



**HAL**  
open science

# Apport des outils d'intelligence artificielle pour le diagnostic des lésions intra-épithéliales et invasives des voies aérodigestives supérieures

Yaëlle Bellahsen-Harrar

► **To cite this version:**

Yaëlle Bellahsen-Harrar. Apport des outils d'intelligence artificielle pour le diagnostic des lésions intra-épithéliales et invasives des voies aérodigestives supérieures. Médecine humaine et pathologie. 2022. dumas-04152015

**HAL Id: dumas-04152015**

**<https://dumas.ccsd.cnrs.fr/dumas-04152015v1>**

Submitted on 24 Oct 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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## **AVERTISSEMENT**

Cette thèse d'exercice est le fruit d'un travail approuvé par le jury de soutenance et réalisé dans le but d'obtenir le diplôme d'Etat de docteur en médecine. Ce document est mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt toute poursuite pénale.

Code de la Propriété Intellectuelle. Articles L 122.4

Code de la Propriété Intellectuelle. Articles L 335.2 - L 335.10

**UNIVERSITÉ PARIS CITÉ**

Faculté de Santé

UFR de Médecine

Année 2022

N°

**THÈSE POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN MÉDECINE**

Présentée et soutenue publiquement le : 16/06/2022

Par

**Yaëlle BELLAHSEN-HARRAR**

**Apport des outils d'intelligence artificielle pour le diagnostic des  
lésions intra-épithéliales et invasives des voies aérodigestives  
supérieures**

**Dirigée par Mme la Professeure Cécile BADOUAL**

JURY

M. le Professeur Eric Deutsch

Président

Mme la Professeure Cécile Badoual

Directrice

M. le Professeur Philippe Bertheau

Membre du jury

M. le Professeur Julien Calderaro

Membre du jury

Mme la Docteure Marion Classe

Membre du jury

## Résumé :

**Introduction** Les lésions dysplasiques malpighiennes, précurseurs des carcinomes épidermoïdes des voies aéro-digestives supérieures (VADS), sont de diagnostic particulièrement difficile avec une reproductibilité inter et intra-observateur peu satisfaisante. Depuis quelques années, des algorithmes d'intelligence artificielle (IA) de plus en plus nombreux et performants sont développés pour assister le pathologiste dans ses diagnostics. L'objectif de ce projet est le développement d'un algorithme d'IA pour le diagnostic automatisé des lésions dysplasiques et infiltrantes des VADS, selon la classification OMS 2017. Ce travail a été réalisé en partenariat avec une ingénieure en *data science* en thèse Cifre (Ecole des Mines de Paris), dans le cadre d'une collaboration entre l'AP-HP et la société Keen Eye Technologies. **Matériel et Méthodes** L'algorithme a été développé sur 2360 lames scannées issues du service d'anatomie pathologique de l'Hôpital Européen Georges Pompidou, concernant au total plus de 600 patients pris en charge pour des lésions dysplasiques ou de carcinomes épidermoïdes infiltrant diagnostiqués entre 2000 et 2013, avec la relecture de 127 lames de dysplasies dont la classification ne correspondait pas aux catégories de 2017 (dysplasie légère à modérée). Les nombreux échanges et partages de connaissance entre interne en pathologie et *data scientist* ont permis de proposer un modèle d'architecture originale, nommé *Spatial Attention*. Ses prédictions prennent en compte le contexte spatial des lésions et l'analyse combine une lecture différenciée de la surface du prélèvement et de sa profondeur. Ses performances ont été comparées à celles des modèles de référence dans la littérature, suivant la métrique d'exactitude pondérée proposée par la Société Française de Pathologie lors du *data challenge* de 2020. Sa reproductibilité a été comparée aux données publiées dans les études de reproductibilité inter-pathologistes. **Résultats** Notre modèle de *Spatial Attention* est plus performant que les modèles de référence, avec une exactitude pondérée de 0,879. Sa reproductibilité est bonne, de 0,757 et supérieure à celles observées dans la littérature. Il reste perfectible, notamment pour la classification des dysplasies de bas grade qui sont moins représentées dans le jeu de données et plus subtiles. **Conclusion** La collaboration entre pathologiste et *data scientist* a permis de proposer un algorithme original et performant, adapté à la problématique des lésions malpighiennes des VADS.

Spécialité : DES Anatomie et Cytologie Pathologiques

**Mots clés français** : voies aérodigestives supérieures, dysplasie malpighienne, carcinome, intelligence artificielle, *deep learning*, algorithme, lames numérisées

### Forme ou Genre :

fMeSH : Dissertation universitaire

Rameau : Thèses et écrits académiques

## Abstract :

**Introduction** Diagnosis of pre-neoplastic lesions of the head and neck is a real challenge for pathologists, with a poor inter- and intra-reliability. In recent years, many works achieved to create AI algorithms which assist the pathologists in their practice. The aim of this project is the development of an algorithm for the automatic diagnosis of squamous dysplasia and carcinomas following the 2017 WHO Classification. This work was realized as a pair of a pathology resident (Université Paris Cité) and a data scientist (Thèse Cifre – Ecole des Mines de Paris), with a collaboration between AP-HP and Keen Eye Technologies. **Materials and Methods** The algorithm was trained on 2360 virtual slides from the pathology department of Hopital European Georges Pompidou, coming from more than 600 patients diagnosed between 2000 and 2013. 127 slides of « mild to moderate » dysplasias were reviewed by an expert in order to follow the WHO 2017 classification. The sharing of our expertise allowed us to create a new model that we named *Spatial Attention*. Its predictions take into account the local context of the lesion and its presence at the surface or at the depth of the sample. Its performance was assessed with the weighted accuracy used during the 2020 data challenge organised by the French Pathology Society, compared to the gold standards described in the literature. Its reliability was calculated and compared with the inter-rater agreement between pathologists published in the literature. **Results** Our model has a weighted accuracy of 0.879 and a kappa index of 0.757, which are higher than the best performances reported in the literature. Its performances can be perfected, particularly for low grade dysplasia which is a difficult diagnosis and less represented in our dataset. **Conclusion** The collaboration between pathologist and data scientist resulted in the development of a new model, which fits the particularities of the head and neck squamous lesions.

**English keywords :** head and neck, dysplasia, carcinoma, artificial intelligence, algorithm, virtual slides

**Publication type :**

MeSH : Academic Dissertation

# Table des matières

<b>REMERCIEMENTS</b> .....	<b>3</b>
<b>TABLE DES MATIERES</b> .....	<b>7</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>9</b>
<b>INTRODUCTION</b> .....	<b>10</b>
<b>CHAPITRE 1. GENERALITES SUR LES CARCINOMES DES VOIES AERODIGESTIVES SUPERIEURES</b> .....	<b>11</b>
1. ÉPIDEMIOLOGIE .....	11
2. FACTEURS DE RISQUE .....	12
3. DIAGNOSTIC .....	13
4. PRISE EN CHARGE DES LESIONS MALPIGHIENNES DES VADS.....	15
<b>CHAPITRE 2. CLASSIFICATION DES DYSPLASIES MALPIGHIENNES DES VADS</b> .....	<b>16</b>
1. OBJECTIF DES CLASSIFICATIONS DES LESIONS DYSPLASIQUES.....	16
2. PREMIER MODELE DE CLASSIFICATION DES DYSPLASIES MALPIGHIENNES : LE COL UTERIN .....	16
3. PREMIERE CLASSIFICATION DES LESIONS MALPIGHIENNES DES VADS.....	18
4. MULTIPLICATION DES CLASSIFICATIONS .....	19
5. SIMPLIFICATION DE LA CLASSIFICATION ET DES DENOMINATIONS .....	21
6. CLASSIFICATION OMS 2017 : ETAT DES LIEUX, PERSPECTIVES .....	22
<b>CHAPITRE 3. L'INTELLIGENCE ARTIFICIELLE : UNE TECHNOLOGIE MISE AU SERVICE DU DIAGNOSTIC ANATOMO-PATHOLOGIQUE</b> .....	<b>26</b>
1. PRINCIPES GENERAUX D'INTELLIGENCE ARTIFICIELLE .....	26
2. RESEAUX DE NEURONES ET <i>DEEP LEARNING</i> .....	27
3. RESEAU DE NEURONES CONVOLUTIFS.....	30
1.1 <i>Extraction de features</i> .....	31
1.2 <i>Conclusion de l'analyse</i> .....	33
4. APPRENTISSAGE DE L'ALGORITHME .....	35
5. CONSTRUCTION D'UN JEU DE DONNEES .....	37
6. TACHES REALISABLES PAR UN CNN – APPLICATIONS A L'IMAGE HISTOLOGIQUE .....	37
6.1 <i>Classification</i> .....	38
6.2 <i>Régression</i> .....	38
6.3 <i>Segmentation</i> .....	38
6.4 <i>Détection d'objets</i> .....	38
7. UTILISATION DE L'IA EN ANATOMIE PATHOLOGIQUE : REVUE DE LA LITTERATURE .....	39
7.1 <i>L'algorithme comme assistant du pathologiste</i> .....	39
7.2 <i>L'algorithme comme outil de recherche</i> .....	41
<b>CHAPITRE 4. UTILISATION DE L'IA DANS LE DIAGNOSTIC DES LESIONS MALPIGHIENNES DES VADS : ETAT DE L'ART, PERSPECTIVES OBJECTIF DE LA THESE</b> .....	<b>43</b>
1. UTILISATION DE L'IA POUR LES LESIONS DES VADS : ETAT DE L'ART ET PERSPECTIVES .....	43
2. OBJECTIF DE LA THESE .....	44
<b>MATERIEL ET METHODES</b> .....	<b>45</b>
<b>CHAPITRE 1. CONSTITUTION DU PROJET PREINV</b> .....	<b>46</b>
1. CADRE JURIDIQUE – PROTECTION DES DONNEES.....	46
2. CONSTITUTION DE LA BASE DE DONNEES.....	47
3. SELECTION DES LAMES .....	50
4. NUMERISATION DES LAMES ET ANONYMISATION .....	51
5. CONTROLE DE LA QUALITE ET RELECTURE DES LAMES NUMERISEES .....	52

<b>CHAPITRE 2. PREPARATION DES DONNEES POUR L'ALGORITHME DE CLASSIFICATION DES LESIONS.....</b>	<b>53</b>
1. ATTRIBUTION D'UN LABEL GLOBAL ET <i>SPLIT</i> DES DONNEES .....	53
2. PRETRAITEMENT DES IMAGES LAME ENTIERE .....	54
3. TUILAGE .....	54
<b>CHAPITRE 3. ARCHITECTURES ET CONFIGURATIONS TESTEES .....</b>	<b>56</b>
1. INITIALISATION DES POIDS DE L'EXTRACTEUR DE <i>FEATURES</i> .....	56
2. ARCHITECTURES TESTEES .....	57
2.1 <i>État de l'art</i> .....	57
2.2 <i>Nouvelle approche : le modèle de Spatial Attention</i> .....	61
3. METHODES DE CALCUL DE LA PERTE .....	66
<b>CHAPITRE 4. ÉVALUATION DE LA PERFORMANCE DES MODELES .....</b>	<b>68</b>
1. METRIQUES.....	68
1.1 <i>Exactitude pondérée</i> .....	68
1.2 <i>Sensibilité, spécificité, précision</i> .....	69
1.4 <i>AUC de la courbe ROC</i> .....	69
1.5 <i>Coefficient kappa de reproductibilité</i> .....	70
2. INTERPRETATION DE LA CLASSIFICATION DES LAMES VIRTUELLES .....	70
2.1 <i>Cartographie des scores d'attention superposée à l'image histologique (heatmap)</i> .....	70
2.2 <i>Visualisation des features extraites par le modèle</i> .....	71
<b>CARACTERISATION DE LA COHORTE ET DU JEU DE DONNEES.....</b>	<b>72</b>
1. TABLEAUX DESCRIPTIFS DE LA COHORTE ET DU JEU DE DONNEES .....	72
2. EXEMPLES DE CAS RELUS PAR L'EXPERT .....	74
<b>RESULTATS.....</b>	<b>75</b>
<b>CHAPITRE 1. PERFORMANCES DES MODELES .....</b>	<b>76</b>
1. METRIQUES, MATRICES DE CONFUSION, COURBES ROC .....	76
2. REPRODUCTIBILITE DU MODELE DE <i>SPATIAL ATTENTION</i> A QUATRE BRANCHES.....	79
<b>CHAPITRE 2. ANALYSE DES <i>HEATMAPS</i>.....</b>	<b>80</b>
1. COMPORTEMENT DU MODELE DE <i>SPATIAL ATTENTION</i> A DEUX BRANCHES .....	80
2. COMPORTEMENT DU MODELE DE <i>SPATIAL ATTENTION</i> A QUATRE BRANCHES .....	82
<b>CHAPITRE 3. ANALYSE DES <i>FEATURES</i> EXTRAITES APRES INITIALISATION DES POIDS PAR SSL VERSUS IMAGENET .....</b>	<b>85</b>
<b>DISCUSSION .....</b>	<b>86</b>
1. FORCES ET LIMITES DU MODELE DE <i>SPATIAL ATTENTION</i> .....	87
2. VERS UNE NOUVELLE CLASSIFICATION DES DYSPLASIES ? .....	90
3. MISE EN PERSPECTIVE DE NOTRE TRAVAIL DE THESE AVEC LE <i>DATA CHALLENGE</i> DE LA SFP .....	92
3.1 <i>Apports du challenge</i> .....	92
3.2 <i>Différences entre les jeux de données</i> .....	93
3.3 <i>Différences de moyens techniques et administratifs</i> .....	94
<b>CONCLUSION .....</b>	<b>96</b>
<b>LEGENDES DES FIGURES .....</b>	<b>97</b>
<b>LEGENDES DES TABLEAUX .....</b>	<b>98</b>
<b>BIBLIOGRAPHIE.....</b>	<b>99</b>
<b>ANNEXE .....</b>	<b>104</b>
<b>PERMIS D'IMPRIMER .....</b>	<b>105</b>

## Liste des abréviations

---

ADN : acide désoxyribonucléique  
AP-HP : assistance publique hôpitaux de Paris  
AUC : *area under the curve* (aire sous la courbe)  
CEVADS : carcinome épidermoïde des voies aérodigestives supérieures  
CIN : *cervical intraepithelial neoplasia* (néoplasie intra-épithéliale cervicale)  
CNN : *convolutional neural network* (réseau de neurones convolutifs)  
CPP : comité de protection des personnes  
DPO : département de protection des données  
DRCI : direction de la recherche clinique et de l'innovation  
EBO : endobrachyoesophage  
HEGP : hôpital européen Georges Pompidou  
HPV : *human papillomavirus*  
HSIL : *high grade intraepithelial lesion* (lésion malpighienne intraépithéliale de haut grade)  
IA : intelligence artificielle  
IRM : imagerie par résonance magnétique  
KE : Keen Eye Technologies  
LIN : *laryngeal intraepithelial neoplasia* (néoplasie intraépithéliale laryngée)  
LSIL : *low grade squamous intraepithelial lesion* (lésion malpighienne intraépithéliale de bas grade)  
OMS : organisation mondiale de la santé  
ROI : *region of interest* (région d'intérêt)  
SFP : société française de pathologie  
SIN : *squamous intraepithelial neoplasia* (néoplasie malpighienne intraépithéliale)  
SOSR : *One-sided Regression Loss*  
SSL : *self-supervised learning* (apprentissage auto-supervisé)  
URC : unité de recherche Clinique  
VADS : voies aérodigestives supérieures  
WCE : *weighted cross-entropy*

## Introduction

---

# Chapitre 1. Généralités sur les carcinomes des voies aérodigestives supérieures

---

## 1. Épidémiologie

Les carcinomes épidermoïdes des voies aérodigestives supérieures (CEVADS) constituent un problème majeur de santé publique, classés mondialement 6<sup>ème</sup> cancer en termes d'incidence et de mortalité. Ils font partie des cancers au pronostic le plus sombre, avec une survie globale à 5 ans variant de 50 à 60% selon la localisation anatomique <sup>(1)</sup> et une survie médiane en cas de récurrence ou de métastase de 10 mois <sup>(2)</sup>. Ces chiffres s'expliquent par une prise en charge souvent à un stade avancé de la maladie et par une réponse médiocre aux traitements anti-cancéreux conventionnels. Lorsqu'une prise en charge curative est possible, la morbidité reste élevée avec des séquelles fonctionnelles handicapantes : aphonie, troubles de la déglutition et de la mastication en particulier <sup>(3)</sup>.

Les lésions cancéreuses tirent leur agressivité de plusieurs caractéristiques : leur capacité à envahir et détruire leur tissu d'origine, puis à atteindre des vaisseaux sanguins ou des filets nerveux et à travers eux à métastaser dans d'autres organes à distance, qui seront détruits à leur tour. Le problème des CEVADS est qu'il existe peu de signes d'alerte cliniques au stade initial. En effet, lorsque la lésion est de petite taille, celle-ci est rarement symptomatique. En particulier dans le larynx, l'atteinte d'une corde vocale peut donner lieu à une dysphonie, mais les atteintes supra- et infraglottiques sont en général asymptomatiques au début de la maladie <sup>(4)</sup>. Ainsi, les patients consultent leur médecin le plus souvent lorsqu'ils ont déjà une symptomatologie marquée ou handicapante, notamment une dysphagie ou une dyspnée, ou lorsqu'ils palpent une adénopathie cervicale métastatique... C'est-à-dire souvent à un stade où le cancer est déjà localement avancé, donc potentiellement non opérable et pouvant avoir

déjà disséminé à distance. Cette difficulté à prendre en charge précocement l'atteinte laryngée s'illustre par l'absence d'amélioration de la survie des patients pour cette localisation anatomique depuis les années 1990 <sup>(1)</sup>.

## **2. Facteurs de risque**

Le principal agent causal des CEVADS est l'intoxication tabagique, qui possède des effets cancérigènes multiples : des effets mutagènes sur l'ADN <sup>(1)</sup>, l'induction d'une immunotolérance locale favorisant l'échappement tumoral à la pression immunitaire notamment en inhibant l'activité des lymphocytes Natural Killer <sup>(5)</sup> ou l'induction d'espèces réactives de l'oxygène. Des mutations s'observent dans la muqueuse non tumorale des personnes tabagiques <sup>(6)</sup>. Il est à noter que les effets du tabac peuvent être potentialisés par une intoxication éthylique associée <sup>(7)</sup>. En Asie, le mâchage de noix de bétel joue également un rôle notable dans la carcinogénèse des CEVADS. D'autres facteurs de risque peuvent être évoqués comme une immuno-déficience, un antécédent de radiothérapie <sup>(8)</sup>.

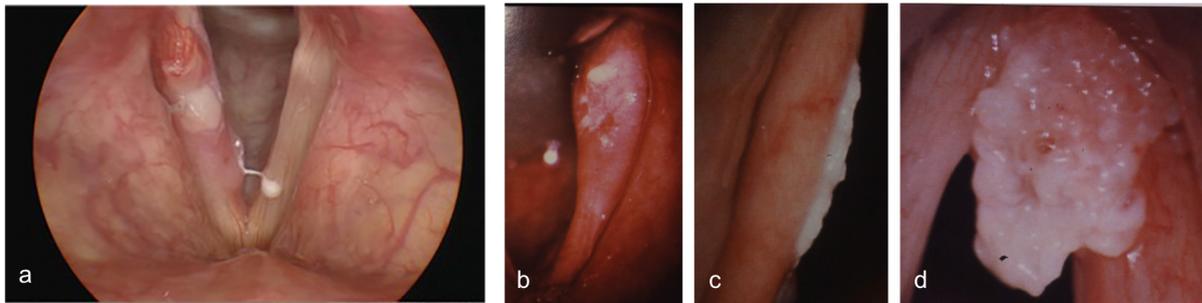
L'infection par les papillomavirus humains (HPV) joue également un rôle dans l'épidémiologie des CEVADS. Cependant, la grande majorité des cas induits par HPV se localise dans l'oropharynx. Ils constituent alors une entité clinique à part entière, avec des aspects histologiques différents des cas induits par le tabac, notamment l'absence de kératinisation des lésions <sup>(9)</sup>. Ils sont également de meilleur pronostic <sup>(1),(10)</sup>.

L'acquisition de mutations oncogéniques se répercute sur l'aspect morphologique des cellules épithéliales et sur l'architecture de l'épithélium malpighien tapissant les VADS. Ces aspects précancéreux sont également nommés lésions dysplasiques ou lésions intra-épithéliales. La première description de lésion dysplasique dans le larynx a été faite par Stout et al en 1952 <sup>(11)</sup>. A noter que dans les carcinomes oropharyngés HPV-induits, l'épithélium de surface ne

présente pas de caractéristiques dysplasiques ou de façon exceptionnelle <sup>(12)</sup>. L'existence de lésions dysplasiques dans ce contexte fait donc débat et ne sera pas traitée plus en détail dans ce travail.

### 3. Diagnostic

A l'examen clinique, les lésions dysplasiques ou de carcinomes infiltrants se traduisent par des plaques érythémateuses ou blanchâtres de la muqueuse, planes ou en relief, ou par un épaissement de la muqueuse. Aucun de ces aspects n'est spécifique de la présence ou de l'absence d'une invasion, ou indicateur d'un plus mauvais pronostic <sup>(4),(13)</sup>.



**Figure 1. Photographies de lésions des cordes vocales durant une laryngoscopie.** a. Leucoplasie correspondant à une dysplasie malpighienne légère à modérée, d'après Park et al <sup>(14)</sup> ; b. érythroleucoplasie, c. leucoplasie, d. carcinome épidermoïde verruqueux (Pr Laccourreye – service ORL, HEGP)

A l'examen microscopique, on observe que les lésions dysplasiques des VADS comportent des éléments communs à toute lésion dysplasique épithéliale : augmentation du rapport nucléocytoplasmique, anisocaryose, noyaux aux contours irréguliers et hyperchromatiques, présence d'un nombre de mitoses supérieur à la normale, cellules apoptotiques. La maturation épithéliale est altérée, ce qui se traduit par la présence de cellules dyskératosiques. La grande majorité des dysplasies des VADS sont kératinisées <sup>(15)</sup> ; et inversement, la présence de kératinisation ne traduit pas nécessairement un état de dysplasie et peut se voir dans un contexte inflammatoire ou infectieux.

Le carcinome infiltrant se définit par l'effraction de la lame basale par les cellules et leur pénétration dans la profondeur du tissu. Le carcinome micro-invasif, c'est-à-dire le tout début de l'invasion, n'est pas précisé dans la classification OMS (Organisation Mondiale de la Santé) ; il est généralement défini par la présence de cellules isolées ou en nids dans le chorion superficiel avec une distance d'au moins 0,5 mm à la lame basale <sup>(16)</sup>. Le diagnostic peut être difficile sur des prélèvements biopsiques de petite taille, fragmentés ou avec une incidence de coupe tangentielle. Or ces lésions possèdent un réel potentiel métastatique. En effet, les carcinomes micro-invasifs laryngés supra glottiques sont associés à des métastases dans près de 20 % des cas <sup>(15)</sup>.

Ainsi, le challenge diagnostique se fait à deux niveaux : la détection correcte des dysplasies pour prévenir le cancer infiltrant, et l'identification de la micro-invasion pour une exérèse précoce et moins délabrante.

On estime que la détection précoce des lésions dysplasiques pourrait permettre la prévention de 88 % des CEVADS, soit la quasi-totalité d'entre eux <sup>(8)</sup>. Cette détection est possible dans deux cadres : d'une part, chez un patient vu pour la première fois pour des symptômes dus à des lésions strictement dysplasiques (par exemple, dysphonie secondaire à des lésions dysplasiques d'une corde vocale qui modifient ses capacités vibratoires). D'autre part, au décours du suivi d'un patient ayant un antécédent de CEVADS : la toxicité tabagique étant pan-muqueuse et pouvant avoir des effets persistant plusieurs années après l'arrêt de toute consommation <sup>(17),(18)</sup>, notamment sur l'expression de certains gènes comme ceux de la voie Wnt/bêta-caténine (ces effets ont surtout été démontrés sur l'épithélium bronchique) <sup>(19)</sup>. Par ailleurs, l'évolution des lésions induites par le tabac suit le schéma de la *field cancerization* <sup>(20)</sup>. En effet, l'exposition au carcinogène va induire des mutations différentes dans les cellules (et pas dans

100% de celles-ci) ; celles qui auront des mutations *driver* de l'oncogenèse vont se multiplier de façon clonale et créer des "patches" dysplasiques qui peuvent être synchrones ou apparaître successivement. Ainsi, une deuxième lésion peut apparaître plus tard, ainsi un premier diagnostic de carcinome exige un dépistage d'un potentiel nouveau cancer par un examen clinique régulier de l'ensemble des VADS.

#### **4. Prise en charge des lésions malpighiennes des VADS**

Actuellement, le suivi des patients ayant un carcinome infiltrant fait l'objet de recommandations, avec un suivi régulier pour la réalisation d'un examen clinique voire endoscopique au moindre doute de récurrence, et la réalisation d'examens d'imagerie (scanner, IRM) dans les premiers mois post-opératoires <sup>(21)</sup>. Cependant, la prise en charge thérapeutique et le suivi des lésions dysplasiques sans invasion sont variables selon les centres, aussi bien dans les techniques (biopsies répétées, exérèse large, voire radiothérapie) que dans le rythme des consultations de suivi <sup>(22),(23)</sup>. L'aspect clinique des lésions dysplasiques n'étant pas strictement corrélé au risque d'évolution en cancer invasif <sup>(4)</sup>, une analyse plus fine des lésions, à l'échelle histologique, devrait en théorie permettre de mieux évaluer leur potentiel évolutif. Cependant, cette analyse est complexe et fait l'objet de plusieurs difficultés, que nous allons expliciter ci-après.

## Chapitre 2. Classification des dysplasies malpighiennes des VADS

---

### 1. Objectif des classifications des lésions dysplasiques

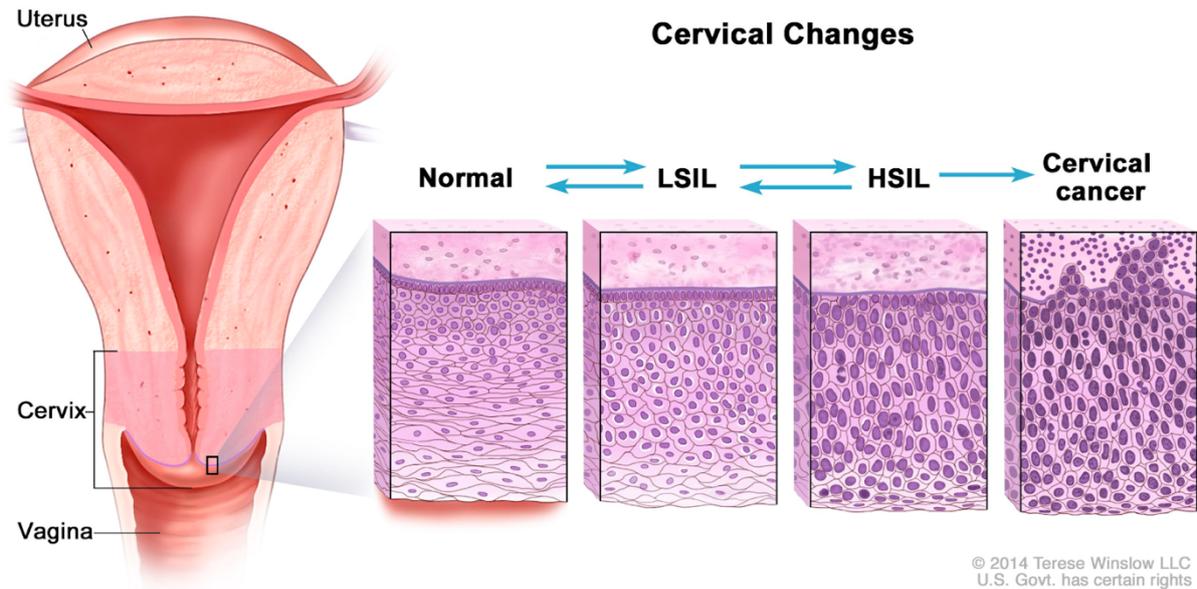
Afin d'accompagner au mieux la prise en charge des patients, les pathologistes tentent de classer les lésions dysplasiques en définissant des « degrés de sévérité », ou « grades histologiques ». Pour pouvoir remplir ces objectifs, une classification doit se traduire par une distinction nette des lésions selon leurs conséquences pronostiques pour les patients, tout en utilisant des critères bien définis et reproductibles entre les pathologistes.

Les critères généraux utilisés pour les lésions épithéliales dysplasiques, dans les VADS comme dans les autres épithéliums (tube digestif, peau...) sont à la fois cytologiques : atypies cytonucléaires, activité mitotique, apoptoses, dysmaturations cellulaires ; et architecturaux : désorganisation de l'architecture épithéliale. Tous ces critères sont combinés pour grader la dysplasie.

### 2. Premier modèle de classification des dysplasies malpighiennes : le col utérin

Historiquement, la première classification de dysplasies malpighiennes a concerné un autre organe que les VADS : le col de l'utérus. En effet, comme les VADS, le col utérin est revêtu en partie par un épithélium malpighien qui peut également se transformer en lésion dysplasique puis cancéreuse lorsqu'il est infecté de façon chronique par HPV. Ainsi, le premier modèle de classification des dysplasies en plusieurs grades de sévérité a été proposé pour cet organe en 1953 <sup>(11)</sup>. Actuellement, la classification de ces lésions est robuste et le pathologiste peut s'aider d'immunomarquages pour affiner le grade histologique. Cette classification a pour point

fort de reposer sur des éléments physiopathologiques. En effet, la dysplasie de bas grade concerne l'atteinte des cellules basales de l'épithélium, qui sont les premières cellules infectées par le virus, et la présence d'HPV se traduit par un effet cytopathogène viral donnant aux cellules malpighiennes en surface un aspect particulier nommé koïlocyte. L'analyse du tissu par PCR ou hybridation *in situ* permet également de documenter la présence de virus HPV à haut risque de transformation cancéreuse. La progression vers la dysplasie de haut grade correspond à l'atteinte progressive de l'épaisseur de l'épithélium, qui est corrélée à l'intégration de l'ADN viral à celui de la cellule épithéliale et à l'expression forte des oncoprotéines virales E6 et E7 <sup>(24)</sup>. Ainsi, cette méthode permet de préciser le degré de progression des lésions dans leur voie vers la cancérisation et donc de donner une estimation assez précise du risque de cancer. Les premières classifications comportaient 3 grades allant du moins au plus sévère selon le degré d'atypies cytonucléaires, l'épaisseur de l'épithélium atteinte et l'activité mitotique, les CIN (*cervical intraepithelial neoplasia*) de grade 1 à 3. La nouvelle classification adoptée en 2014 <sup>(25)</sup> reconnaît deux grades, un léger (LSIL pour *low grade squamous intraepithelial lesion*) et un sévère (HSIL pour *high grade squamous intraepithelial lesion*). Cette dernière classification est censée mieux discriminer le potentiel évolutif des lésions que le système à trois grades, les LSIL ayant un potentiel évolutif très faible.



**Figure 2. Chronologie des lésions malpighiennes du col utérin - National Cancer Institute**

Paradoxalement, ces aspects dysplasiques distincts dans le col utérin ne sont pas observés dans l'oropharynx, bien que les carcinomes épidermoïdes y soient très majoritairement HPV-induits.

### **3. Première classification des lésions malpighiennes des VADS**

La première classification des lésions précancéreuses des VADS fut proposée par Kleinsasser pour la muqueuse laryngée en 1963 <sup>(26)</sup>. Celle-ci reposait sur le degré d'atypies nucléaires et classait les lésions en trois catégories : hyperplasie simple, hyperplasie atypique, carcinome *in situ*. Ce n'est que vingt ans plus tard, en 1974, que la notion de dysplasie laryngée fit consensus, lors de la *Toronto Centennial Conference of Laryngeal Cancer*. Cette conférence fut le point de départ d'autres travaux qui aboutirent en 1978 à la première classification des dysplasies laryngées de l'OMS, reprécisée en 1991. Celle-ci proposait une classification en quatre grades : dysplasie légère, modérée, sévère et carcinome *in situ*.

Cette classification avait plusieurs problématiques. D'une part, il s'agissait d'une adaptation de la classification des lésions dysplasiques du col utérin à la muqueuse des VADS <sup>(15),(27)</sup>. Or même si les muqueuses malpighiennes du col utérin et des VADS partagent des caractéristiques communes, dans la très grande majorité des cas les dysplasies des VADS sont induites par le tabac, l'alcool... et non par HPV. La corrélation entre les lésions histologiques et la progression vers le cancer dans les cas induits par HPV de la classification des lésions du col ne s'applique donc pas <sup>(4),(26),(28)</sup>. Les lésions sont également très différentes, celles induites par le tabac étant en général kératinisées, au contraire des lésions induites par HPV <sup>(15)</sup>. D'autre part, la présence de multiples catégories n'aidait pas à la reproductibilité, en induisant une variabilité à la fois inter et intra-observateur élevée <sup>(4)</sup>. Enfin, l'entité carcinome *in situ* ne faisait pas consensus même entre les experts de la pathologie des VADS, puisque ces derniers ne sont pas d'accord entre eux sur les critères la définissant (encore le cas actuellement) <sup>(29)</sup>.

#### **4. Multiplication des classifications**

Depuis cette première classification internationale, de nombreuses autres classifications furent proposées par des groupes de pathologistes, avec chacune une autre terminologie : *squamous intraepithelial neoplasia* (SIN) pour Friedmann et Osborn en 1976, *intraepithelial neoplasia of the larynx* pour Crissman et Fu en 1986, *laryngeal intraepithelial neoplasia* (LIN) pour Friedmann et Ferlito en 1988 <sup>(11)</sup>. La multiplication des possibilités pour classer les lésions dysplasiques a été une source de confusions et de discordes sans qu'aucune méthode ne se montre clairement supérieure. Ceci rendant la comparaison des études très difficile.

Nomenclature	Author(s)	Year	References
Intraepithelial carcinoma	Altmann et al.	1952	[1]
Carcinoma in situ	Reagan et al.	1953	[9]
Proliferative squamous cell lesions	Kleinsasser	1963	[7]
Carcinoma in situ	Miller and Batsakis	1974	[10]
Squamous intraepithelial neoplasia (SIN)	Friedmann and Osborn	1976	[13]
Dysplasia, carcinoma in situ	WHO classification	1978	[11]
Intraepithelial neoplasia of the larynx	Crissman and Fu	1986	[42]
Laryngeal intraepithelial neoplasia (LIN)	Friedmann and Ferlito	1988	[36]
Squamous intraepithelial lesion (SIL)	Gale et al.	2014	[55]
Dysplasia: low-grade, high-grade	WHO classification	2017	[44]

**Tableau 1. Nomenclature des lésions dysplasiques laryngées, d'après Hellquist et al**

Après treize ans de débats, l'OMS a publié en 2005 ses recommandations sur la classification des dysplasies. Celles-ci proposaient aux pathologistes de choisir entre trois méthodes différentes, dont aucune n'a prouvé par la suite être supérieure en termes de reproductibilité - qui était d'ailleurs moyenne - ni de prédiction pronostique <sup>(30),(31),(32),(33)</sup>. Cette absence de choix reflétait la difficulté de trouver des critères objectifs pour grader les aspects dysplasiques, même entre experts de la pathologie des VADS.

Les lésions dysplasiques de la cavité orale sont par ailleurs classées par un autre système proposé par l'OMS, et de multiples autres classifications ont été proposées par différents groupes de pathologistes <sup>(29)</sup> sans explication solide du rationnel de cette distinction des dysplasies orales du reste des VADS. Les agents causaux étant les mêmes que pour le larynx, les modifications dysplasiques devraient être semblables et donc gradées de façon similaire. La seule différence évoquée par l'OMS est la formation plus fréquente de "crêtes épithéliales" dans les lésions de dysplasie orale <sup>(28)</sup>.

Malgré cette hétérogénéité, les classifications des dysplasies malpighiennes laryngées parvenaient tout de même à différencier grossièrement les lésions les plus graves : de façon générale la probabilité de transformation en cancer était plus forte si le grade lésionnel était plus

élevé. Cela indique qu'au moins une partie des critères utilisés était pertinente. Notamment, une méta-analyse de la littérature de l'époque de Weller et al en 2010 pour les dysplasies laryngées <sup>(22)</sup> retrouvait une évolution vers un cancer infiltrant pour 10,6% des lésions de dysplasie légère à modérée et 30,4% des lésions de dysplasie sévère ou carcinome *in situ*. L'évolution moyenne de toutes les lésions, peu importe le grade de dysplasie, était de 13,6%. L'aspect lésionnel ne permettait pas de prédire le délai d'évolution vers un cancer, celles-ci évoluant toutes sur une large plage temporelle s'étendant de 3 à 14 ans. A noter également que dans cette étude, le traitement par radiothérapie des dysplasies jugées sévères n'a pas démontré de différence statistiquement significative dans la prévention de l'évolution en carcinome infiltrant.

## **5. Simplification de la classification et des dénominations**

Un travail publié en 2014 par Gale et al <sup>(29)</sup> a cherché à unifier les différentes classifications proposées par l'OMS et à les simplifier, en limitant à deux le nombre de catégories : une catégorie de dysplasie de bas grade, et une autre de haut grade regroupant les dysplasies modérées, sévères et les lésions de carcinome *in situ*. Elle a choisi de décrire les lésions par le terme *squamous intraepithelial neoplasia* (SIL). Cette étude rétrospective a porté sur des prélèvements de patients avec un suivi allant jusqu'à 31 ans. Ce travail a servi de base pour la dernière classification des dysplasies laryngées de l'OMS publiée en 2017. Cette classification est utilisée par analogie dans les autres localisations des VADS hors oropharynx.

Épaisseur atteinte	Ljubljana (2000)	SIN (2001)	OMS (2005)	LIN (2012)	Ljubljana (2014)	OMS (2017)
1/3 inférieur	Hyperplasie malpighienne	Hyperplasie malpighienne	Hyperplasie malpighienne		SIL de bas grade	Dysplasie de bas grade
	Hyperplasie basale/parabasale	SIN 1	Dysplasie légère	LIN 1		
1/3 - 1/2	Hyperplasie atypique	SIN 1 ou 2	Dysplasie modérée		SIL de haut grade	
2/3 inférieurs				LIN 2		
1/2 - ¾		SIN 2				Dysplasie de haut grade
Plus de 2/3				LIN 3		
Épaisseur totale	Carcinome <i>in situ</i>		Dysplasie sévère			
			Carcinome <i>in situ</i>		Carcinome <i>in situ</i>	

**Tableau 2. Résumé des différentes classifications des lésions dysplasiques publiées depuis 2000**

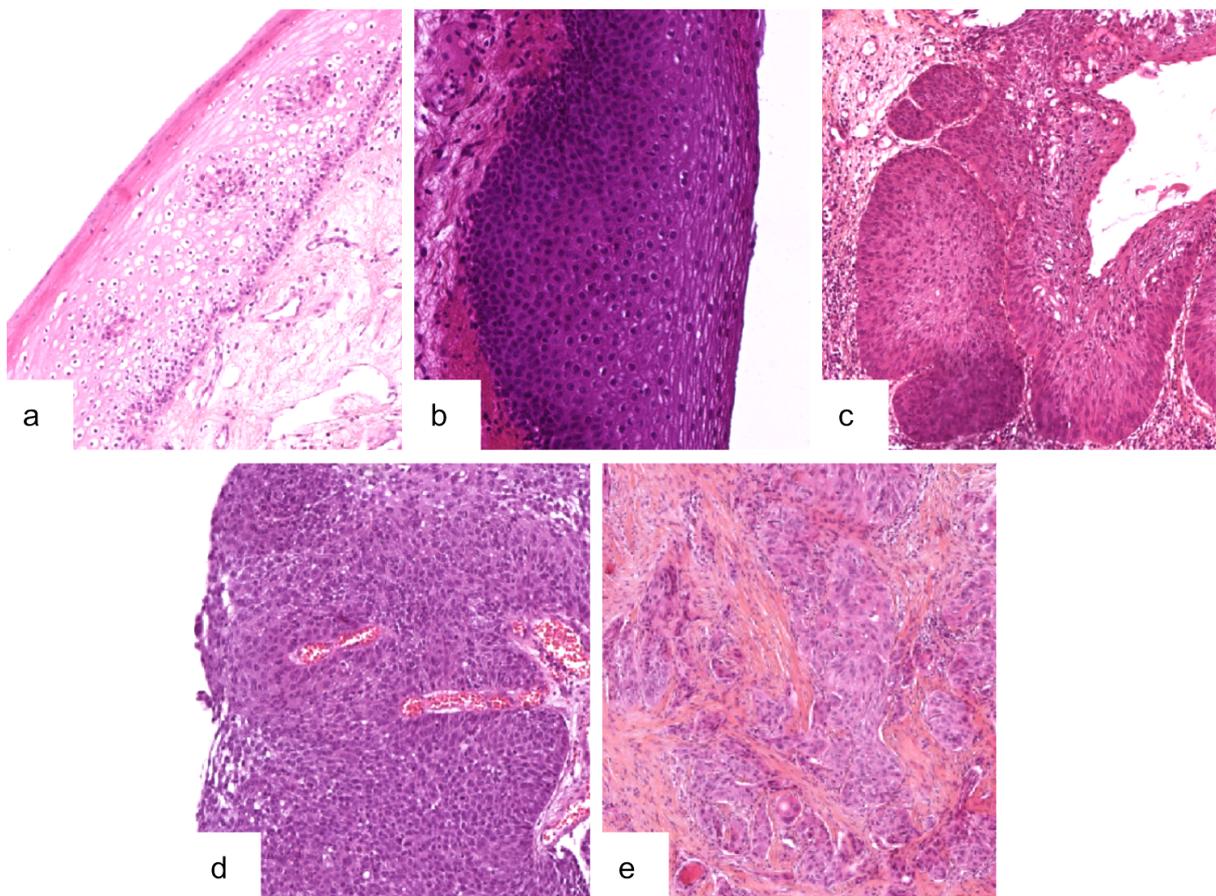
## 6. Classification OMS 2017 : état des lieux, perspectives

La classification OMS 2017 des dysplasies laryngées insiste sur les atypies cytonucléaires comme marqueurs de dysplasie, ainsi que sur la désorganisation architecturale qui est plus marquée dans les lésions de haut grade. Elle conserve une prise en compte de l'épaisseur de l'épithélium atteinte.

Cette simplification a permis un gain de reproductibilité significatif entre les pathologistes de l'étude, avec un coefficient de reproductibilité global très bon de 0,75. Ces deux grades lésionnels présentent une différence statistiquement significative dans le risque de transformation en cancer, avec 1,6 % d'évolution des lésions classées en bas grade et 12,5 % des lésions de haut grade sur la cohorte de l'étude. Le délai entre le diagnostic de dysplasie et de cancer invasif allait jusqu'à plus de vingt ans.

<b>Dysplasie de bas grade</b>	
Faible potentiel de malignité. Spectre de modifications morphologiques allant de l'hyperplasie à l'épaississement des couches parabasales ne dépassant pas la moitié inférieure de l'épithélium avec persistance de la maturation de surface	
Critères architecturaux	Conservation de la stratification : transition entre la couche des cellules basale ou basales/parabasales épaissie, perpendiculaire à la couche basale, et les cellules épineuses horizontales de la partie supérieure de l'épithélium
Critères cytologiques	Au plus, atypies cellulaires minimales Cellules parabasales : cytoplasme légèrement augmenté de volume comparé à cellules des cellules basales, noyau augmenté de volume, chromatine homogène, absence de pont intercellulaires Mitoses rares et normales dans la couche basale ou à proximité de celle-ci Peu de cellules dyskératosiques
<b>Dysplasie de haut grade</b>	
Lésion précancéreuse. Spectre de modifications comprenant des cellules épithéliales immatures occupant au moins la moitié inférieure de l'épithélium, jusqu'à toute son épaisseur	
Critères architecturaux	Maturation anormale Désorganisation architecturale et de la polarité à des degrés variables pouvant intéresser tout l'épithélium. Cellules épithéliales altérées occupant de la moitié à la totalité de l'épaisseur de l'épithélium Deux sous-types : kératinisant (« spinocellulaire ») et non-kératinisant (à « cellules basaloïdes ») Degré variable d'irrégularité dans la forme des crêtes épithéliales, avec persistance de l'intégrité de la membrane basale Absence de modification du stroma
Critères cytologiques*	Atypies cytologiques et nucléaires évidentes, incluant des variations marquées de taille et forme, des variations marquées d'affinité tinctoriale avec hyperchromatisme fréquent et une augmentation du nombre et de la taille des nucléoles Augmentation du rapport nucléo-cytoplasmique Augmentation des mitoses basales et suprabasales (avec ou sans mitose atypique) Cellules dyskératosiques et apoptotiques nombreuses dans tout l'épithélium
* La perte complète de la stratification et de la polarité et/ou la présence d'atypies cytonucléaires sévères ainsi que de mitoses atypiques font classer la lésion en carcinome <i>in situ</i> dans un système en 3 classes	

**Tableau 3. Classification OMS 2017 des dysplasies malpighiennes du larynx**  
(traduction en français par le Dr. Charles LEPINE)



**Figure 3. Exemples de lésions malpighiennes selon la classification OMS 2017, issues du jeu de données du projet PREINV.** a. preinv\_950 : épithélium malpighien normal ; b. preinv\_2259 : dysplasie de bas grade ; c. preinv\_1985 : dysplasie de haut grade (modérée) ; d. preinv\_1279 : dysplasie de haut grade (sévère) ; e. preinv\_1807 : carcinome épidermoïde infiltrant

Cette classification conserve une catégorie intermédiaire optionnelle entre la dysplasie légère et la dysplasie sévère. Cette formulation permet aux pathologistes de faire une description moins dichotomique des lésions, pour ne pas faire entrer à tout prix une image dans une case, tout en gardant une vraie significativité dans le risque d'évolution en cancer. L'usage est de nommer ces lésions intermédiaires dans les comptes rendus d'anatomie pathologique "dysplasie de haut grade (dysplasie modérée)". Ce choix est conforté par d'autres études qui n'ont pas retrouvé de différence statistique d'évolutivité entre les dysplasies modérées, sévères et les lésions de carcinome *in situ* <sup>(28)</sup>.

Malgré ces progrès, la classification des lésions malpighiennes dysplasiques présente des limites.

En effet, même si l'épaisseur de l'épithélium atteinte participe au grade, ce critère est difficile d'interprétation et présente des limites. L'épaisseur atteinte est difficile à estimer du fait de la variabilité physiologique de l'épaisseur de l'épithélium. Par exemple, au sein-même de la corde vocale, l'épaisseur de l'épithélium malpighien varie de 5 à 20 couches cellulaires <sup>(34)</sup>. Cette évaluation est d'autant plus problématique qu'il existe des discordances dans le « Blue Book » de l'OMS : selon les tableaux, l'épaisseur à considérer pour passer du bas grade au haut grade est soit le tiers (Table 3.02) soit la moitié de l'épaisseur (Table 3.03), ce qui ne facilite pas la tâche du pathologiste ! <sup>(35)</sup>.

D'autre part, la classification de l'OMS 2017 est difficile à effectuer car elle nécessite la prise en compte combinée de nombreux critères cytologiques et architecturaux subtils. Les aspects dysplasiques se situent sur un continuum lésionnel, avec des aspects très variés et progressifs, ce qui limite la reproductibilité entre pathologistes. C'est pourquoi la reproductibilité des pathologistes est moyenne : l'étude de Mehlum et al <sup>(35)</sup> qui a évalué la reproductibilité inter-

observateur de la classification OMS a retrouvé un coefficient kappa de reproductibilité de 0,46 entre les pathologistes participants, ce qui est bien inférieur à celui obtenu par Gale et al dans l'article princeps. A noter que la classification OMS 2022, qui sera publiée très prochainement <sup>(36)</sup>, ne semble pas apporter de modification majeure à la classification des dysplasies.

## **Chapitre 3. L'Intelligence Artificielle : une technologie mise au service du diagnostic anatomo-pathologique**

---

Depuis quelques années, l'utilisation d'outils d'analyse d'images par Intelligence Artificielle (IA) est en plein essor dans notre discipline. Elle apporte une aide précieuse dans l'évaluation des lésions tissulaires et assiste le pathologiste dans l'analyse fine des prélèvements. Les lésions malpighiennes dysplasiques et infiltrantes des VADS étant d'analyse complexe, l'IA a le potentiel de fournir des outils d'aide diagnostique précieux pour les pathologistes.

Nous allons introduire ci-dessous les principes généraux de l'analyse d'images par IA, ses applications actuelles en pathologie et ses potentiels apports pour l'analyse des lésions malpighiennes des VADS.

### **1. Principes généraux d'Intelligence Artificielle**

L'Intelligence Artificielle (IA) désigne un ensemble de méthodes cherchant à imiter l'intelligence humaine, pour effectuer à grande échelle des tâches variées aboutissant à un certain degré d'abstraction. Quelques exemples d'utilisation d'outils d'IA dans la vie quotidienne sont bien connus du grand public : les traducteurs de langue, la reconnaissance d'images, les systèmes de conduite automatique, les adversaires virtuels dans les jeux en ligne, etc. Les outils d'IA gagnent également une place croissante dans la médecine. Ils sont particulièrement intéressants dans ce domaine car ils arrivent à exploiter et tirer profit de la grande quantité et diversité des données disponibles, qui sont maintenant archivées informatiquement. Notamment, le séquençage génétique, l'imagerie médicale et l'anatomo-pathologie digitalisée sont des domaines de la médecine pour lesquels des algorithmes d'IA sont en plein développement <sup>(37)</sup>.

Les méthodes d'IA s'inspirent du fonctionnement du cerveau des animaux et de l'Homme afin de créer des algorithmes qui apprennent à réaliser leur tâche en acquérant de l'expérience, au contraire des logiciels standards qui sont codés de façon à s'exécuter de façon rigide, pré-déterminée et sans être modelés par les données qu'ils traitent. C'est le concept de *machine learning* : la machine apprend sa tâche et optimise ses performances à partir des données. Ces données peuvent prendre plusieurs formes, des tableaux de valeurs, des textes, des données génomiques ou encore des images ou des vidéos.

## 2. Réseaux de neurones et *deep learning*

Pour l'analyse d'images, les premiers algorithmes ont été développés en s'inspirant de la physiologie animale, grâce aux découvertes médicales sur le fonctionnement de la vision <sup>(38)</sup> et la structure d'un neurone. Schématiquement, un neurone biologique traite des données entrant dans la cellule via de multiples dendrites, puis émet un signal de sortie via un unique axone connecté à un dendrite d'un autre neurone. Les neurones impliqués dans la vision sont nombreux et organisés dans le cortex visuel des animaux et des humains en couches connectées les unes aux autres.

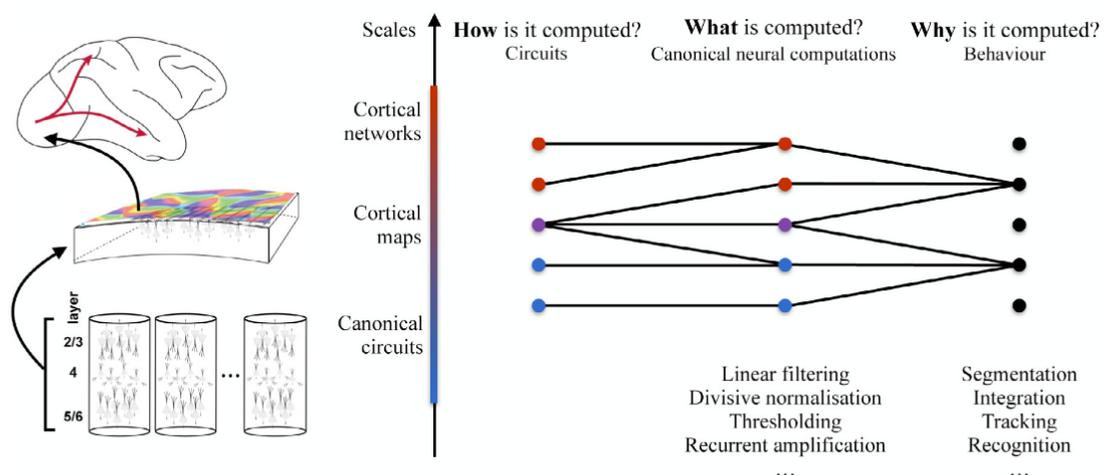


Figure 4. Analogies entre les réseaux de neurones cérébraux et informatiques par Medathati et al <sup>(38)</sup>

Prenons l'exemple de la visualisation d'un chat. Dans le cerveau, les informations perçues par les yeux : contours, couleurs, textures, sont captées par les neurones superficiels, puis ces informations sont encodées en signaux électriques propagés aux neurones des couches intermédiaires. Ces neurones intermédiaires traitent toutes ces informations successivement, transformant les données "brutes" en données plus abstraites ou *patterns* (oreilles pointues, moustaches, yeux, queue...) jusqu'à ce que ces signaux arrivent aux neurones de la couche profonde et soient traduits en une donnée sémantique : "c'est un chat".

Le réseau de neurones artificiels est une adaptation mathématique et informatique de ce processus. Il comporte donc de multiples fonctions mathématiques, ou neurones artificiels, organisés en couches :

- la couche d'entrée qui extrait les informations de l'image, ou *features* ;
- des couches intermédiaires multiples qui comportent des neurones interconnectés, qui traitent les informations ;
- une couche de sortie qui aboutit à l'information finale.

Le premier neurone artificiel a été créé à la fin des années 1950 par Rosenblatt <sup>(39)</sup>, nommé « perceptron ». Il s'agit d'une fonction mathématique qui traite un vecteur de données d'entrée ( $w_x$ ) pour aboutir à une donnée de sortie ( $y$ ).

Les premiers réseaux neuronaux utilisant des perceptrons multicouches ont été développés au début des années 1960. Ils comportaient des neurones tous connectés les uns aux autres (*fully connected*) pour permettre de traiter des données complexes, notamment des images. Cependant ils n'étaient pas optimaux pour l'analyse d'images car celles-ci sont beaucoup trop riches en informations : en effet, le nombre de connexions entre neurones augmente de façon

exponentielle avec la taille de l'image, ce qui implique un nombre beaucoup trop important de calculs <sup>(40)</sup>. De plus, ces réseaux nécessitent d'être supervisés dans leur apprentissage, c'est-à-dire que c'est l'humain qui doit fournir à l'algorithme les caractéristiques discriminantes à rechercher pour chaque catégorie d'images, via des annotations sur celles-ci. Cela implique un temps alloué à l'annotation conséquent, le design d'extracteurs de *features* par l'ingénieur (ce qui est très technique et chronophage) et une restriction à des *features* connues.

Pour pallier ce problème, des *data scientists* ont cherché à développer une méthode qui pourrait apprendre automatiquement quelles sont les *features* d'intérêt dans une image (*representation learning*) : c'est l'ambition des algorithmes de *deep learning*. Cette approche est basée sur la combinaison de multiples représentations à un niveau d'abstraction croissant : par exemple, le premier niveau d'analyse sera la définition des bordures de l'image ; le deuxième, la recherche de contours dans l'image ; et ainsi de suite jusqu'à la détection d'objets d'intérêt. Cette méthode permet l'apprentissage de *features* complexes <sup>(41)</sup>. Cet apprentissage successif n'est pas pré-établi par l'ingénieur mais modelé par les données présentées à l'algorithme, ce qui permet de simplifier le codage. Le corollaire est que ces algorithmes nécessitent une très grande quantité de données. Ce n'est qu'à la fin des années 1990 grâce au développement du *big data* et de processeurs graphiques puissants que le *deep learning* a connu une grande popularité, en étant autant ou plus performant que les autres types de *machine learning* (40).

En particulier, les réseaux de neurones convolutifs sont devenus incontournables. Ce type de réseau a connu une très grande popularité depuis la compétition *ImageNet Large Scale Visual Recognition Competition* en 2012. Ce challenge international consistait en le développement du meilleur algorithme pour la classification d'images naturelles (animaux, objets, paysages...). L'équipe gagnante a obtenu d'excellents résultats grâce à son réseau de neurones convolutif

AlexNet, qui est depuis utilisé par beaucoup d'équipes pour l'analyse d'images et a inspiré de nouveaux algorithmes <sup>(42)</sup>.

### 3. Réseau de neurones convolutifs

Un réseau de neurones convolutif (CNN) est un type d'algorithme de *deep learning* optimisé pour le traitement de signaux et notamment d'images. Le point fort du CNN est qu'il extrait, sans apprentissage préalable par un humain, les caractéristiques discriminantes dans les images pour prédire la classe, réduisant ainsi le nombre de calculs à réaliser. L'information contenue dans les pixels est ainsi traduite en *features*. Ce type de réseau a déjà fait ses preuves en médecine, de l'analyse de lésions cutanées <sup>(43)</sup> à la lecture de mammographies <sup>(44)</sup>.

Les CNN sont constitués de deux blocs : un premier bloc pour l'extraction de *features* nommé *convolutional backbone*, un deuxième bloc de neurones *fully connected* pour l'analyse des *features* extraites et la classification finale. Ainsi, ils limitent le nombre de connexions entre un neurone et les neurones des couches adjacentes, ce qui diminue drastiquement le nombre de paramètres à apprendre et rend l'analyse d'images volumineuses faisable.

Cette architecture est particulièrement efficace pour découvrir des structures complexes dans les données, car elle permet d'aboutir à un haut niveau d'abstraction tout en réduisant le nombre de calculs <sup>(41)</sup>.

Différents modèles de CNN ont été élaborés pour l'analyse d'images, chacun ayant ses spécificités. Nous abordons ci-dessous le principe global de l'analyse d'une image histologique par un CNN. Les spécificités des méthodes utilisées dans le projet sont détaillées dans le chapitre Matériel et Méthodes.

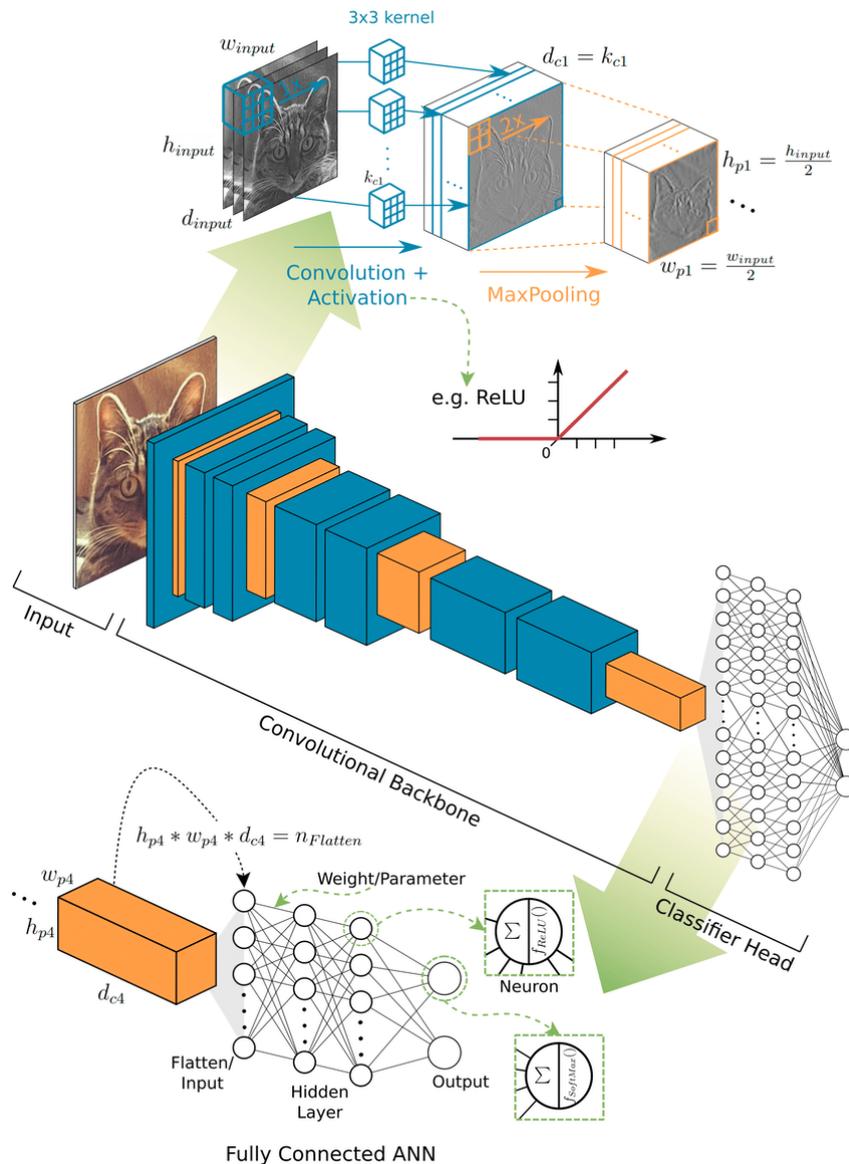


Figure 5. Structure globale d'un réseau de neurones convolutifs pour l'analyse d'images <sup>(45)</sup>

### 1.1 Extraction de features

L'extraction de *features* fait intervenir deux processus successifs : convolution et *pooling*. La donnée d'entrée est un objet mathématique plan (à deux dimensions).

La convolution consiste en l'analyse des pixels à travers des filtres (exemples de filtres : filtres de couleurs, de luminosité, de lignes ou de courbes, de textures...) et attribuer des valeurs aux pixels. Plus le pixel a activé l'extraction de la *feature* car il y eu une bonne convolution entre le filtre et le pixel, plus la valeur attribuée à celui-ci est élevée.

Chaque neurone analyse un champ récepteur de l'image. Les champs récepteurs se chevauchent (*stride*) afin d'obtenir une plus grande cohérence dans l'analyse.

Le *pooling* synthétise l'information obtenue par la convolution. Cette étape est nécessaire afin de permettre les calculs ultérieurs sur une quantité moins massive de données, tout en restant descriptive de l'image. Plusieurs méthodes de *pooling* sont possibles, par exemple le calcul de la moyenne des valeurs obtenues sur le champ récepteur (*receptive field*), la sélection de la valeur maximale... La donnée de sortie finale est un vecteur, donc un objet mathématique à une dimension.

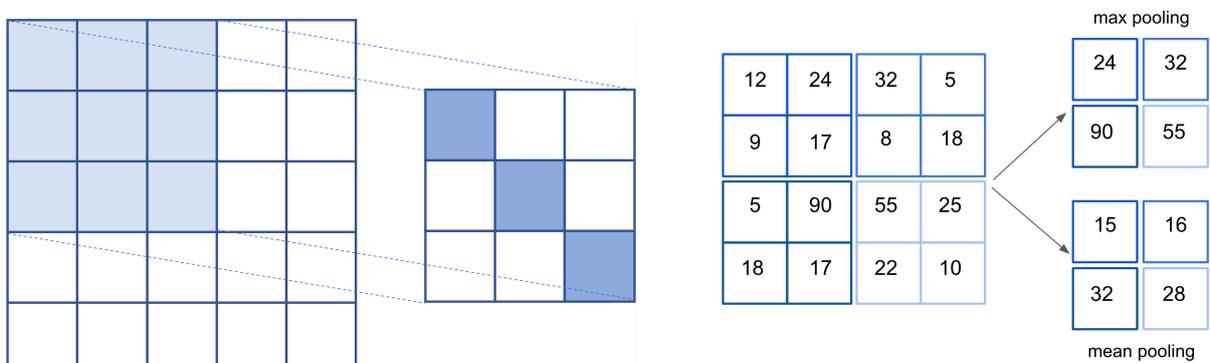


Figure 6. Convolution et pooling

Plusieurs couches de convolution et de *pooling* peuvent s'enchaîner, pour pouvoir extraire des *features* de plus en plus complexes. Par exemple, les *features* extraites par les premières couches sont des lignes, les couches suivantes combinent ces lignes pour en extraire des formes, et les dernières couches en extraient des objets complexes : visages, animaux...

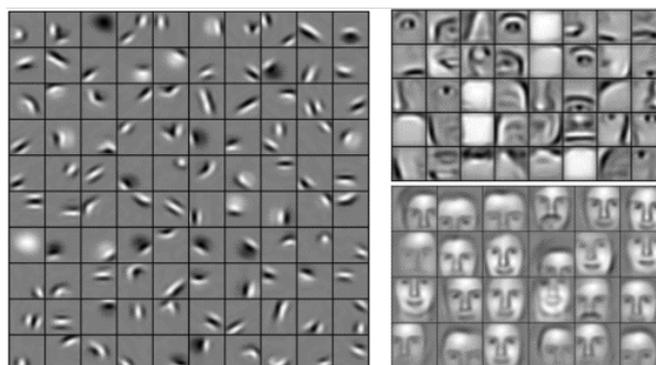


Figure 7. Niveaux d'abstraction, d'après Li et al <sup>(46)</sup>

## 1.2 Conclusion de l'analyse

La conclusion de l'analyse ou prédiction se fait à l'aide de couches classiques de neurones entièrement connectées (*fully-connected layers*). Ces couches sont organisées de la façon suivante : couche d'entrée, couches intermédiaires, couche de sortie.

Chaque neurone de la couche d'entrée reçoit le vecteur descriptif d'une partie de l'image. Des valeurs élevées vont activer la transmission d'un signal aux neurones de la couche suivante ; des valeurs faibles ou nulles vont induire la transmission d'un signal plus faible, voire nul, c'est-à-dire qui n'active pas le neurone suivant.

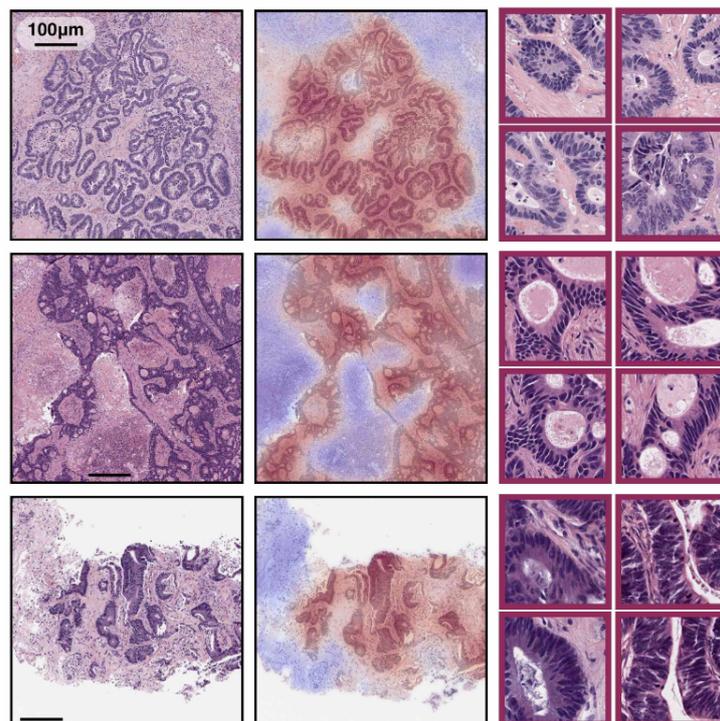
Les neurones des couches intermédiaires sont interconnectés. Ils captent les signaux émis par les neurones de la première couche et s'activent plus ou moins selon l'intensité des signaux d'entrée. Eux-mêmes émettent des signaux de sortie pour les neurones des couches suivantes, et ainsi de suite.

Enfin, les neurones de la couche de sortie reçoivent les signaux de la dernière couche intermédiaire, les traitent et émettent à leur tour un signal. Chaque neurone de cette couche correspond à une prédiction possible : par exemple classification en épithélium normal, dysplasie, ou cancer infiltrant. C'est le neurone de sortie ayant le signal le plus fort qui détermine la prédiction, donc le diagnostic final ou label.

Les couches intermédiaires multiples sont dites "couches cachées" ou *hidden layers*, car elles ne sont pas en contact avec "l'extérieur". La multiplicité des *features* et des calculs dans ces *hidden layers* font qu'il est très difficile pour l'observateur humain de comprendre pourquoi telle *feature* a été sélectionnée par l'algorithme comme plus discriminante que les autres.

C'est ce que l'on appelle le phénomène de boîte noire, ou *black box* : l'explicabilité des CNN est très difficile.

Pour pallier ce problème, des méthodes d'interprétation des modèles de *deep learning* existent. Il est ainsi possible, par exemple, de visualiser les parties de l'image ayant le plus contribué à la conclusion de l'algorithme. Cela permet d'interpréter les prédictions, à défaut de les expliquer mathématiquement.



**Figure 8. Cartographie des poids d'attention, d'après Lu et al <sup>(47)</sup>**

Par exemple, sur ces lames de métastase d'adénocarcinome colorectal, le masque rouge indique quelles zones de l'image ont contribué à la classification. Le fait que le carcinome soit surligné permet de dire que le modèle base sa prédiction sur des éléments tangibles et pas grâce au hasard ou à un facteur confondant.

Nous avons vu jusqu'à présent en quoi un algorithme utilisant un CNN peut imiter le cortex cérébral dans son organisation. Mais qu'en est-il de la capacité d'apprentissage ?

#### 4. Apprentissage de l'algorithme

Dans le CNN comme dans le cortex visuel humain, chaque neurone transmet des signaux avec une intensité variable. Mathématiquement, cette variabilité se traduit par un poids ( $w$  pour *weight*) attribué à chaque neurone. Cette modélisation cherche à imiter la plasticité des réseaux de neurones biologiques.

Un des points cruciaux d'un CNN est donc l'attribution des bons poids à ses neurones. C'est ce à quoi sert l'apprentissage : à force de voir des images et de gagner en expérience, le CNN ajuste les poids de chaque neurone jusqu'à obtention de paramètres optimaux. Cet apprentissage ou entraînement se fait sur un échantillon des images du jeu de données dédié, ou *training set*.

L'entraînement de l'algorithme consiste en la confrontation répétée entre la prédiction faite par celui-ci sur les images d'entraînement et la *ground truth* c'est-à-dire la bonne réponse. Chaque confrontation se traduit en le calcul d'un coût induit par la différence entre la valeur prédite par l'algorithme et la valeur vraie de l'image. Ce coût (*cost function / loss*) peut-être calculé de différentes manières (cf. Matériel et Méthodes). La *loss* est ensuite rétropropagée dans le CNN pour ajuster les poids des neurones à l'aide d'un algorithme de *backpropagation*. Ainsi, au fur et à mesure de son entraînement, le CNN est forcé à apprendre les *features* qui sont censées être pertinentes car elles minimisent le plus la perte.

L'apprentissage se fait en plusieurs étapes ou itérations, où l'algorithme analyse plusieurs images, ajuste ses paramètres internes selon ses performances puis analyse de nouvelles images. L'apprentissage est considéré comme optimal lorsque la *loss* est minimale.

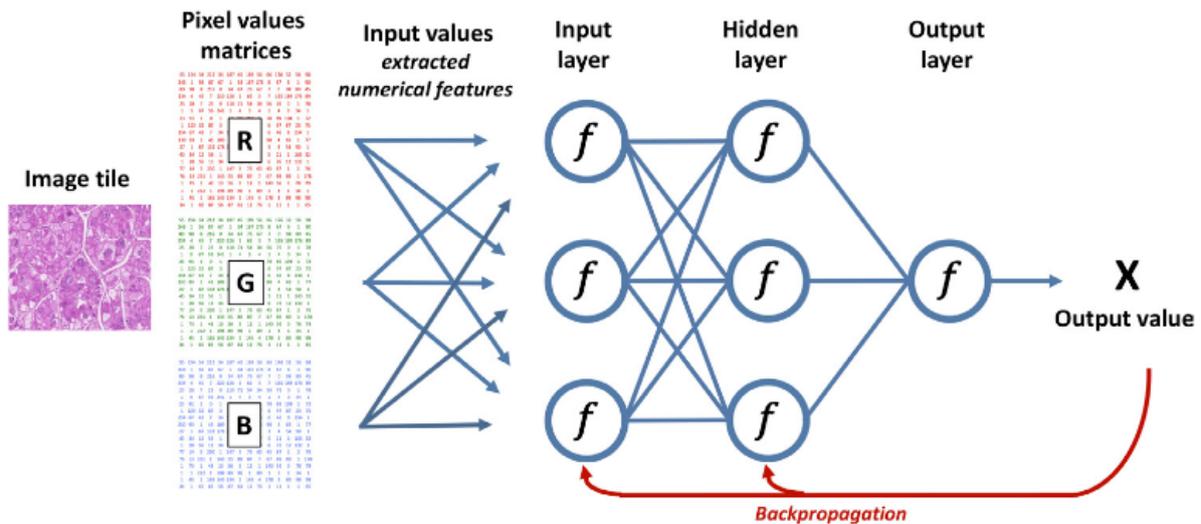


Figure 9. Principe de rétropropagation, d'après Calderaro et al<sup>(48)</sup>

Il est cependant important de ne pas faire un apprentissage excessif car à partir d'un moment, l'algorithme apprend à reconnaître les images par cœur au lieu d'utiliser judicieusement les *features* extraites. C'est ce qu'on appelle l'*overfitting*. L'algorithme ne sera pas capable de faire des prédictions exactes sur des images qu'il n'a jamais vues.

La figure suivante représente l'évolution de la *loss* après plusieurs itérations de l'apprentissage. L'*overfitting* est visualisé à droite : lorsque l'algorithme apprend par cœur les données du *training set*, la *loss* est quasiment nulle sur l'entraînement mais les performances réelles mesurées sur le *test set* sont médiocres.

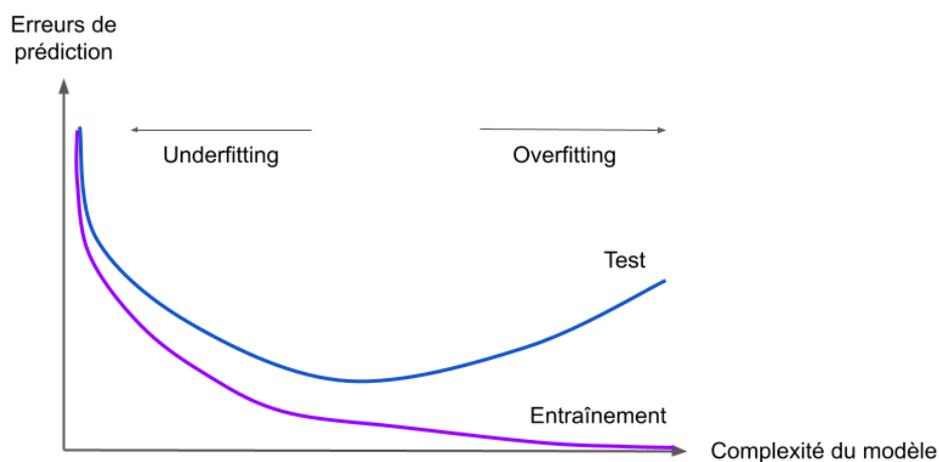


Figure 10. Conséquences du surentraînement sur les performances du modèle

## 5. Construction d'un jeu de données

La bonne construction du jeu de données est donc cruciale pour pouvoir avoir un bon entraînement. De plus, il est nécessaire de pouvoir vérifier les performances de l'algorithme sur des images qu'il n'a jamais vues, donc de réserver des images hors du jeu d'apprentissage.

Ainsi, un jeu de données se divise classiquement en trois parties :

- le *training set* ou set d'apprentissage : ce sont les images qui sont présentées à l'algorithme pour l'ajustement des poids des neurones (60-70% du total du jeu de données)
- le *validation set* qui permet de vérifier l'absence d'*overfitting* (10-20%)
- le *test set* qui permet d'évaluer les performances de l'algorithme (20%)

Il est indispensable de maintenir un équilibre entre les classes dans chaque set d'images. En effet, si le *training set* ne comporte que des carcinomes et que le *test set* ne comporte que des dysplasies, il sera impossible à l'algorithme de faire les bonnes prédictions. De plus, il est nécessaire de proposer à l'algorithme une variété d'images suffisante pour lui permettre d'extraire les *features* utiles et non de se baser sur de potentiels biais pour faire ses prédictions. Par exemple, si toutes les lames de dysplasie sont colorées par un automate et toutes les lames de cancer épidermoïde par un autre, ou qu'elles sont scannées par des scanners différents, alors l'algorithme va potentiellement apprendre les éléments différenciant ces images basées sur ces différences et non sur le diagnostic lui-même.

## 6. Tâches réalisables par un CNN – applications à l'image histologique

Plusieurs types de tâches sont réalisables avec un algorithme à CNN pour l'analyse d'images histologiques.

## 6.1 Classification

Les données d'entrée sont l'image histologique et la classe associée (le diagnostic) ; la donnée de sortie est la classification entre plusieurs possibilités diagnostiques, par exemple la classification entre types histologiques de carcinomes broncho-pulmonaires.

De nombreuses publications ont démontré la capacité des algorithmes de *deep learning* de classer des images selon le pronostic des patients ou la réponse attendue à un traitement<sup>(49)</sup> et même à prédire des mutations<sup>(50)</sup> et des signatures moléculaires<sup>(51)</sup> ; nous détaillerons les conséquences et perspectives de ces découvertes ci-après.

## 6.2 Régression

Les données d'entrée sont l'image histologique et une valeur continue, les données de sortie également. Par exemple, ce type d'algorithme peut aider à la stadification d'une tumeur en évaluant le degré d'infiltration en profondeur de la paroi de l'organe.

## 6.3 Segmentation

Les données d'entrée sont l'image histologique et une zone d'intérêt délimitée, la donnée de sortie est une zone délimitée sur l'image. Par exemple, ce type d'algorithme peut segmenter les zones tumorales infiltrantes par rapport au stroma.

## 6.4 Détection d'objets

Les données d'entrée correspondent à l'image histologique et la localisation des objets, les données de sortie sont les localisations des objets d'intérêt sur l'image. On peut par exemple détecter des cellules marquées par un anticorps, des cellules en mitose...

## 7. Utilisation de l'IA en anatomie pathologique : revue de la littérature

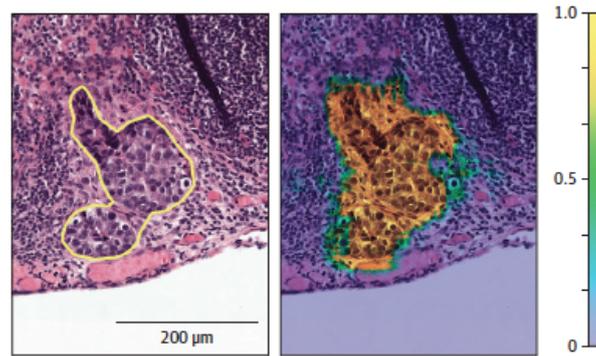
L'analyse d'image à l'aide d'algorithmes en anatomopathologie a pour ambition de répondre à de multiples défis qui se posent dans la discipline, notamment la prise en charge de prélèvements toujours plus nombreux du fait de la croissance de la population mondiale, avec un nombre de pathologistes disponibles restreint, ce d'autant plus dans les pays en développement.

De nombreuses publications ont montré le potentiel de l'IA en anatomopathologie, en voici quelques exemples.

### 7.1 L'algorithme comme assistant du pathologiste

L'une des applications principales des algorithmes d'IA est l'aide au diagnostic : des cas difficiles, pour la détection des lésions de petite taille, pour l'accélération du workflow. L'algorithme peut alors devenir un véritable assistant du pathologiste, comme un second lecteur.

La détection de métastases ganglionnaires de carcinome mammaire a fait l'objet d'une compétition internationale mettant en concurrence de très nombreuses équipes, en 2016 pour le Cancer Metastases in Lymph Nodes Challenge 2016 (Camelyon16) <sup>(52)</sup>. La figure suivante illustre un exemple de détection de métastase, avec à gauche l'image de la lame scannée et à droite la *heatmap* obtenue qui est bien superposée à la zone métastatique. L'algorithme gagnant, développé par une équipe de la Harvard Medical School et du Massachusetts Institute of Technology, avait une performance (AUC) de 0,994 meilleure que celle des pathologistes de l'étude (0,810).



**Figure 11. Exemple de heatmap de l'algorithme gagnant du challenge Camelyon16 <sup>(52)</sup>**

De nombreux autres algorithmes de détection de zones cancéreuses ont été développés dans d'autres organes : la prostate<sup>(53),(54)</sup>, le tube digestif<sup>(55)</sup>, le poumon<sup>(50),(56)</sup>, le foie<sup>(57)</sup> notamment.

A l'échelle cellulaire, le développement d'outils d'analyse automatisée des frottis cervico-utérins est en plein essor. L'utilisation de *CNN* permet de s'affranchir d'annotations fastidieuses et chronophages de cellules anormales tout en étant très performant<sup>(58)</sup>. Ce développement répond à une demande majeure, le nombre de frottis cervico-utérins étant très élevé et le nombre de pathologistes pouvant les analyser étant restreint, en particulier dans les pays en développement. Notamment, l'entreprise Microsoft développe un outil nommé Cervical Cancer Image Detection API, qui permet de différencier les frottis normaux des anormaux, en collaboration avec des pathologistes indiens. Cet outil permettrait de réaliser des diagnostics jusqu'à quatre fois plus vite qu'un pathologiste. La classification de lésions bénignes de diagnostic difficile est également facilitée par les algorithmes de *deep learning*, par exemple dans une étude classifiant les polypes coliques hyperplasiques des adénomes festonnés sessiles<sup>(59)</sup> ou des lésions gliales rares<sup>(60)</sup>.

Se pose alors la question suivante : ces algorithmes étant au moins aussi performants que le pathologiste voire plus, vont-ils remplacer les médecins ? Une première étude de Kiani et al<sup>(57)</sup>

évaluant les performances de l'algorithme associé à l'analyse médicale apporte un premier élément de réponse. En effet, les performances de la double lecture informatique et médicale surpassent celles de la lecture médicale seule comme l'analyse informatique seule. Le futur proche de l'emploi de ces outils va dans le sens d'une véritable "pathologie augmentée".

## 7.2 L'algorithme comme outil de recherche

En dehors de l'application en pratique courante de diagnostic, les algorithmes de *deep learning* sont également un outil de recherche prometteur pour la découverte de nouvelles caractéristiques histologiques. Plusieurs études ont démontré le pouvoir des algorithmes pour détecter des *features* liées aux mutations génétiques des cellules tumorales qui ne sont pas vues par les pathologistes. Notamment, l'étude de Coudray et al<sup>(50)</sup> a permis de classifier les images d'adénocarcinomes pulmonaires selon leur statut moléculaire uniquement avec la coloration HE. L'étude de Kather et al<sup>(61)</sup> a démontré l'existence de *features* liées au statut d'instabilité microsatellite sur cette seule coloration standard. L'étude de Naik et al<sup>(62)</sup> a permis de prédire la surexpression des récepteurs aux œstrogènes par les carcinomes mammaires sur la coloration HE. Enfin, le travail de Schmauch et al<sup>(51)</sup> a réussi à développer un algorithme classifiant les images en prédisant le profil transcriptionnel (RNASeq) de la tumeur.

Il est également possible de prédire le pronostic du patient avec l'image histologique. Par exemple, un travail de Saillard et al<sup>(63)</sup> a démontré cette compétence sur la prédiction de la survie pour les carcinomes hépatocellulaires.

Enfin, l'équipe de Mahmood et al<sup>(47)</sup> a développé un algorithme de diagnostic des carcinomes de primitif inconnu, qui arrive notamment à prédire l'origine des métastases de carcinomes épidermoïdes, chose en général impossible pour un pathologiste. Toutes ces études font la preuve qu'un algorithme peut détecter des *features* invisibles pour l'œil humain, grâce au fait

que l'apprentissage est modelé par les données elles-mêmes et non par des éléments connus des pathologistes et prédéfinis dans l'entraînement.

Ainsi, les algorithmes de classification peuvent donc aussi bien servir d'assistant pour le pathologiste en routine, comme d'outil de recherche à la découverte de nouvelles caractéristiques d'intérêt dans les images. En oncologie en particulier, ces outils contribuent et contribueront à une prise en charge toujours plus personnalisée des patients.

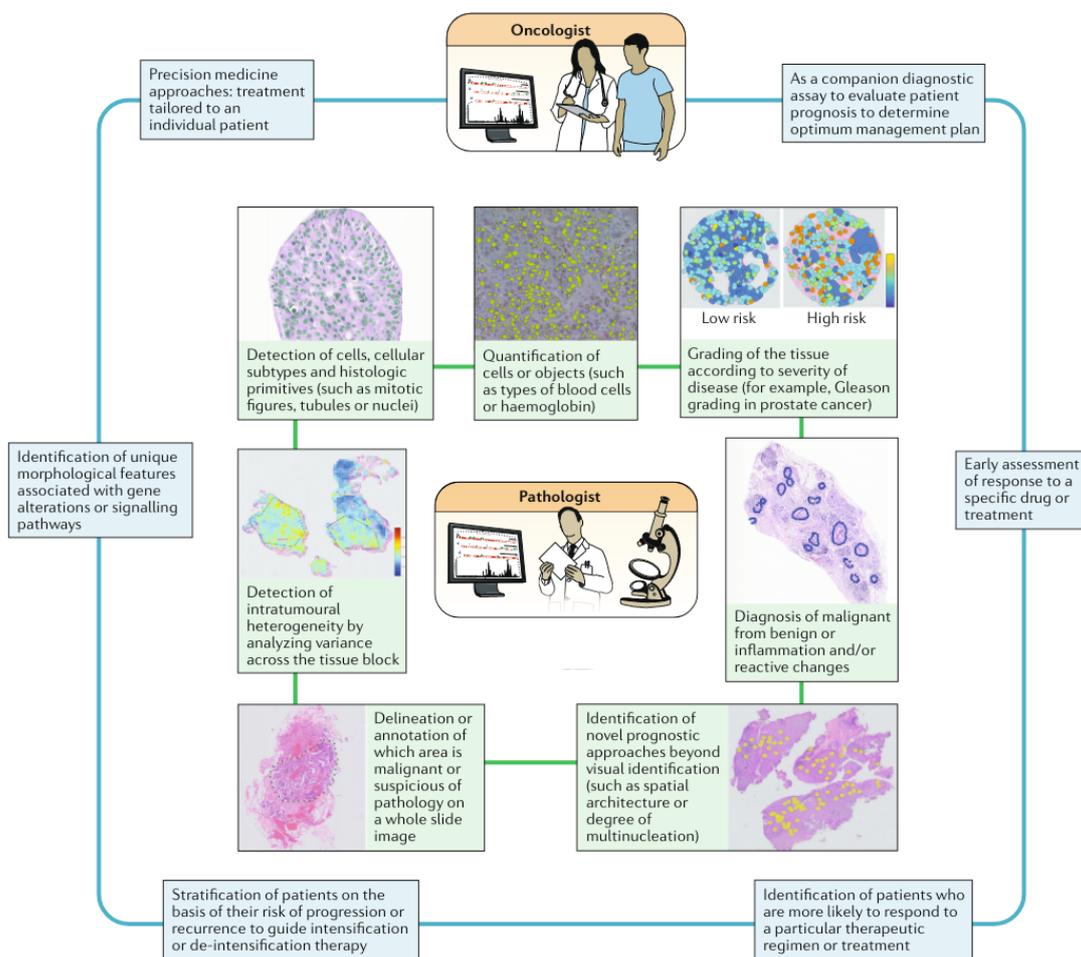


Figure 12. La pathologie augmentée au cœur de la prise en charge personnalisée des cancers, d'après Bera <sup>(64)</sup>

# Chapitre 4. Utilisation de l'IA dans le diagnostic des lésions malpighiennes des VADS : Etat de l'art, perspectives Objectif de la thèse

---

## 1. Utilisation de l'IA pour les lésions des VADS : état de l'art et perspectives

En comparaison aux autres organes, peu d'études ont concerné l'application de l'IA aux pathologies malpighiennes des VADS. Une méta-analyse réalisée par Mahmood et al<sup>(8)</sup> publiée fin 2020 recense douze articles traitant du sujet. Toutes ces études portent sur les lésions de la cavité orale ou de l'oropharynx. De plus, elles utilisent des méthodes d'IA peu modernes, notamment pas de *deep learning*. Ces méthodes ont pour avantage de nécessiter moins de données, mais sont beaucoup moins performantes.

Ainsi, il existe de nombreuses pistes à explorer pour créer de nouveaux outils d'IA appliqués à la pathologie malpighienne des VADS. Des algorithmes pourraient classer les lésions dysplasiques et infiltrantes selon la dernière classification de l'OMS avec une meilleure reproductibilité que les pathologistes. Notamment pour les dysplasies, un algorithme pourrait aider le pathologiste à prendre en compte de multiples paramètres en simultané (mitoses, apoptoses, atypies cytonucléaires, dyskératoses, architecture globale...). Ces algorithmes pourraient devenir un véritable outil compagnon du pathologiste, comme un deuxième lecteur, en affichant sous la forme de *heatmap* les zones dysplasiques, les zones micro-invasives... Cela permettrait d'avoir un second avis et de réduire le temps d'analyse en pointant les zones suspectes au pathologiste.

De plus, des algorithmes pourraient mettre en évidence dans les lésions dysplasiques des *features* prédictives de leur évolution en cancer. Cela pourrait permettre de grader ces lésions de

façon plus fine que celle proposée par les classifications actuelles. Les algorithmes arrivent à combiner l'analyse de multiples *features* plus aisément que le pathologiste humain ; ainsi, ils ont le potentiel de donner une gradation plus fine des lésions. Plus encore, ils arrivent à extraire des *features* invisibles ou difficilement décelables pour l'œil humain, comme l'ont prouvé les études prédictives du statut moléculaire sur la seule coloration HES. Étant donné que l'épithélium semblant normal pour le pathologiste peut parfois évoluer en carcinome <sup>(15)</sup>, il serait envisageable de pouvoir détecter de nouvelles *features* prédictives de l'évolution avec des algorithmes.

## **2. Objectif de la thèse**

L'objectif de ce travail est l'application de modèles d'IA pour le diagnostic automatisé des lésions malpighiennes des VADS selon la classification OMS 2017. Il utilise les dernières méthodologies d'IA proposées dans la littérature dans d'autres domaines de la pathologie, et s'appuie sur le dialogue entre le pathologiste le *data scientist* pour proposer un modèle plus original.

## Matériel et Méthodes

---

# Chapitre 1. Constitution du projet PREINV

---

Le travail exposé dans cette thèse constitue une partie du projet PREINV, fruit d'une collaboration de recherche entre l'Assistance Publique – Hôpitaux de Paris et Keen Eye Technologies (KE), une start-up française spécialisée dans l'analyse d'images médicales par algorithmes d'intelligence artificielle. Le projet de recherche s'intitule "Classification des lésions malpighiennes dysplasiques et invasives des voies aérodigestives supérieures : développement d'un algorithme d'analyse d'image automatisée par deep learning pour la classification automatisée des lésions et la prédiction de l'évolution des dysplasies".

Dans ce cadre, j'ai étroitement travaillé avec Mélanie Lubrano di Scandalea (ML), ingénieure spécialisée en intelligence artificielle et doctorante dans le cadre d'une thèse Cifre supportée par Keen Eye et encadrée sur le plan académique par le Pr Thomas Walter (TW). Ainsi, le projet PREINV bénéficie d'une double expertise médicale et scientifique.

## 1. Cadre juridique – protection des données

Ce projet a fait l'objet d'un contrat entre l'AP-HP et KE, encadrant chaque partie et notamment la gestion des lames virtuelles sur la plateforme. Les aspects juridiques et la protection des données côté A-HP ont été traités par la Direction de la Recherche Clinique et de l'Innovation (DRCI), l'Unité de Recherche Clinique (URC), le Département de Protection des Données (DPO) et le service juridique de l'Hôpital Européen Georges Pompidou (HEGP).

Il s'agit d'une recherche médicale hors loi Jardé n'impliquant pas la personne humaine. Elle est basée sur la réutilisation de données recueillies dans le cadre du soin, dont le traitement est encadré par la méthodologie de référence MR004.

Dans le respect de la MR004, tous les patients ont reçu une information générale durant leur prise en charge médicale et les patients non décédés au moment de l'étude ont également été informés par une note d'information explicitant le projet, les données utilisées et la possibilité de refuser de participer, avec un délai de trois semaines pour y répondre. La note d'information a été validée par l'Unité de Recherche Clinique de l'Hôpital Européen Georges Pompidou. Au total, 398 courriers ont été envoyés et 7 refus ont été reçus et donc exclus de l'étude (dont un refus de l'épouse d'un patient décédé).

De plus, toutes les données médicales ont été dé-identifiées et la table de correspondance sauvegardée à l'HEGP. Chaque lame et chaque patient ont eu l'attribution d'un numéro ne permettant pas de remonter à son identité ni au numéro de prélèvement. Enfin, un résumé du projet a été déposé sur un registre public hébergé sur le site du Health Data Hub.

Le projet PREINV a obtenu l'avis favorable du Comité de Protection des Personnes (CPP) le Sud-Est VI le 10 mars 2021 et du Comité d'Ethique pour la Recherche CERAPHP-Centre le 19 novembre 2021.

## **2. Constitution de la base de données**

Les critères d'inclusion sont les suivants :

- patient âgé d'au moins 18 ans ;
- patient diagnostiqué avec une lésion malpighienne dysplasique (bas grade, grade intermédiaire, haut grade/carcinome *in situ*) et/ou un carcinome épidermoïde infiltrant (micro-invasif, bien différencié, moyennement différencié, peu différencié) selon la classification OMS 2017 des lésions dysplasiques et invasives des VADS ;

- lésions localisées au niveau du larynx, du pharynx, du nasopharynx, de l'hypopharynx, du bord latéral de la langue mobile ou de la cavité buccale ;
- patient ayant été suivi pendant au minimum 7 ans, ce qui correspond à des prélèvements réalisés entre début 2000 et fin 2013.

Les critères d'exclusion sont :

- la localisation à l'oropharynx (car pas de lésion dysplasique reconnue dans cette localisation anatomique, physiopathologie différente) ;
- la mention dans le compte-rendu de l'évaluation impossible d'une éventuelle infiltration (du fait d'une inclusion tangentielle ou d'un prélèvement insuffisant) ;
- la mention dans le compte-rendu ou le dossier patient d'une chimiothérapie néo-adjuvante dans le cas des lésions invasives car le traitement modifie l'aspect de la tumeur (hors sujet pour notre étude puisque le diagnostic de cancer est déjà connu pour le prélèvement).

Suivant ces critères a été réalisée une requête de la base de données du logiciel Diamic pour les cas diagnostiqués entre janvier 2000 et décembre 2013, sur les codes ADICAP des régions anatomiques d'intérêt et des lésions d'intérêt. La codification ADICAP consiste en la classification nationale des lésions en anatomie cytologie pathologiques, utilisée pour tous les diagnostics. L'extraction des données a permis l'obtention d'un tableau Excel avec les codes ADICAP renseignés par le pathologiste lors de la validation du compte-rendu, comportant : les codes de mode de prélèvement (B, P, O), d'organe (AC, AH, AL, AP, BL, BX) et de lésions (voir Tableau 4 ci-dessous).

L'extraction des données de Diamic a dû être faite en plusieurs étapes pour s'adapter aux performances des ordinateurs disponibles à l'hôpital. La première extraction faite de 2000 à

fin 2002 a abouti à une surreprésentation des carcinomes et des dysplasies sévères-carcinomes *in situ*. Afin de garder un équilibre entre les classes, tous les cas de ces lésions n'ont pas été ajoutés à la base de données sur les extractions de 2003 à fin 2013.

Tous les comptes-rendus histologiques ont été relus pour vérification des codes ADICAP renseignés par le pathologiste (ces codes étant renseignés manuellement, des erreurs sont possibles) et l'ajout des ADICAP non renseignés de lésions pourtant mentionnées dans le texte, notamment : ajout d'une dysplasie mentionnée dans le compte-rendu d'un carcinome invasif mais non codée par le pathologiste, degré de différenciation du carcinome qui a été en général codé avec un code générique E7T0, présence de carcinome micro-invasif...

Il a également été ajouté un code pour l'épithélium normal quand il était présent, car la comparaison de l'épithélium lésionnel à l'épithélium normal est indispensable pour l'apprentissage de l'algorithme.

Durant cette lecture, il a été identifié un certain nombre de comptes-rendus où les lésions dysplasiques ont été classées en dysplasie légère à modérée, ce diagnostic étant tout à fait possible de 2000 à 2013 étant donné que la nouvelle classification de l'OMS a été établie en 2017. Au total 127 lames ont été listées pour reclassification par pathologiste expert (CB).

Le logiciel DxCare a ensuite été consulté pour le recueil des données cliniques d'intérêt : statut vital au moment de l'étude, décès lié au cancer ou non, évolution en cancer qui ne serait pas mentionnée dans Diamic, etc. Un patient a été considéré perdu de vue si les derniers éléments de son dossier DxCare datent d'avant 2015 ; à partir de 2015, il a été considéré comme vivant. Un fichier anonymisé a été créé pour pouvoir partager les données avec Keen Eye par la suite. Un identifiant numérique (de 1 à 610) a été attribué à chaque patient, permettant de savoir quelles lames correspondent à quel patient, donnée indispensable pour le *split* des données

tout en anonymisant l'identité patient. En effet, les lames d'un même patient ne doivent pas être présentes à la fois dans le jeu d'apprentissage, de validation ou de test pour éviter un biais de reconnaissance du patient par l'algorithme.

### **3. Sélection des lames**

Toutes les lames des cas identifiés dans Diamic ont été désarchivées : au total, près de 3000 lames. Un tri de ces lames était indispensable car en général plusieurs niveaux de coupe étaient disponibles pour les biopsies et plusieurs prélèvements de tumeur pour les pièces opératoires. De plus, il fallait retrouver les lames lésionnelles dans les cas de pièces opératoires, où de nombreux prélèvements sont réalisés et pour lesquels le pathologiste ne mentionne pas toujours dans son compte-rendu quelle lésion est présente sur quelle lame.

Pour les cas où plusieurs niveaux de coupe étaient disponibles, tous les niveaux ont été revus par YB afin de sélectionner la lame la plus représentative de la lésion et exclure les niveaux ayant trop d'artéfacts. Cette relecture a également permis d'exclure des prélèvements sans tumeur ou sans épithélium (hors sujet). Une seule lame par prélèvement biopsique a été conservée, pour ne pas créer de biais avec plusieurs fois la même image dans le jeu de données.

Afin de garder un équilibre entre les classes, seules quelques lames de tumeurs issues de pièces opératoires ont été sélectionnées, car prendre toutes les lames tumorales aurait largement surreprésenté les échantillons de carcinome invasif ainsi que les coupes avec une grande surface tissulaire par-rapport aux biopsies.

#### 4. Numérisation des lames et anonymisation

La première étape pour pouvoir analyser l'image est de la rendre "lisible" par l'algorithme, c'est-à-dire de la numériser.

L'image perçue par l'œil humain est une image analogique, c'est-à-dire liée à un support matériel. Dans le cadre du travail de ma thèse, il s'agit de la coupe tissulaire colorée sur lame de verre. Elle contient une quantité d'informations infinie, le cerveau humain lui-même ne parvient pas à voir séparément chaque point coloré de l'image. La numérisation consiste à transformer ces données continues en données discrètes, en découpant l'image en pixels. Chaque pixel possède des valeurs chiffrées qui le décrivent sur plusieurs paramètres : sa couleur, sa luminosité, sa texture, son intensité, etc. L'image devient donc un objet mathématique à deux dimensions.

Les lames ont été numérisées par un scanner Hamamatsu Nanozoomer® 360, à la résolution de 20X, au format .ndpi. Les détails de la numérisation sont précisés en annexe.

Il a été attribué à chaque lame un identifiant (pseudo\_ID lame : preinv\_x (x = 1,..., n)) avec x un nombre sans aucun lien avec le numéro de la lame et sans retour possible vers celui-ci, ni avec le pseudo\_ID du patient précédemment établi.

Après la numérisation, la métadonnée de l'étiquette de la lame a été supprimée de façon définitive, rendant les lames histologiques tout à fait anonymisées.

Enfin, les lames ont été téléchargées sur la plateforme sécurisée de Keen Eye. Cette plateforme permet de visualiser les lames, de les annoter, de leur attribuer des labels globaux et de les utiliser pour faire tourner des algorithmes, visualiser des cartographies (*heatmaps*).

Les données partagées avec Keen Eye sont les suivantes : lames numérisées, pseudo\_ID du patient et de la lame, code ADICAP (qui comporte les informations suivantes : type de prélèvement, localisation anatomique, diagnostic histologique).

## 5. Contrôle de la qualité et relecture des lames numérisées

Une fois sur la plateforme, toutes les lames du jeu de données ont été revues par YB pour vérification des labels et contrôle de la qualité des lames qui n'avaient pas encore été relues.

Au total, 168 lames ont été exclues après la numérisation pour des raisons de mauvaise qualité (problème de coloration, artéfacts d'électrocoagulation trop délétères pour l'analyse) ou d'absence d'épithélium pour certains prélèvements codés normaux dans Diamic.

Cette relecture a également permis d'ajouter des codes de lésions dysplasiques et d'épithélium normal qui n'étaient pas mentionnés dans les comptes-rendus des cas qui n'avaient pas encore été relus par YB.

Une fois les lames numérisées, les dysplasies classées en "légère à modérée" ont été relues par le binôme interne (YB) - pathologiste expert (CB) ; soit au total 127 lames. Des cas où il existait un doute entre le diagnostic de carcinome *in situ* et de micro-invasion, listés par YB lors de sa relecture, ont également été relus ; il s'agissait de 23 lames.

Les lames ont été labellisées selon les codes ADICAP de toutes les lésions présentes, comme précisé dans le **Tableau 4**. La granularité entre les dysplasies modérées et sévères a été conservé dans la base de données.

Label	Code ADICAP lésion
Épithélium normal	0000
Dysplasie de bas grade	E5S1, 6811, 6812
Dysplasie modérée	E5S2, 6813
Dysplasie sévère	E5S3, E5T0, 6814
Carcinome invasif	E6T0, E7T1, E7T2, E7T3, E7T4

**Tableau 4. Codes des lésions selon le thésaurus ADICAP**

## Chapitre 2. Préparation des données pour l'algorithme de classification des lésions

---

### 1. Attribution d'un label global et *split* des données

Nous avons attribué comme label global à chaque lame la classe de la lésion la plus grave présente sur la lame. Ce label global est par la suite considéré comme la vérité de terrain (*ground truth*).

Label global	Lésion(s)
Classe 0	Épithélium normal
Classe 1	Dysplasie de bas grade
Classe 2	Dysplasie de haut grade (modérée ou sévère)
Classe 3	Carcinome invasif

**Tableau 5. Classes diagnostiques basées sur la classification OMS 2017**

Une séparation du jeu de données d'entraînement (*split*) a été ensuite effectuée pour :

- ne pas retrouver les lames d'un même patient dans les jeux d'entraînement, de validation et de test ;
- garder un équilibre entre les classes dans ces trois parties ;
- permettre une validation croisée en 5 plis (*cross validation 5-folds*) : chaque lame aura été vue par l'algorithme pour le test.

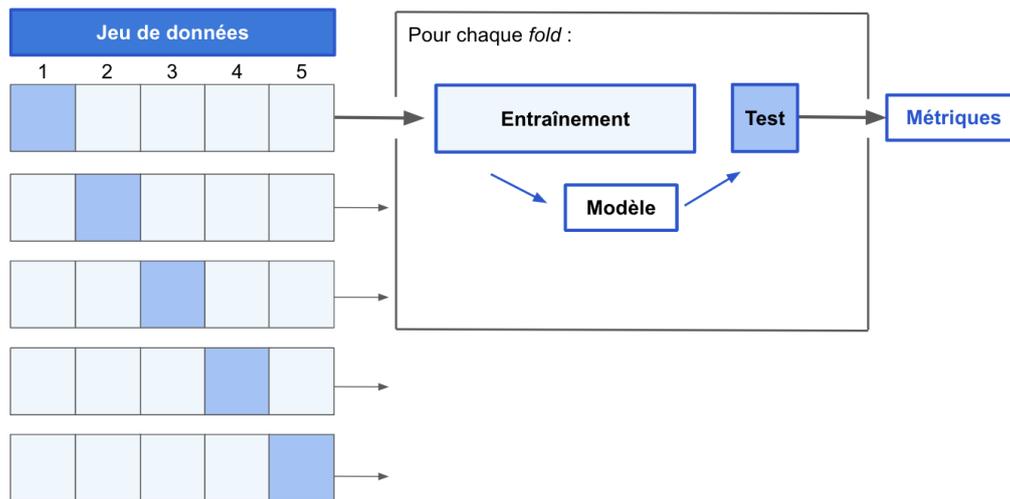


Figure 13. Schéma de la cross-validation

## 2. Prétraitement des images lame entière

Afin d'optimiser les entraînements d'algorithmes et de faciliter l'analyse des lames, il est nécessaire de retirer un maximum de zones vides de tissu dans l'image. Pour ce faire, nous calculons l'écart-type entre les valeurs des canaux RGB de la lame, sur chaque pixel (la lame étant réduite à une taille minimale). Nous obtenons une carte des écarts-types et appliquons la méthode d'Otsu sur celle-ci <sup>(65)</sup> qui permet de distinguer le premier plan (tissu) et l'arrière-plan (vide blanc autour, poussières, traces de feutre mal effacées...).

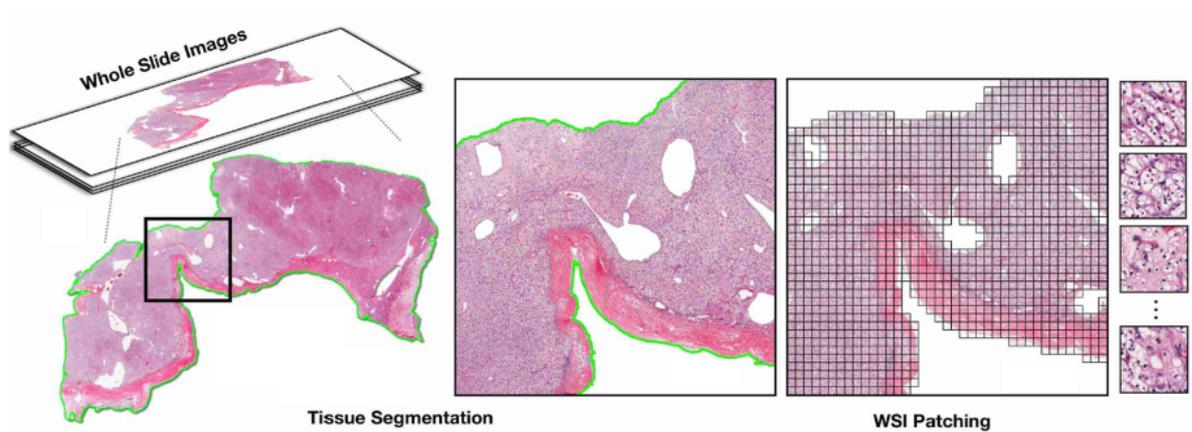
## 3. Tuilage

Une image numérisée avec une haute résolution est très lourde en données, parfois jusqu'à plusieurs gigabits (GB) et peut contenir des centaines de milliers de pixels en largeur et hauteur. Même les processeurs les plus puissants ne sont pas capables d'analyser tous les pixels d'une telle image d'une traite. Cela est valable en particulier pour les lames histologiques numérisées. La lame scannée est découpée en tuiles (*tiles*) de taille identique comportant un nombre fixe de pixels, dont les dimensions sont suffisamment petites pour être traitées par le

réseau. Chaque tuile est analysée puis les résultats obtenus pour toutes les tuiles sont synthétisés (plusieurs méthodes existent, celles que nous avons utilisées sont explicitées dans le chapitre Matériel et Méthodes).

Le choix de la taille des tuiles a un impact sur la façon dont l’algorithme va “voir” l’image. Si les tuiles sont de grande taille, elles vont recouvrir une large partie de l’image. Pour les lames histologiques numérisées, cela correspond à une visualisation au microscope au faible grossissement, avec la visualisation de l’architecture. Au contraire, si les tuiles sont de petite taille, elles permettent une analyse de l’image plus dans les détails, à l’échelle cellulaire, comme l’analyse par le pathologiste au fort grossissement.

Après avoir testé plusieurs tailles de tuiles, nous avons choisi un tuilage à 10X c’est-à-dire des tuiles de dimension 224x224 pixels à la résolution de 1 mpp.



**Figure 14. Tuilage, d’après Lu et al <sup>(66)</sup>**

## Chapitre 3. Architectures et configurations testées

---

### 1. Initialisation des poids de l'extracteur de *features*

L'extracteur de *features* choisi est DenseNet 121<sup>(67)</sup>. Il s'agit d'un extracteur dont les poids sont gelés durant l'entraînement du modèle, car son paramétrage directement durant l'entraînement du modèle sur le jeu de données serait trop lourd. Son utilisation nécessite donc l'emploi de poids prédéfinis. Nous avons comparé deux méthodes de pré-entraînement pour obtenir ces poids.

La première méthode a été de pré-entraîner l'extracteur sur ImageNet, qui est une banque d'images publique contenant plusieurs millions d'images naturelles (paysages, animaux...). Cette méthode est connue comme étant plus performante que l'initialisation aléatoire du réseau<sup>(68)</sup>. L'avantage de cette méthode est que l'énorme quantité d'images et leur variété permet d'extraire des *features* très génériques. L'inconvénient est qu'il ne s'agit pas d'images histologiques, les *features* extraites ne sont donc pas complètement adaptées au type d'image de notre étude.

La deuxième méthode a été le *self-supervised learning (SSL)* avec le module SimCLR. Il s'agit d'un pré-entraînement sur 1 million de tuiles non annotées extraites du jeu de données, modifiées par *data augmentation* (réduction de la taille de l'image, rotation, *flipping*, modification des couleurs...). Chaque image est transformée en deux nouvelles images grâce à une *data augmentation* aléatoire. L'objectif est de maximiser l'accord entre les scores de classification obtenus pour ces deux images. Cette méthode a été évaluée comme supérieure au pré-entraînement sur ImageNet par certains auteurs<sup>(69)</sup> mais elle est beaucoup plus gourmande en ressources de calculs.

## 2. Architectures testées

### 2.1 État de l'art

La première étape a été de tester des modèles de référence dans la classification de lames histologiques, afin d'évaluer les performances de ceux-ci sur le jeu de données et d'établir des métriques de référence.

La structure globale utilisée est la suivante :

- 1<sup>ère</sup> étape : extraction des *features* par le *convolutional backbone (CNN)* DenseNet 121 : ce CNN est appliqué à chaque tuile, effectue plusieurs étapes de convolution et *Pooling* pour aboutir à un vecteur de 1024 valeurs pour chaque tuile ;
- 2<sup>ème</sup> étape : synthèse du vecteur par un réducteur de valeurs (*fully connected reducer*) qui aboutit à un vecteur de 64 valeurs pour chaque tuile (le choix de cette valeur est empirique) ;
- 3<sup>ème</sup> étape : synthèse de toutes les tuiles pour obtenir un unique vecteur de 64 valeurs résumant toute la lame. Cette synthèse a été effectuée par plusieurs descripteurs de tuiles que nous avons comparés : Mean Pooling, Max Pooling, Chowder, Attention Pooling et CLAM ;
- 4<sup>ème</sup> étape : synthèse du vecteur de 64 valeurs par le classifieur qui aboutit finalement à un vecteur à 4 dimensions (une dimension par classe).

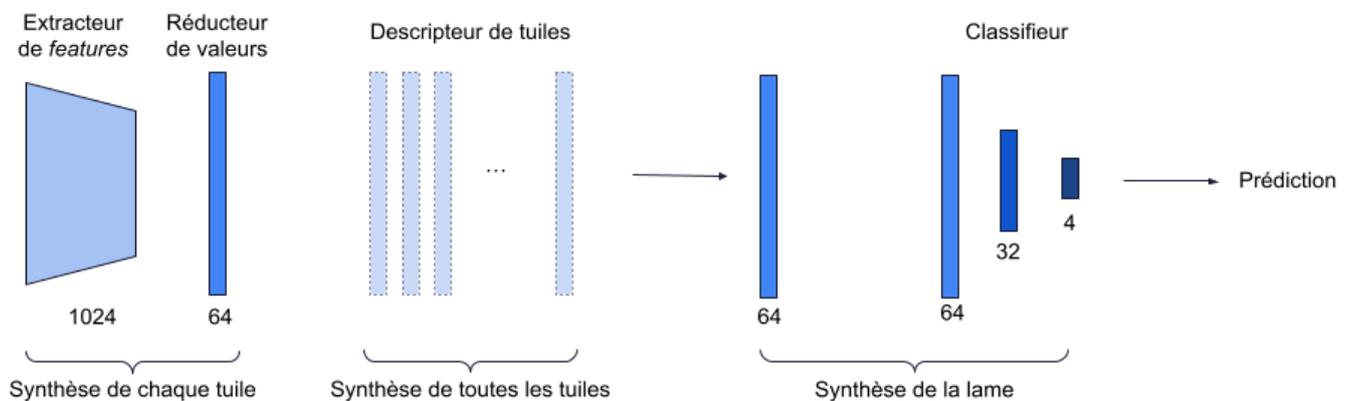


Figure 15. Principe général des modèles testés

### **Modèles non paramétrables : Mean et Max Pooling**

Les méthodes de *Pooling* correspondent à une synthèse des tuiles reposant uniquement sur leur valeur suite à la convolution, sans mise en jeu d'autres paramètres.

Dans notre modèle, chaque tuile est synthétisée en un vecteur de 64 valeurs ou dimensions. Le *Pooling* se fait sur chaque dimension des tuiles : avec le *Max Pooling*, pour la première dimension, c'est la valeur maximale retrouvée parmi toutes les tuiles qui est gardée, et ainsi de suite jusqu'à la 64ème dimension. Avec le *Mean Pooling* c'est la valeur moyenne.

### **Modèles entraînés par apprentissage profond : Chowder, Attention, CLAM**

Contrairement aux modèles de *Pooling* qui sont non paramétriques, ces modèles sont entraînés à reconnaître et sélectionner les tuiles les plus importantes pour effectuer la bonne classification. Les fonctions mathématiques (ou neurones) contenues dans ces modèles sont pondérées par des poids correspondant à l'importance de tel ou tel élément contenu dans chaque tuile ; l'entraînement du modèle permet d'ajuster ces poids par rétro-propagation du gradient pour qu'ils soient les plus précis possibles.

Ces méthodes permettent ainsi une approche holistique de la lame grâce à une analyse dynamique des tuiles. Elles permettent de s'affranchir des annotations manuelles et sont très efficaces pour la classification sur label global. De plus, il est possible de générer la cartographie des tuiles à l'origine des plus forts scores d'attention superposées à l'image de la lame histologique (*heatmap*).

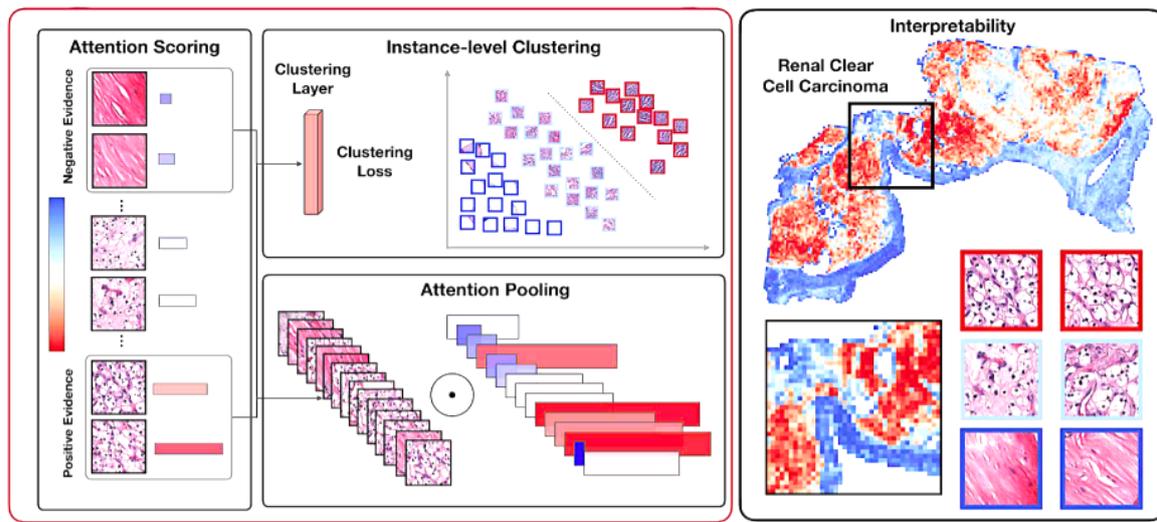


Figure 16. Construction des heatmaps, d'après Lu et al<sup>(47)</sup>

Le modèle **Chowder** a été proposé par Courtiol et al. en 2020<sup>(56)</sup>. Le module de sélection des tuiles est nommé *MinMax layer*. Il permet de sélectionner pour le classifieur les tuiles les plus importantes pour choisir la classe ainsi que les exemples "négatifs" pour la classe. Cette méthode a été développée pour la classification correcte des lésions de petite taille, qui pourraient être sous-estimées par un calcul de la valeur moyenne des tuiles comme dans le *Mean Pooling*. Elle a fait ses preuves sur une problématique binaire (présence ou absence de lésion) sur des lames de métastases ganglionnaires de carcinomes mammaires du jeu de données de Camelyon16.

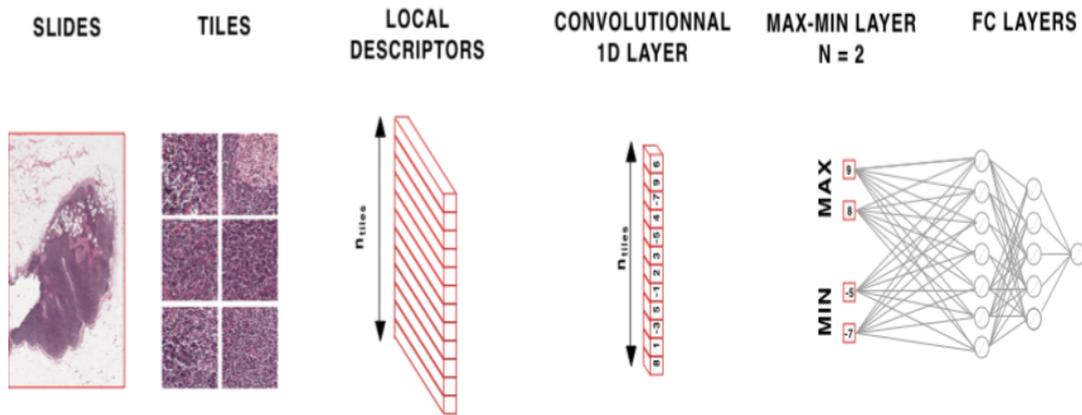


Figure 17. Modèle Chowder, figure adaptée de Courtiol et al

Le modèle **Attention Pooling** a été proposé par Ilse et al en 2018 <sup>(70)</sup>. Le principe est que le modèle ajuste le poids de ses neurones grâce à un élément supplémentaire dans le module de prédiction, nommé branche d'attention. Cette branche permet de donner plus de poids dans le choix de classe aux tuiles qui ont beaucoup contribué à la classification correcte, avec une *loss* minimale ; et inversement de diminuer l'importance des tuiles non pertinentes. Ce modèle a été testé et validé pour la détection de zones cancéreuses sur une lame, sur plusieurs types de carcinomes.

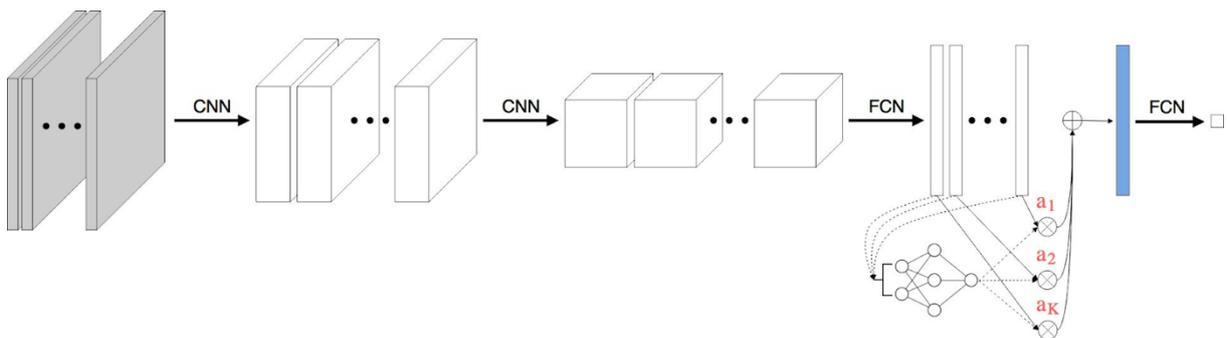


Figure 18. Modèle Attention Pooling, figure adaptée de Ilse et al

Le modèle CLAM a été proposé par Lu et al en 2020 <sup>(66)</sup>. Il repose également sur l'Attention Pooling mais il intègre une branche d'attention par classe diagnostique possible. Cette méthode s'adapte ainsi aux problématiques multiclassées (classifications non binaires, c'est-à-dire qu'il y a plus de deux classes possibles).

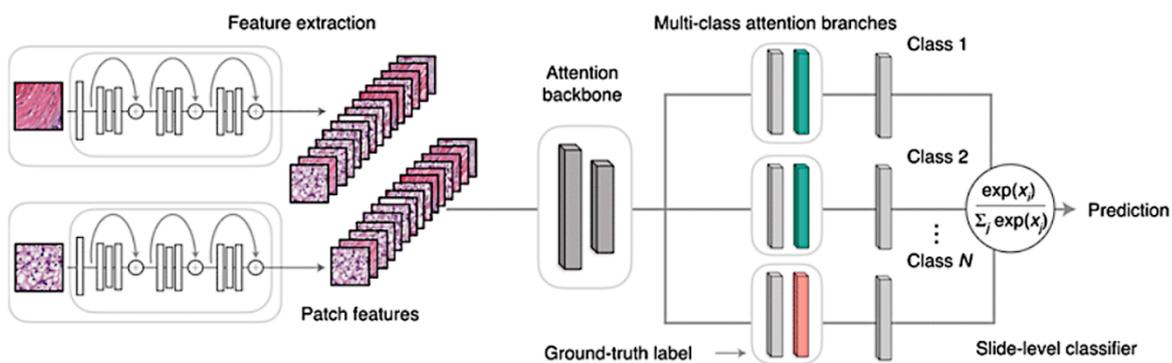


Figure 19. Modèle CLAM, figure adaptée de Lu et al

## 2.2 Nouvelle approche : le modèle de *Spatial Attention*

Les échanges entre pathologiste et *data scientist* ont permis de proposer un modèle original vis-à-vis des travaux déjà publiés sur l'analyse de lames histologiques virtuelles. Ce modèle a pour objectif de répondre aux limites des modèles classiques de *multiple instance learning*.

### **Limites des modèles de *multiple instance learning* classiques**

L'une des principales limites de ces derniers est qu'ils ne prennent pas en compte le contexte dans l'image. En effet, chaque lame est considérée comme un sac de tuiles indépendantes les unes des autres : chaque tuile est analysée individuellement par le module d'attention. Or cette façon de voir l'image est à l'exact opposé de celle du pathologiste, qui combine tous les éléments de l'image pour faire un diagnostic global. Ce constat est d'autant plus important dans les lésions les plus subtiles. Par exemple, différencier un épithélium normal d'une dysplasie de bas grade est aidé par le fait que l'on peut comparer la zone suspecte de dysplasie à celle, voisine, qui paraît normale sans aucun doute. Un autre exemple est celui de la dysplasie

sévère, qu'il est possible de différencier du carcinome invasif car on verra dans ce dernier le franchissement de la lame basale et la présence de stroma. Sur une tuile centrée sur l'épithélium dysplasique, sans chorion visible, il est presque impossible de trancher entre ces deux diagnostics. Or, tous ces éléments de contexte très importants pour le pathologiste sont ignorés dans les modèles classiques.

De plus, ces modèles ne prennent pas en compte la position des tuiles pour l'analyse. Or cette information est toujours utilisée par les pathologistes : l'épithélium de revêtement, qui peut être normal ou dysplasique, se trouve à la périphérie du prélèvement, alors que le carcinome infiltrant sera plutôt au centre. Ces éléments importants sont totalement ignorés par les architectures de *multiple instance learning* classiques, qui n'apprennent à reconnaître les lésions qu'en patterns isolés.

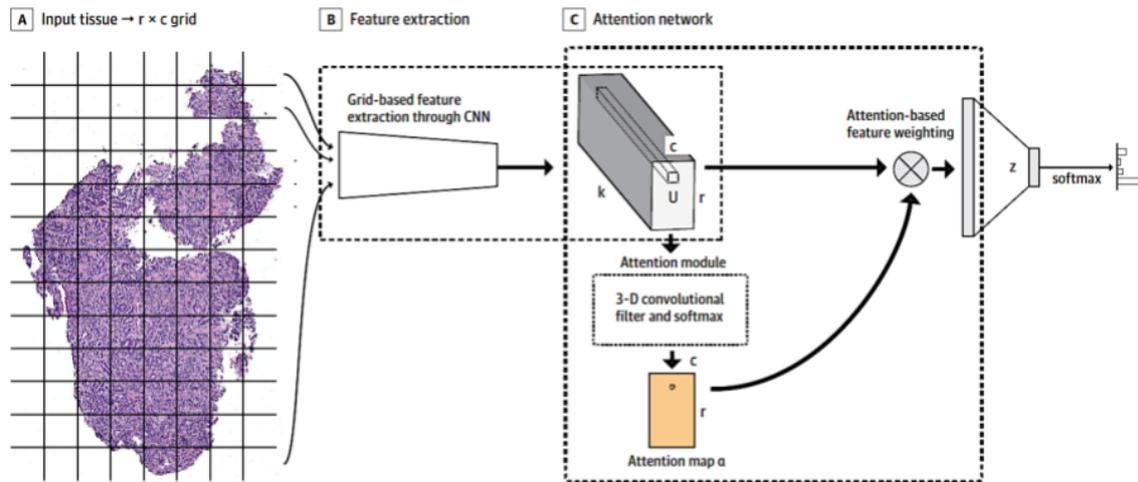
Enfin, mis à part le modèle CLAM, tous les autres modèles de *multiple instance learning* sont pensés pour des classifications binaires (présence ou absence de lésion), ce qui n'est pas adapté à notre problématique.

À la suite de ces constats, l'hypothèse qui a été formulée est que les performances de l'algorithme pourraient être améliorées en prenant en compte les lésions dans leur contexte spatial.

### ***Une prise en compte du contexte spatial***

A notre connaissance, la première publication d'un modèle prenant en compte le contexte spatial a été proposé par Tomita et al en 2019 <sup>(71)</sup> pour la détection des lésions d'endobrachyoesophage (EBO), une maladie où se succèdent la dysplasie glandulaire de bas grade et haut grade puis l'adénocarcinome. Tomita avait constaté que dans les modèles actuels, chaque tuile est analysée indépendamment de ses voisines, ce qui est très différent de l'analyse par un pathologiste. Le modèle proposé a pour force de déterminer les régions d'intérêt dans l'image et les analyse exclusivement, sans annotations préalable, avec uniquement le

label global indiqué dans l'entraînement. Le module d'attention est en trois dimensions, ce qui permet de s'adapter à toute taille d'image.

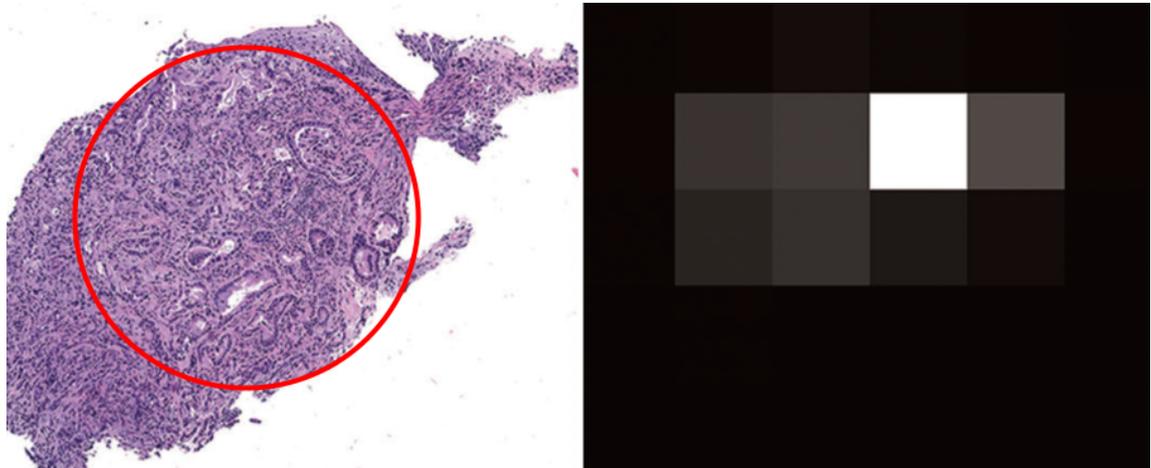


**Figure 20. Modèle proposé par Tomita et al**

Cependant, l'étude de Tomita comporte plusieurs limites notables.

D'une part, au niveau jeu de données qui comportait peu de lames (180). Pour pallier ce problème, les auteurs ont choisi de fusionner leurs dysplasies de bas grade avec celles de haut grade dans une même classe "EBO avec dysplasie", ainsi que les dysplasies de haut grade suspectes d'invasion avec les adénocarcinomes invasifs dans une même classe "adénocarcinome". Ces catégories diagnostiques ne suivent pas la classification en vigueur et manquent de précision, ce qui est d'autant plus dommageable pour le mélange de lésions dysplasiques et de lésions infiltrantes qui n'ont pas les mêmes conséquences pour le patient. Les auteurs ont également choisi de considérer chaque niveau de coupe présent sur une même lame comme une image à part entière, ce qui permet d'augmenter le jeu de données en quantité mais nuit à sa qualité. En effet, cela induit des informations redondantes dans l'entraînement et dans le test et améliore faussement les performances de l'algorithme, qui aura été confronté plusieurs fois à quasiment la même image. Les performances ne sont d'ailleurs pas très bonnes, avec une exactitude moyenne semblant correcte (0,83) mais des performances

faibles par ailleurs (F1-score moyen de 0,59 ; performances très basses pour les catégories “EBO avec dysplasie” et “adénocarcinome”). Les *heatmaps* générées sont également grossières, inexploitable pour l’interprétabilité des résultats.



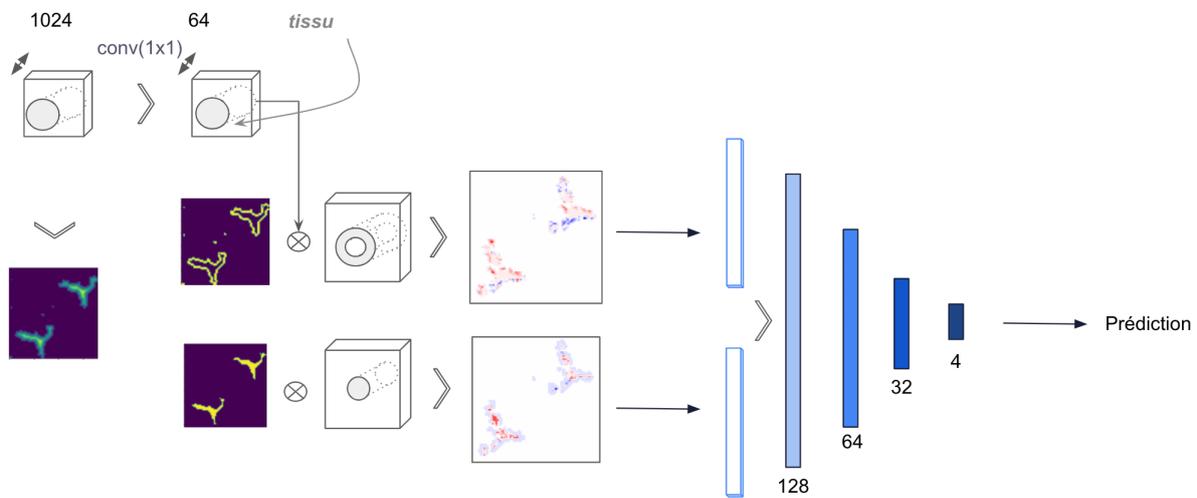
**Figure 21. Heatmap de Tomita et al (zone cerclée en rouge : adénocarcinome)**

D’autre part, le modèle en lui-même est perfectible. En effet, l’utilisation de convolutions synthétisant les pixels (*kernel* des filtres de 3 x 3, 5 x 5...) ne permet de prendre en compte que les tuiles voisines les unes des autres et non pas le contexte dans lequel se trouve la lésion : à la surface du tissu ou dans la profondeur.

### ***Proposition d’un nouveau modèle de Spatial Attention***

Le modèle que nous avons développé cherche à prendre en compte ces problématiques. Il comporte deux branches d’attention, l’une se focalisant sur les tuiles à la périphérie du prélèvement, l’autre sur la profondeur. Afin d’éviter une perte d’information du contexte spatial, la convolution est faite avec un filtre de dimension 1x1. Les vecteurs extraits des tuiles sont réorganisés sur la lame, ce qui aboutit à un cube de 1024 de profondeur. À la suite de cela, une cartographie des vecteurs est réalisée et les tuiles sont adressées à deux branches d’attention indépendantes : une pour la périphérie du tissu, l’autre pour le centre. On obtient

donc deux vecteurs descripteurs de la lame. Ces deux descripteurs sont concaténés et le vecteur final obtenu est utilisé par le classifieur (*fully connected layers*).



**Figure 22. Schéma du modèle de Spatial Attention à deux branches**

### ***Prise en compte du continuum lésionnel des lésions dysplasiques***

Un autre élément qui est ressorti des échanges entre *data scientist* et pathologiste est que les modèles actuellement décrits dans la littérature ne prennent pas en compte l'aspect progressif des lésions dysplasiques dans l'analyse. En effet, l'une des principales difficultés dans l'attribution d'un grade de dysplasie est lié au fait que ces lésions se situent sur un spectre lésionnel, sans aspect toujours nettement tranché entre le bas grade et le haut grade, ou même entre un épithélium normal et une dysplasie de bas grade. L'analyse de l'épithélium de surface est subtile et complexe, à la fois qualitative et quantitative. C'est pourquoi, en l'absence de carcinome invasif évident, le pathologiste va concentrer son interprétation sur l'épithélium de surface pour établir si une dysplasie est présente, et le cas échéant lui attribuer un grade.

La séparation de l'attention entre la périphérie du prélèvement et le centre suit cette volonté d'accorder une attention particulière à la surface ; elle permet de limiter le bruitage du signal

induit par les tuiles du centre du prélèvement, pour accorder une attention particulière aux *features* de la surface.

Afin de s'adapter à la complexité plus élevée de l'analyse de l'épithélium de surface, nous avons également testé d'adresser les tuiles de la périphérie du tissu à trois branches d'attention plutôt qu'une seule, avec l'hypothèse que l'utilisation d'un nombre de branches égal au nombre de classes (épithélium normal, dysplasie de bas grade, dysplasie de haut grade) pourrait permettre d'apporter une plus grande précision dans l'analyse et donc dans la prédiction.

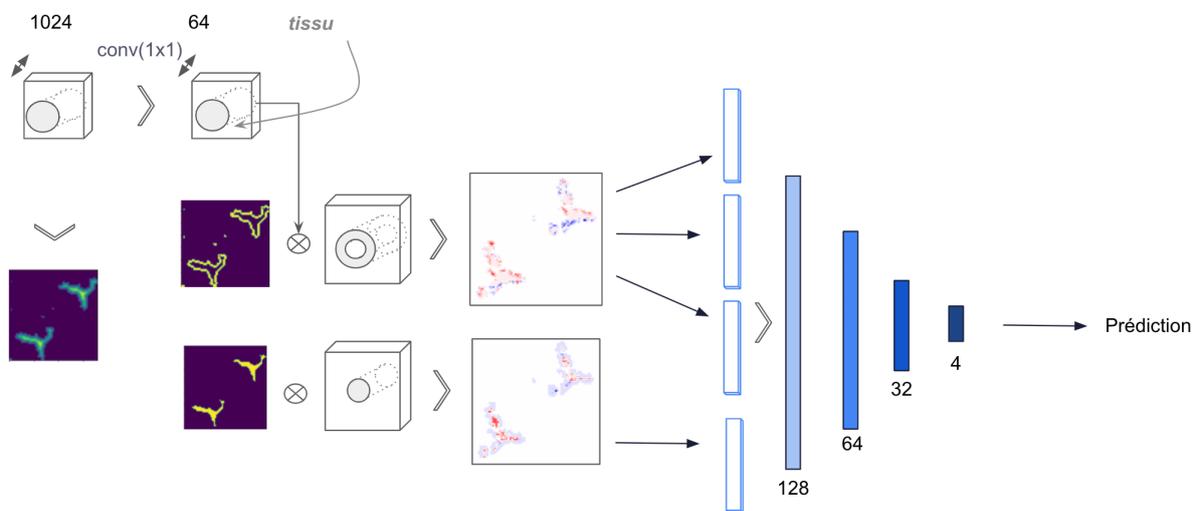


Figure 23. Schéma du modèle de Spatial Attention à quatre branches

### 3. Méthodes de calcul de la perte

Deux méthodes ont été comparées pour le calcul de la perte (*loss*) durant l'entraînement des modèles. Ces méthodes prennent en compte la gravité différente de l'erreur selon la classe prédite par rapport à la vérité de terrain : par exemple, pour un carcinome invasif, prédire une dysplasie de bas grade sera plus pénalisé que de prédire une dysplasie de haut grade.

La première méthode est la *weighted cross-entropy (WCE)*. Il s'agit d'une adaptation de la fonction *cross-entropy* avec une pénalisation plus importante de l'erreur si elle est très éloi-

gnée de la classe à prédire. Elle est très couramment utilisée en analyse d'images histologiques. Elle permet également un entraînement efficace malgré un jeu de données déséquilibré.

La deuxième méthode est la *One-sided Regression Loss (SOSR)*, proposée par Tu et Lin <sup>(72)</sup> puis adaptée par Chung et al aux réseaux de neurones convolutifs <sup>(73)</sup>. Cette méthode consiste en la prédiction de la pénalité plutôt qu'en le calcul d'une probabilité diagnostique. Elle pousse le modèle à apprendre à faire les prédictions les moins pénalisantes. Au contraire de la *WCE*, cette méthode ne permet pas de compenser un déséquilibre du jeu de données ; c'est pourquoi nous avons également réalisé un *upsampling* de la classe 1 durant l'entraînement. Cela consiste à montrer plus souvent au modèle les lames de la classe la moins représentée. Il s'agit d'une méthode couramment utilisée en *machine learning* lorsque l'on dispose d'un jeu de données déséquilibré.

# Chapitre 4. Évaluation de la performance des modèles

## 1. Métriques

Les métriques pour mesurer les performances des algorithmes sont calculées à partir des valeurs obtenues dans la matrice de confusion, dont un modèle est présenté ci-dessous.

	Prédiction	
Vérité de terrain	Vrais Négatifs (VN)	Faux Positifs (FP)
	Faux Négatifs (FN)	Vrais Positifs (VP)

**Figure 24. Matrice de confusion**

### 1.1 Exactitude pondérée

L'exactitude permet de connaître la proportion de bonnes prédictions :  $Acc = \frac{VN+VP}{VN+FN+VP+FP}$

Nous avons choisi comme métrique de référence pour l'évaluation des performances des modèles une exactitude pondérée, qui a été proposée par la Société Française de Pathologie pour le *data challenge 2020* pour le diagnostic des lésions dysplasiques et invasives du col utérin. Le calcul comporte une pondération de la performance selon que le modèle commet des erreurs de prédiction plus ou moins graves. Par exemple, le diagnostic d'un carcinome à la place d'une dysplasie de bas grade sera plus pénalisé que celui d'une dysplasie de haut grade <sup>(74)</sup>.

Vérité de terrain	Prédiction classe 0	Prédiction classe 1	Prédiction classe 2	Prédiction classe 3
Classe 0	0	0,1	0,7	1
Classe 1	0,1	0	0,3	0,7
Classe 2	0,7	0,3	0	0,3
Classe 3	1	0,7	0,3	0

**Tableau 6. Grille de pondération des erreurs utilisée pour le calcul de l'exactitude pondérée <sup>(74)</sup>**

## 1.2 Sensibilité, spécificité, précision

La sensibilité mesure la proportion des prédictions positives correctement identifiées.

La spécificité mesure, à l'inverse, la proportion d'éléments négatifs correctement identifiés.

La précision (ou valeur prédictive positive) correspond au nombre de documents correctement attribués à la classe par rapport au nombre total de documents prédits comme appartenant à la classe.

$$Se = \frac{VP}{VP+FN} \quad ; \quad Sp = \frac{VN}{VN+FP} \quad ; \quad Precision = \frac{VP}{VP+FP}$$

## 1.3 F1-score

Le F1-Score combine la précision et la sensibilité, sans prise en compte des faux négatifs. Il s'agit donc d'une métrique intéressante dans les cas où le jeu de données est déséquilibré

entre les classes.  $F1 = 2 \cdot \frac{Pr \cdot Se}{Pr + Se}$

## 1.4 AUC de la courbe ROC

Une courbe ROC (*receiver operating characteristic*) permet une représentation graphique des performances d'un modèle de classification pour tous les seuils de classification. Le calcul de l'aire sous la courbe (AUC) permet d'évaluer la performance de l'algorithme. Pour nos modèles, nous avons produit des courbes ROC moyennes ainsi que pour chaque classe diagnostique pour évaluer d'éventuelles différences de performances entre les classes.

## 1.5 Coefficient kappa de reproductibilité

Nous avons évalué la reproductibilité de notre modèle le plus performant en calculant le coefficient de reproductibilité kappa de Cohen, en prenant comme éléments de comparaison la vérité de terrain (diagnostic fait par le pathologiste au moment où le cas a été pris en charge) et la prédiction de l’algorithme. Afin de pouvoir comparer la reproductibilité de notre modèle à celle des pathologistes mentionnées dans la littérature, nous avons calculé le kappa non pondéré et le kappa pondéré : pondérations linéaire et quadratique. Le coefficient de kappa quadratique est le plus représentatif car la pondération des erreurs prend en compte l’impact différent d’une erreur entre des classes proches ou éloigné (une lésion de classe 2 prédite en classe 1 versus classe 3).  $K = \frac{Cobs - Cal}{1 - Cal}$

La valeur des coefficients de kappa a été interprétée suivant le classement de Landis et Koch, précisé dans le tableau ci-dessous.

Coefficient Kappa	Interprétation
< 0	Très mauvais
0-0,20	Mauvais
0,21-0,40	Passable
0,41-0,60	Moyen
0,61-0,80	Bon
0,81-1,00	Très bon

Tableau 7. Interprétation des valeurs du coefficient Kappa selon Landis et Koch

## 2. Interprétation de la classification des lames virtuelles

### 2.1 Cartographie des scores d’attention superposée à l’image histologique (*heatmap*)

Pour les modèles de référence et pour le modèle de *Spatial Attention* à deux branches, les *heatmaps* ont été générées sur la plateforme de Keen Eye et ont permis de visualiser les poids d’attentions générés directement superposés sur la lame virtuelle.

Pour le modèle de *Spatial Attention* à quatre branches, il n'est pas possible de créer une seule *heatmap* avec la superposition des quatre *heatmaps* obtenues car elle serait illisible. Des *heatmaps* ont été générées grâce au logiciel Python et lues en parallèle de la lame numérisée.

## **2.2 Visualisation des *features* extraites par le modèle**

Afin de pouvoir visualiser les *features* extraites par le modèle, nous avons généré des pseudo-images par la méthode d'*Activation Maximization* telle que décrite par Nguyen et al<sup>(75)</sup>. Cette méthode permet de créer des images de synthèse qui maximisent l'activation de la *feature* extraite par le modèle. Elle nous a permis de comparer l'extraction de *features* par ImageNet et par *SSL*.

## Caractérisation de la cohorte et du jeu de données

---

### 1. Tableaux descriptifs de la cohorte et du jeu de données

La description de la cohorte des patients est résumée dans le **Tableau 8**. Les relectures de lames sont résumées dans les **Tableaux 9 et 10**. La description du jeu de données final est résumée dans le **Tableau 11**.

Caractéristiques	Total
<b>Patients</b>	610
Hommes	494
<i>Âge moyen au diagnostic</i>	61
Femmes	116
<i>Âge moyen au diagnostic</i>	60
<b>Prélèvements</b>	895
Biopsies	620
Pièces opératoires	275
<b>Localisation anatomique</b>	
Larynx	677
Pharynx	63
Nasopharynx	6
Hypopharynx	112
Bord latéral de la langue ou cavité buccale	71
<b>Diagnostic (lésion la plus grave identifiée sur la lame)</b>	
Épithélium normal	37
Dysplasie de bas grade	92
Dysplasie de haut grade	301
<i>dont dysplasie modérée</i>	91
<i>dont dysplasie sévère / carcinome in situ</i>	210
Carcinome invasif	465
<i>dont micro-invasif</i>	140
<i>dont bien différencié</i>	266
<i>dont moyennement différencié</i>	107
<i>dont peu différencié</i>	37
<b>Statut vital</b>	
Vivant	120
Perdu de vue	300
Décédé	190

**Tableau 8. Descriptif de la cohorte**

<b>Nombre de lames relues</b>	<b>127</b>
Lames reclassées en épithélium normal	43
Lames reclassées en dysplasie de bas grade	56
Lames reclassées en dysplasie de haut grade	23
<i>dont dysplasie modérée</i>	15
<i>dont dysplasie sévère</i>	7
Lames combinant deux lésions	5
<i>Dysplasie légère et CIS</i>	
<i>Dysplasies modérée et sévère</i>	
<i>Dysplasie légère et modérée</i>	
<i>Epithélium normal et dysplasie légère</i>	
<i>Epithélium normal et dysplasie modérée</i>	

**Tableau 9. Relecture des dysplasies légères à modérées**

<b>Nombre de lames relues</b>	<b>23</b>
Micro-invasifs réfutés par la relecture	10
Micro-invasifs détectés par la relecture	4
Micro-invasifs confirmés par la relecture	2
Carcinomes <i>in situ</i> sans micro-invasion confirmés par la relecture	7

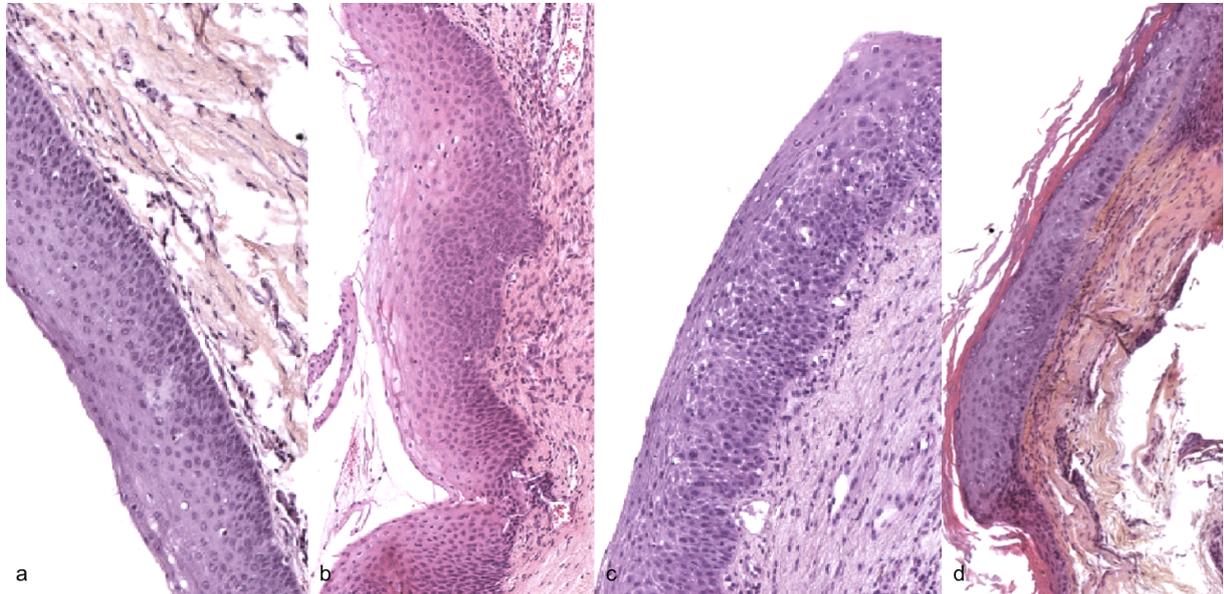
**Tableau 10. Relecture des lames difficiles pour estimer la micro-invasion**

<b>Caractéristiques</b>	<b>Nombre de lames</b>
<b>Type de prélèvement</b>	
Biopsie	1626
Pièce opératoire	734
<b>Labels présents</b>	
Épithélium normal	800
Dysplasie de bas grade	358
Dysplasie de haut grade	1135
<i>dont dysplasie modérée</i>	324
<i>dont dysplasie sévère / carcinome in situ</i>	811
Carcinome infiltrant	761
<i>dont micro-invasif</i>	181
<i>dont bien différencié</i>	390
<i>dont moyennement différencié</i>	137
<i>dont peu différencié</i>	53
<b>Classe lésionnelle de la lame</b>	
Épithélium normal	593
Dysplasie de bas grade	267
Dysplasie de haut grade	770
<i>dont dysplasie modérée</i>	232
<i>dont dysplasie sévère</i>	538
Carcinome infiltrant	730
<b>Nombre total de lames</b>	<b>2360</b>

**Tableau 11. Descriptif du jeu de données**

## 2. Exemples de cas relus par l'expert

Les exemples ci-dessous sont issus des cas relus pour reclassification des dysplasies « légères à modérées ».



**Figure 25. Exemples de lésions classées initialement en dysplasies « légères à modérées » relues par l'expert.**  
a. Preinv\_1480, reclassée en épithélium normal ; b. preinv\_1507, reclassée en dysplasie de bas grade ; c. preinv\_1218, reclassée en dysplasie de haut grade (modérée) ; d. preinv\_1493, reclassée en dysplasie de haut grade (sévère)

## Résultats

---

# Chapitre 1. Performances des modèles

## 1. Métriques, matrices de confusion, courbes ROC

Les performances des modèles testés sont détaillées ci-dessous. Hormis les matrices de confusion, les figures ont été réalisées avec les données issues du premier *split*.

Architecture	Initialisation du CNN	Fonction perte	Exactitude pondérée	Exactitude*	AUC*	Précision*	Sensibilité*	Spécificité*	F1-score*
Mean Pooling	ImageNet	WCE	0,799	0,763	0,77	0,519	0,491	0,837	0,486
Max Pooling	ImageNet	WCE	0,827	0,71	0,798	0,539	0,515	0,846	0,5
CLAM	ImageNet	WCE	0,839	0,802	0,822	0,546	0,554	0,864	0,544
CHOWDER	ImageNet	WCE	0,842	0,765	0,78	0,539	0,503	0,845	0,496
Attention Pooling	ImageNet	WCE	0,846	0,785	0,817	0,566	0,545	0,858	0,532
Attention Pooling	SSL	WCE	0,867	0,819	0,854	0,609	0,606	0,877	0,597
Attention Pooling	SSL	SOSR	0,874	0,827	0,86	0,613	0,603	0,88	0,602
Spatial Attention	SSL	WCE	0,865	0,797	0,826	0,592	0,557	0,865	0,558
Spatial Attention	SSL	SOSR	0,876	0,806	0,847	0,603	0,59	0,872	0,58
Spatial Attention 4 Branches	SSL	SOSR	0,879	0,808	0,85	0,619	0,608	0,875	0,595

\* Macro average par classe, moyenne sur les *folders*

Tableau 12. Métriques des modèles testés

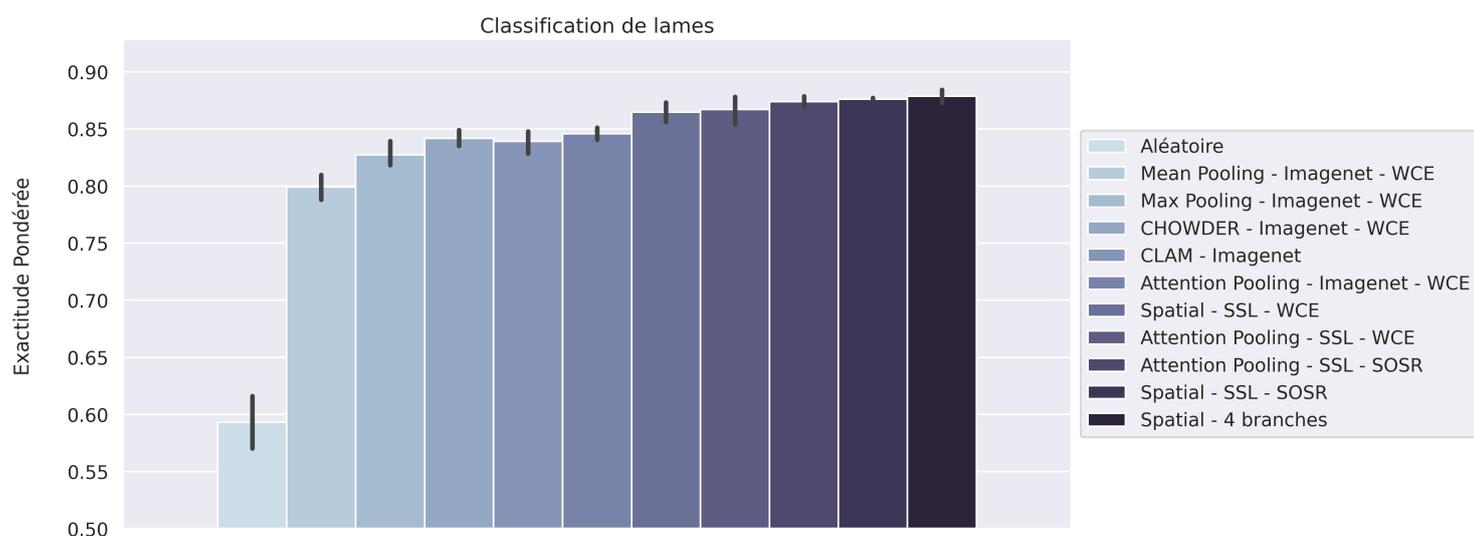
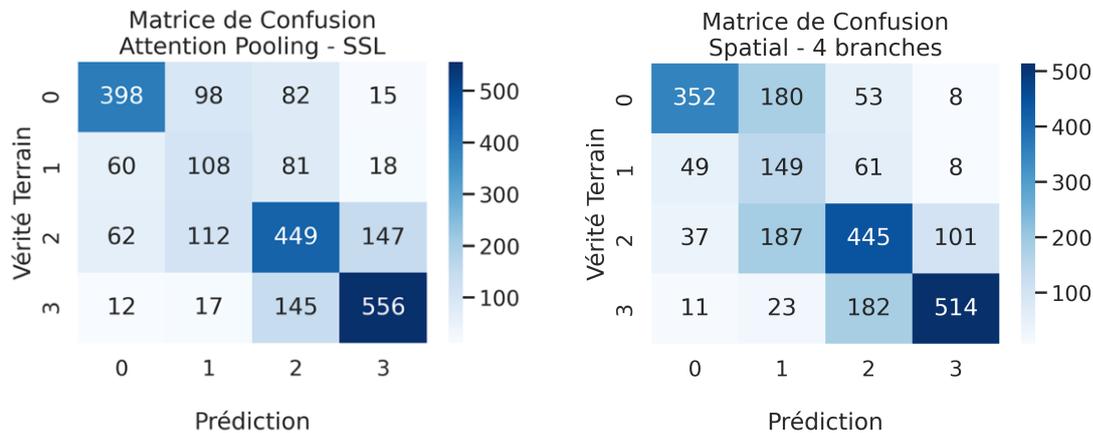


Figure 26. Exactitude pondérée des modèles testés

Les matrices de confusion ci-dessous illustrent les performances des deux meilleurs modèles que nous avons testés, c'est-à-dire la *Spatial Attention* à quatre branches et l'*Attention Pooling*.



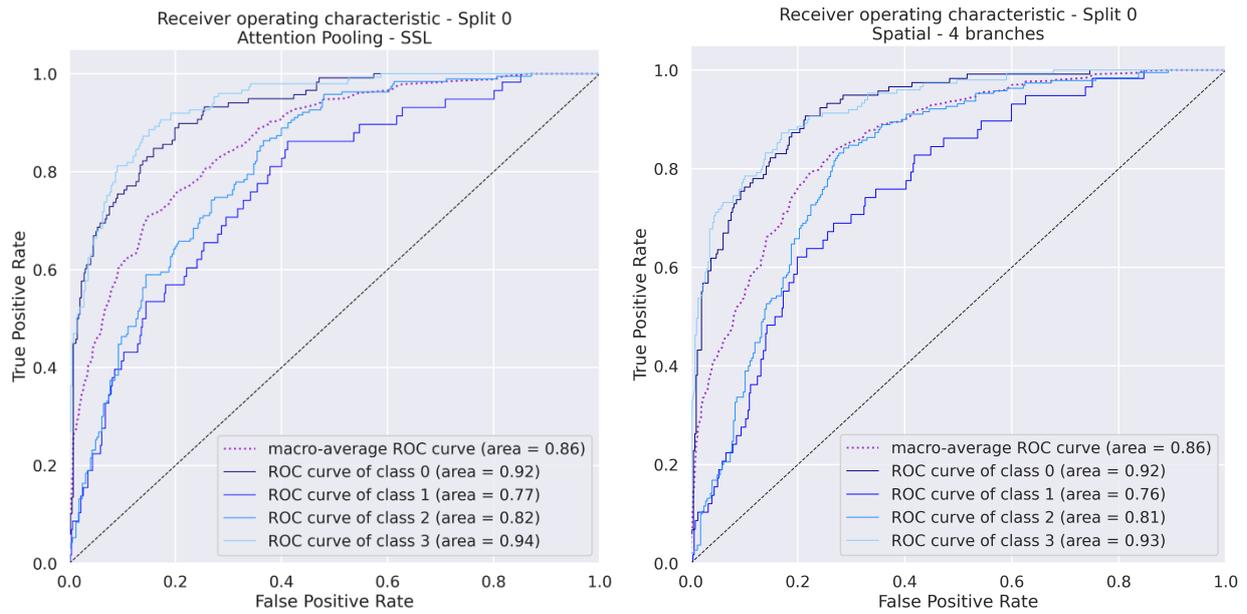
**Figure 27. Matrices de confusion de l'Attention Pooling et de la Spatial attention**

On constate que, comparativement à l'*Attention Pooling*, la *Spatial Attention* à quatre branches :

- classe mieux les dysplasies de bas grade ;
- prédit moins de fausses lames normales ;
- distingue mieux les dysplasies selon leur grade ;
- mais prédit davantage de de carcinomes infiltrants en label "dysplasie de haut grade".

Ainsi, le modèle de *Spatial Attention* ne fait pas beaucoup moins d'erreurs sur le strict plan numérique, mais l'écart entre les erreurs et la vérité terrain est resserré, induisant une meilleure exactitude pondérée que pour le modèle *Attention Pooling*.

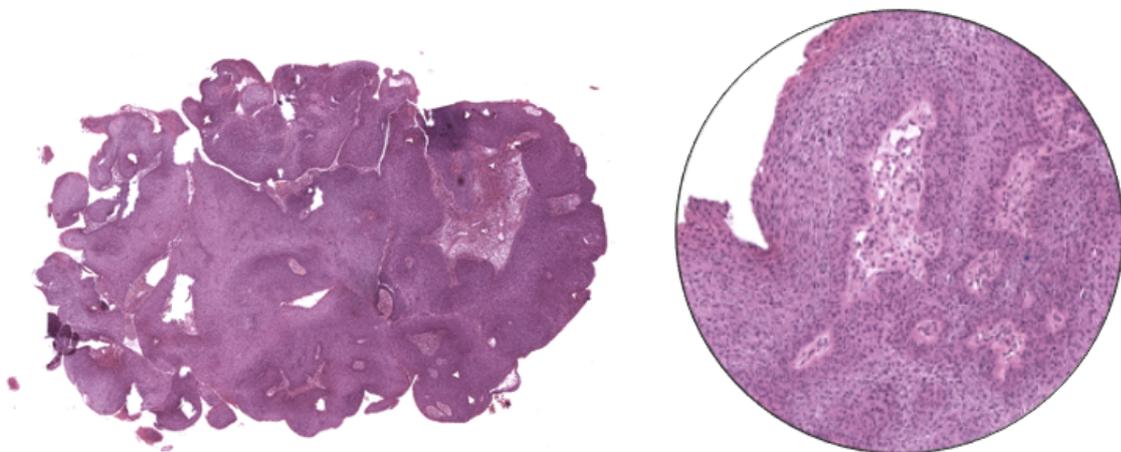
Dans tous les cas, les modèles testés sont toujours moins performants sur la classe 1 - dysplasie de bas grade, comme l'illustrent les courbes ROC ci-après :



**Figure 28. Courbes ROC de l'Attention Pooling et de la Spatial attention**

Le modèle à quatre branches commet davantage d'erreurs entre les classes 2 et 3 que le modèle d'Attention Pooling. Les erreurs concernent deux types de cas :

- des cas où l'incidence de coupe tangentielle rend difficile la distinction entre dysplasie de haut grade et carcinome invasif ;
- des cas où de la dysplasie de haut grade était bien présente, ainsi qu'une micro-invasion qui n'a pas été détectée.



**Figure 29. preinv\_1202 labellisée classe 2, prédite en classe 3**

## 2. Reproductibilité du modèle de *Spatial Attention* à quatre branches

Les coefficients kappa de reproductibilité du modèle sont donnés dans le **Tableau 13** qui permet leur comparaison aux études de reproductibilité déjà publiées.

Etude	Nombre d'échantillons	Nombre de lecteurs	Classes diagnostiques	Reproductibilité brute (%)	Kappa non pondéré	Kappa linéaire	Kappa quadratique
Gale et al 2014	38	9	1, 2, 3		0,73**		
Mehlum et al 2018	211	2	1, 2, 3	0,83	0,45	0,46	
Spatial Attention	2300	2*	1, 2, 3	0,662	0,462	0,537	0,626
4 branches	2300	2*	0, 1, 2, 3	0,619	0,485	0,636	0,757

**Tableau 13. Reproductibilité du modèle *Spatial Attention* à quatre branches, comparaison à la littérature**

\*Sont considérés le label (vérité terrain) et la prédiction de l'algorithme

\*\*Pour les six relecteurs internationaux de l'étude

La *Spatial Attention* à quatre branches a un coefficient kappa plus élevé que dans l'étude de reproductibilité de Mehlum et al. Le kappa quadratique pour le modèle à quatre branches est bon, et supérieure si l'on prend en compte la classe 0 – épithélium normal (qui n'est pas prise en compte dans les articles de Gale et al, Mehlum et al).

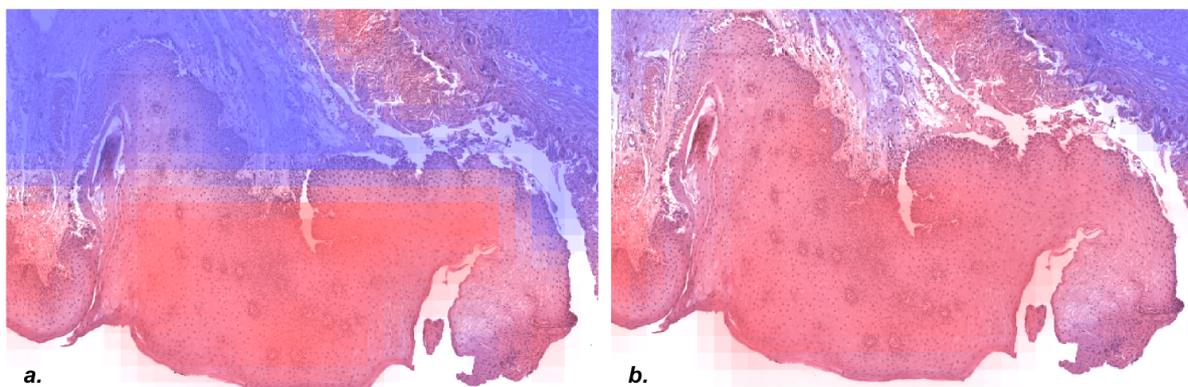
## Chapitre 2. Analyse des *heatmaps*

---

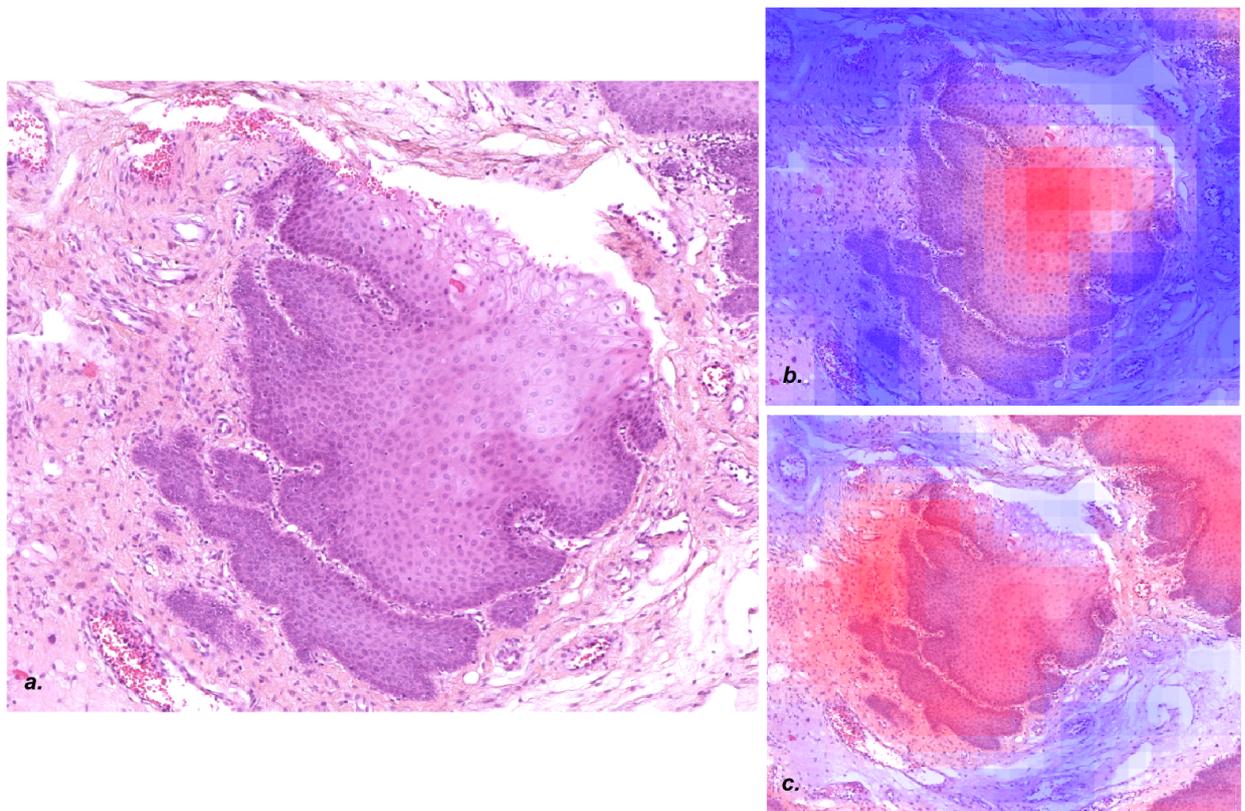
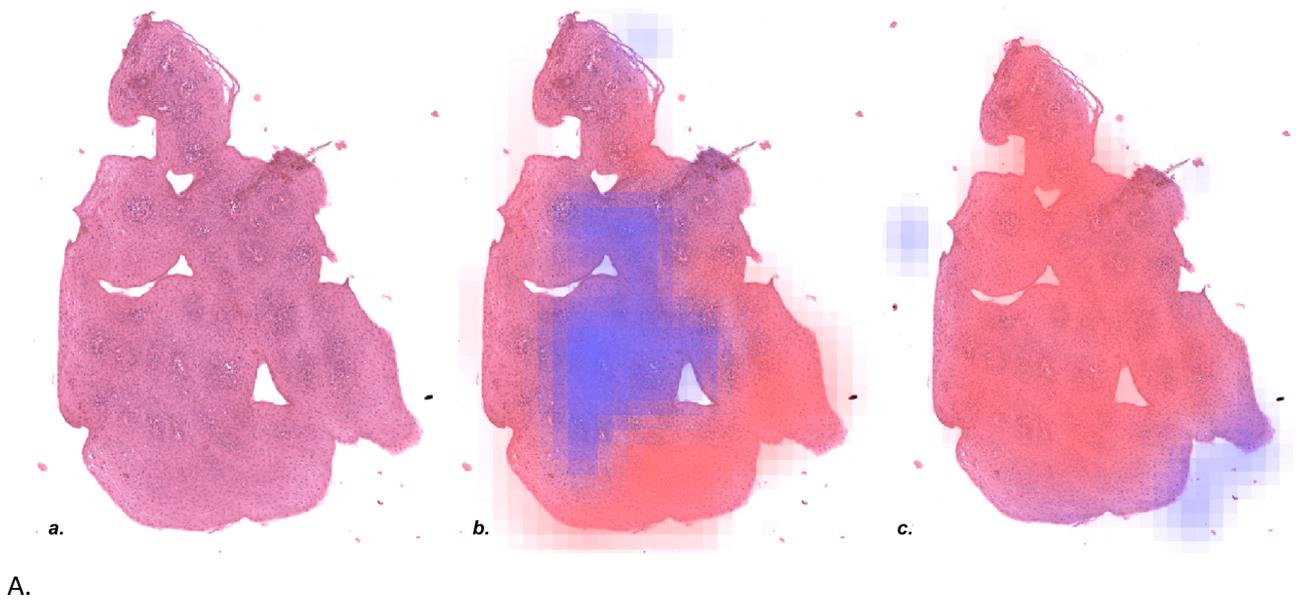
Les *heatmaps* ci-dessous ont été générées pour les cas du *split zero*. La couleur rouge indique les zones de l'image avec de forts poids d'attention ; le bleu les zones de faible poids.

### 1. Comportement du modèle de *Spatial Attention* à deux branches

Comparativement au modèle d'*Attention Pooling*, le modèle de *Spatial Attention* à deux branches a permis de rectifier quelques erreurs de classification d'épithélium normal en dysplasie de bas grade. Il s'agissait de cas où il existait une couche basale épaisse dans un épithélium lui-même épais, ou avec une incidence de coupe tangentielle. Les *heatmaps* permettent de constater un comportement différent des modèles dans l'analyse des lames d'épithélium normal.



**Figure 30. preinv\_876. *Spatial attention* à deux branches (a) vs *Attention Pooling* (b). Cas labellisés classe 0. Le modèle de *Spatial Attention* se focalise sur la surface de l'épithélium et ignore les assises basales. L'*Attention Pooling* accorde la même attention à toute l'épaisseur épithéliale.**

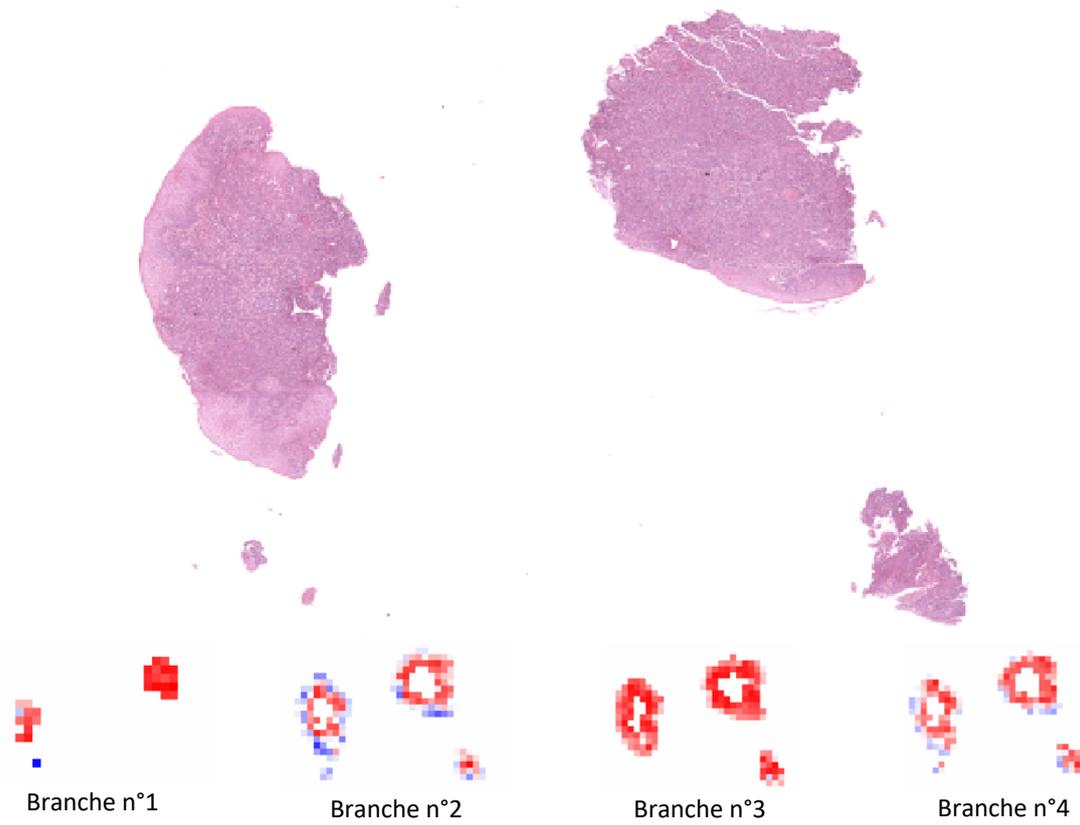


**Figure 31. Spatial attention à deux branches vs Attention Pooling. Cas labellisés classe 0, prédits à tort en classe 1 par le modèle Attention Pooling ; a. HES, b. heatmap Spatial Attention\_SSL\_SOSR, c. heatmap Attention Pooling\_SSL\_SOSR**  
 A. preinv\_773 : la Spatial Attention a réussi à ne pas prendre en compte l'incidence de coupe tangentielle, qui donne un aspect plus basophile aux cellules au centre du prélèvement.  
 B. preinv\_2657 : la Spatial Attention s'est concentrée sur les kératinocytes matures et normaux et a ignoré les aspects plus basophiles des assises basales coupées de biais, contrairement au modèle d'Attention Pooling.

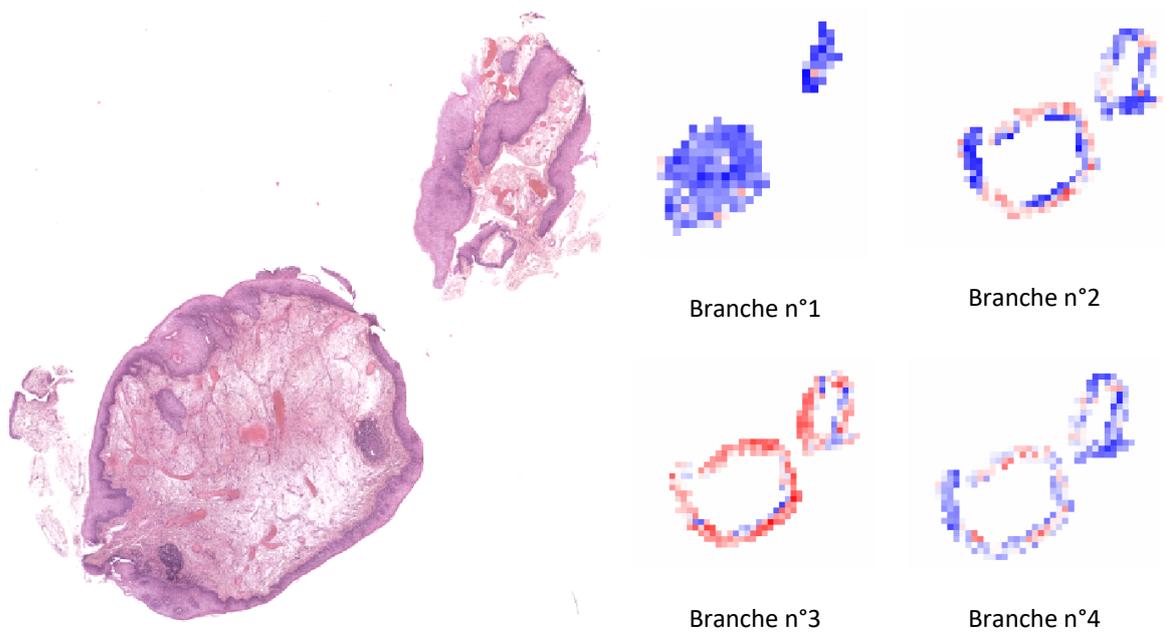
## 2. Comportement du modèle de *Spatial Attention* à quatre branches

Pour rappel, ce modèle est construit avec trois branches d'attention sur la périphérie du prélèvement et une branche au centre. Chaque branche d'attention aboutit à un vecteur, donc à une *heatmap*. Nous avons constaté que chaque branche a un comportement distinct et stable entre les différentes lames analysées. Ces comportements sont illustrés dans les figures suivantes :

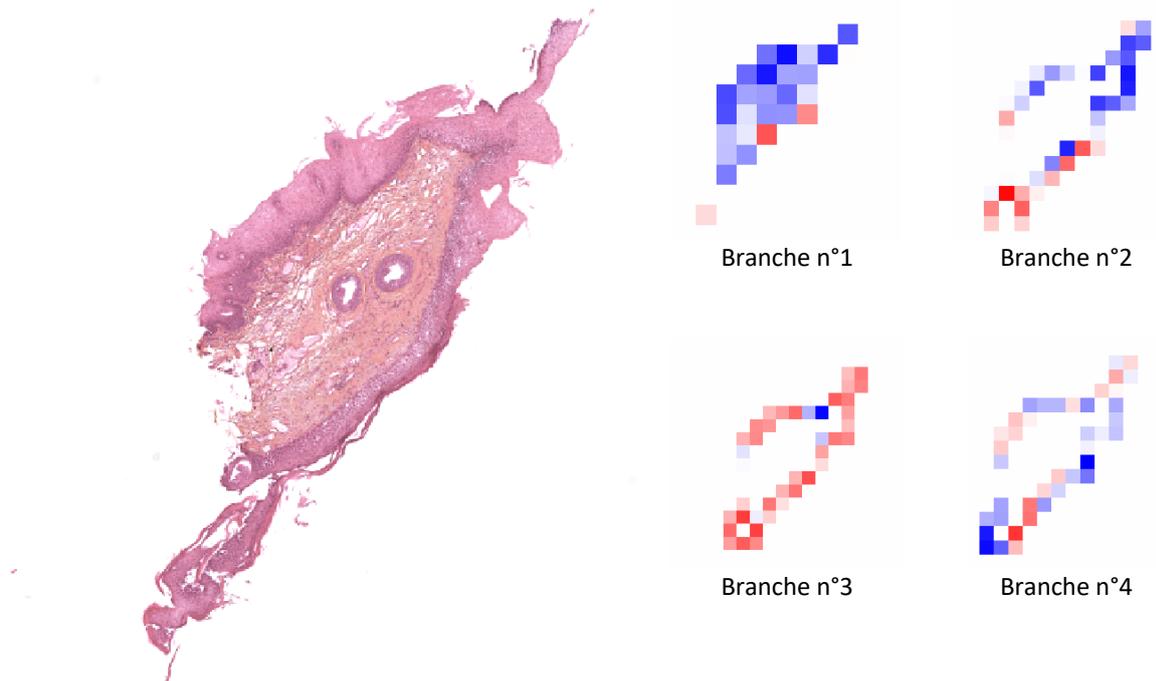
- la branche n°1, qui analyse le centre du prélèvement, accorde de forts poids d'attention aux images de carcinome invasif ;
- la branche n°3 accorde de forts poids d'attention à tout l'épithélium en surface. C'est la seule branche activée dans les cas de classe 0 – épithélium normal ;
- les branches n°2 et n°4 sont activées dans les cas de dysplasie. La branche n°2 semble activée dans les cas de dysplasie de haut grade et accorde de forts poids d'attention à la surface de l'épithélium ; la branche n°4 semble activée dans les cas de dysplasie de bas grade et accorde de l'attention aux assises profondes.



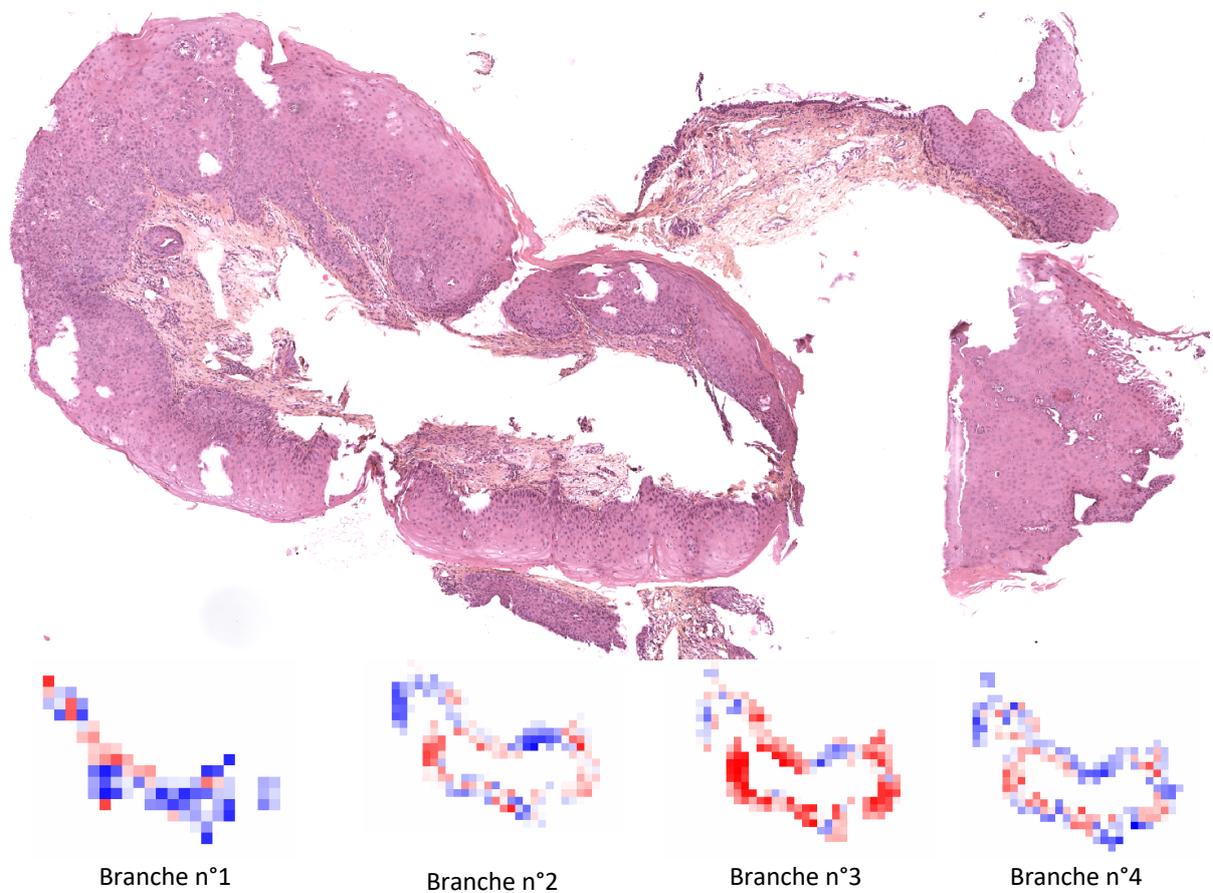
**Figure 32. preinv\_2607, labellisée classe 3, comportant du carcinome infiltrant et de la dysplasie de haut grade. La branche d'attention n°1, qui est celle focalisée sur le centre du prélèvement, a donné de forts poids d'attention, ainsi que les autres branches focalisées sur la périphérie.**



**Figure 33. preinv\_541, labellisée classe 0. Seule la branche n°3 donne de forts poids d'attention**



**Figure 34. preinv\_461, labellisée classe 1.** La branche n°3 est activée sur toute la surface épithéliale ; les branches n°2 et n°4 sur les aspects dysplasiques. La branche n°2 semble être activée pour les assises superficielles alors que la branche n°4 plutôt pour les assises basales.



**Figure 35. preinv\_76, labellisée classe 2.** La branche n°2 est très activée sur les aspects dysplasiques de haut grade.

### Chapitre 3. Analyse des *features* extraites après initialisation des poids par SSL versus ImageNet

---

L'analyse des métriques montre que le self-supervised learning rend les modèles plus performants que le pré-entraînement par ImageNet. Avec la méthode d'*Activation Maximization* qui permet de générer des images de synthèse activant le plus le réseau, on constate que les *features* extraites après SSL ressemblent davantage au tissu histologique. En effet, on distingue des cellules, des noyaux, peut-être parfois des nucléoles...



**Figure 36. Images de synthèse générées par Activation Maximization**  
En haut : features après pré-entraînement par SSL ; en bas, avec ImageNet

Ces éléments suggèrent que le pré-entraînement grâce au SSL permet d'extraire des *features* plus pertinentes, aboutissant à de meilleures performances des modèles.

## Discussion

---

Ce travail avait pour objectif le développement de modèles de *deep learning* pour le diagnostic automatisé des lésions malpighiennes des VADS selon la classification OMS 2017. Il s'est appuyé sur le partage des connaissances et des compétences entre le pathologiste et le *data scientist* pour rechercher des axes d'amélioration aux modèles déjà proposés dans la littérature. Ce partage a abouti au développement d'un nouveau modèle, que nous avons nommé *Spatial Attention*.

Dans un premier temps, nous détaillerons les forces et les limites de la *Spatial Attention* et les points d'amélioration que nous pourrions apporter pour améliorer ses performances. Dans un deuxième temps, nous aborderons la difficulté intrinsèque de la classification des lésions dysplasiques par un algorithme et proposerons une potentielle nouvelle méthode d'évaluation du risque de leur évolution en carcinome invasif. Enfin, nous mettrons en perspective notre travail avec le *data challenge* organisé par la Société Française de Pathologie en 2020, auquel nous avons participé et dont la problématique était très proche de celle de notre projet.

### **1. Forces et limites du modèle de *Spatial Attention***

Le modèle de *Spatial Attention* avait pour objectif d'intégrer le contexte spatial des tuiles dans l'analyse des lames numérisées. En effet, les modèles de référence dans la littérature analysent les images tuile par tuile indépendamment de leur position dans l'image, ce qui n'est pas du tout le cas lors de la lecture de lame par un pathologiste qui situe l'épithélium de surface et la profondeur du prélèvement. En nous inspirant du modèle de Tomita, nous avons développé un algorithme de *Spatial Attention* plus précis et avec des *heatmaps* exploitables pour l'analyse des résultats.

En séparant l'attention du modèle entre la périphérie de l'image et son centre, nous avons pour objectif de permettre une analyse plus fine de l'épithélium de surface, avec un signal

moins bruité par les éléments du chorion n'entrant pas en compte dans la classification des dysplasies. En utilisant ensuite une branche d'attention par classe possible sur la surface (soit trois branches : épithélium normal, dysplasie de bas grade, dysplasie de haut grade), les performances de la *Spatial Attention* ont été améliorées, en particulier pour la détection des dysplasies de bas grade qui est la classe sur laquelle tous les modèles commettent le plus d'erreurs. Cependant, la *Spatial Attention* est moins performante pour la distinction entre dysplasie de haut grade et carcinome, en particulier dans les cas où une incidence de coupe tangentielle rend l'interprétation délicate. Le fait que l'*Attention Pooling* arrive à passer outre ces incidences de coupe, et non la *Spatial Attention*, pourrait être lié au fait que le centre du prélèvement où se situent les carcinomes bénéficie de moins d'attention que ce dont il a besoin. Cela s'illustre en particulier par la comparaison entre les modèles à deux et quatre branches : le fait de tripler les branches d'attention sur la surface dégrade la performance pour la distinction des carcinomes.

L'idée de cette analyse séparée de la surface et de la profondeur du prélèvement est le fruit de nombreux échanges entre pathologiste et *data scientist* : grâce aux explications des problématiques médicales par YB à ML, celle-ci a pu proposer des expériences originales, que nous avons testées puis dont les résultats, succès comme échecs, ont été analysés par YB via l'interprétation des *heatmaps*. Ce travail en binôme a porté ses fruits puisque les métriques obtenues pour la *Spatial Attention* sont meilleures que celles du meilleur modèle de référence (*Attention MIL*). En effet, malgré la présence d'erreurs dans les classifications, l'écart entre les erreurs et la *ground truth* s'est resserré, induisant une meilleure exactitude pondérée.

Malgré de meilleures performances, la *spatial attention* reste en difficulté pour la classification des dysplasies de bas grade. Ce modèle, comme les autres modèles testés, pâtit du manque de lames de classe 1 - dysplasie de bas grade dans le jeu de données. En effet, elles

ne représentent que 15% du total des lames, malgré un recrutement des cas sur treize années. Cela s'explique par le fait que ces lésions sont plus rares : les patients consultent le plus souvent lorsqu'ils sont symptomatiques et les lésions biopsiées sont alors plus évoluées, donc plus souvent de haut grade ou déjà invasives.

Afin de pallier cette difficulté, plusieurs leviers d'action sont envisageables. Premièrement, nous pourrions augmenter le nombre de lames de classe 1 en réalisant un recrutement plus large dans notre centre (cas ultérieurs à 2013) ou dans d'autres services d'anatomie pathologique. Deuxièmement, nous pourrions optimiser l'extraction des *features* : en effet, on peut supposer que le modèle est en difficulté pour extraire les *features* pertinentes pour la dysplasie de bas grade car elles sont plus subtiles. Le *self-supervised learning* a bien amélioré les performances des modèles en permettant l'utilisation de *features* totalement adaptées aux images histologiques, puisque issues directement du jeu de données. Un pré-entraînement plus poussé par le module SSL pourrait permettre d'améliorer les performances du modèle. L'optimisation de l'extraction de *features* pourrait également être bénéfique pour une meilleure distinction entre les carcinomes et les dysplasies de haut grade dans les cas les plus difficiles (incidence de coupe tangentielle, micro-invasion). En effet, l'acquisition de propriétés infiltrantes par les cellules devrait se répercuter sur des différences morphologiques subtiles. Si elles existent, un meilleur extracteur pourrait permettre au modèle d'être plus performant dans ces cas.

Pour la validation finale du modèle qui aura les meilleures performances, notre objectif est de tester sa robustesse sur une cohorte prospective, préférentiellement multicentrique. Grâce à un recrutement étendu sur treize ans, nous devrions disposer d'une variété suffisante dans la

coloration des lames et leur qualité pour que notre algorithme fonctionne aussi sur des lames techniques dans d'autres laboratoires.

## **2. Vers une nouvelle classification des dysplasies ?**

Comme nous l'avons constaté, le diagnostic des dysplasies de bas grade est intrinsèquement le plus difficile, même pour les pathologistes. Cela s'est notamment illustré lors de la relecture des lésions classées en "dysplasie légère à modérée" par YB et CB, après laquelle près d'un tiers des lames relues (43/127) a été reclassé en épithélium normal. Ce problème de robustesse des labels, qui s'ajoute à une faible quantité des données disponibles, constitue un frein à l'apprentissage optimal du modèle qui se base sur des labels imparfaits. Cette difficulté se répercute logiquement sur les performances. En effet, malgré un jeu de données assez exhaustif, le modèle commet toujours un certain nombre d'erreurs dans la classification des dysplasies et en particulier de bas grade.

Notre analyse est que le système de classification lui-même présente des limites qui ne peuvent être entièrement résolues par une meilleure reproductibilité "artificielle". Comme nous l'avons montré dans l'introduction de ce travail, les dysplasies malpighiennes des VADS se trouvent sur un continuum lésionnel et non dans des catégories aussi tranchées que celles de la classification OMS. Comme l'expliquent Mahmood et al <sup>(8)</sup> dans leur revue de la littérature sur la classification des dysplasies orales, les critères histologiques utilisés pour attribuer un grade à une lésion sont relativement peu spécifiques pris individuellement, et peuvent se retrouver à la fois dans les dysplasies de bas grade et de haut grade, et dont seulement une partie va se transformer en carcinome invasif sans que l'on puisse prédire lesquelles. Néanmoins, cette classification reste incontournable à l'heure actuelle pour que les pathologistes aient sur quoi s'appuyer pour faire un diagnostic intelligible et le plus reproductible possible.

Dans la deuxième phase du projet PREINV, qui n'a pas pu être avancée dans ce travail de thèse, nous chercherons à exploiter les potentialités du *deep learning* pour dépasser les limites des classifications actuelles. En effet, grâce à leur apprentissage directement sur les données contenues dans l'image, les algorithmes de *deep learning* permettent de découvrir des *features* encore mal définies, inconnues voire invisibles pour l'œil humain. Dans la deuxième phase du projet PREINV, nous travaillerons donc à développer un modèle qui classera les dysplasies non pas en reproduisant une classification histologique prédéfinie, mais en apprenant à reconnaître celles ayant réellement évolué en carcinome invasif. Pour ce faire, au lieu d'utiliser le grade de la classification OMS pour *ground truth*, nous utiliserons la notion de l'évolution ou non de la lésion en carcinome invasif sur la même lame ou sur un prélèvement ultérieur du même patient au même site anatomique. En cas de succès, l'analyse des *features* extraites pourrait potentiellement mettre en lumière des éléments visibles par l'œil humain qui seraient discriminants pour juger du potentiel évolutif d'une lésion dysplasique. Ces critères pourraient ensuite être utilisés par les pathologistes en routine. Nous essaierons également de grader les dysplasies non pas selon des catégories tranchées, mais avec un scoring continu, par exemple la probabilité d'évolution, qui serait donc chiffrée de 0 à 1. Un tel score continu refléterait mieux le risque d'évolution de ces lésions elles-mêmes situées sur un continuum lésionnel.

A long terme, l'objectif de notre projet est de proposer un outil compagnon du pathologiste pour le diagnostic des lésions malpighiennes dysplasiques et infiltrantes selon la classification OMS et la prédiction de l'évolution des dysplasies en carcinome invasif, utilisable dans la pratique quotidienne, en temps réel. Cet outil serait déployé sur la plateforme de Keen Eye qui est accessible depuis tout ordinateur avec une connexion à internet.

### 3. Mise en perspective de notre travail de thèse avec le *data challenge* de la SFP

Ce travail de thèse a bénéficié de notre participation (YB, ML) au premier *data challenge* organisé par la SFP en octobre 2020 (74). Notre participation au challenge fut une vraie source d'apprentissage et d'encouragement pour notre travail, ainsi que d'enthousiasme devant cette première compétition organisée par des pathologistes français avec une telle envergure et une telle qualité.

#### 3.1 Apports du challenge

L'objectif du challenge était de développer un algorithme pour le diagnostic des lésions malpighiennes, dysplasiques et infiltrantes, du col utérin. Il s'agit de lésions proches de celles des lésions malpighiennes des VADS. Notre participation nous a permis d'aborder une problématique proche de mon sujet de thèse, avec de bons résultats puisque notre équipe s'est classée 14ème sur 547 avec un algorithme dont les prédictions avaient une exactitude pondérée de 0,8991.

Cette méthode de calcul de l'exactitude pondérée, qui a été proposée par l'équipe du challenge pour évaluer les performances des algorithmes concurrents, a été utilisée pour comparer les modèles testés dans mon travail de thèse car il s'agit de la méthode la plus adaptée pour évaluer un algorithme de classification multiclasse avec un impact différent des erreurs diagnostiques. En effet, les métriques "classiques" de sensibilité, spécificité... ne sont pas les plus adaptées pour évaluer ce type d'algorithme car elles ne reflètent pas les performances globales du modèle : ces métriques ne prennent pas en compte l'aspect ordinal de nos classes. Dans la suite du challenge, ML a testé l'initialisation des poids du modèle par *self-supervised learning (SSL)*, qui a bien amélioré ses performances avec une exactitude pondérée de 0,927(76). Cette meilleure performance du SSL par rapport à l'initialisation avec ImageNet

s'est vérifiée dans notre travail, où le SSL a été comparé à une initialisation du modèle par ImageNet sur le modèle d'*Attention Pooling* et a également amélioré les performances.

Cependant, nos performances pour notre projet restent inférieures à celles obtenues pour le challenge, tous modèles confondus. Ceci peut s'expliquer par plusieurs différences entre les jeux de données du challenge et de notre travail.

### **3.2 Différences entre les jeux de données**

Tout d'abord, bien que les problématiques du challenge et de la thèse soient très semblables, la classification OMS des dysplasies malpighiennes du col utérin est plus robuste que celle des VADS. En effet, la classification des lésions du col repose sur des données physiopathologiques bien décrites des effets d'HPV sur les cellules aux différents stades de la maladie. Cette séquence physiopathologique se reflète dans l'aspect des lésions avec des différences plus nettes entre le bas grade et le haut grade. De plus, dans les cas plus difficiles, le *grading* des lésions dysplasiques du col utérin peut être facilité par l'utilisation de marqueurs immunohistochimiques : p16, Ki67. Ainsi, intrinsèquement la problématique de notre travail est plus complexe que celle de la classification des lésions malpighiennes du col.

De plus, la constitution du jeu de données du challenge, qui était d'envergure nationale et soutenue par plusieurs institutions, était plus robuste. D'une part, la quantité des données était supérieure : en effet, à nombre de lames très proche (2542 lames pour le challenge dont 1015 lames pour le jeu d'apprentissage versus notre projet : 2360 lames dont 1500 pour l'apprentissage) le jeu de données du challenge était équilibré entre les classes, ce qui n'est pas le cas du nôtre avec une sous-représentation des lésions dysplasiques de bas grade, qui sont aussi les plus subtiles à détecter. De plus, les annotations du jeu d'apprentissage ont été faites sur des régions lésionnelles ciblées (ROI pour *region of interest*) et non sur label global : cela

a permis d'obtenir au total 5941 annotations graphiques. Les équipes du top 3 du challenge ont toutes utilisé ces annotations pour l'entraînement de leurs modèles. Dans notre travail, l'entraînement se fait sur lame entière avec un label global, ce qui est plus difficile car les tuiles discriminantes pour la classe sont bruitées par les autres. D'autre part, la qualité des données du challenge est également supérieure : chaque lame du challenge a bénéficié d'une relecture par deux pathologistes issus d'un panel d'experts. Dans notre travail, étant donné le nombre important de lames et les moindres moyens disponibles, nous avons ciblé la relecture par l'expert (CB) pour les cas les plus difficiles.

### **3.3 Différences de moyens techniques et administratifs**

Contrairement au challenge, notre travail n'a pas bénéficié de moyens alloués par des institutions d'envergure. Tout le travail de recueil des données, de sélection, de relecture et de numérisation des lames a été réalisé sans aucun organisme partenaire ni sous-traitant. Les démarches administratives à effectuer, qui ne furent pas facilitées par l'épidémie de Covid-19, prirent près d'un an pour être entièrement résolues. Ce délai s'explique en partie par l'aspect relativement précurseur de notre projet, l'AP-HP n'ayant pas encore réalisé beaucoup de partenariats avec des start-ups dans ce domaine, et notre spécialité médicale et ses spécificités étant assez mal connues par ses personnels administratifs et juridiques.

Le *data challenge* a démontré la faisabilité en France du partage de lames histologiques à très grande échelle, avec des participants aux profils variés, français ou étrangers, médecins, ingénieurs au sein d'une entreprise privée ou chercheurs indépendants, et cela avec une mise en ligne sécurisée des données médicales. Nous espérons que notre projet apportera sa pierre à l'édifice pour faciliter les démarches de futurs projets d'IA, le partage de données et les collaborations entre médecins et ingénieurs des secteurs publics comme privés, pour permettre à

la discipline de progresser en France. En gardant toujours comme principe indispensable la protection des données des patients, qui est une des clés de voûte de notre système de santé.

## Conclusion

---

Notre travail a porté sur le développement de modèles de *deep learning* pour la classification des lésions malpighiennes dysplasiques et infiltrantes des voies aérodigestives supérieures, suivant la classification de l'OMS 2017. Ce travail s'est effectué dans le cadre d'un partenariat entre l'AP-HP et la start-up Keen Eye Technologies. Il a bénéficié du travail collégial de deux doctorantes issues de deux domaines d'expertise : pour ma part, ma thèse d'exercice en médecine, sous la direction du Professeur Badoual, et du côté *data science* de la thèse Cifre de Mélanie Lubrano sous la direction du Professeur Walter. Il a également profité de l'inclusion d'un très grand nombre de lames virtuelles (2360) numérisées avec un scanner de grande qualité, et d'un support technique solide via la plateforme de Keen Eye. A notre connaissance, il s'agit du premier travail de ce type et de cette envergure sur les lésions malpighiennes des VADS. Le dialogue et le partage des connaissances entre médecin et *data scientist* a été la clé pour proposer un modèle original de *deep learning*, que nous avons nommé *Spatial Attention*. Ce modèle cherche à s'adapter aux problématiques des VADS car il prend à la fois en compte le contexte spatial des lésions et leur continuum lésionnel. Il est plus performant que les modèles de référence, avec une exactitude pondérée élevée de 0,879 et une reproductibilité de 0,757 ; mais il reste perfectible, aussi bien dans sa structure que par l'utilisation d'un jeu de données plus grand pour son apprentissage. La difficulté intrinsèque de la classification des dysplasies ne reste que partiellement résolue, et de nouvelles approches de *deep learning* pourraient permettre de créer de nouveaux outils de prédiction de l'évolution des dysplasies en carcinome invasif. Ces nouvelles approches feront l'objet de la deuxième partie de notre projet.

## Légendes des figures

---

Figure 1. Photographies de lésions des cordes vocales durant une laryngoscopie	13
Figure 2. Chronologie des lésions malpighiennes du col utérin - National Cancer Institute	18
Figure 3. Exemples de lésions malpighiennes selon la classification OMS 2017, issues du jeu de données du projet PREINV	23
Figure 4. Analogies entre les réseaux de neurones cérébraux et informatiques par Medathati et al	27
Figure 5. Structure globale d'un réseau de neurones convolutifs pour l'analyse d'images	31
Figure 6. Convolution et pooling	32
Figure 7. Niveaux d'abstraction, d'après Li et al	32
Figure 8. Cartographie des poids d'attention, d'après Lu et al	34
Figure 9. Principe de rétropropagation, d'après Calderaro et al	36
Figure 10. Conséquences du surentraînement sur les performances du modèle	36
Figure 11. Exemple de heatmap de l'algorithme gagnant du challenge Camelyon16	40
Figure 12. La pathologie augmentée au cœur de la prise en charge personnalisée des cancers, d'après Bera	42
Figure 13. Schéma de la cross-validation	54
Figure 14. Tuilage, d'après Lu et al	55
Figure 15. Principe général des modèles testés	58
Figure 16. Construction des heatmaps, d'après Lu et al	59
Figure 17. Modèle Chowder, figure adaptée de Courtiol et al	60
Figure 18. Modèle Attention Pooling, figure adaptée de Ilse et al	60
Figure 19. Modèle CLAM, figure adaptée de Lu et al	61
Figure 20. Modèle proposé par Tomita et al	63
Figure 21. Heatmap de Tomita et al	64
Figure 22. Schéma du modèle de Spatial Attention à deux branches	65
Figure 23. Schéma du modèle de Spatial Attention à quatre branches	66
Figure 24. Matrice de confusion	68
Figure 25. Exemples de lésions classées initialement en dysplasies « légères à modérées » relues par l'expert	74
Figure 26. Exactitude pondérée des modèles testés	76
Figure 27. Matrices de confusion de l'Attention Pooling et de la Spatial attention	77
Figure 28. Courbes ROC de l'Attention Pooling et de la Spatial attention	78
Figure 29. preinv_1202 labellisée classe 2, prédite en classe 3	78
Figure 30. preinv_876. Spatial attention à deux branches (a) vs Attention Pooling (b). Cas labellisés classe 0	80
Figure 31. Spatial attention à deux branches vs Attention Pooling. Cas labellisés classe 0, prédits à tort en classe 1 par le modèle Attention Pooling	81
Figure 32. preinv_2607, labellisée classe 3, comportant du carcinome infiltrant et de la dysplasie de haut grade	83
Figure 33. preinv_541, labellisée classe 0	83
Figure 34. preinv_461, labellisée classe 1	84
Figure 35. preinv_76, labellisée classe 2	84
Figure 36. Images de synthèse générées par Activation Maximization	85

## Légendes des tableaux

---

Tableau 1. Nomenclature des lésions dysplasiques laryngées, d'après Hellquist et al	20
Tableau 2. Résumé des différentes classifications des lésions dysplasiques publiées depuis 2000	22
Tableau 3. Classification OMS 2017 des dysplasies malpighiennes du larynx	23
Tableau 4. Codes des lésions selon le thésaurus ADICAP	52
Tableau 5. Classes diagnostiques basées sur la classification OMS 2017	53
Tableau 6. Grille de pondération des erreurs utilisée pour le calcul de l'exactitude pondérée	69
Tableau 7. Interprétation des valeurs du coefficient Kappa selon Landis et Koch	70
Tableau 8. Descriptif de la cohorte	72
Tableau 9. Relecture des dysplasies légères à modérées	73
Tableau 10. Relecture des lames difficiles pour la micro-invasion	73
Tableau 11. Descriptif du jeu de donnée	73
Tableau 12. Métriques des modèles testés	76
Tableau 13. Reproductibilité du modèle Spatial Attention à quatre branches, comparaison à la littérature	79

## Bibliographie

---

1. Johnson DE, Burtneß B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers*. déc 2020;6(1):92.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. nov 2018;68(6):394-424.
3. Liao LJ, Hsu WL, Lo WC, Cheng PW, Shueng PW, Hsieh CH. Health-related quality of life and utility in head and neck cancer survivors. *BMC Cancer*. déc 2019;19(1):425.
4. Odell E, Eckel HE, Simo R, Quer M, Paleri V, Klussmann JP, et al. European Laryngological Society position paper on laryngeal dysplasia Part I: aetiology and pathological classification. *Eur Arch Otorhinolaryngol*. juin 2021;278(6):1717-22.
5. Sopori M. Effects of cigarette smoke on the immune system. *Nat Rev Immunol*. mai 2002;2(5):372-7.
6. Marcos CÁ, Alonso-Guervós M, Prado NR, Gimeno TS, Iglesias FD, Hermsen M, et al. Genetic model of transformation and neoplastic progression in laryngeal epithelium. *Head Neck*. févr 2011;33(2):216-24.
7. Hashibe M, Brennan P, Benhamou S, Castellsague X, Chen C, Curado MP, et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *J Natl Cancer Inst*. 16 mai 2007;99(10):777-89.
8. Mahmood H. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions\_ A systematic review. *Oral Oncology*. 2020;9.
9. Westra WH. The Morphologic Profile of HPV-Related Head and Neck Squamous Carcinoma: Implications for Diagnosis, Prognosis, and Clinical Management. *Head and Neck Pathol*. juill 2012;6(S1):48-54.
10. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J Natl Cancer Inst*. 20 févr 2008;100(4):261-9.
11. Hellquist H, Ferlito A, Mäkitie AA, Thompson LDR, Bishop JA, Agaimy A, et al. Developing Classifications of Laryngeal Dysplasia: The Historical Basis. *Adv Ther*. juin 2020;37(6):2667-77.
12. Organisation mondiale de la santé, Centre international de recherche sur le cancer, éditeurs. WHO classification of head and neck tumours. 4th ed. Lyon: International agency for research on cancer; 2017. (World health organization classification of tumours).
13. Thompson LDR. Laryngeal Dysplasia, Squamous Cell Carcinoma, and Variants. *Surgical Pathology Clinics*. mars 2017;10(1):15-33.
14. Park JC, Altman KW, Prasad VMN, Broadhurst M, Akst LM. Laryngeal Leukoplakia: State of the Art Review. *Otolaryngol Head Neck Surg*. juin 2021;164(6):1153-9.
15. Wenig BM. Squamous cell carcinoma of the upper aerodigestive tract: dysplasia and select variants. *Mod Pathol*. janv 2017;30(S1):S112-8.
16. Thompson LDR, Wenig BM. *Head and Neck*. 2. Edition. Philadelphia: Elsevier; 2016. 1121 p. (Diagnostic pathology).
17. *Smoking Cessation: A Report of the Surgeon General*. :700.
18. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent

effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biology*. 2007;17.

19. Wang G, Wang R, Strulovici-Barel Y, Salit J, Staudt MR, Ahmed J, et al. Persistence of Smoking-Induced Dysregulation of MiRNA Expression in the Small Airway Epithelium Despite Smoking Cessation. *PLOS ONE*. 2015;18.
20. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer*. janv 2018;18(1):19-32.
21. Initial staging for squamous cell carcinoma of the mouth, larynx and pharynx (except nasopharynx). Part 3: General assessment. 2012 SFORL recommendations | Elsevier Enhanced Reader [Internet]. [cité 20 déc 2019]. Disponible sur: <https://reader.elsevier.com/reader/sd/pii/S1879729612001196?to-ken=BD367B6A9BF41B980C289CEEEB8BD1E9562C3F859BFEB0AEADB5CD02CB899C79EDFA570BA0ABCBC15A8DD63053691A7>
22. Weller MD, Nankivell PC, McConkey C, Paleri V, Mehanna HM. The risk and interval to malignancy of patients with laryngeal dysplasia; a systematic review of case series and meta-analysis: Risk and interval to malignancy of patients with laryngeal dysplasia. *Clinical Otolaryngology*. oct 2010;35(5):364-72.
23. Eckel HE, Simo R, Quer M, Odell E, Paleri V, Klussmann JP, et al. European Laryngological Society position paper on laryngeal dysplasia Part II: diagnosis, treatment, and follow-up. *Eur Arch Otorhinolaryngol*. juin 2021;278(6):1723-32.
24. Stanley M. Pathology and epidemiology of HPV infection in females. *Gynecologic Oncology*. mai 2010;117(2):S5-10.
25. Organisation mondiale de la santé, Centre international de recherche sur le cancer, éditeurs. *Female genital tumours*. 5th ed. Lyon: International agency for research on cancer; 2020. (World health organization classification of tumours).
26. Gale N, Cardesa A, Hernandez-Prera JC, Slootweg PJ, Wenig BM, Zidar N. Laryngeal Dysplasia: Persisting Dilemmas, Disagreements and Unsolved Problems—A Short Review. *Head and Neck Pathol*. déc 2020;14(4):1046-51.
27. Xiao HD, Bifulco CB. Cellular and Molecular Pathology. In: *Oral, Head and Neck Oncology and Reconstructive Surgery* [Internet]. Elsevier; 2018 [cité 12 sept 2021]. p. 57-78. Disponible sur: <https://linkinghub.elsevier.com/retrieve/pii/B9780323265683000026>
28. Cho KJ, Song JS. Recent Changes of Classification for Squamous Intraepithelial Lesions of the Head and Neck. *Archives of Pathology & Laboratory Medicine*. juill 2018;142(7):829-32.
29. Gale N, Blagus R, El-Mofty SK, Helliwell T, Prasad ML, Sandison A, et al. Evaluation of a new grading system for laryngeal squamous intraepithelial lesions—a proposed unified classification. *Histopathology*. oct 2014;65(4):456-64.
30. Sarioglu S, Cakalagaoglu F, Elagoz S, Ersoy U, Etit D, Hucumenoglu S, et al. Inter-observer Agreement in Laryngeal Pre-neoplastic Lesions. *Head and Neck Pathol*. déc 2010;4(4):276-80.
31. Fleskens SAJHM, Bergshoeff VE, Voogd AC, van Velthuysen MLF, Bot FJ, Speel EJM, et al. Interobserver variability of laryngeal mucosal premalignant lesions: a histopathological evaluation. *Mod Pathol*. juill 2011;24(7):892-8.
32. McLaren KM, Burnett RA, Goodlad JR, Howatson SR, Lang S, Lee FD, et al. Consistency of histopathological reporting of laryngeal dysplasia. *Histopathology*. nov 2000;37(5):460-3.
33. Eversole LR. Dysplasia of the Upper Aerodigestive Tract Squamous Epithelium. *Head Neck Pathol*. 22 févr 2009;3(1):63-8.
34. Eckel HE. European Laryngological Society position paper on laryngeal dysplasia Part II:

diagnosis, treatment, and follow-up. 2021;10.

35. Mehlum CS, Larsen SR, Kiss K, Groentved AM, Kjaergaard T, Möller S, et al. Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment: Assessment of Laryngeal Precursor Lesions. *The Laryngoscope*. oct 2018;128(10):2375-9.
36. Zidar N, Gale N. Update from the 5th Edition of the World Health Organization Classification of Head and Neck Tumors: Hypopharynx, Larynx, Trachea and Parapharyngeal Space. *Head and Neck Pathol*. mars 2022;16(1):31-9.
37. Kochanny SE, Pearson AT. Academics as leaders in the cancer artificial intelligence revolution. *Cancer*. 29 oct 2020;cncr.33284.
38. Medathati NVK, Neumann H, Masson GS, Kornprobst P. Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision. *Computer Vision and Image Understanding*. sept 2016;150:1-30.
39. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958;65(6):386-408.
40. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. mai 2015;521(7553):436-44.
41. LeCun Y, Bengio Y, Laboratories TB. Convolutional Networks for Images, Speech, and Time-Series. :14.
42. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 24 mai 2017;60(6):84-90.
43. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2 févr 2017;542(7639):115-8.
44. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2 janv 2020;577(7788):89-94.
45. Hoerer T, Kuenzer C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing*. 22 mai 2020;12(10):1667.
46. Li X, Liu Z, Cui S, Luo C, Li C, Zhuang Z. Predicting the effective mechanical property of heterogeneous materials by image based modeling and deep learning. *Computer Methods in Applied Mechanics and Engineering*. avr 2019;347:735-53.
47. Lu MY, Zhao M, Shady M, Lipkova J, Chen TY, Williamson DFK, et al. Deep Learning-based Computational Pathology Predicts Origins for Cancers of Unknown Primary. *Nature*. 3 juin 2021;594(7861):106-10.
48. Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut*. juin 2021;70(6):1183-93.
49. Dodington DW, Lagree A, Tabbarah S, Mohebpour M, Sadeghi-Naini A, Tran WT, et al. Analysis of tumor nuclear features using artificial intelligence to predict response to neoadjuvant chemotherapy in high-risk breast cancer patients. *Breast Cancer Res Treat*. avr 2021;186(2):379-89.
50. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. oct 2018;24(10):1559-67.
51. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat Commun*. déc 2020;11(1):3877.
52. Bejnordi BE, Veta M, Diest PJ van, Ginneken B van, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases

in Women With Breast Cancer. *JAMA*. 12 déc 2017;318(22):2199-210.

53. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *The Lancet Oncology*. févr 2020;21(2):222-32.

54. Li J, Li W, Gertych A, Knudsen BS, Speier W, Arnold CW. An attention-based multi-resolution model for prostate whole slide image classification and localization. arXiv:190513208 [cs, eess] [Internet]. 30 mai 2019 [cité 10 sept 2020]; Disponible sur: <http://arxiv.org/abs/1905.13208>

55. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. *JAMA Netw Open*. 6 nov 2019;2(11):e1914645.

56. Courtiol P, Tramel EW, Sanselme M, Wainrib G. Classification and Disease Localization in Histopathology Using Only Global Labels: A Weakly-Supervised Approach. arXiv:180202212 [cs, stat] [Internet]. 20 févr 2020 [cité 23 mars 2020]; Disponible sur: <http://arxiv.org/abs/1802.02212>

57. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *npj Digit Med*. déc 2020;3(1):23.

58. Zhang L, Le Lu, Nogues I, Summers RM, Liu S, Yao J. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE J Biomed Health Inform*. nov 2017;21(6):1633-43.

59. Wei JW, Suriawinata AA, Vaickus LJ, Ren B, Liu X, Lisovsky M, et al. Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on Histopathologic Slides. *JAMA Netw Open*. 23 avr 2020;3(4):e203398.

60. Kubach J, Muhlebner-Fahrngruber A, Soylemezoglu F, Miyata H, Niehusmann P, Honavar M, et al. Same same but different: A Web-based deep learning application revealed classifying features for the histopathologic distinction of cortical malformations. *Epilepsia*. 20 févr 2020;epi.16447.

61. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. juill 2019;25(7):1054-6.

62. Naik N, Madani A, Esteva A, Keskar NS, Press MF, Ruderman D, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat Commun*. déc 2020;11(1):5727.

63. Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep-learning on histological slides. *Hepatology*. 28 févr 2020;hep.31207.

64. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol*. nov 2019;16(11):703-15.

65. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst, Man, Cybern*. janv 1979;9(1):62-6.

66. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images. arXiv:200409666 [cs, eess, q-bio] [Internet]. 21 mai 2020 [cité 30 nov 2021]; Disponible sur: <http://arxiv.org/abs/2004.09666>

67. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. 2016 [cité 28 avr 2022]; Disponible sur: <https://arxiv.org/abs/1608.06993>

68. Huh M, Agrawal P, Efros AA. What makes ImageNet good for transfer learning? 2016 [cité 28 avr 2022]; Disponible sur: <https://arxiv.org/abs/1608.08614>
69. Chen T, Kornblith S, Norouzi M, Hinton G. A Simple Framework for Contrastive Learning of Visual Representations. arXiv:2002.05709 [cs, stat] [Internet]. 30 juin 2020 [cité 20 févr 2022]; Disponible sur: <http://arxiv.org/abs/2002.05709>
70. Ilse M, Tomczak JM, Welling M. Attention-based Deep Multiple Instance Learning. arXiv:1802.04712 [cs, stat] [Internet]. 28 juin 2018 [cité 20 févr 2022]; Disponible sur: <http://arxiv.org/abs/1802.04712>
71. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. JAMA Netw Open. 6 nov 2019;2(11):e1914645.
72. Tu HH, Lin HT. One-sided Support Vector Regression for Multiclass Cost-sensitive Classification. :8.
73. Chung YA, Lin HT, Yang SW. Cost-aware Pre-training for Multiclass Cost-sensitive Deep Learning. arXiv:1511.09337 [cs] [Internet]. 24 mai 2016 [cité 28 avr 2022]; Disponible sur: <http://arxiv.org/abs/1511.09337>
74. Delaune A, Valmary-Degano S, Loménie N, Zryouil K, Benyahia N, Trassard O, et al. Le premier data challenge organisé par la Société Française de Pathologie : une compétition internationale en 2020, un outil de recherche en intelligence artificielle pour l'avenir ? Annales de Pathologie. janv 2022;S0242649821002467.
75. Nguyen A, Dosovitskiy A, Yosinski J, Brox T, Clune J. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. arXiv:1605.09304 [cs] [Internet]. 23 nov 2016 [cité 28 avr 2022]; Disponible sur: <http://arxiv.org/abs/1605.09304>
76. Lubran di Scandalea M, Lazard T, Balezo G, Bellahsen-Harrar Y, Badoual C, Berlemont S, et al. Automatic grading of cervical biopsies by combining full and self-supervision [Internet]. Cancer Biology; 2022 janv [cité 18 avr 2022]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2022.01.14.476330>

## Annexe

---

### Détails de numérisation

En moyenne, une lame de biopsie pèse 346 MB et une lame de pièce opératoire 354 MB. Les lames les plus lourdes pesaient 1,2 Go. Le jeu de données pèse au total 1,68 To. Les premières lames ont été numérisées en mode semi-automatique. Ce mode nécessite le contrôle des points de focus, ce qui rend le processus beaucoup plus long car chaque lame est pré-numérisée pour cela. Devant ce problème de temps de scan, nous avons demandé au support technique de Hamamatsu si la numérisation en mode automatique (sans contrôle des points de focus) était adaptée à nos prélèvements, ce qui a été confirmé. Ainsi le reste du jeu de données a été numérisé en mode automatique.

A la fin de chaque numérisation, les lames floues sont indiquées par un point d'exclamation à côté de la vignette de la lame. Il est possible de contrôler avec le logiciel de visualisation des lames virtuelles NDPView la qualité de l'image et de re-numériser uniquement les lames imparfaites. Ces lames floues ont été refaites en semi-automatique en plaçant des points de focus sur les zones floues.

## Permis d'imprimer

---

Vu, le Président de thèse  
Université .....  
[Professeur/Docteur] Prénom NOM

Vu, le Doyen de l'UFR de Médecine  
Université de Paris  
Professeur Philippe RUSZNIEWSKI

Date :

Vu, la Présidente d'Université de Paris  
Professeure Christine CLERICI

Pour la Présidente et par délégation,  
Le Doyen Philippe RUSZNIEWSKI