



HAL
open science

Apprentissage profond pour l'analyse d'images histopathologiques

Robin Heckenauer

► **To cite this version:**

Robin Heckenauer. Apprentissage profond pour l'analyse d'images histopathologiques. Biotechnologie. Université de Haute Alsace - Mulhouse, 2023. Français. NNT : 2023MULH5410 . tel-04296605

HAL Id: tel-04296605

<https://theses.hal.science/tel-04296605>

Submitted on 20 Nov 2023

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

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

UNIVERSITÉ DE HAUTE-ALSACE
UNIVERSITÉ DE STRASBOURG

Thèse

Pour l'obtention du grade de
Docteur de l'Université de Haute-Alsace

École Doctorale : Mathématiques, Sciences de l'Information et de l'Ingénieur
(ED 269) Discipline : Informatique

Présentée et soutenue publiquement

par

Robin HECKENAUER

Le 11 Juillet 2023

Apprentissage Profond pour l'Analyse d'Images Histopathologiques

Sous la direction de Prof. Michel HASSENFORDER

Jury

Rapporteur	Prof. Mustapha LEBBAH, Université Paris-Saclay
Rapporteur	Dr. HDR. Camille KURTZ, Université Paris Cité
Examinatrice	Dr. HDR. Aurélie BERTAUX, Université de Bourgogne
Examineur	Prof. Bertrand KERAUTRET, Université Lumière Lyon 2
Directeur de thèse	Prof. Michel HASSENFORDER, Université de Haute-Alsace
co-Directeur de thèse	Prof. Pierre-Alain MULLER, Université de Haute-Alsace
co-Directeur de thèse	Prof. Germain FORESTIER, Université de Haute-Alsace
Encadrant de thèse	Dr. HDR. Jonathan WEBER, Université de Haute-Alsace

UNIVERSITÉ DE HAUTE-ALSACE

Résumé

Apprentissage Profond pour l'Analyse d'Images Histopathologiques

La découverte des cellules a permis de développer le concept de pathologie cellulaire, qui dit qu'un diagnostic peut être posé en examinant des cellules. Aujourd'hui, l'analyse microscopique de tissu est utilisée quotidiennement dans des routines de travail des pathologistes pour établir des diagnostics de différentes pathologies sur tous types d'organes. Étant donné que les tissus prélevés chez les patients sont souvent conservés pour tracer l'évolution de la pathologie, chaque année, le nombre de tissus stockés ne cesse d'augmenter. Dans les années 1990, l'émergence des scanners de lames capables de numériser des tissus pour obtenir des images histopathologiques (WSI), facilite notamment le partage et le stockage des tissus des patients. Ainsi, de nombreuses bases de données d'images histopathologiques se sont constituées. Dès lors, des méthodes automatiques d'analyse d'images ont été proposées pour détecter des structures biologiques d'intérêts. Parmi ces méthodes on retrouve celles issues du domaine de l'intelligence artificielle (IA) et plus particulièrement les approches d'apprentissage profond (DL). L'utilisation de ces méthodes a le potentiel d'aider les pathologistes dans leurs routines de diagnostic sur tout type d'organe et tout type de pathologie, mais également de faire avancer notre compréhension des pathologies.

Toutefois, les challenges sont nombreux. L'un d'entre eux, est de pouvoir assister en temps réel les pathologistes lors de leurs séances d'analyse microscopique. Dans ce but, nous proposons une méthode basée sur YOLOv3 capable de détecter en temps réel des glomérules, qui sont une structure particulière dans le rein, dans des cas de rejet de greffe. Lorsque l'on souhaite détecter des structures biologiques à différents niveaux de grossissement, comme c'est le cas, quand l'on veut suivre les déplacements et les zooms d'un pathologiste lors d'une analyse microscopique, il faut que les méthodes soient robustes aux changements de niveaux. Pour résoudre ce problème, nous étudions une grande plage de différents niveaux de grossissement pour identifier les niveaux dans lesquels les caractéristiques liées au cancer du sein sont les plus intéressants, et pour proposer des approches robustes aux changements de niveaux.

Dans les images histopathologiques, il existe des structures qui sont interprétables par les méthodes d'apprentissage profond, mais pas par les pathologistes. C'est le cas des mutations génétiques. Pourtant, connaître les mutations à l'origine d'un cancer peut permettre aux médecins de diriger les patients vers des traitements adaptés, et ainsi augmenter les chances de survie. L'utilisation systématique de méthode automatique pour trouver les mutations peut économiser l'utilisation de séquenceur ADN dans les pays en voie de développement, ou peut permettre de réserver l'utilisation des séquenceurs pour des tâches plus importantes dans les pays développés. C'est pourquoi nous explorons l'utilisation de méthodes d'apprentissage profond pour la détection de la mutation KRAS G12C dans le cancer du côlon.

UNIVERSITY OF HAUTE-ALSACE

*Abstract***Deep Learning for Histopathological Image Analysis**

The discovery of cells led to the development of the concept of cellular pathology, which states that a diagnosis can be made by examining cells. Today, microscopic analysis of tissue is used daily in the work routines of pathologists to establish diagnosis of various pathologies in all types of organs. Since tissues collected from patients are often stored to follow the evolution of the pathology, every year the number of stored tissues is increasing. In the 1990s, the emergence of slide scanners capable of scanning tissues to obtain whole slide images (WSI), in particular, facilitated the sharing and storage of patient tissues. Thus, many databases of histopathological images have been created. Since then, automatic image analysis methods have been proposed to detect biological structures of interest. Among these methods, we find those from the field of artificial intelligence (AI) and more especially deep learning (DL) approaches. The use of these methods has the potential to assist pathologists in their diagnosis routines on any type of organ and any type of pathology, but also to improve our understanding of pathologies.

However, the challenges are numerous. One of them is to be able to assist pathologists in real time during their microscopic analysis sessions. For this purpose, we propose a method based on YOLOv3 able to detect in real time glomeruli, which are a particular structure in the kidney, in cases of graft rejection. When one wants to detect biological structures at different magnification levels, as it is the case when we want to follow the movements and zooms of a pathologist during a microscopic analysis, the methods must be robust to changes in levels. To address this problem, we study a wide range of different magnification levels to identify the levels in which breast cancer-related features are most interesting, and to propose approaches that are robust to level changes.

In histopathological images, there are structures that are interpretable by deep learning methods, but not by pathologists. This is the case of genetic mutations. However, knowing the mutations that cause a cancer can help doctors to guide patients towards the right treatment, and thus increase the chances of survival. The systematic use of automatic methods to find mutations could save the use of DNA sequencers in developing countries, or could save the use of sequencers for more important tasks in developed countries. Therefore, we are exploring the use of deep learning methods for the detection of the KRAS G12C mutation in colon cancer.

Remerciements

D'abord, je tiens à remercier mon directeur de thèse, Michel Hassenforder, ainsi que mon co-directeur Pierre-Alain Muller, pour m'avoir fait confiance en acceptant de diriger ma thèse.

Je remercie tout particulièrement mon co-directeur Germain Forestier et mon encadrant Jonathan Weber pour m'avoir proposé ce sujet de thèse. Ils m'ont apporté leur soutien sans défaut et m'ont guidé contre vents et marées tout au long de ma thèse.

Merci à Cédric Wemmert pour ses précieux conseils dans le cadre des projets Sys-MIFTA et AiCOLO.

Je remercie également mes anciens et actuels collègues stagiaires, dodoctorants, post-doc, ingés et autres du laboratoire IRIMAS pour tous ces moments de partage autour de la recherche (*et pas seulement ...*) et notamment : Hassan, Mounir, Baptiste, Bastien, Milos, Tsega, Gautier, Ali, Maxime.

Une pensée pour ceux que j'ai embarqués malgré eux dans cette aventure, ma famille et mes amis.

Mes remerciements les plus sincères vont à ma famille, ma belle-famille et à ma compagne Léa pour le soutien indéfectible dont ils ont fait preuve à mon égard pendant ces longues années d'études. Vous êtes les premiers coupables de ma réussite.

Je remercie mes amis les plus chers, Pascal, Victor, Jules, Lucas et Nicolas, sans qui j'aurai sans doute terminé ma thèse plus tôt.

Enfin, je remercie chaleureusement tous ceux qui m'ont accompagné de près ou de loin durant ces années passées au laboratoire IRIMAS.

Table des matières

Résumé	iii
Abstract	v
Remerciements	vii
Introduction	1
1 Pathologie numérique et apprentissage profond	3
1.1 Pathologie numérique	3
1.1.1 Pathologie	3
1.1.2 Pathologie microscopique	4
1.1.3 Pathologie moderne	4
1.1.4 Pathologie numérique	4
1.2 L’intelligence artificielle	5
1.2.1 De l’imaginaire au domaine de recherche	5
1.2.2 Les débuts de l’apprentissage profond	6
1.3 Méthodologie	6
1.3.1 Les étapes de production de données médicales	6
1.3.2 Les difficultés d’acquisition des données médicales	7
1.3.3 La préparation du jeu de données	8
1.3.4 L’apprentissage profond appliqué aux WSI histopathologiques	11
2 Exploitation des différents niveaux de grossissement des images de lames	17
2.1 Introduction	17
2.2 Classification de tumeurs dans le cancer du sein	18
2.2.1 Matériels et méthodes	18
2.2.2 Étude d’une plage de niveaux de grossissement	20
2.2.3 Comparaison de méthodes robustes aux changements de niveaux de grossissement	22
2.3 Détection de glomérules dans le rein en temps réel	24
2.3.1 Le projet Sys-MIFTA	24
2.3.2 Matériels et méthodes	25
2.3.3 La détection de glomérules à différents niveaux de grossissement	28
2.4 Conclusion	31
3 Détection de mutations génétiques dans le cancer du côlon	33
3.1 Introduction	33
3.1.1 Le projet AiCOLO	33
3.1.2 Les mutations génétiques	33
3.1.3 Travaux similaires	36
3.2 Matériels et méthodes	37

3.2.1	Matériels	37
3.2.2	Méthodes	45
3.3	Expérimentations	46
3.3.1	Apprentissage profond pour la classification de la mutation du gène KRAS	47
3.3.2	Apprentissage profond pour la classification du sous-type G12C du gène KRAS	51
3.3.3	Diagnostic assisté par ordinateur du sous-type G12C	51
3.4	Vers l'explicabilité des mutations génétiques	52
3.4.1	Réflexion sur l'explicabilité et l'interprétabilité des mutations génétiques	52
3.4.2	La mutation KRAS G12C vue par CAM et LIME	55
3.5	Conclusion	56
	Conclusion Générale	57
	Conclusion	57
	Contributions	58
	Perspectives	58
	Liste des communications	61
	Article	61
	Communications internationales	61
	Communications nationales	61
	Divers	62
	Financements	63
	Bibliographie	65

Table des figures

1.1	Tendance de la pathologie numérique. Occurrence historique des termes <i>digital pathology</i> dans les articles universitaires sur Google Scholar, ces 50 dernières années (STROBEL, 2018).	5
1.2	Exemple classique des étapes de production des WSI. Dans cet exemple, un patient (1.) subit une biopsie colorectale (2.). Le tissu est plongé dans une solution fixatrice (3.), puis est déshydraté avant d'être emprisonné dans de la paraffine qui est solidifiée pour obtenir un bloc contenant le tissu (4.). Le bloc est découpé en fines tranches à l'aide d'un microtome (5.). Ensuite, un colorateur de lame colore le tissu (6.) avant d'être numérisé pour obtenir une WSI (7.). Enfin, les pathologistes annotent les WSI (8.).	6
1.3	Exemples des différents types d'annotation.	7
1.4	Création des patchs dans une zone annotée. La méthode de la fenêtre glissante crée des patchs dans les zones d'annotations des WSI.	8
1.5	Exemples de normalisation de la coloration. Les méthodes Macenko et Vahadane sont utilisées sur un patch coloré en H&E. (source : https://pathml.readthedocs.io/en/latest/examples/link_stain_normalization.html).	9
1.6	Exemples d'augmentation des données. Le patch original (a) subit un retournement horizontal (b), un retournement vertical (c), une rotation (d), un cisaillement (e) et un zoom (f).	9
1.7	Exemple de validation croisée à cinq blocs ($K=5$). D'abord, les patchs sont répartis dans 5 blocs. Ensuite, cinq modèles sont entraînés pour tester chacun des blocs. Ainsi, après agrégation, l'ensemble du jeu de données est testé.	10
1.8	Réseau de neurones simplifiés.	11
1.9	Détail du calcul d'un neurone.	11
1.10	Impact du taux d'apprentissage sur la descente de gradient.	13
1.11	Impact de la taille de lot sur la descente de gradient.	13
1.12	Exemples de sous-apprentissage et de sur-apprentissage en fonction de la complexité d'un modèle.	14
1.13	Exemple de réseau de neurones convolutifs. Des couches de convolutions extraient des cartes de caractéristiques d'une image donnée en entrée du réseau. Ces caractéristiques sont utilisées par un classifieur pour calculer une probabilité d'appartenance à chaque classe.	14
1.14	Exemples de cartes de caractéristiques obtenues par convolution de filtres horizontaux et verticaux sur un patch.	15
1.15	Détail du calcul d'une convolution sur un patch. Le filtre parcourt le patch d'un pas fixe, de gauche à droite et de haut en bas. À chaque déplacement, une convolution est réalisée entre le filtre et les pixels du patch concernés. Le résultat devient la valeur d'un pixel de la carte de caractéristique.	16

1.16	Détail d'une convolution sur un patch RGB.	16
2.1	Création d'un patch d'un niveau intermédiaire 28X à partir d'un patch de niveau de grossissement 40X. Le patch 40X est découpé puis mis à l'échelle avec une interpolation bicubique pour obtenir un patch 28X.	19
2.2	Exemples de patchs créés à différents niveaux de grossissement à partir des patchs du jeu de données BreakHis. Le jeu de données BreakHis contient des patchs 40X, 100X, 200X et 400X. Des patchs de 20 autres niveaux de grossissement intermédiaire sont créés à partir des patchs originaux de BreakHis.	20
2.3	Carte de chaleur des précisions des réseaux Inception-ResNet-v2. Indique les précisions (%) obtenues par des réseaux entraînés testés individuellement sur tous les jeux de données dans une plage de niveaux de grossissement de 20X à 400X. Plus un réseau obtient une précision élevée sur un jeu de données, plus la case est sombre, et inversement.	21
2.4	Comparaison des approches. Comparaison des précisions (%) de trois approches de prédiction de tumeurs sur différents niveaux de grossissement : Entraîner des réseaux sur plusieurs niveaux ; Entraîner des réseaux sur chaque niveau ; Entraîner des réseaux sur un niveau intermédiaire.	23
2.5	L'architecture YOLOv3 avec Darknet-53 comme <i>backbone</i> (i.e. réalise l'extraction de caractéristiques). L'architecture est composée de 106 couches entièrement convolutives.	27
2.6	La première ligne (1) montre des exemples de patchs d'une taille de 736 x 1280 pixels pris à différentes échelles. La deuxième ligne (2) montre la vérité terrain associée avec en noir le tissu et en blanc les glomérules (les colonnes (a) à (d) sont respectivement à un grossissement de 5X, 10X, 20X et 40X).	28
2.7	Exemples de détections réalisées par l'un de nos modèles sur des patchs à partir d'une WSI colorée en PAS. Nous dessinons une ligne pointillée noire pour montrer la limite des patchs. Les encadrés verts indiquent les prévisions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.	29
2.8	Exemples de détections réalisées par le réseau entraîné et testé sur des images grossis 5X. Les images sont fusionnées pour reconstruire la WSI d'une biopsie. Les boîtes de délimitation vertes montrent les prédictions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.	30
2.9	Exemples de détections réalisées par le réseau entraîné et testé sur des images grossis 5X. Les images sont fusionnées pour reconstruire la WSI d'une néphrectomie du patient 15. Les boîtes de délimitation vertes montrent les prédictions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.	30
3.1	Du corps humain au gène KRAS. Aperçu d'un colon, d'une biopsie, d'une cellule, d'un chromosome et d'un gène KRAS à leurs échelles.	34

3.2	Notre méthode. Une collection de WSI de cancer colorectal contenant des mutations KRAS et WT (1) est annotée par tumeur (2). Nous utilisons une fenêtre glissante (3) pour créer des patchs de taille 224×224 (4). Ces patchs sont organisés en cinq blocs (5) pour effectuer une validation croisée. Ensuite, nous entraînons un réseau de neurones profonds (6) pour prédire la mutation et le sous-type (7). Enfin, nous utilisons les prédictions pour trouver le diagnostic (8).	38
3.3	Exemples de régions annotées des différentes classes sur les images histopathologiques du jeu de données AiCOLO. Ces annotations des classes graisse, stroma, immunitaire, déchets, tumeur, nécrosé, arrière-plan et tissu ont été réalisées manuellement avec Cytomine par des experts en pathologie. Source : (HAMIDA et al., 2021).	39
3.4	Les différentes étapes pour créer un masque de vérité terrain en utilisant le logiciel QuPath. Une zone tumorale sélectionnée est annotée (a) à l'aide de l'outil baguette (b). On élimine les fragments et les trous (c), puis on simplifie la forme de l'annotation (d) en réduisant le nombre d'arêtes du polygone. Enfin, une image et son masque de vérité terrain associé de la zone tumorale (e) sont créés.	41
3.5	Exemples de cas problématiques rencontrés lors du processus d'annotation. Le tissu (a) est flou. (b) contient de la poussière. Dans le tissu (c) on voit une variation dans la coloration. (d) présente des déchirures verticales.	41
3.6	Exemple de patchs découpés à l'intérieur de l'annotation. Une fenêtre glissante découpe les patchs dans le tissu (a), tout en respectant la zone de vérité de base donnée par le masque (b). (c) est une visualisation de tous les patchs créés.	42
3.7	Les histogrammes RVB des WSI avant (a) et après (b) la normalisation de la coloration avec Macenko. Chaque courbe représente la moyenne des histogrammes pour tous les patchs d'une WSI. Les courbes bleues représentent les WSI qui appartiennent à la classe WT et les courbes orange sont les WSI de la classe KRAS.	43
3.8	Exemples d'augmentation des données. Le patch original (a) subit un retournement horizontal (b), un retournement vertical (c), une rotation (d), un cisaillement (e) et un zoom (f).	43
3.9	Le nombre de patchs par classe (b) et par WSI (a). Le bleu correspond à la classe WT, le vert à la classe KRAS.	44
3.10	Le nombre de patchs par classe (b) et par WSI (a). Le bleu correspond à la classe WT, le vert à la classe G12C et le beige aux autres sous-types.	44
3.11	Nombre de patchs bien et mal classés par WSI. (a) montre les prédictions de patchs pour la tâche de classification KRAS effectuée par MobileNetV2, et (b) est pour la tâche de classification du sous-type G12C. Les patchs bien classés sont représentés en clair tandis que les patchs mal classés sont en foncé.	46
3.12	Tendance de l'explicabilité. Occurrence historique des termes <i>explainability</i> dans les articles universitaires sur Google Scholar depuis les années 2000.	54
3.13	Exemple de résultats de CAM (a) et de LIME (b) sur les prédictions du réseau. Les zones surlignées montrent où le réseau se concentre pour classer KRAS G12C.	55

- 3.14 Prédiction de dépôts de mucine comme étant des variants de la mutation KRAS. Les réseaux confondent des dépôts de mucine avec des structures biologiques appartenant à des variants (G12D, G12V, G13D etc). À gauche, le tissu histopathologique. À droite, la zone tumorale annotée. Au milieu, les prédictions du réseau, avec en bleu, les prédictions de type-sauvage, en vert, les prédictions du variant G12C et en beige, les prédictions des autres variants de la mutation KRAS. 56

Liste des tableaux

2.1	Le nombre de glomérules par WSI et par bloc.	25
2.2	La distribution des blocs dans les divisions.	25
2.3	Nombre d'images dans chaque jeu de données avant et après l'augmentation des données.	28
2.4	Résultats des F1-scores pour YOLOv3 entraîné et testé sur un jeu de données à différentes échelles.	29
3.1	Nombre de WSI par mutation. Le nombre et le pourcentage de WSI par mutation BRAF, KRAS, NRAS et WT.	40
3.2	Nombre de WSI par variant. Le nombre et le pourcentage de WSI par variant (G12C, G12D, G12V etc).	40
3.3	Nombre de WSI annotées. Le nombre et le pourcentage de WSI par mutation (WT, KRAS) et par variant (G12C, Autres).	41
3.4	Les blocs par jeu. Répartition des blocs dans les jeux d'entraînement, de validation et de test pour chaque réseau de la validation croisée.	45
3.5	Comparaison des architectures pour la classification de la mutation KRAS. Comparaison de la précision des architectures sur la tâche de classification des patches avec ou sans mutations du gène KRAS.	50
3.6	Comparaison des architectures pour la classification du sous-type G12C, et pour le diagnostic assisté par ordinateur par patient. Comparaison de la précision des architectures sur la tâche de classification par patch du sous-type G12C du gène KRAS. Un diagnostic assisté par ordinateur par patient est calculé à partir des résultats de la classification par patch.	51

Liste des abréviations

ACM	<i>Authority for Consumers and Markets</i> Autorité pour les consommateurs et les marchés
ACS	<i>American Cancer Society</i> Société américaine du cancer
ADN	Acide Désoxyribonucléique
AR	<i>Augmented Reality</i> Réalité augmentée
ARM	<i>Augmented Reality Microscope</i> Microscope à réalité augmentée
CAM	<i>Class Activation Map</i> Carte d'activation des classes
CNIL	Commission Nationale de l'Informatique et des Libertés
CNN	<i>Convolutional Neural Network</i> Réseau de neurones convolutifs
COCO	<i>Common Objects in Context</i> Objets communs dans leur contexte
CT	<i>Computed Tomography</i> Tomodensitométrie
DARPA	<i>Defense Advanced Research Projects Agency</i> Agence des projets de recherche avancée de défense
DL	<i>Deep Learning</i> Apprentissage profond
FCN	<i>Fully Convolutional Network</i> Réseau entièrement convolutif
GACT	Guanine Adénine Cytosine Thymine
GAN	<i>Generative Adversarial Network</i> Réseaux antagonistes génératifs

GPU	<i>Graphics Processing Unit</i> Unité de traitement graphique
H&E	Héματοxyline et Éósine
HES	Héματοxyline Éósine Safran
IA	Intelligence Artificielle
IF/TA	Fibrose Interstitielle et Atrophie Tubulaire
IHC	Immunohistochimie
IOU	<i>Intersection Over Union</i> Intersection de l'union
LIME	<i>Local Interpretable Model-agnostic Explanations</i> Explications locales interprétables et agnostiques du modèle
NMS	<i>Non-Maximum Suppression</i> Suppression non maximale
NVG	<i>Niveau de Gris</i>
PAS	<i>Periodic Acid-Schiff</i>
RAM	<i>Random Access Memory</i> Mémoire vive
RNN	<i>Recurrent Neural Network</i> Réseau de neurones récurrents
RGPD	Règlement Général sur la Protection des Données
RVB	Rouge Vert Bleu
TILS	<i>tumor-infiltrating lymphocytes</i> lymphocytes infiltrant la tumeur
t-SNE	<i>t-distributed Stochastic Neighbor Embedding</i>
VR	<i>Virtual Reality</i> Réalité virtuelle
WSI	<i>Whole Slide Image</i> Image de la lame
WT	<i>Wild-Type</i> Type sauvage
XR	<i>eXtended Reality</i>

Réalité étendue

YOLO *You Only Look Once*

Introduction

L'émergence des scanners de lames à la fin des années 90 a rendu possible la numérisation des lames de verre en une image à très haute résolution appelée *Whole Slide Images* (WSI). Ainsi, il a été possible de conserver, partager et analyser des images de tissus humain de différentes pathologies provenant de différents organes. Dès lors, des approches d'analyse d'image automatique ont été proposées pour détecter des structures biologiques d'intérêt. Parmi ces méthodes, on trouve les approches par apprentissage profond, issues du domaine de l'intelligence artificielle (IA), qui ont montré leur supériorité sur les tâches classiques de vision par ordinateur. Or, les approches d'IA peuvent être appliquées sur les WSI pour détecter, classer et segmenter des structures biologiques et ainsi effectuer différents types de tâches, i.e. identifier des tumeurs dans un tissu. Ces méthodes ont le potentiel d'aider les pathologistes dans différentes tâches comme le diagnostic.

Toutefois, les challenges sont nombreux. Dans cette thèse, nous résolvons des problèmes précis par l'utilisation de méthodes d'apprentissage profond entraînées pour détecter des structures biologiques d'intérêts sur des images histopathologiques. Ces travaux s'intègrent notamment dans les projets Sys-MIFTA et AiCOLO, qui réunissent des experts en pathologie et des experts en IA sur des thématiques respectivement liées aux rejets de greffe dans le rein et de cancer du côlon.

Avant de présenter nos travaux, nous introduisons la pathologie numérique dans le chapitre 1.1, puis l'intelligence artificielle dans le chapitre 1.2. Ensuite, dans le chapitre 1.3, nous présentons les usages standards en matière d'analyse de WSI par apprentissage profond, depuis la production des images de lames, jusqu'à la création de modèles pour une pathologie.

Les WSI sont des fichiers particuliers qui peuvent contenir des représentations des tissus à plusieurs niveaux de grossissement. De ce fait, ces images sont particulièrement volumineuses et nécessitent des outils dédiés pour leurs manipulations. C'est à la fois une contrainte quand il s'agit d'appliquer des méthodes d'analyse, mais c'est également un avantage étant donné que l'on dispose de plusieurs niveaux de détails. Ainsi, on peut fortement zoomer et dézoomer dans les tissus pour apprécier les structures biologiques à différentes échelles. Cependant, en pratique, les travaux de recherche actuelles qui appliquent des algorithmes d'apprentissage profond pour l'analyse d'image médicale se concentrent généralement sur un seul niveau de grossissement tel que 20X ou 40X. Autrement dit, tous les autres niveaux présents dans les WSI sont inutilisés, quand bien même ils contiendraient des informations supplémentaires. En effet, observer des tissus à des niveaux supérieurs ou inférieurs aux niveaux traditionnellement utilisés permet par exemple de faire apparaître de nouvelles frontières et d'ajouter ou enlever du contexte dans les tissus. De plus, exploiter différents niveaux de grossissement peut rendre les méthodes robustes aux changements d'échelle. Cela est particulièrement utile lorsque l'on souhaite créer des méthodes capables de suivre les mouvements d'un pathologiste dans son analyse microscopique afin de l'enrichir d'informations en temps réel.

Dans le chapitre 2, nous proposons d'étudier des images de cancer du sein à différents niveaux de grossissement à l'aide du réseau de neurones profond InceptionResNet-v2. Pour cela, nous utilisons le jeu de données BreakHis, qui contient des images histopathologiques de sein avec et sans tumeurs. Ainsi, nous déterminons les niveaux dans lesquels les caractéristiques des tumeurs sont les plus intéressantes. Ensuite, nous utilisons ces résultats pour comparer différentes stratégies à des fins de classification de tumeurs dans le sein, robuste aux changements de niveaux de grossissement et en temps réel.

Dans un second temps, nous présentons nos travaux réalisés dans le cadre du projet européen Sys-MIFTA¹. Nous proposons une preuve de concept en utilisant YOLOv3, une méthode basée sur l'apprentissage profond pour détecter à différents niveaux de grossissement en temps réel des glomérules dans le rein dans le contexte du suivi de greffe. Ensuite, nous explorons différentes stratégies d'entraînement sur différents niveaux de grossissement pour rendre nos réseaux invariants aux changements de niveau.

Dans le chapitre 3, nous traitons du cancer du côlon. Il existe différents types de cancer et différents stades en fonction de l'évolution de la maladie. Connaître le type de cancer est essentiel pour personnaliser le traitement des patients et ainsi augmenter les chances de survie. Dans le cadre d'une analyse microscopique, un pathologiste est capable de déterminer de nombreuses informations sur le cancer et son évolution à partir d'une lame de verre. Toutefois, l'analyse visuelle des tissus n'est pas suffisante pour savoir quels gènes ont mutés et quels variants sont présents. Cette étape est cruciale pour déterminer certains types de cancer et orienter les patients vers le meilleur traitement. Dans le cas du cancer colorectal, 3% des cancers présentent une mutation KRAS G12C. Cette mutation rare joue un rôle clé dans la prolifération cellulaire et est responsable d'une forme de cancer extrêmement agressive. Il existe cependant des traitements qui peuvent améliorer les chances de survie des patients et qui peuvent être prescrits dès l'identification de la mutation. Les informations sur le génome peuvent être obtenues par un séquenceur ADN, mais cela reste long et coûteux. Or, depuis 2019, certains travaux de recherche laissent penser qu'il est possible de déterminer les mutations génétiques directement depuis les WSI à l'aide de méthodes d'apprentissage profond.

Dans le cadre du projet AiCOLO², nous proposons de comparer différentes architectures d'apprentissage profond sur la tâche de classification de la mutation KRAS directement sur des images histopathologiques de cancer colorectal. Ces images sont issues du jeu de données AiCOLO qui contient des WSI de biopsie du côlon ainsi que des informations génétiques. Ensuite, nous classons le variant G12C du gène KRAS compte tenu de son intérêt dans la sélection d'un traitement augmentant les chances de survie des patients. Enfin, nous utilisons les méthodes d'explicabilité CAM (OQUAB et al., 2015) et LIME (RIBEIRO, SINGH et GUESTRIN, 2016) pour visualiser les zones des patchs qui motivent la prise de décision d'un réseau.

1. Sys-MIFTA : <https://sys-mifta.uha.fr/>

2. AiCOLO : <https://wemmertc.github.io/aicololo/>

Chapitre 1

Pathologie numérique et apprentissage profond

1.1 Pathologie numérique

1.1.1 Pathologie

La Pathologie est l'étude du développement des maladies, et l'examen des causes, des symptômes, de l'évolution ainsi que les lésions et les complications éventuelles des maladies (*Le Larousse médical* 2022). Les plus anciennes traces historiques connues de l'étude de cette science remontent au 3ème siècle av. J.-C., où les médecins grecs Hérophile de Chalcédoine et Erasistrate de Chios ont réalisé les premières dissections humaines recensées malgré les tabous de l'époque (VON STADEN, 1992 ; RACE, TILLERY et DYSERT, 2004 ; BAY et BAY, 2010). Durant l'antiquité, d'autres Médecins grecs tels que Herophilus, Hippocrate, Galen et Erasistrate ont également réalisé des autopsies pour caractériser l'évolution des maladies (JACQUES JOUANNA et EIJK, 2012). Bien que d'autres dissections aient été répertoriées en Chine, en Inde et en Perse (BRENNNA, 2022), très peu d'archives avant l'an 1000 nous sont parvenues (VON STADEN, 1992). De plus, entre l'antiquité et la renaissance, l'abandon des pratiques d'autopsie des corps humains a été à l'origine d'une pause de 1800 ans dans l'étude de la Pathologie (BAY et BAY, 2010 ; BRENNNA, 2022). Pendant cette période, les médecins prodiguaient des soins en s'appuyant sur la théorie des humeurs qui est une croyance selon laquelle le sang, la lymphe, la bile jaune et la bile noire correspondent à quatre éléments (le feu, l'air, la terre, l'eau) et doivent être équilibrés pour être en bonne santé (KARPOZILOS et PAVLIDIS, 2004). Malgré ce coup d'arrêt, certains médecins réalisent tout de même des dissections post mortem pour en apprendre davantage sur le corps humain tels que Avenzoar (1091-1161), Ibn al-Nafis (1213-1288), Antonio Benivieni (1443-1502). C'est ainsi qu'apparaissent les premières applications de la méthode scientifique au domaine de la médecine au Moyen-Orient pendant l'âge d'or islamique et en Europe occidentale pendant la Renaissance italienne (VON STADEN, 1992). Il faudra attendre un regain d'intérêt pour les sciences au début du 18ème siècle, pour que Giovanni Morgagni (1682-1771) réalise des observations pathologiques d'environ 700 dissections après autopsie de patients qu'il avait traités de son vivant. Ces travaux ont permis une avancée majeure dans le diagnostic des maladies, dans le protocole de prise en charge patient et dans le pronostic, marquant ainsi le début de la médecine moderne (GHOSH, 2017). Toujours au 18ème siècle, le développement des hôpitaux et l'amélioration de l'éducation médicale permet l'avancée rapide des connaissances en pathologie grâce au meilleur suivi des patients en hôpital et à l'augmentation des autopsies des patients suivis pour comprendre leurs maladies (RACE, TILLERY et DYSERT, 2004). Ces avancées ont permis notamment de réfuter la théorie des humeurs qui prévalait jusqu'alors. Au 19ème siècle, les études du corps humain se

poursuivent avec par exemple le médecin Carl Rokitansky (1804-1878) qui a réalisé environ 20 000 autopsies et en a supervisé près de 60 000 (WEYERS, 2004). Dans le même temps, Théodore Schwann fait une avancée majeure en découvrant des cellules dans des tissus humains (RACE, TILLERY et DYSERT, 2004). Ces avancées sont poussées par des réformes promouvant la pratique en médecine, comme celles adoptées par l'université de Harvard, de Pennsylvanie et du Michigan.

1.1.2 Pathologie microscopique

La découverte des cellules va permettre dans les années 1850 à Rudolf Virchow (1821-1902), père de la pathologie microscopique (VON STADEN, 1992; SCHULTZ, 2008), de développer le concept de pathologie cellulaire qui dit qu'un diagnostic peut être posé en examinant les cellules (RACE, TILLERY et DYSERT, 2004). Ce sont les prémisses de l'histopathologie, qui encore aujourd'hui est utilisée en routine pour poser des diagnostics cliniques.

1.1.3 Pathologie moderne

À partir de 1931, des microscopes électroniques permettent d'obtenir des grossissements jusqu'à 5 000 fois plus élevés comparés aux microscopes traditionnels (RUSKA, 1987). Dans le même temps, l'immunohistochimie (IHC) fait son apparition, permettant ainsi d'identifier des protéines dans les cellules et les tissus (KABIRAJ et al., 2015).

1.1.4 Pathologie numérique

Dans les années 1960, la popularisation d'internet et l'essor du domaine de la télécommunication (HUURDEMAN, 2003) rendent possible la télépathologie (pathologie à distance) (FARAHANI et PANTANOWITZ, 2015). L'apparition de scanners capables de numériser des lames de verre en image haute résolution va révolutionner le domaine. En 1994, le premier scanner de lames commerciales arrive sur le marché (PANTANOWITZ et al., 2018). Il faudra attendre plusieurs années afin que cette nouvelle technologie gagne en maturité pour que les scanners soient moins chers, pour que les systèmes informatiques soient plus rapides et pour stocker facilement de grandes quantités de données. Dès lors, on assiste à un engouement autour du domaine de la pathologie numérique comme le montre la figure 1.1¹. L'apparition des scanners de lames et la possibilité de numériser les tissus apportent de nombreux avantages (PANTANOWITZ et al., 2018; NIAZI, PARWANI et GURCAN, 2019; DENG et al., 2020). D'abord, il est maintenant possible de conserver sans dégradation cellulaire des images des tissus des patients. Il devient alors possible de comparer l'avancée des maladies. Ensuite, les données numérisées peuvent être plus facilement stockées et partagées contrairement à des échantillons de tissu. Ainsi, le partage d'images de tissus facilite à la fois les tâches d'évaluation par les pairs et d'enseignement. Enfin, le stockage massif des images de tissus permet de constituer des bases de données de différentes pathologies, qui peuvent ensuite servir à des fins d'éducation, de consultation et de recherche. À partir du moment où il a été possible d'avoir des représentations numérisées de tissus, des méthodes d'analyse automatique ont été appliquées (DUNCAN et AYACHE, 2000). Parmi ces approches, on trouve notamment celles basées sur l'intelligence artificielle.

1. Sur la figure 1.1, la baisse d'intérêt pour la *digital pathology* à partir de 2014 est dû à l'utilisation de termes plus précis à mesure que le domaine évolue et à l'ajout tardif d'articles sur Google Scholar.

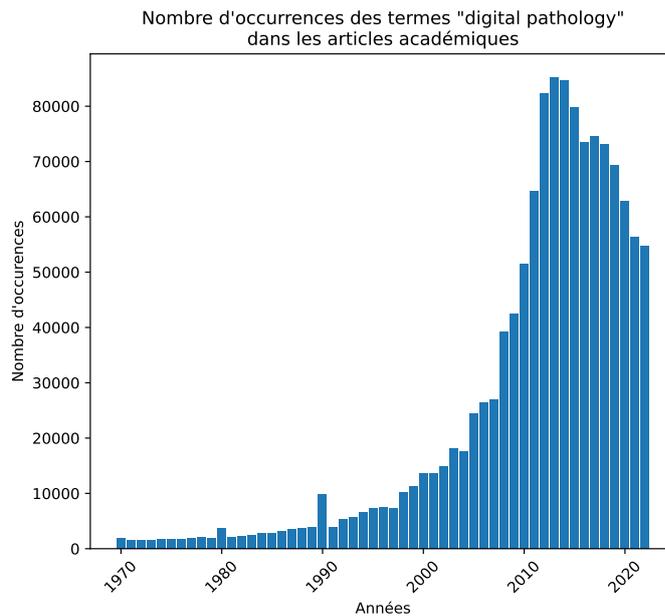


FIGURE 1.1 – Tendance de la pathologie numérique. Occurrence historique des termes *digital pathology* dans les articles universitaires sur Google Scholar, ces 50 dernières années (STROBEL, 2018).

1.2 L'intelligence artificielle

1.2.1 De l'imaginaire au domaine de recherche

L'intelligence artificielle est un terme général qui implique l'utilisation d'un ordinateur pour modéliser un comportement intelligent avec une intervention humaine minimale (HAMET et TREMBLAY, 2017). Cependant, le terme est controversé comme détaillé dans (MIMS, 2021) car les algorithmes que l'on appelle IA n'ont pas réellement de comportement intelligent, il est plus juste de dire qu'ils sont capables de traiter des informations complexes comme proposé en 1956 dans (NEWELL et SIMON, 1956). Ce concept a d'abord été développé dans l'imaginaire des auteurs et des inventeurs. Des auteurs de toutes les époques ont imaginé des automates animés par des comportements intelligents tels que : Homère dans L'Iliade (8ème siècle av. J.-C.) (POPE et SHANKMAN, 2009), Thomas Hobbes dans Léviathan (1651) (HOBBS et MISSNER, 2016) ou encore Isaac Asimov dans ses nombreux ouvrages de science-fiction sur les robots (1950-1993) (ASIMOV, 2004). Des inventeurs ont également tenté de créer des automates tels que Léonard de Vinci et ses croquis de chevalier robot capable de réaliser des actions simples (1495), ou Jacques de Vaucanson et son automate de canard (1738) (NILSSON, 2009). La recherche scientifique en intelligence artificielle ne débute réellement qu'à partir de 1956 à la suite d'une conférence au Dartmouth College (DICK, 2019). Des successions de phases d'engouement, de désillusion et d'arrêt de financement ont fait évoluer le domaine par à-coup. Deux périodes nommées "hiver de l'IA" (en 1973, puis en 1987) sont notamment célèbres pour leur perte d'intérêt de la part des investisseurs et d'une partie de la communauté scientifique. En 1997, l'IA Deep Blue d'IBM bat le champion du monde d'échecs de l'époque Garry Kasparov, et replace le domaine sur le devant de la scène (NILSSON, 2009).

1.2.2 Les débuts de l'apprentissage profond

Dans les années 2010, l'association de plusieurs avancées antérieures mènent au développement de l'apprentissage profond (*deep learning*). D'abord, en 1957, Frank Rosenblatt crée le premier algorithme d'apprentissage supervisé de classificateurs binaires nommé perceptron permettant de résoudre des problèmes linéaires. Il faudra attendre 1986 pour que David Rumelhart mette au point un perceptron multicouche qui utilise la méthode de rétropropagation du gradient pour résoudre des problèmes non linéaires. Ensuite, des bases de données suffisamment volumineuses pour des tâches d'apprentissage ont commencé à voir le jour (LECUN et al., 1998). Enfin, l'évolution rapide des performances de calcul des machines (SHALF, 2020) a rendu possible le traitement de ces grandes quantités de données pour entraîner les modèles d'apprentissage profond. Ainsi, ces dernières années, de nombreuses approches basées sur l'IA ont été proposées pour modéliser des problèmes, notamment liés au domaine médical.

1.3 Méthodologie

De nombreuses étapes complexes sont nécessaires pour produire des lames de verre histopathologiques comme montré en figure 1.2 avant de pouvoir les analyser ou de les utiliser dans des approches basées sur l'IA.

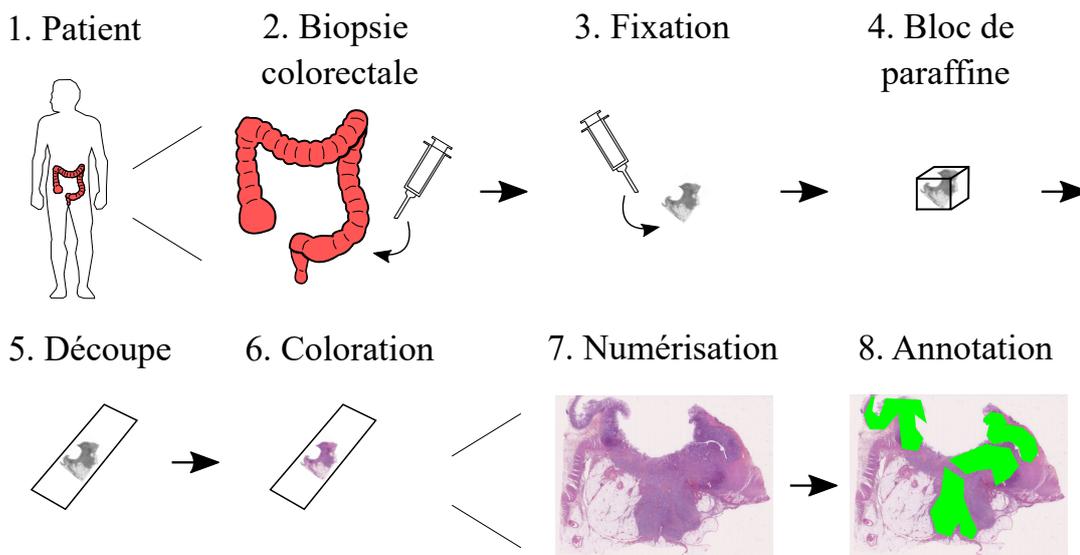


FIGURE 1.2 – Exemple classique des étapes de production des WSI. Dans cet exemple, un patient (1.) subit une biopsie colorectale (2.). Le tissu est plongé dans une solution fixatrice (3.), puis est déshydraté avant d'être emprisonné dans de la paraffine qui est solidifiée pour obtenir un bloc contenant le tissu (4.). Le bloc est découpé en fines tranches à l'aide d'un microtome (5.). Ensuite, un colorateur de lame colore le tissu (6.) avant d'être numérisé pour obtenir une WSI (7.). Enfin, les pathologistes annotent les WSI (8.).

1.3.1 Les étapes de production de données médicales

Tout d'abord, des patients souffrant de la pathologie étudiée sont sélectionnés. Chacun des patients subit une biopsie où un morceau de tissu est prélevé dans l'organe atteint. Suite à la biopsie, les cellules commencent à mourir et à relâcher des enzymes

qui décomposent les tissus. On appelle ce phénomène l'autolyse. Afin d'empêcher la détérioration des tissus et l'autolyse, le tissu est placé dans une solution fixatrice. Puis, le tissu est progressivement déshydraté, dans un processus où l'eau est remplacée par de la cire (e.g. paraffine). La cire est ensuite solidifiée pour obtenir un bloc contenant le tissu. Puis, un microtome (appareil utilisé pour découper des blocs de paraffine en fines tranches) est utilisé pour découper une tranche de quelques micromètres (μm), qui sera placée sur une lame de verre. À cette étape, le tissu sur la lame de verre peut déjà être observé avec un microscope, cependant le tissu apparaîtra transparent. L'immunohistochimie (IHC) permet de mettre en lumière des éléments d'intérêt en les colorants, à l'aide d'un appareil appelé colorateur de lames. Il existe de nombreux colorants qui permettent de localiser différentes structures (ALTURKISTANI, TASHKANDI et MOHAMMEDSALEH, 2016). Par exemple, l'hématoxyline et l'éosine (H&E), qui est la coloration la plus utilisée en histopathologie, permet de colorer les noyaux en bleu/violet et les muscles en rose foncé. Ensuite, la lame de verre est numérisée à l'aide d'un scanner de lames pour obtenir une image haute résolution appelée image de lame (WSI), généralement avec un niveau de grossissement de X20 ou X40. Les WSI font plusieurs dizaines de milliers de pixels par côté et nécessitent l'utilisation de formats de fichier spécialisés tels que NDPI ou SVS. Enfin, les WSI sont annotées à partir des connaissances expertes des pathologistes en prévision des futures analyses automatiques des images. Il existe différents types d'annotations comme montré en figure 1.3 qui sont plus ou moins chronophage et apportent différents niveaux de quantité d'informations.

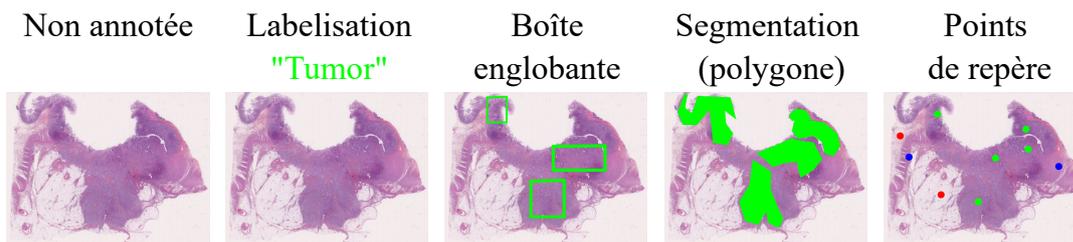


FIGURE 1.3 – Exemples des différents types d'annotation.

1.3.2 Les difficultés d'acquisition des données médicales

De nombreuses difficultés entravent l'acquisition de jeux de données de WSI de bonne qualité et suffisamment volumineux pour être utilisés par des algorithmes d'apprentissage. D'abord, comme présenté dans la partie précédente, beaucoup d'étapes sont nécessaires pour obtenir des WSI à partir de biopsies. Ces tâches redondantes et coûteuses requièrent des connaissances expertes dans l'utilisation de machines spécialisées, la manipulation de lames de verre ou encore en histopathologie. Aussi, il faut garder à l'esprit qu'un jeu de données reflète partiellement une application et que les conditions cliniques d'acquisition sont différentes d'un laboratoire à un autre. Ces effets de bord ajoutent des biais aux jeux de données. De manière générale, les biais apparaissent quand les données utilisées pour entraîner un modèle ont une distribution différente des données utilisées pour tester ce même modèle. De plus, il arrive que les pathologistes aient des avis différents sur des tissus identiques. Ses différences proviennent du niveau d'expertise, mais également de la complexité de l'étude des structures biologiques en elle-même. Ainsi, des biais peuvent apparaître lors de l'annotation des WSI tels que la segmentation d'éléments non désirés ou un mauvais diagnostic de pathologie. Pour contourner une partie de ces contraintes, il est possible d'exploiter des jeux de données publiques tels que BreakHis (SPANHOL et al.,

2015) pour le cancer du sein ou TCGA (*The cancer genome atlas program s. d.*) pour les mutations génétiques responsables du cancer. Cependant, malgré quelques initiatives, les bases de données publiques dans le domaine médical restent rares. Cette pénurie s'explique par la complexité de production des WSI, mais aussi par d'autres facteurs tels que la difficulté des démarches pour que l'usage des données patients soit conforme à la CNIL, ou encore la volonté de garder les données privées pour publier en priorité dessus. D'une manière similaire, les codes sources des articles dans le domaine sont rarement mis à disposition. L'une des raisons est la création de brevets à partir des travaux de recherche. À cause de ces diverses raisons, dans le domaine de la pathologie numérique, il y a un retard dans le partage de jeux de données et de codes sources comparé à d'autres domaines qui utilisent aussi des IA. Ce retard a pour conséquence d'empêcher la reproductibilité des travaux de recherche.

1.3.3 La préparation du jeu de données

1.3.3.1 Création des patches

Les images histopathologiques sont trop grandes pour être directement envoyées en entrée d'un réseau de neurones. Par conséquent, il faut découper des patches dans les WSI pour réduire la complexité de calcul et permettre l'entraînement dans un délai raisonnable. Il existe différentes méthodes pour découper des patches dans une WSI telles que la sélection aléatoire ou la fenêtre glissante. La méthode de la fenêtre glissante utilise une région rectangulaire qui se déplace d'un pas fixe sur la WSI. À chaque déplacement, un patch est créé à partir du contenu de la région rectangulaire comme montré en figure 1.4.

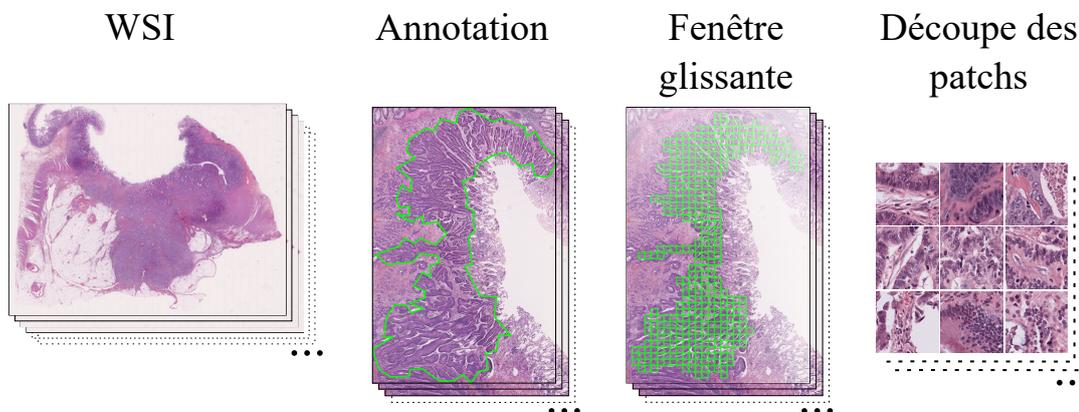


FIGURE 1.4 – Création des patches dans une zone annotée. La méthode de la fenêtre glissante crée des patches dans les zones d'annotations des WSI.

1.3.3.2 Normalisation de la coloration

Lors du processus de coloration des tissus à l'aide d'un colorateur de lames, des différences de couleurs entre les lames apparaissent. Ces différences sont sources de biais dans les jeux de données. Pour atténuer les variations de couleurs qui entraînent des biais et pour permettre la comparaison des lames, il est possible d'appliquer des méthodes de normalisation de couleurs (ROY et al., 2018) telles que la normalisation Macenko (MACENKO et al., 2009) ou Vahadane (VAHADANE et al., 2016), montrée en figure 1.5.

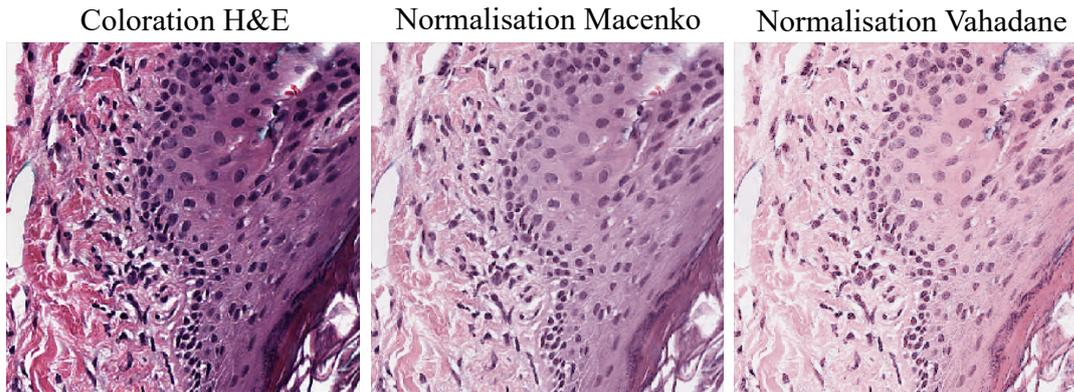


FIGURE 1.5 – Exemples de normalisation de la coloration. Les méthodes Macenko et Vahadane sont utilisées sur un patch coloré en H&E. (source : https://pathml.readthedocs.io/en/latest/examples/link_stain_normalization.html)

1.3.3.3 Augmentation de données

L’augmentation de données (GARCEA et al., 2022) permet d’augmenter artificiellement la taille des données d’entraînement en appliquant des transformations aux données d’origine ou en créant des données artificielles. Ces transformations ont pour effet de lutter contre le sur-apprentissage d’un modèle défini dans la section 1.3.4.1. De plus, dans le cas de données médicales, il est d’usage d’utiliser des transformations qui apparaissent pendant les étapes de production des WSI. Par exemple, lors de la découpe des blocs de paraffine en fines tranches à l’aide d’un microtome, des transformations géométriques peuvent apparaître sur les tissus tels que des cisaillements. D’une autre manière, les tissus peuvent être placés sur les lames de verre avec différents degrés de rotation et peuvent également être positionnés face recto ou verso, ce qui entraîne des rotations et des retournements. En utilisant ces transformations sur les données d’origines, les biais du jeu de données liés à la manipulation des tissus sont réduits. Parmi les méthodes de transformations des données d’origine les plus utilisées en histopathologie, il existe quatre sous-ensembles. Premièrement, les transformations affines, où les distances et les angles sont modifiés tout en conservant les lignes et le parallélisme de l’image (e.g. translation, rotation, retournement, mise à l’échelle, recadrage, cisaillement). Des exemples de telles transformations sont montrées en figure 1.6. Deuxièmement, les transformations par effacement remplacent des régions par du bruit aléatoire. Troisièmement, les transformations élastiques permettent de réaliser des déformations spatiales. Quatrièmement, les transformations de niveau de pixel changent les valeurs des pixels pour modifier la luminosité, le contraste, la saturation et le bruit.

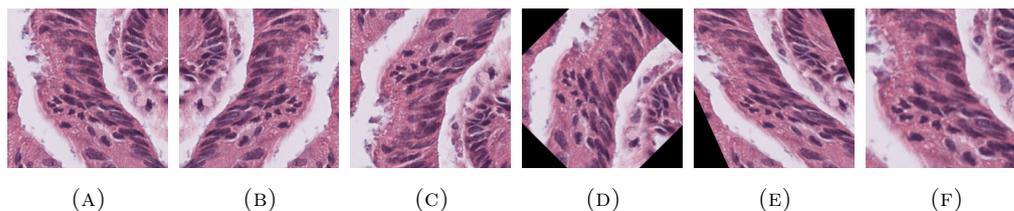


FIGURE 1.6 – Exemples d’augmentation des données. Le patch original (a) subit un retournement horizontal (b), un retournement vertical (c), une rotation (d), un cisaillement (e) et un zoom (f).

Bien que les transformations des données d’origines permettent de réduire les biais d’un jeu de données, il n’est pas possible de généraliser au-delà des données d’origines comme le permettent les méthodes de généralisation de données artificielles. En effet, les données créées artificiellement apportent davantage de variété. Cependant, elles nécessitent des coûts de calcul importants étant donné que ces méthodes sont basées sur des approches d’apprentissage comme c’est le cas avec les réseaux antagonistes génératifs (GAN) (GOODFELLOW et al., 2020).

1.3.3.4 Validation croisée

Pour s’assurer de la robustesse des modèles et de leurs capacités à généraliser, il est possible d’utiliser des méthodes d’estimation de leur fiabilité telle que la validation croisée (STONE, 1978). Un exemple de validation croisée avec cinq blocs ($K = 5$) est montré en figure 1.7. Dans cet exemple, les patches sont distribués entre K blocs. À leur tour, les K blocs sont distribués dans les dossiers *entraînement*, *validation* et *test*, qui sont utilisés pour la phase d’apprentissage et d’évaluation du réseau. Par ailleurs, il est préférable de répartir les patches de chaque classe de manière équitable entre les blocs pour faciliter l’apprentissage du réseau. En effet, des classes mal équilibrées peuvent pousser le réseau à prédire la classe majoritaire au lieu de la tâche souhaitée. Chaque bloc est utilisé une fois en tant que *test* pour évaluer le réseau, les autres blocs servent alors pour la phase d’apprentissage. Ainsi, grâce à la variance des cinq modèles obtenus, on peut mieux évaluer les performances individuelles des modèles.

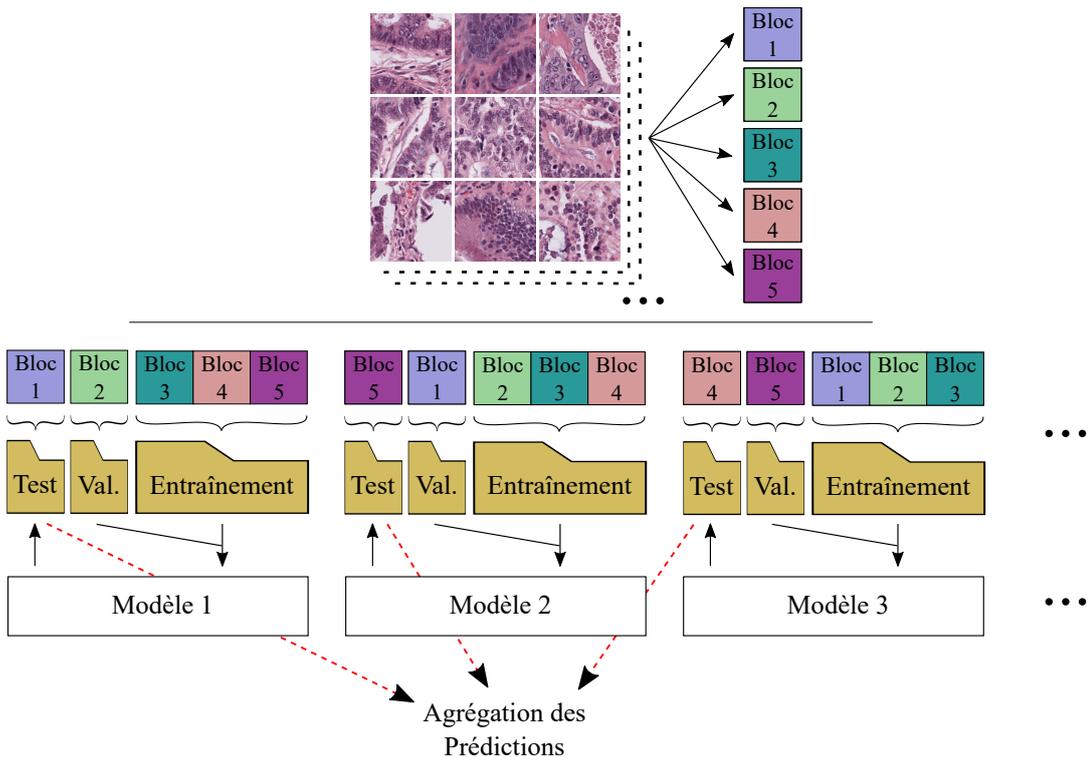


FIGURE 1.7 – Exemple de validation croisée à cinq blocs ($K=5$). D’abord, les patches sont répartis dans 5 blocs. Ensuite, cinq modèles sont entraînés pour tester chacun des blocs. Ainsi, après agrégation, l’ensemble du jeu de données est testé.

À noter que l’utilisation de données médicales amène des contraintes supplémentaires dans l’utilisation de la validation croisée. Effectivement, les données d’un patient ne doivent être présentes que dans un seul bloc pour éviter les phénomènes de fuite

de données, où un modèle est testé sur des données d'un patient qui a déjà été utilisé pendant la phase d'apprentissage.

1.3.4 L'apprentissage profond appliqué aux WSI histopathologiques

1.3.4.1 Réseaux de neurones

Les réseaux de neurones sont inspirés du fonctionnement du cerveau humain où l'on modélise des neurones sous forme de poids et des connexions synaptiques sous forme d'arêtes entre les poids comme montré en figure 1.8. Ainsi, on appelle couche de neurones une série de neurones connectés aux neurones de la couche précédente et de la couche suivante. Les poids du réseau reflètent une modélisation d'un problème. Lorsqu'une donnée est envoyée dans un réseau de neurones, cette modélisation est utilisée pour calculer couche par couche les valeurs des neurones. En sortie de réseau, les neurones sont agrégés afin de prédire au choix une tâche de détection, de classification ou de segmentation.

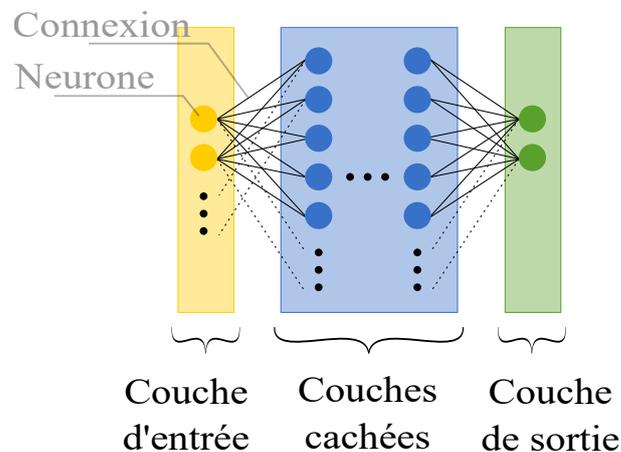


FIGURE 1.8 – Réseau de neurones simplifiés.

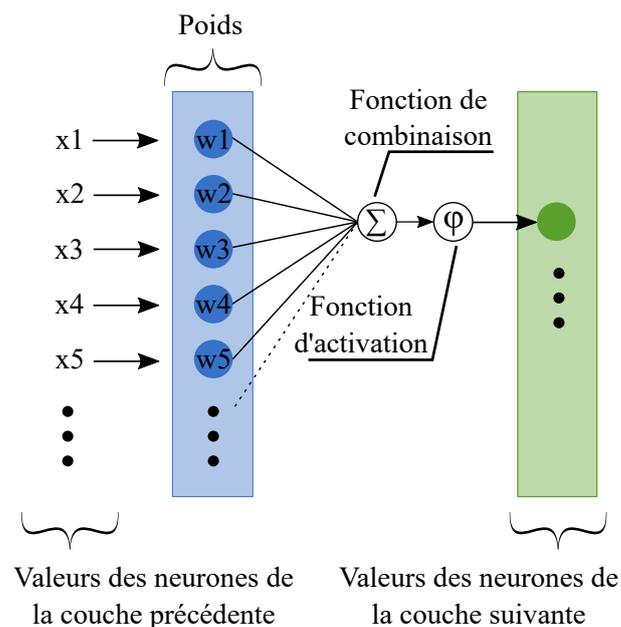


FIGURE 1.9 – Détail du calcul d'un neurone.

Il existe différentes méthodes d'apprentissage telles que l'apprentissage supervisé, non supervisé, par renforcement. Dans le cas de l'apprentissage supervisé, la prédiction du réseau est comparée à la vérité terrain pour améliorer le modèle. À l'inverse, en apprentissage non supervisé, la vérité terrain des données n'est pas disponible. Dans ce cas, l'amélioration du modèle se fait sans prendre en compte la vérité terrain dans les calculs. On trouve également des méthodes d'apprentissage par renforcement où un agent apprend une tâche par l'expérience en optimisant une récompense. Pour calculer la valeur d'un neurone comme montré en figure 1.9, il faut d'abord appliquer une fonction de combinaison présentée dans l'Équation 1.1 qui somme le produit de chaque neurone de la couche précédente x_i par le poids w_i . Puis, le résultat de la fonction de combinaison est utilisé par une fonction d'activation afin de donner la nouvelle valeur de neurone. L'utilisation de fonctions d'activation permet de résoudre des problèmes non linéaires.

$$\text{Fonction de combinaison} = \sum x_i w_i \quad (1.1)$$

Pour modéliser un problème à l'aide d'un réseau de neurones, il faut passer par une étape d'apprentissage pendant laquelle les poids du réseau seront mis à jour dans le but de minimiser une fonction de coût (également appelés perte) détaillée dans l'équation 1.2. La fonction de coût L nous donne une distance entre la prédiction du réseau y et l'objectif réel \hat{y} . En minimisant cette fonction, on tend à améliorer notre modèle sur les données d'apprentissage.

$$\text{Fonction de coût} = L(\hat{y}, y) \quad (1.2)$$

Lors de la rétropropagation du gradient, on utilise une fonction d'optimisation pour corriger les poids du réseau. Il existe de nombreuses fonctions d'optimisation telles que SGD, RMSprop, Adam (CHOI et al., 2019) qui proposent différentes stratégies pour réaliser une descente de gradient. Les hyperparamètres d'un réseau tels que le taux d'apprentissage, le nombre d'époques et la taille des lots permettent d'influer sur la vitesse et la qualité de l'apprentissage. Tandis que les paramètres évoluent pendant la rétropropagation du gradient tels que les poids des neurones. Le taux d'apprentissage est un hyperparamètre qui représente la vitesse à laquelle le réseau remplace ce qu'il a appris par de nouvelles informations. Il est donné par l'utilisateur ou obtenu par optimisation à l'aide de méthodes telles que la recherche exhaustive, la recherche aléatoire, la descente de gradient, l'optimisation bayésienne. (YANG et SHAMI, 2020) La recherche exhaustive consiste à énumérer toutes les combinaisons possibles d'hyperparamètres et à évaluer leur performance. Cette méthode peut être coûteuse en matière de temps de calcul ($O(n^k)$) selon le nombre d'hyperparamètres à optimiser. La méthode de recherche aléatoire consiste à générer aléatoirement des combinaisons d'hyperparamètres, contrairement à la recherche exhaustive, toutes les possibilités ne sont pas testées. Cette méthode est moins coûteuse en temps de calcul ($O(n)$) que la recherche exhaustive, mais il n'est pas garanti que les meilleurs hyperparamètres soient trouvés. En pratique, lors de la minimisation de la fonction de coût, le gradient représente la direction dans laquelle se déplacer dans le champ des solutions, et le taux d'apprentissage contrôle la distance du déplacement. De plus, il est essentiel de choisir correctement le taux d'apprentissage afin de trouver la meilleure solution, le plus rapidement possible comme montré en figure 1.10. En effet, un taux d'apprentissage trop faible change faiblement les poids réseaux et augmente le temps de calcul pour se rapprocher d'une solution. D'une autre manière, un taux trop élevé peut provoquer des divergences.

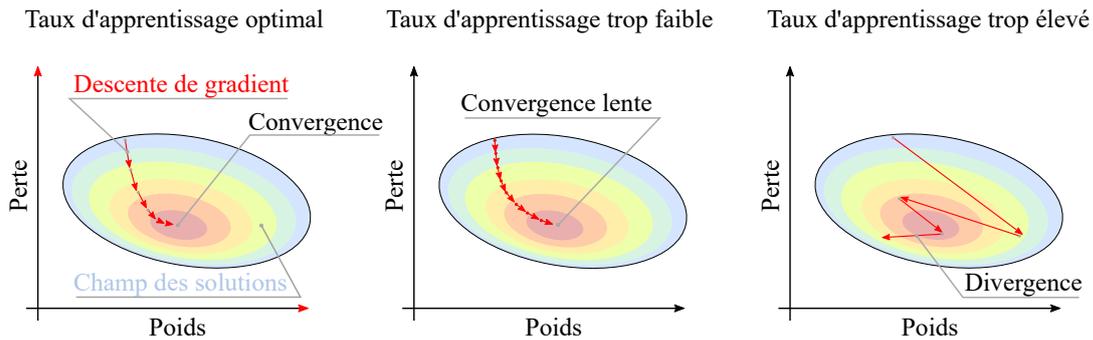


FIGURE 1.10 – Impact du taux d'apprentissage sur la descente de gradient.

Il existe une catégorie de méthodes nommée taux d'apprentissage *annealing* (NAKAMURA et al., 2021) qui proposent de réduire la valeur du taux d'apprentissage progressivement durant l'entraînement. Cela permet d'éviter les minimums locaux et de converger vers une solution globale optimale. On appelle époque le nombre de fois où le réseau va apprendre toutes les données d'entraînement. Pendant une époque les données sont divisées en lots qui alimentent le réseau. C'est après qu'un lot soit traité par le réseau que la rétropropagation est effectuée et que les poids sont corrigés. Une taille de lot trop faible augmente le temps d'entraînement à cause des nombreuses rétropropagations effectuées après chaque lot. De plus, le faible nombre de données dans un lot peut introduire du bruit dans la descente de gradient. Dans le meilleur des cas, ce bruit peut aider à sortir de minimum locaux, mais à l'inverse, il peut aussi empêcher la convergence vers un optimum. D'un autre côté, une taille de lot trop élevée nécessite une mémoire RAM importante. Ensuite, le traitement de nombreuses données avant rétropropagation ne permettra pas au modèle d'être mis à jour suffisamment de fois pour obtenir une bonne modélisation. Des exemples de descentes de gradient avec différentes tailles de batch sont montrés en figure 1.11.

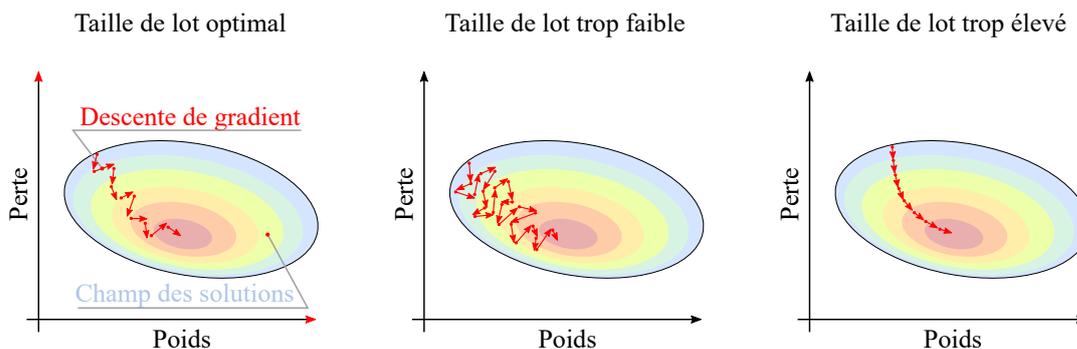


FIGURE 1.11 – Impact de la taille de lot sur la descente de gradient.

Pour obtenir un modèle qui généralise correctement un problème, il faut éviter les phénomènes de sur-apprentissage et de sous-apprentissage présentés en figure 1.12. Le sur-apprentissage se caractérise par un très bon ajustement du modèle aux données d'entraînement, mais une faible capacité à généraliser sur des données inconnues. D'une manière différente, le sous-apprentissage intervient quand le modèle ne parvient pas à comprendre les données d'entraînement et échoue également sur les données inconnues. Pour atténuer ces phénomènes, il est nécessaire d'avoir des jeux de données de qualité et volumineux qui représentent au mieux le problème étudié pour que les réseaux puissent généraliser correctement. Cela passe notamment par les étapes de

prétraitement des données détaillées précédemment en section 1.3.3.

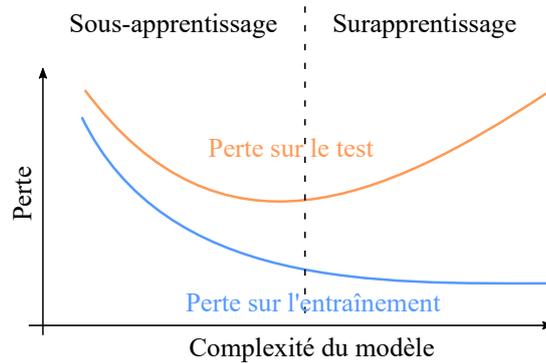


FIGURE 1.12 – Exemples de sous-apprentissage et de sur-apprentissage en fonction de la complexité d'un modèle.

1.3.4.2 Réseaux de neurones convolutifs

Les réseaux de neurones traditionnels peinent à reconnaître des objets dans des images quelque soit leurs positions, on dit qu'ils ne sont pas invariants aux translations. L'introduction des réseaux de neurones convolutifs (CNN) (LECUN, BENGIO et HINTON, 2015), inspirés par le fonctionnement du cortex visuel des animaux (ALOYSIUS et GEETHA, 2017), permet de pallier ce problème. En effet, ce type de réseau spécifiquement créé pour la reconnaissance d'image dans le domaine de la vision par ordinateur détecte les objets indépendamment de leurs positions dans les patches. Pour ce faire, les CNN extraient et apprennent des caractéristiques, avant de les classifier comme dans un réseau de neurones classique, comme montré dans la figure 1.13.

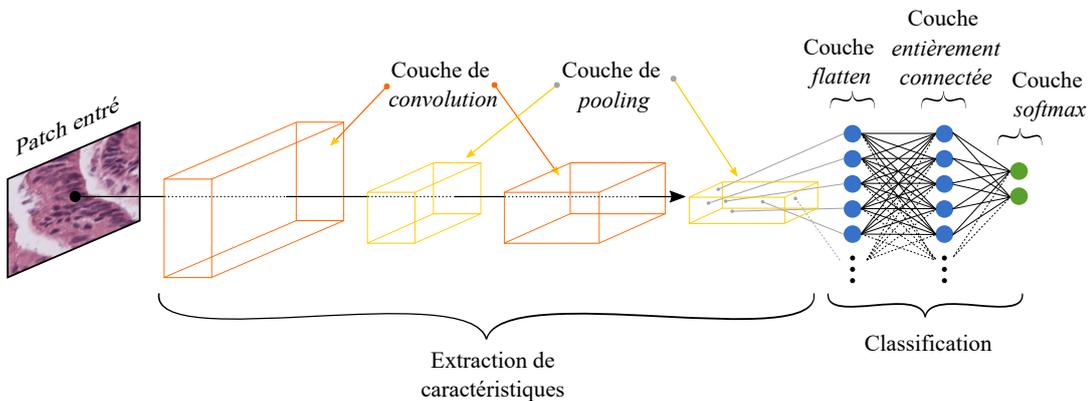


FIGURE 1.13 – Exemple de réseau de neurones convolutifs. Des couches de convolutions extraient des cartes de caractéristiques d'une image donnée en entrée du réseau. Ces caractéristiques sont utilisées par un classifieur pour calculer une probabilité d'appartenance à chaque classe.

L'extraction de caractéristiques est possible grâce à l'utilisation d'opérations de convolution qui permettent par exemple de faire apparaître des arêtes verticales et horizontales comme montrées en figure 1.14.

On appelle filtres, les matrices qui lors des convolutions, permettent de faire apparaître des caractéristiques telles que des arêtes et des contours. Dans un cas classique, le filtre est fictivement positionné sur le coin supérieur gauche de l'image, puis est

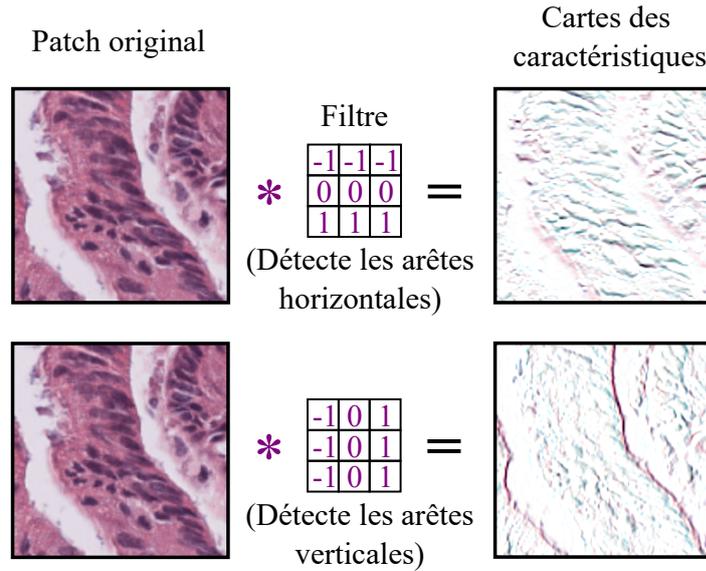


FIGURE 1.14 – Exemples de cartes de caractéristiques obtenues par convolution de filtres horizontaux et verticaux sur un patch.

déplacé d'un pas fixe jusqu'à ce que l'image soit entièrement parcourue. À chaque déplacement, le résultat de la convolution entre le filtre et les pixels concernés de l'image est utilisé pour créer une nouvelle image appelée carte de caractéristique. Un exemple d'un tel calcul est présenté en figure 1.15.

Habituellement, dans un CNN, les couches de convolutions sont suivies de couches de *max pooling* qui sous-échantillonne les cartes de caractéristiques pour gagner en temps de traitement. Les couches de convolutions et de *max pooling* s'enchaînent pour extraire des caractéristiques de plus en plus complexes. Ainsi, les premières couches de convolution d'un CNN extraient des caractéristiques simples telles que des arêtes et des gradients de couleurs, tandis que les couches les plus proches de la sortie du réseau extraient des caractéristiques abstraites telles que des ellipses. Une fois que les caractéristiques sont extraites, la couche *flatten* transforme les cartes de caractéristiques trois dimensions de la précédente couche en vecteur une dimension. Cette étape de réduction de dimensions permet l'utilisation d'un classifieur classique en fin de réseau constitué de couches entièrement connectées. Enfin, une dernière couche *softmax* calcule une probabilité d'appartenance à chaque classe bornée entre 0 et 1 comme détaillé en Équation 1.3. Où σ est la fonction *softmax*, qui prend en paramètre le vecteur des valeurs d'appartenance à chaque classe. Pour chaque classe, la valeur correspondante dans le vecteur est divisée par la somme des valeurs de toutes les classes (i.e. somme du vecteur). Ainsi, on obtient un nouveau vecteur des probabilités d'appartenance à chaque classe.

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (1.3)$$

À noter que dans le cas des images RGB, on décompose l'image en trois canaux distincts. Les filtres sont simplement appliqués à chaque canal. Au final, une image RGB aura trois fois plus de cartes de caractéristiques comme montré en figure 1.16.

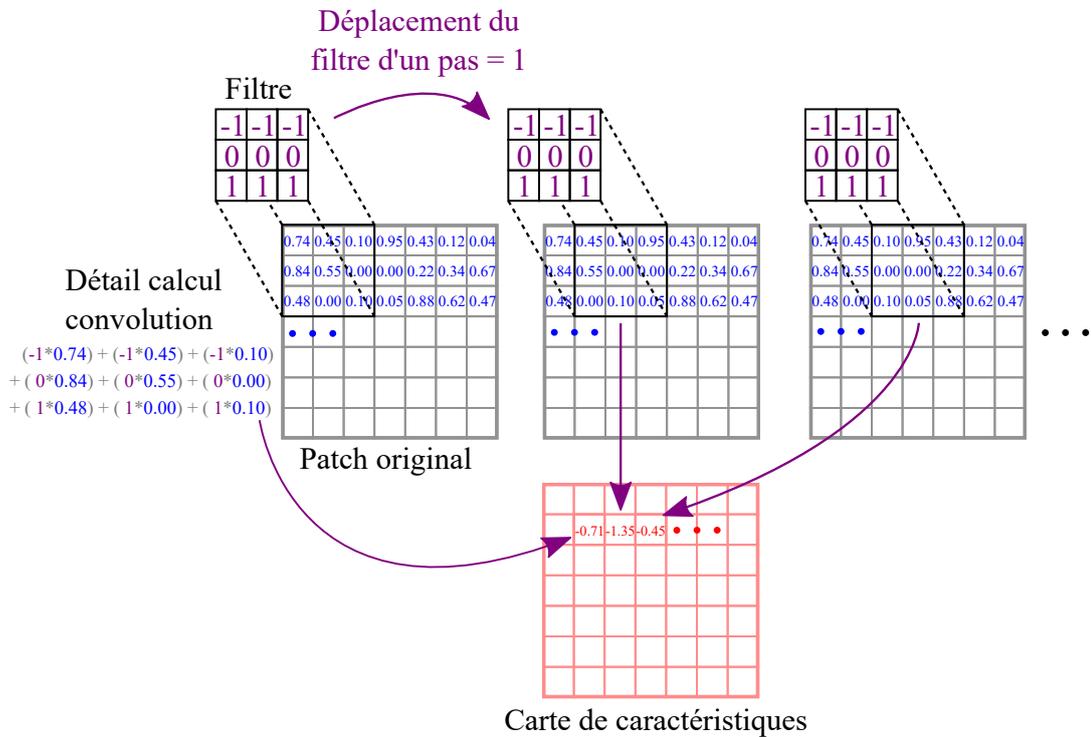


FIGURE 1.15 – Détail du calcul d’une convolution sur un patch. Le filtre parcourt le patch d’un pas fixe, de gauche à droite et de haut en bas. À chaque déplacement, une convolution est réalisée entre le filtre et les pixels du patch concernés. Le résultat devient la valeur d’un pixel de la carte de caractéristiques.

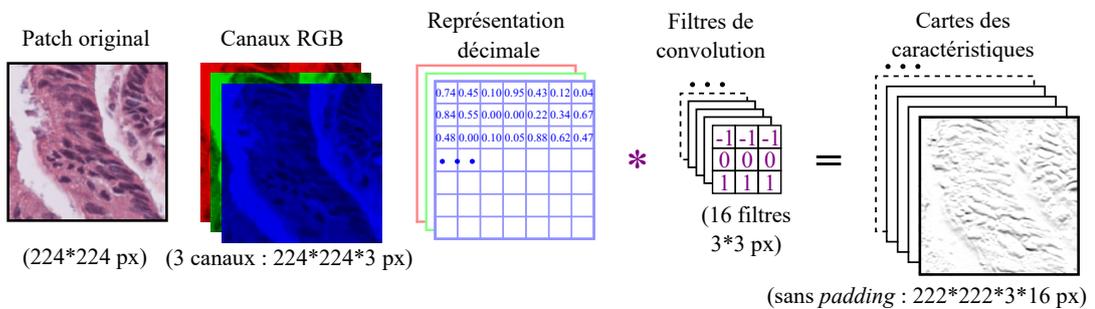


FIGURE 1.16 – Détail d’une convolution sur un patch RGB.

Chapitre 2

Exploitation des différents niveaux de grossissement des images de lames

2.1 Introduction

Les images de lames sont des fichiers spécifiques qui peuvent contenir des représentations des tissus à plusieurs niveaux de grossissement. De ce fait, ces images sont particulièrement volumineuses et nécessitent des outils dédiés pour leurs manipulations. Cela représente à la fois une contrainte quand il s’agit d’appliquer des méthodes d’analyse, mais également un avantage étant donné que l’on dispose de plusieurs niveaux de détails. Ainsi, on peut fortement zoomer et dézoomer dans les tissus pour apprécier les structures biologiques à différentes échelles. Cependant, en pratique, les travaux de recherche actuels qui appliquent des algorithmes d’apprentissage profond pour l’analyse d’image médicale se concentrent généralement sur un seul niveau de grossissement i.e. 20X. Autrement dit, tous les autres niveaux présents dans les WSI sont ignorés, quand bien même ils contiendraient des informations utiles. En effet, observer des tissus à des niveaux supérieurs ou inférieurs aux niveaux traditionnellement utilisés permet par exemple de faire apparaître de nouvelles frontières et d’ajouter ou enlever du contexte dans les tissus.

Bien que la majorité des approches se limite à utiliser un niveau de grossissement unique, cela ne signifie pas que seule des caractéristiques de ce niveau seront extraites par le réseau. Par exemple, il est d’usage d’appliquer des méthodes d’augmentation de données qui peuvent affecter le niveau de grossissement des patches pour limiter les biais dans le jeu de données. Ainsi, il est possible de bénéficier des informations supplémentaires qu’apporte l’ajout de contexte dans un patch ou l’apparition de nouvelles frontières. De plus, certaines architectures de réseaux de neurones profonds sont capables de redimensionner les données en entrée pour les représenter à des niveaux de grossissement plus faibles. De ce fait, les caractéristiques de différents niveaux peuvent être extraites. C’est par exemple le cas avec l’architecture YOLO (REDMON et FARHADI, 2018). Cependant, cela reste insuffisant au vu de tous les autres niveaux qui restent inexploités.

De plus, réaliser un diagnostic à partir d’un seul niveau de grossissement n’est pas conforme aux procédures habituelles employées par les pathologistes. Effectivement, lors d’une séance d’analyse de lames de verre à l’aide d’un microscope, un pathologiste regarde le tissu dans son ensemble avant de zoomer dans les zones d’intérêts (HAMILTON, WANG et MCCULLOUGH, 2012). Les structures biologiques sont donc appréciées à différents niveaux de grossissement avant qu’un diagnostic ne soit rendu.

Ensuite, exploiter différents niveaux de grossissement contribue à rendre les méthodes robustes aux changements d'échelle. Ce qui est particulièrement utile lorsque l'on souhaite créer des méthodes capables de suivre les mouvements d'un pathologiste dans son analyse microscopique afin de fournir des informations en temps réel.

Dans la littérature, on trouve des travaux qui utilisent plusieurs niveaux de grossissement. D'abord, RAZN (DONG et al., 2018) et (WANG et al., 2017) se focalisent sur les zones ayant le plus grand nombre d'objets d'intérêt pour améliorer les résultats des réseaux et économiser du temps de calcul en réduisant les zones de tissus à analyser. Cette approche est également motivée par sa similarité avec la manière dont les pathologistes réalisent leurs séances de diagnostic. D'une manière différente, HAZN (XIA et al., 2016) propose de détecter des objets d'intérêt à différents niveaux de grossissement avant de fusionner les prédictions. D'autres travaux proposent d'entraîner directement des réseaux avec des patches de plusieurs niveaux pour les rendre invariants aux changements de niveaux (SONG et al., 2016; ALKASSAR et al., 2021; ZAIDI, 2021).

Dans un premier temps, dans le chapitre 2.2, nous proposons d'étudier des images de cancer du sein à différents niveaux de grossissement en utilisant le réseau de neurones profond InceptionResNet-v2 (SZEGEDY et al., 2017). Pour cela, nous utilisons le jeu de données BreakHis (SPANHOL et al., 2015), qui contient des images histopathologiques de sein avec et sans tumeurs. Ainsi, nous déterminons les niveaux dans lesquels les caractéristiques des tumeurs sont les plus intéressantes. Ensuite, nous utilisons ces résultats pour proposer et comparer différentes stratégies à des fins de classification de tumeurs dans le sein, robustes aux changements de niveaux de grossissement et en temps réel.

Enfin, dans le chapitre 2.3, nous présentons une preuve de concept en utilisant YOLOv3, une méthode basée sur l'apprentissage profond pour détecter à différents niveaux de grossissement en temps réel des glomérules dans le rein dans le contexte du suivi de greffe. À nouveau, nous explorons différentes stratégies d'entraînement sur différents niveaux de grossissement pour rendre nos réseaux invariants aux changements de niveau.

2.2 Classification de tumeurs dans le cancer du sein

2.2.1 Matériels et méthodes

Dans cette partie, nous souhaitons d'abord utiliser des réseaux de neurones sur des images de lames à différents niveaux de grossissement pour déterminer les niveaux ou les caractéristiques des tumeurs sont les plus pertinentes. Ainsi nous choisissons de travailler avec le jeu de données public BreakHis (SPANHOL et al., 2015) qui contient des images histopathologiques de biopsie de sein avec et sans tumeurs, à quatre niveaux de grossissement 40X, 100X, 200X et 400X. BreakHis est composé de 7 909 patches colorés en H&E (Hématoxyline et Éosine), de taille 700x460 pixels. Parmi ces patches, on compte 2 480 patches de tumeurs bénignes et 5 429 patches de tumeurs malignes, provenant de 82 patients. Ce jeu de données est destiné à la tâche de classification du cancer du sein.

Afin de couvrir une plus large plage de niveaux de grossissement, nous interpolons des patches de 20 niveaux de grossissement intermédiaires à partir des quatre niveaux (40X, 100X, 200X et 400X) du jeu de données d'origine. Pour cela, chaque patch original est redimensionné. Afin d'obtenir des patches de même taille après le redimensionnement à différentes échelles, nous découpons préalablement nos patches à la bonne dimension pour que la taille du patch soit la même après la mise à l'échelle.

Nous utilisons l'équation 2.1 pour calculer la taille de découpe d'un patch en fonction du nouveau niveau de grossissement souhaité. Ainsi, *taille découpe* est la taille du patch à conserver lors du recadrage, *taille patch* est la taille du patch qui sera utilisé par le réseau de neurones profonds et *plus proche niveau* est le niveau de grossissement supérieur le plus proche qui existe dans les patches du jeu de données d'origine.

$$taille\ découpe = \frac{taille\ patch \times plus\ proche\ niveau}{niveau\ souhaité} \quad (2.1)$$

Ensuite, nous choisissons l'interpolation bicubique (KEYS, 1981), qui présente un bon compromis entre qualité de l'interpolation et temps de traitement (PARSANIA et VIRPARIA, 2016), pour dimensionner les patches à la dimension *taille patch*. Un exemple de création d'un patch 28X à partir d'un patch 40X est présenté en figure 2.1. Ainsi, on obtient des nouveaux patches de taille 350x230 pixels¹ pour chacun des 24 niveaux de grossissement de notre plage (de 20X à 400X), comme montré en figure 2.2. Pour chaque niveau, nous créons un jeu de données. De plus, nous créons un dernier jeu de données multi-échelles nommé allX à partir des patches 40X, 100X, 200X et 400X.

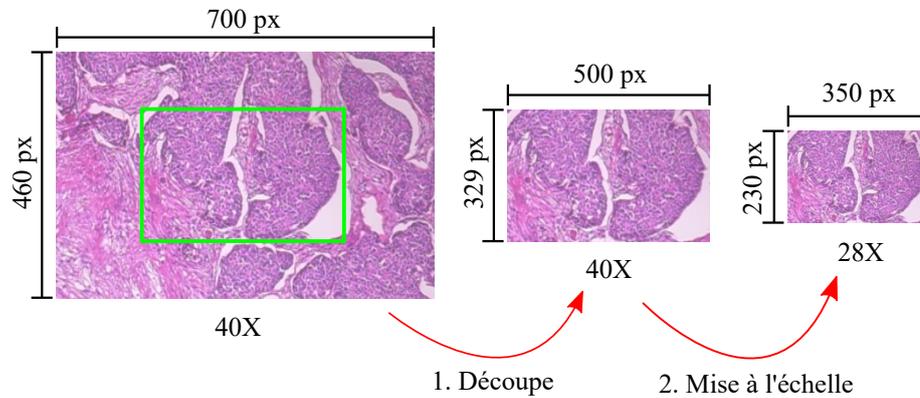


FIGURE 2.1 – Création d'un patch d'un niveau intermédiaire 28X à partir d'un patch de niveau de grossissement 40X. Le patch 40X est découpé puis mis à l'échelle avec une interpolation bicubique pour obtenir un patch 28X.

Ensuite, nous appliquons des méthodes d'augmentation de données aux jeux de données pour améliorer la diversité des données et lutter contre le manque de données qui est un problème récurrent lors de l'utilisation de données médicales. Les patches subissent aléatoirement des rotations, des retournements horizontaux et verticaux.

Puis, nous entraînons un réseau de neurones profond sur chaque jeu de données (i.e. un réseau par niveau), avant de tester nos modèles sur tous les jeux de données, pour en déduire les niveaux où les caractéristiques des tumeurs sont les plus intéressantes. Parmi les architectures qui ont montré les meilleurs résultats sur la tâche de classification de patches de tumeurs bénignes et malignes dans le cas du cancer du sein, on trouve : Inception-v3, ResNet, GoogleNet, VGG. Inception-v3 semble montrer des résultats légèrement supérieurs aux autres méthodes bien que selon les jeux de données, cette différence peut s'avérer minime. Enfin, nous utilisons ces connaissances pour proposer des approches robustes aux changements de niveaux de grossissement.

1. La taille des patches du dataset BreakHis est de 700x460 pixels. Afin d'obtenir des patches de niveau de grossissement plus faible, il faut découper et redimensionner les patches d'origine. De ce fait on perd une partie du contexte du patch d'origine.

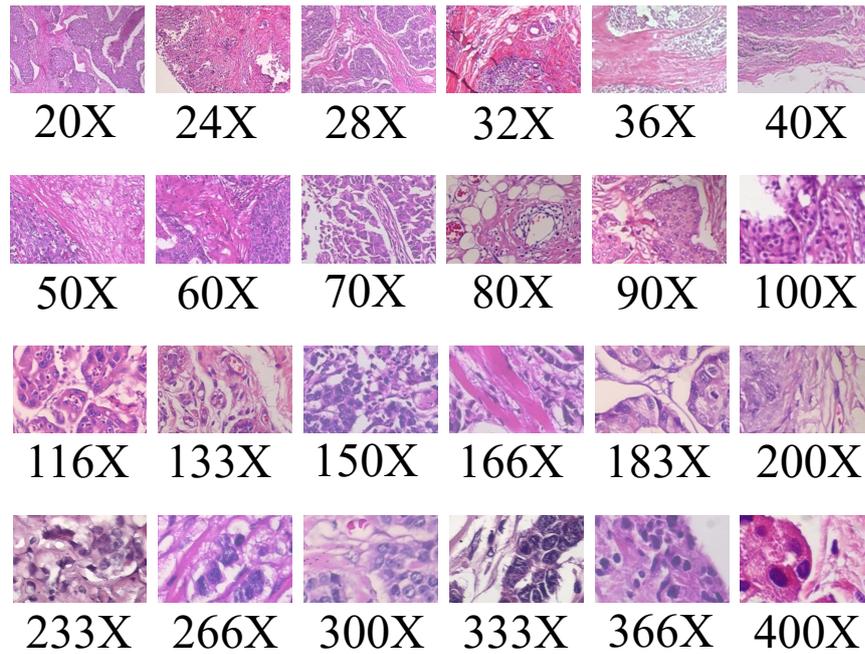


FIGURE 2.2 – Exemples de patches créés à différents niveaux de grossissement à partir des patches du jeu de données BreakHis. Le jeu de données BreakHis contient des patches 40X, 100X, 200X et 400X. Des patches de 20 autres niveaux de grossissement intermédiaire sont créés à partir des patches originaux de BreakHis.

2.2.2 Étude d'une plage de niveaux de grossissement

Nous choisissons d'utiliser l'architecture Inception-ResNet-v2, qui est une évolution de l'architecture Inception-v3, et qui propose d'ajouter des connexions résiduelles aux modules Inception, afin d'accélérer l'entraînement et de surpasser les architectures Inception standards sans connexions résiduelles.

Avant la phase d'entraînement, nous transférons les poids d'un modèle pré-entraîné sur ImageNet (DENG et al., 2009) (ImageNet est la plus grande base de données d'images, couramment utilisée pour les tâches de classification, contenant 14 197 122 images organisées en 21 841 classes) à nos modèles. Nous gelons les poids de nos modèles, avant de réentraîner les cinq dernières couches du réseau sur nos données. Cette méthode est une procédure standard pour affiner l'entraînement des modèles avec peu de données. Ainsi, nous économisons du temps d'entraînement et nous réduisons le sur-apprentissage.

Nous entraînons les réseaux avec les mêmes paramètres et hyperparamètres pour comparer plus facilement les résultats : taux d'apprentissage = 0,01 ; *momentum* = 0,9 ; époques = 30 ; taille de *batch* = 64 ; optimiseur = *SGD* ; fonction de coût = *softmax* ; régularisation = *dropout*. De plus, nous utilisons *EarlyStopping* pour arrêter l'entraînement lorsqu'aucune amélioration n'a lieu sur la *loss* de validation au cours des 15 dernières époques pour éviter le sur-apprentissage. Afin de tirer pleinement parti des données dont nous disposons et de lutter contre le sur-apprentissage, nous effectuons une validation croisée cinq blocs ($K = 5$) pour chaque niveau de grossissement. Les patches sont répartis dans chaque bloc en respectant l'organisation donnée par BreakHis. Chaque bloc est utilisé pour tester une fois les résultats obtenus sur un réseau entraîné sur tous les autres blocs. Ainsi, chaque patch du jeu de données est classé une fois.

Nous effectuons tous les traitements informatiques sur des cartes GPU GTX 1080 fournies par le centre de calcul de l'Université de Strasbourg. Au total, à raison de 10 minutes d'entraînement par réseau, il a fallu environ un jour de calcul cumulé pour entraîner les 125 réseaux (validation croisée $K = 5 \times 25$ jeux de données) que compte cette expérimentation.

Enfin, nous testons chaque modèle sur tous les niveaux de grossissement. Pour cette étape, nous choisissons d'utiliser la précision définie dans l'équation 2.2 comme métrique d'évaluation puisque nos classes sont équilibrées.

$$\text{précision} = \frac{\text{patches bien classés}}{\text{nombre total de patches}} \tag{2.2}$$

Nous construisons une carte de chaleur illustrée en figure 2.3 à partir des précisions obtenues par les réseaux.

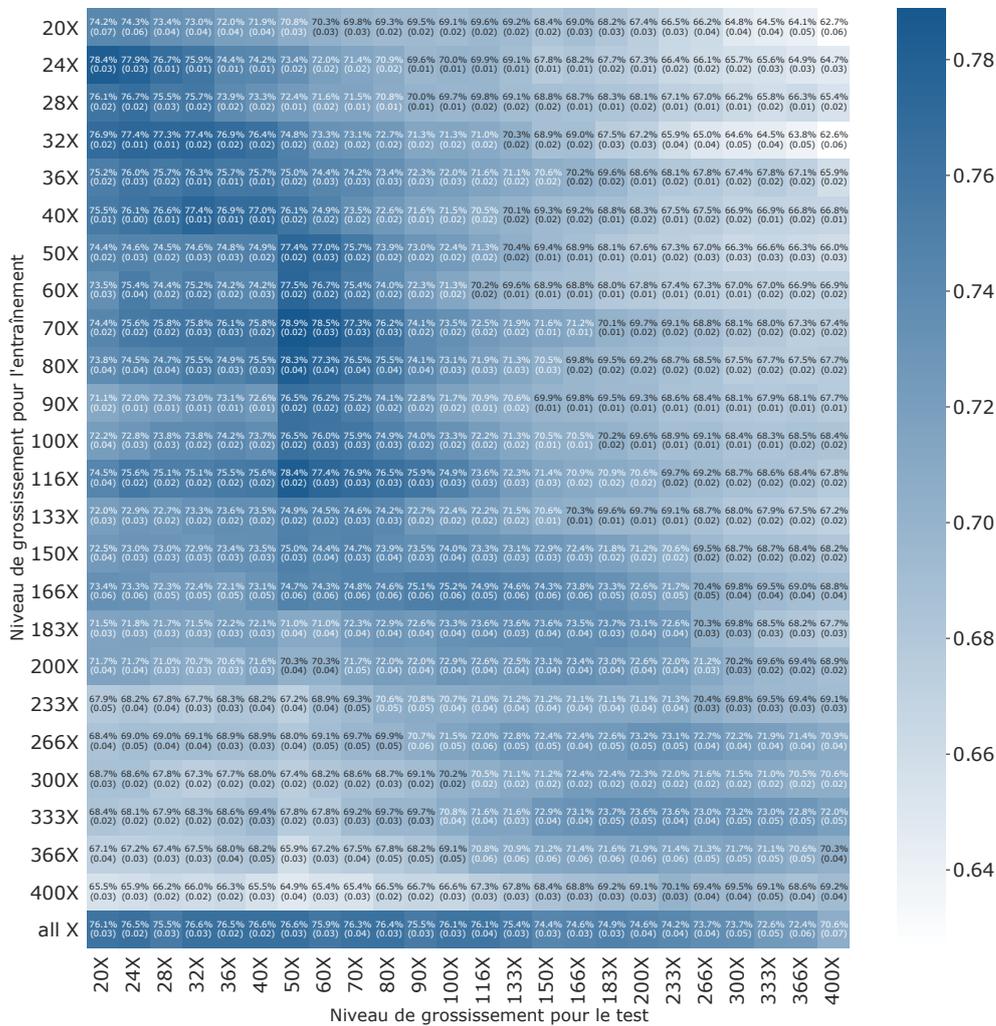


FIGURE 2.3 – Carte de chaleur des précisions des réseaux Inception-ResNet-v2. Indique les précisions (%) obtenues par des réseaux entraînés testés individuellement sur tous les jeux de données dans une plage de niveaux de grossissement de 20X à 400X. Plus un réseau obtient une précision élevée sur un jeu de données, plus la case est sombre, et inversement.

Tout d’abord, il est évident que les réseaux entraînés sur un niveau de grossissement ont les meilleures précisions sur les niveaux de grossissement proches, comme le montre la figure 2.3. Il apparaît clairement qu’un réseau sera généralement meilleur sur un niveau de grossissement proche de celui utilisé lors de l’étape d’apprentissage. Cependant, en regardant la diagonale centrale, on peut voir que la meilleure précision à un niveau de grossissement n’est pas toujours trouvée par le réseau entraîné sur ce niveau de grossissement. En d’autres termes, un léger zoom avant ou arrière pendant l’entraînement améliore les résultats du réseau. Il semble que le contexte proche contient des caractéristiques riches qui améliorent les résultats d’un réseau. Les résultats montrent également qu’à un niveau de grossissement donné, les réseaux entraînés à ce niveau et les niveaux proches obtiennent les meilleures performances. De plus, ces réseaux ne se trompent pas forcément sur les mêmes cas. Cela signifie que chaque niveau de grossissement ne fournit pas les mêmes caractéristiques.

La deuxième observation concerne le gradient de part et d’autre de la diagonale de cases les plus foncées. Plus le niveau de grossissement d’un jeu de données testé est éloigné du niveau de grossissement de l’ensemble de données utilisé pour l’entraînement, plus la précision est faible. Ceci semble confirmer l’intuition que le contexte proche contient les caractéristiques les plus utiles pour un réseau.

Troisièmement, nous remarquons que plus le niveau de grossissement est élevé, moins bonnes seront les précisions des réseaux entraînés sur un seul niveau de grossissement. Avec un niveau de grossissement élevé, on perd le contexte spatial qui permettait de comprendre la structure d’une tumeur. De ce fait, observer des tissus à un niveau de grossissement élevé n’est pas pertinent dans la classification de tumeurs dans le cas du cancer du sein. On remarque que les niveaux de grossissement entre 20X et 80X sont plus riches en caractéristiques pour nos réseaux dans notre cas. En pratique, cela coïncide avec le comportement du pathologiste, qui passe plus de temps dans les niveaux de grossissement faibles (RANDELL et al., 2013).

Quatrièmement, notons que les meilleurs réseaux donnent des précisions proches de 80%. Ces performances sont plus faibles que l’état de l’art en matière de classification de patch de tumeur dans le cas du cancer du sein sur BreakHis (TUFAIL et al., 2021). Cela s’explique par la manière dont nous construisons nos jeux de données à partir de BreakHis. En effet, afin d’étudier une grande plage de niveaux de grossissement, nous construisons des patches de nouveaux niveaux, de ce fait une partie du contexte a été supprimé, ce qui impacte les performances des réseaux. On remarque également que dans le pire des cas, la plus faible précision obtenue est de 62,6%. Cela signifie que dans le cas du cancer du sein, les caractéristiques permettant de distinguer les tumeurs bénignes et malignes sont présentes à tous les niveaux de grossissement de la plage d’étude.

2.2.3 Comparaison de méthodes robustes aux changements de niveaux de grossissement

La première approche consiste à entraîner un seul réseau sur plusieurs grossissements, afin qu’il généralise sa compréhension des structures à plusieurs niveaux. Cette méthode ne nécessite qu’un seul modèle à entraîner, cependant, les performances sont largement liées à la capacité de l’architecture à généraliser sur plusieurs grossissements.

La deuxième approche, consiste à entraîner un réseau par niveau. Dans ce cas, plusieurs modèles sont nécessaires, selon la plage de niveaux de grossissement étudiée.

Nous proposons ensuite une troisième approche où nous entraînons un réseau sur un niveau proche du niveau utilisé pour la phase de test.

Nous comparons les trois approches en étudiant les résultats de différents modèles Inception-ResNet-v2, sur une large plage de niveaux de grossissement pour pouvoir sélectionner les meilleurs réseaux sur une plage donnée.

À partir de la carte de chaleur de la figure 2.3, nous pouvons comparer notre approche avec deux méthodes naïves comme illustré dans la figure 2.4.

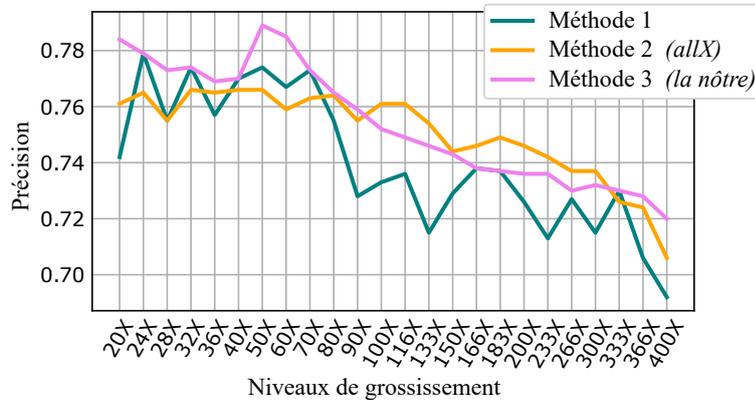


FIGURE 2.4 – Comparaison des approches. Comparaison des précisions (%) de trois approches de prédiction de tumeurs sur différents niveaux de grossissement : Entraîner des réseaux sur plusieurs niveaux ; Entraîner des réseaux sur chaque niveau ; Entraîner des réseaux sur un niveau intermédiaire.

Premièrement, nous constatons que la première approche, où l'on teste les réseaux sur le niveau de grossissement utilisé pour l'entraînement, parvient à obtenir des précisions égales aux autres meilleurs modèles sur certains niveaux de la plage 20X à 70X. Deuxièmement, les réseaux entraînés sur tous les niveaux de grossissement obtiennent des meilleures précisions sur la plage de 100X à 300X. Cette deuxième approche réussit à généraliser les caractéristiques apprises sur plusieurs niveaux, obtenant ainsi des meilleures précisions sur des niveaux de grossissement élevés. Cependant, cette méthode échoue à surpasser les réseaux spécialisés sur des niveaux de grossissement plus faibles (20X à 100X). Troisièmement, nous constatons que l'approche proposée, où nous sélectionnons les meilleurs réseaux entraînés sur un niveau de grossissement quel qu'il soit, obtient les meilleures précisions pour la plage 20X à 90X. Un niveau de grossissement contient des caractéristiques qui permettent parfois d'obtenir de meilleurs résultats sur des niveaux de grossissement proches.

En moyenne, l'approche proposée (75,4%) surpasse ou équivaut les approches naïves (74%, 75,1%) sur la plage de niveaux de grossissement étudié avec l'architecture Inception-ResNet-v2 sur le jeu de données BreakHis. Notre approche montre de meilleurs résultats sur des niveaux de grossissement faibles (20X à 90X). À partir de 100X, l'approche allX devient meilleure. Au-delà d'un certain palier de niveau de grossissement, les caractéristiques des tumeurs présentes dans l'image perdent de leur pertinence. Cependant, les caractéristiques propres à un niveau de grossissement faible aident significativement la compréhension des caractéristiques sur des niveaux de grossissement élevés.

En observant les meilleurs réseaux par niveaux de grossissement montrés dans la figure 2.4, on observe que les réseaux 24X, 32X, 40X, 70X, 116X, 333X et allX permettent de couvrir toute la plage de niveau de grossissement étudié, tout en obtenant les meilleurs résultats. Dans le cadre d'une méthode ARM, on souhaite afficher des informations supplémentaires sur le tissu observé sur les quelques niveaux de grossissement que propose le microscope optique. Prenons par exemple un microscope avec

des niveaux de grossissement 20X, 40X et 100X. Afin d'obtenir les meilleurs résultats dans notre cas, il faut utiliser : le modèle 24X sur le niveau 20X ; le modèle 40X sur le niveau 40X ; le modèle allX sur le niveau 100X. Il paraît intéressant d'utiliser des réseaux entraînés sur un seul niveau de grossissement pour les niveaux de grossissement les plus faibles jusqu'à 90X. Ensuite, à partir de 100X, il convient d'utiliser des réseaux entraînés sur plusieurs niveaux de grossissement. Enfin, il ne faut pas oublier que ces résultats ont été obtenus avec l'architecture Inception-ResNet-v2 sur un jeu de données de cancer du sein. Selon la méthode et les données utilisées, les conclusions peuvent changer.

2.3 Détection de glomérules dans le rein en temps réel

2.3.1 Le projet Sys-MIFTA

Le projet Sys-MIFTA² réunit des mathématiciens qui prédisent le comportement des cellules, des informaticiens qui appliquent des IA à l'analyse d'images microscopique de biopsies, des experts en biologie des macrophages et des médecins expérimentés dans les soins aux patients après une transplantation. Les partenaires qui participent à ce projet proviennent de sept instituts : WIS³ ; HZI/BRICS⁴ ; Radboudumc⁵ ; IH⁶ ; IRIMAS⁷ ; ICube⁸ ; MHH⁹. Sys-MIFTA est financé par ERACoSys-Med.

Des milliers de patients en Europe meurent encore chaque année alors qu'ils sont sur liste d'attente pour une transplantation rénale. La principale raison de rejet de greffe de rein provient d'un processus appelé fibrose interstitielle et atrophie tubulaire (IF/TA). Les macrophages jouent un rôle notamment dans la cicatrisation des plaies, la régulation immunitaire et dans IF/TA. Les récents développements en immunothérapie ouvrent de nouvelles voies pour cibler des macrophages qui ont des effets négatifs sur les organes transplantés. L'objectif du projet Sys-MIFTA est de cibler précisément les macrophages qui ont des effets négatifs sur les organes transplantés, tout en préservant le délicat équilibre immunologique nécessaire à la survie à long terme des greffons.

Dans ce contexte, j'ai travaillé sur la détection en temps réel de structures particulières du rein, appelées glomérules, dont le rôle est de filtrer le sang pour former les urines. Effectivement, en comptant et en déterminant l'état (sain, sclérosé, partiellement sclérosé) des glomérules, un pathologiste est capable d'établir un diagnostic concernant le rejet ou non du greffon, ainsi que l'avancement du rejet. L'objectif est de proposer une méthode qui assiste un pathologiste dans sa routine de travail.

2. Sys-MIFTA : <https://sys-mifta.uha.fr/>

3. WIS, Institut Weizmann des sciences, Rehovot, Israël (<https://www.weizmann.ac.il/>)

4. HZI/BRICS, Centre de recherche sur les maladies infectieuses, Brunswick, Allemagne (<https://www.helmholtz-hzi.de/>)

5. Radboudumc, Centre médical de l'Université Radboud, Nimègue, Pays-bas (<https://www.radboudumc.nl/en/patient-care>)

6. Istituto Humanitas, Milano, Italy (<https://www.humanitas.it/>)

7. IRIMAS, Institut de Recherche en Informatique, Mathématiques, Automatique et Signal, Mulhouse, France (<https://www.irmas.uha.fr/>)

8. ICube, laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie, Strasbourg, France (<https://icube.unistra.fr/>)

9. MHH, École de médecine de Hanovre, Hanovre, Allemagne (<https://www.mhh.de/en/>)

2.3.2 Matériels et méthodes

Les WSI utilisées dans les expériences ont été collectées par la faculté de médecine de Hanovre (MHH) auprès de 10 patients ayant reçu une greffe de rein qui a échoué. Les reins ont été extraits par néphrectomie après perte complète de fonction. Les tissus collectés ont été conservés dans de la paraffine, puis les échantillons ont été coupés en tranches de trois micromètres. Un instrument de coloration (Ventana Benchmark Ultra) a été utilisé pour colorer les lames avec une coloration PAS¹⁰. Ce processus met en évidence les structures tissulaires. Un scanner numérique de diapositives entières (Aperio AT2) a créé des images avec un grossissement de 40X à partir de lames. La plus grande taille d'image est 113 543 x 76 898 pixels. Chaque WSI a été annotée par des experts en pathologie grâce à Cytomine (MARÉE et al., 2016).

Pour nos expériences, nous utilisons 10 WSI contenant entre 86 et 443 glomérules comme présenté dans le tableau 2.1. Pour lutter contre le sur-apprentissage, nous utilisons la validation croisée avec 5 blocs ($K = 5$). Le tableau 2.1 montre également le nombre de glomérules par bloc de validation croisée qui varie entre 400 et 718 glomérules. Pendant le processus de validation croisée, chaque WSI est présentée une seule fois sous forme de patches au réseau pour être testée, comme indiqué dans le tableau 2.2.

ID du patient	#glomérules par WSI	Numéro de bloc	#glomérules par bloc
10	192	1	400
11	208		
12	86	2	506
13	420		
14	208	3	436
15	228		
16	443	4	718
17	275		
18	360	5	461
19	101		

TABLE 2.1 – Le nombre de glomérules par WSI et par bloc.

Numéro de division	Entraînement	Validation	Test
Première division	3, 4, 5	2	1
Deuxième division	1, 4, 5	3	2
Troisième division	1, 2, 5	4	3
Quatrième division	1, 2, 3	5	4
Cinquième division	2, 3, 4	1	5

TABLE 2.2 – La distribution des blocs dans les divisions.

Ensuite, nous créons 5 jeux de données distincts pour alimenter nos réseaux. Quatre d'entre eux contiennent des images (736 x 1280 pixels) d'une seule échelle (grossissement 40X, 20X, 10X et 5X). Le dernier jeu de données contient des images multi-échelles de tous les jeux de données précédents.

10. Periodic Acid-Schiff (PAS) est une méthode de coloration utilisée pour mettre en évidence les glucides dans des coupes histologiques.

De nombreuses méthodes ont été proposées pour la détection d’objets en temps réel : les détecteurs d’objets en temps réel SSD (LIU et al., 2016) et RetinaNet (LIN et al., 2017) détectent plusieurs objets dans les images en prenant des informations en une seule fois. Contrairement à SSD, R-CNN (GIRSHICK et al., 2014) utilise une architecture en deux étapes pour extraire 2000 propositions de région et détecter un objet dans chaque proposition. Les auteurs de R-CNN utilisent un CNN suivi d’un SVM pour détecter et classer les objets. Pour surmonter le long temps nécessaire à l’entraînement du CNN, les auteurs de R-CNN retirent les 2000 propositions de régions et alimentent directement le CNN avec l’image entière. Cette amélioration conduit à l’architecture de détection d’objets en temps réel de pointe appelée Faster-RCNN (GIRSHICK, 2015). De plus, Kawazoe et al. (KAWAZOE et al., 2018) montrent la supériorité de Faster-RCNN pour la tâche de détection des glomérules. En raison de son détecteur en deux étapes, Faster-RCNN est une méthode en temps réel lente par rapport aux SSD, RetinaNet et YOLO (REDMON et al., 2016) comme indiqué dans (REDMON et FARHADI, 2018). Notre objectif étant d’effectuer une détection en temps réel sur des images affichées par un microscope, nous souhaitons une architecture rapide. Nous avons fait un compromis entre la qualité et la vitesse de détection. Par conséquent, nous privilégions YOLO pour ses performances proches de Faster-RCNN et sa vitesse supérieure à Faster-RCNN.

Pour effectuer la détection, YOLO utilise des caractéristiques de 3 échelles différentes comme le montre la figure 2.5. Pour chaque image d’entrée, l’architecture utilise plusieurs couches de sous-échantillonnage pour réduire la taille de l’image à l’échelle 1. Les objets d’intérêts sont détectés à ce niveau avec un pas de 32. Ensuite, l’image est suréchantillonnée à l’échelle 2 et le réseau effectue la détection avec un pas de 16. Enfin, l’image est à nouveau suréchantillonnée et la détection est effectuée avec un pas de 8 à l’échelle 3. Les cartes des caractéristiques des trois échelles différentes sont extraites et concaténées. À chaque échelle, le réseau prédit trois emplacements de boîtes englobantes x, y, w, h avec une probabilité de classe associée. Pour filtrer toutes ces prédictions, un score de confiance est défini (voir l’équation 2.4) comme le produit de la probabilité de classe multipliée par l’intersection de l’union (IOU détaillée en 2.3) entre les prédictions et la vérité terrain :

$$IOU = \frac{\text{prédictions} \cap \text{vérité terrain}}{\text{prédictions} \cup \text{vérité terrain}} \quad (2.3)$$

$$\text{Score de Confiance} = Pr(\text{Classe}_i) \times IOU \quad (2.4)$$

En d’autres termes, un score de confiance élevé signifie une forte probabilité qu’une boîte englobante appartienne à un objet donné par le réseau. Par conséquent, les détections avec un faible score de confiance sont souvent des détections de faible qualité. Un seuil est établi pour supprimer ces détections avec une faible confiance. Après cette étape, il reste plusieurs détections sur le même objet. Pour résoudre ce problème, l’algorithme de suppression non maximale (NMS) (BODLA et al., 2017) est utilisé pour supprimer la détection redondante appartenant au même objet. De cette manière, il ne reste qu’une seule boîte englobante par objet dans l’image.

Nous utilisons l’implémentation YOLOv3 fournie par (QQWEEEE, 2019) qui est programmée en Python en utilisant Keras (CHOLLET et al., 2015). Nous utilisons une approche d’apprentissage par transfert appelée *fine-tuning* pour transférer les poids du réseau neuronal du modèle pré-entraîné sur le jeu de données COCO (LIN et al., 2014) à notre jeu de données (COCO est l’un des plus grands jeu de données d’images pour la tâche de détection, regroupant 123 287 images qui contient 886 284 instances). Ensuite, le modèle est entraîné sur des images de glomérules pour

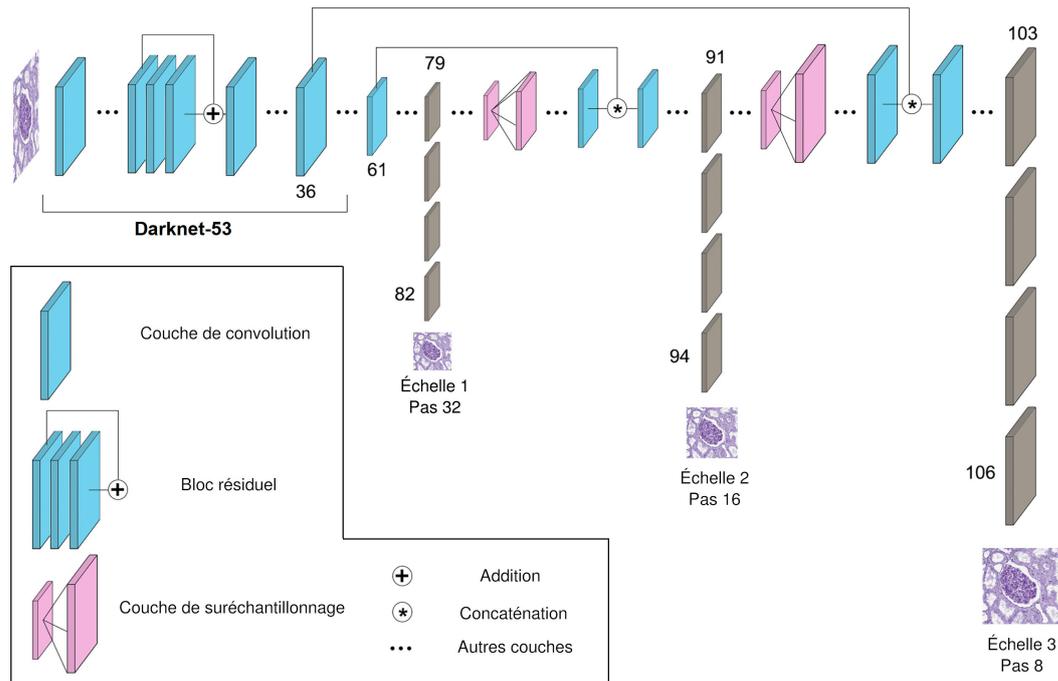


FIGURE 2.5 – L’architecture YOLOv3 avec Darknet-53 comme *backbone* (i.e. réalise l’extraction de caractéristiques). L’architecture est composée de 106 couches entièrement convolutives.

être adapté à notre tâche. Affiner notre réseau est très utile dans notre contexte où les données sont difficiles à obtenir. Ainsi, cela nous permet d’utiliser un jeu de données de taille modeste tout en conservant un modèle efficace et pertinent comme dans (TAJBAKHSI et al., 2016).

Dans notre implémentation, nous formons notre réseau avec des extraits de lames entières scannées car cela imite mieux la façon dont un/une pathologiste voit une partie d’une lame sous un microscope virtuel lors d’un diagnostic. Nous prenons une taille de moniteur standard de 720p comme référence pour alimenter notre réseau. Ainsi, chaque WSI est divisée en patches de même taille (736¹¹ x 1280 pixels). Afin d’analyser les performances du réseau et l’impact de chaque échelle, nous créons 5 jeux de données distincts pour mener nos expériences. Quatre jeux de données contiennent des images à une seule échelle (grossissement 40X, 20X, 10X et 5X) et le dernier jeu de données contient des images de toutes ces échelles. Un exemple de patches et de leurs vérités terrain associées est présenté dans la figure 2.6.

Nous prévoyons que l’utilisation d’images multi-échelles aidera le réseau à généraliser sa capacité à détecter les glomérules à différents niveaux de grossissement en temps réel, comme indiqué dans (SONG et al., 2016) car nous voulons que notre modèle soit capable de détecter les glomérules à différentes échelles.

Pour atténuer le manque de données et augmenter la diversité du jeu de données, nous avons utilisé la bibliothèque Augmentor (BLOICE, ROTH et HOLZINGER, 2019) pour effectuer l’augmentation des données sur notre jeu de données. Nous privilégions les transformations qui pourraient apparaître dans la réalité du fait de manipulations humaines ou mécaniques comme dans (LAMPERT et al., 2019) :

11. Le réseau YOLOv3 sous-échantillonne les images d’entrée par 32. Par conséquent, les images d’entrée doivent être multiples de 32. Nous choisissons donc 736 x 1280 pixels au lieu de 720 x 1280 pixels

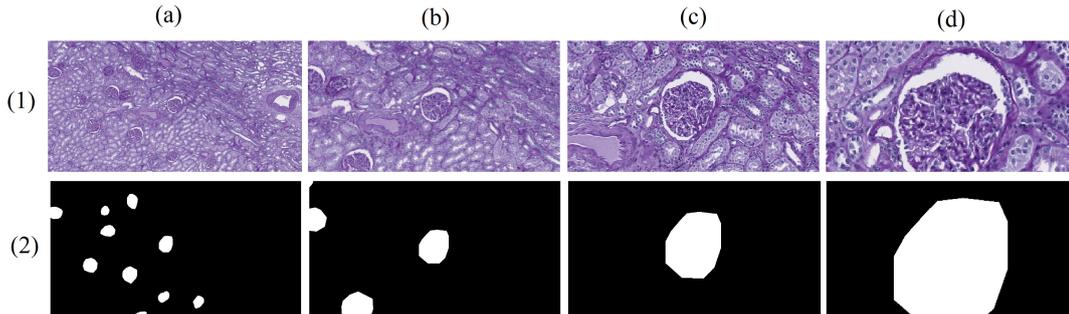


FIGURE 2.6 – La première ligne (1) montre des exemples de patches d’une taille de 736 x 1280 pixels pris à différentes échelles. La deuxième ligne (2) montre la vérité terrain associée avec en noir le tissu et en blanc les glomérules (les colonnes (a) à (d) sont respectivement à un grossissement de 5X, 10X, 20X et 40X).

- affine : rotation aléatoire et retournement horizontal/vertical ;
- contraste : variation aléatoire de l’ombre et de la lumière.

Nous appliquons les méthodes d’augmentation de données choisies à toutes les images de nos jeux de données et nous supprimons les images produites sans glomérule. Ce processus multiplie notre nombre total d’images par 6 pour chaque jeu de données, comme indiqué dans le tableau 2.3.

Nombre d’images	Jeux de données				Total
	40X	20X	10X	5X	
Avant l’augmentation des données	5,701	2,768	1,130	412	10,011
Après l’augmentation des données	35,185	16,851	7,372	2,883	62,272

TABLE 2.3 – Nombre d’images dans chaque jeu de données avant et après l’augmentation des données.

2.3.3 La détection de glomérules à différents niveaux de grossissement

Nous entraînons un réseau avec l’architecture YOLOv3 sur chaque jeu de données (40X, 20X, 10X, 5X, multi-échelle) pour observer les impacts de l’utilisation d’images multi-échelles. Nous définissons les mêmes paramètres et hyper-paramètres lors de l’entraînement pour comparer les performances des réseaux. Le modèle COCO fourni par les auteurs YOLO est utilisé comme modèle pré-entraîné. Pour obtenir une perte stable, nous gelons toutes les couches sauf les 3 dernières. Ensuite, nous entraînons nos réseaux avec un taux d’apprentissage de 0,001, un *batch* de 1 et 50 époques. Nous affinons ensuite notre réseau en dégelant les couches et en le ré-entraînant avec un taux d’apprentissage inférieur de 0,0001, un *batch* de 1 et 50 époques. Pour forcer YOLOv3 à apprendre les informations correctes, nous choisissons expérimentalement une IOU de 0,3 et un score de confiance de 0,2 pendant le processus d’entraînement. Ces valeurs sont les meilleures pour filtrer les mauvaises prédictions et conserver la majorité des meilleures détections, comme indiqué dans (PENG et al., 2016) et (ALMASNI et al., 2018). Nous évaluons le modèle avec les mêmes valeurs d’IOU et de confiance. L’optimiseur Adam et la perte YOLO proposés par les auteurs YOLO sont utilisés.

L'ensemble du processus est effectué sur des cartes GPU Nvidia GTX 1080 fournies par le Centre de Calcul de l'Université de Strasbourg et dure de 4 à 72 heures selon le jeu de données utilisé. Pour évaluer le modèle, nous calculons la précision et le rappel à partir des instances Vrai Positif (TP), Faux Positif (FP) et Faux Négatif (FN). Nous montrons un exemple de détections effectuées par le modèle sur plusieurs patches dans la figure 2.7.

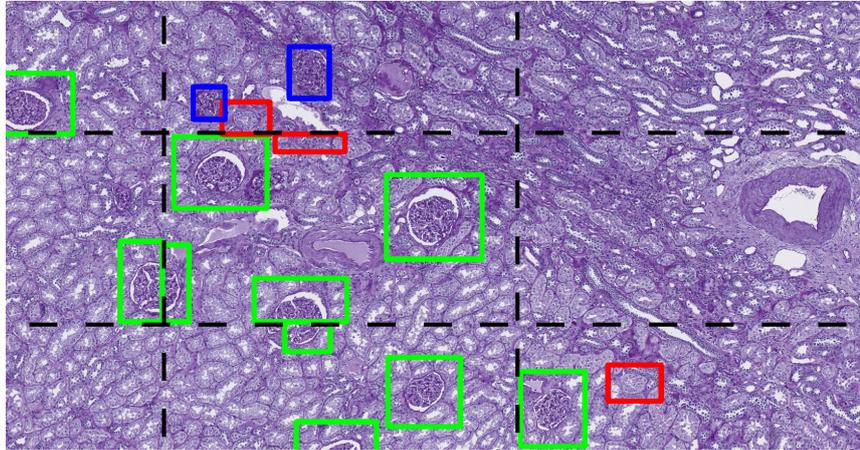


FIGURE 2.7 – Exemples de détections réalisées par l'un de nos modèles sur des patches à partir d'une WSI colorée en PAS. Nous dessinons une ligne pointillée noire pour montrer la limite des patches. Les encadrés verts indiquent les prévisions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.

Une fois l'entraînement terminé, chaque réseau est testé sur chaque jeu de données (40X, 20X, 10X, 5X, multi-échelles). Pour l'évaluation, nous utilisons l'IOU et le score de confiance utilisés pendant l'entraînement des réseaux. Ensuite, les TP, FP et FN obtenus par nos modèles sur les images de test sont dessinés pour évaluer visuellement les résultats. Un exemple de telles prédictions est montré sur une biopsie en figure 2.8, et sur une néphrectomie en figure 2.9.

Puis, nous calculons les F1-scores (validation croisée $K=5$) indiqués dans le tableau 2.4. Des exemples de détections en temps réel effectuées par notre réseau entraîné avec des images agrandies 5X sur une vidéo sont disponible ici ¹².

		Jeux de données utilisés pour les tests					
		Grossissement	40X	20X	10X	5X	40X,20X,10X,5X
Jeux de données utilisés pour entraîner	40X		0.38	0.18	0	0	0.32
	20X		0.44	0.40	0.03	0	0.35
	10X		0.35	0.44	0.56	0.18	0.43
	5X		0.18	0.53	0.78	0.77	0.48
	40X,20X,10X,5X		0.43	0.49	0.61	0.46	0.48

TABLE 2.4 – Résultats des F1-scores pour YOLOv3 entraîné et testé sur un jeu de données à différentes échelles.

Les résultats montrent que le réseau entraîné sur des images à grossissement 20X donne le meilleur F1-score sur des images à grossissement 40X. Selon ces résultats, les images à grossissement 20X ont des informations plus pertinentes pour améliorer la capacité de compréhension du réseau à un grossissement 40X.

12. <http://jonathan-weber.eu/cp/cbms2020/>

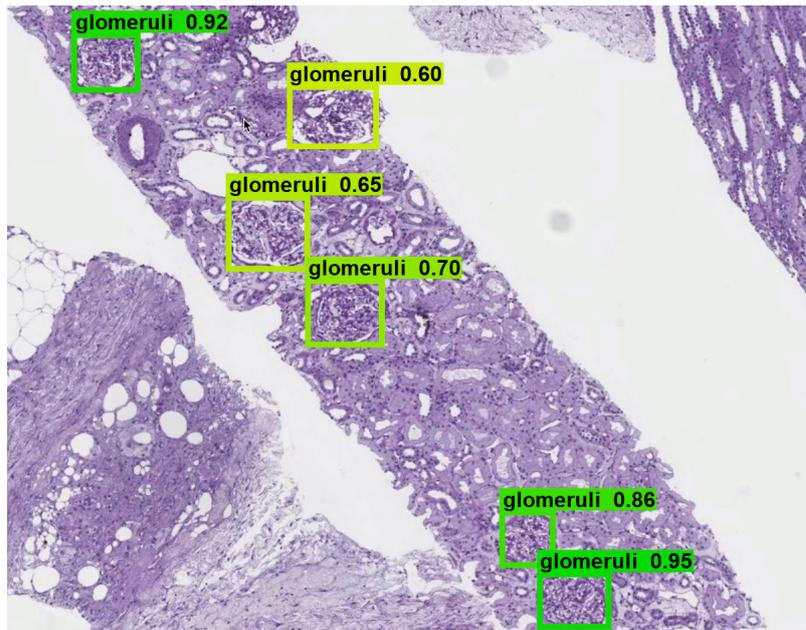


FIGURE 2.8 – Exemples de détections réalisées par le réseau entraîné et testé sur des images grossies 5X. Les images sont fusionnées pour reconstruire la WSI d’une biopsie. Les boîtes de délimitation vertes montrent les prédictions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.

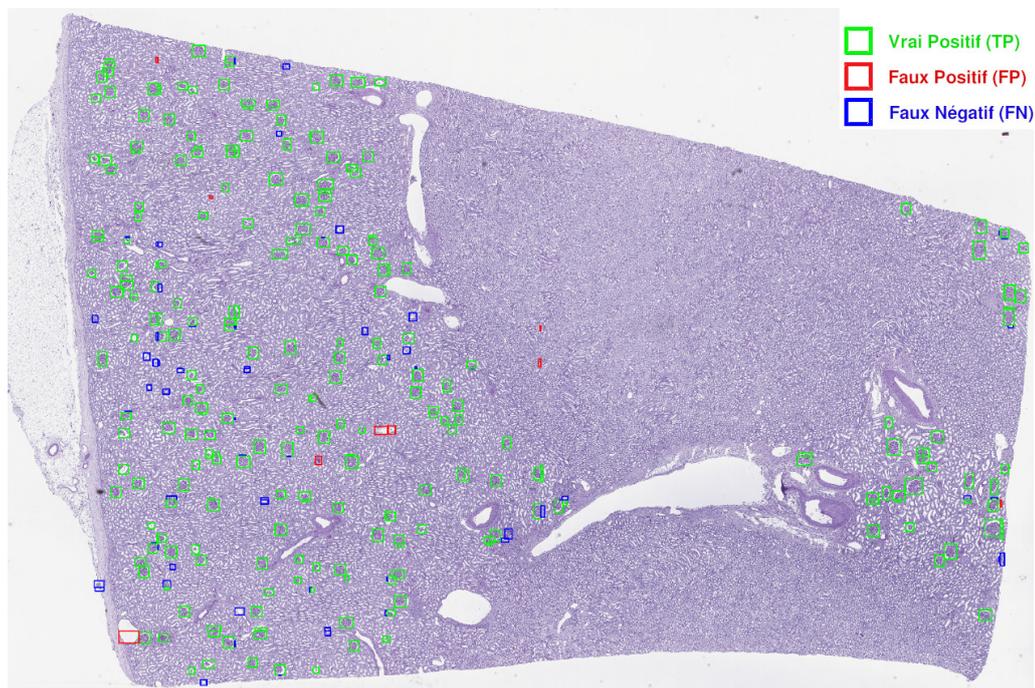


FIGURE 2.9 – Exemples de détections réalisées par le réseau entraîné et testé sur des images grossies 5X. Les images sont fusionnées pour reconstruire la WSI d’une néphrectomie du patient 15. Les boîtes de délimitation vertes montrent les prédictions correctes faites par le réseau, les rouges sont une détection incorrecte et les glomérules manqués sont bleus.

D'une manière différente, le réseau entraîné sur des images à grossissement 5X est le meilleur pour les jeux de données contenant des images à grossissement 20X, 10X et 5X. Encore une fois, cette observation montre que certains niveaux de grossissement sont plus pertinents pour le réseau et il semble judicieux d'entraîner le réseau en utilisant des images avec un grossissement plus faible.

Contrairement aux résultats trouvés par Song et al. (SONG et al., 2016) où l'approche multi-échelle a obtenu les meilleurs F1-scores, dans nos expériences, les réseaux 5X et multi-échelles ont le même F1-score lorsqu'ils sont utilisés sur le jeu de données multi-échelles. Le réseau multi-échelles semble ne pas se généraliser à plusieurs niveaux de grossissement. De plus, certains F1-scores du tableau 2.4 sont proches ou égaux à 0. Cela se produit lorsque nous testons un modèle avec un jeu de données contenant des images avec un grossissement beaucoup plus faible. Cette observation montre que nos réseaux peuvent facilement détecter les glomérules avec un niveau de grossissement plus élevé plutôt qu'un niveau de grossissement inférieur.

En conséquence, pour bien détecter les objets d'intérêt avec un microscope assisté par l'IA, il semble intéressant d'utiliser des images à faible grossissement pour l'entraînement du réseau. Une dernière chose à noter, les meilleures performances sont données par le réseau 5X qui est entraîné avec moins de 500 images originales (un faible niveau de grossissement produit moins d'images).

En pratique, lorsque le/la pathologiste se déplace dans un tissu à l'aide d'un microscope, il arrive que des glomérules en bord de patch soit coupés comme montré sur la figure 2.7. Comme nos réseaux ont été entraînés sur des patches contenant des glomérules coupés, ils arrivent à correctement détecter des parties de glomérule.

Dans la figure 2.9, nous observons de nombreux glomérules manqués. Ceci est dû au fait que les détections effectuées par le réseau sont filtrées par les seuils de confiance et d'IOU. Le filtrage des détections diminue le nombre de glomérules trouvés, mais il diminue également le nombre de fausses détections trouvées par le réseau. Il y a donc un compromis pour obtenir le meilleur F1-score possible.

2.4 Conclusion

Dans ce chapitre, nous avons commencé par étudier une plage de 24 niveaux de grossissement entre 20X et 400X dans des images histopathologiques de sein du jeu de données BreakHis à l'aide de réseaux de neurones InceptionResNet-v2. Nous avons rapporté dans une carte de chaleur les précisions des réseaux à classer correctement des tumeurs sur les différents niveaux. Ainsi nous avons pu montrer les plages de niveaux dans lesquelles les caractéristiques des structures biologiques liées aux tumeurs du sein étaient les plus présentes. Ensuite, nous avons utilisé ces résultats pour proposer et comparer différentes stratégies pour classer des tumeurs dans le sein à différents niveaux de grossissement et en temps réel.

À partir des précédents résultats, nous avons introduit une preuve de concept basée sur l'architecture d'apprentissage profond YOLOv3, dans le cadre du projet européen Sys-MIFTA. Nous avons utilisé notre méthode pour détecter à différents niveaux de grossissement et en temps réel, des glomérules dans le rein dans le contexte du suivi de greffe. À nouveau, nous avons exploré différentes stratégies d'entraînement sur différents niveaux de grossissement pour rendre nos réseaux invariants aux changements de niveau.

Chapitre 3

Détection de mutations génétiques dans le cancer du côlon

3.1 Introduction

3.1.1 Le projet AiCOLO

Le projet AiCOLO¹ réunit des informaticiens, spécialisés en analyse de données et d'images, des laboratoires ICube² et IRIMAS³, et des pathologistes et biostatisticiens, spécialisés en pathologie numérique et cancer du côlon, du laboratoire CGFL⁴. Le projet est financé par Inserm / Plan Cancer (2014-2019).

Le projet AiCOLO vise à contribuer au développement de nouvelles techniques d'intelligence artificielle entraînées sur une cohorte de patients atteints de cancer du côlon cliniquement annotés (e.g. tumeurs, structures stromales, vaisseaux, muscles, infiltrat immunitaire, tissu normal, informations génétiques).

Le premier objectif est le développement d'outils d'aide aux pronostics pour la classification automatique des tissus des WSI et d'énumérer les différents marqueurs de pronostics connus tels que l'infiltration immunitaire TILS, le contenu stromal ou le nombre d'éosinophiles dans les différentes zones. Ces méthodes doivent être entraînées pour trouver des motifs d'images dans le tissu tumoral associés au pronostic des patients.

Le second objectif du projet AiCOLO sur lequel j'ai travaillé, propose de développer une approche nouvelle pour prédire le statut génétique RAS et BRAF à partir de WSI. Le but est de déterminer si l'intelligence artificielle peut détecter des motifs associés à de telles caractéristiques génétiques. Ensuite, l'idée est d'étudier les couches d'activation d'un réseau profond entraîné pour classer les patients afin d'extraire des informations pour expliquer sa décision.

3.1.2 Les mutations génétiques

Le corps humain est composé de cellules, elles-mêmes composées d'un noyau qui contient les 46 chromosomes d'un individu. Un chromosome est constitué d'acide désoxyribonucléique (ADN) organisé, comme montré en figure 3.1. l'ADN est une succession de quatre nucléotides (guanine, adénine, cytosine et thymine) que l'on

1. AiCOLO : <https://wemmertc.github.io/aicolo/>

2. ICube, laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie, Strasbourg, France (<https://icube.unistra.fr/>)

3. IRIMAS, Institut de Recherche en Informatique, Mathématiques, Automatique et Signal, Mulhouse, France (<https://www.irimas.uha.fr/>)

4. CGFL : Centre Georges François Leclerc, Dijon, France (<https://www.cgfl.fr/>)

représente le plus souvent par leur première lettre en majuscule (GACT). Il contient les informations génétiques qui permettent le développement et le bon fonctionnement des organismes. On appelle gène un morceau d'ADN qui produit des protéines, chez l'homme on dénombre plus de 25 000 gènes.

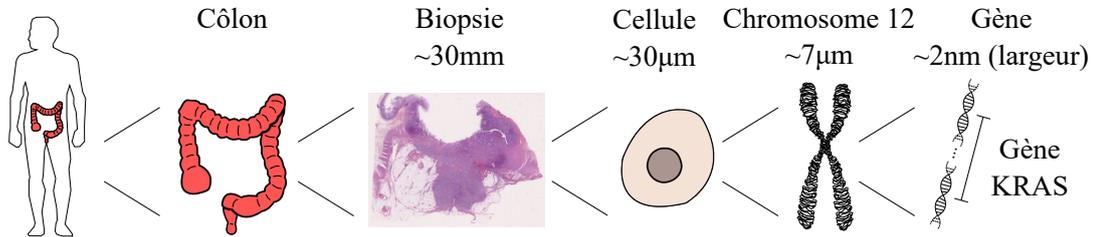


FIGURE 3.1 – Du corps humain au gène KRAS. Aperçu d'un colon, d'une biopsie, d'une cellule, d'un chromosome et d'un gène KRAS à leurs échelles.

Le bon fonctionnement de notre corps dépend d'un renouvellement permanent des cellules appelé le cycle cellulaire. Pendant ce cycle, les cellules vont dupliquer leur matériel génétique (interphase) avant de se diviser (mitose) pour produire des cellules filles, puis mourir. D'un côté, une insuffisance de divisions cellulaires provoque le dysfonctionnement des organes. De l'autre, trop de divisions cellulaires entraînent une prolifération anormale de cellules qui se traduit par des tumeurs bénignes et malignes. Lors de la division cellulaire, il y a un risque d'apparition de modification de l'ADN transmis entre la cellule mère et la cellule fille. On appelle ses modifications de l'information génétique, des mutations génétiques. Pour chaque mutation, il existe des sous-types de variants qui peuvent influencer par exemple sur l'agressivité du cancer.

Dans notre ADN, certains gènes sont responsables de la prolifération cellulaire comme les proto-oncogènes, et d'autres inhibent la prolifération comme les anti-oncogènes. Etant donné que les chromosomes vont par paire, nous possédons deux copies de chaque gène. Dans le cas des anti-oncogènes, si une copie subit une mutation, alors le patient a des prédispositions au cancer. Il faut que les deux copies soient mutées pour qu'une tumeur apparaisse. D'une manière différente, pour les proto-oncogènes, il suffit qu'une seule copie soit mutée pour qu'une tumeur se développe. Lorsque les gènes responsables du contrôle de la prolifération cellulaire sont mutés et qu'une tumeur maligne se développe, les cellules ne meurent plus et prolifèrent. Elles se regroupent en amas et des vaisseaux sanguins apparaissent pour les alimenter. A noter que les mutations des gènes qui n'ont pas d'influence sur la prolifération n'entraînent pas de tumeur.

Selon (ANAND et al., 2008), 90 à 95% des mutations génétiques sont liées à notre environnement comme l'exposition à des produits chimiques ou des polluants, le stress, l'obésité, l'inactivité physique, les ultraviolets. Parmi les facteurs environnementaux, à elle seule, l'alimentation est responsable de 30 à 35% des mutations. Le tabac, quant à lui, entraîne 25 à 30% des mutations. En moindre mesure, 5 à 10% des mutations sont transmises de manière héréditaire. Ce phénomène se produit lorsque les mutations apparaissent sur des cellules reproductrices. Enfin, les cancers touchent différents organes selon les comportements, par exemple, 90% des cancers du poumon sont liés au tabac (AMES, GOLD et WILLETT, 1995).

Pour connaître le génome d'un individu et ensuite déterminer les mutations dont il est atteint, il faut d'abord séquencer son ADN. Avant la phase de séquençage, plusieurs étapes en laboratoire sont nécessaires pour révéler l'ADN, réduire le nombre

d'artefacts et améliorer le séquençage par la suite. Le séquençage permet de déterminer l'ordre des nucléotides (GACT) d'une séquence ADN. Puis, la succession de nucléotides est analysée afin de trouver les parties qui correspondent aux gènes étudiées. Pour chaque gène, il est possible de déterminer ses mutations et ses variants. Les premières méthodes de séquençage ADN sont apparues dans les années 70 (GILBERT et MAXAM, 1973 ; SANGER et COULSON, 1975). Il faudra attendre 1987 pour que cette tâche soit automatisée avec l'invention des séquenceurs ADN (HOOD, HUNKAPILLER et SMITH, 1987). Jusqu'à aujourd'hui, trois générations de séquenceurs ADN se sont succédé permettant ainsi de lire l'ADN plus vite, à moindre coût, avec moins d'erreurs.

Dans son rapport annuel (SIEGEL et al., 2022), l'*American Cancer Society* (ACS) estimait qu'en 2022, il y aurait $\sim 1\,900\,000$ nouveaux cas de cancer et $\sim 600\,000$ décès liés au cancer aux États-Unis. À lui seul, le cancer colorectal est responsable de $\sim 100\,000$ nouveaux cas et de $\sim 50\,000$ décès. Néanmoins, tous les cancers colorectaux ne sont pas identiques. En effet, il existe différents types de cancer et différents stades en fonction de l'évolution de la maladie. Connaître le type de cancer est essentiel pour personnaliser le traitement des patients et ainsi augmenter les chances de survie. Dans le cadre d'une analyse microscopique, un pathologiste est capable de déterminer de nombreuses informations sur le cancer et son évolution à partir d'une lame de verre. Cependant, l'analyse visuelle des tissus n'est pas suffisante pour savoir quels gènes sont mutés et quels variants sont présents. Cette étape est cruciale pour déterminer certains types de cancer et orienter les patients vers le meilleur traitement. Dans le cas du cancer colorectal, 3% des cancers présentent une mutation KRAS G12C (Glycine 12 to Cysteine). Cette mutation rare joue un rôle clé dans la prolifération cellulaire et est responsable d'une forme de cancer extrêmement agressive. Il existe cependant des traitements qui peuvent améliorer les chances de survie des patients et qui peuvent être prescrits dès l'identification de la mutation (NAGASAKA et al., 2020). Les informations sur le génome peuvent être obtenues par un séquenceur d'ADN, mais cela reste long et coûteux.

En 2018, N. Coudray et al. (COUDRAY et al., 2018) ont été les premiers à appliquer avec succès une méthode d'apprentissage profond pour prédire les mutations directement dans des WSI de poumons. Dans le sillage de ces travaux, de nombreuses approches basées sur l'apprentissage profond ont été proposées pour prédire l'information génétique (CHEN et al., 2020a ; BIAN et al., 2021 ; WANG et al., 2019 ; HE et al., 2020 ; LIU et al., 2022).

Parmi elles, certaines approches se concentrent sur les mutations dans le cancer colorectal (BIAN et al., 2021 ; HE et al., 2020). Ces résultats encourageants sont surprenants car les pathologistes ne sont pas capables de déterminer visuellement la présence ou l'absence d'une mutation dans un gène à partir de l'analyse microscopique d'un tissu. Pourtant, les réseaux de neurones profonds semblent y parvenir. Par conséquent, les méthodes d'apprentissage profond ont le potentiel de déterminer l'information génétique sans passer par le séquençage de l'ADN, long et coûteux. En outre, les réseaux pourraient également nous permettre de comprendre les structures biologiques associées aux mutations grâce au domaine de l'explicabilité. Malgré les initiatives visant à comprendre les prédictions des réseaux à l'aide de méthodes d'explicabilité (BIAN et al., 2021 ; WANG et al., 2019), les structures qui conduisent à la prédiction des gènes mutés et des variants sont encore mal comprises (WANG et al., 2019).

Dans ce contexte, il est intéressant de se concentrer sur la prédiction de la mutation KRAS G12C dans le cancer du côlon à l'aide de méthodes d'apprentissage profond, compte tenu de l'agressivité du cancer qui en résulte et de la possibilité de guider les patients vers un traitement spécifique qui augmente leurs chances de survie.

3.1.3 Travaux similaires

Ces dernières années, des approches basées sur des réseaux neuronaux profonds ont été étudiées pour prédire des informations génétiques à partir d'images histopathologiques. Cependant, on comprend encore mal comment les réseaux réalisent leurs prédictions. Le domaine de l'explicabilité apporte des réponses à ces questions.

Suite aux travaux de (COUDRAY et al., 2018), qui présentent un cas de prédiction d'informations génétiques à partir d'images histopathologiques, d'autres approches ont été proposées pour prédire de telles informations (CHEN et al., 2020a; BIAN et al., 2021; WANG et al., 2019; SCHAUMBERG, RUBIN et FUCHS, 2018; LIAO et al., 2020; QU et al., 2021; WANG et al., 2021; HE et al., 2020; CHEN et al., 2020b; TRIPATHI et al., 2022; HONG et al., 2021; TABIBU, VINOD et JAWAHAR, 2019; ZENG et al., 2022). Ces travaux se concentrent principalement sur les organes présentant les cancers les plus meurtriers (FERLAY et al., 2021) tels que le poumon (COUDRAY et al., 2018; WANG et al., 2019; TRIPATHI et al., 2022), le côlon (BIAN et al., 2021; HE et al., 2020), le foie (CHEN et al., 2020a; LIAO et al., 2020; ZENG et al., 2022), le sein (QU et al., 2021; WANG et al., 2021), le pancréas (CHEN et al., 2020b), la prostate (SCHAUMBERG, RUBIN et FUCHS, 2018), l'utérus (HONG et al., 2021) et le rein (TABIBU, VINOD et JAWAHAR, 2019). À partir de ces organes, il existe deux types de méthodes pour recueillir des images médicales. D'une part, les méthodes invasives permettent d'obtenir des images histopathologiques comme dans (COUDRAY et al., 2018; CHEN et al., 2020a; BIAN et al., 2021; SCHAUMBERG, RUBIN et FUCHS, 2018; LIAO et al., 2020; QU et al., 2021; WANG et al., 2021; HONG et al., 2021; TABIBU, VINOD et JAWAHAR, 2019). D'autre part, il existe également des méthodes non invasives qui permettent de créer des images de tomographie par ordinateur (CT) comme dans (WANG et al., 2019; HE et al., 2020; CHEN et al., 2020b; TRIPATHI et al., 2022). Dans les deux cas, les images médicales sont utilisées de manière similaire pour entraîner des réseaux de neurones profonds.

Parmi les approches proposées, on retrouve principalement des méthodes basées sur ResNet (SCHAUMBERG, RUBIN et FUCHS, 2018; QU et al., 2021; WANG et al., 2021; HE et al., 2020; CHEN et al., 2020b; TRIPATHI et al., 2022; HONG et al., 2021; TABIBU, VINOD et JAWAHAR, 2019), qui utilisent des *skip connections* pour réduire le phénomène de *vanishing gradient*. Ensuite, l'architecture Inception-v3, qui a notamment été utilisée par (COUDRAY et al., 2018) pour montrer la faisabilité de la prédiction de mutations dans des images histopathologiques, a également été reprise (CHEN et al., 2020a; HONG et al., 2021). La spécificité des architectures Inception est d'utiliser des filtres de tailles différentes pour chaque niveau afin de réduire le sur-apprentissage. En plus de ces travaux, principalement basés sur ResNet et Inception, il existe des articles qui utilisent d'autres architectures : CNN (LEE et al., 2018; LIAO et al., 2020), DNN (LEUR et al., 2021), DenseNet (WANG et al., 2019), EfficientNet (TRIPATHI et al., 2022). Parmi les approches qui montrent des résultats encourageants, on trouve des tentatives fusions de plusieurs idées, comme dans (ZENG et al., 2022), où les chercheurs proposent notamment d'utiliser un mécanisme d'attention sur des caractéristiques extraites à l'aide d'un réseau ResNet. Notons que ces réseaux de neurones profonds sont principalement utilisés pour classer directement les patchs en fonction de leurs mutations et sous-types. Cependant, certains travaux tentent d'inclure des données de patients en plus des patchs afin de donner plus de contexte au réseau, comme dans (TRIPATHI et al., 2022). D'une manière différente, l'article (VINÇON et al., 2022) présente une approche pour réduire le nombre de séquençage ADN à réaliser en laboratoire pour économiser et réserver les ressources pour les cas cliniques les plus complexes, en éliminant les cas triviaux (i.e. score de

précision et de rappel élevés) à l'aide d'un réseau EfficientNet. Prêt d'un tiers des séquences peuvent ainsi être évités, en réduisant la précision de seulement 5%. Au vu des travaux actuels, les nombreuses couches des architectures d'apprentissage profond semblent pouvoir extraire des caractéristiques abstraites spécifiques aux mutations, cependant à notre connaissance, il n'existe pas de comparaison entre ces architectures sur la tâche de prédiction de mutation.

Bien que les résultats des réseaux sur la tâche de prédiction d'information génétique soient prometteurs, les travaux existants peinent à justifier les prédictions des réseaux de neurones profonds, comme décrit dans (CHEN et al., 2020a). Afin de contourner cet effet de boîte noire, (BIAN et al., 2021) étudie les relations entre certains gènes en utilisant t-SNE (MAATEN et HINTON, 2008). En parallèle, (LEUR et al., 2021 ; WANG et al., 2019) a exploré l'utilisation de CAM (OQUAB et al., 2015) pour tenter d'expliquer visuellement les prédictions des réseaux. En effet, les méthodes d'explicabilité telles que CAM et LIME (RIBEIRO, SINGH et GUESTRIN, 2016) permettent de visualiser l'impact de chaque partie d'un patch dans son résultat final de classification. Ces méthodes sont particulièrement utiles pour comprendre les décisions des réseaux qui tendent à devenir plus complexes pour traiter des tâches plus difficiles (BURKART et HUBER, 2021), comme la classification des mutations. Malgré ces initiatives, le comportement des réseaux sur la tâche de prédiction de mutation n'est pas bien compris, notamment en raison de la nature abstraite des caractéristiques due au grand nombre de paramètres dans les réseaux, comme décrit dans (WANG et al., 2019).

Dans le cadre du projet AiCOLO financé par l'INSERM / Plan Cancer, qui réunit des experts en IA et en pathologie provenant trois laboratoires, nous proposons de détecter la mutation KRAS G12C à l'aide de réseaux de neurones profonds directement depuis des images histopathologiques issues du jeu de données AiCOLO contenant des WSI de biopsie du côlon ainsi que des informations génétiques. Ensuite, nous comparons différentes architectures d'apprentissage profond sur la tâche de classification des mutations et des variants dans le cancer colorectal. Enfin, les modélisations ainsi créées ont été utilisées par les méthodes d'explicabilité CAM et LIME pour visualiser les zones des patches qui motivent la prise de décision d'un réseau.

3.2 Matériels et méthodes

Dans cette section, nous présentons notre approche pour prédire des informations génétiques à l'aide d'algorithmes d'apprentissage profond, détaillée dans la figure 3.2. Tout d'abord, nous décrivons nos données dans la section 3.2.1. Ensuite, dans la section 3.2.2, nous expliquons notre méthode, avant de présenter les résultats obtenus dans la partie suivante, dans la section 3.3.

3.2.1 Matériels

3.2.1.1 AiCOLO

Le jeu de données privé AiCOLO contient 143 WSI de cancer colorectal provenant de 105 patients. Toutes les WSI ont été colorées en hématoxyline éosine safran (HES), avant d'être numérisées à l'aide d'un scanner photonique Hamamatsu à un grossissement X20 et une résolution spatiale de $0,454\mu\text{m}/\text{pixel}$. Le scanner numérise ensuite le tissu à très haute résolution pour produire des WSI de taille allant de 10 000 à 100 000 pixels par côté. Les WSI proviennent du Centre Georges-François Leclerc de Dijon.

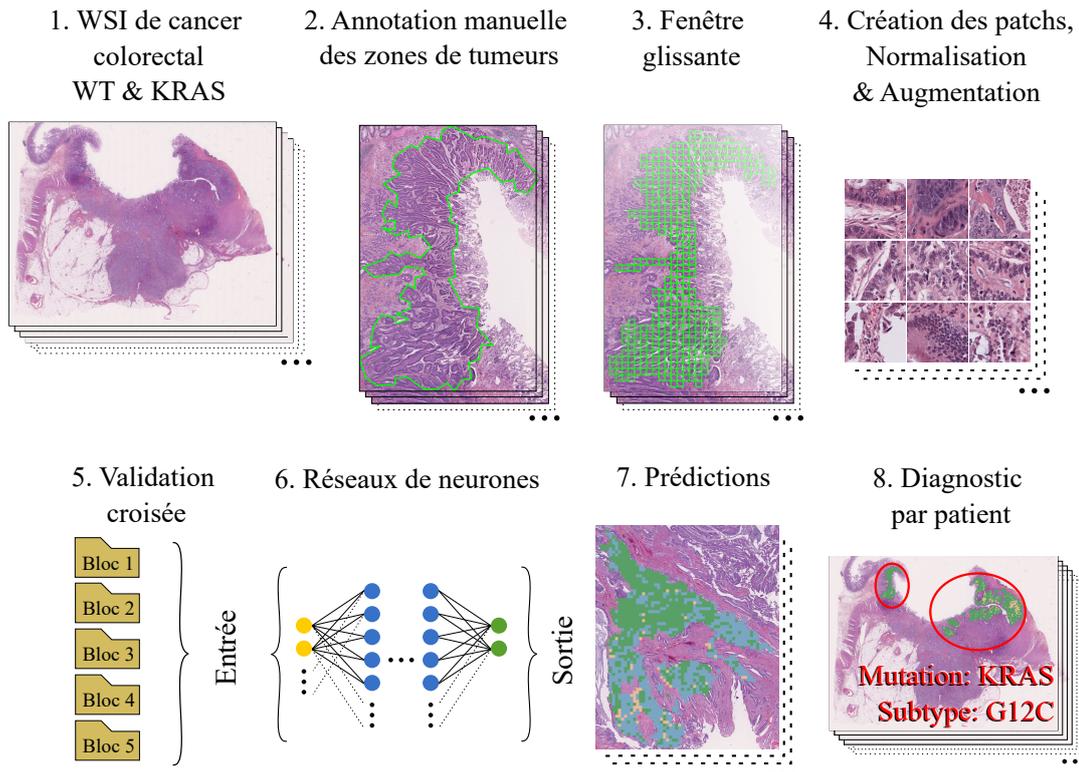


FIGURE 3.2 – Notre méthode. Une collection de WSI de cancer colorectal contenant des mutations KRAS et WT (1) est annotée par tumeur (2). Nous utilisons une fenêtre glissante (3) pour créer des patches de taille 224×224 (4). Ces patches sont organisés en cinq blocs (5) pour effectuer une validation croisée. Ensuite, nous entraînons un réseau de neurones profonds (6) pour prédire la mutation et le sous-type (7). Enfin, nous utilisons les prédictions pour trouver le diagnostic (8).

Le jeu de données AiCOLO contient également des annotations partielles de différentes classes de tissu telles que tumeur, stroma, nécrose, immunitaire, graisse, tissu, arrière-plan, comme montré en figure 3.3. Le logiciel d'analyse d'images collaborative Cytomine (MARÉE et al., 2016) a été utilisé par les experts en pathologie de notre projet pour obtenir ces annotations. 45 WSI contiennent des annotations complètes de toutes les zones tumorales, produites à l'aide du logiciel QuPath (BANKHEAD et al., 2017). Le procédé d'annotation est détaillé en section 3.2.1.2. De plus, des informations génétiques obtenues par séquençage de l'ADN, telles que les gènes mutés (WT, BRAF, KRAS, NRAS, etc) présentés dans le tableau 3.1, sont disponibles pour toutes les WSI. Pour chaque WSI contenant le gène KRAS, nous disposons également d'un variant spécifique (G12C, G12D, G12V, etc) présentés dans le tableau 3.2.

Toutes les WSI du jeu de données AiCOLO contiennent des zones tumorales. Celles qui ne présentent pas de mutation sont dites de type sauvage (WT). Les autres WSI sont identifiées par le gène muté qu'elles contiennent (BRAF, KRAS, NRAS). Il en va de même pour les variants.

Dans la suite de notre travail sur les mutations et les variants, nous choisissons de n'utiliser que les WSI KRAS et WT car elles représentent la majorité de nos WSI (94%), mais aussi parce que si le sous-type de la mutation de ce gène est G12C, il provoque un type de cancer particulièrement agressif qui nécessite un traitement spécifique. Les classes BRAF et NRAS qui ne contiennent pas suffisamment de données

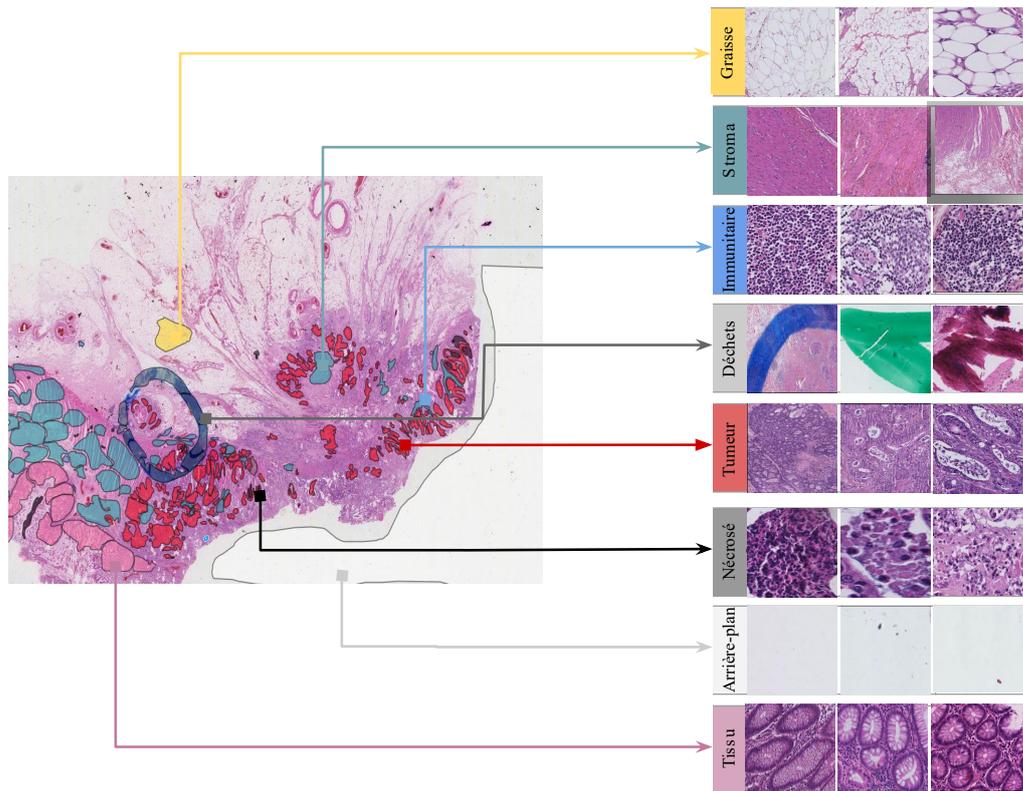


FIGURE 3.3 – Exemples de régions annotées des différentes classes sur les images histopathologiques du jeu de données AiCOLO. Ces annotations des classes graisse, stroma, immunitaire, déchets, tumeur, nécrosé, arrière-plan et tissu ont été réalisées manuellement avec Cytomine par des experts en pathologie. Source : (HAMIDA et al., 2021).

pour entraîner correctement les méthodes d'apprentissage profond avec de la validation croisée sont donc écartées de notre jeu de données. On obtient ainsi 95 WSI KRAS et 39 WSI WT. D'une manière différente, compte tenu du peu de données disponibles pour certains sous-types de mutation KRAS présentés dans le tableau 3.2, nous choisissons de regrouper les classes contenant le moins de données dans une autre classe, pour nous concentrer sur le sous-type G12C tout en conservant des connaissances liés à d'autres mutations. De cette manière on obtient un groupe de 33 WSI G12C et un groupe de 62 WSI d'un autre variant.

3.2.1.2 Annotation complète des zones tumorales

Les mutations génétiques ne peuvent être identifiées que dans les zones tumorales. Bien que les tâches de classification ou de segmentation utilisant des méthodes d'apprentissage profond donnent d'excellents résultats sur les WSI de côlons (HAMIDA et al., 2021), de mauvaises prédictions peuvent introduire des biais dans la suite du processus. De plus, l'étape intermédiaire de prédiction de la zone tumorale nécessite des ressources de calcul importantes. Par conséquent, nous annotons manuellement les WSI en utilisant le logiciel QuPath (BANKHEAD et al., 2017). Tout d'abord, nous segmentons toutes les zones tumorales dans une WSI en utilisant l'outil baguette. Ensuite, nous éliminons les fragments et les trous plus petits que $10e4\mu m^2$. Nous avons simplifié la forme de l'annotation avec un seuil de 300 pixels par arête. Ces

Gène	#WSI	%WSI
KRAS	95	66,4
WT	39	27,3
BRAF	8	5,6
NRAS	1	0,7
Total	143	100

TABLE 3.1 – Nombre de WSI par mutation. Le nombre et le pourcentage de WSI par mutation BRAF, KRAS, NRAS et WT.

Sous-type KRAS	#WSI	%WSI
G12C	33	34,7
G12D	17	17,9
G12V	13	13,7
G13D	12	12,6
A146T	5	5,3
G12S	4	4,2
Q61H	4	4,2
A146V	2	2,1
V14I	2	2,1
Q61K	1	1,1
Q61L	1	1,1
G12P	1	1,1
Total	95	100

TABLE 3.2 – Nombre de WSI par variant. Le nombre et le pourcentage de WSI par variant (G12C, G12D, G12V etc).

procédés nous ont permis de réduire le nombre d’arêtes et de sommets du polygone qui constitue l’annotation, et donc de réduire la complexité et le temps des calculs qui seront effectués ultérieurement lors des étapes de création des patches. Enfin, pour chaque annotation, nous exécutons un script dans l’éditeur de script QuPath afin de générer une image et son masque de vérité terrain à un niveau de grossissement X20. Pour chaque masque, les zones tumorales sont représentées par des pixels blancs et les zones sans tumeur sont noires. Un exemple des différentes étapes d’annotation est présenté dans la figure 3.4.

Lors des séances d’annotation, les cas problématiques ont été répertoriés. Effectivement, pendant le processus de création d’une WSI, des éléments extérieurs peuvent polluer les tissus. Cela explique la présence d’artefacts, de poussière, d’ombre, de tâche ou encore de variation dans la coloration. Des exemples de tels cas sont montrés en figure 3.5. Dans un premier temps, nous écartons de nos expérimentations ces cas particuliers afin de minimiser les biais dans notre jeu de données pour réduire la complexité de l’apprentissage des réseaux. Dans un second temps, après avoir constaté que l’apprentissage des mutations se déroulait correctement sans les cas particuliers, nous les avons réintroduits pour rendre notre méthode robuste à une utilisation en condition réelle.

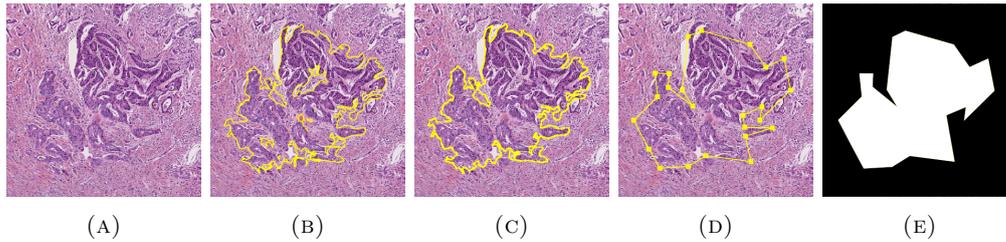


FIGURE 3.4 – Les différentes étapes pour créer un masque de vérité terrain en utilisant le logiciel QuPath. Une zone tumorale sélectionnée est annotée (a) à l’aide de l’outil baguette (b). On élimine les fragments et les trous (c), puis on simplifie la forme de l’annotation (d) en réduisant le nombre d’arêtes du polygone. Enfin, une image et son masque de vérité terrain associé de la zone tumorale (e) sont créés.

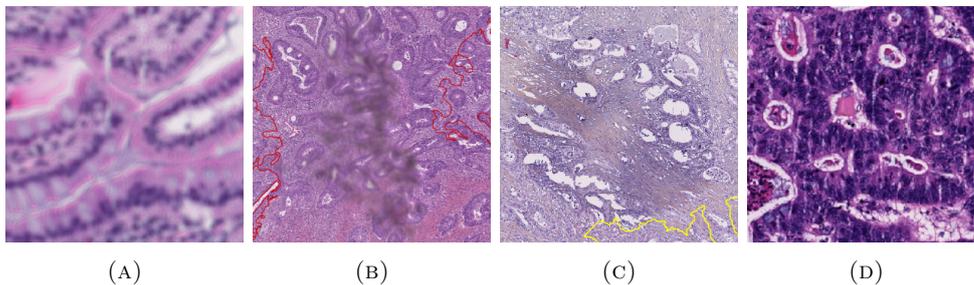


FIGURE 3.5 – Exemples de cas problématiques rencontrés lors du processus d’annotation. Le tissu (a) est flou. (b) contient de la poussière. Dans le tissu (c) on voit une variation dans la coloration. (d) présente des déchirures verticales.

Parmi les 134 WSI de notre jeu de données, 60 WSI étaient inutilisables (mauvaise qualité ou trop peu de zones tumorales) et 38 WSI étaient redondantes (i.e. 2 WSI du même patient). Les 45 WSI restantes ont été entièrement annotées comme le montre le tableau 3.3.

Gène	Sous-type	#WSI	%WSI
WT	NA	19	42,2
KRAS	G12C	13	28,9
KRAS	Autres	13	28,9
Total		45	100

TABLE 3.3 – Nombre de WSI annotées. Le nombre et le pourcentage de WSI par mutation (WT, KRAS) et par variant (G12C, Autres).

3.2.1.3 Création des patches

Nous utilisons une fenêtre glissante sans recouvrement comme montré dans la figure 3.6 pour créer des patches de taille 224×244 pixels⁵ à un niveau de grossissement X20. Les patches sont créés uniquement s’ils sont entièrement dans une zone annotée, afin d’éviter de créer des patches en dehors des zones tumorales qui ne contiennent pas d’information sur les mutations.

5. 224×224 pixels est une taille de patch standard utilisée par exemple par ResNet.

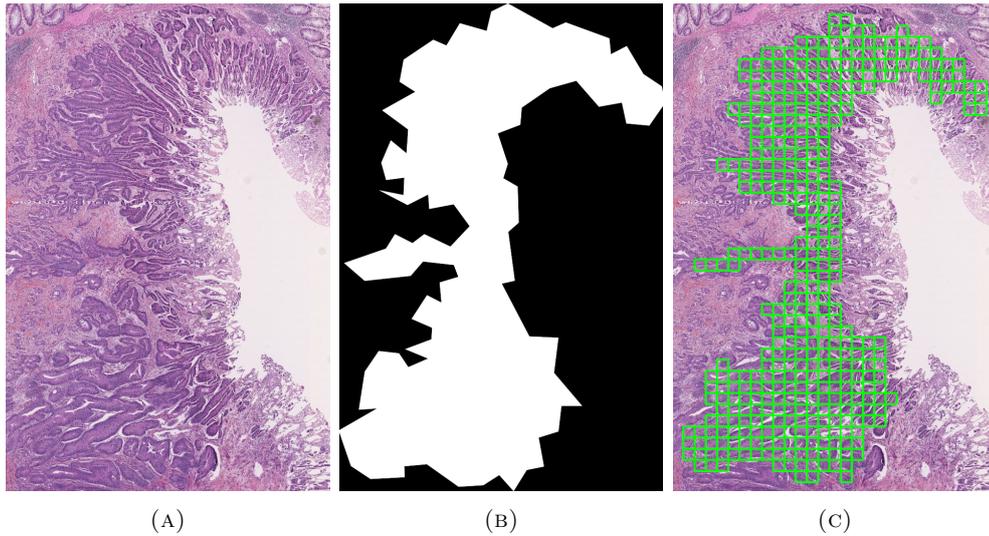


FIGURE 3.6 – Exemple de patches découpés à l'intérieur de l'annotation. Une fenêtre glissante découpe les patches dans le tissu (a), tout en respectant la zone de vérité de base donnée par le masque (b). (c) est une visualisation de tous les patches créés.

Dans un second temps, nous appliquons la méthode de normalisation de la coloration Macenko (MACENKO et al., 2009) implémentée dans la bibliothèque PathML (ROSENTHAL et al., 2022). Même si les WSI sont toutes colorées en HES, elles peuvent avoir des apparences différentes comme montrées en figure 3.7 (a). Cela vient des manipulations effectuées lors de la création des WSI. Ces différences sont un frein à la compréhension des méthodes d'apprentissage profond. C'est pourquoi, cette étape de normalisation est indispensable afin d'obtenir des patches de couleurs similaires comme présenté en figure 3.7 (b).

3.2.1.4 Augmentation de données

Afin de contrer le manque de données et le phénomène de sur-apprentissage, nous utilisons des méthodes d'augmentation des données. Ces méthodes appliquent des transformations aux images pour obtenir une nouvelle représentation des données avec plus de diversité, afin de forcer le réseau à se concentrer sur la généralisation et non sur l'apprentissage des données.

De plus, de nombreux artefacts sont introduits lors des manipulations nécessaires à la création de WSI (fixation des tissus, déshydratation, inclusion en paraffine, sectionnement, coloration...) (KANWAL et al., 2022). Nous prenons soin d'appliquer des transformations qui pourraient se produire dans des conditions réelles afin de rendre nos réseaux robustes à ces artefacts.

Ainsi, nous appliquons aléatoirement les transformations suivantes aux patches :

- Retournement horizontal ;
- Retournement vertical ;
- Rotation entre 0 et 45 degrés ;
- Cisaillement entre 0 et 20 degrés le long d'un axe ;
- Zoom sur une partie du patch.

Un exemple de ces transformations sur un patch est illustré dans la figure 3.8.

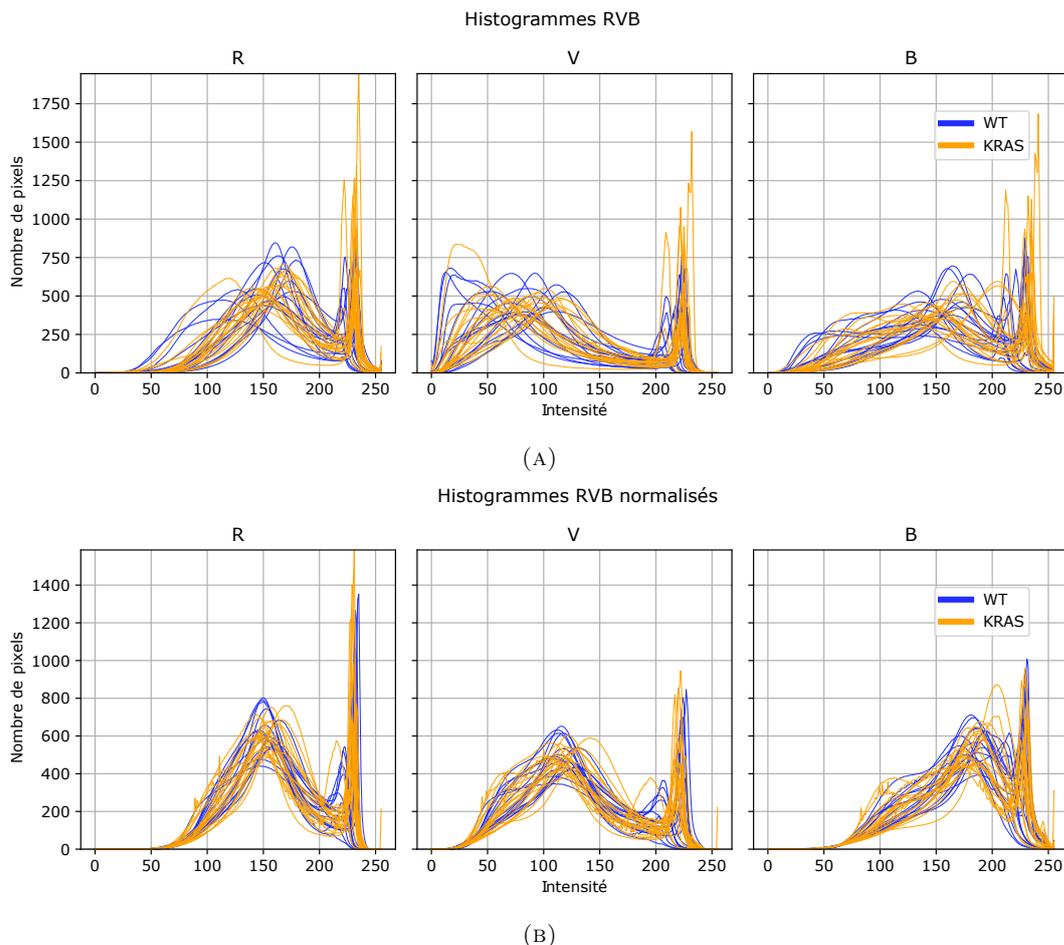


FIGURE 3.7 – Les histogrammes RVB des WSI avant (a) et après (b) la normalisation de la coloration avec Macenko. Chaque courbe représente la moyenne des histogrammes pour tous les patches d’une WSI. Les courbes bleues représentent les WSI qui appartiennent à la classe WT et les courbes orange sont les WSI de la classe KRAS.

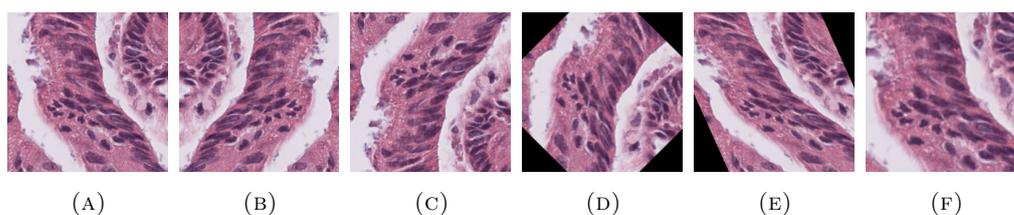


FIGURE 3.8 – Exemples d’augmentation des données. Le patch original (a) subit un retournement horizontal (b), un retournement vertical (c), une rotation (d), un cisaillement (e) et un zoom (f).

3.2.1.5 Jeux de données mutation : WT vs KRAS

Dans un premier temps, nous étudions la présence ou l’absence de mutation du gène KRAS dans les WSI de cancer colorectal. Pour cela, nous créons un premier jeu de données à partir des patches créés en 3.2.1.3, puis augmentées en 3.2.1.4.

A cette étape, le jeu de données est déséquilibré (i.e. le nombre de patches extraits par WSI et par classe est différent). Un réseau entraîné sur ces données prédira la classe majoritaire et échouera à généraliser la tâche souhaitée (JOHNSON et

KHOSHGOFTAAR, 2019). Afin de limiter les biais dans notre jeu de données, nous supprimons aléatoirement et proportionnellement des patchs dans chaque WSI.

Ainsi, nous obtenons 76 000 patchs comme présentés en figure 3.9 (a), dont 38 000 patchs où le gène KRAS est muté et 38 000 patchs où le gène KRAS est de type sauvage (WT). Notons que le nombre de patchs extraits par WSI est hétérogène comme présenté en figure 3.9 (b).

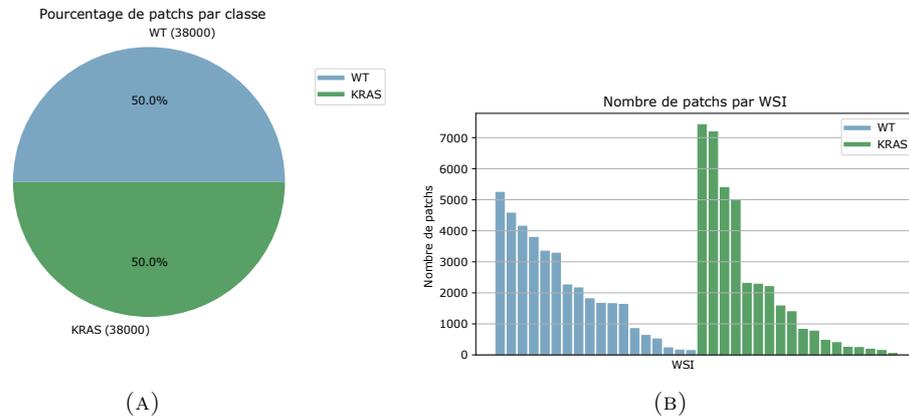


FIGURE 3.9 – Le nombre de patchs par classe (b) et par WSI (a). Le bleu correspond à la classe WT, le vert à la classe KRAS.

3.2.1.6 Jeux de données variant : WT vs G12C vs Autres

Dans un second temps, nous nous concentrons sur le sous-type G12C du gène KRAS. Nous procédons de la même manière que dans la partie précédente pour créer un jeu de données comprenant 70 500 patchs comme présenté en figure 3.10 (a), dont 23 500 patchs où le sous-type du gène KRAS muté est G12C, 23 500 patchs où le sous-type du gène KRAS muté n'est pas G12C, et 23 500 patchs où le gène KRAS est de type sauvage (WT). Notons également que le nombre de patchs extrait par WSI est hétérogène comme présenté en Figure 3.10 (b).

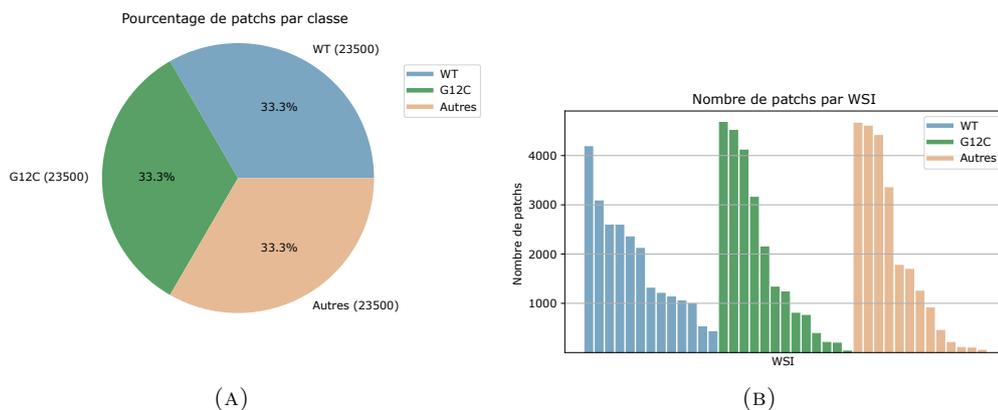


FIGURE 3.10 – Le nombre de patchs par classe (b) et par WSI (a). Le bleu correspond à la classe WT, le vert à la classe G12C et le beige aux autres sous-types.

3.2.2 Méthodes

3.2.2.1 Validation croisée

Pour toutes nos expérimentations, nous utilisons la validation croisée avec cinq blocs (trois pour l’entraînement, un pour la validation et un pour le test) comme présenté dans le tableau 3.4 afin de tirer pleinement parti de nos données et de réduire le sur-apprentissage. Nous distribuons les WSI des différentes classes de manière égale entre les cinq blocs pour éviter que nos réseaux associent la classe la plus représentée à toutes les données. De plus, nous prenons soin d’éviter les fuites de données en plaçant les patches d’une WSI (i.e. patient) dans un seul bloc afin qu’il soit impossible pour un réseau de tester des patches d’un patient qu’il aurait pu utiliser pendant l’étape d’apprentissage. Chaque bloc est utilisé une fois pour tester un modèle entraîné sur tous les autres blocs.

#Réseau	#Blocs dans		
	Test	Validation	Train
1	1	2	3, 4, 5
2	2	3	4, 5, 1
3	3	4	5, 1, 2
4	4	5	1, 2, 3
5	5	1	2, 3, 4

TABLE 3.4 – Les blocs par jeu. Répartition des blocs dans les jeux d’entraînement, de validation et de test pour chaque réseau de la validation croisée.

3.2.2.2 Réseaux de neurones profonds

Comme nous voulons détecter le sous-type G12C de la mutation KRAS, nous procédons en deux étapes. Tout d’abord, nous entraînons notre modèle à discriminer la mutation KRAS (jeu de données KRAS vs WT). Ensuite, nous répétons nos expériences pour discriminer le sous-type G12C de la mutation KRAS (jeu de données G12C vs Autres vs WT). Nous choisissons de comparer les performances des architectures d’apprentissage profond que nous sélectionnons en fonction de l’état de l’art et des travaux similaires dans la section 3.1.3 : DenseNet201 (HUANG et al., 2017), EfficientNetV2L (TAN et LE, 2021), Inception-v3 (SZEGEDY et al., 2016), InceptionResNetV2 (SZEGEDY et al., 2017), MobileNetV2 (SANDLER et al., 2018), NASNetLarge (ZOPH et al., 2018) et ResNet152V2 (HE et al., 2016). À l’exception de l’étape de normalisation par la méthode de Macenko qui est réalisée en prétraitement, toutes les méthodes qui transforment ou augmentent les patches sont réalisées à la volée pendant l’apprentissage.

Nous utilisons ces hyperparamètres trouvés par optimisation :

- Taux d’apprentissage : 0.01
- Epoques : 30
- Taille de *batch* : 32
- Fonction de coût : Categorical crossentropy
- Fonction d’optimisation : RMSprop (GRAVES, 2013)
- Métrique : Précision
- *Dropout* : 0.1
- Regularisation L2 : 0.0001

— Apprentissage par transfert : Aucun

3.2.2.3 Prédictions

Les prédictions des réseaux sur les cinq blocs de test sont ensuite agrégées pour avoir la performance sur l'ensemble du jeu de données. Un exemple de ces prédictions est montré dans la figure 3.2 (7) pour les tâches de classification de mutation du gène KRAS et de classification du sous-type G12C du gène KRAS. Nous attribuons pour chaque WSI (i.e. patient) la classe la plus prédite parmi les patches qui la composent, comme le montre la figure 3.11. Puis, nous calculons la précision sur toutes les WSI bien et mal classées par le réseau. Nous détaillons d'abord la précision obtenue pour chaque classe (G12C, Autres et WT), puis nous présentons la précision pour les tâches G12C vs Autres vs WT et G12C vs NonG12C.

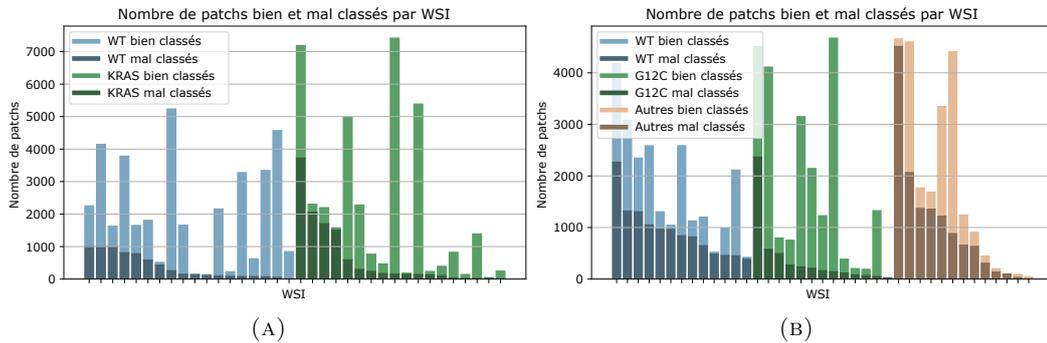


FIGURE 3.11 – Nombre de patches bien et mal classés par WSI. (a) montre les prédictions de patches pour la tâche de classification KRAS effectuée par MobileNetV2, et (b) est pour la tâche de classification du sous-type G12C. Les patches bien classés sont représentés en clair tandis que les patches mal classés sont en foncé.

3.2.2.4 Explicabilité

Comprendre les prédictions des réseaux est l'un des défis majeurs dans la tâche de prédiction des mutations génétiques. Dans ce contexte, les méthodes d'explicabilité les plus couramment utilisées, telles que CAM (Class Activation Map) (ZHOU et al., 2016), permettent de contourner l'effet boîte noire des réseaux. Dans les travaux similaires sur les mutations, on retrouve l'utilisation de CAM. Ainsi, nous choisissons d'implémenter CAM, mais également LIME (RIBEIRO, SINGH et GUESTRIN, 2016) pour son côté agnostique qui nous permet d'étudier l'importance des zones des patches dans la prédiction donnée par le réseau. En effet, LIME fait varier l'entrée d'un réseau en masquant certaines parties d'un patch donné, afin de déterminer les zones qui influencent le plus le choix de la prédiction en sortie.

3.3 Expérimentations

À présent, nous appliquons notre méthode décrite en section 3.2.2 sur nos jeux de données présentés en section 3.2.1 afin de prédire des informations génétiques. D'abord, nous commençons par prédire en section 3.3.1 la présence ou l'absence de mutation du gène KRAS dans des patches du jeu de données KRAS vs WT. Une fois cette expérimentation concluante, nous avons poussé nos investigations à une échelle plus fine sur les sous-types du gène KRAS. Ainsi, d'une manière similaire

à la première tâche, nous utilisons le jeu de données G12C vs Autres vs WT pour classer les sous-types du gène KRAS dans la section 3.3.2. Enfin, nous exploitons en section 3.3.3 les prédictions des réseaux de neurones des précédentes expérimentations afin de diagnostiquer automatiquement la mutation du gène KRAS et le sous-type G12C pour les patients de notre étude.

3.3.1 Apprentissage profond pour la classification de la mutation du gène KRAS

Dans un premier temps, nous entraînons des réseaux de neurones profonds sur la tâche de classification de patches KRAS vs WT. Nous commençons par optimiser les paramètres de nos réseaux en section 3.3.1.1, avant de comparer différentes architectures d'apprentissage profond dans le tableau 3.5 de la partie 3.3.1.2.

3.3.1.1 Optimisation des hyperparamètres

Un réseau de neurones a de nombreux hyperparamètres qui influent sur la capacité du modèle à trouver une solution optimale qui minimise la fonction de coût lors de l'apprentissage. Il est essentiel de trouver la meilleure configuration de paramètres afin d'améliorer les performances d'un modèle comme le souligne (YANG et SHAMI, 2020). La méthode d'optimisation d'hyperparamètres *grid search* nécessitent d'entraîner de nombreux réseaux afin de déterminer les meilleures combinaisons de valeurs. Cette méthode garantie de trouver les meilleurs couples d'hyperparamètres, mais cela nécessite de trop grandes ressources de calcul pour être réalisées dans notre cas où nos données sont des images et où nous utilisons de la validation croisée. De ce fait, pour garder un temps de calcul réalisable, nous choisissons d'utiliser la méthode d'optimisation *random search*, où l'on teste aléatoirement des couples d'hyperparamètres. À noter que bien que cette méthode réduise considérablement la complexité de recherche des hyperparamètres, il n'est pas garanti que nous trouvions les meilleurs valeurs. Pour choisir les valeurs par défaut, nous nous appuyons sur les articles qui appliquent des réseaux de neurones sur des tâches de classification de mutation, ceux présentés en section 3.1.3. Dans la suite de cette partie, nous présentons les hyperparamètres les plus impactants dans notre cas d'étude.

L'Apprentissage par transfert est une procédure standard dans le domaine de la vision par ordinateur pour diminuer le temps de calcul. Elle consiste à transférer les connaissances apprises par un premier modèle sur un jeu de données source à un second modèle qui sera affiné sur un jeu de données cible. L'apprentissage par transfert repose sur l'idée que les images contiennent des caractéristiques communes, quelque soit leur type. En partant de cette idée, nous pré-entraînons des réseaux de neurones sur deux jeux de données : ImageNet (DENG et al., 2009) et BreakHis (SPANHOL et al., 2015). Dans un premier temps, nous utilisons les 14 197 122 images du jeu de données ImageNet, car l'important volume de données permet d'extraire des caractéristiques primaires pertinentes. Dans un second temps, nous avons recours au jeu de données BreakHis, qui contient 9 109 patches de tumeurs bénignes et malignes dans le cas du cancer du sein. Nous justifions ce choix par la similarité des domaines entre les données de AiCOLO, et celles de BreakHis. En effet, dans les deux cas, les données sont des images histopathologiques colorées avec des colorants similaires (H&E et HES). Les structures des tissus ont des caractéristiques spécifiques que l'on ne retrouve pas dans les images de scène du quotidien comme dans ImageNet. De ce fait, en utilisant un jeu de données d'un domaine proche, nous espérons transmettre aux réseaux des caractéristiques propres à l'identification de structures biologiques. Nos

expérimentations sur l'utilisation des jeux de données ImageNet et BreakHis montrent que les réseaux de neurones échouent à apprendre la tâche de classification de la mutation du gène KRAS. Cependant, lorsque nous entraînons un réseau de neurones à partir de poids initialisés aléatoirement (*from scratch*), nous obtenons une précision de 67%. Il semble que les caractéristiques utiles à la prédiction d'informations génétiques s'éloignent des caractéristiques pertinentes dans les tâches plus visuelles où l'on peut déterminer à l'oeil nu une vérité terrain. Par conséquent, nous choisissons d'entraîner nos réseaux *from scratch*. Avant d'arriver à cette conclusion, nous avons également varié les méthodes d'affinage, sans parvenir à de meilleurs résultats.

L'affinage (*fine-tuning*) permet de choisir le niveau de complexité des caractéristiques qui sont transférées à un réseau lors d'un apprentissage par transfert. Il est admis que les couches les plus basses d'un réseau contiennent les caractéristiques les plus simples. De plus, la succession des opérations de convolutions complexifie la représentation des caractéristiques. Ainsi, les couches proches de la sortie d'un réseau extraient des caractéristiques complexes, spécifiques à la tâche demandée. Suivant la tâche souhaitée et la similarité des données utilisées, il peut être intéressant de conserver des caractéristiques primaires ou plus complexes pour améliorer l'apprentissage. En partant de cette idée, nous avons varié le nombre de couches gelées provenant d'ImageNet et de BreakHis avant de ré-entraîner le reste des couches sur AiCOLO. Nous observons que ré-entraîner toutes les couches d'un réseau permet d'obtenir des performances légèrement meilleures. Cependant, nous retenons la méthode *from scratch* qui obtient de meilleurs résultats. Effectivement, il semble que même si le réseau n'est pas gelé, le simple fait d'avoir appris à reconnaître des caractéristiques trop éloignées de sa tâche initiale, ne permet pas d'apprendre par la suite correctement des informations génétiques.

L'augmentation de données que nous effectuons est décrite en section 3.2.1.4. Nous expérimentons indépendamment l'utilisation de chaque méthode (retournement, rotation, cisaillement, zoom) afin de comparer les performances. À noter que les transformations sélectionnées sont toutes susceptibles de se produire dans un cas réel. D'abord, nous observons que les méthodes sélectionnées améliorent les précisions de 6.4% à 7.1% comparé à un réseau qui n'utilise pas d'augmentation de données. Ensuite, il apparaît qu'utiliser en même temps toutes les méthodes d'augmentation de données améliore la précision de 2.1% comparé aux réseaux qui utilisent une seule méthode d'augmentation de données. L'application de transformations variées permet d'améliorer la robustesse d'un réseau. De ce fait, nous choisissons d'appliquer l'ensemble des méthodes d'augmentation de données testées. Selon ces résultats, il est possible qu'utiliser davantage de méthodes d'augmentation permette d'augmenter encore les résultats des réseaux.

La fonction d'optimisation est utilisée pour minimiser la fonction de coût en diminuant le taux d'apprentissage et en modifiant les poids d'un réseau. Nous comparons les résultats de nos réseaux obtenus en variant la fonction d'optimisation. Pour cela, nous sélectionnons des *optimizers state-of-the-art* et des *optimizers* utilisées dans les travaux appliquant des algorithmes d'apprentissage sur des mutations, présentés en section 3.1.3. Ainsi, nous comparons Adadelta (ZEILER, 2012), Adagrad (DUCHI, HAZAN et SINGER, 2011), Adamax (KINGMA et BA, 2014), FTRL (MCMAHAN et al., 2013), Nadam (DOZAT, 2016), RMSprop (GRAVES, 2013), SGD (KIEFER et WOLFOVITZ, 1952). Nous observons que la fonction d'optimisation Adam et ses variantes se démarquent. Par exemple, Adam bat SGD de 10%. Nous sélectionnons RMSprop, qui obtient les meilleurs résultats en battant SGD de 13.3%.

La taille du *batch* est le nombre de patchs qui seront envoyés au réseau en une fois. D'une part, une taille de *batch* important empêche un réseau de généraliser. À l'inverse, un *batch* trop faible augmente le temps de calcul. Il convient de faire un compromis entre le temps de calcul et la généralisation du réseau. Nous obtenons 1.1% de précision de différence entre le meilleur et le moins bon réseau testé sur les *batches* 16, 32, 64 et 128. Considérant ce faible écart ainsi que les *batches* utilisés par des travaux de recherches similaires présentés dans la section 3.1.3, nous choisissons un *batch* de 32.

Le *dropout* et la régularisation L1 L2 permettent de combattre le sur-apprentissage et aident les réseaux à généraliser. Avec la méthode du *dropout*, certains poids sont ignorés aléatoirement pendant l'apprentissage. La régularisation quant à elle, applique des pénalités sur les poids. Nous expérimentons l'utilisation de ces méthodes indépendamment afin de les comparer. De plus, nous varions les valeurs de ces paramètres pour impacter plus ou moins les poids des réseaux. Nous constatons qu'ignorer 10% des poids améliore de 2% la précision d'un réseau. Concernant la régularisation L2, nous trouvons qu'appliquer une pénalité de 0.0001 améliore de 8% la précision.

Le seuil de zone tumorale dans un patch est le pourcentage de l'annotation requise pour créer un patch. Lors de l'étape d'annotation manuelle, les frontières des annotations peuvent être imprécises. Ainsi, il arrive que les frontières des tumeurs soient à l'intérieur ou à l'extérieur de la zone annotée. Or, ces frontières tumorales contiennent peut-être des informations utiles à la prédiction de mutation. En variant le seuil de zone tumorale, nous observons que plus le seuil de zone tumorale est faible et plus les résultats baissent (jusqu'à -7%). Effectivement, un des risques de diminuer ce seuil est que des patchs dans des zones qui ne contiennent pas de tumeurs soient créés. Cela peut également s'expliquer par le fait que les caractéristiques utiles à la prédiction d'informations génétiques se trouvent majoritairement au milieu d'une zone tumorale. Pour avoir les meilleures performances, il convient de créer des patchs pleinement dans une zone tumorale.

Les niveaux de grossissement permettent d'apprécier les structures biologiques à différentes échelles. Le niveau utilisé impacte directement la représentation des tissus en modifiant ses frontières et le contexte environnant. Ainsi, certains niveaux permettent l'extraction de caractéristiques plus pertinente suivant la tâche demandée. Nous comparons les résultats obtenus par des niveaux de grossissement 5X, 10X et 20X. Il apparaît que plus le niveau de grossissement est élevé, et plus la précision est haute. Nous choisissons donc le niveau de grossissement 20X.

3.3.1.2 Comparaison des architectures

Une fois l'optimisation des hyper-paramètres réalisée, nous obtenons :

- Taux d'apprentissage : 0,01
- Époques : 30
- Taille de lot : 32
- Fonction de coût : Entropie croisée
- Fonction d'optimisation : RMSprop
- Taux de *dropout* : 0,1
- Régularisation L2 : 0,0001

À présent, nous passons à la comparaison d’architecture de réseaux de neurones profonds. À nouveau, nous nous appuyons sur l’état de l’art et les travaux de recherche listés en section 3.1.3 pour choisir les architectures. Ainsi, nous comparons sur la tâche de classification de la mutation du gène KRAS, les architectures DenseNet201 (HUANG et al., 2017), EfficientNetV2L (TAN et LE, 2021), Inception-v3 (SZEGEDY et al., 2016), InceptionResNetV2 (SZEGEDY et al., 2017), MobileNetV2 (SANDLER et al., 2018), NASNetLarge (ZOPH et al., 2018) et ResNet152V2 (HE et al., 2016). Nous présentons les résultats dans le tableau 3.5.

Architectures	Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5	All Folds
InceptionResNetV2	0.782	0.745	0.784	0.681	0.850	0.769 (± 0.056)
MobileNetV2	0.832	0.667	0.816	0.648	0.866	0.766 (± 0.090)
DenseNet201	0.783	0.650	0.845	0.681	0.833	0.758 (± 0.079)
ResNet152V2	0.780	0.670	0.830	0.646	0.853	0.756 (± 0.084)
NASNetLarge	0.861	0.571	0.811	0.670	0.845	0.752 (± 0.113)
Inception-v3	0.788	0.763	0.528	0.710	0.809	0.720 (± 0.101)
EfficientNetV2L	0.764	0.496	0.709	0.626	0.500	0.619 (± 0.108)

TABLE 3.5 – Comparaison des architectures pour la classification de la mutation KRAS. Comparaison de la précision des architectures sur la tâche de classification des patches avec ou sans mutations du gène KRAS.

De manière générale, l’entraînement des réseaux semble instable, les réseaux peinent à converger vers une solution et sur-apprennent rapidement les données. On remarque d’ailleurs de grande disparité dans les précisions obtenues sur les différents blocs (e.g. écart-type allant de 0,056 à 0,113).

Malgré tous les efforts fournis lors des étapes de prétraitement pour réduire les biais dans le jeu de données (Normalisation de la coloration, augmentation de données, équilibrage des données etc), il apparaît que la tâche de classification de la mutation KRAS reste difficile pour les réseaux d’apprentissage profond. Les difficultés rencontrées par nos réseaux peuvent être expliquées par les données d’entraînement mal labellisées que nous donnons aux réseaux. En effet, nous attribuons à tous les patches d’une WSI la même classe, mais les signatures des mutations et des variants ne sont pas nécessairement présentes dans chaque patch. Nous remarquons que dans nos expérimentations, toutes les architectures utilisant des connexions résiduelles obtiennent de meilleurs résultats. De manière surprenante, MobileNetV2 se classe parmi les meilleures architectures sur la tâche de classification de mutation du gène KRAS, alors que sur ImageNet, MobileNetV2 peine à rivaliser avec les autres architectures. Il semble que les convolutions *depth-wise separable* qui font la spécificité de MobileNet permettent d’extraire des caractéristiques pertinentes, tout en garantissant un temps de calcul bien moindre.

Nos expérimentations montrent que les architectures InceptionResNetV2, MobileNetV2, DenseNet201, ResNet152V2, NASNetLarge, Inception-v3 et EfficientNetV2L parviennent tout de même à extraire certaines des caractéristiques liées à la mutation KRAS.

Bien que nos résultats montrent que l’extraction de caractéristiques génétiques à partir d’images de lames est possible, nous pensons que les réseaux de neurones profonds que nous avons utilisés sont trop complexes (i.e. énormément de paramètres) et pas assez spécialisés pour la tâche de classification de mutation.

3.3.2 Apprentissage profond pour la classification du sous-type G12C du gène KRAS

Suite aux résultats encourageants obtenus sur la classification de patch de mutation du gène KRAS, nous continuons nos investigations à une échelle encore plus fine. Effectivement, nous tentons maintenant de déterminer le sous-type G12C de la mutation du gène KRAS. Nous conservons la même méthode et les mêmes hyperparamètres que nous appliquons au jeu de données sur le sous-type G12C. On observe dans le tableau 3.6 des tendances similaires comparé aux résultats de la tâche de classification de la mutation KRAS. Cette fois encore, les réseaux MobileNetV2 et InceptionResNetV2 se démarquent des autres architectures avec 57,4% et 53,8% de précision pour la tâche de classification G12C vs Autres vs WT.

Architectures	Classification des patches	Diagnostic par patient				
	Average accuracy (5 blocs)	G12C	Autres	WT	G12C vs Autres vs WT	G12C vs NonG12C
MobileNetV2	0.574 (± 0.135)	0.923	0.462	0.462	0.615	0.897
InceptionResNetV2	0.538 (± 0.113)	0.923	0.538	0.538	0.667	0.821
NASNetLarge	0.483 (± 0.111)	0.846	0.385	0.538	0.590	0.769
ResNet152V2	0.478 (± 0.151)	0.769	0.231	0.462	0.487	0.746
Inception-v3	0.482 (± 0.079)	0.692	0.154	0.692	0.513	0.641
DenseNet201	0.428 (± 0.166)	0.615	0.231	0.385	0.410	0.641
EfficientNetV2L	0.463 (± 0.115)	0.846	0.462	0.000	0.536	0.615

TABLE 3.6 – Comparaison des architectures pour la classification du sous-type G12C, et pour le diagnostic assisté par ordinateur par patient. Comparaison de la précision des architectures sur la tâche de classification par patch du sous-type G12C du gène KRAS. Un diagnostic assisté par ordinateur par patient est calculé à partir des résultats de la classification par patch.

3.3.3 Diagnostic assisté par ordinateur du sous-type G12C

Enfin, nous proposons un diagnostic assisté par ordinateur en utilisant les résultats de classification de patches comme présenté dans la section 3.3.2. Le tableau 3.6 montre les précisions des diagnostics obtenues sur 39 patients. Lorsque nous examinons les résultats de la classification des WSI, toujours sur la tâche G12C vs Autres vs WT, nous constatons que la précision est plus élevée que pour la classification des patches. Si nous examinons en détail la précision pour chacune des classes de la classification par WSI (comme présenté dans les colonnes G12C, Autres et WT), nous voyons que les classes contenant les autres variants (Autres) et le type sauvage (WT) sont confondues par les réseaux. Nous pensons que cela est dû aux données limitées dont nous disposons pour ces classes. Puisque nous nous concentrons sur la détection du variant G12C de la mutation KRAS en raison des formes de cancer agressives qu'elle crée, nous pouvons réduire notre problème à deux classes seulement en combinant les classes Autres et WT. Ainsi, dans la colonne G12C vs NonG12C, nous pouvons voir que nos réseaux reconnaissent correctement le variant G12C.

Plusieurs réseaux de neurones profonds atteignent une précision élevée dans la tâche de classification des mutations KRAS G12C par WSI (G12C vs NonG12C) avec 74,6% pour ResNet152V2, 76,9% pour NASNetLarge, 82,1% pour InceptionResNetV2 et 89,7% pour MobileNetV2. Nous notons que la précision avec laquelle nos réseaux classifient la mutation KRAS G12C est comparable aux résultats rapportés par (BIAN et al., 2021) sur les WSI HES de cancer colorectal. A notre connaissance,

nous proposons les meilleurs résultats de la littérature sur la tâche de détection de la mutation KRAS G12C par patient en utilisant nos modèles MobileNetV2.

Dans cette étude, nous avons réussi à prédire le sous-type G12C du gène KRAS. Cependant, cela n'a pas été le cas pour les autres variants disponibles. En plus des limitations évoquées précédemment concernant la classification de la mutation KRAS, il n'est pas évident si tous les variants peuvent être détectés dans des WSI, ou si cela est dû à une limitation de nos conditions d'expérimentation. Le manque de données peut également être à l'origine de l'échec de la classification des mutations et variants les moins représentés. Nous pensons que nos réseaux sont plus performants pour la classification basée sur les patients plutôt que sur les patches, car cette approche lissent les erreurs du réseau.

3.4 Vers l'explicabilité des mutations génétiques

3.4.1 Réflexion sur l'explicabilité et l'interprétabilité des mutations génétiques

3.4.1.1 L'importance des données et leurs biais

Les algorithmes d'intelligence artificielle ont besoin de données en quantité et de qualité pour apprendre correctement une tâche. C'est d'autant plus vrai pour les réseaux de neurones profonds qui surpassent les autres méthodes d'apprentissage sur des jeux de données volumineux et à nombre de dimensions élevés (LECUN, BENGIO et HINTON, 2015). Lorsque l'on souhaite modéliser une tâche, on doit d'abord disposer de données acquises en condition réelle. Pendant la phase de récolte des données, l'objectif est de retranscrire le plus fidèlement la réalité dans les données. Cependant, les technologies d'acquisition qui transforment des signaux du monde réel en données numériques compréhensible pas des ordinateurs, ont des limitations liées aux techniques d'acquisition, aux supports de stockage et aux types de données. De ce fait, les jeux de données comportent des biais. C'est pourquoi, il est d'usage de réaliser des étapes de prétraitement pour diminuer le nombre de biais présents dans un jeu de données afin de pousser les réseaux de neurones à se focaliser sur la tâche souhaitée.

3.4.1.2 Les limites de l'évaluation d'un réseau sur un jeu de données

Pour juger un réseau de neurones et s'assurer que l'entraînement s'est déroulé correctement, on utilise d'abord des métriques qui vont nous donner des indications sur la performance d'un réseau. Suivant le domaine d'application, les standards d'utilisation de métrique changent. En effet, les métriques ont des avantages et des inconvénients qu'il convient de considérer avant de les utiliser. Ainsi, chaque métrique à ses limitations et il est possible d'obtenir de très bons résultats sur une métrique à cause d'un jeu de données biaisé (KOROTCOV et al., 2017). De plus, en prenant en compte la perte qui est un score de pénalité calculé sur les mauvaises prédictions, on peut diagnostiquer des effets de sous-apprentissage ou de sur-apprentissage. Ces pratiques font office de standard dans l'entraînement de réseaux de neurones profonds, malgré les limitations dont elles souffrent. Elles permettent d'évaluer une modélisation et fournissent des informations qui peuvent servir à améliorer les prochaines modélisations. Lorsque l'on analyse en détail les prédictions d'un réseau et qu'on les compare avec les résultats d'un expert du domaine de la tâche, il arrive que des données jugées faciles à classer par un expert soient mal classées par le réseau. A l'inverse, il existe des cas sur lesquels les experts passent beaucoup de temps et hésitent, alors que le

réseau prédit rapidement la bonne classe. Dans ces exemples, il est intéressant de comprendre ce que fait précisément un réseau pendant une prédiction. Or, en utilisant les méthodes classiques d'évaluation, on considère le réseau comme une boîte noire. Dès lors, la donnée en entrée couplée à la prédiction en sortie sont les seuls éléments à notre disposition. Ainsi, on peut avoir une première intuition du comportement d'un réseau, mais il y a une limite à l'interprétation que l'on peut en faire. A ce stade, à cause du caractère opaque d'un réseau de neurones, et dans l'impossibilité d'évaluer les caractéristiques apprises, il est possible d'obtenir une modélisation imparfaite qui obtiendra des performances parfaites sur un jeu de données biaisé. C'est pourquoi, les méthodes d'évaluations traditionnelles des réseaux de neurones ne suffisent pas à juger les performances des intelligences artificielles.

3.4.1.3 Une absence de considération de l'éthique

En plus d'être techniquement complexes à comprendre, les décisions des intelligences artificielles souffrent d'un autre problème majeur qui préoccupe aussi bien les scientifiques que l'opinion publique : l'absence d'éthique dans les IA. Avant toute chose, on pourrait penser que l'automatisation d'une tâche sensible à la discrimination, tel que le tri de candidature à un poste, permet de réduire les discriminations qu'une personne effectue inconsciemment. Effectivement, les algorithmes ont le potentiel de prendre des décisions à notre place en restant constant, impartial et insensible aux émotions ou biais pouvant discriminer. Cependant, pour être à même de réaliser une tâche, les IA ont besoin d'apprendre sur des données. Ainsi, il est possible à partir de données biaisées, d'obtenir des décisions discriminantes pour certaines catégories de la population. À partir de ce constat, on peut penser qu'il suffit alors de travailler à partir de jeux de données qui ne contiennent pas de biais. Or, les données utilisées par les IA sont acquises à partir du monde réel qui n'échappe pas aux discriminations. La collecte de données biaisées, suivie de l'utilisation de celles-ci par une IA entraîne la création de modèles biaisés. Lorsque ces modèles sont utilisés sans vérifications supplémentaires en production, cela renforce les discriminations déjà existantes et peut créer des scandales lors de la découverte du problème. Parmi les scandales qui ont éclaté sur des IA qui discriminaient une population, on retrouve celui d'Amazon en 2018 (DASTIN, 2018), où un outil de tri automatique de candidature privilégiait des CV d'hommes au profit de CV de femmes. Le biais à l'origine de cette discrimination vient d'un déséquilibre existant dans les effectifs actuels d'Amazon, qui a servi à entraîner l'IA pour trier les candidatures. En dehors des biais dans les données, d'autres questions se posent sur la capacité des IA à faire des choix moraux. Lorsque l'on parle de morale dans un choix, on pense souvent au problème du tramway (JARVIS THOMSON, 1985) introduit en 1985, qui présente des situations où des décisions doivent être prises pour sauver une ou plusieurs personnes aux dépens d'autres personnes. Suivant la nature des événements qui précèdent ses situations, les actions entreprises pour arriver à un même résultat de personnes sauvées et sacrifiées, les actions peuvent être jugées comme moralement acceptables ou inacceptables. Compte tenu de la complexité que l'on rencontre pour déterminer si une décision est morale ou non, on peut alors se questionner sur la faisabilité d'intégrer cette notion dans une IA. Ces cas concrets montrent les failles éthiques des IA. Les questionnements et les craintes légitimes amenées par les intelligences artificielles ne sont pas récents. Dès les années 1950, on trouve dans les œuvres de science-fiction de l'auteur Isaac Asimov des portraits de sociétés où les robots peuvent faire des choix (ASIMOV, 2004). A travers la fiction, de nombreuses questions éthiques sont abordées et trois lois de la robotique sont établies pour garder le contrôle des humains sur les robots.

Ces récits précurseurs sont les fondations des questions éthiques que l'on se pose encore aujourd'hui et qui n'ont cessé de gagner en intérêt avec l'essor de l'intelligence artificielle de ces dernières années. De la même façon que Asimov a imaginé des règles pour que les robots servent l'humanité de manière éthique, à présent, le défi est de définir un cadre qui garantisse de l'équité, de l'explicabilité, de la robustesse, de la transparence tout en protégeant des données personnelles, dans les décisions des intelligences artificielles. Ainsi, le nombre d'études sur l'IA éthique a augmenté ces dernières années (SIAU et WANG, 2020 ; POWERS et GANASCIA, 2020 ; MORLEY et al., 2020 ; DUBBER, PASQUALE et DAS, 2020 ; HAGENDORFF, 2020) pour répondre à ces nouveaux besoins comme montré en figure 3.12. Pour tendre vers davantage d'éthique, les méthodes d'explicabilité et d'interprétabilité émergent et améliorent notre compréhension des prédictions des IA.

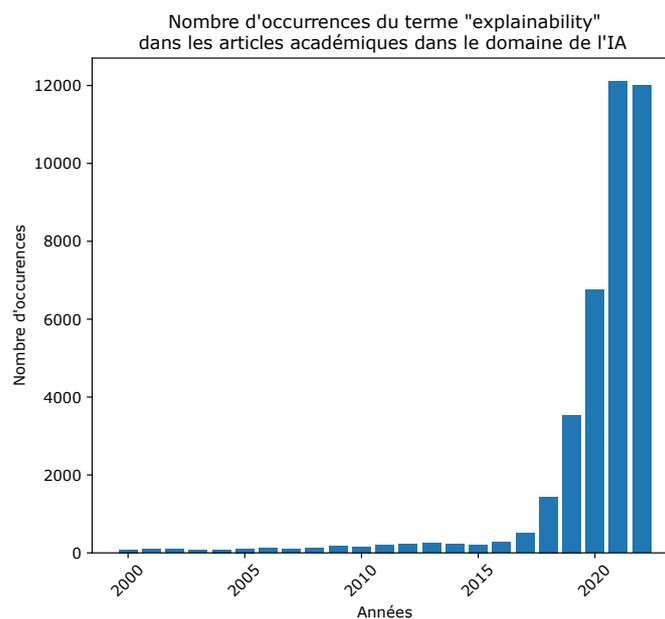


FIGURE 3.12 – Tendence de l'explicabilité. Occurrence historique des termes *explainability* dans les articles universitaires sur Google Scholar depuis les années 2000.

3.4.1.4 Vers des méthodes qui apportent de l'explicabilité et de l'interprétabilité

Sans compter les problèmes éthiques, il existe d'autres raisons qui augmentent le besoin d'explicabilité des prédictions des IA. Dans l'article (BURKART et HUBER, 2021), les chercheurs mettent en avant huit raisons qui incitent à vouloir davantage d'explication sur les choix des IA. Parmi elles, on trouve une volonté de comprendre les effets de causalité entre une donnée d'entrée et la prédiction en sortie de réseau pour notamment estimer la confiance, la transférabilité, le caractère informatif, l'équité, l'éthique, la responsabilité. Pour aller plus loin, l'explication des prédictions peut également servir pour vérifier la présence de biais dans les données qui peut entraîner une mauvaise discrimination d'une partie des données. Il devient également possible d'utiliser l'explicabilité pour ajuster notre compréhension d'une tâche. Pour toutes ces raisons, plusieurs initiatives ont vu le jour ces dernières années telles que : le programme IA explicable de la DARPA en 2017 (GUNNING, 2017) ; la liste des bonnes

pratiques pour éviter les discriminations dans les données publiées par l'ACM, également en 2017 (COUNCIL, 2017) ; la loi RGPD (*LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles* 2018) pour protéger les données personnelles en 2018. Pour aller plus loin dans le diagnostic d'un modèle généré par un réseau, on trouve des méthodes que l'on peut réunir en deux grandes catégories : Les méthodes d'explicabilité et les méthodes d'interprétabilité. En premier lieu, les méthodes d'explicabilité s'intéressent aux opérations mathématiques réalisées dans le cœur du réseau de neurones, c'est-à-dire, dans les couches et dans les neurones. Ainsi, ces méthodes qui apportent de la transparence, nous donnent des clés pour déconstruire l'effet de boîte noire dans les réseaux, afin de les visualiser comme des boîtes blanches.

En second lieu, Les méthodes d'interprétabilité regroupent les méthodes capables d'associer une cause à un effet dans un modèle d'apprentissage automatique. De ce fait, on peut visualiser les zones d'intérêts directement sur les données.

3.4.2 La mutation KRAS G12C vue par CAM et LIME

Nous avons appliqué les méthodes d'explicabilité CAM et LIME pour essayer de comprendre quelles caractéristiques visuelles ont aidé le réseau à prendre sa décision (voir figure 3.13). Nous observons que CAM et LIME mettent en évidence les zones qui sont associées aux tumeurs. Cependant, après analyse de ces résultats par un pathologiste, nous n'avons pas pu déterminer visuellement les structures qui discriminent la mutation KRAS ou le variant G12C. Nous pensons que la difficulté d'interprétation des résultats donnés par les méthodes d'explicabilité provient des biais de nos réseaux. Il semble que la prolifération des cellules produise des motifs qui permettent aux IA de prédire les mutations. Toutefois, ces motifs sont trop complexes ou trop petits pour être identifiés visuellement par un pathologiste muni d'un microscope.

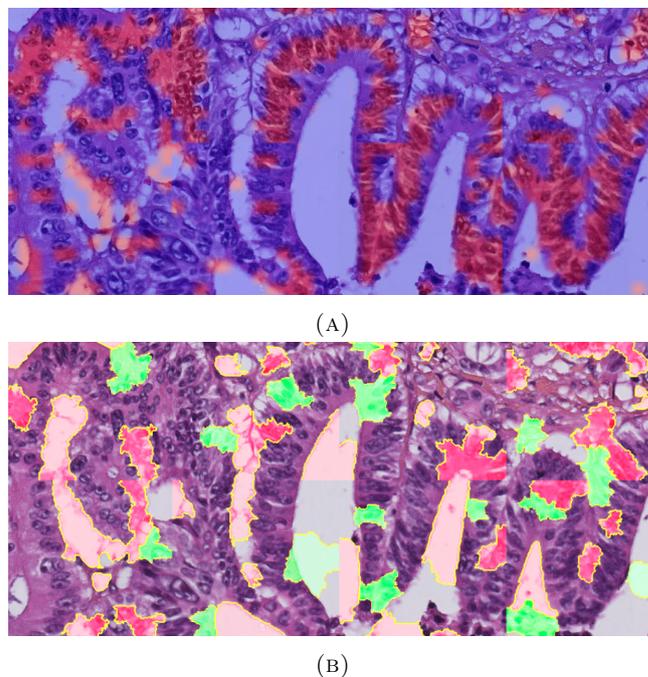


FIGURE 3.13 – Exemple de résultats de CAM (a) et de LIME (b) sur les prédictions du réseau. Les zones surlignées montrent où le réseau se concentre pour classer KRAS G12C.

Parmi les biais présents dans nos modèles, nous avons constaté que les traces de mucines 3.14 sur les tissus sont souvent interprétées comme un variant de la classe Autres par les réseaux.

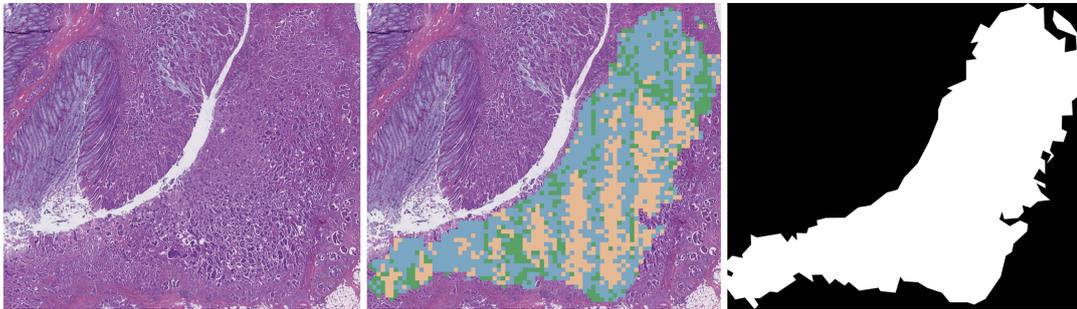


FIGURE 3.14 – Prédiction de dépôts de mucine comme étant des variants de la mutation KRAS. Les réseaux confondent des dépôts de mucine avec des structures biologiques appartenant à des variants (G12D, G12V, G13D etc). À gauche, le tissu histopathologique. À droite, la zone tumorale annotée. Au milieu, les prédictions du réseau, avec en bleu, les prédictions de type-sauvage, en vert, les prédictions du variant G12C et en beige, les prédictions des autres variants de la mutation KRAS.

3.5 Conclusion

Dans le cadre du projet AiCOLO, nous nous concentrons sur la détection de la mutation KRAS, puis du variant G12C à l'aide de réseaux de neurones profonds. Nous justifions ce choix de variant pour son intérêt clinique compte tenu de la forme particulièrement agressive de cancer qui peut en découler, mais également parce que notre jeu de données contient suffisamment de diversité pour l'apprentissage. Néanmoins, nos expériences subsidiaires suggèrent que l'utilisation d'un plus grand nombre de gènes et de variants lors de l'entraînement d'un réseau peut améliorer les performances du réseau.

Dans cette partie nous avons investigué l'utilisation de méthodes d'explicabilité comme CAM et LIME, pour essayer de comprendre visuellement les structures biologiques associées aux mutations génétiques. Malgré nos efforts, les caractéristiques qui permettent aux réseaux de détecter la mutation KRAS G12C restent incomprises.

Bien qu'il soit possible que tous les patches ne présentent pas une signature du gène que nous recherchons, une majorité de patches avec un niveau de grossissement de X20 est susceptible de contenir des traces d'une mutation car ils permettent tout de même aux réseaux d'apprendre les caractéristiques de la mutation. Ce type d'approche pourrait être utilisé dans les routines de laboratoire pour obtenir rapidement et à faible coût des informations génétiques sur une lame de verre, comme c'est le cas pour d'autres méthodes d'analyse d'images médicales décrites par (PANDA, JANGID et JAIN, 2021).

Conclusion Générale

Conclusion

L'analyse d'images histopathologiques par des méthodes d'intelligence artificielle et plus spécifiquement d'apprentissage profond demeure une tâche complexe. La première raison est la difficulté d'acquisition des données médicales qui constituent les jeux de données utilisés par les réseaux. En effet, la création de telles données requiert des machines spécialisées, du personnel qualifié, des patients volontaires atteints de la pathologie étudiée. Deuxièmement, une des difficultés réside dans la présence de biais dans les jeux de données, ces biais peuvent tromper le réseau dans son apprentissage et ainsi obtenir une mauvaise modélisation du problème souhaité. Par exemple, les manipulations effectuées ou les machines utilisées peuvent faire apparaître des biais dans les étapes de création des images histopathologiques telles que le prélèvement de tissus, la fixation, la découpe, la colorisation, la numérisation. Troisièmement, l'histopathologie est une science complexe, les experts peuvent avoir des avis divergents sur un même sujet. Là encore, ces divergences peuvent biaiser un réseau lors de son entraînement et ainsi créent une modélisation qui ne reflète pas la réalité du problème. Quatrièmement, dans le domaine de la pathologie numérique, les jeux de données et les codes sources sont souvent gardés privés. L'une des raisons est la volonté de créer des brevets. Cela empêche la reproductibilité des travaux et de surcroît, limite la comparaison entre les nouvelles méthodes. Ces difficultés sont un frein à l'avancée du domaine. Toutefois, des initiatives existent pour donner libre accès à de grandes quantités d'image médicale telle que TCGA (*The cancer genome atlas program s. d.*). De plus, des lois relatives aux données personnelles telles que la RGPD sont mises en place pour réguler l'utilisation des données privées. Donc, dans les années qui viennent, on peut s'attendre à voir le nombre de données médicales publiques de tous types d'organes et de pathologies augmenter, tout en étant mieux encadrées. Plus il y aura de données publiques de différentes pathologies, et plus il sera facile de proposer des méthodes automatiques pour des applications médicales afin d'assister les médecins dans leur routine.

Ces outils d'assistance basée sur les IA ont le potentiel de faire baisser les erreurs de diagnostic et donc d'augmenter les chances de survie des patients. Dans le chapitre 2 et dans le cadre du projet Sys-MIFTA, nous avons présenté une méthode de détection de glomérules dans le rein en temps réel pour assister des pathologistes dans leur séance de diagnostic. Pour que les méthodes de prédiction puissent suivre les déplacements et les zooms d'un pathologiste dans sa routine, il faut que les approches soient robustes aux changements de niveaux dans les tissus. Pour cela, nous avons également exploré l'utilisation de différents niveaux de détails des images histopathologiques.

L'utilisation de méthodes d'apprentissage profond sur des images histopathologiques peut également servir à simplifier l'acquisition de certaines données patientes. En effet, il est possible de prédire des informations génétiques telles que les mutations et les variants, directement depuis des images histopathologiques, sans passer par l'utilisation de séquenceur ADN. Ces méthodes ont le potentiel d'aider les médecins qui

n'ont pas accès à des séquenceurs ADN, à diriger les patients vers des traitements mieux adaptés, et augmenter ainsi leurs chances de survie. Ainsi, nous avons présenté une méthodologie pour la détection de la mutation KRAS G12C avec des approches d'apprentissage profond dans le cadre du projet AiCOLO, dans le chapitre 3.

Contributions

Dans le chapitre 2.2, nous avons proposé d'étudier des images de cancer du sein à différents niveaux de grossissement à l'aide du réseau de neurones profond InceptionResNet-v2. Pour cela, nous utilisons le jeu de données BreakHis, qui contient des images histopathologiques de sein avec et sans tumeurs. Ainsi, nous déterminons les niveaux dans lesquels les caractéristiques des tumeurs sont les plus intéressantes. Ensuite, nous reprenons ces résultats pour comparer différentes stratégies à des fins de classification de tumeurs dans le sein, robuste aux changements de niveaux de grossissement et en temps réel.

Dans un second temps, dans le chapitre 2.3, nous avons présenté nos travaux dans le cadre du projet européen Sys-MIFTA. Nous proposons une preuve de concept en utilisant YOLOv3, une architecture basée sur l'apprentissage profond, pour détecter à différents niveaux de grossissement en temps réel des glomérules dans le rein dans le contexte du suivi de greffe. Nous explorons différentes stratégies d'entraînement sur différents niveaux de grossissement pour rendre nos réseaux invariants aux changements de niveau.

Enfin, dans le chapitre 3, dans le cadre du projet AiCOLO, nous avons proposé de détecter la mutation KRAS G12C à l'aide de réseaux de neurones profonds directement depuis des images histopathologiques issues du jeu de données AiCOLO contenant des WSI de biopsie du côlon ainsi que des informations génétiques. Nous avons comparé différentes architectures d'apprentissage profond sur la tâche de classification des mutations et des variants dans le cancer colorectal. Les modélisations ainsi créées ont été utilisées par les méthodes d'explicabilité CAM et LIME pour visualiser les zones des patches qui motivent la prise de décision d'un réseau.

Afin de permettre la reproductibilité de nos expérimentations, nous avons rendu publics tous les codes sources qui ont servi à réaliser nos travaux (HECKENAUER, 2020; HECKENAUER, 2022; HECKENAUER, 2023)⁶.

Perspectives

Exploitation des différents niveaux de grossissement des images de lames

Les travaux présentés dans le chapitre 2 ont le potentiel d'être utilisés dans le développement d'outils pour assister les pathologistes dans leur routine basés sur la réalité étendue (XR) (i.e. Réalité Augmentée (AR), Réalité Virtuelle (VR)), tels que (VENKATESAN et al., 2021; CHEN et al., 2018; CHEN et al., 2019; RAZAVIAN, 2019). Par exemple, dans un contexte où l'on souhaite afficher des informations supplémentaires sur un microscope à réalité augmentée (ARM), il est nécessaire d'utiliser

6. Real-time detection for renal pathology : <https://github.com/RobinHCK/Real-time-detection-for-renal-pathology>

Breast cancer classification methods for augmented reality microscopes : <https://github.com/RobinHCK/Breast-cancer-classification-methods-for-augmented-reality-microscopes>

Comparison of deep learning architectures for colon cancer mutation detection : <https://github.com/RobinHCK/Comparison-of-deep-learning-architectures-for-colon-cancer-mutation-detection>

des approches d'analyse d'images en temps réel. Ensuite, il est intéressant de connaître les niveaux dans lesquels les structures biologiques ciblées sont les plus pertinentes. Les outils augmentés ont le potentiel d'aider différents champs de la santé à des fins de formation ou de diagnostics sur des tâches telles que les analyses microscopiques ou encore les opérations chirurgicales. Néanmoins, les avantages et les inconvénients que peuvent apporter ces méthodes restent méconnus. Effectivement, les études qui proposent des outils augmentés pour l'éducation (CHEN et al., 2022; ZHOU et al., 2020), manquent souvent de retours utilisateurs.

Les travaux futurs pourraient dans un premier temps, continuer d'exploiter les niveaux de grossissement des images histopathologiques avec de l'apprentissage ensembliste pour fusionner les connaissances des réseaux entraînés à différents niveaux. D'une manière différente, il serait souhaitable d'avoir des données de retours utilisateurs avec différents degrés de connaissance (e.g. néophyte, étudiant en médecine, pathologiste etc), utilisant des méthodes XR tel que l'ARM. L'identification des bénéfices et des risques rencontrés par les pathologistes pourrait nous aider à mieux identifier les challenges du domaine, et ainsi mieux cibler les prochains travaux sur les niveaux de grossissement.

Détection de mutations génétiques dans le cancer du côlon

La détection de mutation directement sur les images de lames présenté en chapitre 3, permet d'envisager des applications qui pourraient à la fois permettre de diagnostiquer à moindre coût les patients des pays les plus pauvres, mais également permettre de réserver l'utilisation des séquenceurs ADN à des tâches plus critiques dans les pays les plus riches. L'utilisation de telles applications permet également d'économiser le temps des procédures en laboratoire nécessaires à l'obtention d'informations génétiques.

Les futurs travaux pourraient se concentrer sur l'utilisation de méthodes non supervisées pour éviter le problème des données mal labellisées qui introduisent des biais dans le jeu de données. Des entraînements plus stables pourraient ainsi permettre d'appliquer plus facilement des méthodes d'explicabilité pour comprendre quelles structures biologiques sont associées à quelles mutations.

Liste des communications

Article de journal

- Soumission à *Computer Methods and Programs in Biomedicine Update* : "Deep learning for KRAS G12C mutation detection in colon cancer.", R. Heckenauer, J. Weber, C. Wemmert, C. Truntzer, V. Derangère, F. Ghiringhelli, M. Hassenforder, P-A. Muller, G. Forestier.

Communications internationales

- Heckenauer, Robin, Jonathan Weber, Cedric Wemmert, Friedrich Feuerhake, Michel Hassenforder, Pierre-Alain Muller, and Germain Forestier. "Real-time detection of glomeruli in renal pathology." In 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pp. 350-355. IEEE, 2020. (HECKENAUER et al., 2020)
- Heckenauer, Robin, Jonathan Weber, Cédric Wemmert, Michel Hassenforder, Pierre-Alain Muller, and Germain Forestier. "Breast cancer classification methods for augmented reality microscopes." Applications in Medicine and Manufacturing : 124. (HECKENAUER et al., s. d.)
- Heckenauer, Robin, Jonathan Weber, Cédric Wemmert, Caroline Truntzer, Valentin Derangere, François Ghiringhelli, Michel Hassenforder, Pierre-Alain Muller, and Germain Forestier. "Comparison of deep learning architectures for colon cancer mutation detection." In 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), pp. 360-365. IEEE, 2023. (HECKENAUER et al., 2023b)

Communications nationales

- Heckenauer, Robin, Jonathan Weber, Cédric Wemmert, Friedrich Feuerhake, Michel Hassenforder, Pierre-Alain Muller, and Germain Forestier. "Détection en temps réel des glomérules en pathologie rénale." In ORASIS 2021. (HECKENAUER et al., 2021)
- Heckenauer, R., J. Weber, C. Wemmert, M. Hassenforder, P. A. Muller, and G. Forestier. "Méthodes de classification du cancer du sein pour les microscopes à réalité augmentée." In RFIAP 2022 (Congrès Reconnaissance des Formes, Image, Apprentissage et Perception). 2022. (HECKENAUER et al., 2022)
- Heckenauer, R., Weber, J., Wemmert, C., Truntzer, C., Derangère, V., Ghiringhelli, F., Hassenforder, M., Muller, P.A. and Forestier, G., 2023, May. Comparaison des architectures d'apprentissage profond pour la détection des mutations du cancer du côlon. In ORASIS 2023 : journées francophones des jeunes chercheurs en vision par ordinateur, (Vol. 2023). (HECKENAUER et al., 2023a)

Divers

- Prix de la meilleure présentation de thèse lors de la journée doctorale sciences exactes du 15 juin 2022.
- Présentation d'un poster "Détection de la mutation KRAS G12C dans le cancer du côlon." à la Journée d'étude autour de l'IA (JSIA 2023) à Nancy.

Financements

Les travaux sur la détection de glomérules dans le rein en temps réel présentés en section 2.3 ont été réalisés dans le cadre du projet SysMIFTA. Le projet ERACoSys-Med "SysMIFTA", cofinancé par EU H2020 et les agences nationales de financement du ministère allemand de l'Éducation et de la Recherche (BMBF) gestion de projet PTJ (FKZ : 031L-0085A), et Agence nationale de la recherche (ANR), numéro de projet ANR-15-CMED-0004.

Les recherches sur la détection de la mutation KRAS G12C dans le cancer du côlon présenté en section 3, font parti du projet AiCOLO. Le projet AiCOLO est financé par l'INSERM/Plan Cancer.

Le Centre de Calcul Haute Performance de l'Université de Strasbourg a soutenu ce travail en fournissant un soutien scientifique et un accès aux ressources informatiques. Une partie des ressources informatiques ont été financées par le projet Equipex Equip@Meso (Programme Investissements d'Avenir) et le CPER Alsacalcul / Big Data.

Bibliographie

- AL-MASNI, Mohammed A et al. (2018). « Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system ». In : *Computer methods and programs in biomedicine* 157, p. 85-94 (cf. p. 28).
- ALKASSAR, S et al. (2021). « Going deeper : magnification-invariant approach for breast cancer classification using histopathological images ». In : *IET Computer Vision* 15.2, p. 151-164 (cf. p. 18).
- ALOYSIUS, Neena et M GEETHA (2017). « A review on deep convolutional neural networks ». In : *2017 international conference on communication and signal processing (ICCSP)*. IEEE, p. 0588-0592 (cf. p. 14).
- ALTURKISTANI, Hani A, Faris M TASHKANDI et Zuhair M MOHAMMEDSALEH (2016). « Histological stains : a literature review and case study ». In : *Global journal of health science* 8.3, p. 72 (cf. p. 7).
- AMES, Bruce N, Lois Swirsky GOLD et Walter C WILLETT (1995). « The causes and prevention of cancer. » In : *Proceedings of the National Academy of Sciences* 92.12, p. 5258-5265 (cf. p. 34).
- ANAND, Preetha et al. (2008). « Cancer is a preventable disease that requires major lifestyle changes ». In : *Pharmaceutical research* 25, p. 2097-2116 (cf. p. 34).
- ASIMOV, Isaac (2004). *I, robot*. T. 1. Spectra (cf. p. 5, 53).
- BANKHEAD, Peter et al. (2017). « QuPath : Open source software for digital pathology image analysis ». In : *Scientific reports* 7.1, p. 1-7 (cf. p. 38, 39).
- BAY, Noel Si-Yang et Boon-Huat BAY (2010). « Greek anatomist herophilus : the father of anatomy ». In : *Anatomy & cell biology* 43.4, p. 280-283 (cf. p. 3).
- BIAN, Chang et al. (2021). « ImmunoAIzer : A deep learning-based computational framework to characterize cell distribution and gene mutation in tumor microenvironment ». In : *Cancers* 13.7, p. 1659 (cf. p. 35-37, 51).
- BLOICE, Marcus D, Peter M ROTH et Andreas HOLZINGER (avr. 2019). « Biomedical image augmentation using Augmentor ». In : *Bioinformatics*. btz259. ISSN : 1367-4803 (cf. p. 27).
- BODLA, Navaneeth et al. (2017). « Soft-NMS—improving object detection with one line of code ». In : *Proceedings of the IEEE international conference on computer vision*, p. 5561-5569 (cf. p. 26).
- BRENNAN, Connor TA (2022). « Bygone theatres of events : A history of human anatomy and dissection ». In : *The Anatomical Record* 305.4, p. 788-802 (cf. p. 3).
- BURKART, Nadia et Marco F HUBER (2021). « A survey on the explainability of supervised machine learning ». In : *Journal of Artificial Intelligence Research* 70, p. 245-317 (cf. p. 37, 54).
- CHEN, Jenny X et al. (2022). « Augmented Reality in Otolaryngology/Neurotology : A Scoping Review with Implications for Practice and Education ». In : *The Laryngoscope* (cf. p. 59).
- CHEN, Mingyu et al. (2020a). « Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning ». In : *NPJ precision oncology* 4.1, p. 1-7 (cf. p. 35-37).

- CHEN, Po-Hsuan Cameron et al. (2018). « Microscope 2.0 : An augmented reality microscope with real-time artificial intelligence integration ». In : *arXiv preprint arXiv :1812.00825* (cf. p. 58).
- CHEN, Po-Hsuan Cameron et al. (2019). « An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis ». In : *Nature medicine* 25.9, p. 1453-1457 (cf. p. 58).
- CHEN, Xiahan et al. (2020b). « Combined spiral transformation and model-driven multi-modal deep learning scheme for automatic prediction of tp53 mutation in pancreatic cancer ». In : *IEEE Transactions on Medical Imaging* 40.2, p. 735-747 (cf. p. 36).
- CHOI, Dami et al. (2019). « On empirical comparisons of optimizers for deep learning ». In : *arXiv preprint arXiv :1910.05446* (cf. p. 12).
- CHOLLET, François et al. (2015). *Keras* (cf. p. 26).
- COUDRAY, Nicolas et al. (2018). « Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning ». In : *Nature medicine* 24.10, p. 1559-1567 (cf. p. 35, 36).
- COUNCIL, US Public Policy (2017). « Statement on algorithmic transparency and accountability ». In : *Commun. ACM* (cf. p. 55).
- DASTIN, Jeffrey (2018). « Amazon scraps secret AI recruiting tool that showed bias against women ». In : *Ethics of Data and Analytics*. Auerbach Publications, p. 296-299 (cf. p. 53).
- DENG, Jia et al. (2009). « Imagenet : A large-scale hierarchical image database ». In : *2009 IEEE conference on computer vision and pattern recognition*. Ieee, p. 248-255 (cf. p. 20, 47).
- DENG, Shujian et al. (2020). « Deep learning in digital pathology image analysis : a survey ». In : *Frontiers of medicine* 14.4, p. 470-487 (cf. p. 4).
- DICK, Stephanie (2019). « Artificial intelligence ». In : (cf. p. 5).
- DONG, Nanqing et al. (2018). « Reinforced auto-zoom net : towards accurate and fast breast cancer segmentation in whole-slide images ». In : *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer, p. 317-325 (cf. p. 18).
- DOZAT, Timothy (2016). « Incorporating nesterov momentum into adam ». In : (cf. p. 48).
- DUBBER, Markus Dirk, Frank PASQUALE et Sunit DAS (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks (cf. p. 54).
- DUCHI, John, Elad HAZAN et Yoram SINGER (2011). « Adaptive subgradient methods for online learning and stochastic optimization. » In : *Journal of machine learning research* 12.7 (cf. p. 48).
- DUNCAN, James S et Nicholas AYACHE (2000). « Medical image analysis : Progress over two decades and the challenges ahead ». In : *IEEE transactions on pattern analysis and machine intelligence* 22.1, p. 85-106 (cf. p. 4).
- FARAHANI, Navid et Liron PANTANOWITZ (2015). « Overview of telepathology ». In : *Surgical pathology clinics* 8.2, p. 223-231 (cf. p. 4).
- FERLAY, Jacques et al. (2021). « Cancer statistics for the year 2020 : An overview ». In : *International journal of cancer* 149.4, p. 778-789 (cf. p. 36).
- GARCEA, Fabio et al. (2022). « Data augmentation for medical imaging : A systematic literature review ». In : *Computers in Biology and Medicine*, p. 106391 (cf. p. 9).

- GHOSH, Sanjib Kumar (2017). « Giovanni Battista Morgagni (1682–1771) : father of pathologic anatomy and pioneer of modern medicine ». In : *Anatomical science international* 92.3, p. 305-312 (cf. p. 3).
- GILBERT, Walter et Allan MAXAM (1973). « The nucleotide sequence of the lac operator ». In : *Proceedings of the National Academy of Sciences* 70.12, p. 3581-3584 (cf. p. 35).
- GIRSHICK, Ross (2015). « Fast R-CNN ». In : *Proceedings of the IEEE International Conference on Computer Vision*, p. 1440-1448 (cf. p. 26).
- GIRSHICK, Ross et al. (2014). « Rich feature hierarchies for accurate object detection and semantic segmentation ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 580-587 (cf. p. 26).
- GOODFELLOW, Ian et al. (2020). « Generative adversarial networks ». In : *Communications of the ACM* 63.11, p. 139-144 (cf. p. 10).
- GRAVES, Alex (2013). « Generating sequences with recurrent neural networks ». In : *arXiv preprint arXiv :1308.0850* (cf. p. 45, 48).
- GUNNING, David (2017). « Explainable artificial intelligence (xai) ». In : *Defense advanced research projects agency (DARPA), nd Web 2.2*, p. 1 (cf. p. 54).
- HAGENDORFF, Thilo (2020). « The ethics of AI ethics : An evaluation of guidelines ». In : *Minds and Machines* 30.1, p. 99-120 (cf. p. 54).
- HAMET, Pavel et Johanne TREMBLAY (2017). « Artificial intelligence in medicine ». In : *Metabolism* 69, S36-S40 (cf. p. 5).
- HAMIDA, A Ben et al. (2021). « Deep learning for colon cancer histopathological images analysis ». In : *Computers in Biology and Medicine* 136, p. 104730 (cf. p. 39).
- HAMILTON, Peter W, Yinhai WANG et Stephen J MCCULLOUGH (2012). « Virtual microscopy and digital pathology in training and education ». In : *Apmis* 120.4, p. 305-315 (cf. p. 17).
- HE, Kaiming et al. (2016). « Identity mappings in deep residual networks ». In : *European conference on computer vision*. Springer, p. 630-645 (cf. p. 45, 50).
- HE, Kan et al. (2020). « Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging ». In : *BMC medical imaging* 20.1, p. 1-9 (cf. p. 35, 36).
- HECKENAUER, R et al. (2022). « Méthodes de classification du cancer du sein pour les microscopes à réalité augmentée ». In : *RFIAP 2022 (Congrès Reconnaissance des Formes, Image, Apprentissage et Perception)* (cf. p. 61).
- HECKENAUER, R et al. (2023a). « Comparaison des architectures d'apprentissage profond pour la détection des mutations du cancer du côlon ». In : *ORASIS 2023 : journées francophones des jeunes chercheurs en vision par ordinateur*, t. 2023 (cf. p. 61).
- HECKENAUER, Robin (2020). *Real-time detection for renal pathology*. URL : <https://github.com/RobinHCK/Real-time-detection-for-renal-pathology> (cf. p. 58).
- (2022). *Breast cancer classification methods for augmented reality microscopes*. URL : <https://github.com/RobinHCK/Breast-cancer-classification-methods-for-augmented-reality-microscopes> (cf. p. 58).
- (2023). *Comparison of deep learning architectures for colon cancer mutation detection*. URL : <https://github.com/RobinHCK/Comparison-of-deep-learning-architectures-for-colon-cancer-mutation-detection> (cf. p. 58).
- HECKENAUER, Robin et al. (s. d.). « Breast cancer classification methods for augmented reality microscopes ». In : *Applications in Medicine and Manufacturing* (), p. 124 (cf. p. 61).

- HECKENAUER, Robin et al. (2020). « Real-time detection of glomeruli in renal pathology ». In : *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, p. 350-355 (cf. p. 61).
- HECKENAUER, Robin et al. (2021). « Détection en temps réel des glomérules en pathologie rénale ». In : *ORASIS 2021* (cf. p. 61).
- HECKENAUER, Robin et al. (2023b). « Comparison of deep learning architectures for colon cancer mutation detection ». In : *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, p. 360-365 (cf. p. 61).
- HOBBS, Thomas et Marshall MISSNER (2016). *Thomas Hobbes : Leviathan (Longman library of primary sources in philosophy)*. Routledge (cf. p. 5).
- HONG, Runyu et al. (2021). « Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models ». In : *Cell Reports Medicine* 2.9, p. 100400 (cf. p. 36).
- HOOD, Leroy E, Michael W HUNKAPILLER et Lloyd M SMITH (1987). « Automated DNA sequencing and analysis of the human genome ». In : *Genomics* 1.3, p. 201-212 (cf. p. 35).
- HUANG, Gao et al. (2017). « Densely connected convolutional networks ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 4700-4708 (cf. p. 45, 50).
- HURDEMAN, Anton A (2003). *The worldwide history of telecommunications*. John Wiley & Sons (cf. p. 4).
- JACQUES JOUANNA, By et Philip Van der EIJK (2012). *Greek medicine from Hippocrates to Galen : selected papers*. Brill (cf. p. 3).
- JARVIS THOMSON, Judith (1985). « The trolley problem ». In : (cf. p. 53).
- JOHNSON, Justin M et Taghi M KHOSHGOFTAAR (2019). « Survey on deep learning with class imbalance ». In : *Journal of Big Data* 6.1, p. 1-54 (cf. p. 43).
- KABIRAJ, Arpita et al. (2015). « Principle and techniques of immunohistochemistry—a review ». In : *Int J Biol Med Res* 6.3, p. 5204-5210 (cf. p. 4).
- KANWAL, Neel et al. (2022). « The devil is in the details : Whole Slide Image acquisition and processing for artifacts detection, color variation, and data augmentation. A review. » In : *IEEE Access* (cf. p. 42).
- KARPOZILOS, Apostolos et Nicholas PAVLIDIS (2004). « The treatment of cancer in Greek antiquity ». In : *European journal of cancer* 40.14, p. 2033-2040 (cf. p. 3).
- KAWAZOE, Yoshimasa et al. (2018). « Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images ». In : *Journal of Imaging* 4.7, p. 91 (cf. p. 26).
- KEYS, Robert (1981). « Cubic convolution interpolation for digital image processing ». In : *IEEE transactions on acoustics, speech, and signal processing* 29.6, p. 1153-1160 (cf. p. 19).
- KIEFER, Jack et Jacob WOLFOWITZ (1952). « Stochastic estimation of the maximum of a regression function ». In : *The Annals of Mathematical Statistics*, p. 462-466 (cf. p. 48).
- KINGMA, Diederik P et Jimmy BA (2014). « Adam : A method for stochastic optimization ». In : *arXiv preprint arXiv :1412.6980* (cf. p. 48).
- KOROTCOV, Alexandru et al. (2017). « Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets ». In : *Molecular pharmaceutics* 14.12, p. 4462-4475 (cf. p. 52).
- LAMPERT, Thomas et al. (2019). « Strategies for training stain invariant CNNs ». In : *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, p. 905-909 (cf. p. 27).
- Le larousse médical* (2022). Larousse (cf. p. 3).

- LECUN, Yann, Yoshua BENGIO et Geoffrey HINTON (2015). « Deep learning ». In : *nature* 521.7553, p. 436-444 (cf. p. 14, 52).
- LECUN, Yann et al. (1998). « Gradient-based learning applied to document recognition ». In : *Proceedings of the IEEE* 86.11, p. 2278-2324 (cf. p. 6).
- LEE, Kyubum et al. (2018). « Deep learning of mutation-gene-drug relations from the literature ». In : *BMC bioinformatics* 19.1, p. 1-13 (cf. p. 36).
- LEUR, Rutger R van de et al. (2021). « Discovering and visualizing disease-specific electrocardiogram features using deep learning : proof-of-concept in phospholamban gene mutation carriers ». In : *Circulation : Arrhythmia and Electrophysiology* 14.2, e009056 (cf. p. 36, 37).
- LIAO, Haotian et al. (2020). « Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma ». In : *Clinical and translational medicine* 10.2 (cf. p. 36).
- LIN, Tsung-Yi et al. (2014). « Microsoft coco : Common objects in context ». In : *European conference on computer vision*. Springer, p. 740-755 (cf. p. 26).
- LIN, Tsung-Yi et al. (2017). « Focal loss for dense object detection ». In : *Proceedings of the IEEE International Conference on Computer Vision*, p. 2980-2988 (cf. p. 26).
- LIU, Meili et al. (2022). « A Lung-Parenchyma-Contrast Hybrid Network For EGFR Gene Mutation Prediction In Lung Cancer ». In : *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE, p. 1-5 (cf. p. 35).
- LIU, Wei et al. (2016). « Ssd : Single shot multibox detector ». In : *European conference on computer vision*. Springer, p. 21-37 (cf. p. 26).
- LOI n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles (2018). Légifrance. URL : <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037085952> (cf. p. 55).
- MAATEN, Laurens Van der et Geoffrey HINTON (2008). « Visualizing data using t-SNE. » In : *Journal of machine learning research* 9.11 (cf. p. 37).
- MACENKO, Marc et al. (2009). « A method for normalizing histology slides for quantitative analysis ». In : *2009 IEEE international symposium on biomedical imaging : from nano to macro*. IEEE, p. 1107-1110 (cf. p. 8, 42).
- MARÉE, Raphaël et al. (2016). « Collaborative analysis of multi-gigapixel imaging data using Cytomine ». In : *Bioinformatics* 32.9, p. 1395-1401 (cf. p. 25, 38).
- MCMAHAN, H Brendan et al. (2013). « Ad click prediction : a view from the trenches ». In : *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 1222-1230 (cf. p. 48).
- MIMS, Christopher (2021). « Why Artificial Intelligence Isn't Intelligent ». In : *The Wall Street Journal* (cf. p. 5).
- MORLEY, Jessica et al. (2020). « The ethics of AI in health care : a mapping review ». In : *Social Science & Medicine* 260, p. 113172 (cf. p. 54).
- NAGASAKA, Misako et al. (2020). « KRAS G12C Game of Thrones, which direct KRAS inhibitor will claim the iron throne? » In : *Cancer treatment reviews* 84, p. 101974 (cf. p. 35).
- NAKAMURA, Kensuke et al. (2021). « Learning-Rate Annealing Methods for Deep Neural Networks ». In : *Electronics* 10.16, p. 2029 (cf. p. 13).
- NEWELL, Allen et Herbert SIMON (1956). « The logic theory machine—A complex information processing system ». In : *IRE Transactions on information theory* 2.3, p. 61-79 (cf. p. 5).
- NIAZI, Muhammad Khalid Khan, Anil V PARWANI et Metin N GURCAN (2019). « Digital pathology and artificial intelligence ». In : *The lancet oncology* 20.5, e253-e261 (cf. p. 4).

- NILSSON, Nils J (2009). *The quest for artificial intelligence*. Cambridge University Press (cf. p. 5).
- OQUAB, Maxime et al. (2015). « Is object localization for free?-weakly-supervised learning with convolutional neural networks ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 685-694 (cf. p. 2, 37).
- PANDA, Shubhajit, Mahesh JANGID et Ashish JAIN (2021). « A comprehensive review on the significance and impact of deep learning in medical image analysis ». In : *2021 International Conference on Technological Advancements and Innovations (ICTAI)*. IEEE, p. 358-366 (cf. p. 56).
- PANTANOWITZ, Liron et al. (2018). « Twenty years of digital pathology : an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives ». In : *Journal of pathology informatics* 9.1, p. 40 (cf. p. 4).
- PARSANIA, Pankaj S et Pares V VIRPARIA (2016). « A comparative analysis of image interpolation algorithms ». In : *International Journal of Advanced Research in Computer and Communication Engineering* 5.1, p. 29-34 (cf. p. 19).
- PENG, Qiwei et al. (2016). « Pedestrian detection for transformer substation based on Gaussian mixture model and YOLO ». In : *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. T. 2. IEEE, p. 562-565 (cf. p. 28).
- POPE, Alexandar et Steven SHANKMAN (2009). *The Iliad of Homer : Translated by Alexander Pope*. T. 1. Wipf et Stock Publishers (cf. p. 5).
- POWERS, Thomas M et Jean-Gabriel GANASCIA (2020). *The ethics of the ethics of AI* (cf. p. 54).
- QQWWEEE (2019). *keras-yolo3 : A Keras implementation of YOLOv3 (Tensorflow backend)* (cf. p. 26).
- QU, Hui et al. (2021). « Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning ». In : *NPJ precision oncology* 5.1, p. 1-11 (cf. p. 36).
- RACE, George J, G Weldon TILLERY et Peter A DYSERT (2004). « A history of pathology and laboratory medicine at Baylor University Medical Center ». In : *Baylor University Medical Center Proceedings*. T. 17. 1. Taylor & Francis, p. 42-55 (cf. p. 3, 4).
- RANDELL, Rebecca et al. (2013). « Virtual reality microscope versus conventional microscope regarding time to diagnosis : an experimental study ». In : *Histopathology* 62.2, p. 351-358 (cf. p. 22).
- RAZAVIAN, Narges (2019). « Augmented reality microscopes for cancer histopathology ». In : *Nature Medicine* 25.9, p. 1334-1336 (cf. p. 58).
- REDMON, Joseph et Ali FARHADI (2018). « Yolov3 : An incremental improvement ». In : *arXiv preprint arXiv :1804.02767* (cf. p. 17, 26).
- REDMON, Joseph et al. (2016). « You only look once : Unified, real-time object detection ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 779-788 (cf. p. 26).
- RIBEIRO, Marco Tulio, Sameer SINGH et Carlos GUESTRIN (2016). « " Why should i trust you?" Explaining the predictions of any classifier ». In : *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, p. 1135-1144 (cf. p. 2, 37, 46).
- ROSENTHAL, Jacob et al. (2022). « Building tools for machine learning and artificial intelligence in cancer research : best practices and a case study with the PathML toolkit for computational pathology ». In : *Molecular Cancer Research* 20.2, p. 202-206 (cf. p. 42).

- ROY, Santanu et al. (2018). « A study about color normalization methods for histopathology images ». In : *Micron* 114, p. 42-61 (cf. p. 8).
- RUSKA, Ernst (1987). « The development of the electron microscope and of electron microscopy ». In : *Reviews of modern physics* 59.3, p. 627 (cf. p. 4).
- SANDLER, Mark et al. (2018). « Mobilenetv2 : Inverted residuals and linear bottlenecks ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 4510-4520 (cf. p. 45, 50).
- SANGER, Fred et Alan R COULSON (1975). « A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase ». In : *Journal of molecular biology* 94.3, p. 441-448 (cf. p. 35).
- SCHAUMBERG, Andrew J, Mark A RUBIN et Thomas J FUCHS (2018). « H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer ». In : *BioRxiv*, p. 064279 (cf. p. 36).
- SCHULTZ, Myron (2008). « Rudolf virchow ». In : *Emerging infectious diseases* 14.9, p. 1480 (cf. p. 4).
- SHALF, John (2020). « The future of computing beyond Moore's Law ». In : *Philosophical Transactions of the Royal Society A* 378.2166, p. 20190061 (cf. p. 6).
- SIAU, Keng et Weiyu WANG (2020). « Artificial intelligence (AI) ethics : ethics of AI and ethical AI ». In : *Journal of Database Management (JDM)* 31.2, p. 74-87 (cf. p. 54).
- SIEGEL, Rebecca L. et al. (2022). « Cancer statistics, 2022 ». In : *CA : A Cancer Journal for Clinicians* 72.1, p. 7-33. DOI : <https://doi.org/10.3322/caac.21708>. eprint : <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21708>. URL : <https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21708> (cf. p. 35).
- SONG, Youyi et al. (2016). « Accurate cervical cell segmentation from overlapping clumps in pap smear images ». In : *IEEE transactions on medical imaging* 36.1, p. 288-300 (cf. p. 18, 27, 31).
- SPANHOL, Fabio A et al. (2015). « A dataset for breast cancer histopathological image classification ». In : *Ieee transactions on biomedical engineering* 63.7, p. 1455-1462 (cf. p. 7, 18, 47).
- STONE, Mervyn (1978). « Cross-validation : A review ». In : *Statistics : A Journal of Theoretical and Applied Statistics* 9.1, p. 127-139 (cf. p. 10).
- STROBEL, Volker (2018). *Historic word occurrence in academic papers*. <http://doi.org/10.5281/zenodo.1218409> (cf. p. 5).
- SZEGEDY, Christian et al. (2016). « Rethinking the inception architecture for computer vision ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 2818-2826 (cf. p. 45, 50).
- SZEGEDY, Christian et al. (2017). « Inception-v4, inception-resnet and the impact of residual connections on learning ». In : *Thirty-first AAAI conference on artificial intelligence* (cf. p. 18, 45, 50).
- TABIBU, Sairam, PK VINOD et CV JAWAHAR (2019). « Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning ». In : *Scientific reports* 9.1, p. 1-9 (cf. p. 36).
- TAJBAKSHI, Nima et al. (2016). « Convolutional neural networks for medical image analysis : Full training or fine tuning ? ». In : *IEEE transactions on medical imaging* 35.5, p. 1299-1312 (cf. p. 27).
- TAN, Mingxing et Quoc LE (2021). « Efficientnetv2 : Smaller models and faster training ». In : *International Conference on Machine Learning*. PMLR, p. 10096-10106 (cf. p. 45, 50).

- The cancer genome atlas program* (s. d.). URL : <https://www.cancer.gov/tcga> (cf. p. 8, 57).
- TRIPATHI, Satvik et al. (2022). « Radgennets : Deep learning-based radiogenomics model for gene mutation prediction in lung cancer ». In : *bioRxiv* (cf. p. 36).
- TUFAIL, Ahsan Bin et al. (2021). « Deep learning in cancer diagnosis and prognosis prediction : a minireview on challenges, recent trends, and future directions ». In : *Computational and Mathematical Methods in Medicine 2021* (cf. p. 22).
- VAHADANE, Abhishek et al. (2016). « Structure-preserving color normalization and sparse stain separation for histological images ». In : *IEEE transactions on medical imaging* 35.8, p. 1962-1971 (cf. p. 8).
- VENKATESAN, Mythreye et al. (2021). « Virtual and augmented reality for biomedical applications ». In : *Cell reports medicine* 2.7, p. 100348 (cf. p. 58).
- VINÇON, Nathan et al. (2022). « A deep learning solution to predict tumor mutational status : an evaluation as a first-line diagnostic tool ». In : (cf. p. 36).
- VON STADEN, Heinrich (1992). « The discovery of the body : human dissection and its cultural contexts in ancient Greece. » In : *The Yale journal of biology and medicine* 65.3, p. 223 (cf. p. 3, 4).
- WANG, Shuo et al. (2019). « Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning ». In : *European Respiratory Journal* 53.3 (cf. p. 35-37).
- WANG, Xiaoxiao et al. (2021). « Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images ». In : *Frontiers in Genetics*, p. 1147 (cf. p. 36).
- WANG, Zhe et al. (2017). « Zoom-in-net : Deep mining lesions for diabetic retinopathy detection ». In : *Medical Image Computing and Computer Assisted Intervention-MICCAI 2017 : 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*. Springer, p. 267-275 (cf. p. 18).
- WEYERS, Wolfgang (2004). « Carl Rokitansky : his life, work, and contributions to dermatopathology on the occasion of his 200th birthday ». In : *The American Journal of Dermatopathology* 26.5, p. 431-438 (cf. p. 4).
- XIA, Fangting et al. (2016). « Zoom better to see clearer : Human and object parsing with hierarchical auto-zoom net ». In : *Computer Vision-ECCV 2016 : 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*. Springer, p. 648-663 (cf. p. 18).
- YANG, Li et Abdallah SHAMI (2020). « On hyperparameter optimization of machine learning algorithms : Theory and practice ». In : *Neurocomputing* 415, p. 295-316 (cf. p. 12, 47).
- ZAIDI, Fatima (2021). « Deep learning-based scale-invariant cancer detection from whole slide image ». Thèse de doct. Murdoch University (cf. p. 18).
- ZEILER, Matthew D (2012). « Adadelta : an adaptive learning rate method ». In : *arXiv preprint arXiv :1212.5701* (cf. p. 48).
- ZENG, Qinghe et al. (2022). « Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology ». In : *Journal of Hepatology* (cf. p. 36).
- ZHOU, Bolei et al. (2016). « Learning deep features for discriminative localization ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 2921-2929 (cf. p. 46).
- ZHOU, Xiang et al. (2020). « Virtual & augmented reality for biological microscope in experiment education ». In : *Virtual Reality & Intelligent Hardware* 2.4, p. 316-329 (cf. p. 59).

ZOPH, Barret et al. (2018). « Learning transferable architectures for scalable image recognition ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 8697-8710 (cf. p. [45](#), [50](#)).