

Le Deep Learning pour la prédiction du Cancer et des réponses à son traitement

Mémoire présenté par

Mathieu RIDET

Pour l'obtention du Master **1** MIAGE

De l'université

Paris 1 Panthéon - Sorbonne

Année Universitaire : **2020 - 2021**

Date de soutenance : **03 septembre 2021**

Directeur de mémoire : **Rebecca DENECKERE**

Membre du jury : **Irina RYCHKOVA**

Remerciements

Tout d'abord, je voudrais remercier ma tutrice enseignante madame DENECKERE Rebecca pour l'aide qu'elle m'a apportée tout au long de l'année que ce soit pour la réflexion, la conception ou la rédaction de ce mémoire de recherche à l'état de l'art.

Je souhaiterais également remercier toute l'équipe pédagogique de l'université Paris 1 de m'avoir apporté autant de connaissances en cette année de Master 1.

Aussi, j'exprime toute ma reconnaissance et ma sympathie à l'ensemble des membres du service TSI Habilitations du groupe Covéa qui me permettent d'évoluer professionnellement mais aussi sur le plan personnel depuis deux ans maintenant. Plus précisément, je remercie mon équipe MOE RSA qui participe grandement à mon évolution à tous les niveaux : technique, fonctionnel et relationnel.

Je tiens à remercier tout particulièrement mon manager TORRE Eric qui me suit quotidiennement avec bienveillance depuis le début de mon alternance mais aussi mon maître d'apprentissage DANG Liem qui est l'acteur le plus impliqué dans mon processus d'apprentissage et m'apporte toute l'aide et l'assistance dont j'ai besoin.

Table des matières

| | |
|---|----|
| Remerciements | 1 |
| Résumé | 3 |
| Abstract | 3 |
| Contexte | 4 |
| Méthodologie de recherche de la littérature | 5 |
| Introduction | 6 |
| I. Détecter en amont la maladie | 7 |
| a. Le Machine Learning pour la prédiction du cancer | 7 |
| b. Le Deep Learning pour pallier certains manques | 12 |
| c. L'association du Machine Learning et du Deep Learning | 16 |
| II. Prédire la réponse aux traitements | 20 |
| a. A l'aide du Deep Learning et de jeux de données d'expressions génétiques | 20 |
| b. A l'aide du Deep Learning pour l'analyse d'imageries médicales | 23 |
| Conclusion | 27 |
| Annexes | 29 |
| Table des Figures | 29 |
| Glossaire | 29 |
| Références | 30 |

Résumé

L'intelligence Artificielle (IA) se développe depuis plusieurs décennies maintenant. Son utilisation, de plus en plus fréquente dans de plus en plus de domaines, se diversifie car les industries profitent des avancées technologiques incessantes. C'est le cas de l'industrie médicale qui tend à utiliser l'IA afin de faciliter certaines tâches ou même pour en effectuer d'autres qui sont, jusqu'à présent, irréalisables.

En effet, actuellement certaines maladies demeurent incurables comme c'est le cas du cancer qui, détecté bien souvent à un stade déjà avancé, ne laisse pas grande chance aux patients. Ainsi, tant qu'il n'existe pas de traitement efficace contre cette maladie, l'objectif est double : trouver un moyen de détecter la maladie le plus tôt possible afin d'augmenter les chances de survie mais aussi déterminer un traitement capable de soigner le cancer. Pour cela, il faut pouvoir tester un nombre infini de combinaisons de médicaments et d'interventions médicales afin d'analyser les réponses cellulaires et immunitaires en fonction des différents types de patients, de cellules et de médicaments.

Le Deep Learning (DL), branche de l'IA composée de plusieurs méthodes permettant un apprentissage autonome efficace grâce à tout type et format de données, se présente comme un atout potentiel majeur dans cet objectif de prédiction. Ainsi, dans quelles mesures le Deep Learning pourrait permettre de détecter l'apparition du cancer et de prédire la réponse au traitement de celui-ci ?

Abstract

Artificial Intelligence (AI) has been developing exponentially since many decades. Its increasingly frequent use in many business areas is diversifying because industries take advantage of unceasing technological advances. Medical industry tries to include more and more AI tools to facilitate or even proceed unfeasible tasks.

At this moment, some diseases are still incurable such as cancer which, often detected at an advanced stage, is very murderous. Therefore, as long as it doesn't exist effective treatment against this disease, there is a double purpose: find a way to detect the disease earlier in order to increase survival chances and find a treatment to heal cancer. To do it, we have to be able to test an infinite number of drugs and medical interventions combinations to analyze cell responses in function of patient, cell or drug types.

Deep Learning, which is an AI branch composed of many methods allowing unsupervised learning from all data types and formats, could be a potential major asset in this prediction purpose. Thus, in which measures could DL detect cancer appearance and also predict treatment response?

Mots clés

cancer, treatment, prediction, machine learning, deep learning

Contexte

Le Cancer, maladie génétique causée par un dysfonctionnement de la division cellulaire, est une des principales causes de mortalité depuis plusieurs dizaines d'années. A la fin du vingtième siècle, 1 personne sur 3 est atteinte du Cancer et, pire encore, 1 homme sur 4 et 1 femme sur 5 en meurent [1]. En 2015, il est à l'origine d'environ 8.8 millions de décès dans le monde [13].

Détecté à un stade avancé, il réduit considérablement l'espérance de vie à moyen et long terme : moins de 50% des patients vivent plus de 5 ans après la détection de la maladie [54].

Ainsi, le cancer est un problème sanitaire majeur qui semble faire l'objet de nombreuses recherches afin de trouver une solution permettant de réduire cette forte mortalité, notamment avec l'avancée de la technologie et l'essor de l'IA.

Le Machine Learning (ML), branche de l'IA dont l'objectif est de prédire un résultat grâce à un processus d'apprentissage [4], est très utilisé depuis plusieurs décennies pour les recherches concernant le diagnostic du cancer [4].

L'algorithme de ML utilise le concept d'inférence : à partir de données récupérées en entrée, il s'entraîne pour être capable de fournir un résultat en sortie [41 - 43].

Il existe trois principales méthodes de ML : l'apprentissage supervisé (SL), non supervisé (NSL) et semi-supervisé (SSL) [4]. Dans le premier cas, les données fournies en entrée doivent être "étiquetées" (*labeled data*). En d'autres termes, elles doivent suivre un *pattern* connu par l'algorithme afin d'être traitées [4]. Dans le cas de l'apprentissage non supervisé, il est laissé au modèle la responsabilité de reconnaître un *pattern* dans le jeu de données qu'il reçoit de façon autonome ; cela permet d'éviter cette tâche de *preprocessing* avant le traitement [4]. Enfin, l'apprentissage semi-supervisé a une valeur ajoutée par rapport aux 2 précédents car il admet, en entrée, n'importe quel type de données : il sait gérer les données suivant – ou pas – un *pattern* précis [4].

Le ML contient des sous-classes qui peuvent s'avérer plus adaptées à certaines situations pour effectuer certaines actions [27].

C'est le cas du Deep Learning (DL) qui se distingue du ML par sa capacité à gérer de grandes quantités de données à multiples dimensions ainsi que des données « bruyantes » [27]. Également, les méthodes du DL sont capables de détecter efficacement des relations non linéaires et/ou complexes entre différentes données [27].

De manière générale, les méthodes spécifiques au DL sont souvent plus efficaces que les algorithmes de ML et ces méthodes ont beaucoup été utilisées pour effectuer des prédictions [27].

Méthodologie de recherche de la littérature

J'ai principalement utilisé le moteur Google Scholar pour rechercher les articles nécessaires à la confection de mon état de l'art.

Traitant un thème évoluant énormément actuellement, j'ai décidé d'appliquer un filtre sur la date de publication afin de ne retenir que les articles les plus récents possible. J'ai également dû appliquer un filtre sur le type d'article/de revue afin de sélectionner les articles parus dans des revues informatiques et non dans des revues médicales.

Aussi, le choix des mots clés a été une étape cruciale car je me suis rendu compte qu'il est très difficile de trouver ce que l'on recherche sans renseigner les bons mots. Pour cette raison, cette étape de recherche m'a posé quelques soucis et m'a pris pas mal de temps. Voici les mots clés et expressions utilisées: "cancer", "machine learning prediction cancer", "machine learning cancer", "deep learning prediction cancer", "deep learning cancer", "cancer prediction", "cancer treatment prediction", "deep learning treatment prediction".

Finalement, 12 articles ont été retenus comme base de rédaction de mon mémoire dont un qui a servi exclusivement pour l'introduction avec des informations relatives au cancer et ses caractéristiques.

Introduction

Depuis que les Hommes réussissent à contrôler les maladies infectieuses - principales causes de mortalité dans le passé -, le nombre de personnes atteintes de cancer ou qui risque d'être touchée par cette mortalité a fortement augmenté [1]. Ainsi, il semble de plus en plus important de trouver une solution pour lutter contre ce fléau sanitaire et ainsi sauver des millions de vies.

Pour cela, l'Homme mise de plus en plus sur les avancées technologiques pour trouver de nouvelles méthodes efficaces dans cette lutte [4]. C'est dans cette idée qu'est beaucoup utilisé l'IA dans le secteur médical depuis quelques années [4].

Dans ce contexte, dans quelles mesures le Deep Learning pourrait permettre de détecter l'apparition précoce d'un cancer et de prédire la réponse au traitement de celui-ci ?

La détection du Cancer à un stade peu avancé ou avant même son apparition serait une évolution révolutionnaire qui permettrait la diminution de son impact meurtrier via un rôle à la fois préventif mais aussi en proposant un traitement adapté en fonction des besoins du patient.

D'autre part, pour tenter de réduire les effets secondaires et déterminer l'efficacité d'un traitement, le DL permettrait de savoir ce que serait la réaction du système immunitaire du patient avant même de le lui administrer.

I. Détecter en amont la maladie

a. Le Machine Learning pour la prédiction du cancer

Depuis plusieurs décennies, la recherche scientifique concernant le traitement du cancer s'est intensifiée notamment grâce aux voies technologiques et à des méthodes comme l'analyse d'images via le ML [13]. Aussi, l'utilisation de certains outils a été possible grâce à l'augmentation considérable du nombre de données relatives à la maladie [16]. C'est évidemment le cas des algorithmes de ML qui jouent un rôle de plus en plus important quant à la prédiction d'un probable développement de la maladie chez différents patients [4]. En effet, le taux de précision de la prédiction du cancer a augmenté de 15 à 20% sur les dernières années grâce, notamment, à l'utilisation du ML [4]. Ainsi, de nombreux chercheurs à travers le monde se sont intéressés à la prédiction du cancer grâce au ML [4].

L'idée de toutes les études qui nous intéressent ici est de pouvoir prédire le cancer. Certaines [2, 3] se sont concentrées sur cette idée générale tandis qu'une autre [4] s'est penchée plus en détails sur les façons de prédire cette maladie. En effet, *Kourou, et al.* [4] ont ajouté une couche de subtilité et de précision à cet objectif car, selon eux, trois tâches principales de prédiction se démarquent lorsque l'on s'intéresse à la prédiction du cancer :

- La susceptibilité/le risque d'être atteint de la maladie [5],
- La récurrence [6, 7, 8],
- La survie ou non - et pendant combien de temps - du patient [9, 10].

Néanmoins, l'objectif final commun étant la prédiction, les techniques utilisées dans ces différentes parties sont plus ou moins les mêmes. C'est ce que nous allons voir par la suite.

Dans l'ensemble, l'objectif commun à toutes ces études est de déterminer le modèle prédictif le plus efficace possible afin de prédire le cancer.

Pour construire le modèle de prédiction le plus efficace possible, *Khourdifi, et al.* [2] se sont attelés à la comparaison de 4 méthodes différentes de ML : *Naïve Bayes* (NB) [74], *Support Vector Machine* (SVM) [75], *K-Nearest Neighbors* (KNN) [76] et *Rain Forest* (RF) [77].

Tout comme [2], *Li, et al.* [3] ont mis en place un modèle - en comparant les 5 méthodes de ML suivantes : *Decision Tree* (DT) [78], RF, SVM, *Neural Network* (NN) [79], *Linear Regression* (LR) [80] - pour classer les patients présentant une tumeur bénigne au sein d'une classe B et ceux ayant une tumeur maligne dans une classe M. Les enregistrements représentant les patients proviennent de 2 jeux de données différents issus de base de données médicales (BCCD et WBCD [11]).

Ayer T, et al. [5] ont réalisé la même classification - via leur modèle basé sur la méthode d'*Artificial Neural Networks* (ANNs) - en séparant les 48 774 images mammographiques à leur disposition en fonction de la présence d'anormalités malignes ou bénignes.

Exarchos KP, et al. [6] - dont l'étude porte sur la prédiction de la récurrence de carcinomes épidermoïdes buccal (OSCC) sur 86 patients (dont 73 sont complètement guéris sans rechute possible) - ont construit 2 classes pour les patients qui ont connu une rechute et ceux qui n'en ont pas connu à l'aide des méthodes de ML suivantes : *Bayesian Networks* (BNs) [81], ANNs, SVM, DTs, RF.

L'étude [7], dont le but est la prédiction de la récurrence du cancer du sein dans les 5 années suivant l'intervention chirurgicale subie par le patient, a, de façon similaire, catégorisé les patients à faible et haut risque grâce au modèle BCRSVM basé sur SVM.

Pour ce qui est de [9], une classe, appelée *survivability*, a été utilisée afin de classer les enregistrements selon que les patients ont survécu ou non. Cette étude vise à comparer les méthodes de ML SVM, ANN et une méthode d'apprentissage semi-supervisé à l'aide d'un jeu de données composé de 162 500 enregistrements et de 16 variables clés.

Chen Y-C, et al. [10] n'explicitent pas les classes que leur modèle ANN utilise tandis que Park C, et al. [8] ne mentionnent pas l'utilisation de cette méthode ni d'une autre méthode équivalente car leur modèle se base sur l'apprentissage semi-supervisé et ne nécessite pas cette étape. Dans ce cas, le jeu de données utilisé provient à la fois des expressions de gènes de patients atteints de NSCLC mais aussi de données cliniques issues d'une base de données externes [8].

Lors du traitement effectué par un modèle de ML, beaucoup d'erreurs rencontrées sont dues à la qualité des données en entrée [2]. Il est donc nécessaire de s'attarder sur une méthode de *preprocessing* afin de formater parfaitement les données selon les contraintes en entrée imposées par les différents modèles concernés. Il existe différentes méthodes de *preprocessing* comme la réduction de dimension (*dimensionality reduction*), la sélection de variables (*feature selection*) ou encore - la plus utilisée - l'extraction de variables (*feature extraction*). Effectivement, les auteurs de [2, 3] ont choisi d'appliquer une méthode de *feature extraction* dans le but de ne garder que les données de qualité en éliminant les données inutiles pour le traitement ou encore les données non valorisées (vide ou "NULL") qui pourraient fausser les résultats des différents modèles.

Exarchos KP, et al. [6] ne se sont pas contentés de la *feature selection* "classique" pour formater les données utilisées. En effet, leur étape de *preprocessing* des données a été plus complexe afin d'éviter tous les biais possibles. Ils ont procédé comme suit :

- Une première étape de *feature selection* qui permet de réduire la dimension des données. Cela signifie qu'on sélectionne uniquement les variables les plus importantes de chaque enregistrement avant de réduire le nombre d'attributs sur lesquels les traitements sont effectués par la suite,
- Une seconde étape pendant laquelle deux algorithmes de *feature extraction* sont utilisés.

Avant cette étape de *preprocessing*, les auteurs de [6] disposaient de 65 variables de type cliniques, 17 variables de type imageries et de 40 variables de type génomiques. À la suite du tri effectué par les différents algorithmes, il ne leur reste que, respectivement, 8, 6 et 7 variables. Concernant les données cliniques, par exemple, les attributs considérés comme intéressants sont de savoir si le patient est fumeur, l'épaisseur de la tumeur ainsi que des informations sur la tache p53 (*p53 stain*) [6].

Les méthodes de *preprocessing* - non détaillées - utilisées par [10] leur ont permis de sélectionner les signatures de gènes apportant le plus d'informations en rapport à la survie des patients.

Plus surprenant encore, les auteurs de [5] ont fait confiance aux Hommes pour réaliser cette tâche en demandant à des radiologistes d'analyser les différentes images utilisées pour l'étude. D'ailleurs, pour essayer de pallier certaines limites propres à leur modèle, ils proposent l'utilisation de méthodes de *preprocessing* pour formater convenablement les données en entrée du modèle.

Park K, et al. [9] non plus n'ont pas utilisé de méthode de *preprocessing* ; ce qui explique possiblement les résultats plus ou moins faibles obtenus par leur modèle que nous verrons par la suite.

Ces résultats sont mesurables grâce à différentes métriques dont les plus connues et utilisées sont :

- *Accuracy* : c'est la précision de la détection qui est mesurée par le taux de données correctement identifiées. En d'autres termes, c'est le nombre de prédictions correctes sur le nombre total de données dans le jeu de données [4],
- *Sensitivity or Recall* : la sensibilité est le taux de prédictions données positives et qui le sont vraiment [4],
- *Precision* : la précision correspond au taux d'éléments correctement classifiés au sein d'une même classe (déterminée par l'étape de classification) [4].

Nous pouvons voir sur les Tableaux 1 à 3 ci-dessous les résultats obtenus par les modèles des différentes études. Ce genre de métriques permet aux auteurs d'effectuer des comparaisons numériques entre les différents modèles analysés. Par exemple, on voit sur le Tableau 1 que, dans le cadre de l'étude [2], le modèle de ML optimal est SVM.

Tableau 1 : Performance des différents modèles de ML étudiés dans [2].

| Evaluation criteria | Classifiers | | | |
|---------------------------------|-------------|-------|-------|-------|
| | K-NN | SVM | RF | NB |
| Time to build model (s) | 0 | 0.08 | 0.28 | 0.01 |
| Correctly classified instances | 547 | 557 | 546 | 527 |
| Incorrectly classified instance | 22 | 12 | 23 | 42 |
| Accuracy (%) | 96.1 | 97.9 | 96 | 92.6 |
| TP Rate | 0,961 | 0,979 | 0,960 | 0,926 |
| FP Rate | 0,046 | 0,034 | 0,055 | 0,086 |
| Recall | 0,961 | 0,979 | 0,960 | 0,926 |
| Precision | 0,961 | 0,979 | 0,960 | 0,926 |

Il existe d'autres métriques numériques comme la zone sous la courbe (AUC) qui permet, bien souvent, de déterminer le meilleur modèle lors d'une comparaison. On peut voir sur les Tableaux 2 et 3 ci-dessous que la solution optimale de l'étude [3] est le modèle RF pour les 2 jeux de données mis à disposition.

Tableau 2 : Valeurs AUC obtenues sur le jeu de données BCCD [3].

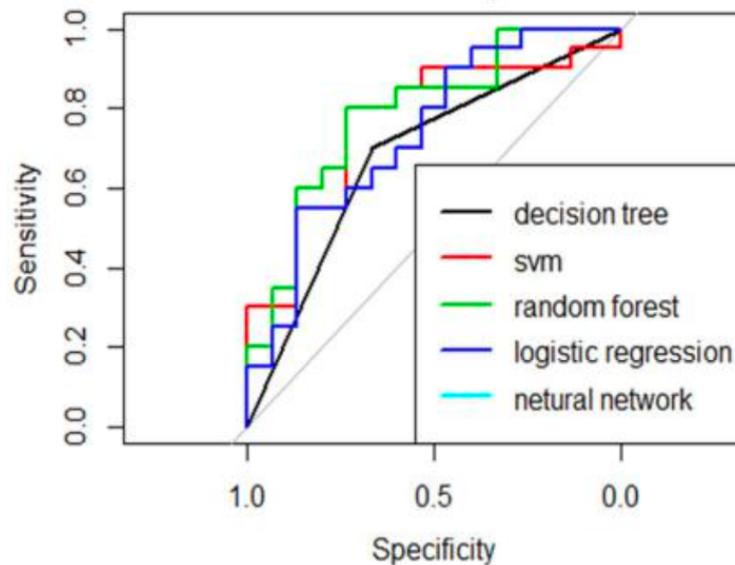
| Classification model | DT | SVM | RF | LR | NN |
|----------------------|-------|-------|-------|-------|-------|
| AUC | 0.683 | 0.763 | 0.785 | 0.737 | 0.748 |

Tableau 3 : Valeurs AUC obtenues sur le jeu de données WBCD [3].

| Classification model | DT | SVM | RF | LR | NN |
|----------------------|-------|-------|-------|-------|-------|
| AUC | 0.963 | 0.977 | 0.989 | 0.981 | 0.985 |

Ci-dessous, sur la Figure 1, nous pouvons voir la courbe ROC du modèle de *Li, et al.* [3] qui permet d'avoir une représentation graphique des résultats obtenus par les différents modèles de ML étudiés. Notons que cette technique d'évaluation est utilisée dans la majorité des études auxquelles nous nous intéressons ici.

Figure 1 : Courbes ROC des différents modèles étudiés sur le jeu de données BCCD [3].



Finalement, il existe plusieurs techniques permettant d'améliorer l'efficacité d'un modèle comme *S-fold-cross-validation* [2, 3, 5, 6, 7] dont le but est de diviser le jeu de données normalement utilisé par le modèle en *S* échantillons afin d'avoir des échantillons d'entraînement pour construire le modèle et d'autres de test [2]. Cela permet de mieux évaluer la performance du modèle notamment en termes de temps de traitement pour construire le modèle, classer les enregistrements puis calculer les métriques de mesure de performance [2]. Alors que *Khourdifi, et al.* [2], par exemple, ont utilisé la technique *10-fold-cross-validation*, *Park K, et al.* [9], eux, ont choisi de diviser leur jeu de données en 5 échantillons en appliquant la méthode suivante : *5-fold-cross-validation*.

L'évaluation du modèle obtenu par *Chen Y-C, et al.* [10] a été réalisée avec l'analyse de survie Kaplan-Meier [67] qui leur a permis de montrer que les patients à haut risque présentent un taux de survie médian inférieur à celui des patients classés à faible risque. Néanmoins, malgré la production de résultats robustes, ce modèle présente une limite majeure : selon les auteurs [10] leur modèle n'est pas généralisable à d'autres types de cancer et se limite au NSCLC.

Kim W, et al. [7] ne semblent pas avoir utilisé de méthode d'échantillonnage pour tester l'efficacité de leurs modèles. Cela pourrait être dû aux jeux de données à disposition qui sont déjà relativement petits.

En somme, nous avons vu que le ML est très largement utilisé dans la recherche de prédiction de l'apparition du cancer depuis de nombreuses années. Néanmoins, de nombreuses limites sont souvent rencontrées par les chercheurs, comme le fait que - comme on peut le voir via les résultats des différentes études ci-dessus - les technologies les plus efficaces ne sont jamais les mêmes en fonction des critères de comparaison appliqués ou des jeux de données utilisés. De plus, pour obtenir ces

données, de nombreuses techniques de *preprocessing* sont nécessaires afin de formater les données dont certaines qui entraînent une perte d'informations [2, 3] potentiellement intéressantes : les données non étiquetées. Ainsi, pour combler ces nombreuses failles, on pourrait penser à l'utilisation du DL pour tenter de répondre à cet objectif de prédiction dans le milieu médical ; et plus particulièrement pour la prédiction du cancer.

b. Le Deep Learning pour pallier certains manques

Ausawalaithong, et al. [13] ont développé un modèle de DL permettant la prédiction du cancer du sein à l'aide de l'analyse d'images produites par rayons X. Certes la qualité de ces images est plus faible qu'en utilisant d'autres méthodes comme la tomodensitométrie, mais cette méthode a le mérite d'être disponible en zones rurales [13] et est donc plus réalisable. De plus, les auteurs ont combiné cette méthode d'imagerie aux réseaux de neurones convolutifs - ou *convolutional neural networks* (CNN) [82] – qui ont prouvé être très efficaces dans un but de reconnaissance d'images ainsi que pour la tâche de classification [13]. Ils sont donc utilisés par les auteurs afin de détecter des anomalies dans les imageries produites par les rayons X thoraciques [13].

Xiao, et al. [12] ont misé sur l'analyse de données d'expression génique pour développer un modèle prédictif de DL "*stacked sparse auto-encoder*" (SSAE) basé sur le SSL. Ils ont comparé leur modèle à trois méthodes de classification supervisée communément utilisées : SVMs, RFs et NNs mais aussi à une méthode de classification non supervisée : les *Auto-Encoders* (AE) [83]. Les modèles ont été comparés à l'aide de 3 jeux de données contenant des informations sur les séquences ADN de 3 types de cancers différents : le cancer du poumon (LUAD), le cancer de l'estomac (STAD) et le cancer du sein (BRCA) [12].

Pour ce qui est du modèle de DL (CNN) en lui-même [13], sa structure est un bloc dense qui lui permet d'assurer la continuité du flux de données entre ses couches mais aussi d'éviter certains problèmes très connus lors de l'utilisation de réseaux de neurones profonds comme la perte d'informations entre les différentes couches du réseau (*vanishing gradient problem*) [15].

Comme on peut le voir sur la Figure 2 ci-dessous, le modèle SSAE de *Xiao, et al.* [12] est constitué de plusieurs couches, chacune représentée par un AE. La couche finale (de sortie) est le point d'entrée d'un réseau de neurones qui a été branché à la place de la couche "Décodeur" que l'on trouve normalement dans un modèle AE classique [12].

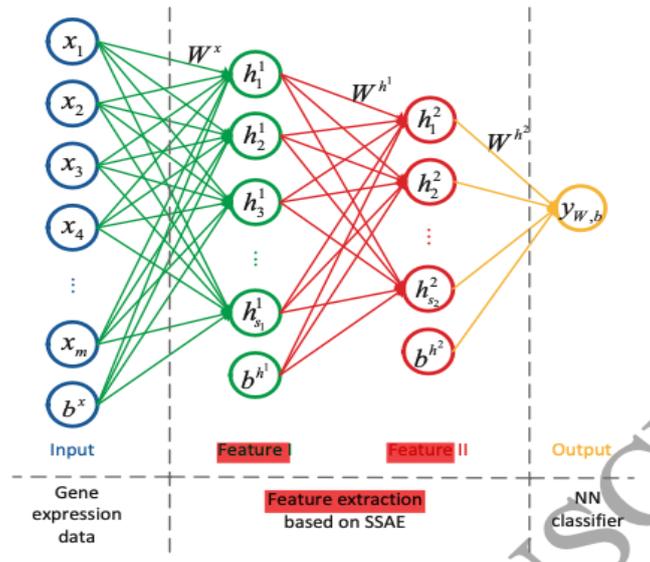


Figure 2 : Modèle de classification proposé par Xiao, et al. [12] composé de SSAE et NN.

Pour ce qui est de l'étape de *preprocessing*, Xiao, et al. [12] ont combiné deux méthodes afin d'utiliser toutes les données mises à disposition (qu'elles soient étiquetées ou non) :

- Une procédure de *feature extraction* non supervisée - ne prenant en compte que les données étiquetées - dont le but est d'entraîner chaque couche du réseau de neurones récursivement en prenant en entrée les résultats obtenus à la suite de l'entraînement de la couche précédente. Un algorithme de descente stochastique de gradient (SGD) [14] est ensuite appliqué pour affiner le modèle dans son ensemble [12].
- Un algorithme de classification supervisé qui prend en entrée tout type de données et dont l'objectif est de distinguer les patients atteints de cancer des personnes saines. Cela est fait par un jeu sur les réseaux de neurones avant d'appliquer un algorithme de propagation inverse pour tenter d'éliminer les possibles erreurs de classification commises par le modèle. Pour finir, on affine l'ensemble du modèle pour minimiser l'apparition de ces erreurs dans le futur [12].

Dans le cas de l'étude [13], aucune méthode de *preprocessing* « classique » n'est utilisée. Néanmoins, il est tout de même nécessaire de préparer au mieux les images qui seront fournies en entrée au modèle. Pour cela, il y a plusieurs étapes à suivre comme l'augmentation du contraste des images, l'utilisation de filtres pour éliminer le maximum d'imperfections, etc. Aussi, le traitement de données a été réalisé par une méthode appelée "apprentissage par transfert" (*transfer learning*) [13]. Dans ce contexte, étant donné la très petite taille du jeu de données, les auteurs l'ont appliqué deux fois : la première fois permettant aux images d'être classées comme "avec ou sans nodule(s)" tandis que la seconde fois leur permet d'être qualifiées plus précisément en tant que "avec ou sans nodule(s) **maligne(s)**". Grâce à cette méthode d'apprentissage automatique, le système devient de plus en plus spécifique au cancer du poumon au fil du temps [13].

La performance du modèle d'Ausawalaithong, et al. [13], évaluée avec les méthodes d'accuracy, specificity et sensitivity, a permis de s'assurer que la méthode d'entraînement utilisée dans cette étude est plus performante que l'apprentissage par transfert classique. En effet, le fait d'entraîner ce modèle à l'architecture dense plusieurs fois à la suite, à partir d'un jeu de données très petit, permet d'obtenir de meilleurs résultats moyens notamment grâce à une bonne identification visuelle de la localisation des nodules cancéreuses sur les images médicales utilisées [13]. Ainsi, dans l'ensemble le modèle répond convenablement au problème de jeux de données quantitativement faibles [13].

Concernant [12], les résultats obtenus par les différents modèles sur les 3 jeux de données utilisés sont visibles dans les Tableaux 4, 5 et 6 ci-dessous. Comme on peut le voir, le modèle SSAE proposé par Xiao, et al. [12] obtient quasiment toujours les meilleurs résultats quel que soit le jeu de données ou la métrique prise en compte. On peut donc dire que le modèle de DL SSAE basé sur le SSL permet d'améliorer les performances quant à la prédiction du cancer. Cela est due à l'architecture dense et profonde qu'offrent les réseaux de neurones qui permet aux algorithmes de DL de déceler automatiquement et de façon autonome les relations intrinsèques entre les structures de données [12].

Tableau 4 : Les résultats de prédiction via le jeu de données LUAD [12].

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1 (%) |
|-------|---------------------|---------------------|---------------------|---------------------|
| SVM | 97.54(±2.60) | 99.68(±1.45) | 96.80(±3.35) | 98.35(±1.74) |
| RF | 93.86(±3.71) | 96.18(±3.78) | 96.00(±4.90) | 95.98(±2.51) |
| NN | 87.77(±12.46) | 98.92(±2.87) | 84.00(±16.49) | 90.53(±10.86) |
| AE | 99.37(±1.40) | 99.29(±0.82) | 99.20(±1.79) | 99.19(±0.91) |
| SSAE | 99.89(±0.18) | 99.85(±0.21) | 99.90(±0.15) | 99.92(±0.34) |

Tableau 5 : Les résultats de prédiction via le jeu de données STAD [12].

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1 (%) |
|-------|---------------------|---------------------|---------------------|---------------------|
| SVM | 92.64(±1.72) | 92.65(±2.28) | 99.57(±0.95) | 95.97(±0.91) |
| RF | 94.09(±2.73) | 95.66(±3.59) | 97.88(±2.61) | 96.69(±1.50) |
| NN | 90.02(±3.20) | 98.85(±0.19) | 88.64(±3.59) | 93.95(±2.05) |
| AE | 94.46(±6.13) | 95.19(±6.60) | 99.17(±1.86) | 97.02(±3.25) |
| SSAE | 98.15(±0.04) | 99.58(±0.93) | 98.32(±0.94) | 98.94(±0.01) |

Tableau 6 : Les résultats de prédiction via le jeu de données BRCA [12].

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1 (%) |
|-------|---------------------|---------------------|---------------------|---------------------|
| SVM | 91.23(±1.17) | 90.97(±1.12) | 99.67(±0.12) | 95.27(±0.61) |
| RF | 93.96(±0.85) | 97.06(±1.71) | 96.13(±2.41) | 96.56(±0.54) |
| NN | 91.01(±4.80) | 99.57(±0.63) | 90.19(±5.27) | 94.59(±3.06) |
| AE | 94.08(±5.29) | 94.81(±5.95) | 99.10(±0.87) | 96.81(±2.77) |
| SSAE | 96.23(±0.80) | 99.08(±0.74) | 97.03(±0.35) | 98.05(±0.46) |

Remarquons que les différentes études analysées ci-dessus ne désignent quasiment jamais la même méthode de ML comme étant la meilleure. Ainsi, étant donné les résultats hétérogènes obtenus en fonction des jeux de données ou des critères de comparaison utilisés, il serait intéressant d'éviter les défauts ou les points faibles de certaines méthodes [16] pour améliorer les performances quant à la prédiction du cancer. Pour cela, comme suggéré par [2], on pourrait confectionner un multi-modèle de prédiction composé de plusieurs méthodes de ML associées aux architectures plus complexes des méthodes de DL.

c. L'association du Machine Learning et du Deep Learning

Xiao, et al. [16] ont poussé cette idée de combinaison de différentes méthodes en utilisant des réseaux de neurones profonds - *deep neural networks* (DNN) [84] - afin d'associer 5 méthodes de ML de classification différentes ; à savoir kNN, SVMs, DTs, RFs et "*gradient boosting decision trees*" (GBDTs) pour construire un multi-modèle permettant la prédiction du cancer. Les données utilisées en entrée du modèle sont des informations génétiques provenant de trois jeux de données contenant des séquences d'ADN [16].

Arya, et al. [17] se sont inspirés du *Multi-modal Deep Neural Network by integrating Multidimensional Data* (MDNNMD) de *Sun et al.* [18] pour construire leur modèle en l'adaptant afin de pallier certains manques présumés du modèle précité.

Leur modèle est constitué d'un CNN pour l'extraction de données et d'un réseau "*stacked-based*" [17] qui est en réalité un ensemble de méthodes de ML.

Les auteurs [17] ont utilisé les trois mêmes types de données en entrée que le modèle de *Sun et al.* ; à savoir le profil d'expression génique, la copie du nombre d'altérations (CNA) ainsi que les données cliniques du patient. Le jeu de données de l'étude étant le même que celui utilisé par *Sun et al.*, aucune étape de *preprocessing* des données n'a été nécessaire car tout a été fait pour le projet MDNNMD ; à savoir l'exécution d'un algorithme de *feature selection* connu - mRMR [21-23] - pour réduire la dimension du jeu de données [17].

Pour ce qui est de l'étude [16], *Xiao, et al.* ont utilisé une méthode de *feature selection* - la méthode DESeq [24] - pour ne sélectionner que les informations génétiques les plus pertinentes et ainsi réduire le *scope* des jeux de données. Par la suite, la méthode de *cross validation* [25] est utilisée pour diviser le jeu de données global en trois sous-ensembles différents : l'ensemble d'entraînement, l'ensemble de validation et l'ensemble de test. Cela permet de réduire le nombre d'erreurs de généralisation et d'éviter le surentraînement [25].

Dans le cas de [17], *Arya, et al.* ont décidé d'utiliser un CNN [19, 20] pour extraire chaque type de données différent : un CNN pour les données cliniques (CNN-Clinical), un CNN pour les données CNA (CNN-CNA) et un CNN pour les données d'expression génique (CNN-Expr). Les données résultant de ces modèles CNNs sont concaténées pour former les variables d'entrée du modèle "*stacked-based*" qui peut être un ensemble de méthodes de ML ou alors une unique méthode comme c'est le cas sur la Figure 3 ci-dessous - qui présente l'architecture du modèle de *Arya, et al.* [17] dans son ensemble - sur laquelle on peut voir l'utilisation d'un RF. Ce modèle "*stacked-based*" a été choisi car il est reconnu pour avoir de bonnes performances lorsqu'il s'agit de traiter des données "déséquilibrées" [26] ; ce qui est le cas de [17].

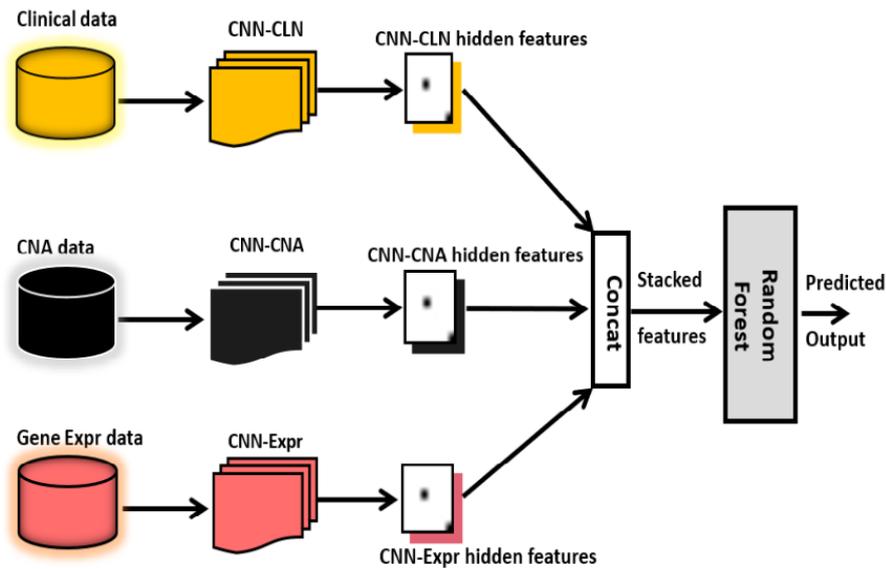


Figure 3 : Architecture du modèle “stacked-based” pour la prédiction du cancer du sein [17].

Xiao, et al. [16] ont fait l'inverse en utilisant d'abord les méthodes de ML afin de classer les données en formant des échantillons d'enregistrements représentant des personnes saines et d'autres représentant des patients atteints de cancer. Pour cette étape, le choix des méthodes de ML ainsi que leur assemblage a été réfléchi afin de tirer parti au maximum des avantages de chaque méthode en essayant d'éliminer les inconvénients [16]. Par exemple, DT ne semble pas être une méthode efficace de classification pour distinguer les personnes saines des patients atteints de cancer car elle cause souvent le surentraînement du modèle [16]. A l'inverse, les récentes méthodes de RFs et GBDTs permettent en général d'éviter ce problème [16]. Les résultats générés par cet ensemble de méthodes de classification sont fournis à un second modèle d'apprentissage [16].

Dans le cas de *Xiao, et al.* [16], ce modèle est un multi-modèle basé sur le DL (plus précisément des DNNs) qui combine les prédictions résultant de la première étape pour construire un ensemble prédictif final optimal. Cette prédiction finale est donnée par la couche de sortie des DNNs qui ne contient qu'un neurone dont la valeur peut être 0 (personne saine) ou 1 (patient présentant une tumeur maligne) [16]. La Figure 4 ci-dessous représente l'architecture du modèle utilisé par *Xiao, et al.*.

Notons que dans le cas des méthodes de ML, les algorithmes SGD sont utilisés pour tenter de réduire les erreurs produites par le modèle [16] tandis qu'ici c'est l'ajustement automatique des comportements du modèle de DL (des paramètres d'entrée et de sortie des couches internes du DNN en particulier) qui effectue ce travail de façon autonome.

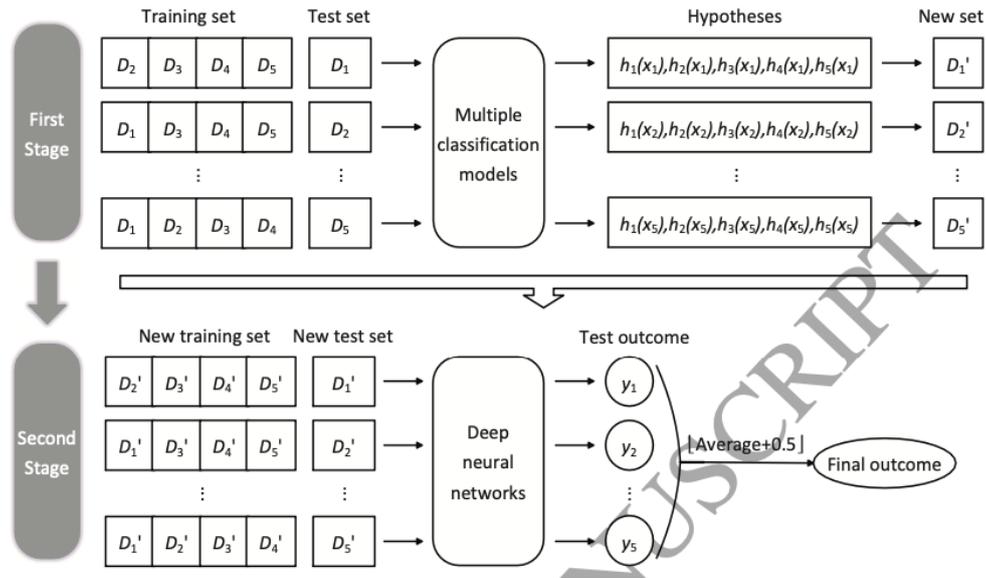


Figure 4 : Architecture du multi-modèle basé sur le DL proposé par Xiao, et al. [16].

Xiao, et al. [16] ont utilisé la technique de *5-fold-cross validation* pour réduire la dimension des jeux de données et augmenter la taille des échantillons utilisés ; permettant ainsi une utilisation efficace de leur méthode de DL.

Pour ce qui est des résultats, ils ont utilisé les cinq méthodes de ML prévues dans leur modèle de façon individuelle afin de récolter les résultats avant de lancer leur propre modèle sur les mêmes jeux de données. Les résultats (*l'accuracy* uniquement) ont été stockés dans un tableau pour effectuer les comparaisons entre les différents modèles. Comme on peut le voir sur le Tableau 7 ci-dessous, le modèle des auteurs [16] est le plus performant sur les trois jeux de données utilisés dans cette étude. Arya, et al. [17] ont, eux aussi, procédé à des comparaisons de performance entre leur modèle et d'autres méthodes de ML (comme SVM, RF et LR) ou avec des méthodes utilisées dans leur modèle de façon individuelle (CNN-Clinical, CNN-Expr, CNN-CNA). Comme on peut le voir sur le Tableau 8 ci-dessous, le modèle proposé par les auteurs [17] surpasse les autres méthodes que ce soit en termes d'AUC, d'*accuracy* (Acc), de *precision* (Pre) ou de *sensitivity* (Sn).

Notons que les deux modèles étudiés ici sont constitués à la fois de méthodes de ML mais aussi de DL. Néanmoins, le modèle de Arya, et al. [17] utilise le DL (plus particulièrement des CNNs) au début du traitement afin de mettre en forme les données à passer en entrée du modèle de ML qui produit le résultat final. Dans le cas de Xiao, et al. [16] c'est l'inverse : après les étapes de *preprocessing* c'est le ML qui est utilisé pour réaliser l'étape de classification et le DL (DNNs) qui donne le verdict.

Même si les deux modèles n'ont pas été étudiés sur les mêmes jeux de données, comparons leurs résultats. On se rend compte que le modèle de Xiao, et al. [16] obtient un taux d'*accuracy* moyen de 98.66% (+-1.21)¹ bien supérieur à celui du modèle de Arya, et al. [17] qui est de 88.1%.

Aussi, l'AUC moyen du modèle du modèle de [16] est de 0.985² alors que celui de [17] n'est que de 0.968.

Certes, la comparaison de deux métriques de performance n'est pas assez significative mais nous pouvons quand même penser que le modèle de [16] est plus efficace que celui de [17]. Cela est peut-être dû à l'ordre d'utilisation des méthodes de ML et de DL.

¹ En effectuant une moyenne des taux obtenus par la méthode proposée dans le Tableau 7 ci-dessous.

² En faisant, comme précédemment, une moyenne des AUCs obtenus sur les différents jeux de données de l'étude.

Il serait intéressant de faire fonctionner ces deux modèles avec les mêmes jeux de données afin de pouvoir, peut-être, tirer des conclusions.

Tableau 7 : Comparaison des métriques de performance de différents modèles comparés dans l'étude [16].

| Classification algorithm | LUAD | STAD | BRCA |
|--------------------------|---------------------|---------------------|---------------------|
| kNN | 88.00(±5.10) | 93.90(±2.99) | 95.08(±0.89) |
| SVM | 97.20(±2.28) | 81.22(±22.50) | 79.55(±19.22) |
| DT | 96.80(±2.28) | 96.59(±3.80) | 95.76(±0.94) |
| RF | 93.20(±1.79) | 96.83(±1.85) | 94.17(±1.53) |
| GBDT | 96.80(±2.28) | 96.59(±2.64) | 95.76(±4.46) |
| Majority voting | 97.20(±1.79) | 98.54(±1.34) | 98.18(±0.73) |
| Proposed method | 98.80(±1.79) | 98.78(±1.44) | 98.41(±0.41) |

Tableau 8 : Comparaison des métriques de performance de différents modèles comparés dans l'étude [17].

| Model | AUC | Acc | Pre | Sn |
|--------------------------|--------------|--------------|--------------|--------------|
| STACKED RF | | | | |
| (Hidden features) | 0.968 | 0.881 | 0.949 | 0.950 |
| CNN-Clinical | 0.761 | 0.843 | 0.116 | 0.047 |
| CNN-Expr | 0.941 | 0.837 | 0.317 | 0.360 |
| CNN-CNA | 0.811 | 0.848 | 0.278 | 0.158 |
| SVM | 0.709 | 0.840 | 0.846 | 0.844 |
| RF | 0.835 | 0.843 | 0.713 | 0.843 |
| LR | 0.683 | 0.735 | 0.788 | 0.721 |

Lorsque l'on s'intéresse au traitement du cancer, une part importante du travail consiste à pouvoir prédire l'apparition de la maladie en amont [2] ; pourquoi pas avant même l'apparition de symptômes afin d'être capable d'entamer un traitement rapidement et d'éviter le pire ou même de soigner le patient. Néanmoins, un autre aspect de ce travail pourrait s'avérer être utile : la prédiction de la réponse du système immunitaire d'un patient à un traitement qu'il pourrait potentiellement suivre [28]. En effet, cela permettrait aux médecins de savoir à l'avance les effets secondaires possiblement néfastes que pourrait avoir l'application d'un traitement. Ainsi, ils seraient en mesure de trouver le meilleur traitement au cas par cas. Dans cette optique, le DL s'avérerait être un outil d'aide à la décision précis et efficace [44].

II. Prédire la réponse aux traitements

a. A l'aide du Deep Learning et de jeux de données d'expressions génétiques

Pouvoir fournir des traitements personnalisés de grande efficacité en déterminant les interventions thérapeutiques nécessaires [28] est une tâche essentielle dans le cadre du traitement du cancer. Pour cela, il est nécessaire d'analyser les données génétiques des patients [27, 28]. Or, les jeux de données génétiques sont de trop grandes dimensions pour pouvoir être traités par des algorithmes de ML "classiques" [27, 28] comme ceux qu'on a vu dans la première partie. Ainsi, l'utilisation du DL pourrait être intéressante pour effectuer cette tâche [28].

Selon *Baptista et al.* [27], le dépistage de tous les traitements possibles n'est pas faisable par l'Homme car il existe un trop grand nombre de possibilités. Il faut donc profiter de la puissance de calcul des machines pour effectuer ce travail mais aussi pour convertir les connaissances tirées de ces données en réponse cellulaire potentielle [27] afin d'établir des traitements personnalisés en fonction des différents types de cancer et de patients. Le DL semble être le parfait outil pour cela car, en plus de pouvoir gérer des jeux de données à grande dimension et de travailler avec des données "bruyantes", il permet bien souvent d'éviter les phases de *preprocessing* en déterminant, de façon autonome durant sa phase d'apprentissage [27], les variables pertinentes pour le traitement.

La prédiction des réponses cellulaires aux traitements peut être divisée en deux sous-tâches :

- La prédiction de la sensibilité à un seul médicament [27] pour laquelle les auteurs ont analysé plusieurs études visant à construire des modèles répondant à cet objectif. En effet, ils se sont intéressés à un modèle mis en place par *Ding et al.* [29] basé sur des AEs profonds (*deep AEs*) capables d'extraire des informations importantes de données comme des cellules tumorales afin d'entraîner des modèles de classification tels qu'elastic net (Enet) [30] pour prédire l'effet d'un médicament sur les cellules cancéreuses du patient. Ce modèle a été entraîné sur le jeu de données « *Genomics of Drug Sensitivity in Cancer* » (GDSC) [33]. Certaines métriques de performance montrent la capacité du modèle à déterminer l'efficacité ou l'inefficacité du médicament tandis que d'autres prouvent que le système connaît des difficultés de généralisation lorsqu'il est testé avec un autre jeu de données [29]. Également, *Baptista et al.* [27] ont étudié le modèle DeepR [31] ; modèle de DL capable de prédire la sensibilité d'un médicament en se basant sur les mutations et les profils d'expression des cellules ou tumeurs cancéreuses. Ce modèle est constitué de trois sous-réseaux [31] et utilise des données cellulaires provenant des jeux de données « *Cancer Cell Line Encyclopedia* » (CCLE) [32] et GDSC afin de prédire la réponse d'un patient atteint de cancer à un médicament donné. DeepR a été comparé à des méthodes de ML comme LR et SVM mais aussi à des méthodes de DL comme DNNs (sans pré-entraînement) et a montré sa supériorité [31]. Finalement, les auteurs ont démontré que leur modèle est également capable de découvrir de potentielles nouvelles utilisations d'un médicament ou d'un composant d'un médicament [31] ; découverte que l'on verra par la suite. De leur côté, *Sakellaropoulos, et al.* [28] se sont attelés à la construction d'un modèle de DL - dont on peut voir la structure sur la Figure 5 ci-dessous - basé sur la méthode DNN et qui prend en entrée des données génétiques également issues du jeu de données GDSC.

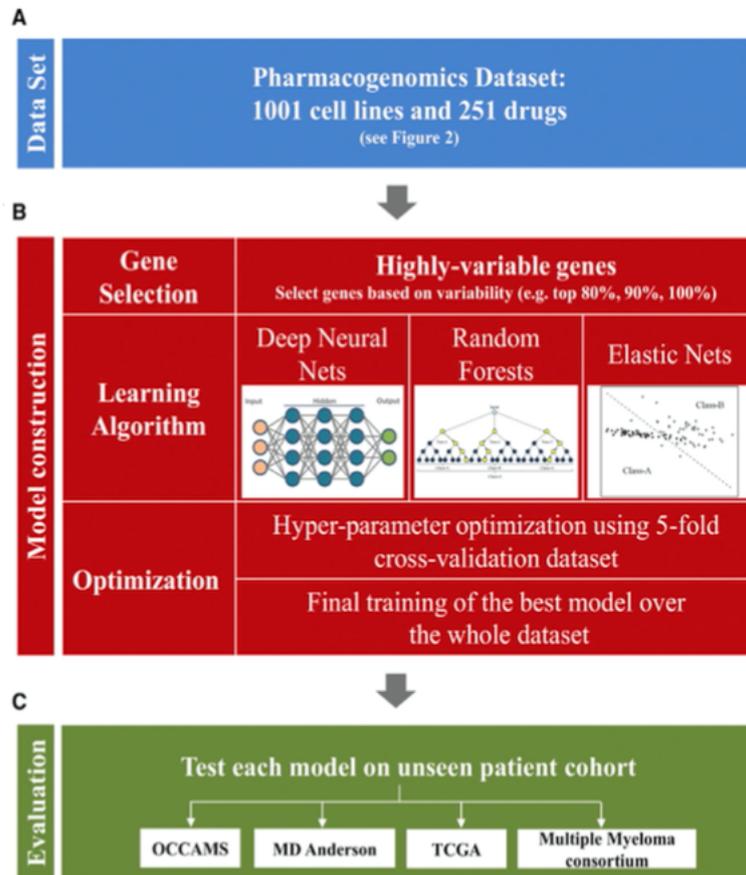


Figure 5 : Architecture du modèle de DL proposé par Sakellaropoulos, et al. [28].

Les auteurs [28] ont comparé les performances de leur modèle à des méthodes d'apprentissage fréquemment utilisées comme RF et Enet [28]. Pour cela, ils ont utilisé la technique *5-fold cross validation* pour sélectionner les paramètres optimaux afin de déterminer les capacités de chaque modèle à identifier les patients qui répondent - ou pas - à un médicament [28]. En d'autres termes, ils ont appliqué une étape de *feature selection* dans un premier temps (n'utilisant que les données optimales des données cellulaires à disposition) avant d'appliquer leurs différents modèles sur l'ensemble du jeu de données [28].

Pour ce qui est de l'évaluation des modèles, les auteurs se sont concentrés sur deux métriques principales : la taille d'effet (*effect size*) et la signification statistique (*statistical significance*) [28]. En effectuant les comparaisons à l'aide de ces deux métriques, ils [28] ont conclu que leur modèle est le plus performant. En effet, par exemple, aucun autre modèle n'obtient de bonnes performances concernant la signification statistique sur le jeu de données relatif au médicament paclitaxel [28]. Cela est sûrement dû au fait que leur modèle utilise tous les gènes à disposition dans le jeu de données [28], contrairement aux autres méthodes.

De plus, grâce à l'apprentissage autonome, le modèle de Sakellaropoulos, et al. [28] est capable de reconnaître des liaisons biologiques importantes entre médicaments ou entre molécules pour déterminer la réponse possible d'un patient à un médicament donné.

Baptista et al. [27] mentionnent même certaines études dans lesquelles les auteurs n'hésitent pas à combiner plusieurs méthodes de DL complémentaires afin d'obtenir des résultats encore plus efficaces lors de la prédiction des réponses aux médicaments [34]. C'est le cas de ELDAP [35] ; un ensemble hétérogène composé de quatre modèles de base différents (dont un DNN multitâche) qui a montré de meilleures performances lors d'une comparaison avec d'autres modèles simples ou des ensembles composés d'une seule base d'apprentissage via les jeux de données GDSC et CCLE.

- La prédiction des effets produits par un ensemble de médicaments. *Baptista et al.* [27] ont, également dans ce cas, étudié plusieurs modèles à l'état de l'art comme le modèle DeepSynergy [36] ou encore celui de *Xia et al.* [37] composé de sous-réseaux encodant chacun un type de données d'entrée mais aussi d'un sous-réseau responsable de la production de la prédiction finale. Ce modèle n'ayant pas été testé sur des données autres que celles utilisées lors de son entraînement, il est impossible de savoir si ses performances de prédiction seront aussi élevées sur des données inconnues [37]. De plus, une comparaison avec d'autres modèles à l'état de l'art entraînés sur les mêmes jeux de données aurait été intéressante [37].
Le modèle DeepSynergy [36] est l'objet de la première étude ayant été publiée dans le cadre de la recherche pour la prédiction de l'effet des combinaisons de médicaments. C'est un DNN qui utilise en entrée des données de réponse aux médicaments, des caractéristiques chimiques mais aussi des données d'expressions génétiques pour prédire un score de synergie [36]. Lui aussi s'est révélé performant sur les jeux de données utilisés pour son entraînement mais a également connu des difficultés de prédiction lors de l'utilisation de données inconnues [36].

En plus de prédire la réponse des cellules à un nouveau médicament ou à une nouvelle combinaison de médicaments, certains modèles de DL sont capables de répondre à une problématique de réutilisation de médicaments (*drug repurposing* [38]) [27]. Par exemple, un médicament ou une combinaison de médicaments qui est utilisé(e) pour le traitement d'un certain type de cancer peut être testé(e) par le modèle sur un autre type de cancer pour voir la réaction des cellules du patient [27]. Si le modèle perçoit une réponse, cela signifie que le médicament ou la combinaison est utilisable pour traiter le type de cancer en question [27] ; ainsi nous sommes dans un cas de *drug repurposing*. C'est ce qu'ont tenté de faire les auteurs du modèle CDRScan [39] qui ont étudié le potentiel de *drug repurposing* de leur modèle en l'utilisant pour prédire la réponse de cellules cancéreuses à environ 1000 médicaments provenant de la base de données DrugBank [40]. Parmi ces médicaments, leur modèle [39] a déterminé 23 cas de *drug repurposing*.

La plupart des études analysées par *Baptista et al.* [27] ont, elles aussi, utilisé la technique de *cross fold validation* pour évaluer la performance des modèles de prédiction de réponses aux médicaments. Néanmoins, comme vu dans le cas du modèle DeepSynergy [36] ou encore celui de *Xia et al.* [37], cette technique d'évaluation ne garantit pas la performance des modèles sur des données inconnues au préalable mais uniquement sur les données utilisées lors des entraînements [27].

Ainsi, il serait plus judicieux d'utiliser d'autres métriques de performance comme cela a été fait par *Sakellaropoulos, et al.* [28] ou alors de mettre en place d'autres techniques d'évaluation de performance comme "*leave-drug-out*" ou "*leave-cell-line-out*" [27]. Également, une autre possibilité serait de comparer les différents résultats générés par un modèle à l'aide de jeux de données différents [27]. Cette comparaison peut aussi être faite entre les résultats du modèle et des associations (des paires gène-médicament) déjà bien étudiées dans le monde de la recherche [27]. Finalement, beaucoup d'études comparent les modèles avec des méthodes de ML ne prenant pas exactement les mêmes données en entrée et/ou nécessitant une phase de *preprocessing* [27] or il serait plus intéressant d'effectuer cette comparaison plutôt avec des modèles de prédiction de réponses aux médicaments à l'état de l'art.

Toujours dans l'optique de fournir des solutions personnalisées permettant de traiter le cancer, le DL est aussi largement utilisé dans un autre domaine depuis plusieurs années maintenant : l'analyse d'imageries médicales [44, 54].

b. A l'aide du Deep Learning pour l'analyse d'imageries médicales

Malheureusement de nombreux patients atteints de certains types de cancer sont déjà à un stade assez avancé de la maladie lorsque celle-ci est détectée [44]. C'est le cas, par exemple, des personnes atteintes de NSCLC dont le taux de survie supérieur à 5 ans après la détection n'est que d'environ 18% [44]. Aussi, les cancers de la tête et du cou laissent aux personnes malades une chance de survie à long terme assez faible : moins de la moitié vivent aussi longtemps après la détection [54].

A ce stade, le cancer est traité avec des méthodes non-chirurgicales comme la radiation, la chimiothérapie, l'immunothérapie, etc. [44]. Ainsi, il est crucial de trouver un moyen d'identifier les patients à risque élevé et faible avant l'application d'un traitement afin de pouvoir adapter celui-ci [54] : un patient à haut risque aura besoin d'un traitement agressif pour traiter efficacement la maladie tandis qu'un patient à faible risque se verra assigner un traitement plus doux pour réduire les effets secondaires que peuvent produire ce genre de techniques [55].

L'IA - et plus particulièrement le ML - est très utilisée depuis des décennies pour tenter de résoudre ce genre de problématique [54] notamment avec une technique reconnue appelée "*radiomics*" (ou "*biomarqueurs d'image*") [45, 54] dont le principe est la caractérisation du phénotype des tumeurs cancéreuses grâce à l'extraction de données à partir d'images médicales [56]. Plus récemment, c'est le DL qui fait surface au sein du monde de la recherche [54] notamment car il a fait ses preuves quant à l'analyse d'images [44]. Ainsi, de nombreux chercheurs se sont intéressés au développement de modèles de DL pour résoudre ce problème en utilisant des CNNs ; technique de DL très efficace pour l'analyse d'images [54] grâce à sa capacité d'extraction d'informations importantes d'une image et d'identification de relations non-linéaires complexes entre différentes données [44].

Ayant pour objectif la prédiction de la survie d'un patient ou encore la réponse de celui-ci après une chimiothérapie ou une thérapie de radiation, *Xu, et al.* [44] ont développé un modèle de DL constitué de CNNs et de réseaux de neurones récurrents (RNNs) [85] en utilisant des pré-traitements ainsi que du suivi médical à l'aide d'images de tomodensitométrie (*CT images*). Comme on peut le voir sur la Figure 6 ci-dessous, ils [44] ont utilisé un modèle CNN déjà entraîné afin d'avoir une base de connaissances et qu'il soit plus performant quant à l'analyse des images passées en entrée. Ces images correspondent à un unique point sélectionné présentant la localisation d'une lésion cancéreuse (*timepoint*) [44].

Les algorithmes RNNs servent à s'assurer de pouvoir gérer le cas où il manquerait des *CT images* en entrée [46, 47] en amalgamant des informations provenant d'autres *timepoints* [44].

Selon *Diamant, et al.* [54], il existe trois principales approches concernant la prédiction à partir d'images : l'utilisation d'"*handcrafted features*" [57], de "*deep features*" [58-62] ou encore la technique de *transfer learning* comme utilisé par *Xu, et al.* [44] pour affiner le poids d'un NN existant [63, 64]. Dans ce cas, les auteurs vont mettre en place une nouvelle approche : ils vont utiliser un unique CNN entraîné "*de novo*" à l'aide de la technique de *gray-scale CT images* [54].

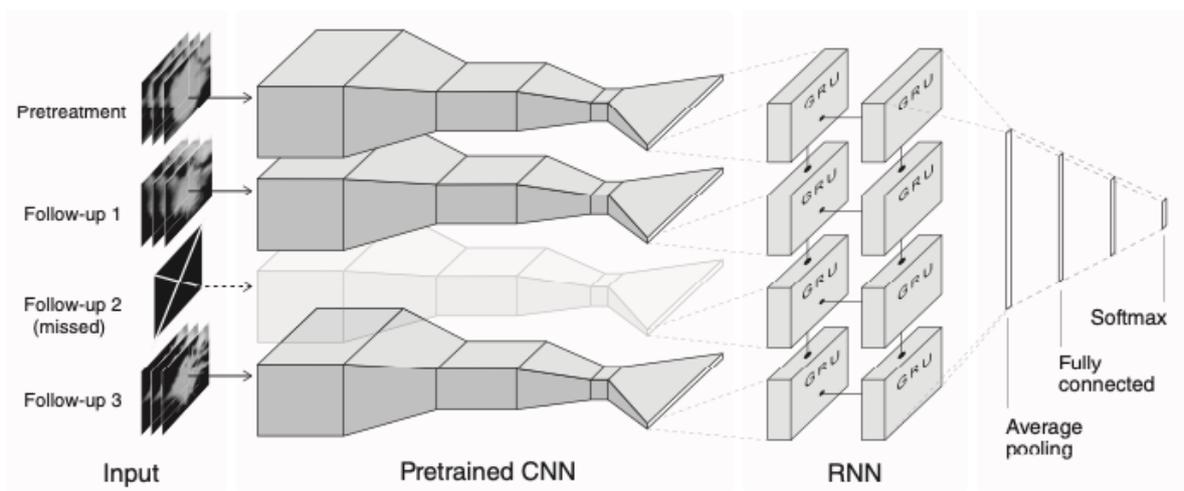


Figure 6 : Architecture du modèle proposé par Xu, et al. [44].

Pour cela, ils vont reprendre les données utilisées par l'étude (similaire) de Vallières et al. [65] comme jeu de données de validation afin de nourrir leur *framework* CNN ; l'objectif étant que le modèle soit capable de reconnaître les caractéristiques radiomiques d'une image [54]. Tout au long de cette étude [54] ils vont comparer leur modèle à celui de Vallières et al. [65]. Nous pouvons, d'ailleurs, voir les résultats des deux modèles dans le Tableau 9 ci-après.

Pour ce qui est des données utilisées par Xu, et al. [44], ils ont analysé deux jeux de données différents (*datasets* A et B) constitués de patients avec des maladies et des modalités de traitement différentes. Le premier, constitué de patients traités par une thérapie de chimio radiation définitive, leur a permis de développer et d'évaluer leur modèle. Le second jeu de données, composé de patients traités par de la chimio radiation suivie d'une intervention chirurgicale, leur a été utile pour évaluer les capacités de généralisation et de validations pathologiques de leur modèle.

| | Specificity | | Sensitivity | | Balanced Accuracy | |
|-----|---------------|--------------------------------|---------------|--------------------------------|-------------------|--------------------------------|
| | Present study | Vallières et al. ²⁵ | Present study | Vallières et al. ²⁵ | Present study | Vallières et al. ²⁵ |
| DM | 0.89 | 0.77 | 0.86 | 0.79 | 88% | 77% |
| LRF | 0.67 | 0.61 | 0.65 | 0.39 | 66% | 58% |
| OS | 0.67 | 0.67 | 0.68 | 0.55 | 68% | 62% |

Tableau 9 : Comparaison des résultats de l'étude menée par Diamant, et al. [54] avec celle de Vallières et al. [65].

Comme on peut le voir sur le Tableau 9 ci-dessus, Diamant, et al. [54] ont utilisé trois critères de comparaison différents. Tout d'abord, l'acronyme DM signifie "Distant metastasis" qui représente le taux de développement de métastases [54]. La seconde métrique est la défaillance logo-régionale (*loco-regionale failure*) qui est le taux d'échec de maîtrise locale de la maladie [54]. Finalement, la survie globale (*overall survival*) est le taux de survie des patients correspondants aux données étudiées [54].

Ces critères sont évalués grâce à trois métriques d'évaluation de performance vues précédemment (*specificity*, *sensitivity* et *accuracy*). Comme on peut le voir, le modèle de *Diamant, et al.* [54] est le plus performant en tout point.

De leur côté, *Xu, et al.* [44] ont mis en œuvre plusieurs techniques afin d'évaluer leur modèle selon différents angles : l'observation de l'AUC, le test Wilcoxon [66] ou encore la méthode d'analyse Kaplan-Meier [67]. Par exemple, grâce au test Wilcoxon ils [44] se sont rendu compte que les modèles de base auxquels on fournit des scans d'entraînement uniquement lors de la phase de pré-traitement produisent de faibles performances quant à la prédiction de survie sur 2 ans (AUC = 0.58). Néanmoins, l'ajout de scans de suivi fréquents a permis d'améliorer ces performances (AUC = 0.64 pour le scan à 1 mois, AUC = 0.69 pour celui à 3 mois et AUC = 0.74 pour celui à 1 an) [44]. Notons que cette courbe de progression a été aussi visible sur d'autres résultats comme la survie à 1 an, le développement de métastases, etc [44]. A titre de comparaison, un modèle similaire prenant en entrée des données cliniques semble être moins performant (AUC = 0.51 pour la prédiction de la survie à 2 ans). De plus, la méthode Kaplan-Meier leur a permis de pousser les résultats d'analyse de survie explicités ci-dessus en distinguant des groupes de patients à faible et haut risque de mortalité [44].

Pour savoir si leur modèle est capable de reconnaître les caractéristiques radiomiques convenablement, *Diamant, et al.* [54] ont pensé à inclure un ensemble de filtres (*convolutional filters*) dans leur modèle CNN "*de novo*". Ces filtres, disposés en nombre à chaque couche du réseau de neurones, permettent au modèle de reconnaître de plus en plus de caractéristiques dans les données reçues au fil du temps et des nombreux apprentissages.

Finalement, les auteurs [54] ont utilisé un PReLU (*parametrized rectified linear unit*) pour apprendre au modèle à détecter des erreurs (la présence de valeurs négatives, par exemple) notamment dans le cas d'opérations non linéaires. Cela est une étape importante dans l'objectif d'utiliser, un jour, le modèle dans l'industrie car celui-ci sera capable de reconnaître les erreurs humaines dans les données d'entrée.

Comme tout modèle composé de réseaux de neurones, celui de *Xu, et al.* [44] est susceptible de connaître le problème du surentraînement. Pour pallier cela, les auteurs ont utilisé certaines techniques [48, 49].

Quant aux limites du modèle de *Diamant, et al.* [54], certes il est capable de reconnaître les caractéristiques radiomiques, mais il ne les représente pas. Une inconnue demeure toujours entre les valeurs quantitatives de ces caractéristiques et les résultats fournis par les filtres utilisés par les auteurs dans leur modèle [54].

Aussi, les auteurs [54] estiment qu'il est possible d'améliorer la généralisabilité et les performances de prédiction de leur modèle ; notamment en retravaillant la préparation des *CT images* lors de l'étape de *preprocessing*. Néanmoins, selon eux, cela améliorerait la capacité d'apprentissage de leur modèle tout en le compliquant donc c'est une idée à réfléchir.

Finalement, dans ce cas, *Diamant, et al.* [54] ont mis en place une méthode de combinaison simple (LR) mais estiment qu'il serait intéressant d'étudier de meilleurs moyens permettant d'associer caractéristiques radiomiques et informations relatives aux CNNs.

Selon *Xu, et al.* [44], après avoir été bien entraîné sur plusieurs jeux de données de grande taille et diversifié, pourrait servir aux médecins dans leur choix de traitement.

Néanmoins, ils remontent certaines limites existantes de leur modèle.

Celui-ci prenant en entrée des images de localisation de tumeurs, il est nécessaire de disposer d'un outil capable de les détecter et de rogner les images [50] pour les fournir au modèle.

De plus, il a été montré [51] que, pour prédire le risque de développement de métastases, l'utilisation de méthodes "simples" de ML pourrait se révéler plus efficace qu'un modèle de DL comme celui de *Xu, et al.* [44] surtout dans le cas où on ne dispose pas d'échantillons de données assez importants.

Un autre problème se pose et c'est celui de l'interprétabilité : comme beaucoup de modèles de DL, le fonctionnement est une boîte noire [44] ce qui est un frein important comparé à des méthodes de ML qui sont clairement compréhensibles par le plus grand nombre.

Également, malgré l'utilisation de certaines techniques pour éviter ce problème, le modèle de [44] est exposé au surentraînement.

Les auteurs [44] remontent aussi un souci important quant à la définition des images utilisées par leur modèle qui sont des images 2D pour pouvoir utiliser le *transfer learning*. Pour améliorer les performances du modèle, il serait préférable d'utiliser des images 3D [44].

D'autres soucis relatifs à leur modèle sont remontés dont on ne parlera pas ici mais que vous pouvez retrouver dans les études [44, 51, 54].

Selon *Xu, et al.* [44], un réseau CNN entraîné sur un grand nombre d'images peut être utilisé pour traiter les images médicales grâce à la technique du *transfer learning* [52]. A l'inverse, le modèle "*de novo*" de *Diamant, et al.* [54] prouve qu'il n'est pas toujours nécessaire d'avoir des milliers - voir des millions - d'images à disposition pour obtenir des résultats robustes. Dans leur cas, un *training set* composé d'uniquement 200 images leur a permis de développer un modèle performant.

De plus, les auteurs [54] affirment que l'interprétabilité de leur modèle est assez intuitive comparé à d'autres. Selon eux, malgré certaines limites remontées, leur modèle pourrait être utilisé comme un outil d'analyse d'images fiable et pourrait également être adapté pour servir à d'autres problématiques d'imageries [54].

Ainsi, l'IA peut être utilisée dans l'industrie médicale en tant que support avec un rôle d'outil d'aide à la décision précis et efficace [45, 53].

Conclusion

Ayant pour objectif principal la recherche du moyen permettant de traiter le cancer, les chercheurs ne cessent de profiter des avancées technologiques constantes [2]. Dans ce contexte, après avoir utilisé différentes méthodes comme le dépistage à un stade peu avancé [2], l'augmentation des données en relation avec le cancer mises à disposition leur a permis de se tourner vers l'utilisation du ML ; qui est devenu un outil populaire dans le milieu de la recherche [2]. Ainsi, de nombreux modèles se basant sur de nombreuses méthodes de ML ont été testés. Néanmoins, dans une optique de recherche d'optimisation des performances, ces modèles présentent des inconvénients [16] qu'il est nécessaire d'éviter. C'est en suivant cette idéologie que sont nés des modèles associant méthodes de classification de ML classiques et méthodes de DL plus avancées [16, 17] et même des modèles combinant plusieurs méthodes de DL entre elles [35].

Tous ces modèles - permettant de prédire l'apparition d'un certain type de cancer, la probable récurrence de la maladie, le taux de survie dans les années suivant un traitement ou encore la prédiction des effets que peut avoir un traitement sur un individu en particulier - ont besoin de données en entrée afin de pouvoir fonctionner. Certains ont besoin de données formatées et/ou étiquetées tandis que d'autres acceptent tout type de données en fonction de leur méthode d'apprentissage (non supervisé, semi supervisé, supervisé). Ces données peuvent être de plusieurs types : cliniques, génétiques, images, etc.

Les modèles sont évalués à l'aide de métriques d'évaluation de performance numériques (*accuracy*, *sensitivity*, *precision*, AUC, etc) ou visuelles (courbe ROC) afin de pouvoir les comparer et déterminer le meilleur modèle en fonction du besoin. Il est aussi possible d'optimiser les performances d'un modèle en appliquant certaines techniques comme celles de type *fold cross validation* qui ont beaucoup été utilisées dans les modèles étudiés ici.

Malgré toutes ces techniques à disposition, la plupart des modèles proposés présentent toujours des limites. Pour les modèles de ML à l'apprentissage supervisé, certaines limites sont souvent relatives à la quantité des données qui, une fois passée les étapes de *preprocessing* pour formatage, ne sont plus suffisamment nombreuses pour avoir des résultats assez robustes [6, 7]. Concernant les modèles de DL, il a été souvent remonté dans les études que les modèles produisent des résultats intéressants sur des jeux de données connus (utilisés lors des entraînements) mais ne sont plus aussi performants lorsqu'ils doivent traiter des données inconnues (de nouveaux jeux de données) [36, 37]. Il existe d'autres limites comme le temps de traitement nécessaire à certains modèles de DL proposés comme celui de l'étude [16], le manque de données disponibles - notamment les données génétiques ou en relation avec les réponses cellulaires aux traitements - afin d'entraîner convenablement les modèles efficaces de DL [28] ou encore le manque d'interprétabilité qui a empêché l'utilisation de certains modèles dans l'industrie pour répondre à des problèmes biologiques ou relatifs à la santé [68-70].

Pour nombre de modèles, il reste des choses à étudier, à essayer et à analyser. Par exemple, le modèle de Xu, *et al.* [44] mériterait d'être testé avec des images de lésion en 3D plutôt qu'en 2D afin de voir si cela améliorerait - ou pas - les performances.

Selon de Baptista *et al.* [27], certes les modèles de ML sont performants quant à la prédiction des réponses aux médicaments mais il reste du travail lorsqu'il s'agit de généraliser cette prédiction à de nouvelles cellules ou de nouveaux médicaments. Également, selon les auteurs [27] il serait intéressant de remplacer les méthodes de DL classiquement utilisées pour la tâche de prédiction de réponse aux médicaments (DNNs simples, CNNs, etc) par d'autres méthodes de DL plus avancées [71-73] qui proposent d'autres avantages et caractéristiques potentiellement intéressantes dans ce cas.

Le modèle de *Diamant, et al.* [54] est capable de reconnaître les caractéristiques radiomiques mais n'est pas en mesure de les représenter. Cela doit être étudié dans le futur [54]. Aussi, de manière générale, *Diamant, et al.* [54] estiment qu'il serait intéressant d'étudier de nouvelles façons d'associer caractéristiques radiomiques et informations relatives aux NNs.

En somme, nous pouvons penser que même s'il existe encore aujourd'hui un certain nombre d'incertitudes, de limites et de découverte concernant les modèles de prédiction du cancer, les avancées réalisées par les chercheurs sur les dernières décennies nous laissent de l'espoir quant à l'utilisation de ces modèles dans l'industrie d'ici quelques années ou quelques décennies. Pour y arriver, la recherche de nouvelles techniques, de nouveaux principes et modèles toujours plus performants est un point clé qu'il est crucial de développer afin d'atteindre cet objectif de prédiction du cancer.

Annexes

Table des Figures

| | |
|---|----|
| TABLEAU 1 : PERFORMANCE DES DIFFERENTS MODELES DE ML ETUDIES DANS [2]..... | 9 |
| TABLEAU 2 : VALEURS AUC OBTENUES SUR LE JEU DE DONNEES BCCD [3]. | 9 |
| TABLEAU 3 : VALEURS AUC OBTENUES SUR LE JEU DE DONNEES WBCD [3]..... | 9 |
| FIGURE 1 : COURBES ROC DES DIFFERENTS MODELES ETUDIES SUR LE JEU DE DONNEES BCCD [3]..... | 10 |
| FIGURE 2 : MODELE DE CLASSIFICATION PROPOSE PAR XIAO, ET AL. [12] COMPOSE DE SSAE ET NN..... | 13 |
| TABLEAU 4 : LES RESULTATS DE PREDICTION VIA LE JEU DE DONNEES LUAD [12]. | 14 |
| TABLEAU 5 : LES RESULTATS DE PREDICTION VIA LE JEU DE DONNEES STAD [12]..... | 14 |
| TABLEAU 6 : LES RESULTATS DE PREDICTION VIA LE JEU DE DONNEES BRCA [12]..... | 14 |
| FIGURE 3 : ARCHITECTURE DU MODELE "STACKED-BASED" POUR LA PREDICTION DU CANCER DU SEIN [17]. | 17 |
| FIGURE 4 : ARCHITECTURE DU MULTI-MODELE BASE SUR LE DL PROPOSE PAR XIAO, ET AL. [16]. | 18 |
| TABLEAU 7 : COMPARAISON DES METRIQUES DE PERFORMANCE DE DIFFERENTS MODELES COMPARES DANS L'ETUDE [16]. | 19 |
| TABLEAU 8 : COMPARAISON DES METRIQUES DE PERFORMANCE DE DIFFERENTS MODELES COMPARES DANS L'ETUDE [17]. | 19 |
| FIGURE 5 : ARCHITECTURE DU MODELE DE DL PROPOSE PAR SAKELLAROPOULOS, ET AL. [28]. | 21 |
| FIGURE 6 : ARCHITECTURE DU MODELE PROPOSE PAR XU, ET AL. [44]..... | 24 |
| TABLEAU 9 : COMPARAISON DES RESULTATS DE L'ETUDE MENEES PAR DIAMANT, ET AL. [54] AVEC CELLE DE VALLIERES ET AL. [65]. | 24 |

Glossaire

IA: Intelligence Artificielle
DL: Deep Learning
ML: Machine Learning
SL (*Supervised Learning*): Apprentissage Supervisé
SSL (*Semi-supervised Learning*): Apprentissage semi-supervisé
NSL (*uNSupervised Learning*): Apprentissage non supervisé
SVM: Support Vector Machine
NB: Naïve Bayes
KNN: K-nearest Neighbors
RF: Rain Forest
DT: Decision Trees
NN: Neural Networks
ANN: Artificial Neural Networks
CNN: Convolutional Neural Networks
LR: Linear Regression
BN: Bayesian Networks
NSCLC: Non-Small Cell Lung Cancer
AE: Auto-Encoders
SGD: Stochastic Gradient Descent
AUC: Area Under the Curve
ROC: Receiver Operating Characteristic

Références

- [1] Knowles, Margaret, and Selby. "What is cancer?" *Introduction to the cellular and molecular biology of cancer*: 1-24. Oxford university press, 2005.
- [2] Khourdifi, Youness, and Bahaj. "Applying best machine learning algorithms for breast cancer prediction and classification." *2018 International conference on electronics, control, optimization and computer science (ICECOCS)*. IEEE, 2018.
- [3] Li, Yixuan, and Chen. "Performance evaluation of machine learning methods for breast cancer prediction." *Appl Comput Math* 7.4 (2018): 212-216.
- [4] Kourou, Konstantina, et al. "Machine learning applications in cancer prognosis and prediction." *Computational and structural biotechnology journal* 13 (2015): 8-17.
- [5] Ayer, Alagoz, Chhatwal, Shavlik, Kahn, Burnside. "Breast cancer risk estimation with artificial neural networks revisited." *Cancer* 2010; 116:3310–21.
- [6] Exarchos, Goletsis, Fotiadis. "Multiparametric decision support system for the prediction of oral cancer reoccurrence." *IEEE Trans Inf Technol Biomed* 2012;16: 1127–34.
- [7] Kim, Kim, Lee, Noh, Kim, Jung, et al. "Development of novel breast cancer recurrence prediction model using support vector machine." *J Breast Cancer* 2012; 15:230–8.
- [8] Park, Ahn, Kim, Park. "Integrative gene network construction to analyze cancer recurrence using semi-supervised learning." *PLoS One* 2014; 9:e86309.
- [9] Park, Ali, Kim, An, Kim, Shin. "Robust predictive model for evaluating breast cancer survivability." *Engl Appl Artif Intell* 2013; 26:2194–205.
- [10] Chen, Ke, Chiu. "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories." *Comput Biol Med* 2014; 48:1–7.
- [11] Breast Cancer Wisconsin Dataset, Kaggle. Accessed on: Feb 13, 2020. Available: <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
- [12] Xiao, Yawen, et al. "A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-seq data." *Computer methods and programs in biomedicine* 166 (2018): 99-105.
- [13] Ausawalaithong, Worawate, et al. "Automatic lung cancer prediction from chest X-ray images using the deep learning approach." *2018 11th Biomedical Engineering International Conference (BMEICON)*. IEEE, 2018.
- [14] Bottou, Léon. "Stochastic gradient descent tricks." *Neural networks: Tricks of the trade*. Springer, Berlin, Heidelberg, 2012. 421-436.
- [15] Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 6.02 (1998): 107-116.
- [16] Xiao, Yawen, et al. "A deep learning-based multi-model ensemble method for cancer prediction." *Computer methods and programs in biomedicine* 153 (2018): 1-9.
- [17] Arya, Nikhilanand, and Sriparna Saha. "Multi-modal classification for human breast cancer prognosis prediction: Proposal of deep-learning based stacked ensemble model." *IEEE/ACM transactions on computational biology and bioinformatics* (2020).
- [18] Sun, Wang, and Li, "A multimodal deep neural net- work for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2018.

- [19] Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, June 2015, pp. 1–9. [Online]. Available: doi.ieeecomputersociety.org/10.1109/CVPR.2015.7298594
- [20] Tan, Yuan, and Su, "Content-based similar document image retrieval using fusion of cnn features," *Internet Multimedia Computing and Service*, p. 260–270, 2018. [Online]. Available: http://dx.doi.org/10.1007/978-981-10-8530-7_25
- [21] Peng, Long, and Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug 2005.
- [21] Ding and Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Proceedings of the IEEE Computer Society Conference on Bioinformatics*, ser. CSB '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 523–. [Online]. Available : <http://dl.acm.org/citation.cfm?id=937976.938050>
- [23] Cai, Huang, Hu, Shi, Xie, and Li, "Prediction of lysine ubiquitination with mrmr feature selection and analysis," *Amino Acids*, vol. 42, no. 4, pp. 1387–1395, Apr 2012. [Online]. Available: <https://doi.org/10.1007/s00726-011-0835-0>
- [24] Anders and Huber. "Differential expression analysis for sequence count data." *Genome Biology*, 11(10):1–12, 2010.
- [25] Bishop. "Pattern Recognition and Machine Learning", chapter 1.3. Springer Science+Business Media, 2006.
- [26] Zhang, Li, He, and Wang, "A novel mkl method for gbm prognosis prediction by integrating histopathological image and multi-omics data," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, pp. 1–1, 02 2019.
- [27] Baptista, Delora, Pedro Ferreira, and Rocha. "Deep learning for drug response prediction in cancer." *Briefings in Bioinformatics* 22.1 (2021): 360-379.
- [28] Sakellaropoulos, Theodore, et al. "A deep learning framework for predicting response to therapy in cancer." *Cell reports* 29.11 (2019): 3367-3373.
- [29] Ding, Chen, Cooper, et al. "Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics." *Mol Cancer Res* 2018; 16(2): 269–78.
- [30] Li, Qing, and Lin. "The Bayesian elastic net." *Bayesian analysis* 5.1 (2010): 151-170.
- [31] Chiu, Chen, Zhang, et al. "Predicting drug response of tumors from integrated genomic profiles by deep neural networks." *BMC Med Genomics* 2019; 12(S1): 18.
- [32] Barretina, Caponigro, Stransky, et al. "The Cancer Cell Line Encyclopedia enables predictive modelling of anti-cancer drug sensitivity." *Nature* 2012; 483(7391): 603–307.
- [33] Yang, Soares, Greninger, et al. "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells." *Nucleic Acids Res* 2013; 41(D1): D955–61.
- [34] Costello, Heiser, Georgii, et al. "A community effort to assess and improve drug sensitivity prediction algorithms." *Nat Biotechnol* 2014; 32(12): 1202–12.
- [35] Tan, Özgül, Bardak, et al. "Drug response prediction by ensemble learning and drug-induced gene expression signatures." *Genomics* 2018; 111(5): 1078–1088.
- [36] Preuer, Lewis, Hochreiter, et al. "DeepSynergy: predicting anti-cancer drug synergy with deep learning." *Bioinformatics* 2018 ; 34(9): 1538–46.
- [37] Xia, Shukla, Brettin, et al. "Predicting tumor cell line response to drug pairs with deep learning." *BMC Bioinformatics* 2018; 19(S18): 486.
- [38] Pulley, Rhoads, Jerome, et al. "Using what we already have: uncovering new drug repurposing strategies in existing omics data." *Annu Rev Pharmacol Toxicol* 2020; 60(1): annurev-pharmtox-010919-023537.
- [39] Chang, Park, Yang, et al. "Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature." *Sci Rep* 2018; 8(1): 8857.

- [40] Wishart, Feunang, Guo, et al. "DrugBank 5.0: a major update to the DrugBank database for 2018." *Nucleic Acids Res* 2018; 46(D1): D1074–82.
- [41] Bishop. "Pattern recognition and machine learning." New York: Springer; 2006.
- [42] Mitchell. "The discipline of machine learning: Carnegie Mellon University." Carnegie Mellon University, School of Computer Science, Machine Learning Department; 2006.
- [43] Witten, Frank. "Data mining: practical machine learning tools and techniques." Morgan Kaufmann; 2005.
- [44] Xu, Yiwen, et al. "Deep learning predicts lung cancer treatment response from serial medical imaging." *Clinical Cancer Research* 25.11 (2019): 3266-3275.
- [45] Hosny, Parmar, Quackenbush, Schwartz, Hugo. "Artificial intelligence in radiology." *Nat Rev Cancer* 2018; 18:500–10.
- [46] Rubins, Unger, Colice. "Follow-up and surveillance of the lung cancer patient following curative intent therapy." *Chest* 2007;132: 355S–367S.
- [47] Calman, Beaver, Hind, Lorigan, Roberts, Lloyd-Jones. "Survival benefits from follow-up of patients with lung cancer: a systematic review and meta-analysis." *J Thorac Oncol* 2011; 6:1993–2004.
- [48] Ioffe, Szegedy. "Batch normalization: accelerating deep network training by reducing internal covariate shift." arXiv Preprint 2015. [arxiv.org/abs/ 1502.03167](https://arxiv.org/abs/1502.03167).
- [49] Srivastava, Hinton, Krizhevsky, Sutskever, Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." *J Mach Learn Res* 2014; 15:1929–58.
- [50] Valente, Cortez, Neto, Soares, de Albuquerque, Tavares. "Automatic 3D pulmonary nodule detection in CT images: a survey." *Comput Methods Programs Biomed* 2016; 124:91–107.
- [51] Coroller, Grossmann, Hou, Rios Velazquez, Leijenaar, Hermann, et al. "CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma." *Radiother Oncol* 2015; 114:345–50.
- [52] Shin, Roth, Gao, Lu, Xu, Nogues, et al. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." *IEEE Trans Med Imaging* 2016;35: 1285–98.
- [53] Parmar, Barry, Hosny, Quackenbush, Aerts. "Data analysis strategies in medical imaging." *Clin Cancer Res* 2018; 24:3492–9.
- [54] Diamant, André, et al. "Deep learning in head & neck cancer outcome prediction." *Scientific reports* 9.1 (2019): 1-10.
- [55] Tolentino, Ed et al. "Oral adverse effects of head and neck radiotherapy: literature review and suggestion of a clinical oral care guideline for irradiated patients." *Journal of applied oral science: revista FOB* 19, 448–54 (2011).
- [56] Gillies, R. J., Kinahan, P. E. & Hricak, H. "Radiomics: Images Are More than Pictures, They Are Data." *Radiology* 278, 563–577 (2016).
- [57] Zwanenburg, A., Leger, S., Vallières, M. & Löck, S. "Image biomarker standardisation initiative." *Arxiv* (2016).
- [58] Huynh, B. Q., Li, H. & Giger, M. L. "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks." *Journal of Medical Imaging* 3, 034501 (2016).
- [59] Antropova, N., Huynh, B. Q. & Giger, M. L. "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets." *Med. Phys.* 44 (2017).
- [60] Paul, R. et al. "Deep Feature Transfer Learning in Combination with Traditional Features Predicts Survival Among Patients with Lung Adenocarcinoma." *Tomography (Ann Arbor, Mich.)* 2, 388–395 (2016).
- [61] Paul, R. et al. "Predicting malignant nodules by fusing deep features with classical radiomics features." *Journal of Medical Imaging* 5, 1 (2018).
- [62] Paul, R., Hawkins, S. H., Hall, L. O., Goldgof, D. B. & Gillies, R. J. "Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic CT." In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002570–002575 (IEEE, 2016).

- [63] Goodfellow, I., Bengio, Y. & Courville, A. "*Deep Learning*" (MIT Press, 2016).
- [64] Buduma, N. "*Fundamentals of Deep Learning*" (O'Reilly, 2015).
- [65] Vallières, M. *et al.* "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer." *Scientific Reports* 7, 1–33 (2017).
- [66] Woolson, R. F. "Wilcoxon signed-rank test." *Wiley encyclopedia of clinical trials* (2007): 1-3.
- [67] Stel, Vianda, *et al.* "Survival analysis I: the Kaplan-Meier method." *Nephron Clinical Practice* 119.1 (2011): c83-c88.
- [68] Goh, Hodas, Vishnu. "Deep learning for computational chemistry." *J Comput Chem* 2017 ; 38(16): 1291–307.
- [69] Ching, Himmelstein, Beaulieu-Jones, *et al.* "Opportunities and obstacles for deep learning in biology and medicine." *J R Soc Interface* 2018 ; 15(141): 20170387.
- [70] Mamoshina, Vieira, Putin, *et al.* "Applications of deep learning in biomedicine." *Mol Pharm* 2016 ; 13(5): 1445–54.
- [71] Duvenaud, Maclaurin, Aguilera-Iparraguirre, *et al.* "Convolutional networks on graphs for learning molecular fingerprints." *J Chem Inf Model* 2015; 56(2): 399–411.
- [72] Kearnes, McCloskey, Berndl, *et al.* "Molecular graph convolutions: moving beyond fingerprints." *J Comput Aided Mol Des* 2016; 30(8): 595–608.
- [73] Xu, Wang, Zhu, *et al.* "Seq2seq Fingerprint." In: Proc. 8th ACM Int. Conf. Bioinformatics, Comput. Biol. Heal. Informatics - ACM-BCB '17. New York, USA: ACM Press2017, pp. 285–294.
- [74] Webb, Geoffrey, Keogh, and Miikkulainen. "Naïve Bayes." *Encyclopedia of machine learning* 15 (2010): 713-714.
- [75] Hearst, Marti, *et al.* "Support vector machines." *IEEE Intelligent Systems and their applications* 13.4 (1998): 18-28.
- [76] Peterson, Leif. "K-nearest neighbor." *Scholarpedia* 4.2 (2009): 1883.
- [77] Gehrke, Johannes, Ramakrishnan, and Ganti. "Rainforest-a framework for fast decision tree construction of large datasets." *VLDB*. Vol. 98. 1998.
- [78] Freund, Yoav, and Mason. "The alternating decision tree learning algorithm." *icml*. Vol. 99. 1999.
- [79] Wang, Sun-Chong. "Artificial neural network." *Interdisciplinary computing in java programming*. Springer, Boston, MA, 2003. 81-100.
- [80] Groß, Jürgen. "*Linear regression*." Vol. 175. Springer Science & Business Media, 2012.
- [81] Friedman, Nir, Dan Geiger, and Goldszmidt. "Bayesian network classifiers." *Machine learning* 29.2 (1997): 131-163.
- [82] Kim, Phil. "Convolutional neural network." *MATLAB deep learning*. Apress, Berkeley, CA, 2017. 121-147.
- [83] Ng, Andrew. "Sparse autoencoder." *CS294A Lecture notes* 72.2011 (2011): 1-19.
- [84] Liu, Weibo, *et al.* "A survey of deep neural network architectures and their applications." *Neurocomputing* 234 (2017): 11-26.
- [85] Medsker, Larry, and Jain. "Recurrent neural networks." *Design and Applications* 5 (2001) : 64-67.