-digitalisation-de-l-officine>
5. Le Parisien, Paracétamol: l'ANSM alerte sur les risques de surdosages. [Internet] 2019. Disponible sur : <http://www.leparisien.fr/societe/sante/paracetamol-l-an-sm-alerte-sur-les-risques-de-surdosage-09-07-2019-8112603.php>
6. Les Echos, Alerte sur les dangers de l'ibuprofène et du kétoprofène. [Internet] 2019. Disponible sur : <https://www.lesechos.fr/industrie-services/pharmacie-sante/alerte-sur-les-dangers-de-libuprofene-et-du-ketoprofene-1013002>
7. Viola, P. Rapid Object Detection using a Boosted Cascade of Simple Features. [Internet] 2001. Disponible sur: <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf>

8. Cen, K. (n.d.). Study of Viola-Jones Real Time Face Detector. [Internet] 2016. Disponible sur: https://web.stanford.edu/class/cs231a/prev_projects_2016/cs231a_final_report.pdf

9. Jean-Emmanuel Bibault. Comment réguler l'intelligence artificielle. [Internet] 2019. Disponible sur : <http://sante.lefigaro.fr/article/comment-reguler-l-intelligence-artificielle-en-medecine/>

11. Redmon, J., & Farhadi, A. YOLO9000: Better, Faster, Stronger. [Internet] 2016. Disponible sur : <https://doi.org/10.1109/CVPR.2017.690>

12. Vikashraj Iuhaniwal. "Why gradient descent is not enough, a comprehensive introduction to optimisation algorithm [Internet] 2015. Disponible sur : "<https://towardsdatascience.com/why-gradient-descent-isnt-enough-a-comprehensive-introduction-to-optimization-algorithms-in-59670fd5c096>

13. Babar, V. S., & Ade, RA Review on Imbalanced Learning Methods. [Internet] 2015. Disponible sur : <https://research.ijcaonline.org/ncac2015/number2/ncac175029.pdf>

14. SIL-LAB. IA et Biologie Médicale cas réel. [Internet] 2019. Disponible sur : <http://www.sil-lab-experts.com/images/SIL-LAB/actualites/IA-et-Biologie-Medicale-Cas-Reel-V4.pdf>

15. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. Going Deeper with Convolutions. ArXiv:1409.4842. [Internet]. 2014. Disponible sur : <https://doi.org/10.1109/CVPR.2015.7298594>
16. DNDI . DNDI Achievements. [Internet] 2019. Disponible sur : <https://www.dndi.org/achievements/>
17. Bernard Grainier. L'Afrique face à la révolution technologique: un potentiel de développement formidable! [Internet] 2017. Disponible sur : <https://www.linkedin.com/pulse/lafrique-face-%C3%A0-la-r%C3%A9volution-technologique-un-de-bernard-gainnier/?originalSubdomain=fr>
18. MMC Ventures, The state of AI: Divergence [Internet] 2019. [Internet] Disponible sur : <https://www.mmcentures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf>
19. Cédric Villani. Rapport Villani. [Internet] 2018. Disponible sur : https://aiforhumanity.fr/pdfs9782111457089_Rapport_Villani_accessible.pdf
20. Elisa Braun, 40% des start-up européennes d'intelligence artificielle n'utilisent pas d'intelligence artificielle.[Internet] 2019. Disponible sur : <http://www.lefigaro.fr/secteur/high-tech/2019/03/06/32001-20190306ARTFIG00121-40-des-start-up-europeennes-d-intelligence-artificielle-n-utilisent-pas-d-intelligence-artificielle.php>
21. Les échos. Comment la pharmacie s'approprie le big Data et l'intelligence artificielle, [Internet] 2019. Disponible sur : <https://www.lesechos.fr/industrie-services/pharmacie-sante/comment-la-pharmacie-sapproprie-le-big-data-et-lintelligence-artificielle-1003161>

RESUME

Le rôle du pharmacien est traditionnellement multiple: Il développe, fabrique et garantit la qualité des médicaments. En outre, il analyse les prescriptions et délivre les médicaments. Il délivre des conseils sur les médicaments (leur action, risques et interactions), il participe aussi à des actions ponctuelles de prévention et suit les patients atteints de maladies chroniques (patients diabétiques, sous anticoagulants oraux par voie directe, ou asthmatiques) et/ou poly-médiqués. Il peut également réaliser des prélèvements, effectuer des analyses et commenter les résultats de biologie médicale. L'intelligence artificielle ouvre de nouvelles possibilités d'exercice au pharmacien et peut contribuer à une redéfinition complète de son rôle. En effet, le pharmacien est en train de devenir un intermédiaire de santé polyvalent, garantissant une santé de proximité.

Ce travail, illustre comment cette révolution peut s'appliquer au pharmacien d'officine au travers d'un prototype minimaliste de reconnaissance d'interaction médicamenteuse. Cet algorithme peut facilement s'étendre à plusieurs classes de médicaments, inclure la reconnaissance vocale et ou être jumelé à un assistant conversationnel.

L'intelligence artificielle peut apporter au pharmacien biologiste de nouveaux outils, par exemple l'automatisation de la classification de clichés issus de l'atlas protéique. Enfin, cette thèse inclut le pharmacien situé en milieu défavorisé qui peut, lui aussi, bénéficier de cette révolution technologique. Cette réalité est illustrée via le développement d'un algorithme léger qui peut détecter les cas de malaria et qui peut facilement être embarqué sur un pico ordinateur de type Nvidia NANO dont le coût ne dépasse pas une centaine d'euros.

Mots-clés:

PROTOTYPE
PREVENTION
CLASSIFICATION
DETECTION

MACHINE LEARNING
DEEP LEARNING
EDUCATION
INTELLIGENCE ARTIFICIELLE