



UNIVERSITÉ DE LILLE
FACULTÉ DE MÉDECINE HENRI WAREMBOURG
Année : 2021

THÈSE POUR LE DIPLÔME D'ÉTAT
DE DOCTEUR EN MÉDECINE

**L'Intelligence Artificielle au service du diagnostic histologique
d'adénocarcinome sur matériel de cytoponction pancréatique**

Présentée et soutenue publiquement le 10 novembre à 16 h
au Pôle Recherche

par Oriane KARLESKIND

JURY

Président :

Monsieur le Professeur Claude Alain MAURAGE

Assesseurs :

Madame le Docteur Florence RENAUD

Monsieur le Docteur Romain GERARD

Directrice de thèse :

Madame le Professeur Emmanuelle LETEURTRE

AVERTISSEMENT

La faculté n'entend donner aucune approbation aux opinions émises dans les thèses : celles-ci sont propres à leurs auteurs.

Liste des abréviations

ACP	Anatomie Cytologie pathologiques
ADCP	Adénocarcinome du pancréas
CHU	Centre Hospitalier Universitaire
CNIL	Commission Nationale de l'Informatique et des Libertés
CZI	Carl Zeiss Image
DL	Deep learning
FN	Faux Négatif
FNA	Fine Needle Aspiration
FNB	Fine Needle Biopsy
FP	Faux Positif
HES	Hématoxyline Eosine Safran
IA	Intelligence Artificielle
INSEE	Institut National de la Statistique et des Études Économiques
IRM	Imagerie par Résonance Magnétique
ISEN	Institut Supérieur de l'Électronique et du Numérique
ML	Machine Learning
NT	Non Tumoral
PNG	Portable Network Graphic
RGPD	Règlement Général sur la Protection des Données
RNA	Réseau de Neurones Artificiels
RNAC	Réseau de Neurones Artificiels à Convolution
TDM	Tomodensitométrie
TNE	Tumeur Neuroendocrine
VN	Vrai Négatif
VP	Vrai Positif

Tables des matières

Résumé	1
Introduction	2
1. L'adénocarcinome du pancréas.....	2
1.1 <i>Un enjeu de santé publique</i>	2
1.2 <i>Un enjeu diagnostique</i>	3
2. Le Deep Learning	10
2.1 <i>Définitions</i>	10
2.2 <i>Les réseaux de neurones artificiels</i>	12
2.3 <i>Applications dans le domaine de l'anatomie cytologie pathologiques</i>	16
3. Objectifs de notre étude.....	17
Matériels et méthodes	18
1. Création de la base de données	18
1.1 <i>Inclusion des lames</i>	18
1.2 <i>Information des patients</i>	19
1.3 <i>Numérisation et anonymisation</i>	20
1.4 <i>Segmentation et filtre</i>	21
1.5 <i>Labellisation</i>	22
1.6 <i>Transfert des données</i>	24
2. Phase d'entraînement.....	26
2.1 <i>Apprentissage</i>	29
2.1.1 <i>Augmentation artificielle des données</i>	30
2.1.2 <i>Transfert d'apprentissage</i>	31
2.2 <i>Validation</i>	32
3. Phase de test.....	33
4. Mise au point d'un logiciel d'aide à l'interprétation du pathologiste	34
Résultats	35
1. Base de données.....	35
2. Phase d'entraînement.....	38
3. Phase de test.....	41
4. Logiciel d'aide à l'interprétation du pathologiste	43
Discussion	45
1. La donnée : la lame numérique	45
2. La procédure de labellisation.....	47
3. Les biais d'apprentissage	51
4. Perspectives	52
Conclusion	55
Annexe 1 : note d'information pour les patients.....	56
Annexe 2 : exemples de tuiles labellisées	59
Bibliographie	60

Résumé

Contexte :

L'adénocarcinome représente 90 % des tumeurs du pancréas et devrait être la deuxième cause de mortalité par cancer aux Etats-Unis en 2030. La cytoponction par voie écho-endoscopique est la technique non chirurgicale de référence pour le diagnostic des tumeurs pancréatiques. Le déploiement de la pathologie numérique offre de nouvelles perspectives innovantes pour le diagnostic, notamment grâce au Deep Learning. Cette méthode d'Intelligence Artificielle repose sur des algorithmes capables de mimer le cerveau humain grâce à des réseaux de neurones artificiels.

Méthode :

Trois réseaux de neurones artificiels (VGG-19, ResNet-50 et DenseNet-201) ont été entraînés à partir de 68 lames du CHU de Lille, après information des patients. Des techniques de transfert d'apprentissage et/ou d'augmentation artificielle des données ont été appliquées.

À l'issue de la phase d'entraînement, nous avons calculé le taux de bonne prédiction pour chaque réseau de neurones en fonction de chaque stratégie d'apprentissage. Pour la phase de test, un jeu supplémentaire de 4 lames d'adénocarcinome a été prédit par le meilleur algorithme. Le taux de bonne prédiction a été calculé à l'échelle des lames entières puis à l'échelle des tuiles. Pour évaluer l'applicabilité en routine de notre algorithme, nous avons également calculé la précision (valeur prédictive positive) et le rappel (sensibilité).

Un logiciel d'aide à l'interprétation du pathologiste a été conçu pour faciliter la lecture du pathologiste.

Résultats :

Lors de la phase d'entraînement, le taux de bonne prédiction de l'algorithme DenseNet-201 avec augmentation artificielle des données sans transfert d'apprentissage était de 96,5 %.

À l'issue de la phase de test, DenseNet-201 a obtenu un taux de bonne prédiction de 100 % à l'échelle des lames entières. À l'échelle des tuiles, le taux de bonne prédiction était de 69,3 % avec une précision de 13,7 % et un rappel de 84,6 %.

Conclusion :

Les performances de notre algorithme ne permettent pas encore l'application en routine clinique, néanmoins de nombreuses pistes d'amélioration sont envisageables et réalisables.

L'Intelligence Artificielle a toute sa place dans le diagnostic d'adénocarcinome du pancréas. Cela ouvre des perspectives pour l'ensemble des petits prélèvements de cytoponction d'organe profond.

Introduction

1. L'adénocarcinome du pancréas

1.1 Un enjeu de santé publique

L'adénocarcinome du pancréas (ADCP) représente, avec ses variantes histologiques, la très grande majorité des tumeurs solides primitives du pancréas (90 %). Son incidence est en augmentation sans explication claire. En France, ces vingt dernières années, elle a triplé : en 1997, on recensait 4 500 nouveaux cas annuels, 9 040 en 2011 et 14 184 en 2018, ce qui en fait le second cancer digestif dans notre pays. Malgré les progrès de la prise en charge thérapeutique tant sur le plan médical que chirurgical, la mortalité reste élevée chez les patients tous stades confondus. Il est la troisième cause de décès par cancer en France après le cancer du poumon et le cancer du côlon (1). Selon une étude épidémiologique, il pourrait devenir la seconde cause de mortalité par cancer aux Etats Unis en 2030 (2). Le taux de survie globale à 5 ans de suivi est de 9 % et son taux de létalité est le plus élevé parmi toutes les tumeurs solides (3).

Certains facteurs de risque ont été clairement établis : le tabac, l'obésité et le syndrome métabolique, le diabète, la pancréatite chronique calcifiante. Les lésions précancéreuses reconnues sont : les tumeurs kystiques mucineuses, les néoplasies intra-épithéliales pancréatiques et les tumeurs intracanalaires papillaires et mucineuses du pancréas (1).

1.2 Un enjeu diagnostique

Le délai entre le début des symptômes et la mise en route d'un traitement a un impact sur la survie. Malheureusement, les symptômes de l'ADCP sont peu spécifiques. En effet, les symptômes les plus fréquents sont l'asthénie, l'anorexie et la perte de poids, marquant le début de la maladie chez 50 % des patients et présents au diagnostic chez environ 85 % d'entre eux. Les autres signes fréquents sont les douleurs abdominales (80 %) et les signes en rapport avec l'obstruction biliaire (55 %). L'ADCP peut être fortement suggéré par la conjonction d'éléments cliniques (âge, terrain, ictère, altération de l'état général), biologique (CA 19-9) et les résultats de l'imagerie en coupe (TDM et/ou IRM) (4). Néanmoins, ces différents éléments ne permettent pas de poser le diagnostic de certitude d'ADCP.

La cytoponction à l'aiguille fine par voie écho-endoscopique est la technique non chirurgicale de référence pour le diagnostic des tumeurs pancréatiques. Cette technique nécessite un échoendoscope ainsi qu'une aiguille dédiée. Une fois la lésion détectée sur l'image ultrasonographique, la ponction est effectuée sous guidage en temps réel en s'assurant de l'absence de vascularisation interposée par Doppler (5). Elle est indiquée dans tous les cas où une prise en charge thérapeutique optimale nécessite une preuve histologique (6).

Au laboratoire d'anatomie cytologie pathologiques (ACP) du Centre Hospitalier Universitaire de Lille (CHU de Lille), le matériel biopsique recueilli fait l'objet d'une technique de cytologie sur couche mince (ThinPrep®) avec coloration par le Papanicolaou. S'il reste du matériel, il est inclus en paraffine en totalité pour une étude histologique avec la coloration Hématoxyline Eosine Safran (HES).

Sur la lame de cytologie, le pathologiste traque les anomalies cyto-nucléaires en faveur de l'ADCP :

- des amas tridimensionnels de cellules cohésives (figure 1) ;
- des noyaux cellulaires qui se chevauchent (figure 1) ;
- des membranes nucléaires aux contours irréguliers et densifiés (figure 1) ;
- une chromatine hétérogène (figure 1) ;
- des nucléoles volumineux (figure 1) ;
- une augmentation du rapport nucléo-cytoplasmique (figure 1) ;
- des figures de mitoses (condition rare en cytologie) (figure 2) ;
- un fond « sale » avec de la nécrose cellulaire et des débris nucléaires (figure 3) ;
- la présence de mucosécrétion intracytoplasmique (figure 4).

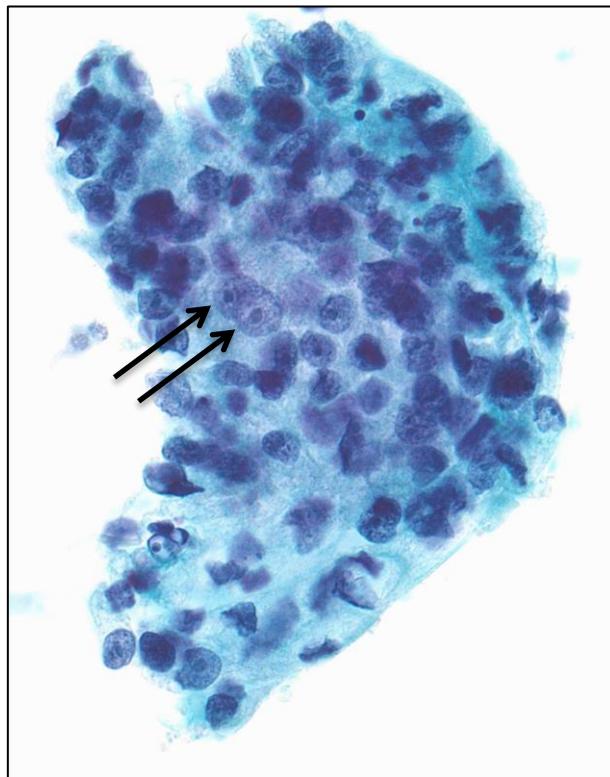


Figure 1. Amas cellulaire tridimensionnel, atypies cyto-nucléaires, nucléoles proéminents (flèches noires) au grandissement x 400, coloration de Papanicolaou

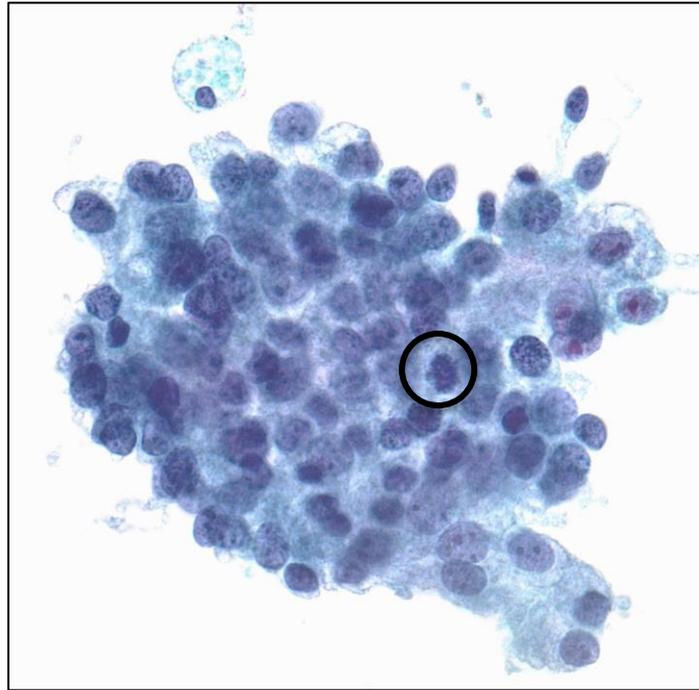


Figure 2. Figure de mitose (cercle noir) sur lame de cytologie au grossissement x 400, coloration de Papanicolaou

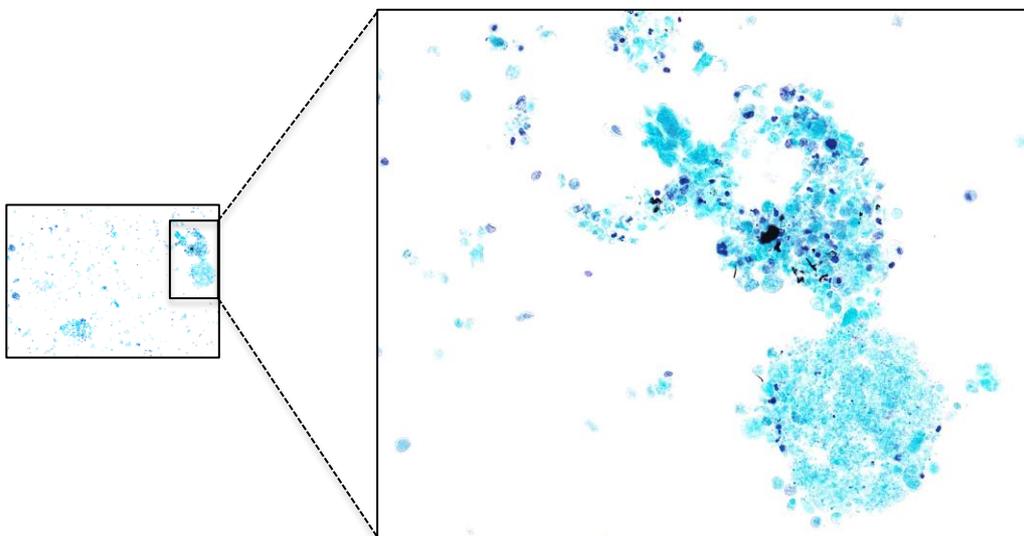


Figure 3. Fond nécrotico-inflammatoire sur lame de cytologie aux grossissements x 100 et x 200, coloration de Papanicolaou

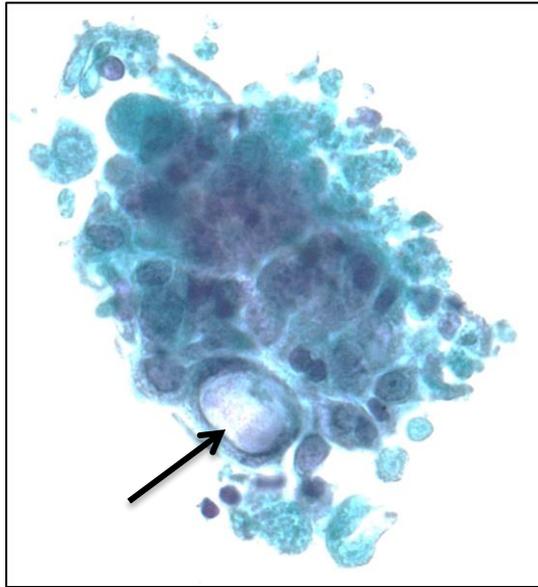


Figure 4. Vacuole de mucus intracytoplasmique (flèche noire) au grandissement x 400, coloration de Papanicolaou

Sur la lame d'histologie, le pathologiste recherche :

- une désorganisation architecturale du parenchyme pancréatique (figures 6 et 7) ;
- des formations canales, glandulaires ou tubulaires plus ou moins irrégulières au sein d'un stroma desmoplasique (figures 5, 6 et 7) ;
- des atypies cyto-nucléaires avec des nucléoles proéminents (figure 8) ;
- des vacuoles de mucus intracytoplasmique (figure 9) ;
- des engainements périnerveux (très rare) (figure 10).

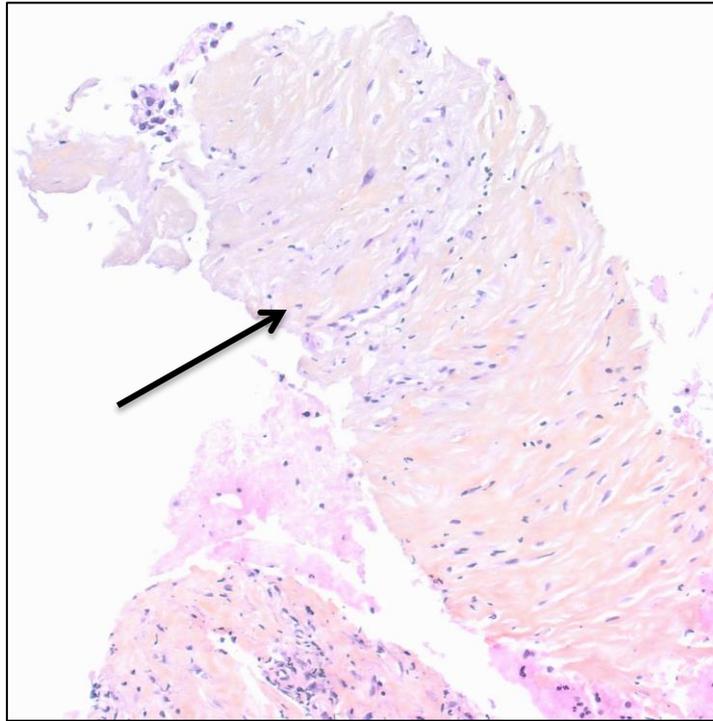


Figure 5. Stroma (flèche noire) au grandissement x 100, coloration HES

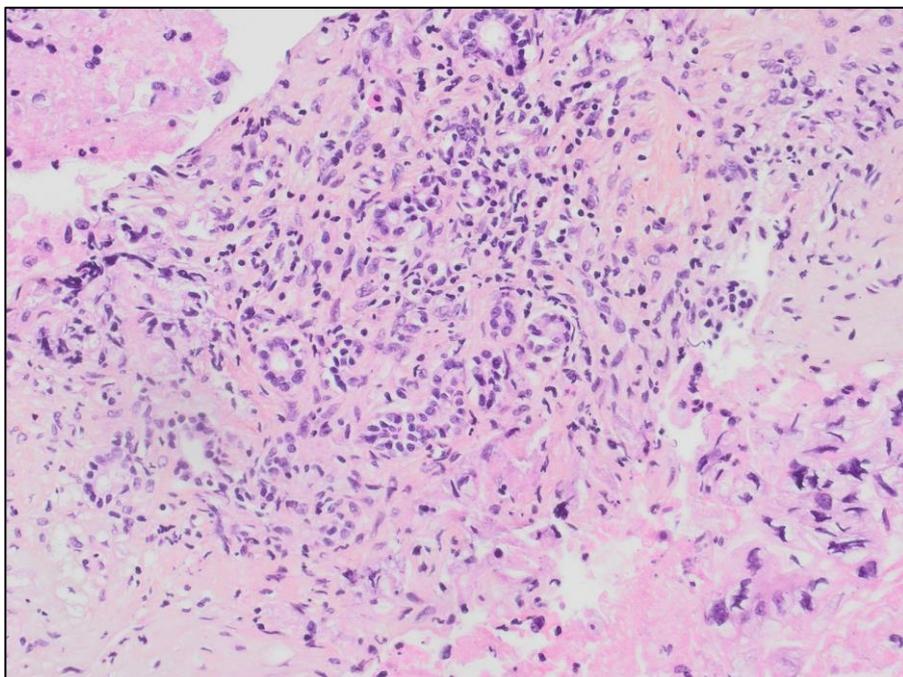


Figure 6. Désorganisation architecturale du parenchyme pancréatique au grandissement x 200, coloration HES

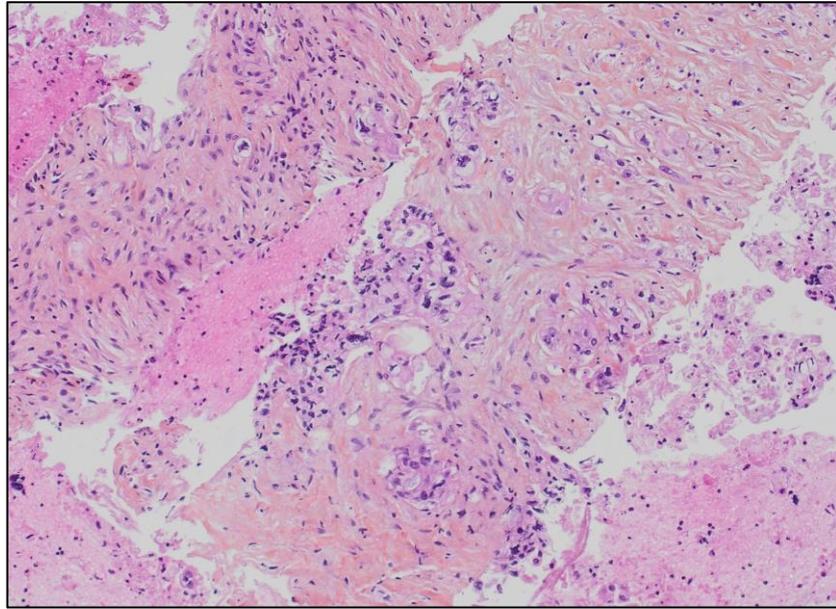


Figure 7. Désorganisation architecturale et stroma au grossissement x 100, coloration HES

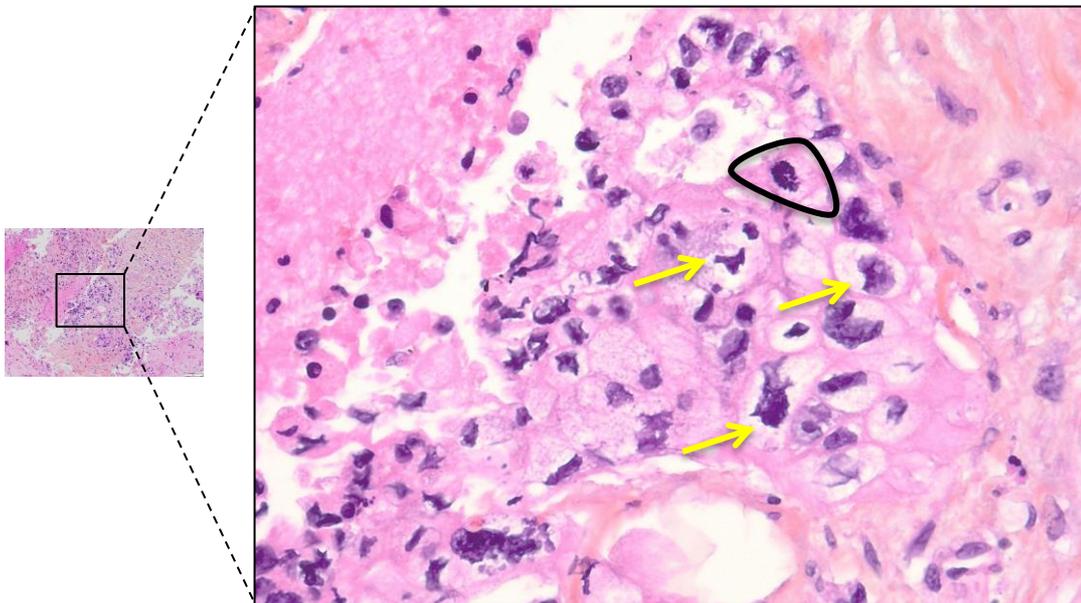


Figure 8. Anisocaryose (flèches jaunes) et figure de mitose (triangle noir) au grossissement x 400, coloration HES

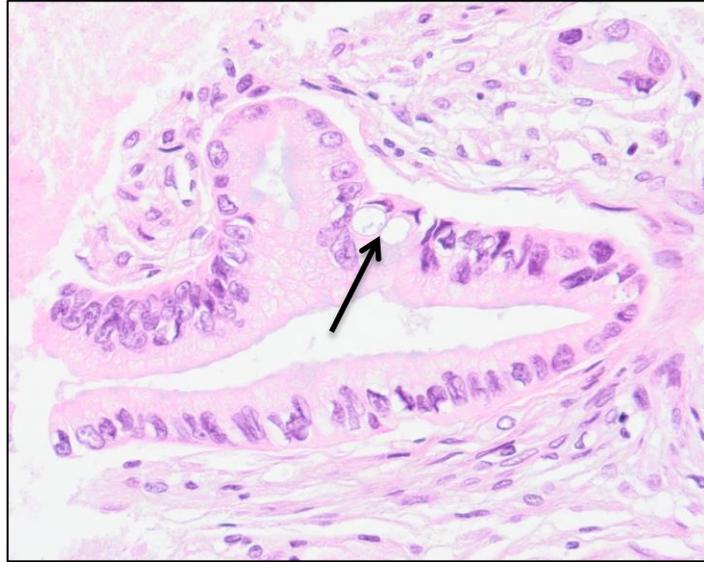


Figure 9. Vacuoles de mucus intra-cytoplasmiques avec noyaux déjetés en périphérie (flèche noire) au grandissement x 400, coloration HES

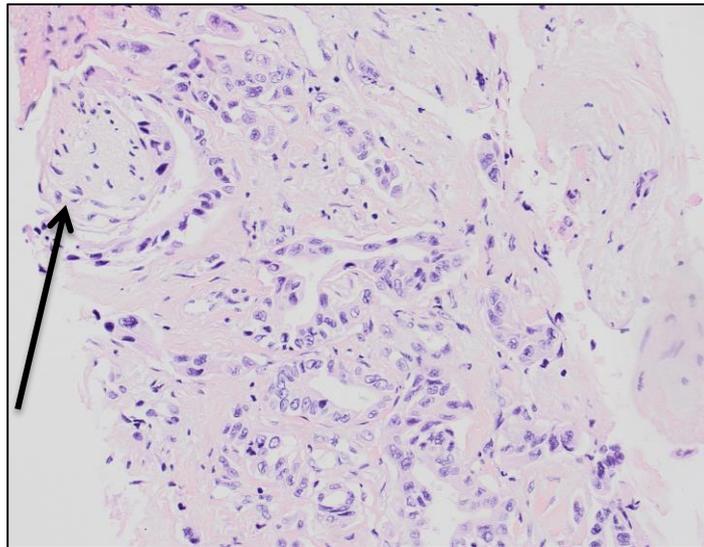


Figure 10. Engainement périnerveux (flèche noire) au grandissement x 200, coloration HES

La recherche de tous ces éléments, souvent focaux, requiert une analyse fine et méthodique ainsi qu'une expérience dans le domaine de la pathologie pancréatique. L'œil et l'avis d'un pathologiste expert sont souvent nécessaires pour conclure au diagnostic d'ADCP.

2. Le Deep Learning

Si l'expression « Deep Learning » (DL) est apparue pour la première fois au début des années 2000, le concept d'utilisation des réseaux de neurones artificiels (RNA) pour permettre aux ordinateurs de prendre des décisions intelligentes, remonte à plusieurs décennies (7).

Les domaines d'application du DL sont très variés et complexes : traitement d'images et de sons (reconnaissance faciale, reconnaissance automatique de la parole pour la transformation d'une voix en texte écrit), vision par ordinateur (imiter la vision humaine), traitement automatique du langage, classification de texte (détection de spams). On peut citer quelques applications connues dont on se sert dans la vie quotidienne : Facebook, Google Traduction, Siri... Les quantités d'applications potentielles sont infinies (8).

2.1 Définitions

Au préalable, il est important de définir l'Intelligence artificielle (IA) et le Machine Learning (ML).

L'Intelligence Artificielle est une discipline qui réunit des sciences, théories et techniques (notamment logique mathématique, statistiques, probabilités, neurobiologie computationnelle et informatique) dont le but est de parvenir à faire imiter par une machine les capacités cognitives d'un être humain (9). Autrement dit, elle est le résultat d'un processus effectué par une machine (terme « artificielle ») en suivant une logique (terme « intelligence »), toutes deux émanant de l'Homme. À chaque fois que la machine exécute ce processus, elle enrichit son « savoir » et

augmente ainsi graduellement sa capacité autonome à effectuer des processus qui suivent des logiques de plus en plus élaborées. Le but n'est pas de copier le cerveau humain, mais de s'inspirer de son fonctionnement cognitif afin de trouver de nouvelles solutions à une large variété de problèmes. Si le terme et le concept sont nés en 1950 dans un article publié par Alan Turing, mathématicien britannique, son application n'aura jamais été aussi développée que depuis ces dernières années (10). Elle s'applique à tous les secteurs d'activités : santé, transports, logistique, industrie, énergie, finance ou encore commerce...

Le Machine Learning (apprentissage automatique en français) est un sous-domaine de l'IA. Il se fonde sur des approches mathématiques et statistiques (des algorithmes) qui donnent aux ordinateurs la capacité « d'apprendre » à partir d'un grand nombre de données afin de prédire des résultats sans avoir été explicitement programmés (11). Le nombre de données doit être assez conséquent pour déduire statistiquement une fonction de prédiction. La particularité de ces algorithmes est que l'on ne sait expliquer comment ils arrivent au résultat. En effet, l'aspect fondamental de cette technique est que l'utilisateur ne spécifie pas les règles du domaine modélisé. Le Deep Learning (apprentissage profond en français) est un sous-type de ML qui s'appuie sur les réseaux de neurones artificiels (figure 11).

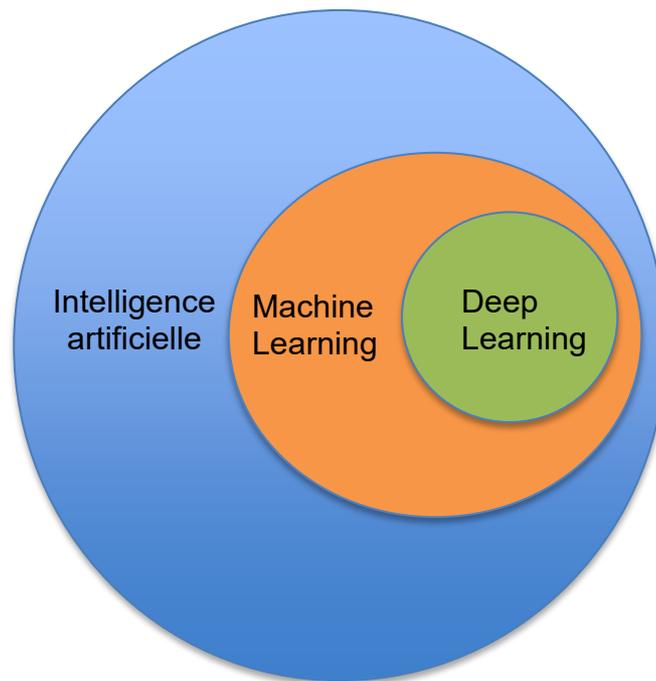


Figure 11. Intelligence Artificielle, Machine Learning et Deep Learning

Les algorithmes de Deep Learning, et plus largement de ML, sont qualifiés de « supervisés » ou « non supervisés ». L'apprentissage supervisé consiste à établir une fonction de prédiction à partir de données annotées tandis que l'apprentissage « non supervisé » s'appuie sur des données brutes. A ce jour, l'apprentissage supervisé est le plus performant et se trouve donc plus utilisé. Le DL profite de l'ère du Big Data avec des accès accrus à des données massives mais également de l'explosion de la puissance de calcul disponible.

2.2 Les réseaux de neurones artificiels

En 1943, le neurophysicien Warren McCullough et le mathématicien Walter Pitts de l'Université de Chicago se sont inspirés des propriétés électriques des neurones du cerveau humain pour élaborer le concept de neurone « artificiel ». Le neurone « biologique » a la particularité de produire, sous certaines conditions, un signal

électrique et de le transmettre. C'est par ce signal électrique que se fait la diffusion du message nerveux. Le neurone artificiel n'a pas de réalité physique ni de propriété biologique. Il est un processeur élémentaire qui fonctionne comme suit : il reçoit un nombre variable d'entrées en provenance des neurones situés en amont (équivalent des synapses du neurone). A chacune de ces entrées est associée un « poids » représentatif de la force de connexion. Dans un premier temps, il calcule la somme pondérée des entrées. Dans un second temps, une fonction de transfert calcule la valeur de l'état du neurone (équivalent du corps cellulaire du neurone). Cette valeur est transmise grâce à une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones situés en aval (équivalent de l'axone pour le neurone) (figure 12).

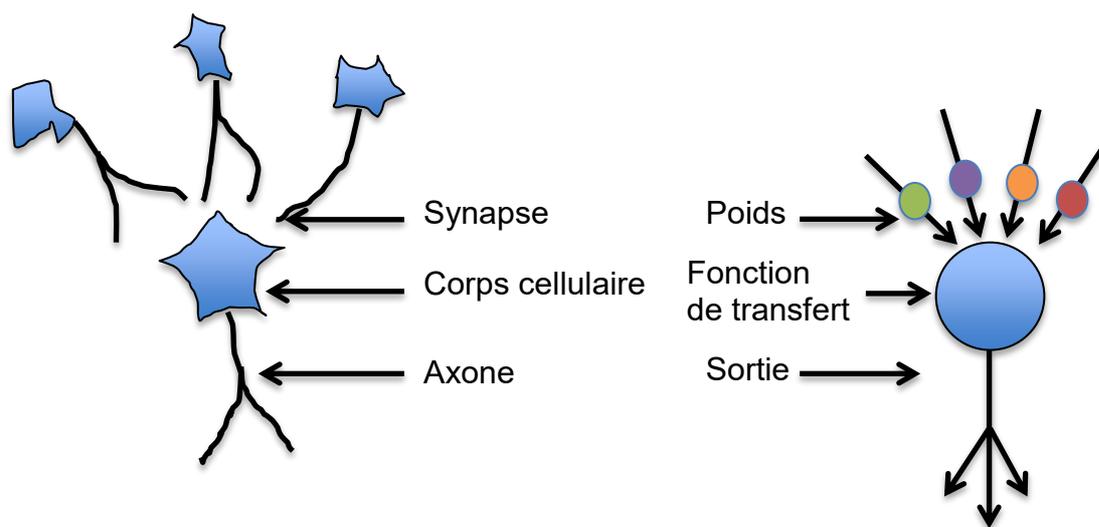


Figure 12. Correspondances neurone « biologique » (à gauche) / neurone « artificiel » (à droite)

Chaque fois que des nouvelles informations sont intégrées, les connexions existantes entre les neurones sont susceptibles d'être modifiées et étendues. Ceci permet au système d'apprendre sans intervention humaine, de manière autonome, tout en améliorant la qualité des prises de décision et de ses prévisions (7).

Plus le problème à résoudre est complexe, plus le RNA doit comporter de couches de neurones successives. Un RNA profond peut contenir des milliers voire des millions de neurones organisés en couches interconnectées. Les connexions entre les neurones qui composent le réseau définissent l'architecture du RNA.

Il y a différents types de RNA : à propagation avant, récurrent ou convolutif. Ici, nous ne détaillerons que les RNA convolutifs (RNAC) car ils sont devenus les architectures de référence dans la quasi totalité des tâches de DL appliquées à la reconnaissance d'images.

Les premiers RNAC à succès ont été mis au point à la fin des années 1990 par le Professeur Yann LeCunn (12) récompensé en 2019 par le prix Turing, équivalent du « prix Nobel » de l'informatique. À l'origine, une convolution est un outil mathématique très utilisé en retouche d'image car il permet de faire ressortir des caractéristiques grâce à un filtre (mise en évidence des traits verticaux, horizontaux, le flou de l'image, etc.).

Les RNAC sont composés de trois couches complémentaires :

- couche de convolution,
- couche de pooling,
- couche fully-connected.

La couche de convolution est la composante clé des RNAC et constitue toujours au moins leur première couche. Elle agit comme un « détecteur de formes ». Dans un réseau multicouche, les premières strates de convolution apprennent des formes élémentaires (par exemple : des bordures obliques, des bordures verticales

etc.). Plus on avance dans les couches du réseau, plus les formes apprises sont complexes (par exemple : des visages humains).

La couche de pooling (couche de mise en commun en français) est souvent placée entre deux couches de convolution. Cette couche a pour objectif de résumer l'information. Le but est de garder uniquement les caractéristiques majeures de notre image. De cette manière, l'image diminue en taille et se trouve simplifiée. Cette méthode permet de réduire une image de grande taille tout en préservant les informations les plus importantes qu'elle contient. Cela permet de diminuer la charge de calcul. La succession des différentes couches de convolution et de pooling permet d'extraire les caractéristiques pertinentes des images.

La couche fully connected (couche totalement connectée en français) constitue toujours la dernière couche d'un RNA, convolutif ou non. Elle n'est donc pas caractéristique de ce type de RNA. Elle permet de classifier l'image reçue en entrée. Tous les RNAC commencent par une couche de convolution et finissent par une couche fully-connected. Les couches intermédiaires peuvent s'empiler de différentes manières, à condition que la sortie d'une couche ait la même structure que l'entrée de la suivante. Plus il y a de couches, plus le réseau de neurones est « profond ». C'est cette grande quantité de neurones qui sont disposés en couches profondes et interconnectées qui confère la puissance de calcul aux RNAC.

Un autre concept clé est la rétropropagation du gradient de l'erreur. Dans le cadre d'un apprentissage supervisé, il est possible de comparer la prédiction de l'algorithme à la réalité. Ainsi, il devient possible de calculer le gradient de l'erreur

pour chaque neurone d'un RNA, de la dernière couche vers la première. Ensuite il convient de corriger les erreurs selon l'importance de la contribution de chaque élément aux erreurs. Les poids qui contribuent le plus à une erreur seront modifiés de manière plus importante que les poids qui provoquent une erreur marginale.

2.3 Applications dans le domaine de l'anatomie cytologie pathologiques

Permettre aux ordinateurs de reconnaître des objets dans des images était jusqu'à récemment considéré comme une tâche très difficile. Cependant, depuis peu, les RNAC ont surpassé les performances humaines.

Des applications de DL spécialisées dans l'analyse d'images médicales ont commencé à apparaître. Le nombre de communications a augmenté rapidement dans les années 2015-2016 et le sujet est aujourd'hui prédominant. En effet, le domaine de la santé génère énormément de données et le ML se révèle être l'un des meilleurs moyens d'intégrer, d'analyser et de faire des prédictions basées sur des vastes ensembles de données hétérogènes.

Les applications sont multiples : prévision d'évènements médicaux (accidents vasculaires cérébraux et arrêts cardiaques), aide au diagnostic, aide à la prise de décision clinique et au pronostic, aide à la sélection du traitement, etc. (13). De nombreux articles ont été publiés sur la détection des diverses tumeurs d'origine pancréatique ou biliaire grâce à l'IA, et plus particulièrement dans le domaine de l'imagerie médicale (14). Depuis plusieurs années déjà, les images de scanner et d'échographie sont largement utilisées pour la détection de l'ADCP grâce à des algorithmes de reconnaissance d'images (15-21).

En ACP, le DL va probablement modifier les pratiques des pathologistes dans le futur. En effet, grâce à la numérisation des lames de plus en plus rapide et performante, l'utilisation des outils de DL se développe afin d'optimiser le diagnostic et sa précision (22). Aujourd'hui, quelques articles scientifiques rapportent de très bons résultats, prometteurs pour l'avenir (23-28). La recherche dans ce domaine est stimulée par l'innovation avec l'émergence de start-up qui collaborent avec les laboratoires d'ACP. Des start-up françaises comme OWKIN se spécialisent dans l'IA au service de la recherche médicale ou PRIMAA dont l'objectif est de développer des outils logiciels permettant d'assister les pathologistes dans les diagnostics de cancer grâce à l'IA. Très récemment, PRIMAA a développé un outil d'aide au diagnostic pour le cancer du sein : CLEO. Avec 200 millions d'images labellisées par des médecins spécialistes en ACP, il s'agit de la plus grande base de données au monde sur le cancer du sein. Ils ont collaboré avec 6 laboratoires d'ACP français et européens et une trentaine de pathologistes pour l'annotation des lames (29).

3. Objectifs de notre étude

L'objectif principal de notre étude est de développer un algorithme de DL pour l'aide au diagnostic d'ADCP sur lame de cytoponction pancréatique.

L'objectif secondaire est de développer un logiciel d'aide à l'interprétation du pathologiste.

Matériels et méthodes

Il s'agit d'une étude rétrospective et monocentrique. Le service d'ACP du CHU de Lille a travaillé en collaboration avec l'Institut Supérieur de l'Électronique et du Numérique (ISEN) et le plateau de microscopie photonique BICeL du Campus Hospitalo-Universitaire de Lille.

1. Création de la base de données

1.1 Inclusion des lames

Sur la période du 1^{er} janvier 2017 au 31 janvier 2019, les examens de cytoponction pancréatique enregistrés dans le service d'ACP ont été recensés à l'aide du logiciel DIAMIC. Ils ont été inclus en fonction du diagnostic posé par le pathologiste.

Les critères d'inclusion étaient :

- diagnostic d'adénocarcinome du pancréas ;
- absence d'élément suspect de malignité.

Les critères de non-inclusion étaient :

- diagnostic de métastase ;
- diagnostic de tumeur neuro-endocrine (TNE) ;
- diagnostic de kyste pancréatique ;
- diagnostic de pancréatite auto-immune.

Les lames ont été désarchivées. La lame de cytologie n'a pas été analysée. Seule la lame d'inclusion en paraffine colorée par l'HES a été exploitée. Par conséquent, les examens pour lesquels le matériel biopsique n'était pas assez abondant ont été exclus.

1.2 Information des patients

Les données de santé sont des données à caractère personnel et considérées comme sensibles (30). A ce titre, elles font l'objet d'une protection particulière par les textes de lois afin de garantir le respect de la vie privée des personnes. Le règlement général sur la protection des données (RGPD) en application depuis le 25 mai 2018 a renforcé les droits des individus par rapport à l'utilisation de leurs données dans tous les domaines et pour tous les Etats membres de l'Union Européenne. La commission nationale de l'informatique et des libertés (CNIL) a élaboré des cadres de référence permettant de guider les organismes dans la mise en conformité du traitement des données. Ils comprennent des référentiels et des méthodologies de référence. Pour notre étude, la méthodologie de référence est MR-004. Elle encadre le traitement des données à caractère personnel à des fins d'études, évaluation ou recherche n'impliquant pas la personne humaine. Il s'agit plus précisément des études portant sur la réutilisation de données (23).

Conformément au référentiel de la méthodologie de référence MR-004, le patient a été informé de manière individuelle avec la possibilité de s'opposer à la recherche et à l'utilisation de ses données. Une note d'information expliquant le but de notre étude et ses intérêts a été envoyée à chaque patient encore vivant en novembre 2020 (annexe 1). Pour ce faire, nous avons consulté le logiciel SILLAGE pour obtenir les adresses postales des patients et le fichier des personnes décédées depuis 1970 disponible sur le site internet de l'Institut National de la Statistique et des Études Économiques (INSEE). En parallèle, une déclaration informatique a été transmise au délégué à la protection des données pour le CHU de Lille.

1.3 Numérisation et anonymisation

L'acquisition des images issues des lames HES a été réalisée grâce au scanner Axio Scan.Z1 (figure 13) de la marque ZEISS (Carl Zeiss Micro Imaging GmbH, 37081 Göttingen) disponible sur le plateau de microscopie photonique BICeL du campus hospitalo-universitaire de Lille (plateforme Lilloise en Biologie & Santé Université de Lille 1). La caméra était une Hamamatsu Orca Flash 4. Le logiciel de numérisation était le logiciel ZEN 2.6 (objectif 20x/0.8).



Figure 13. Scanner de lames Axio Scan.Z1. Crédits photos Zeiss.

Les images issues des lames étaient sous le format Carl Zeiss Image (CZI). Ce format a été développé par la marque ZEISS pour répondre spécifiquement aux exigences de l'imagerie en microscopie. Dans un second temps, elles ont été converties au format Portable Network Graphics (PNG). Ce format avait plusieurs avantages : il permettait de réduire la taille du fichier et de conserver la qualité même après compression.

Un numéro d'anonymat a été attribué à chaque fichier numérique afin de ne plus pouvoir identifier le patient, conformément au référentiel et à la méthodologie de référence MR-004 (31).

1.4 Segmentation et filtre

Une fois numérisées et anonymisées, les images ont été segmentées en territoires de taille égale de 96 x 96 pixels soit 9216 pixels au total, appelées « tuiles ». Chaque tuile était nommée avec le numéro d'anonymat de la lame d'origine et sa position au sein de celle-ci (pour exemple : lame 1 crop_5_30.png).

Toutes les tuiles ainsi créées ne contenaient pas toujours d'information pertinente : les tuiles acellulaires ou correspondantes aux bords de la lame (figure 14). L'équipe de l'ISEN a conçu un programme informatique afin de supprimer ces tuiles. Pour ce faire, un traitement d'image a été appliqué en associant une valeur de niveau de gris à chaque pixel des tuiles. Puis, les pixels ont été classés en deux catégories en fonction de leur niveau de gris (seuillage empirique) : pixel noir ou pixel blanc. Si le nombre total de pixels noirs ou de pixels blancs au sein d'une tuile était supérieur à 8000 (soit 87% des pixels), alors la tuile était considérée comme noire ou blanche. Ainsi, ces dernières ont été supprimées de notre jeu de données.



Figure 14. Tuiles non pertinentes

a. Bord de la lame

b. Tuile acellulaire

1.5 Labellisation

Les pathologistes — expert (Professeur Emmanuelle Leteurtre) et junior (Oriane Karleskind) — ont convenu d’annoter les tuiles en cinq classes :

- non tumoral ;
- adénocarcinome ;
- probable ;
- nécrose ;
- autre.

Le label « non tumoral » était destiné aux tuiles sans élément suspect de malignité.

Le label « adénocarcinome » était destiné aux tuiles sans équivoque. Le stroma sans cellule tumorale a été labellisé « non tumoral ». En effet, seul, il n’est pas suffisant pour conclure au diagnostic « adénocarcinome ».

Le label « probable » a été créé pour les tuiles difficiles à classer entre « adénocarcinome » et « non tumoral ».

Le label « nécrose » était destiné aux tuiles renfermant des cellules en voie de lyse ou lysées.

Le label « autre » a été créé pour annoter la tuile au cas par cas si besoin grâce à un encart de texte libre.

Un programme informatique a été imaginé et conçu spécialement pour faciliter la labellisation des tuiles avec le logiciel ImageJ. Sur l’écran d’ordinateur, le programme affichait simultanément deux fenêtres informatiques : l’une pour la tuile à annoter et l’autre permettant de choisir le label adéquat (figure 15).

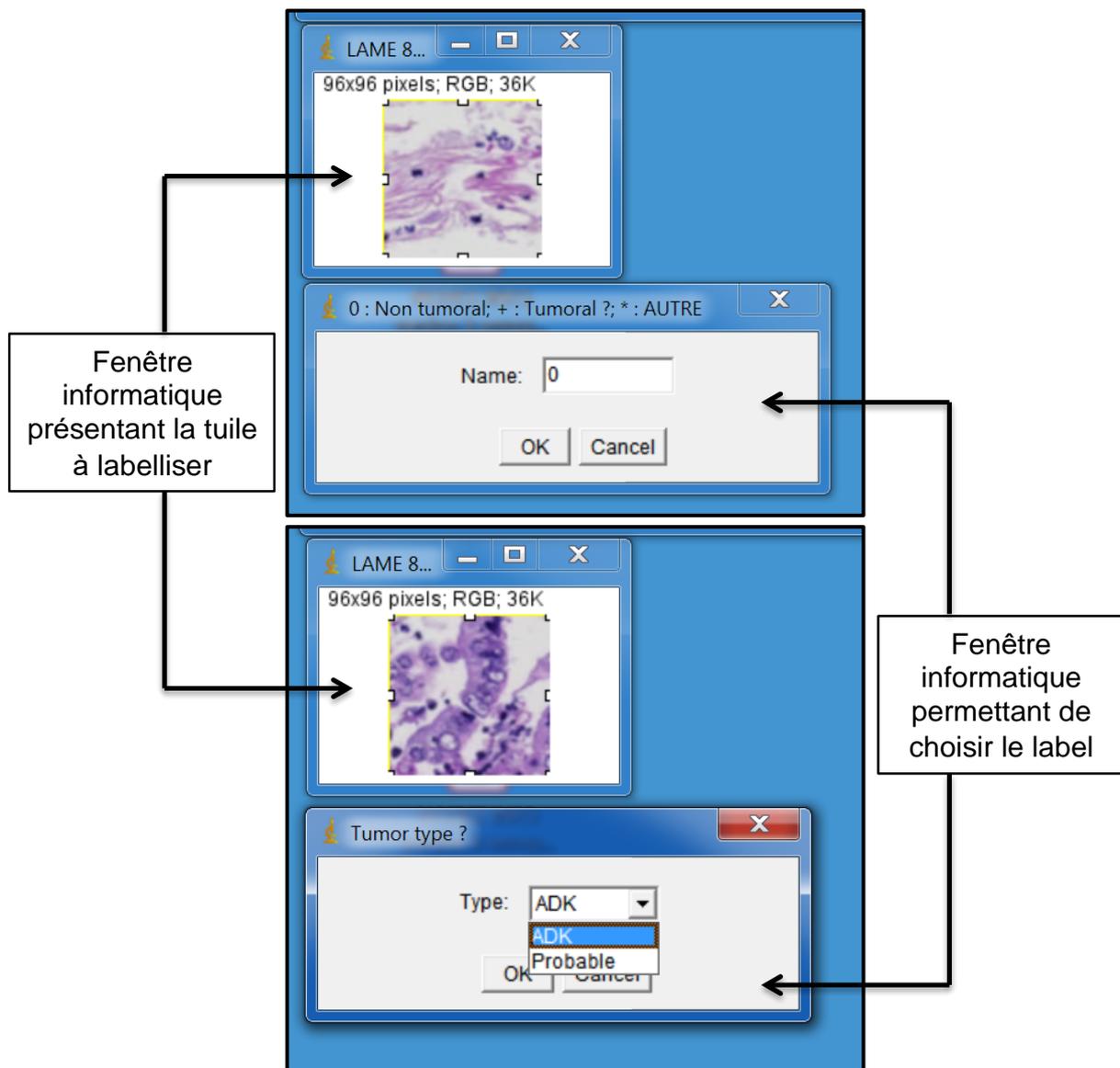


Figure 15. Etape de labellisation

Une copie de la tuile était automatiquement générée dans un nouveau dossier (« Renamed ») avec le label directement intégré au nom du fichier (par exemple : lame 1crop_5_30.png-NON TUMORAL). Cette étape a été réalisée par le pathologiste junior.

1.6 Transfert des données

Le transfert des données entre les différents acteurs du projet s'est fait selon plusieurs modalités, toujours en accord avec le référentiel et la méthodologie de référence MR-004 (31).

Tous les échanges électroniques relatifs aux données de la recherche ont été envoyés par le biais de messageries électroniques sécurisées. Après la numérisation des lames et leur anonymisation, elles ont été stockées sur le serveur IMGENSTOCK du plateau de microscopie photonique BICeL du campus hospitalo-universitaire de Lille. Le dossier contenant les fichiers a été sécurisé par un mot de passe. Il a ensuite été transmis à l'ISEN via le site informatique CHUTransfert qui a permis de transférer les fichiers de façon sécurisée et chiffrée à travers le réseau. Lors du téléchargement des fichiers, un mot passe a été demandé au préalable. Tous les fichiers téléchargés et stockés temporairement sur CHUTransfert (suppression automatique après 30 jours) sont conservés sur du matériel appartenant au CHU et exploité par celui-ci.

Une fois les tuiles créées et filtrées par l'équipe de l'ISEN sur un serveur sécurisé et protégé, le pathologiste junior a récupéré l'ensemble du jeu de données sur un disque dur externe sécurisé par un mot de passe.

Une fois la labellisation effectuée, le pathologiste junior a transféré les tuiles annotées via le site informatique CHUTransfert vers l'ISEN.

Ces différentes étapes ont été résumées ci-après (figure 16).

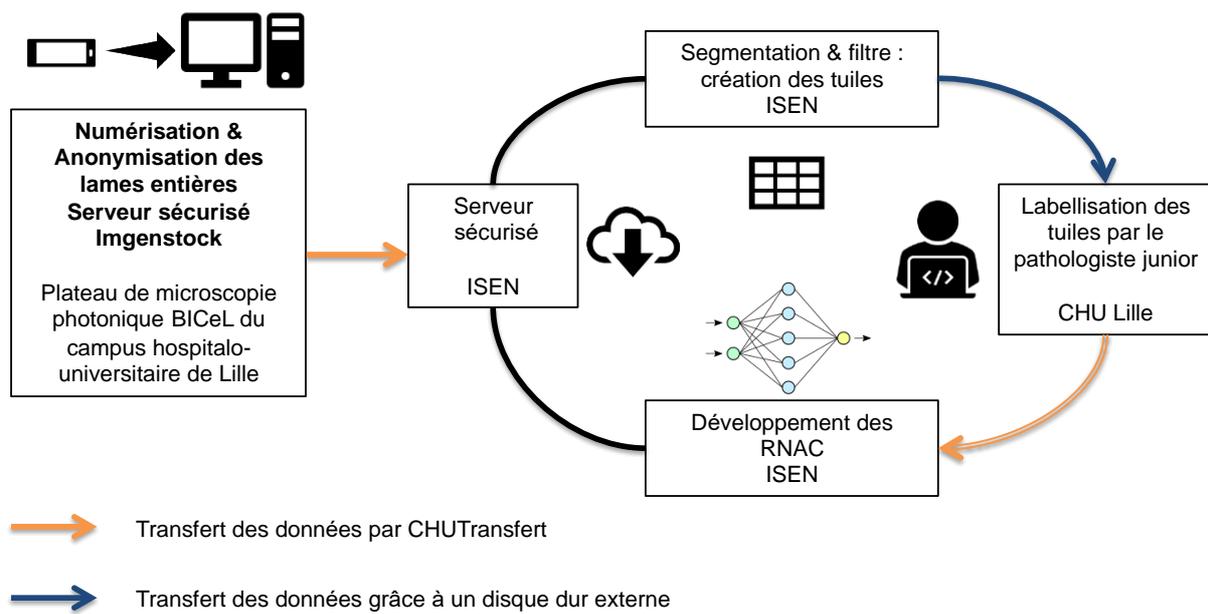


Figure 16. Transferts des données

2. Phase d'entraînement

La phase d'entraînement de des RNAC comporte toujours deux étapes distinctes : l'étape d'apprentissage et l'étape de validation. Dans le cadre d'un apprentissage supervisé, les RNAC ont besoin d'un jeu de données volumineux pour lequel les proportions de classes sont relativement similaires. C'est-à-dire qu'il est nécessaire d'avoir un nombre similaire et suffisant de tuiles dans chacune de nos catégories. Si ce n'est pas le cas, le risque est qu'ils apprennent un biais vis-à-vis de la classe majoritaire. Ils auraient alors tendance à toujours prédire cette dernière. C'est pourquoi nous ne retenons que les catégories « adénocarcinome » et « non tumoral » car dans les catégories « probable », « nécrose » et « autre » le nombre de tuiles n'était pas suffisant (moins de 2000 tuiles par catégorie au total).

Le jeu de données a été d'abord équilibré — autant de tuiles « adénocarcinome » que de tuiles « non tumoral » — puis scindé en deux — un pour l'étape d'apprentissage (75%) et un pour l'étape de validation (25%) (figure 17).

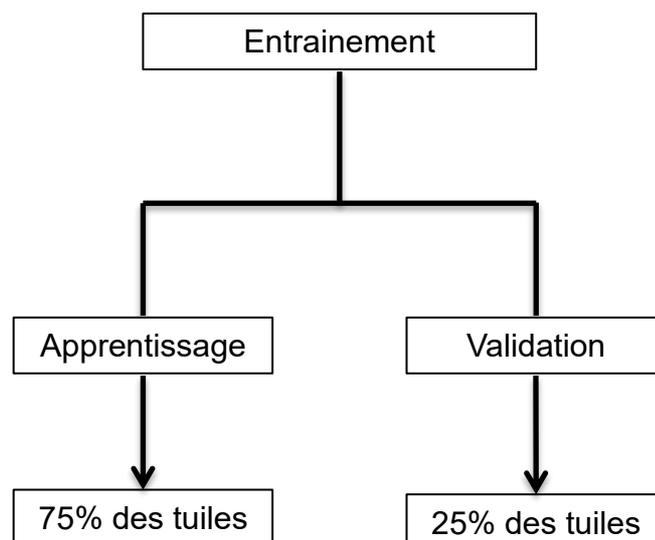


Figure 17. Création du jeu de données de la phase d'entraînement

Trois RNAC ont été entraînés :

- VGG-19 ;
- ResNet-50 ;
- DenseNet-201.

VGG-19 est une architecture développée par le Groupe de Géométrie Visuelle de l'Université d'Oxford en 2014. Elle se compose de 19 couches profondes organisées de manière linéaire : chaque couche est uniquement connectée à la couche suivante. Cette architecture a déjà été utilisée sur des images histo-pathologiques pour reconnaître le carcinome mammaire avec de très bons résultats (24).

Resnet-50 est une architecture développée par Microsoft en 2015. Elle se compose de 50 couches profondes. Différent de VGG-19, ce type de réseau utilise des connexions ou des raccourcis pour sauter certaines couches. On parle de bloc résiduel (figure 18). Cette technique permet de faciliter la formation de réseaux qui sont beaucoup plus profonds. Ils offrent de meilleures capacités d'apprentissage et peuvent gagner en précision (32).

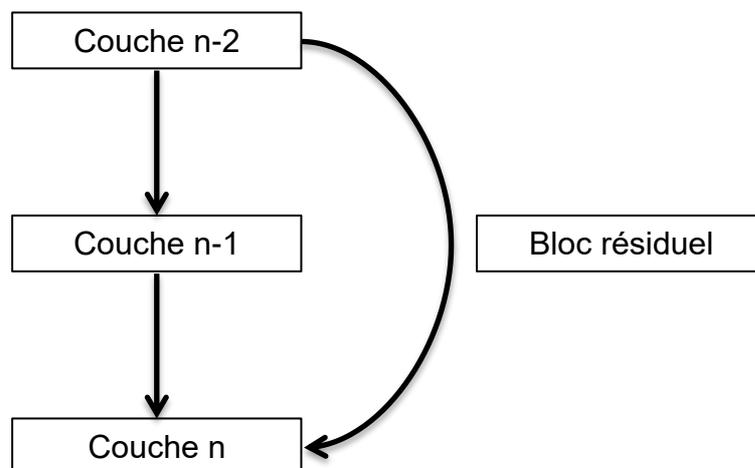


Figure 18. Schéma explicatif d'un RNAC résiduel

DenseNet-201 se compose de 201 couches profondes. Il a été inventé conjointement par l'Université Cornwell, l'Université Tsinghua et Facebook AI Research (FAIR) en 2017. Il s'agit d'un « réseau dense » qui simplifie le modèle de connectivité entre les couches (33). Elles sont directement connectées les unes aux autres permettant un flux maximum d'informations. En effet, chaque couche reçoit l'entièreté des sorties des couches similaires précédentes. Cette particularité augmente le nombre de connexions du réseau (figure 19).

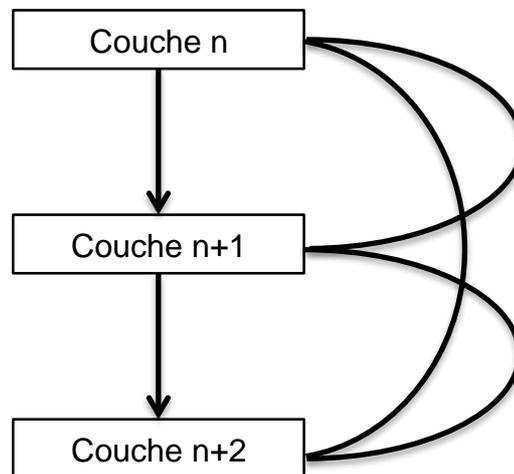


Figure 19. Schéma explicatif d'un RNAC dense

2.1 Apprentissage

Pour notre étude, l'apprentissage des RNAC est dit « supervisé ». Ils ont appris à partir d'exemples annotés que sont nos tuiles préalablement labellisées. Cette étape consiste à fournir aux RNAC :

- en entrée : les tuiles ;
- en sortie : le résultat attendu (label).

Elle se déroule comme suit : chaque tuile est analysée à travers les différentes couches du RNAC. Il fait une prédiction : « adénocarcinome » ou « non tumoral ». En fonction de la véracité de sa prédiction (en comparaison avec le label), il met à jour ses paramètres selon l'algorithme de rétro-propagation du gradient de l'erreur. Avec la totalité des tuiles dédiées à l'apprentissage, ce processus est répété plusieurs fois. En d'autres termes, l'apprentissage consiste à faire varier les paramètres du modèle jusqu'à ce que sa prédiction soit la plus proche possible de la labellisation. Le temps nécessaire pour construire un modèle dépend de la valeur initiale des paramètres, de l'algorithme utilisé pour les mettre à jour, de la taille de la base de données et enfin de la puissance de calcul. Dans notre travail, différentes stratégies d'apprentissage ont été expérimentées : avec ou sans augmentation artificielle du jeu de données ; avec ou sans transfert d'apprentissage (tableau I).

Stratégies d'apprentissage	Augmentation artificielle du jeu de données	Transfert d'apprentissage
Apprentissage 1	non	non
Apprentissage 2	non	oui
Apprentissage 3	oui	non
Apprentissage 4	oui	oui

Tableau I. Description des différentes stratégies d'apprentissage

2.1.1 Augmentation artificielle des données

Pour augmenter la diversité de nos données et donc le champ d'apprentissage de nos RNAC, une technique d'augmentation artificielle des données a été employée. Le principe repose sur la manipulation des tuiles pour générer de nouvelles images labellisées à partir de celles déjà disponibles. Cela peut être des transformations géométriques, des effets de couleurs, de luminosité ou de résolution, etc. Cet outil permet d'augmenter de manière artificielle notre jeu de données en réutilisant plusieurs fois la même image. Bien que le contenu soit le même, c'est techniquement une image différente pour les RNAC. Les transformations appliquées à notre jeu de données étaient les suivantes (figure 20) :

- rotation (0° ; 90°),
- retournement horizontal,
- retournement vertical,
- flou gaussien.

Le flou gaussien fait partie des filtres d'atténuation qui permettent d'uniformiser les parties d'une image en les floutant et donc en harmonisant les détails de celle-ci. Il atténue les changements brusques d'intensité. En effet, ce filtre agit sur chaque pixel de l'image en réglant sa valeur sur la moyenne de tous les pixels présents dans un rayon défini. L'adjectif « gaussien » renvoie à la fonction mathématique qui est utilisée pour réaliser le flou (34). Deux stratégies d'apprentissage ont utilisé cette technique : les apprentissages 3 et 4 (cf. tableau I).

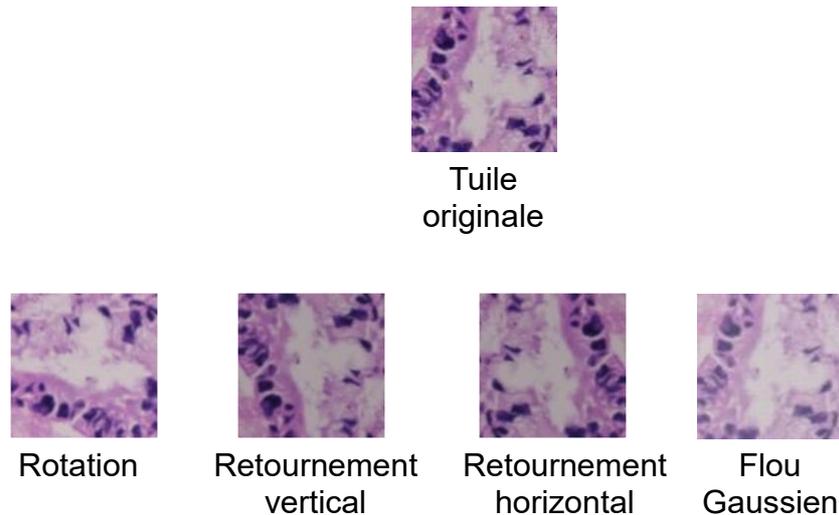


Figure 20. Transformations appliquées aux tuiles de notre jeu de données d'entraînement

2.1.2 Transfert d'apprentissage

Appliquer des connaissances acquises lors d'expériences passées lorsque nous sommes confrontés à une nouvelle tâche est la base du transfert d'apprentissage. Il s'agit d'un procédé souvent utilisé pour la reconnaissance d'images. Il consiste à appliquer des données apprises à partir de tâches antérieures sur de nouvelles tâches partageant des similitudes. Ceci est possible parce que les RNAC sont empilés en couche, chacune apprenant de la précédente : les premières couches sont spécialisées dans la reconnaissance de formes simples (lignes horizontales, lignes verticales etc.) alors que les dernières couches se consacrent à ce qui fait l'objet de l'apprentissage du réseau. Cette technique permet d'optimiser l'apprentissage des RNAC en mutualisant les connaissances d'un modèle à l'autre. Cela permet notamment de réduire le temps, les données et les ressources de calcul nécessaires à la formation d'un nouveau modèle (23).

ImageNet est une gigantesque base de données de plus de 14 millions d'images labellisées et réparties en plus de 1 000 classes (35). L'intérêt de disposer d'une large base de données pour l'apprentissage est de pouvoir entraîner les RNAC sur des tâches assez générales. Ainsi, un réseau de neurones pré-entraîné sur ce large jeu de données possède des caractéristiques de bas niveau (appprises dans les premières couches) potentiellement transférables à de nombreuses nouvelles tâches. Les RNAC auront déjà appris à travers les 14 millions d'images d'ImageNet à extraire des formes, textures et objets. Pour notre étude, deux stratégies d'apprentissage ont utilisé ce procédé : les apprentissages 2 et 4 (cf. tableau I).

2.2 Validation

Il n'est pas rare que les RNAC « sur-apprennent » le jeu de données d'apprentissage. Le phénomène de « sur-apprentissage » est l'hyperspécialisation de l'algorithme sur un jeu de donnée, ici les données d'apprentissage. Pour pouvoir le détecter, on a besoin d'un nouveau jeu de données : le jeu de validation. Ce sont des nouvelles tuiles qui n'ont pas été utilisées pour l'étape d'apprentissage mais qui ont été labellisées. Le but d'un RNAC est de prédire correctement des nouvelles tuiles sur lesquelles il ne s'est pas entraîné. On appelle « généralisation » la capacité d'un modèle à faire des prédictions correctes sur de nouvelles données, qui n'ont pas été utilisées pour le construire. Afin de détecter le sur-apprentissage et d'évaluer la qualité des modèles, nous avons calculé le taux de bonne prédiction sur le jeu de données de validation. Il correspond à la proportion de tuiles bien classées (nombre de prédictions correctes / nombre total de prédictions). Le modèle ayant obtenu le meilleur taux de bonne prédiction a été sélectionné pour la phase de test.

3. Phase de test

Cette étape permet d'évaluer la performance finale du modèle sélectionné à l'issue de la phase d'entraînement. Désormais l'algorithme est « figé » et n'ajuste plus ses paramètres. Les données de validation ne sont pas suffisantes pour apporter la preuve du bon pouvoir de généralisation d'un RNAC entraîné. En effet, le jeu de données utilisé lors de la phase d'entraînement peut être partiellement représentatif de l'objet d'apprentissage. Si la performance du RNAC est correcte sur les jeux de données de validation et de test, alors le modèle possède un bon pouvoir de généralisation sur des données inconnues. Pour ce faire, le meilleur modèle sélectionné à l'issue de la phase d'entraînement a été soumis à un nouveau jeu de données : le jeu de données test. Premièrement, nous avons calculé le taux de bonne prédiction à l'échelle des lames entières (nombre de lames bien prédites / nombre total de lames prédites). Deuxièmement, nous avons calculé le taux de bonne prédiction et à l'échelle des tuiles. Les prédictions ont été présentées à l'aide d'une matrice de confusion. Il s'agit d'un tableau croisé entre les valeurs réelles et les prédictions du modèle (tableau II) :

- Vrai Positif (VP) : la tuile est positive (adénocarcinome) et elle est prédite comme telle (adénocarcinome) ;
- Faux Négatif (FN) : la tuile est positive (adénocarcinome) mais a été prédite négative (non tumoral) ;
- Vrai Négatif (VN) : la tuile est négative (non tumoral) et elle est prédite comme telle (non tumoral) ;
- Faux Positif (FP) : la tuile est négative (non tumoral) mais elle est prédite comme positive (adénocarcinome).

		Meilleur modèle	
		non tumoral	adénocarcinome
Pathologiste	non tumoral	VN	FP
	adénocarcinome	FN	VP

Tableau II. Matrice de confusion

À partir de la matrice de confusion, deux nouveaux paramètres ont été calculés :

- - **la précision** (ou valeur prédictive positive) : elle correspond à la proportion d'identifications positives qui était effectivement correctes ($VP / VP + FP$).
- - **le rappel** (ou sensibilité). Elle correspond à la proportion d'identifications positives qui a été identifié correctement ($VP / VP + FN$).

Ces deux nouveaux paramètres ont permis de prendre en compte les deux types d'erreurs sur le plan quantitatif : les faux négatifs et les faux positifs. Sur le plan qualitatif, les tuiles correspondant aux faux positifs et aux faux négatifs ont été relues par les pathologistes expert et junior.

4. Mise au point d'un logiciel d'aide à l'interprétation du pathologiste

La finalité de notre travail était la mise au point d'un logiciel d'aide à l'interprétation du pathologiste. Sur la lame entière, toutes les tuiles prédites « adénocarcinome » par le RNAC devaient être visibles. L'objectif était de faire gagner du temps au pathologiste lors de la lecture de la lame pour poser le diagnostic d'adénocarcinome.

Résultats

1. Base de données

Au total, 108 cytoponctions pancréatiques ont été réalisées entre le 1^{er} janvier 2017 et le 31 janvier 2019. Vingt-et-un examens ne répondaient pas aux critères d'inclusion. Treize ont du être exclus en raison de l'absence d'inclusion en paraffine. Ainsi, 74 lames étaient disponibles dans le service d'ACP et éligibles pour notre étude (figure 21).

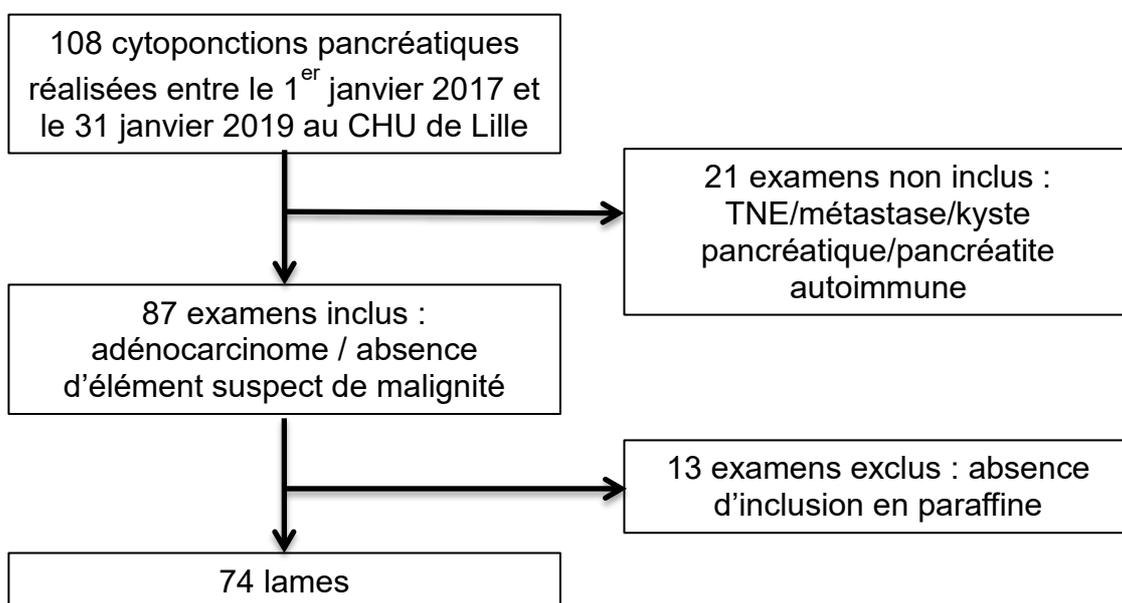


Figure 21. Sélection des lames

Sur les 70 patients (4 patients ont été ponctionnés deux fois), 32 étaient vivants en novembre 2020. La note d'information individuelle a été envoyée à leur domicile :

- Un seul courrier postal n'a pas pu être distribué : destinataire inconnu à l'adresse.
- Un seul patient s'est opposé grâce au cadre réservé au recueil de l'opposition.

Par conséquent, les lames de ces deux patients ont été exclues.

Aucun des patients décédés ne s'était opposé à la recherche de son vivant, ce qui a permis leur inclusion sans démarche complémentaire (figure 22).

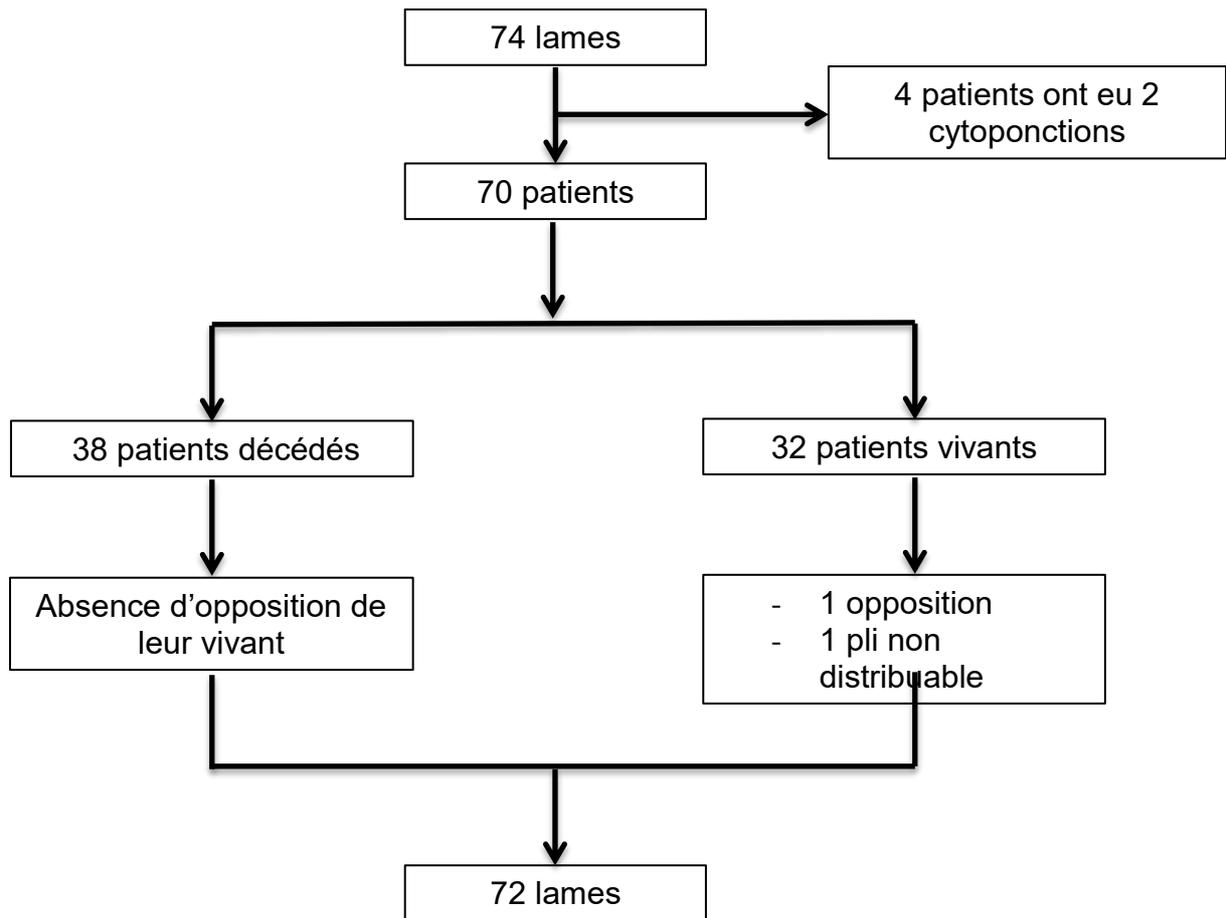


Figure 22. Information des patients

Une lame numérisée faisait en moyenne 9 400 x 9 700 pixels. Après segmentation, une lame pouvait générer jusqu'à 40 000 tuiles avec une moyenne de 12 338 tuiles par lame. Grâce au système de filtre, 60% des tuiles ont été éliminés avant la labellisation : 40% des tuiles contenaient du matériel à analyser.

Au total, 220 557 tuiles ont été labellisées (figure 23) :

- le label « non tumoral » a intéressé 205 483 tuiles (93%),
- le label « adénocarcinome » a intéressé 11 705 tuiles (5%),
- le label « probable » a intéressé 1 166 tuiles (1%),
- le label « nécrose » a intéressé 2 203 tuiles (1%).

Il n'a pas été utile de revoir les labels ou d'en ajouter. Le label « autre » n'a finalement pas été utilisé.

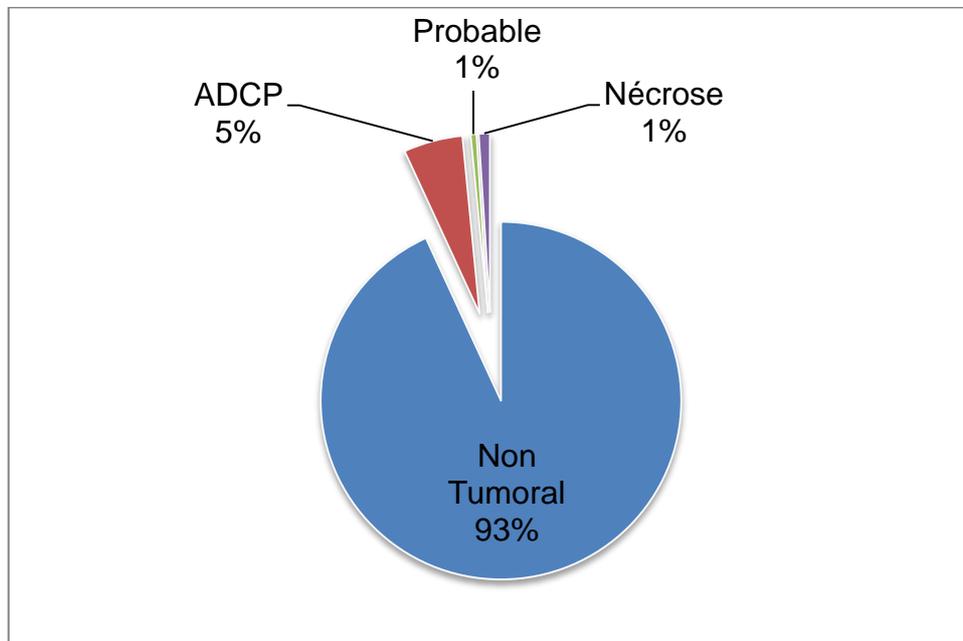


Figure 23. Tuiles labellisées

Des exemples de tuiles pour chaque label sont présentés en Annexe 2.

En raison d'un nombre trop faible de tuiles pour les catégories « probable » et « nécrose », il n'a pas été possible de les utiliser pour entraîner les RNAC. Elles n'ont donc pas été intégrées au jeu de données pour la phase d'entraînement.

2. Phase d'entraînement

Le jeu de données pour la phase d'entraînement a été constitué à partir de 68 lames soit 201 490 tuiles. Quarante six lames (68 %) ont fait l'objet d'un diagnostic d'ADCP. Vingt deux lames (32 %) étaient indemnes de cellule tumorale. Vingt-et-une lames (31 %) étaient issues d'examens réalisés au cours de l'année 2017. Quarante cinq lames (66 %) étaient issues d'examens réalisés au cours de l'année 2018. Deux lames (3 %) étaient issues d'examens réalisés au cours de l'année 2019. Ces cytoponctions pancréatiques ont été réalisées grâce à deux types d'aiguille : soit une aiguille de type Fine Needle Aspiration 22 G (FNA), soit une aiguille de type Fine Needle Biopsy 22 G de type Acquire® (FNB). Vingt-huit examens (41 %) ont été réalisés avec l'aiguille de type FNA et 40 examens (59 %) ont été réalisés avec l'aiguille de type FNB. Le label « non tumoral » a intéressé 190 664 tuiles (95 %). Le label « adénocarcinome » a intéressé 10 826 tuiles (5 %).

Ces données sont résumées dans le tableau III.

Caractéristiques des 68 lames			
Diagnostic	Adénocarcinome	46	(68 %)
	Non tumoral	22	(32 %)
Année	2017	21	(31 %)
	2018	45	(66 %)
	2019	2	(3 %)
Type d'aiguille	FNA	28	(41 %)
	FNB	40	(59 %)
Tuiles	Non tumoral	190 664	(95 %)
	Adénocarcinome	10 826	(5 %)

Tableau III. Phase d'entraînement

La classe « adénocarcinome » était minoritaire (10 826 tuiles au total). Pour éviter le biais de classe majoritaire, nous avons sélectionné au hasard 10 826 tuiles parmi les 190 664 tuiles « non tumoral ».

Les 21 652 tuiles (10 826 tuiles « adénocarcinome » + 10 826 tuiles « non tumoral ») ont été réparties de manière aléatoire pour les étapes d'apprentissage et de validation (figure 24). D'une part, 16 240 tuiles (75%) soit 8 120 tuiles « adénocarcinome » et 8 120 tuiles « non tumoral » ont été utilisées pour l'apprentissage. Avec la procédure d'augmentation artificielle des données, le nombre de tuiles a été multiplié par 5 soit 81 200 tuiles. D'autre part, 5 412 tuiles (25%) soit 2 706 tuiles « adénocarcinome » et 2 706 tuiles « non tumoral » ont été utilisées pour la validation.

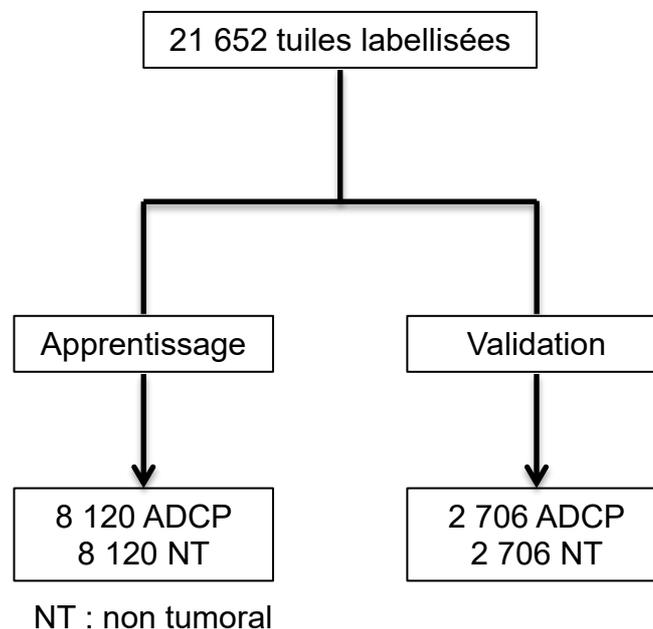


Figure 24. Phase d'entraînement

Pour chaque RNAC, la durée moyenne de la phase d'entraînement était de :

- 3 heures et 17 minutes pour VGG-19 ;
- 4 heures et 54 minutes pour ResNet-50 ;
- 5 heures et 32 minutes pour DenseNet-201.

Pour chaque RNAC et en fonction des différentes stratégies d'apprentissage, les taux de bonne prédiction sur le jeu de données de validation sont détaillés dans le tableau ci-dessous (tableau IV).

RNAC	Stratégies d'apprentissage	Taux de bonne prédiction
VGG-19	Apprentissage 1	91,9 %
	Apprentissage 2	92,3 %
	Apprentissage 3	92,3 %
	Apprentissage 4	93,4 %
ResNet-50	Apprentissage 1	92,5 %
	Apprentissage 2	93,3 %
	Apprentissage 3	90 %
	Apprentissage 4	93,4 %
DenseNet-201	Apprentissage 1	88,4 %
	Apprentissage 2	95,6 %
	Apprentissage 3	96,5 %
	Apprentissage 4	86,9 %

Tableau IV. Taux de bonne prédiction à l'issue de l'étape de validation

DenseNet-201 avec la stratégie d'apprentissage 3 (augmentation artificielle du jeu de données sans transfert d'apprentissage) a obtenu le meilleur taux de bonne prédiction (96,5%). VGG-19 et ResNet-50 avec la stratégie d'apprentissage 4 (augmentation artificielle du jeu de données et transfert d'apprentissage) ont chacun obtenu leur meilleur taux de bonne prédiction (93,4%). DenseNet-201 avec la stratégie d'apprentissage 4 a obtenu le taux de bonne prédiction le plus faible (86,9%).

3. Phase de test

DenseNet-201 avec la stratégie d'apprentissage 3 a été soumis au jeu de données de la phase de test. Il a été constitué à partir de 4 nouvelles lames soit 15 698 tuiles. Ces 4 lames ont fait l'objet d'un diagnostic d'ADCP. Deux d'entre elles étaient issues d'examens réalisés au cours de l'année 2018 et les deux autres étaient issues d'examens réalisés au cours de l'année 2019. Ces cytoponctions pancréatiques ont été réalisées grâce à un seul type d'aiguille : Fine Needle Biopsy 22 G de type Acquire® (FNB). Le label « non tumoral » a intéressé 14 819 tuiles. Le label « adénocarcinome » a intéressé 879 tuiles. Il est à noter que lors de la phase de test, le jeu de données est brut sans rééquilibrage des deux classes.

Ces données sont résumées dans le tableau V.

Caractéristiques des 4 lames			
Diagnostic	Adénocarcinome	4	(100 %)
Année	2018	2	(50 %)
	2019	2	(50 %)
Type d'aiguille	FNB	4	(100 %)
Tuiles	Non tumoral	14 819	(94 %)
	Adénocarcinome	879	(6 %)

Tableau V. Phase de test

À l'échelle des lames entières, DenseNet-201 avec la stratégie d'apprentissage 3 a obtenu un taux de bonne prédiction de 100 % (en prédisant une seule tuile « adénocarcinome », la lame est considérée comme telle).

À l'échelle des tuiles, le taux de bonne prédiction était de 69,3 %. Les erreurs de prédictions commises par notre modèle ont été analysées grâce à la matrice de confusion (tableau VI).

		DenseNet-201	
		non tumoral	adénocarcinome
Pathologiste	non tumoral	10 143 (VN)	4 676 (FP)
	adénocarcinome	135 (FN)	744 (VP)

Tableau VI. Matrice de confusion

Les faux positifs représentaient 30 % (4 676/15 698) des tuiles prédites. Après analyse des tuiles, dans la majorité des cas, elles correspondaient à du stroma sans cellule tumorale ou moins fréquemment des hématies lysées et des cellules inflammatoires (figures 25 et 26).

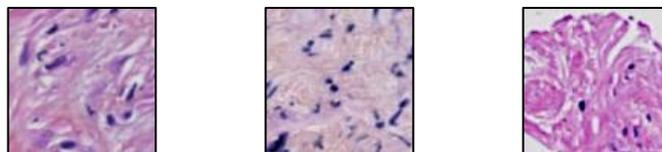


Figure 25. Tuiles de stroma sans cellule tumorale (FP)

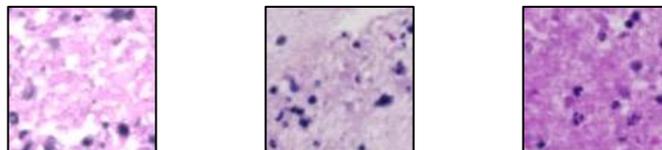


Figure 26. Tuiles de cellules inflammatoires et hématies lysées (FP)

Les faux négatifs représentaient 0,9 % (135/15 698) des tuiles prédites. Dans la majorité des cas, elles renfermaient des lambeaux d'épithélium pancréatique tumoraux sans stroma sous-jacent ou des cellules tumorales isolées (figure 27).

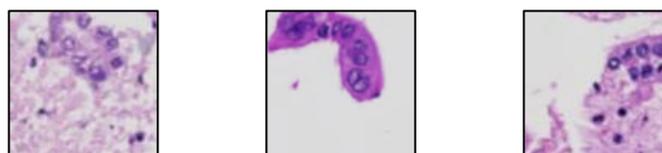


Figure 27. Faux négatifs

La précision (ou valeur prédictive positive) était de 13,7 % ($744 / 744 + 4\ 676$). Autrement dit : 13,7 % des tuiles prédites « adénocarcinome » étaient réellement positives.

Le rappel (ou sensibilité) était de 84,6 % ($744 / 744 + 135$). En d'autre terme, notre modèle a identifié correctement 84,6 % des tuiles « adénocarcinome ».

4. Logiciel d'aide à l'interprétation du pathologiste

Avec les 4 lames de la phase de test, nous avons mis au point le logiciel d'aide à l'interprétation du pathologiste. Sur la lame entière, le logiciel a repéré les tuiles pour lesquelles DenseNet-201 a prédit « adénocarcinome » (figures 28 et 29).

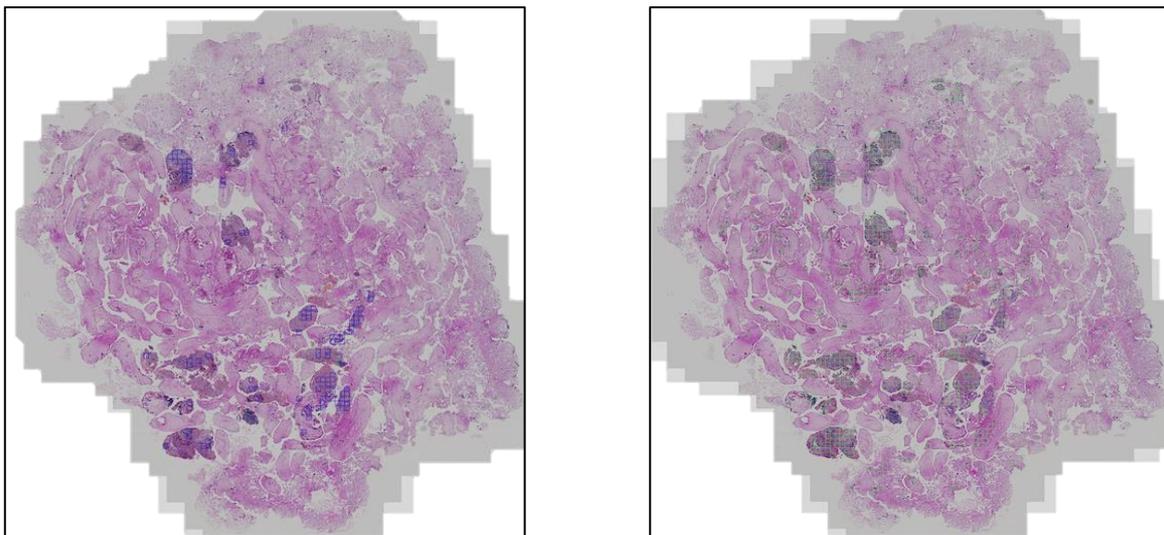


Figure 28. Lame 1 de la phase de test, vrais positifs (carrés bleus) et faux positifs (carrés verts) avec le logiciel d'aide à l'interprétation du pathologiste

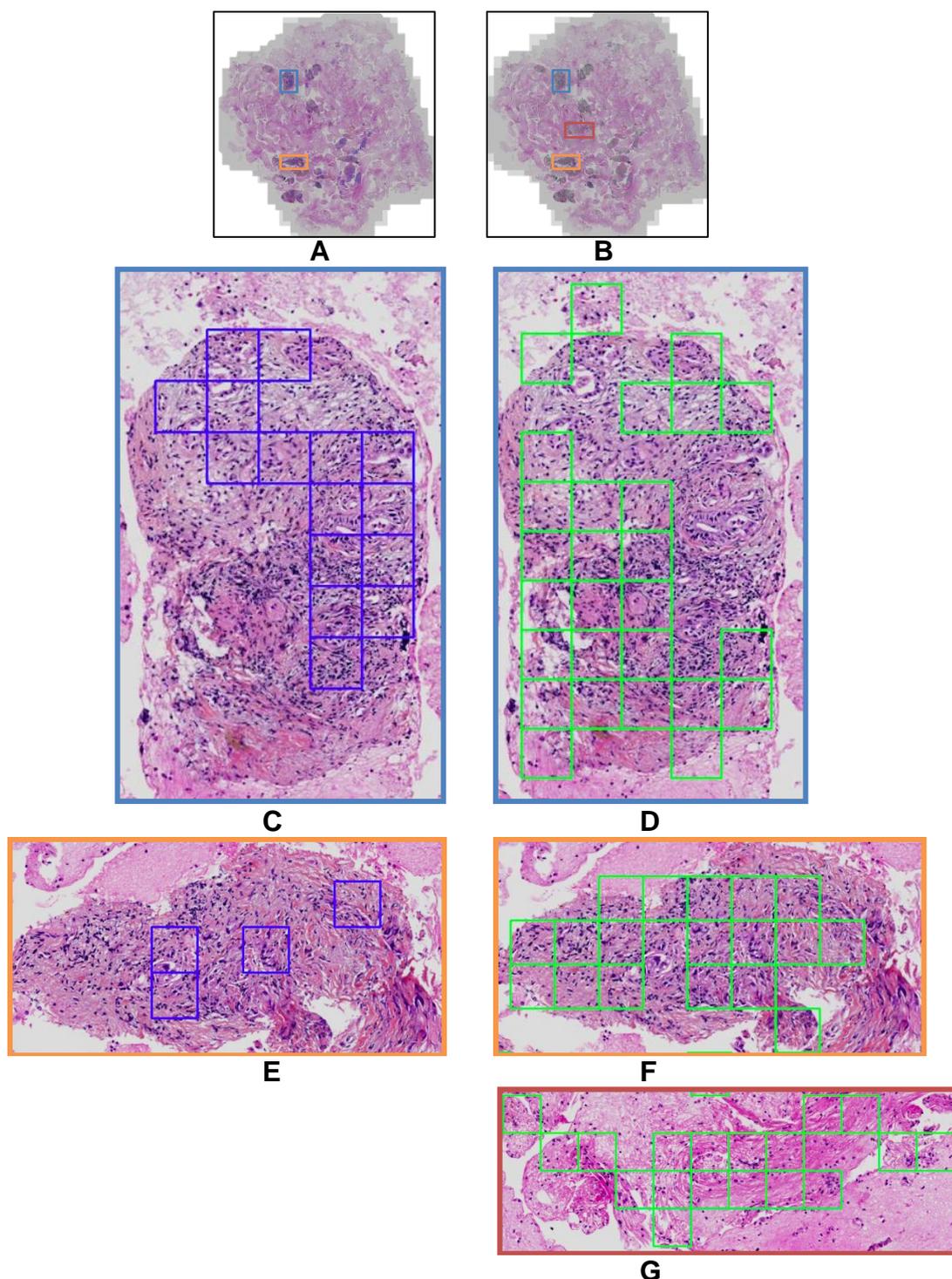


Figure 29 : lame 1 de la phase de test avec le logiciel d'aide à l'interprétation du pathologiste

- A. Vrais positifs à l'échelle de la lame entière
- B. Faux positifs à l'échelle de la lame entière
- C. Carrés bleus : vrais positifs
- D. Carrés verts : faux positifs, stroma sans cellule tumorale
- E. Carrés bleus : vrais positifs
- F. Carrés verts : faux positifs, stroma sans cellule tumorale
- G. Carrés verts : faux positifs : cellules inflammatoires et hématies

Discussion

Notre objectif principal était de développer un algorithme de DL pour l'aide au diagnostic d'ADCP sur lames de cytoponction. Nous avons entraîné 3 types de RNAC selon 4 stratégies d'apprentissage. À l'issue de la phase d'entraînement, nous avons sélectionné le RNA ayant obtenu le meilleur taux de bonne prédiction : DenseNet-201 avec la stratégie d'apprentissage 3 (augmentation artificielle du jeu de données, sans transfert d'apprentissage). Pour la phase de test, DenseNet-201 a été testé sur un nouveau jeu de données. D'une part, nous avons calculé le taux de bonne prédiction à l'échelle des lames entières ; d'autre part, nous avons calculé le taux de bonne prédiction, la précision et le rappel à l'échelle des tuiles.

Notre objectif secondaire était de mettre au point un logiciel présentant les zones prédites « adénocarcinome » par l'algorithme. Le but était de faire gagner du temps au pathologiste à la lecture.

Dans un premier temps, notre discussion va examiner différents facteurs ayant pu impacter la prédiction de l'algorithme et par conséquent nos résultats. Dans un second temps, elle considérera les perspectives de ce travail de recherche.

1. La donnée : la lame numérique

Un des défis inhérents au déploiement de la pathologie numérique est celui lié aux artefacts des images microscopiques. En effet, ces dernières sont acquises après une série d'étapes successives : la fixation, les prélèvements, l'inclusion en paraffine, la coupe, la coloration (36). À chaque étape, des erreurs peuvent survenir et entraîner des artefacts sur la lame, ayant des répercussions lors de la numérisation de la lame (37). Même si notre étude était monocentrique et que les protocoles de coloration

sont entièrement standardisés, un contrôle précis de la couleur et de l'intensité de la coloration n'est généralement pas faisable. De plus, le recrutement des lames a été réalisé sur trois années consécutives. Les colorants ont tendance à s'estomper avec le temps et les couleurs peuvent légèrement différer (38). Si, pour le pathologiste la différence d'intensité des différents colorants n'est que rarement un problème pour l'analyse finale, qu'en est-il pour nos algorithmes ? Ces variations de couleurs ont pu avoir un impact sur l'apprentissage de nos RNAC. En effet, ce facteur a pu les induire en erreur. Dans la littérature, plusieurs équipes se sont penchées sur ce problème (39-41). Elles ont développé différentes techniques de normalisation des couleurs en prouvant que cette étape de prétraitement pouvait être utile pour améliorer les performances des algorithmes de DL. En réduisant les incohérences de couleur dans les images, cela pourrait améliorer l'efficacité et la précision des RNAC. Néanmoins, une équipe travaillant sur des images histo-pathologiques de cancer du sein a fait l'expérience avec deux méthodes de normalisation des couleurs (Macenko Stain Normalization et Reinhard Stain Normalisation) versus les images non normalisées. D'après leurs résultats, aucune normalisation des couleurs ne s'est avérée supérieure à l'autre et pour certains groupes, les performances étaient meilleures sur les images non normalisées (42). L'état actuel de la recherche ne permettant pas de trancher cette question, il aurait été intéressant dans notre étude de pouvoir comparer les performances des RNAC avec et sans normalisation des couleurs pour déterminer si ce facteur impacte les performances de classification des RNAC.

Notre jeu de données était issu d'un seul centre (CHU de Lille) : il n'est sûrement pas représentatif de la diversité des scénarii histologiques auxquels les pathologistes sont confrontés dans leur pratique courante. Pour améliorer la représentativité de notre base de données, il aurait fallu collecter des lames issues de différents centres. De plus, même si l'ADCP est le type histologique le plus fréquent, il aurait également été intéressant de pouvoir inclure d'autres tumeurs pancréatiques dans notre étude comme les tumeurs neuroendocrines du pancréas ou encore les kystes pancréatiques.

2. La procédure de labellisation

En avril 2021, Yoshiki Naito et collaborateurs avaient une base de données nettement plus importante que la nôtre : 594 lames numérisées issues de cytoponctions pancréatiques réalisées grâce à un seul type d'aiguille (type FNB) entre 2010 et 2020 à Kurume (Japon). Dix huit pathologistes spécialisés dans la pathologie pancréatique ont participé à l'étape de labellisation. Ils devaient circonscrire les foyers d'adénocarcinome sur la lame entière. Pour chaque lame annotée, une vérification par un second pathologiste a été mise en place (43). Ils ont entraîné et testé un seul type de RNAC : EfficientNetB127 avec les stratégies de transfert d'apprentissage (ImageNet) et d'augmentation artificielle des données (variation de la luminosité, de la saturation et rotation) (44). Les résultats de la phase de test sur 120 lames étaient de 94,2 % pour le taux de bonne prédiction, 98,8 % pour la précision et de 93 % pour le rappel.

Notre procédure de labellisation était différente de celle utilisée par l'équipe de Yoshiki Naito puisque nous l'avons faite à l'échelle des tuiles et non à l'échelle des lames comme l'équipe japonaise. Plusieurs éléments nous ont incités à labelliser à l'échelle des tuiles.

Une des principales raisons était la quantité limitée de lame disponible pour notre étude. En optant pour des tuiles de 96 x 96 pixels, nous augmentions considérablement le jeu de données tout en réduisant le temps de calcul des RNAC. De plus, cela nous paraissait plus pertinent et plus précis en raison d'une des spécificités de l'ADCP : l'abondance du stroma. Le stroma est un tissu conjonctif vascularisé au contact des cellules tumorales, sans être lui même tumoral. Dans l'ADCP, le stroma peut représenter jusqu'à 90% du volume tumoral (45). Pour rappel, le stroma sans cellule tumorale a été labellisé « non tumoral ». Cela explique la différence de proportion entre les catégories « adénocarcinome » et « non tumoral ». Ainsi, l'algorithme se trouve mis en difficulté immédiatement par le stroma. En effet, le tissu tumoral présente systématiquement et en grande quantité du stroma mais la présence de stroma n'implique pas nécessairement le caractère tumoral du tissu. Notre algorithme aura alors tendance à produire un nombre important de faux positifs ce qui est le cas (30%). Il se pose alors la question suivante : aurions-nous eu intérêt à créer une classe à part entière pour le stroma ? C'est, sans aucun doute, une piste intéressante à suivre car il représente une part non négligeable des tuiles « non tumoral ». Pour rappel, lors de la phase d'entraînement nous avons dû sélectionner aléatoirement 10 826 tuiles parmi les 190 664 tuiles labellisées « non tumoral ». Compte tenu de la diversité et de la complexité des images histo-pathologiques, notre jeu de données d'entraînement n'était probablement pas exhaustif et représentatif des

différents aspects du contingent « non tumoral » et particulièrement du stroma. En outre, ce même stroma était également représenté au sein des tuiles labellisées « adénocarcinome ». Ces constatations nous permettent de mieux comprendre la proportion de faux positifs représentés par des tuiles comprenant du stroma et donc le manque de précision (valeur prédictive positive) de notre algorithme.

Labelliser à l'échelle des tuiles permettait de véritablement cibler les cellules tumorales ce qui était notre prérequis pour développer notre logiciel d'aide à l'interprétation du pathologiste. L'intérêt premier de ce logiciel était le gain de temps lors de la lecture de la lame pour réduire le délai diagnostique. La réalisation de notre objectif secondaire n'a pas été difficile sur le plan informatique et la mise en valeur des zones d'intérêts était opérationnelle. Cependant, l'algorithme n'ayant pas une capacité prédictive suffisante, notre logiciel n'est à ce jour pas utilisable.

Les faux négatifs, bien que peu nombreux, étaient représentés par des tuiles contenant deux ou trois cellules tumorales sans stroma associé. Ce type de tuile n'était pas représenté lors la phase d'entraînement des RNAC : elles représentent les tuiles classées « probable » et non intégrées. Tous confondus, les faux positifs et faux négatifs représentent les tuiles sur lesquelles notre modèle s'est très peu ou pas entraîné.

En parcourant la littérature, une autre procédure de labellisation est envisageable : « noisy label annotation » expérimentée par Han Le et collaborateurs pour détecter l'ADCP. A partir de lames entières numérisées issues de pièces opératoires pancréatiques, les pathologistes devaient délimiter la zone tumorale au

faible grossissement sur la lame entière. Si les avantages de cette méthode étaient de réduire le temps d'annotation et d'augmenter considérablement le jeu de données, elle a introduit la notion de « bruit » : la présence de secteurs non tumoraux au sein de la zone tumorale (faux positifs). Néanmoins, ils ont réussi à montrer qu'un RNAC entraîné avec ce plus grand ensemble de données même « bruitées » surpassait un RNAC entraîné avec moins de données non « bruitées » (46).

Notre Gold Standard était l'annotation faite par le pathologiste junior. Il aurait été souhaitable d'avoir au moins deux, voire plusieurs pathologistes, pour mettre en place une étape de vérification des labels afin de limiter le risque d'erreur et la part de subjectivité. Le taux de concordance entre les pathologistes aurait pu être calculé pour évaluer la reproductibilité inter-observateurs.

La labellisation des données est une étape fondamentale pour l'entraînement des RNAC. Elle requiert des moyens humains conséquents, un temps certain mais également une véritable réflexion en amont en prenant en compte les spécificités du type de prélèvement, du type de lésion à détecter, de l'environnement lésionnel, etc. Si les données d'entraînement sont peu nombreuses et/ou non représentatives de la pathologie à détecter comme du contingent « normal », les RNAC ne pourront jamais être performants. Meilleure est la qualité du jeu de données d'entraînement, meilleures seront les performances du modèle.

3. Les biais d'apprentissage

Les tuiles ont été réparties aléatoirement pour les jeux de données de l'étape d'apprentissage et de validation sans prendre en compte l'origine de la lame. Pour mieux comprendre : les tuiles de la lame 1 ont pu être réparties à la fois dans le jeu d'apprentissage et dans le jeu de validation. Stricto sensu, ces deux jeux de données n'étaient donc pas indépendants. Ils partageaient des tuiles issues des mêmes lames. Cela a certainement contribué aux très bons taux de bonne prédiction calculés à l'issue de l'étape de validation.

Pour 2 des 3 RNAC entraînés (VGG-19 et ResNet-50), la stratégie d'apprentissage 4 (avec transfert d'apprentissage et augmentation artificielle du jeu de données), a donné le meilleur taux de bonne prédiction par rapport aux 3 autres stratégies d'apprentissage. De manière inexplicable, pour le RNAC DenseNet-201, cette stratégie d'apprentissage a donné le plus mauvais taux de bonne prédiction tous RNAC confondus et stratégies d'apprentissage confondues. En revanche, la stratégie d'apprentissage 3 (augmentation artificielle du jeu de données, sans transfert d'apprentissage) a donné le meilleur taux de bonne prédiction tous RNAC confondus et stratégies d'apprentissage confondues. Sans pouvoir l'affirmer avec certitude, le fait que la stratégie d'apprentissage 4 puisse produire un taux de prédiction moins bon que la stratégie d'apprentissage 3 est sûrement dû au phénomène de sur-apprentissage abordé précédemment. En effet, à force d'ajouter artificiellement des données pour nourrir toujours plus l'algorithme, ce dernier devient de plus en plus spécifique aux tuiles d'apprentissage. Ainsi, devant un jeu de donnée complètement nouveau, son taux de prédiction chute. Quoiqu'il en soit, cela illustre parfaitement l'effet « boîte noire » du DL vu en introduction. Bien que l'interrogation des RNAC soit

possible grâce à un raisonnement mathématique, elle ne reflète pas la prise de décision clinique. Or, comprendre la manière dont les RNAC analysent les informations est primordial pour améliorer leurs performances et corriger les erreurs (47).

4. Perspectives

A ce stade de développement, notre algorithme et notre logiciel ne pourront pas être utilisés dans la pratique quotidienne. Un des points forts des méthodes de DL et plus largement d'IA, est la possibilité de ré-entraîner les RNAC autant de fois que nous le souhaitons. Avec ce premier travail, nous avons pu identifier différents axes d'amélioration : augmenter notre base de données, si possible provenant de différents centres avec différents types de tumeur, vérifier la labellisation, utiliser des outils de normalisation des couleurs, créer une nouvelle catégorie avec le stroma...

Pour agrandir plus rapidement et facilement notre base de données de tuiles « adénocarcinome », nous pourrions exploiter des lames entières issues de pièces opératoires pancréatiques et en extraire les tuiles « adénocarcinome ». Cela permettrait d'augmenter les exemples et de créer deux jeux de données strictement indépendants (jeux d'apprentissage et jeux de validation) pour la phase d'entraînement. De nombreux laboratoires d'anatomopathologie en France ont déjà adopté ou sont en transition vers la pathologie numérique dans le diagnostic de routine. Les bases de données de lames numériques vont alors se multiplier et le partage de données d'un laboratoire à l'autre sera facilité, tout en respectant le droit des patients. Ainsi, de nouveaux et nombreux projets basés sur l'IA et le DL pourront être développés.

Ce travail a permis de constituer une base de données rare sur l'adénocarcinome du pancréas dans le cadre du diagnostic sur cytoponctions. En ayant fait le choix de labelliser « non tumoral » le stroma sans cellule tumorale, qui est une spécificité de l'adénocarcinome pancréatique, elle pourra être réemployée pour d'autres travaux de classification d'images histologiques bénin/malin sur des prélèvements biopsiques.

Ces nouveaux outils d'aide au diagnostic basé sur le DL ont un grand avantage par rapport aux différentes techniques qui sont utilisées en routine (l'immunohistochimie ou l'hybridation in situ en fluorescence) : ils ne consomment pas le matériel biopsique. Cela permet de réserver le matériel biopsique pour d'éventuelles techniques de biologie moléculaire. D'autant plus dans ce type de prélèvement qui est souvent de petite taille.

Pour ce premier travail, nous avons fait le choix de ne pas inclure de données cliniques, radiologiques ou biologiques. Nous imaginons que dans le futur, l'enjeu sera de réaliser une intégration multiparamétrique pour exploiter les données variées des patients. Très récemment, lors du congrès de l'American Society of Clinical Oncology 2021, l'Assistance Publique des Hôpitaux de Paris (AP-HP) et OWKIN ont annoncé les résultats de leur collaboration : « Identification des sous-types moléculaires d'adénocarcinome pancréatique sur les lames d'histologie à l'aide de modèles d'apprentissage profonds ». Il s'agit du premier outil basé sur l'IA pour prédire les sous-types moléculaires de l'ADCP à partir de lames d'histologie. Ils ont connecté les informations aux niveaux génomique, cellulaire et tissulaire. Ainsi, pouvoir déterminer le sous-type de signature transcriptomique tumorale permet de

prédire la réponse à la chimiothérapie. Cet outil a été développé en utilisant les ressources histologiques et moléculaires uniques de quatre hôpitaux du réseau AP-HP (Amboise Paré, Beaujon, Pitié Salpêtrière, Saint Antoine) à travers une collaboration unique entre pathologistes de l'AP-HP, bioinformaticiens du groupe Carte d'Identité des Tumeurs de la Ligue Contre le Cancer et des data scientists d'OWKIN. Il peut sous-typer la tumeur à distance en quelques minutes ouvrant la voie à de nombreuses applications issues de la science fondamentale (étude de l'hétérogénéité intratumorale) à la pratique clinique (sous-typage tumoral dans les essais cliniques) (48).

Conclusion

Ce travail est le fruit d'une collaboration pluridisciplinaire, avec l'équipe de l'ISEN et l'équipe du plateau de microscopie photonique BICeL du Campus Hospitalo-Universitaire de Lille.

Il nous a permis de développer un prototype d'algorithme et de logiciel d'aide au diagnostic pour l'ADCP sur lames de cytoponction. À la fin de la phase d'entraînement, le taux de bonne prédiction était de 96,5 %. Lors de la phase de test, le taux de bonne prédiction était de 69,3 % avec une précision (valeur prédictive positive) de 13,7 % et un rappel (sensibilité) de 84,6 %. À ce stade, nos deux outils ne peuvent pas encore être utilisés dans la pratique quotidienne. Néanmoins, de très nombreuses pistes d'amélioration sont envisageables et réalisables. La labellisation spécifique du stroma est l'une de nos pistes privilégiées pour essayer de réduire le nombre de faux positifs et ainsi augmenter notre précision (valeur prédictive positive).

Plus largement, ce projet nous a donné l'opportunité de découvrir l'IA de manière concrète dans le diagnostic anatomopathologique. Elle a toute sa place dans le diagnostic d'ADCP et cela ouvre des perspectives pour l'ensemble des petits prélèvements de cytoponction d'organe profond.

Annexe 1 : note d'information pour les patients

NOTE D'INFORMATION D'UNE RECHERCHE N'IMPLIQUANT PAS LA PERSONNE HUMAINE

« Développement d'un nouvel outil informatique pour l'aide au diagnostic anatomopathologique d'une maladie du pancréas »

Nom du patient : XXXX
Prénom du patient : XXXX
Date de naissance : XXXX

Cette note d'information a pour objectif de vous expliquer le but de cette étude afin que vous puissiez décider d'y participer ou non. Votre participation à cette étude doit être entièrement volontaire. Prenez le temps de lire cette note d'information et n'hésitez pas à poser des questions à votre médecin pour avoir des renseignements complémentaires.

I. Le but de la recherche

Cette recherche porte sur le diagnostic anatomopathologique de votre maladie du pancréas. Elle entre dans le cadre d'un travail de thèse de médecine, qui a pour but d'améliorer les performances diagnostiques des médecins anatomopathologistes en développant un nouvel outil informatique, qui les aidera dans la prise en charge de chaque prélèvement comme le vôtre.

II. Quelles données sont recueillies, pourquoi et comment ?

Vous êtes sollicité car vous avez eu une écho-endoscopie pancréatique dans le but d'identifier votre maladie du pancréas. Au total, 70 patients comme vous sont concernés. Si vous l'acceptez, nous allons relire et numériser votre biopsie pancréatique qui participera à la mise au point de ce nouvel outil informatique. Lequel sera une aide dans la prise de décision diagnostique pour le médecin anatomopathologiste.

Il n'y aura aucun impact sur votre prélèvement. Cette étude n'aura aucune influence sur votre prise en charge et ne comportera aucun risque pour vous.

Améliorer nos performances diagnostiques au quotidien permettra d'améliorer la prise en charge de chaque futur patient.

Le recueil de données se fera à partir des informations présentes dans votre dossier médical informatique ou papier de l'hôpital de Lille.

Les données nécessaires pour la conduite de la recherche comprennent notamment : nom, prénom adresse postale, âge, sexe, compte-rendu anatomopathologique, compte-rendu écho-endoscopique.

III. Confidentialité des données

Les données médicales recueillies dans le cadre de cette étude seront réunies sur un fichier informatique permettant leur traitement par le centre hospitalier sur le fondement de l'intérêt public. Les données vous concernant seront rendues anonymes, c'est-à-dire identifiées par un numéro de code. Les données pourront également, dans des conditions assurant la confidentialité, *être transmises à l'Organisme Responsable de la Recherche ou aux personnes ou société agissant pour son compte, ou menant des projets de recherche conjoints, en France ou à l'étranger, y compris en dehors de l'Union Européenne à condition que le pays de destination soit reconnu par les autorités françaises comme assurant un niveau de protection des données suffisant et approprié, ou à condition que l'organisme destinataire des données ait signé un contrat comportant les clauses contractuelles types, sur la base des modèles de la Commission Européenne.*

La base de données ainsi créée sera conservée pendant la durée de réalisation de l'étude et jusqu'à deux ans après la dernière publication des résultats de l'étude. Puis, la base de données sera archivée pendant la durée légale autorisée. Passé ce délai, la base de données ainsi que tous les documents relatifs à la recherche seront définitivement détruits.

Conformément à la loi « informatique et libertés » du 6 janvier 1978 et au règlement (UE) 2016/679 du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, vous disposez à leur égard d'un droit d'accès, de rectification, d'effacement, de limitation du traitement et du droit de vous opposer au traitement.

Si vous souhaitez exercer vos droits et obtenir communication des informations vous concernant, veuillez-vous adresser au délégué à la protection des données du Groupement Hospitalier Territorial Lille Métropole Flandres Intérieur à l'adresse suivante : dpo@chru-lille.fr.

Si vous considérez que vos droits n'ont pas été respectés, vous avez également la possibilité de saisir la Commission Nationale de l'Informatique des Libertés (CNIL) directement via son site internet : www.cnil.fr.

Vous pouvez accéder directement ou par l'intermédiaire d'un médecin de votre choix à l'ensemble de vos données médicales en application des dispositions de l'article L1111-7 du Code de la Santé Publique. Ces droits s'exercent auprès du médecin qui vous suit dans le cadre de la recherche et qui connaît votre identité.

IV. Acceptation et interruption de votre participation

Vous êtes libre d'accepter ou de refuser de participer à cette recherche. Dans le cas où vous répondriez par la positive, vous disposez de la possibilité d'interrompre votre participation à tout moment sans aucun préjudice et sans engager votre responsabilité. Cela n'affectera évidemment pas votre prise en charge.

En l'absence d'opposition / refus de votre part sous un délai de 3 semaines après réception de la présente note, nous considérerons que vous acceptez de participer à cette recherche.

V. Comment cette recherche est-elle encadrée ?

Le responsable de la recherche, a pris toutes les mesures nécessaires pour mener cette recherche conformément à la loi française aux dispositions de la Loi Informatique et Liberté applicables aux recherches n'impliquant pas la personne humaine (CNIL loi 78-17 du 6 janvier 1978 modifiée) et européenne (Règlement européen ° 2016/679 relatif à la protection des données - RGPD).

Fait à : Lille

Le : 09 Novembre 2020

Signature du responsable de la recherche:

CADRE RESERVE AU RECUEIL DE L'OPPOSITION		
NOM/Prénom	du	patient :
.....		
Je m'oppose à l'utilisation de mes données dans le cadre de cette recherche « Développement d'un nouvel outil informatique pour l'aide au diagnostic anatomopathologique d'une maladie du pancréas ».		
<input type="checkbox"/> Oui		
<input type="checkbox"/> Non		
<u>Signature</u> (Précédée de la mention : Lu, compris et approuvé) :		
<u>Date</u> :		

Si vous souhaitez vous opposer à la recherche, merci de renvoyer le cadre ci-dessus rempli à l'adresse suivante :
Pôle de Biologie Pathologie Génétique
A l'attention du Pr Leteurtre
Service anatomie cytologie pathologique
Rue du Pr Jules LECLERCQ
59 037 LILLE CEDEX

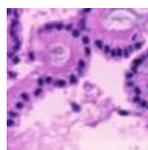
Annexe 2 : exemples de tuiles labellisées



LAME 4
crop_11_19.NON
TUMORAL.png



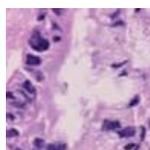
LAME 4
crop_13_15.NON
TUMORAL.png



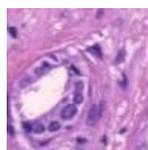
LAME
40_crop_67_65.NON
TUMORAL.png



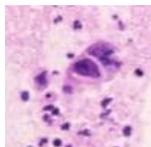
LAME
43_crop_33_18.ADK.png



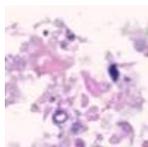
LAME
60_crop_29_53.
ADK.png



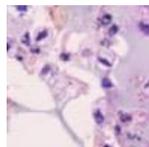
LAME
43_crop_24_107.
ADK.png



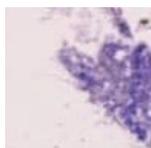
LAME
61_crop_38_15.Probable.png



LAME
64_crop_71_65.Probable.png



LAME
9_crop_40_46.
Probable.png



LAME
20_crop_58_93.
necrose.png



LAME
34_crop_53_28.
necrose.png



LAME
27_crop_45_10.
necrose.png

Bibliographie

1. Facteurs de risque de l'adénocarcinome du pancréas : prévention et dépistage. Disponible sur : <https://www.fmcgastro.org/texte-postu/postu-2021-paris/facteurs-de-risque-de-ladenocarcinome-du-pancreas-prevention-et-dépistage>.
2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting Cancer Incidence and Deaths to 2030 : The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Res.* 1 juin 2014;74(11):2913-21.
3. Christenson ES, Jaffee E, Azad NS. Current and emerging therapies for patients with advanced pancreatic ductal adenocarcinoma: a bright future. *The Lancet Oncology.* 1 mars 2020;21(3):e135-45.
4. Traitement du cancer du pancréas. Disponible sur : <https://www.fmcgastro.org/texte-postu/postu-2019-paris/traitement-du-cancer-du-pancreas-recommandations-en-cours-de-labellisation-inca-2019>.
5. Echoendoscopie interventionnelle : une technique en pleine évolution. *Revue Médicale Suisse.* Disponible sur : <https://www.revmed.ch/revue-medicale-suisse/2009/revue-medicale-suisse-215/echoendoscopie-interventionnelle-une-technique-en-pleine-evolution>.
6. Ponction sous échoendoscopie. Disponible sur : <https://www.fmcgastro.org/postu-main/archives/postu-2004-paris/ponction-sous-echoendoscopie>
7. Deep learning : le cerveau humain, un modèle à imiter. Disponible sur : <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning>.
8. Deep learning pour et par les nuls.
9. L'IA, c'est quoi? Disponible sur : <https://www.coe.int/fr/web/artificial-intelligence/what-is-ai>.
10. Machine learning, deep learning : quelles différences? Disponible sur : <https://siecledigital.fr/2019/01/30/differences-intelligence-artificielle-machine-learning-deep-learning>.
11. Apprentissage automatique. Wikipédia. Disponible sur : https://fr.wikipedia.org/w/index.php?title=Apprentissage_automatique&oldid
12. Intelligence Artificielle Yann LeCun.pdf Disponible sur : https://www.college-defrance.fr/media/yannlecun/UPL4485925235409209505_Intelligence_Artificielle____Y._LeCun.pdf.
13. Fernandez-Maloigne C, Guillevin R. L'intelligence artificielle au service de l'imagerie et de la santé des femmes. *Imagerie de la Femme.* déc 2019;29(4):179-86.
14. Goyal H, Mann R, Gandhi Z, Perisetti A, Zhang Z, Sharma N, et al. Application of artificial intelligence in pancreaticobiliary diseases. *Clin Med Insights Gastroenterol.* 1 janv 2021;14:2631774521993059.
15. Săftoiu A, Vilman P, Gorunescu F, Gheonea DI, Gorunescu M, Ciurea T, et al. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. *Gastrointestinal Endoscopy.* déc 2008;68(6):1086-94.

16. Săftoiu A, Vilmann P, Gorunescu F, Janssen J, Hocke M, Larsen M, et al. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol.* janv 2012;10(1):84-90.e1.
17. Zhu M, Xu C, Yu J, Wu Y, Li C, Zhang M, et al. Differentiation of Pancreatic Cancer and Chronic Pancreatitis Using Computer-Aided Diagnosis of Endoscopic Ultrasound (EUS) Images: A Diagnostic Test. *PLoS One.* 21 mai 2013;8(5):e63820.
18. Săftoiu A, Vilmann P, Dietrich CF, Iglesias-Garcia J, Hocke M, Seicean A, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). *Gastrointestinal Endoscopy.* 1 juill 2015;82(1):59-69.
19. Ozkan M, Cakiroglu M, Kocaman O, Kurt M, Yilmaz B, Can G, et al. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. *Endosc Ultrasound.* 2016;5(2):101-7.
20. Kuwahara T, Hara K, Mizuno N, Okuno N, Matsumoto S, Obata M, et al. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Clin Transl Gastroenterol.* 7 mai 2019;10(5):e00045.
21. Liu S-L, Li S, Guo Y-T, Zhou Y-P, Zhang Z-D, Li S, et al. Establishment and application of an artificial intelligence diagnosis system for pancreatic cancer with a faster region-based convolutional neural network. *Chin Med J (Engl).* 5 déc 2019;132(23):2795-803.
22. Gaubert A, Long-Mira É, Lassalle S, Hofman V, Hofman P, Ilié M. Intelligence artificielle et pathologistes, est-ce bien raisonnable ? *Revue Francophone des Laboratoires.* 1 janv 2020;2020(518):34-9.
23. Shallu, Mehra R. Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express.* 1 déc 2018;4(4):247-54.
24. Hameed Z, Zahia S, Garcia-Zapirain B, Javier Aguirre J, María Vanegas A. Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models. *Sensors (Basel).* 5 août 2020;20(16):4373.
25. Salle FG, Stang NL, Tirode F, Courtiol P, Nicholson AG, Tsao M-S, et al. Comprehensive Molecular and Pathologic Evaluation of Transitional Mesothelioma Assisted by Deep Learning Approach: A Multi-Institutional Study of the International Mesothelioma Panel from the MESOPATH Reference Center. *Journal of Thoracic Oncology.* 1 juin 2020;15(6):1037-53.
26. Pathology Image Analysis Using Segmentation Deep Learning Algorithms | Elsevier Enhanced Reader. Disponible sur : <https://reader.elsevier.com/reader>.
27. 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.
28. Zemouri R, Devalland C, Valmary-Degano S, Zerhouni N. Intelligence artificielle : quel avenir en anatomie pathologique ? *Annales de Pathologie.* 1 avr 2019;39(2):119-29.

29. Primaas prépare la commercialisation de son outil d'aide au diagnostic histologique. Disponible sur : <https://www.ticpharma.com/story/1593/primaas-prepare-la-commercialisation-de-son-outil-d-aide-au-diagnostic-histologique>.
30. RGPD : par où commencer | CNIL. Disponible sur: <https://www.cnil.fr/fr/rgpd-par-ou-commencer>.
31. Recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé Méthodologie de référence MR-004 | CNIL. Disponible sur : <https://www.cnil.fr/fr/declaration/mr-004-recherches-n-impliquant-pas-la-personne-humaine>.
32. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE; 2016. p. 770-8. Disponible sur : <http://ieeexplore.ieee.org/document/7780459>.
33. Huang G, Liu Z, Pleiss G, van der Maaten L, Weinberger KQ. Convolutional Networks with Dense Connectivity. Disponible sur : <http://arxiv.org/abs/2001.02394>.
34. Le flou gaussien dans le traitement d'image. Disponible sur : <https://www.adobe.com/fr/creativecloud/photography/discover/gaussian-blur>.
35. TD6 : Image Classification Architectures. Disponible sur : <https://www.enseignement.polytechnique.fr/informatique>.
36. Farris AB, Cohen C, Rogers TE, Smith GH. Whole Slide Imaging for Analytical Anatomic Pathology and Telepathology: Practical Applications Today, Promises, and Perils. Archives of Pathology & Laboratory Medicine. 1 avr 2017;141(4):542-50.
37. Heeke S, Delingette H, Fanjat Y, Long-Mira E, Lassalle S, Hofman V, et al. La pathologie cancéreuse pulmonaire à l'heure de l'intelligence artificielle : entre espoir, désespoir et perspectives. Annales de Pathologie. avr 2019;39(2):130-6.
38. Niethammer M, Borland D, Marron JS, Woosley J, Thomas NE. Appearance Normalization of Histology Slides. Mach Learn Med Imaging. 2010;6357:58-66.
39. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Xiaojun Guan, et al. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro [Internet]. Boston, MA, USA: IEEE; 2009 p. 1107-10. Disponible sur : <http://ieeexplore.ieee.org/document/5193250>.
40. Khan AM, Rajpoot N, Treanor D, Magee D. A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. IEEE Transactions on Biomedical Engineering. juin 2014;61(6):1729-38.
41. Salvi M, Michielli N, Molinari F. Stain Color Adaptive Normalization (SCAN) algorithm: Separation and standardization of histological stains in digital pathology. Computer Methods and Programs in Biomedicine. sept 2020;193:105506.
42. Munien C, Viriri S. Classification of Hematoxylin and Eosin-Stained Breast Cancer Histology Microscopy Images Using Transfer Learning with EfficientNets. Comput Intell Neurosci. 9 avr 2021;2021:5580914.

43. Naito Y, Tsuneki M, Fukushima N, Koga Y, Higashi M, Notohara K, et al. A deep learning model to detect pancreatic ductal adenocarcinoma on endoscopic ultrasound-guided fine-needle biopsy. *Sci Rep.* 19 avr 2021;11(1):8454.
44. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Disponible sur : <http://arxiv.org/abs/1905.11946>
45. Felix K, Gaida MM. Neutrophil-Derived Proteases in the Microenvironment of Pancreatic Cancer -Active Players in Tumor Progression. *Int J Biol Sci.* 28 janv 2016;12(3):302-13.
46. Le H, Samaras D, Kurc T, Gupta R, Shroyer K, Saltz J. Pancreatic Cancer Detection in Whole Slide Images Using Noisy Label Annotations. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., éditeurs. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* p. 541-9. (Lecture Notes in Computer Science; vol. 11764). Disponible sur : http://link.springer.com/10.1007/978-3-030-32239-7_60.
47. Mendoza Ladd A, Diehl DL. Artificial intelligence for early detection of pancreatic adenocarcinoma: The future is promising. *World J Gastroenterol.* 7 avr 2021;27(13):1283-95.
48. Une collaboration Carnot AP-HP - Owkin pour lutter contre cancer du pancréas grâce à l'IA | Le réseau des Carnot. Disponible sur : <https://www.instituts-carnot.eu/fr/actualite/une-collaboration-carnot-ap-hp-owkin-pour-lutter-contre-cancer-du-pancreas>.

AUTEURE : Nom : KARLESKIND

Prénom : Oriane

Date de soutenance : mercredi 10 novembre 2021

Titre de la thèse : L'Intelligence Artificielle au service du diagnostic histologique d'adénocarcinome sur matériel de cytoponction pancréatique

Thèse - Médecine - Lille 2021

Cadre de classement : *Anatomie cytologie pathologiques*

DES + spécialité : *Anatomie cytologie pathologiques*

Mots-clés : Intelligence Artificielle ; Deep Learning ; pathologie numérique ; adénocarcinome ; pancréas ; cytoponction pancréatique.

Résumé :

Contexte : L'adénocarcinome représente 90 % des tumeurs du pancréas et devrait être la deuxième cause de mortalité par cancer aux Etats Unis en 2030. La cytoponction par voie écho-endoscopique est la technique non chirurgicale de référence pour le diagnostic des tumeurs pancréatiques. Le déploiement de la pathologie numérique offre de nouvelles perspectives innovantes pour le diagnostic, notamment grâce au Deep Learning. Cette méthode d'Intelligence Artificielle repose sur des algorithmes capables de mimer le cerveau humain grâce à des réseaux de neurones artificiels.

Méthode : Trois réseaux de neurones artificiels (VGG-19, ResNet-50 et DenseNet-201) ont été entraînés à partir de 68 lames du CHU de Lille, après information des patients. Des techniques de transfert d'apprentissage et/ou d'augmentation artificielle des données ont été appliquées.

À l'issue de la phase d'entraînement, nous avons calculé le taux de bonne prédiction pour chaque réseau de neurones en fonction de chaque stratégie d'apprentissage. Pour la phase de test, un jeu supplémentaire de 4 lames d'adénocarcinome a été prédit par le meilleur algorithme. Le taux de bonne prédiction a été calculé à l'échelle de lame entière puis à l'échelle des tuiles. Pour évaluer l'applicabilité en routine de notre algorithme, nous avons également calculé la précision (valeur prédictive positive) et le rappel (sensibilité).

Un logiciel d'aide à l'interprétation du pathologiste a été conçu pour faciliter la lecture du pathologiste.

Résultats : Lors de la phase d'entraînement, le taux de bonne prédiction de l'algorithme DenseNet-201 avec augmentation artificielle des données sans transfert d'apprentissage était de 96,5 %.

À l'issue de la phase de test, DenseNet-201 a obtenu un taux de bonne prédiction de 100 % à l'échelle des lames entières. À l'échelle des tuiles, le taux de bonne prédiction était de 69,3 % avec une précision de 13,7 % et un rappel de 84,6 %.

Conclusion : Les performances de notre algorithme ne permettent pas encore l'application en routine clinique, néanmoins de nombreuses pistes d'amélioration sont envisageables et réalisables.

L'Intelligence Artificielle a toute sa place dans le diagnostic d'adénocarcinome du pancréas. Cela ouvre des perspectives pour l'ensemble des petits prélèvements de cytoponction d'organe profond.

Composition du Jury :

Président : Monsieur le Professeur Claude Alain MAURAGE

Assesseurs : Madame le Docteur Florence RENAUD, Monsieur le Docteur Romain GERARD

Directrice de thèse : Madame le Professeur Emmanuelle LETEURTRE