



**HAL**  
open science

# **Evaluation du score mitotique des carcinomes mammaires infiltrants : développement et apport d'un algorithme de détection de mitoses**

Loris Guichard

## ► To cite this version:

Loris Guichard. Evaluation du score mitotique des carcinomes mammaires infiltrants : développement et apport d'un algorithme de détection de mitoses. Anatomie, Histologie, Anatomopathologie [q-bio.TO]. Faculté de médecine Paris-Saclay, 2022. Français. ⟨NNT : ⟩. ⟨tel-03984138⟩

**HAL Id: tel-03984138**

**<https://hal.science/tel-03984138v1>**

Submitted on 12 Feb 2023

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

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



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

UNIVERSITE PARIS-SACLAY  
FACULTE DE MEDECINE PARIS-SACLAY

Année : 2022

N°.....

(Enregistrement bibliothèque)

**Thèse pour le Doctorat en Médecine**

Présentée et soutenue publiquement le : 12 Octobre 2022

Par : Loris GUICHARD

Né le : 19 Août 1992

À Echirolles (38)

**Evaluation du score mitotique des carcinomes mammaires infiltrants :  
développement et apport d'un algorithme de détection de mitoses**

**Membres du Jury:**

**Président du Jury : Madame le Professeur Catherine GUETTIER, PU-PH**

**Directrice de thèse : Madame le Professeur Sophie PREVOT, PU-PH**

**Madame le Professeur Isabelle BROCHERIOU, PU-PH**

**Madame le Docteur Magali LACROIX-TRIKI, Praticien Spécialiste**

**Monsieur le Docteur Nicolas LOMENIE, Professeur Associé**

# Remerciements

Aux membres du Jury,

A Madame le Professeur Catherine GUETTIER, qui nous fait l'honneur de présider ce jury de thèse. Merci de m'avoir confié ce sujet. Votre vision de l'anatomopathologie tournée vers l'avenir est une vraie source d'inspiration.

A Madame le Professeur Sophie PREVOT, pour vos enseignements et votre encadrement durant toute la durée de ce travail. Votre présence, votre bienveillance et votre vision toujours positive de la situation ont été des éléments indispensables et déterminants à la réussite de ce travail.

A Madame le Professeur Isabelle BROCHERIOU, l'année que j'ai passé au sein de votre service a été riche d'enseignements, et c'est un honneur pour moi que vous ayez accepté de faire partie de ce jury.

A Madame le Docteur Magali LACROIX-TRIKI, en tant qu'experte reconnue de la pathologie mammaire et compte tenu de l'intérêt que vous portez à ce projet, c'est pour moi un honneur de vous avoir dans ce jury de thèse.

A Monsieur le Docteur Nicolas LOMENIE, en tant qu'expert reconnu dans le domaine de l'intelligence artificielle c'est pour moi un honneur que vous ayez accepté de faire partie de ce jury.

A tous les développeurs et personnes travaillant chez Prima qui ont rendu ce travail possible, Marie, Nicolas, Eloi que je remercie pour leur implication dans ce projet, ainsi qu'une mention toute particulière pour Clara qui m'a énormément aidé pour la rédaction et la relecture de ce manuscrit notamment pour toutes les parties relatives au fonctionnement de l'algorithme. Ce travail de thèse n'aurait pas été possible sans eux.

A Margot et Suzanne pour leur participation en tant que pathologistes lecteurs au cours de ce travail.

A toutes les personnes qui m'ont formé et accompagné durant ces années d'études,

A l'équipe du service de radiothérapie du Professeur HENNEQUIN de l'hôpital Saint-Louis, pour leur bienveillance et leur accompagnement lors de mon changement de spécialité en premier semestre d'internat.

Au Professeur Antoine MARTIN de l'hôpital Avicenne, pour sa gentillesse et sa bienveillance lors de mes premiers pas en anatomopathologie, pour m'avoir permis de confirmer mon choix d'orientation lors de mon second semestre d'internat.

Aux médecins, chefs et assistants qui ont donné de leur temps pour me permettre d'apprendre l'anatomopathologie: Pr Sophie PREVOT, Pr Sophie FERLICOT, Dr Anne VINCENT-SALOMON, Dr Paul FRENEAUX, Dr Armelle BARDIER-DUPAS, Dr Frédéric CHARLOTTE, Dr Justine GIRAULT, Dr Laura LARNAUDIE, Dr Pauline QUILHOT, Dr Gabrielle DENIZIAUT, Dr Pierre SOHIER, Pr Diane DAMOTTE, Dr Barbara BURRONI.

Aux techniciens des services d'anatomopathologie des hôpitaux Avicenne, Bicêtre, l'Institut Curie, la Pitié-Salpêtrière, Cochin.

Aux secrétaires du service d'anatomopathologie de l'hôpital Bicêtre, Nicole, Sophie, Catherine, Mameess, pour leur gentillesse durant mon stage et l'aide qu'elles m'ont apporté pour la réalisation de ce travail.

A mes co-internes,

Les avengers et panteras de la Pitié, Jérémy, Wiame, Fanny, Sarah, Adam, Flora. Pour les brunchs et les playlists musicales déjantées.

La promotion des anapotes Alice, Olivier, Rodolphe, Valentine, Nicole, Amel.

A mes formidables amis Benji, Jonathan et Geoffrey. Pour tous les bons moments passés ensemble, depuis la P2. Le temps passe et vous êtes toujours présents.

A ma famille,

A mes parents, à qui je dédie ce travail, pour leur amour, leur présence, leur soutien indéfectible et sans faille durant toutes ces longues et difficiles années. Papa, maman, je ne serais jamais arrivé là où j'en suis sans vous. Les mots me manquent pour vous exprimer tout l'amour et la gratitude que j'ai pour vous.

A ma sœur Emma, à qui je souhaite du fond du cœur tout le bonheur du monde.

A mes grands-parents, papi Gil, mamie Nic, papi Alain, ma tante Christiane, pour leur soutien et pour n'avoir, contrairement à moi, jamais douté de ma réussite.

A mon cousin Christophe, pour sa sagesse inspirante.

A la vie, pour ce qu'elle nous apporte.

*A la mémoire de ma grand-mère, à son courage et à sa force de caractère. Elle qui n'a jamais douté  
de ma réussite et qui ne m'a jamais vu passer le cap de la première année.  
Je lui dédie ce travail.*

# Table des matières

Liste des abréviations	7
Table des illustrations	8
Introduction	9
Etat des connaissances	10
1. Le cancer du sein	10
1.1. Epidémiologie	10
1.2. Types histologiques des carcinomes infiltrants	10
1.3. Facteurs pronostiques et prédictifs	10
2. La pathologie numérique	11
2.1. Généralités	11
2.2. Principe de la lame numérique	12
2.3. Equivalence avec la microscopie optique conventionnelle	13
3. L'intelligence artificielle	13
3.1. Contexte	13
3.2. L'apprentissage automatique ou <i>machine learning</i>	14
3.3. Concept du neurone artificiel	14
3.4. L'apprentissage profond ou <i>deep learning</i>	15
3.5. Modes d'apprentissage des algorithmes	15
3.6. Développement d'un algorithme d'intelligence artificielle en anatomopathologie	16
3.7. Défis de l'intelligence artificielle en anatomopathologie	18
3.8. Importance du pré-analytique	19
4. Intelligence artificielle en pathologie mammaire	20
4.1. Identification des lésions	20
4.2. Intelligence artificielle et carcinome infiltrant	20
4.2.1. Détection des métastases ganglionnaires	20
4.2.2. Biomarqueurs immunohistochimiques	21
4.2.3. Algorithmes de prédiction	22
5. Intelligence artificielle et établissement du grade du carcinome mammaire infiltrant	22
5.1. Généralités	22
5.2. Reproductibilité du grade	23
5.3. Différenciation glandulaire	24
5.4. Pléomorphisme nucléaire	24
6. Intelligence artificielle et score mitotique dans le carcinome mammaire infiltrant	24
6.1. Principes généraux	24
6.2. Evaluation des mitoses sur lame numérique	25
6.3. Surface de comptage et impact des moniteurs	27
6.4. Bases de données et grands challenges portant sur la détection des mitoses en pathologie mammaire humaine	28
6.4.1. MITOS12	30
6.4.2. AMIDA13	30
6.4.3. MITOS-ATYPIA14	30
6.4.4. TUPAC16	30
6.4.5. MIDOG2021	31
	5

6.5. Reproductibilité des pathologistes dans l'évaluation des mitoses sur lame numérique à l'aune de l'intelligence artificielle	31
7. Rationnel et objectif de l'étude	33
<b>Matériel et Méthodes</b>	<b>35</b>
1. Phase 1	35
1.1. Constitution de la base de données	35
1.1.1. Préparation du matériel	35
1.1.1.1. Prénnumérisation	35
1.1.1.2. Numérisation	35
1.1.2. Labellisation	36
1.1.2.1. Labellisation initiale	36
1.1.2.2. Relecture consensuelle	37
1.2. Entraînement et test de l'algorithme	39
1.2.1. Description générale du modèle	39
1.2.2. Le détecteur	39
1.2.2.1. Bases de données utilisées	39
1.2.2.2. Entraînement du détecteur	40
1.2.3. Ajout du classifieur	41
1.2.3.1. Principe du classifieur et choix des paramètres d'inférence du détecteur	41
1.2.3.2. Entraînement, test et métriques	43
1.3. Présentation de l'interface de visualisation	45
2. Phase 2 : apport de l'algorithme pour le pathologiste dans la détection des mitoses	48
2.1. Matériel d'étude	48
2.2. Déroulement de l'étude	48
3. Critères de jugements et analyse statistique	49
4. Aspects éthiques et réglementaires	50
<b>Résultats</b>	<b>51</b>
1. Reproductibilité intra-observateur des pathologistes	51
2. Reproductibilité inter-observateur des pathologistes avec et sans utilisation de l'algorithme	52
3. Reproductibilité entre les pathologistes et l'algorithme	54
4. Concordance des pathologistes pour l'évaluation de chaque figure mitotique prise individuellement	55
<b>Discussion</b>	<b>57</b>
1. Développement de l'algorithme et de l'interface de lecture	57
1.1. Étape de labellisation	57
1.2. Métriques du modèle	59
1.3. Impact des paramètres de numérisation et généralisation des performances de l'algorithme	60
1.4. Interface de visualisation de l'algorithme	61
2. Reproductibilité des pathologistes et apport de l'algorithme	62
3. Perspectives	66
<b>Conclusion</b>	<b>68</b>
<b>Bibliographie</b>	<b>69</b>
<b>Annexes</b>	<b>76</b>

# Liste des abréviations

AJCC : *American Joint Committee on Cancer*

AMIDA13 : *Assessment of Mitosis Detection Algorithm challenge 2013*

CCI : coefficient de corrélation intra-classe

CNN : *convolutional neural network* (réseau neuronal convolutif)

CE-IVD : conformité européenne pour le diagnostic in vitro

FDA : *Food and Drug Administration*

FN : faux négatif

FP : faux positif

HER2 : facteur humain de croissance épidermique de type 2

H&E : hématoxyline et éosine

HPF : *high-power field* (champ microscopique à fort grossissement)

IC : intervalle de confiance à 95%

IoU : *Intersection over Union*

KB : Kremlin-Bicêtre (base de données de mitoses de l'hôpital Bicêtre)

MIDOG2021 : *Mitosis DOmain Generalization challenge 2021*

MIN\_PROB : probabilité minimale pour considérer une détection comme valide

NB\_DET : nombre maximal de détection par image

OMS : Organisation Mondiale de la Santé

P1000 : scanner de lames PANNORAMIC 1000 (3DHISTECH)

P250 : scanner de lames PANNORAMIC 250 (3DHISTECH)

pN : *pathological node staging*

PR : *Precision – Recall*

PU-PH : Professeur des Universités – Praticien Hospitalier

RCP : réunion de concertation pluridisciplinaire

ROC : *Receiver Operating Characteristic*

RUO : *research use only*

TCGA : The Cancer Genome Atlas

TIFF : Tag Image File Format

TNM : *Tumor, Node, Metastasis*

TUPAC16 : *TUmor Proliferation Assessment Challenge 2016*

VN : vrai négatif

VP : vrai positif

# Table des illustrations

Tableau I : résumé des challenges portant sur la détection de mitoses .....	29
Tableau II : répartition des mitoses pour l'entraînement, la validation et le test du détecteur .....	40
Tableau III : répartition des mitoses et artefacts utilisés pour entraîner et tester le classifieur .....	43
Tableau IV : métriques obtenues en utilisant le détecteur seul et le détecteur combiné au classifieur .....	44
Tableau V : CCI pour l'évaluation du compte mitotique de chaque pathologiste entre leurs deux lectures .....	51
Tableau VI : kappa pondéré linéaire pour la reproductibilité intra-observateur du score mitotique .....	51
Tableau VII : CCI pour l'évaluation du compte mitotique des pathologistes avec et sans l'algorithme .....	52
Tableau VIII : kappa de Cohen pondérés linéaires portant sur la reproductibilité du score mitotique entre les pathologistes.....	53
Tableau IX : CCI pour l'évaluation du compte mitotique entre les pathologistes et l'algorithme .....	54
Tableau X : kappa pondérés linéaires pour la reproductibilité du score mitotique entre les pathologistes et l'algorithme .....	54
Tableau XI : nombre de mitoses annotées par chaque pathologiste et nombre de mitoses communes pour chaque couple de pathologistes sans assistance de l'algorithme .....	55
Tableau XII : nombre de mitoses annotées par chaque pathologiste et nombre de mitoses communes pour chaque couple de pathologistes avec assistance de l'algorithme .....	55
Tableau XIII : proportion de mitoses annotées par un ou plusieurs pathologistes sans assistance de l'algorithme .....	56
Tableau XIV : proportion de mitoses annotées par un ou plusieurs pathologistes avec assistance de l'algorithme .....	56
Tableau XV : pourcentages de mitoses pour lesquelles chaque couple d'observateur (pathologiste ou algorithme) était d'accord .....	56
Figure 1 : labellisation des mitoses et des "non-mitoses" sur Cytomine ; x20 .....	37
Figure 2 : étapes de relecture des mitoses et des "non-mitoses" .....	38
Figure 3 : changements de labels au cours de la relecture .....	38
Figure 4 : schéma général de l'algorithme .....	39
Figure 5 : cycle de la donnée lors de l'entraînement.....	42
Figure 6 : courbe ROC (à gauche), courbe PR (au centre) et matrice de confusion (à droite) pour le MobileNet.....	44
Figure 7 : échantillon de carcinome infiltrant (à gauche) ; détection du carcinome infiltrant (au centre) ; détection des mitoses et zones de <i>hotspot</i> mitotiques (à droite) .....	45
Figure 8 : détection des mitoses (carrés de petite taille) ; zones de plus forte densité mitotiques proposées par l'algorithme (carrés de grande taille) ; visualisation des mitoses détectées (onglet latéral).....	46
Figure 9 : filtre des mitoses et suppression des faux positifs.....	47
Figure 10 : dix champs contigus déposés par le pathologiste (en bleu), zone de plus forte densité proposée par l'algorithme (carré rouge de grande taille) ; mitoses détectées (carrés de petite taille) .	47

# Introduction

Le grade d'Elston-Ellis, dont l'évaluation du score mitotique est un des éléments, est un facteur pronostique majeur du carcinome mammaire infiltrant. La reproductibilité du score mitotique a fait l'objet de nombreux travaux depuis les débuts de son utilisation et les critères de comptage ont évolué au cours du temps.

Avec l'essor de la pathologie numérique, des études récentes ont montré que la détection des mitoses sur lame numérisée présentait des difficultés différentes de celles constatées sur lame de verre, pouvant être à l'origine d'une sous-estimation du score mitotique. Les progrès récents de l'intelligence artificielle pourraient néanmoins faciliter la détection des mitoses sur lame numérique par le pathologiste.

L'objectif de ce travail de thèse était de développer un algorithme de détection de mitoses et son interface utilisateur, puis de tester l'apport de cet algorithme pour le pathologiste lors de l'évaluation du score mitotique. Il a été réalisé en étroite collaboration avec les *data scientists* et développeurs de la start-up Primaà fortement impliqués sur une période d'un peu plus de 2 ans.

Nous présenterons tout d'abord l'état des connaissances dans le domaine. Après un bref rappel sur le cancer du sein et la pathologie numérique, quelques généralités sur l'intelligence artificielle seront données avant d'aborder sa place en pathologie mammaire. Nous discuterons ensuite plus spécifiquement des problèmes de reproductibilité du score mitotique, des difficultés existantes lors de la détection des mitoses sur lames numériques et de la place actuelle de l'intelligence artificielle face à cette problématique.

Dans une seconde partie, nous détaillerons les matériels et méthodes utilisés au cours de notre travail avant de présenter dans une troisième partie les résultats obtenus au cours d'une étude portant sur la reproductibilité des pathologistes dans l'évaluation du score mitotique.

Enfin, dans une dernière partie, les forces et faiblesses de notre étude seront explicitées et les résultats obtenus mis en perspective avec la littérature existante.

# Etat des connaissances

## 1. Le cancer du sein

### 1.1. Epidémiologie

Le cancer du sein est un problème majeur de santé publique. C'est le cancer le plus fréquent tous sexes confondus et la première cause de mortalité par cancer chez la femme à travers le monde. En 2020, plus de 2,2 millions de cas ont été recensés et on estime qu'il a été la cause d'environ 685 000 décès (1). En France, il est au premier rang des cancers incidents (environ 58 500 nouveaux cas diagnostiqués chaque année) et c'est le plus meurtrier (14% des décès féminins par cancer soit 12 146 décès en 2018). Toutefois, lorsqu'il est pris en charge à temps, c'est un cancer de bon pronostic. La survie nette standardisée sur l'âge a augmenté au fil du temps ; elle est actuellement de 87% à 5 ans et de 76% à 10 ans (2).

### 1.2. Types histologiques des carcinomes infiltrants

La classification OMS 2019 des tumeurs épithéliales mammaires distingue une trentaine de carcinomes de types histologiques différents. Certains sont des variantes morphologiques du carcinome infiltrant sans type spécifique, d'autres correspondent à des sous-types particuliers qu'il convient d'identifier du fait de leur prise en charge spécifique ou de leur pronostic distinct (3).

Les carcinomes infiltrants sans type spécifique sont les plus fréquents (80% des cas). Cette catégorie regroupe de nombreuses entités différentes sur le plan morphologique, immuno-phénotypique et moléculaire. Il s'agit d'un diagnostic d'élimination retenu après exclusion de tous les sous-types spéciaux. Le carcinome lobulaire infiltrant est le premier sous-type spécial en terme de fréquence, regroupant entre 5 et 15% des carcinomes infiltrants. Les autres sous-types, rares, complètent ensemble les 5 à 15% restants.

### 1.3. Facteurs pronostiques et prédictifs

Le rôle du pathologiste dans la prise en charge du cancer du sein n'est pas seulement de donner le sous-type histologique, mais également de déterminer avec précision des paramètres pronostiques et prédictifs permettant de guider la prise en charge thérapeutique des patientes.

Les facteurs pronostiques permettent de prédire l'évolution de la maladie en termes de rechute locale et métastatique ainsi que le risque de décès en présence ou en absence de traitement. Ils permettent d'identifier les patientes requérant une thérapie complémentaire.

Les facteurs prédictifs permettent de prévoir la réponse de la tumeur à un traitement néo-adjuvant ou adjuvant particulier.

Ces facteurs peuvent être cliniques, anatomiques, histologiques, immunohistochimiques ou moléculaires (4). Certains d'entre eux comme le stade TNM ou le grade sont utilisés depuis maintenant plusieurs décennies dans l'évaluation du cancer du sein et les perspectives offertes par la pathologie numérique et l'intelligence artificielle sont à même de révolutionner l'approche du pathologiste dans leur évaluation.

## 2. La pathologie numérique

### 2.1. Généralités

Depuis le développement du premier microscope virtuel en 1997 (5), la pathologie numérique a considérablement progressé et se déploie maintenant dans les laboratoires de pathologie. L'amélioration de la qualité des lames numériques et de la vitesse de numérisation des scanners permet aujourd'hui d'envisager l'utilisation de ces outils en routine diagnostique.

Cependant, elle requiert des investissements matériels importants. En plus des scanners de lames, des moniteurs d'affichage en haute résolution et des serveurs de stockage de grande capacité sont nécessaires pour l'analyse et l'archivage des lames numériques. Les infrastructures réseaux et les processeurs/cartes graphiques des ordinateurs doivent permettre le chargement des données dans des délais suffisamment courts pour ne pas entraver le travail du pathologiste. Le système de gestion d'image doit être ergonomique, pratique d'utilisation et parfaitement compatible avec le système de gestion du laboratoire (6).

Bien que cette approche nécessite des procédures de validation du matériel, un temps d'adaptation et d'apprentissage ainsi qu'une réorganisation du travail pour le pathologiste et l'équipe technique (7), elle peut *in fine* permettre au pathologiste de gagner du temps dans son analyse en améliorant le confort de lecture, facilitant les mesures et la navigation sur l'échantillon et permettant de réaliser des analyses comparatives aisées des images morphologiques et immunohistochimiques (8). De plus les applications potentielles de la pathologie numérique au-delà du diagnostic de routine sont multiples : utilisation en

recherche, partage d'image en ligne pour l'enseignement ou la télé-expertise (9), télédiagnostic extemporané (10,11), facilitation des échanges entre pathologistes ou avec les cliniciens lors des RCP. Il s'agit en outre de la base indispensable au développement des outils d'analyses d'images automatisées (6).

## 2.2. Principe de la lame numérique

Une lame numérique présente une architecture pyramidale obtenue par scan en tuiles ou en bandes contiguës d'une lame de verre au grossissement maximal souhaité. Des algorithmes de redimensionnement permettent d'obtenir les tuiles aux grossissements intermédiaires. L'image finale peut avoir une taille de plusieurs centaines de milliers de pixels et peser plusieurs giga-octets. Elle correspond à une représentation en deux dimensions sur un seul plan focal de la lame de verre (objet tridimensionnel ayant une épaisseur de 3 à 5  $\mu\text{m}$ ). Pour permettre au pathologiste d'accéder aisément à la lame sans avoir à charger l'ensemble du fichier, un système de flux continu (*streaming*) actualise en permanence les données affichées à l'écran (8,12).

Le poids du fichier et le temps de numérisation vont croissant avec la surface de l'échantillon à scanner, la résolution, le grossissement, le niveau de compression du fichier ainsi que le nombre de plans focaux souhaités en cas de numérisation multi-plans (*z-stacking*) (6). Il s'agit donc de trouver un compromis entre ces différents paramètres permettant une utilisation en routine.

La qualité d'une lame numérique dépend avant tout de la qualité de la lame de verre elle-même et donc de toute la phase pré-analytique. La qualité des objectifs et du capteur ainsi qu'un réglage correct du scanner en termes d'exposition, de nombre de points de focus et de compression du fichier seront les garants d'une lame numérique exploitable pour le diagnostic.

Il est possible après la numérisation de faire varier les paramètres de l'image (bruit de fond, contraste, standardisation des couleurs) pour optimiser le rendu et corriger des défauts mineurs. Cependant une même lame de verre peut également avoir un rendu numérique différent en fonction du scanner, du moniteur ou du logiciel de visualisation employé. La perte d'information résultant de la compression peut être acceptable si le rendu visuel de l'image est identique, toutefois elle peut avoir un impact en cas d'analyse d'image secondaire par des algorithmes.

## 2.3. Equivalence avec la microscopie optique conventionnelle

La non infériorité de la lame numérique par rapport à la lame de verre pour le diagnostic a été démontrée par de nombreuses études dans tous les domaines de l'anatomopathologie (13–15). Dans une méta-analyse récente ayant compilé les résultats de 25 études, le taux de concordance était supérieur à 98% (16). Les discordances entre lame de verre et lame numérique s'observent essentiellement sur l'appréciation des lésions épithéliales pré-néoplasiques et des atypies nucléaires, la détection des micro-organismes comme *Helicobacter pylori* ou la détection de micrométastases.

En pathologie mammaire, les discordances concernent majoritairement l'appréciation du grade d'Elston-Ellis et en particulier le comptage des mitoses (17), mais également des lésions d'interprétation difficile pour lesquelles des problèmes de reproductibilité sont connus entre pathologistes sur lame de verre (différenciation entre hyperplasie canalaire simple / atypique / carcinome *in situ*, entre tumeurs phyllodes bénignes et fibroadénomes et identification des foyers de micro-invasion ou d'emboles lympho-vasculaires) (18,19).

## 3. L'intelligence artificielle

### 3.1. Contexte

L'intelligence artificielle définit l'ensemble des techniques permettant à une machine d'imiter ou de surpasser l'intelligence humaine. C'est à la fin des années 50 qu'apparaissent les premiers algorithmes de reconnaissance d'image (20). Les progrès considérables de l'informatique et de l'intelligence artificielle réalisés depuis, particulièrement au cours des deux dernières décennies, ont permis d'élargir les applications de ces outils au domaine médical (21). Des algorithmes sont ainsi capables d'évaluer des lésions cutanées (22) ou des images de fond d'œil (23). En radiologie, des outils d'aide à la détection de lésions mammographiques ou scannographiques suspectes ont récemment été développés (24,25). Les progrès couplés de la pathologie numérique et de l'intelligence artificielle permettent aujourd'hui d'envisager des analyses de lames numérisées par des algorithmes capables d'assister le pathologiste dans son travail (26).

En 2021, la FDA a approuvé le premier algorithme d'intelligence artificielle capable d'assister le pathologiste dans l'évaluation des lésions prostatiques (27).

### 3.2. L'apprentissage automatique ou *machine learning*

L'apprentissage automatique ou *machine learning* est un sous-domaine de l'intelligence artificielle dans lequel des méthodes statistiques sont utilisées pour améliorer les performances d'un algorithme dans une tâche à partir d'un ensemble de données fournies (images, mots). Le but est de développer un modèle capable de prédire un résultat sur de nouvelles données inconnues. Cette approche diffère de la programmation traditionnelle où les règles sont explicitement mises en œuvre au départ et où les performances de l'algorithme sont figées.

En pratique, il s'agit pour un algorithme de prédire une sortie  $y$  correspondant à la classe de l'objet à partir d'une entrée  $x$  correspondant aux différents paramètres analysables de l'image, ce qui se matérialise par une fonction  $y = f(x)$ . L'algorithme analyse les images et apprend par essais-erreurs successifs en faisant varier l'importance de chaque paramètre de l'image avec comme objectif de trouver la fonction  $f()$  optimale à partir des données d'apprentissage. C'est le principe de l'apprentissage supervisé. Lorsque le modèle développé est satisfaisant on peut l'utiliser pour prédire la classe d'une nouvelle image. On mesure alors la qualité du modèle sur des données de test inconnues de l'algorithme (28).

### 3.3. Concept du neurone artificiel

Le cerveau humain contient environ 86 milliards de neurones en réseaux capables d'envoyer un signal à travers leur axone si pendant un intervalle de temps donné le signal d'entrée (somme des signaux excitateurs et inhibiteurs reçus à travers les dendrites) est supérieur au seuil d'activation. C'est le principe du potentiel d'action caractérisant l'influx nerveux.

Le neurone artificiel est une fonction mathématique capable de traiter de multiples variables d'entrées. Les paramètres propres à cette fonction peuvent être modifiés par le programmeur selon les objectifs souhaités et les variables analysées. Ainsi la somme des entrées est effectuée, pondérée de façon variable en fonction des connexions, de la même façon que dans un réseau de neurones biologiques certaines synapses ont plus d'influence que d'autres. Le résultat obtenu permet la transmission au neurone suivant du signal qui est produit via une fonction d'activation si les informations transmises au neurone artificiel en entrée sont suffisamment importantes, à l'image du potentiel d'action du neurone biologique (29).

### 3.4. L'apprentissage profond ou *deep learning*

L'apprentissage profond ou *deep learning* est une approche récente d'apprentissage automatique utilisant des réseaux de neurones artificiels, organisés en couches successives.

L'architecture d'un réseau est principalement définie par le nombre de couches cachées et le nombre de neurones par couche cachée. Ces deux éléments conditionnent le nombre de paramètres (de poids) à estimer et donc la complexité du modèle. Ils participent à la recherche d'un bon compromis entre qualité d'apprentissage et qualité de prévision.

La propagation du signal dans le réseau (*forward propagation*) permet de décomposer l'analyse de l'objet (couleur, texture, forme...) et de transférer à la couche suivante les informations pertinentes extraites (les *features*) dans le but de classer ou d'exploiter les données analysées de façon adéquate.

Lors d'une analyse d'image, le résultat obtenu à la sortie du réseau est confronté au label initial. Un calcul de l'erreur est renvoyé de la sortie vers l'entrée du réseau afin de réajuster le poids attribué à chaque connexion neuronale. C'est ce qu'on appelle la rétro-propagation du gradient (*backward propagation*). En pratique, il s'agit donc de répondre à un problème d'optimisation afin de minimiser l'erreur. On parle aussi de minimisation de fonction de perte ; cette fonction permettant de quantifier les différences entre les prédictions du modèle et les labellisations réelles attendues (30).

Dans cette démarche, il est important d'éviter le sur-apprentissage de l'algorithme. Cela l'empêcherait par la suite de généraliser ses résultats à de nouvelles données (l'algorithme ne serait en quelque sorte capable de classer que des images qu'il connaît déjà).

Pour ce faire, il est par exemple possible d'augmenter les données en utilisant des techniques de déformation d'image ou de modification des couleurs. Il est également possible d'abandonner certaines unités du réseau, à l'image de la plasticité neuronale (31).

### 3.5. Modes d'apprentissage des algorithmes

Trois modes d'apprentissage sont classiquement décrits en apprentissage automatique : l'apprentissage supervisé, non supervisé et semi-supervisé.

L'apprentissage supervisé permet de former l'algorithme en utilisant des données labellisées, le plus souvent par un humain expert du domaine.

L'apprentissage non supervisé permet aux algorithmes d'apprendre sur des données non labellisées. Il répond à des problématiques spécifiques comme l'identification de caractéristiques non visibles à l'œil nu pouvant être corrélées aux données cliniques ou le regroupement de tumeurs en sous-types selon

leurs caractéristiques. Leur évaluation peut s'avérer délicate en l'absence de référentiel et leur méthode de classification difficile à conceptualiser. Les résultats peuvent être imprévisibles et difficiles à comprendre (on ne sait pas toujours quelles sont les caractéristiques et les voies neuronales utilisées par l'algorithme pour effectuer son classement. C'est le phénomène dit de la "boîte noire"). Clarifier le fonctionnement de ces modèles est un sujet d'étude important pour améliorer la transparence des résultats et la confiance de l'utilisateur (6).

L'apprentissage semi-supervisé peut être utile dans les situations où la génération d'un ensemble de données labellisées est difficile. Les algorithmes développés utilisent à la fois des données labellisées et non labellisées. L'idée est de coupler des méthodes non supervisées telles que le partitionnement de données, permettant de regrouper en paquets les objets non labellisés présentant des caractéristiques communes, avec des méthodes supervisées fonctionnant par exemple à partir de quelques images labellisées. Ces modèles partiellement entraînés peuvent être employés pour labelliser eux même les données, ce qui permet de réduire le coût de l'annotation manuelle. L'algorithme demande de façon interactive l'avis de l'expert pour les données ambiguës et un contrôle humain peut donc s'avérer nécessaire pour s'assurer de la pertinence des données labellisées (32).

### 3.6. Développement d'un algorithme d'intelligence artificielle en anatomopathologie

Le développement d'un algorithme d'intelligence artificielle en anatomopathologie requiert une collaboration étroite entre les développeurs et les médecins pathologistes.

L'objectif de l'algorithme est d'assister le pathologiste dans son analyse. Les problématiques sont variées et peuvent intéresser des comptages de cellules, des mesures, des analyses morphologiques, voir l'identification de caractéristiques non visibles à l'œil-nu qui auraient une valeur pronostique ou prédictive. Les éléments analysés par les pathologistes vont être utilisés par les développeurs pour définir le fonctionnement de l'algorithme (33).

Dans un premier temps, il est nécessaire pour les pathologistes de constituer une base de données de référence permettant d'entraîner l'algorithme à la reconnaissance des critères discriminants. Les images labellisées constituant cette base de données sont réparties en plusieurs ensembles. L'ensemble d'entraînement constitue le socle d'images sur lequel l'algorithme va apprendre. Parallèlement un ensemble de validation est utilisé pour vérifier que l'algorithme apprend de façon adéquate et n'entre pas dans une phase de sur-apprentissage qui l'empêcherait alors de généraliser ses performances à

d'autres données. Lorsque le comportement souhaité est obtenu, on peut alors tester le modèle sur un troisième ensemble d'images labellisées inconnues de l'algorithme. Des métriques sont alors mesurées pour évaluer ses performances. Si les résultats sont concluants et que les performances sont constantes, il est possible de tester le modèle sur des lames numériques vierges de toute labellisation (29).

Dans le cadre de la détection d'objets par apprentissage supervisé comme c'est majoritairement le cas pour les mitoses, ces métriques dépendent du nombre de vrais et faux positifs et de vrais et faux négatifs. Ce sont:

- La sensibilité (*recall*) :  $VP / (VP + FN)$ , correspondant au rapport entre les observations positives correctement prédites et toutes les observations de la classe réelle.
- La précision (valeur prédictive positive) :  $VP / (VP + FP)$ , correspondant au rapport entre les observations positives correctement prédites et le total des observations positives.
- L'exactitude (*accuracy*) :  $(VP + VN) / (VP + VN + FP + FN)$ , correspondant au rapport entre les observations correctement prédites et les observations totales.
- Le F1-score :  $2 \times (\text{précision} \times \text{sensibilité}) / (\text{précision} + \text{sensibilité})$ , correspondant à la métrique de référence, la plus utilisée pour évaluer la performance des algorithmes de détection car il prend en compte autant les faux positifs que les faux négatifs. C'est la moyenne pondérée de la précision et de la sensibilité. Lors de l'évaluation du score mitotique, une sous-estimation du nombre de mitoses peut être aussi préjudiciable qu'une surestimation. C'est pourquoi ce critère apparaît comme pertinent car il a l'avantage de combiner la précision et la sensibilité dans un seul résultat. Un F1-score proche de 1 indique une excellente performance de détection.

Les algorithmes peuvent fournir des analyses précises et reproductibles, mais ils peuvent aussi avoir leurs propres biais et produire des résultats inexacts. Ils peuvent néanmoins toujours être utiles pour prédire l'évolution d'une maladie ou la nécessité d'un traitement lorsqu'ils sont validés dans des études cliniques (34). Il n'est pas forcément nécessaire pour un algorithme qu'il ait une précision et une sensibilité de 100% mais un niveau de performance suffisant doit être atteint selon l'utilisation qui en est faite, corrélé avec les objectifs généraux de l'analyse (35).

L'entraînement et le test de l'algorithme sont ainsi répétés jusqu'à atteindre les performances souhaitées. Ceci permet le marquage RUO (*research use only*) de l'algorithme correspondant à une validation interne de ses performances. Pour obtenir le marquage CE-IVD correspondant à une

validation clinique de l'algorithme, il est nécessaire d'évaluer la reproductibilité des pathologistes avec et sans son utilisation. Les résultats se doivent également d'être reproductibles quelles que soient les conditions pré-analytiques du prélèvement et en particulier quel que soit le scanner de lames utilisé.

L'intégration de l'algorithme une fois validé doit être réfléchi pour ne pas ralentir le flux de travail (*workflow*) du laboratoire. Son interface doit être ergonomique et optimisée pour une utilisation en routine. Une fois implémenté, l'expertise du pathologiste reste indispensable pour s'assurer de la corrélation entre les données clinico-pathologiques et l'analyse d'image. La pertinence des résultats de l'algorithme doit être évaluée pour chaque situation (33).

### 3.7. Défis de l'intelligence artificielle en anatomopathologie

L'intelligence artificielle fait face à plusieurs difficultés dans son développement en anatomopathologie (36).

La première est le manque de données labellisées disponibles pour l'entraînement d'algorithmes. La labellisation est une tâche chronophage et peu de laboratoires sont entièrement équipés en pathologie numérique ce qui limite l'existence des bases de données disponibles. La labellisation peut également être ardue pour des images ambiguës où la reproductibilité inter-observateur est faible. Un paradoxe est que les situations pour lesquelles un algorithme est le plus utile sont aussi celles pour lesquelles les pathologistes sont le moins reproductibles. Cela peut être à l'origine de labellisations erronées et compromettre l'apprentissage de l'algorithme à la source.

Un algorithme doit par ailleurs être capable de composer avec l'importante hétérogénéité tissulaire des prélèvements examinés. Le caractère mono-tâche des algorithmes actuellement développés est à mettre en perspective avec le nombre important d'analyses et de mesures simultanées faites par le pathologiste dans sa pratique. La plupart des algorithmes répondent à des problématiques binaires de type bénin ou malin, mitose ou "non-mitose". Cette logique réductrice ne tient pas compte des données contextuelles et en pratique de nombreux diagnostics sont parfois nuancés par le pathologiste pour souligner certaines spécificités propres au cas examiné et ne pouvant être prises en compte à l'heure actuelle par les outils d'intelligence artificielle.

Compte tenu de la grande taille des lames numériques, il est difficile pour un algorithme de procéder à une analyse directe de l'image. Il est en effet nécessaire de fractionner l'analyse à partir de patches de plus petite taille pouvant conduire à perdre des informations contextuelles nécessaires au diagnostic. Aussi la fiabilité des algorithmes peut être mise à mal. La modification imperceptible à l'œil de pixels sur

une image peut totalement biaiser l'analyse d'un algorithme et une potentielle évolution dans la qualité des images analysées entre la phase d'entraînement, celle de test et l'utilisation en routine, (par exemple changement du scanner de lames, modification des paramètres de numérisation) pourrait compromettre ses performances. Il serait nécessaire de s'assurer de façon régulière par un contrôle qualité que les analyses de l'algorithme restent reproductibles dans le temps.

Enfin, l'équipement d'un laboratoire entièrement en pathologie numérique a un coût élevé. Les solutions d'intelligence artificielle développées requièrent pour fonctionner des processeurs puissants et coûteux qui viendront se surajouter et le prix de ces algorithmes pourrait donc freiner leur généralisation, un modèle économique d'utilisation n'étant pas encore défini.

### 3.8. Importance du pré-analytique

Les performances des algorithmes dépendent de la qualité des lames numérisées et des étapes pré-analytiques. Les artefacts susceptibles de perturber les analyses d'un algorithme peuvent être en lien avec les étapes pré-numérisation. Leurs causes sont variées : mauvaise fixation, type de fixateur utilisé, protocole de déshydratation, épaisseur et qualité de la coupe au microtome, type et protocole de coloration, qualité et conservation des colorants, épaisseur de la lamelle, épaisseur et séchage du milieu de montage, présence de plis, de déchirures, de décollements ou de bulles d'air.

Des artefacts peuvent aussi être liés à l'étape de numérisation et alors rattachés aux spécificités intrinsèque du scanner de lames ou au protocole de numérisation faisant intervenir de multiples paramètres (luminosité, contraste, gestion des couleurs et de la balance des blancs, points de focus, taux de compression du fichier, gestion du flou). Différentes techniques de normalisation des couleurs et de déformation des images peuvent être utilisées par les programmeurs pour essayer de pallier cette variabilité des colorations et améliorer les performances des algorithmes en augmentant les bases de données, mais ils ne peuvent compenser des défauts trop importants (37).

Pour les algorithmes évaluant de l'immunohistochimie, le type d'anticorps, sa dilution, le démasquage antigénique, la révélation ainsi que la contre-coloration auront un impact sur l'analyse (18).

Une standardisation rigoureuse des phases pré-analytiques dans chaque laboratoire est donc un prérequis indispensable pour limiter au maximum la fluctuation des performances des algorithmes en aval. Dans tous les cas, une calibration de l'algorithme avec les lames du laboratoire sera nécessaire avant son utilisation (38,39).

## 4. Intelligence artificielle en pathologie mammaire

Les applications de l'intelligence artificielle en pathologie mammaire sont l'objet de nombreux travaux, en particulier dans le cas des carcinomes infiltrants.

### 4.1. Identification des lésions

Un algorithme d'intelligence artificielle intégré de façon automatique dans le flux de travail d'un laboratoire de pathologie et appliqué à une lame entière doit être capable d'identifier et différencier différentes lésions. Cette capacité est un prérequis pour s'assurer dans un second temps que l'évaluation du grade ou d'autres paramètres par différents algorithmes soit bien faite sur le territoire d'intérêt (40). En pathologie mammaire, cette étape peut s'avérer délicate particulièrement dans le contexte d'une tumeur de bas grade qui présenterait des caractéristiques morphologiques proches de celles du tissu mammaire normal ou compte tenu de la grande variété morphologique des lésions bénignes (41).

La reproductibilité des pathologistes pour évaluer la bénignité ou malignité d'une lésion étant excellente pour la grande majorité des cas (42,43), ce type d'algorithme ne serait donc pas tant là pour aider le pathologiste dans l'affirmation du caractère malin d'une lésion mais plutôt pour l'aider à repérer les zones d'intérêt ou l'assister dans l'évaluation de la taille tumorale. Ils pourraient également avoir un intérêt dans la recherche de zones tumorales de petite taille comme par exemple dans l'évaluation d'une tumeur post-chimiothérapie.

Toutefois dans une étude récente, il est apparu qu'un tel algorithme pouvait être générateur de confusion chez les pathologistes moins expérimentés. Des niveaux d'exactitude élevés doivent être requis pour espérer déployer ce type d'outil en routine diagnostique (44).

### 4.2. Intelligence artificielle et carcinome infiltrant

Parallèlement à l'établissement du grade d'Elston-Ellis, dont la reproductibilité et la place de l'intelligence artificielle dans son évaluation sont discutées à la *section 5*, d'autres algorithmes sont en cours de développement pour l'évaluation du carcinome mammaire infiltrant.

#### 4.2.1. Détection des métastases ganglionnaires

Plusieurs challenges portant sur la détection des métastases ganglionnaires sur lame H&E, ont montré que des algorithmes pouvaient obtenir des résultats comparables à ceux obtenus par les pathologistes

pour la détection des macro comme des micro-métastases, mais leurs performances pour la détection des cellules tumorales isolées restent limitées (45,46).

L'intelligence artificielle pourrait à l'avenir trouver sa place en complément des niveaux de coupe et de l'immunohistochimie réalisés en routine pour l'évaluation du statut ganglionnaire. Dans l'étude de Steiner *et al.*, la détection des micro-métastases par les pathologistes était significativement améliorée par l'utilisation de l'algorithme (47).

Toutefois les algorithmes ne peuvent pas contextualiser la métastase sur l'ensemble de la lame (nombre de ganglions par lame, ganglion coupé en deux ou deux ganglions distincts), le pathologiste étant le seul à même de replacer la métastase détectée dans le contexte de l'échantillonnage macroscopique pour donner le stade pN.

#### 4.2.2. Biomarqueurs immunohistochimiques

La détermination systématique en immunohistochimie de l'expression des biomarqueurs pronostiques des carcinomes mammaires (récepteurs aux œstrogènes et à la progestérone, HER2 et Ki67) permet de guider les décisions thérapeutiques pour la prise en charge des patientes (48).

L'évaluation du pourcentage d'expression de ces marqueurs est toutefois sujette à une certaine variabilité inter-observateur (49,50).

La standardisation des techniques d'immunohistochimie et l'amélioration de l'apprentissage de la lecture des lames et des méthodes de comptage par les pathologistes permettrait de réduire cette variabilité (51,52). Des outils d'analyse automatisée basés sur l'intelligence artificielle pourraient assister le pathologiste dans l'évaluation de ces biomarqueurs et plusieurs études ont montré une bonne corrélation entre analyse automatisée et analyse manuelle aussi bien dans le cas des récepteurs hormonaux (53–55) que dans celui de l'évaluation du Ki67 (56,57) et de HER2 (58). Cependant la plupart de ces modèles ne permettaient pas une analyse sur lame entière numérisée ; d'autres algorithmes sont en cours de développement par des industriels pour résoudre cette problématique.

Dans le cas de HER2, l'emploi de ce type d'outil pourrait réduire l'usage de l'hybridation *in-situ* comme technique complémentaire (59,60). Dans le cas de Ki67 l'usage de tels algorithmes permettait d'obtenir une excellente reproductibilité des pathologistes (61).

Quel que soit le biomarqueur analysé, des conditions pré-analytiques optimales sont requises pour effectuer un comptage fiable avec ce type d'outil. Des contrôles qualité et des tests doivent être mis en œuvre en interne pour s'assurer du maintien de la fiabilité de l'algorithme (62).

### 4.2.3. Algorithmes de prédiction

Des approches basées sur l'intelligence artificielle sont en cours de développement pour prédire la classe histo-moléculaire des carcinomes infiltrants et les résultats de tests génomiques, témoignant de l'existence de caractéristiques intrinsèques de l'image invisibles pour l'œil humain pouvant être prédictives de rechute ou de la réponse au traitement (63–65). Des scores pronostics permettant de stratifier les patientes selon le risque de décès pourraient ainsi émerger dans un avenir proche (66). Dans le cas des carcinomes canaux *in situ* de haut grade, d'autres modèles pourraient prédire le risque de récurrence homolatérale à 10 ans et donc permettre d'identifier les patientes à même de bénéficier d'un traitement complémentaire (67).

## 5. Intelligence artificielle et établissement du grade du carcinome mammaire infiltrant

### 5.1. Généralités

L'établissement d'un grade histologique directement corrélé à l'agressivité tumorale et à la survie des patientes constitue une problématique majeure en pathologie mammaire depuis le début du XX<sup>ème</sup> siècle (68). De nombreux travaux ont permis d'aboutir au grade d'Elston-Ellis tel qu'il est utilisé de nos jours en démontrant sa forte valeur pronostique et en améliorant l'objectivité des pathologistes dans son évaluation (69–72). Son utilisation est recommandée quel que soit le sous-type histologique du carcinome infiltrant (73) et il est intégré depuis 2017 dans les items de stadification du cancer selon la 8<sup>ème</sup> édition de l'AJCC (74). Sa détermination nécessite l'évaluation du pourcentage de différenciation glandulaire, du degré d'atypies des noyaux et du nombre de mitoses pour une surface donnée. Chaque item étant scoré de 1 à 3, leur somme comprise entre 3 et 9 permet de définir le grade (I, II ou III) du carcinome infiltrant.

La valeur pronostique du grade a été validée par de nombreuses études de survie à long terme. Il constitue un facteur pronostique corrélé au risque métastatique à distance et à la survie globale des patientes, indépendamment du statut ganglionnaire ou de la taille tumorale (75,76). Il a un rôle clé dans la décision thérapeutique, en particulier pour les tumeurs hormonosensibles HER2 négatives. Dans une étude de cohorte menée aux Pays-Bas entre 2013 et 2016, le grade constituait le facteur décisionnel déterminant d'une chimiothérapie adjuvante et/ou d'une hormonothérapie pour respectivement 35% et 30% des patientes. C'est également un facteur pouvant justifier d'une désescalade thérapeutique (77).

Les quelques algorithmes générés pour déterminer le grade des carcinomes mammaires infiltrants de façon globale se sont avérés décevants et ne résolvent pas à ce jour son manque de reproductibilité (64,78,79). La plupart des travaux se concentrent maintenant sur le critère “mitose”, élément du grade ayant la plus forte valeur pronostique et posant à la fois des problèmes de reproductibilité inter-observateur et de détection sur lame numérique (80–82).

Après une présentation des difficultés de reproductibilité de l'établissement du grade histo-pronostique des carcinomes mammaires infiltrants tant sur lame de verre que sur lame numérique, les travaux actuels en intelligence artificielle pour l'établissement du score de différenciation glandulaire et celui du pléomorphisme nucléaire seront cités avant d'aborder plus en détail dans la *section 6*. l'établissement du score mitotique.

## 5.2. Reproductibilité du grade

La reproductibilité de la détermination du grade dépend de l'expérience du pathologiste et des conditions pré-analytiques des prélèvements (76).

La variation inter-observateur dans une série de 17 publications entre 1990 et 2020 rapporte des valeurs de kappa comprises entre 0,43 et 0,85 (83). Dans une étude d'assurance qualité réalisée au Royaume-Uni en 2016, l'évaluation des mitoses était l'item sur lequel les pathologistes étaient le moins reproductibles ( $\kappa = 0,34$ ) suivi de près par celles des noyaux ( $\kappa = 0,36$ ) et de l'architecture ( $\kappa = 0,49$ ) (43).

Dans l'étude de Rabe *et al.* (84) portant sur l'analyse de 100 lames H&E de carcinomes infiltrants par six pathologistes, une discordance du grade était observée dans près de la moitié des cas.

Plusieurs études ayant comparé la reproductibilité du grade entre lame de verre et lame numérique retrouvent des niveaux de concordance modérés comparable à ceux observés entre plusieurs observateurs sur lame de verre (81,85–87). Dans ces études, l'association significative avec la survie des patientes était conservée. Néanmoins dans les études de Rakha *et al.* des discordances de grade avec l'évaluation sur lame de verre étaient relevées dans 32% des cas avec une proportion de grade I, II et III de 16%, 41% et 43% sur lame de verre et de 28%, 40% et 32% sur lame numérique. Les patientes qui avaient des tumeurs discordantes présentaient une survie intermédiaire entre les deux grades se rapprochant de façon statistiquement significative du grade le plus élevé. Cette observation souligne le fait que le grade constitue un continuum lésionnel dont les *cut-off* ne séparent pas des entités distinctes (81,85).

Dans l'étude de Davidson *et al.* portant sur vingt-deux cas évalués par un panel de 208 pathologistes, la reproductibilité du grade était significativement plus élevée sur lame de verre que sur lame numérique. Néanmoins les auteurs ont suggéré que cette différence pouvait en partie s'expliquer par la faible expérience des pathologistes de l'étude à la lecture des lames numériques (86).

### 5.3. Différenciation glandulaire

Les quelques algorithmes publiés sur la différenciation glandulaire ne permettent pas d'envisager une utilisation en routine diagnostique (88,89). Ils se concentrent essentiellement sur la segmentation des structures glandulaires, prérequis à une bonne estimation du score (90,91). Un algorithme d'évaluation du score architectural devrait tenir compte de l'hétérogénéité morphologique des carcinomes mammaires (glandes de taille et de forme variées, nids cellulaires, ou massifs cribriformes). De plus les adipocytes parfois nombreux, ainsi que les artefacts de rétraction liés à la fixation formolée constituent les principaux pièges à une analyse automatisée.

### 5.4. Pléomorphisme nucléaire

De même, les performances des algorithmes développés jusqu'à présent ayant pour objet l'évaluation des noyaux ne sont pas suffisantes pour envisager une utilisation en routine diagnostique (92). Les études se concentrent aussi sur la segmentation, tâche difficile en particulier compte tenu des chevauchements nucléaires (93–95) et sur la distinction entre noyaux bénins et malins (96). L'utilisation des données morphométriques telles que la surface nucléaire moyenne constitue une piste intéressante pouvant être corrélée à la survie des patientes (97). Une approche, basée sur 6 caractéristiques morphologiques (taille des noyaux, taille des nucléoles, épaisseur de la membrane nucléaire, régularité des contours nucléaires, densité de la chromatine et degré d'anisonucléose), a été utilisée lors du challenge MITOS-ATYPIA 2014, sans que les résultats des algorithmes aient été améliorés (98).

## 6. Intelligence artificielle et score mitotique dans le carcinome mammaire infiltrant

### 6.1. Principes généraux

Le score mitotique est le reflet de l'activité proliférative de la tumeur. Il est le seul des trois items du grade d'Elston-Ellis à avoir une valeur pronostique indépendante quel que soit la taille tumorale ou le statut ganglionnaire (80). Pour autant, les critères morphologiques d'évaluation des mitoses et la

méthode de comptage ont évolué depuis l'introduction du grade de Scarff, Bloom et Richardson à la fin des années 50 (70,71,99).

Le compte mitotique est établi le plus souvent en périphérie de la tumeur mais il n'y a pas de recommandations précises pour le choix de la zone de comptage. Certains critères tels que le *pattern* architectural, la densité cellulaire, la présence de nécrose ou l'absence de carcinome canalaire *in-situ* au voisinage peuvent néanmoins orienter le pathologiste (100).

Ce comptage est parfois fastidieux. Avant l'apparition des premiers scanners de lames, une étude de faisabilité d'une détection automatisée avait été réalisée en 1983, basée sur des images photographiques argentiques en noir et blanc prises à l'objectif x100 en immersion. Les résultats avaient alors été considérés comme décevants (101).

Par la suite, la possibilité d'appliquer un algorithme de détection de mitoses sur des images de lames capturées à l'aide d'une caméra optique a été testée en 1991 avec des résultats encourageants (102), mais la résolution des images analysées et les contraintes matérielles nécessaires ne permettaient pas d'envisager une utilisation de ce type d'outil en routine et il faut attendre les années 2000 pour que la détection automatique des mitoses connaisse un regain d'intérêt.

Avant de définir le rationnel et les objectifs de notre travail, nous allons présenter les résultats des études réalisées en pathologie numérique pour la reconnaissance des mitoses et la détermination de la surface de comptage puis la synthèse des challenges réalisés portant sur la détection des mitoses en pathologie mammaire humaine.

## 6.2. Evaluation des mitoses sur lame numérique

La vis micrométrique du microscope optique permet de naviguer sur la totalité de l'épaisseur de la coupe de façon fluide et instantanée. Cet outil est utilisé par les pathologistes pour affiner leur analyse des mitoses. La numérisation multi-plans (*z-stack*) est proposée sur certains scanners. Elle n'est toutefois pas utilisable actuellement en routine diagnostique du fait d'un allongement considérable du temps de numérisation, d'une augmentation du poids des fichiers posant des problèmes de stockage et d'un manque de fluidité et d'instantanéité à la lecture. De plus, le nombre de plans z numérisés reste très inférieur à celui accessible sur lame de verre par l'utilisation de la vis micrométrique.

Les mitoses peuvent présenter des similarités morphologiques avec d'autres objets tels que les corps apoptotiques, les noyaux carryorhexiques, hyperchromatiques ou artéfactés, voire même avec des cellules inflammatoires. Les critères d'identification des mitoses sur lame de verre édictés par van Diest

en 1992 (103) semblent imparfaitement adaptés au contexte de la pathologie numérique et les contraintes inhérentes à cette méthode d'analyse nécessitent de définir de nouvelles approches et/ou nouveaux critères pour évaluer les mitoses (81). En effet la numérisation sur un seul plan focal rend l'identification des mitoses différente sur lame numérique ; il existe alors, en absence d'apprentissage adapté, un risque de sous-estimation du score mitotique. L'équipe de Nottingham a récemment publié une mise au point sur la détection des mitoses sur lame numérique (104).

Dans une étude comparative entre lames de verre et lame numérique réalisée en 2021 (105), ces derniers ont retenu huit critères morphologiques analysables permettant de déterminer le caractère mitotique d'une cellule sur lame numérique : l'absence de membrane nucléaire, l'irrégularité des contours nucléaires, la texture de la chromatine et sa répartition dans la cellule, la présence d'extensions chromatinienne velues en périphérie, l'aspect du cytoplasme (couleur, granulation, rapport nucléocytoplasmique), la taille relative de la cellule et l'absence de phénomène de compression ou d'étirement par rapport aux cellules tumorales adjacentes.

Les mitoses apparaissent plus sombres sur lame numérique, la basophilie est accentuée et certains détails morphologiques comme les extensions velues périphériques sont moins évidents à repérer même si l'irrégularité des contours persiste le plus souvent. La numérisation a tendance à homogénéiser la texture de la chromatine, rendant parfois délicate la distinction avec les noyaux hyperchromatiques. Le cytoplasme des mitoses est le plus souvent granulaire légèrement éosinophile, parfois basophile ou clair mais peut dans 8% des cas présenter une éosinophilie comparable à celui d'une cellule en apoptose.

Plusieurs travaux ont été réalisés avant la publication de ces critères. Dans une étude portant sur 100 cas de carcinomes infiltrants examinés par trois pathologistes sur lame de verre et lame numérique dans la même zone de comptage avec une période de *wash-out* d'au moins quatre mois, une sous et surestimation du compte mitotique sur lame numérique était relevée dans respectivement 52% et 33% des cas. Il en résultait une modification du score mitotique pour 25% des cas (sous-estimation dans 15% et sur estimation dans 10%) (106).

Dans l'étude de Davidson *et al.* (86), les scores mitotiques étaient significativement plus faibles sur lame numérique par rapport aux lames de verre, il en était de même pour l'accord inter-observateur ( $\kappa = 0,42$  sur lame de verre et  $\kappa = 0,25$  sur lame numérique).

Dans l'étude de Rakha *et al.* (81), la valeur pronostique du score mitotique sur lame numérique était moins nette que sur lame de verre, particulièrement entre les scores 2 et 3. Des discordances de score mitotique entre lame de verre et lame numérique d'un ou deux points étaient retrouvées dans respectivement 24% et 6% des cas avec des proportions de scores 1, 2 et 3 de respectivement 50%, 19% et 31% sur lame de verre et 65%, 15% et 20% sur lame numérique. Ainsi, 60% des score 2 sur lame de verre ont été classés en score 1 sur lame numérique. Les score 3 sur lame de verre ont été classés score 2 et score 1 sur lame numérique dans respectivement 26% et 20% des cas.

Cette tendance à la sous-estimation du score mitotique sur lame numérique se retrouve également dans l'étude de Lashen *et al.* où malgré la plus grande reproductibilité des pathologistes observée en tenant compte des nouveaux critères proposés, l'évaluation de 5807 mitoses sur 341 carcinomes infiltrants montrait une sous-estimation du compte mitotique et du score associé dans respectivement 20% et 19% des cas. Ainsi 37% des cas ayant un score mitotique de 2 en microscopie optique ont été classés en score 1 sur lame numérique et 27% des cas ayant un score 3 sur lame de verre ont eu un score 2 sur lame numérique (105).

### 6.3. Surface de comptage et impact des moniteurs

En microscopie optique, la surface de champ varie en fonction de l'objectif et de l'oculaire et n'est pas identique sur tous les microscopes. En pathologie numérique, la taille de l'écran et sa résolution d'affichage ont un impact significatif sur la surface analysée (le *viewer* et le scanner de lames ayant un impact minimal) (107). Une image affichée sur un moniteur au x40 n'est donc pas équivalente à un HPF en microscopie optique. Sur lame numérique, l'équipe de Nottingham recommande d'évaluer les mitoses sur une surface de 2 ou 3 mm<sup>2</sup> (100). Cette surface peut être déterminée soit directement par déplacement successif du champ analysé (sous réserve que le pathologiste sache exactement quelle surface est représentée sur son moniteur), soit en définissant une surface globale de 2 ou 3 mm<sup>2</sup> ou en disposant dix surfaces contiguës de 0,2 mm<sup>2</sup> ou 0,3 mm<sup>2</sup> avec l'outil de mesure disponible dans le *viewer*. Il n'y a pas à l'heure actuelle de consensus sur la meilleure façon de procéder et le pathologiste demeure libre d'utiliser l'une ou l'autre des méthodes en fonction des cas et de ses préférences. Il faut noter que cette flexibilité d'analyse rendue possible par la pathologie numérique présente un intérêt particulier en cas de matériel tumoral de petite taille là où sur un microscope classique il est impossible de moduler la surface de champ.

## 6.4. Bases de données et grands challenges portant sur la détection des mitoses en pathologie mammaire humaine

On dénombre depuis 2010 une vingtaine de challenges visant à développer des outils d'aide au diagnostic en anatomopathologie. Plus de la moitié de ces compétitions concernent la pathologie mammaire et cinq ont porté spécifiquement sur la détection de mitoses dans les carcinomes infiltrants (108). Les équipes de ces challenges travaillent sur des bases de données communes labellisées. Les participants peuvent ainsi confronter leurs méthodes et innovations en matière d'apprentissage automatisé. Au terme des concours, les bases de données de mitoses sont le plus souvent laissées à la disposition des développeurs pour leur permettre de continuer à améliorer leurs algorithmes en essayant différentes approches pour tenter de dépasser les résultats initialement obtenus pendant la compétition (109,110). Néanmoins développer un outil utilisable en routine diagnostique par les pathologistes nécessite de prendre en considérations des éléments pratiques qui vont au-delà des métriques classiquement présentées lors de ces challenges. Ils permettent tout de même aux pathologistes et aux différents acteurs du domaine de connaître les avancées en matière d'intelligence artificielle et de comprendre les limites de certaines méthodes diagnostiques. L'intelligence artificielle prendra davantage de place en routine dans les années à venir. Il apparaît donc important pour les pathologistes de se tenir informés de ces avancées et de comprendre comment les algorithmes sont conçus et évalués (92,108).

Les cinq principaux challenges consacrés à la détection des mitoses (avec labellisation de mitoses) ou à la détermination du score mitotique (sans labellisation de mitoses) dans le carcinome mammaire et leurs caractéristiques sont présentés par ordre chronologique dans le **tableau I**. La résolution de chaque scanner de lames était la suivante : Leica Aperio ScanScope XT et Aperio CS2 : 0,25  $\mu\text{m}/\text{pixel}$  ; Leica Aperio GT450 : 0,26  $\mu\text{m}/\text{pixel}$  ; Hamamatsu S360, NanoZoomer 2.0-HT, XR et RS : 0,23  $\mu\text{m}/\text{pixel}$  ; 3DHISTECH Panoramic 1000 : 0,24  $\mu\text{m}/\text{pixel}$ . La numérisation des lames était faite dans tous les cas au x40.

Tableau I : résumé des challenges portant sur la détection de mitoses

Challenge	Objectif	Scanners de lame	Matériel	Bases de données	Métriques du meilleur algorithme
MITOS 2012	Détection des mitoses sur aires de taille prédéfinie extraites de lames entières	Aperio ScanScope XT Hamamatsu NanoZoomer 2.0-HT	5 lames 50 HPF de 0,26 mm <sup>2</sup> (2084 x 2084 pixels)	326 mitoses sur les 50 HPF Entraînement : 226 mitoses de 35 HPF Test : 100 mitoses de 15 HPF	Aperio : F1-score de 0,782, recall de 0,70 et précision de 0,89 Hamamatsu : F1-score de 0,701, recall de 0,61 et précision de 0,82
AMIDA13	Détection des mitoses sur aires de taille prédéfinie extraites de lames entières	Aperio ScanScope XT	23 lames Images TIFF de 0,25 mm <sup>2</sup> (2000 x 2000 pixels) surface extraite / lame : [7 - 58] mm <sup>2</sup>	1083 mitoses Entraînement : 550 mitoses (12 lames) Test : 533 mitoses (11 lames)	F1-score, recall et précision de 0,61
MITOS&ATYPIA14	Détection des mitoses sur aires de taille prédéfinie extraites de lames entières	Aperio ScanScope XT Hamamatsu NanoZoomer 2.0-HT	16 lames 1696 tuiles de 0,13 mm <sup>2</sup>	1502 mitoses et 5740 non mitoses Entraînement (11 lames) : * Aperio : 749 mitoses et 2849 non-mitoses * Hamamatsu : 753 mitoses et 2891 non-mitoses Test : 203 mitoses (5 lames)	Aperio : F1-score de 0,356 Hamamatsu : F1-score de 0,279
TUPAC16	Prédiction du score mitotique sur lame entière numérisée	Aperio ScanScope XT	base de donnée du TCGA 821 lames	Entraînement : 500 lames (236 score 1, 117 score 2 et 147 score 3) Test : 321 lames (147 score 1, 77 score 2 et 97 score 3)	kappa de Cohen pondéré de 0,567 36% de discordance entre le label prédit et la vérité établie
MIDOG2021	Développer un algorithme dont les résultats soient reproductible d'un scanner à l'autre	A: Hamamatsu XR NanoZoomer 2.0 B: Hamamatsu S360 C: Leica Aperio CS2 D: Leica Aperio GT450 E: 3DHISTECH Panoramic 1000 F: Hamamatsu NanoZoomer 2.0 RS	280 lames Images TIFF de 2mm <sup>2</sup> 1 image TIFF / lame	Entraînement (scanners A, B, C et D, 200 lames, 50 / scanner) : 1721 mitoses et 2714 non-mitoses Test (scanners A, D, E et F, 80 lames, 20 / scanner) : nombre de mitoses et non-mitoses non publié	F1-score de 0,748 précision de 0,754 recall de 0,742

#### 6.4.1. MITOS12

Les performances obtenues lors de ce challenge étaient bonnes mais reposaient sur seulement cinq lames, la détection n'était pas faite sur l'ensemble de la surface tumorale et les mêmes patients étaient présents en entraînement comme en test ce qui constitue un biais (111). Par ailleurs, le temps d'analyse de l'algorithme le plus performant était de 8 minutes pour une image de 0,26 mm<sup>2</sup> ce qui constituait un véritable obstacle à une utilisation en routine (37).

#### 6.4.2. AMIDA13

Ce challenge a été proposé pour tenter de résoudre les problèmes de MITOS12. Les performances obtenues étaient moins élevées que celles du challenge précédent mais portaient sur une base de donnée plus importante et sans chevauchement de cas entre la phase d'entraînement et celle de test. Néanmoins les images avaient été numérisées sur un seul scanner de lames avec des paramètres optimaux et n'étaient pas présentées aux participants sous forme de lame entière mais de fichiers TIFF de plus petite taille ce qui diffère de la routine diagnostique. Les auteurs soulignent par ailleurs un nombre élevé de faux positifs détectés par les algorithmes (112).

#### 6.4.3. MITOS-ATYPIA14

Le challenge MITOS-ATYPIA 2014 s'inscrit dans la continuité de celui initié en 2012. Les pathologistes pouvaient attribuer lors de la labellisation un label "mitose", "probablement mitose" ou "non-mitose" permettant de pondérer l'indice de confiance lors de l'apprentissage des algorithmes. Quel que soit le scanner utilisé les résultats étaient en retrait par rapport aux deux challenges précédents (98).

#### 6.4.4. TUPAC16

Le challenge TUPAC16 est le premier à avoir eu pour objectif d'évaluer le score mitotique et à avoir proposé de travailler sur lame entière. Les cas numérisés sur un seul scanner avaient été distribués de façon aléatoire à une équipe de 15 pathologistes experts pour évaluer le score mitotique sans labellisation de mitoses. L'algorithme le plus performant pour prédire le score mitotique a obtenu un kappa de Cohen pondéré de 0,57, signifiant un accord modéré avec les pathologistes. Les algorithmes les plus performants ont fait un nombre substantiel d'erreurs avec fréquemment une discordance de deux points, ce qui pourrait avoir de lourdes conséquences dans la prise en charge des patientes. Le fait qu'il n'y ait pas de mitoses labellisées est important car les stratégies utilisées par les compétiteurs utilisaient soit une méthode de détection de la zone la plus mitotique suivie d'un comptage de mitoses

dans ces zones, soit une méthode directe basée sur une évaluation de l'apparence globale de l'image sans comptage de mitoses. Les algorithmes attribuaient un score de probabilité de mitose aux objets détectés sans garantie que ce soit effectivement une mitose car il n'y avait pas de label pour s'en assurer, ou bien ils évaluaient le score mitotique sur d'autres paramètres de l'image que les mitoses rendant l'acceptabilité de la méthode discutable. Néanmoins les performances de détection évaluées secondairement sur une base de données de 1552 mitoses issues de 73 lames retrouvaient un F1-score de 0,65 pour le meilleur algorithme (113,114).

#### 6.4.5. MIDOG2021

L'objectif de ce challenge était de développer un algorithme dont les résultats soient reproductibles d'un scanner à l'autre. Il ne portait pas sur lame entière mais uniquement sur une surface de 2 mm<sup>2</sup> extraite au format TIFF, l'équipe gagnante a obtenu des métriques meilleures que celles obtenues lors des challenges précédents (115,116).

### 6.5. Reproductibilité des pathologistes dans l'évaluation des mitoses sur lame numérique à l'aune de l'intelligence artificielle

Les laboratoires d'anatomopathologie travaillant en 100% numérique en 2022 sont encore peu nombreux. Les outils d'intelligence artificielle n'ont donc pas encore la possibilité d'être intégrés à l'activité de routine des pathologistes. Par ailleurs, les résultats des différents challenges visent seulement à évaluer les performances d'algorithmes indépendamment de leur éventuelle utilité pratique. C'est pour essayer de préciser ce point que des études ont été entreprises pour évaluer la reproductibilité inter-observateur et les performances des pathologistes dans l'évaluation des mitoses prenant en compte les résultats d'algorithmes d'intelligence artificielle.

Dans l'étude de Veta *et al.* (117) portant sur 100 cas, le kappa évaluant l'accord inter-observateur de trois pathologistes pour l'attribution du score mitotique sur une surface de 2 mm<sup>2</sup> prédéfinie variait entre 0,79 et 0,89. Le kappa portant sur l'accord entre l'algorithme et les pathologistes variait entre 0,66 et 0,74. L'impact clinique des discordances est particulièrement important dans les situations où le score mitotique varie de deux points (1 au lieu de 3 et *vice versa*). Seuls 2 cas sur 100 ont fait l'objet d'un désaccord de ce type entre les trois pathologistes contre 13 entre les pathologistes et l'algorithme.

L'accord des pathologistes sur chaque mitose évalué par le coefficient de Dice variait entre 0,57 et 0,67. Entre les pathologistes et l'algorithme, il variait entre 0,50 et 0,56. Par ailleurs l'accord le plus élevé

mesuré entre les pathologistes et l'algorithme arrivait au mieux au niveau de l'accord le moins élevé mesuré entre les pathologistes. Les performances de l'algorithme, inférieures à celles des trois pathologistes étaient en partie dues à la détection de faux positifs.

Dans l'étude de Balkenhol *et al.* (118), portant sur 90 cas et ayant comparé la reproductibilité de deux pathologistes dans l'évaluation du score mitotique sur lame de verre et sur lame numérique assistée d'un algorithme, la prédéfinition d'une zone de plus forte densité mitotique (*hotspot*) de 2 mm<sup>2</sup> par l'algorithme permettait d'améliorer significativement la reproductibilité des pathologistes avec un kappa passant de 0,69 sur lame de verre à 0,81 sur lame numérique. Par ailleurs, l'accord mesuré entre l'algorithme et les pathologistes était comparable à celui observé entre les deux pathologistes. Cependant dans les tumeurs comportant peu de mitoses, l'algorithme avait tendance à générer beaucoup de faux positifs (lymphocytes, fibroblastes, encre de repérage) ce qui entraînait des erreurs dans la délimitation des zones de *hotspot*. De plus, seule la zone de *hotspot* proposée était présentée et les mitoses détectées par l'algorithme n'étaient pas visibles par le pathologiste.

Sur une cohorte annexe de 298 cancers du sein triple négatifs évalués par 3 pathologistes, l'algorithme surestimait le score mitotique d'un facteur x1,5.

Dans l'étude de Pantanowitz *et al.* (119), vingt-quatre pathologistes de niveau d'expertise variés ont compté les mitoses à deux reprises avec ou sans assistance d'un algorithme sur 140 champs de 0,2 mm<sup>2</sup> avec une période de *wash-out* de quatre semaines. La précision et la sensibilité des pathologistes étaient améliorées par l'utilisation de l'algorithme. Le F1-score pour la détection des mitoses était de 0,61 sans l'aide de l'algorithme, et de 0,71 avec l'aide de l'algorithme. Par ailleurs, 21 lecteurs sur 24 (87,5%) ont identifié plus de mitoses avec l'aide de l'algorithme et 13 (54,2%) ont diminué le nombre de faux positifs. Cette étude portait sur la détection des mitoses mais sans évaluation du score mitotique et la reproductibilité des pathologistes n'a pas été évaluée. L'impact potentiel de l'utilisation de l'algorithme sur le score mitotique final n'a donc pas été exploré. Étant donné que l'intelligence artificielle peut générer de nombreux faux positifs, il existe un risque pour le pathologiste d'augmenter artificiellement le grade tumoral. Néanmoins il a été montré que les pathologistes avaient tendance à sous-estimer le score mitotique sur lame numérique. Des études d'impact sur le score mitotique seraient donc pertinentes pour voir si l'utilisation de l'algorithme permettrait de compenser cette sous-estimation. La principale limite des solutions d'intelligence artificielle pour l'évaluation des mitoses est le manque de données d'apprentissage disponibles représentatives de la routine diagnostique en partie lié au

caractère chronophage de la labellisation et aux variations des conditions pré-analytiques. Cela conduit à des algorithmes qui peuvent donner de bons résultats dans un contexte d'étude spécifique mais qui peinent à généraliser les résultats à d'autres images issues de différents laboratoires ou ayant des caractéristiques différentes. L'application dans un laboratoire d'un algorithme développé à partir d'un ensemble de données issues d'un autre laboratoire nécessite de vérifier ses performances pour s'assurer qu'il soit capable de généraliser ses résultats.

A l'heure actuelle, ces algorithmes ont tendance à détecter trop de faux positifs ce qui peut biaiser l'analyse finale du pathologiste. Certains modèles d'algorithmes basés sur une étape de détection (sensible mais peu spécifique) suivie d'une étape de classification (spécifique mais peu sensible) tentent de contourner ce problème (120).

## 7. Rationnel et objectif de l'étude

Les données principales de cette revue sont donc les suivantes. La reconnaissance des mitoses sur lame numérique même après une période d'apprentissage présente des difficultés supplémentaires pour le pathologiste par rapport à la lame de verre pouvant impacter l'évaluation du score mitotique. Parmi les nombreuses applications potentielles de l'intelligence artificielle en pathologie mammaire, des algorithmes sont en cours de développement pour assister le pathologiste dans l'évaluation de ce score. Les études publiées sur le sujet concernent essentiellement les performances des algorithmes en termes de métriques sur des lames numérisées dans des conditions parfois éloignées de la routine diagnostique d'un laboratoire d'anatomopathologie. De plus, il a été montré que les algorithmes avaient tendance à détecter trop de faux positifs ce qui peut être un frein à leur utilisation. L'optimisation de l'architecture de l'algorithme et de l'interface de lecture ou d'analyse pourrait permettre de résoudre ce problème.

Ce travail de thèse s'inscrit dans une démarche collaborative ayant pour objectif le développement d'un algorithme d'intelligence artificielle ainsi que son interface utilisateur permettant d'assister le pathologiste dans l'évaluation du score mitotique sur lame numérique en pratique quotidienne. Le but final est l'intégration de l'outil dans le flux de travail du laboratoire et du pathologiste. Ce travail collaboratif a été effectué en partenariat avec la start-up Primaa, travaillant au développement d'outils d'aide au diagnostic en anatomopathologie.

Il se décompose en trois phases :

- Phase 1 : constitution par deux médecins pathologistes d'une base de données de mitoses labellisées sur lames numériques de routine ; entraînement et test de l'algorithme sur cette base de données par la start-up ; mise au point de l'interface permettant au pathologiste d'interagir avec les données présentées par l'algorithme.
- Phase 2 : évaluation de l'apport de l'algorithme pour la détection des mitoses par le pathologiste. Pour cette étape, une comparaison de la détection des mitoses par quatre pathologistes avec et sans utilisation de l'algorithme a été effectuée sur une zone prédéfinie de 2 mm<sup>2</sup>.
- Phase 3 : évaluation de la reproductibilité inter-observateur dans l'appréciation du score mitotique sur lame entière avec et sans assistance de l'algorithme.

Nous présentons dans ce manuscrit les données de la phase 1 et 2, la phase 3 étant en fin de réalisation pratique.

# Matériel et Méthodes

Remarque préliminaire : pour une lisibilité plus aisée, les métriques de l'algorithme développé ont été intégrées dans ce chapitre Matériel et Méthodes à la *section 1.2. Entraînement et test de l'algorithme*.

## 1. Phase 1

### 1.1. Constitution de la base de données

Le premier temps a été la constitution d'une base de données de mitoses et de "non-mitoses" nécessaires à l'entraînement, à la validation et au test de l'algorithme.

Trente-deux lames issues des dossiers de 11 patientes ayant eu un diagnostic de carcinome mammaire infiltrant entre Octobre 2019 et Mai 2020 dans le service d'anatomopathologie de l'hôpital Bicêtre ont été labellisées entre Novembre 2020 et Septembre 2021.

Les caractéristiques cliniques et anatomopathologiques de ces dossiers sont présentées en annexe **(annexe 1)**.

#### 1.1.1. Préparation du matériel

##### 1.1.1.1. Prénnumérisation

Les biopsies ont été fixées en formol neutre tamponné à 4% de formaldéhyde pendant au minimum 6h. Les pièces opératoires ont été échantillonnées après fixation en formol tamponné à 4% pendant 24 à 48h. Puis les prélèvements en cassettes ont été déshydratés et imprégnés sur automate Sakura Tissue-Tek® VIP® selon le protocole en vigueur. L'inclusion en paraffine a été effectuée sur automate Sakura Tissue-Tek® AutoTEC® pour les échantillons de pièces opératoires et manuellement sur platine Sakura Tissue-Tek® pour les biopsies. Les coupes réalisées au microtome Leica ou Microm de 3 µm d'épaisseur ont été étalées sur lame de verre SuperFrost™ puis mises à sécher 30 min à l'étuve à 60 °C et 15 min dans le colorateur. La coloration en Hématoxyline-Eosine-Safran a été effectuée sur automate Leica ST5020®. Les lames ont ensuite été montées sur l'automate Leica CV5030® avec un milieu de montage Pertex® puis séchées au minimum 5 min dans l'automate.

##### 1.1.1.2. Numérisation

La numérisation a été effectuée sur deux scanners de lames P1000 (pièces opératoires) et P250 (biopsies) de la société 3DHISTECH, tous deux équipés de deux objectifs Plan-Apochromat x20 et x40

et d'une caméra Adimec QUARTZ Q-12A180 d'une définition de 4096 × 3072 pixels - (taille du pixel : 5.5 µm × 5.5 µm) permettant une multiplication numérique x1,6 du grossissement.

Les 32 lames ont été numérisées au format .mrxs avec une définition de 0,24 µm/pixel selon un protocole utilisé en routine diagnostique dans le service à l'objectif x20 avec multiplicateur numérique x1,6. Ce protocole a été choisi car il offre le meilleur compromis entre le poids des fichiers, la vitesse de numérisation et la qualité de la lame numérique. En effet, la numérisation au x40 multiplie en moyenne le temps de numérisation par 1,5 et le poids des fichiers par 3,7 par rapport au protocole standard, tandis que l'utilisation du multiplicateur numérique permet d'approcher avec un objectif x20 le rendu visuel d'une lame numérisée avec un objectif x40. Ce protocole est donc parfaitement adapté à un flux de travail entièrement numérisé.

Les protocoles de numérisation utilisés en routine dans le service de Bicêtre sont disponibles en annexe **(annexe 2)**.

## 1.1.2. Labellisation

### 1.1.2.1. Labellisation initiale

Les 32 lames numérisées ont été pseudonymisées à l'aide d'un code alphanumérique conservé par le seul coordinateur de l'étude avant d'être transférées sur le *viewer* Cytomine, logiciel open-source développé par l'Université de Liège permettant de réaliser des travaux collaboratifs d'analyse d'image numérisée de très haute résolution. Nous avons retenu les critères d'identification des mitoses tels qu'ils avaient été initialement proposés par van Diest en 1992 (103), les critères d'identification des mitoses sur lame numérique dans les cancers du sein décrits par Lashen *et al.* n'ayant été publiés qu'après la fin de cette étape (105).

La première étape de labellisation a été réalisée par un interne en anatomopathologie sur ordinateur Asus C423N équipé d'un écran de 14" avec une résolution native de 1366 x 768 pixels.

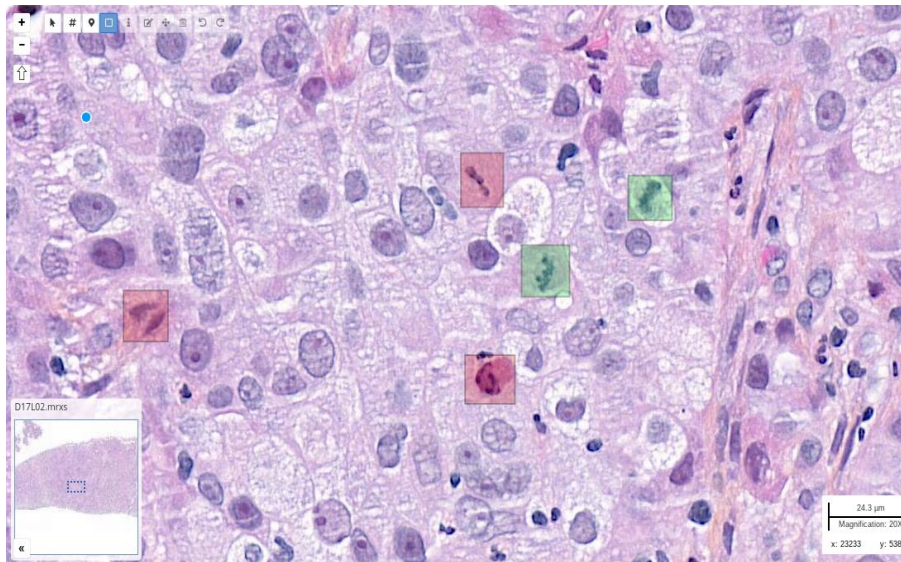
L'objectif était de labelliser le maximum de mitoses identifiables sur les lames.

Deux catégories d'objets ont été retenues pour cette étape de labellisation:

- Les mitoses (en vert)
- Les "non-mitoses" (en rouge) correspondant soit à des objets morphologiquement ambigus et pouvant mimer des mitoses, soit des objets clairement identifiables comme n'étant pas des mitoses.

Les mitoses et "non-mitoses" ont été labellisées à l'aide de l'outil "rectangle" de Cytomine (**figure 1**) permettant une délimitation précise des objets et la constitution de vignettes facilement exploitables par la suite pour l'entraînement, la validation et le test de l'algorithme.

Au total, 2858 mitoses et 3901 "non-mitoses" ont été ainsi labellisées sur les 32 lames de carcinome mammaire infiltrant.

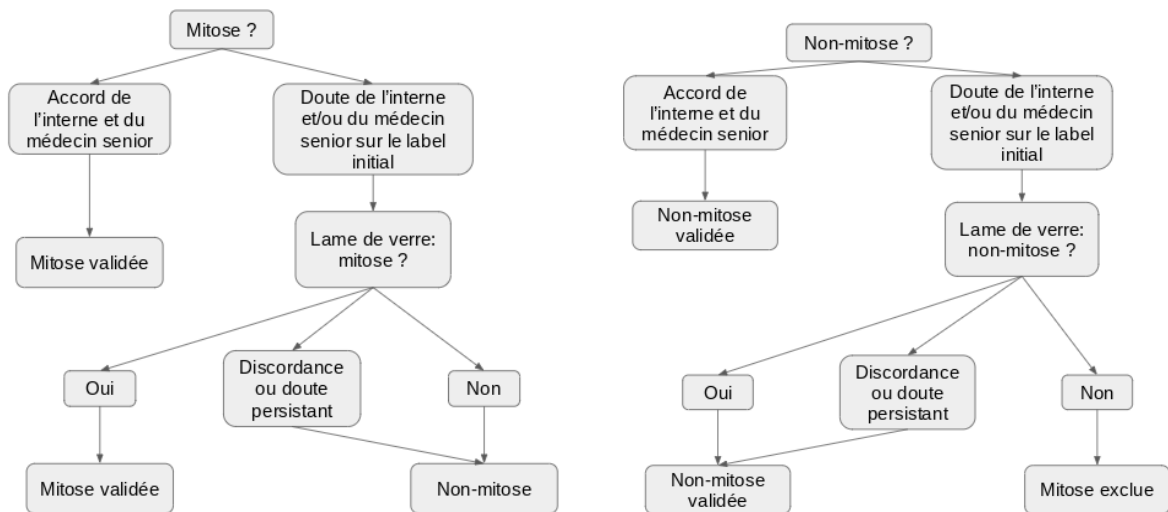


**Figure 1 : labellisation des mitoses et des "non-mitoses" sur Cytomine ; x20**

#### 1.1.2.2. Relecture consensuelle

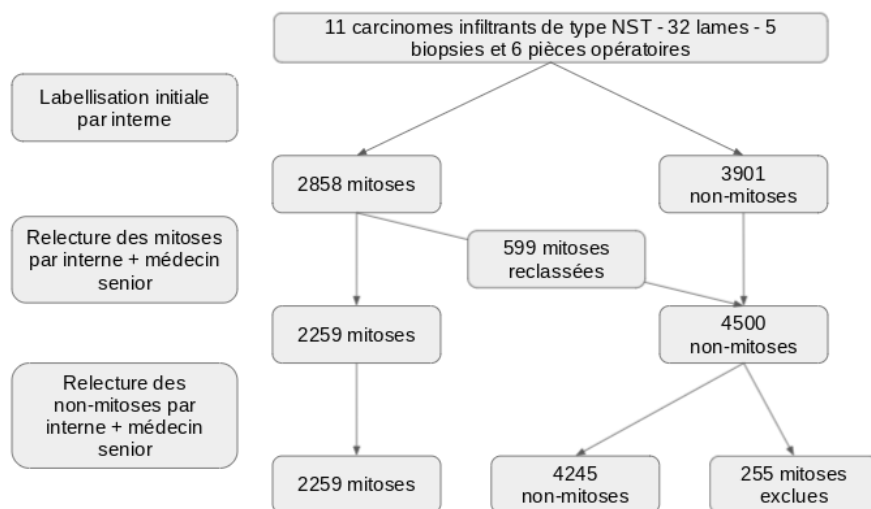
Une relecture conjointe en deux étapes a ensuite été faite : relecture consensus des mitoses puis des "non-mitoses" sur écran 27" haute résolution EIZO FlexScan EV2750 par l'interne et un médecin pathologiste sénior expérimenté en pathologie mammaire et dont l'activité de routine est entièrement réalisée en pathologie numérique. En cas de discordance entre le label attribué par l'interne et l'avis du pathologiste sénior, ou en cas de doute de l'interne a posteriori sur le label initialement attribué, une relecture de l'objet sur la lame de verre correspondante était effectuée.

Concernant les mitoses, en cas de doute persistant sur lame de verre, l'image était basculée dans les "non-mitoses", étant admis en routine diagnostique que les images douteuses ne doivent pas être comptabilisées lors de l'évaluation du score mitotique. En cas d'erreur de label avérée sur lame de verre ou en cas de doute/désaccord persistant sur la lame de verre, l'objet était basculé dans les "non-mitoses" (**figure 2**). Au total, 599 images initialement classées en mitoses ont été reclassées en "non-mitoses" à l'issue de cette phase.



**Figure 2 : étapes de relecture des mitoses et des "non-mitoses"**

Concernant les "non-mitoses", en cas de doute persistant sur lame de verre, l'image était conservée dans les "non-mitoses" étant admis en routine diagnostique que les images douteuses ne doivent pas être comptabilisées lors de l'évaluation du score mitotique. En cas d'erreur de label avérée sur lame de verre, la mitose mal classée était supprimée de la base de données ceci afin d'éviter un éventuel retour en arrière des mitoses basculées en "non-mitoses" à la première relecture (**figure 2**). Au total, 255 mitoses ont été exclues à l'issue de cette phase.



**Figure 3 : changements de labels au cours de la relecture**

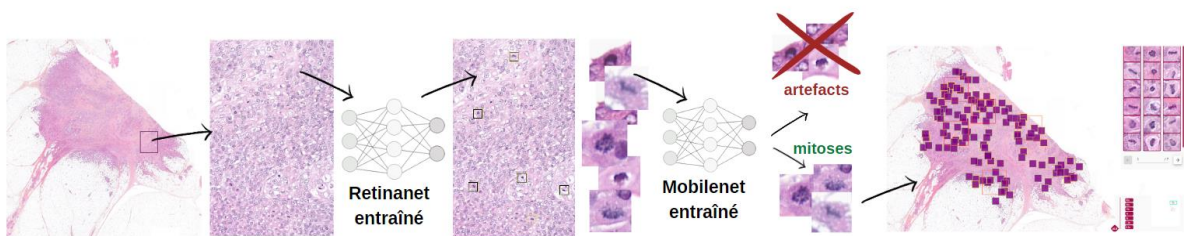
Au terme de la relecture, 2259 mitoses et 4245 "non-mitoses" ont constitué la base de données (**figure 3**). La répartition du nombre de mitoses et de "non-mitoses" pour chaque cas est détaillée en annexe (**annexe 3**).

## 1.2. Entraînement et test de l'algorithme

La mise au point de l'outil, son entraînement et son test ont été réalisés par l'équipe de *data scientists* de la start-up Prima. Au cours de ce développement, les pathologistes sont intervenus régulièrement pour valider ou non sur le plan médical la pertinence des images proposées par l'outil en cours d'élaboration. Dans un souci de compréhension pour le lecteur non spécialisé dans les questions d'intelligence artificielle, nous ne présentons ici que les étapes simplifiées de ce travail. Des éléments plus détaillés sont fournis en annexes.

### 1.2.1. Description générale du modèle

Le fonctionnement de l'algorithme se décompose en deux étapes : la détection d'objets candidats puis la classification de ces objets en mitoses et artefacts correspondant aux "non-mitoses" (**figure 4**). Pour la première étape, l'utilisation d'un réseau de neurones détecteurs de type RetinaNet, entraîné sur des images annotées, permet d'estimer la probabilité que les pixels d'une nouvelle image appartiennent à une mitose. Pour la seconde étape, l'utilisation d'un classifieur de type MobileNet permet d'affiner les objets présentés visuellement au pathologiste. Nous présentons successivement les modalités d'entraînement – validation puis de test du détecteur et du classifieur.



**Figure 4 : schéma général de l'algorithme**

### 1.2.2. Le détecteur

#### 1.2.2.1. Bases de données utilisées

Deux bases de données de mitoses ont été utilisées (**tableau II**). La première (KB) est la base de données obtenue par labellisation des lames de Bicêtre (*cf. ci-dessus*). Après une phase de test préliminaire, deux lames labellisées ont été retirées par les développeurs du fait d'artefacts (pré-

analytiques et /ou de numérisation) qui perturbaient l’entraînement initial du modèle. Trente lames ont finalement été retenues. Treize d’entre elles correspondant aux lames des cas les plus mitotiques ont été utilisées pour l’étape d’entraînement – validation avec une répartition des mitoses à 85% pour l’entraînement et 15% pour la validation. Les dix-sept autres lames ont été utilisées pour l’étape de test (**annexe 3**). Au total, 1677 mitoses ont été réparties dans les trois jeux de données successifs (1322 pour l’entraînement, 234 pour la validation et 121 pour le test). En plus des mitoses que nous avons labellisé, nous avons retenu la base de données en accès libre la plus récente (challenge MIDOG2021) comportant 1721 mitoses dont 1483 ont été utilisées et réparties en 1260 mitoses d’entraînement et 223 mitoses de validation. Dans le but de ne pas ajouter trop de complexité dans l’apprentissage de l’algorithme, le choix a été fait de commencer par travailler sur la base de données KB. Une fois le détecteur stable, les deux bases de données ont été combinées afin d’apporter plus d’informations au modèle.

**Tableau II : répartition des mitoses pour l’entraînement, la validation et le test du détecteur**

	KB	MIDOG
Mitoses (patches d’entraînement)	1322	1260
Mitoses (patches de validation)	234	223
Mitoses (patches de test)	121	0

#### 1.2.2.2. Entraînement du détecteur

Le détecteur RetinaNet (121), composé d’un extracteur de caractéristiques d’image, a été retenu par l’équipe de *data scientists* sur la base de la littérature et de ses performances reconnues lors de différents challenges (98,116). Contrairement aux modèles basés sur des techniques de segmentation, il était approprié à la base de données disponible comportant les coordonnées des mitoses présentes sur la lame sans information sur leur contour.

Les caractéristiques extraites ont été apprises par le modèle afin qu’il soit capable de détecter des mitoses. Les images d’entraînement ont été analysées à de multiples reprises par le réseau suivant des cycles répétés appelés *epoch*. A la sortie du réseau nous avons obtenu une liste de boîtes englobantes (*bounding-boxes*), caractérisées par leurs coordonnées et leur probabilité de contenir une mitose. Un exemple d’image de *bounding-boxe* est présenté visuellement en annexe (**annexe 4**).

Deux entraînements ont été réalisés : le premier sur les données KB, le second sur les données KB et MIDOG. Lors de cette étape, les résultats sur la base de données de test KB étaient meilleurs lorsque le modèle était entraîné avec des images de Bicêtre et de MIDOG (**annexe 5**). Une stratégie d'augmentation de données a également été introduite. Le principe étant d'appliquer des modifications géométriques ou colorimétriques sur l'image (**annexe 6**). Ceci est particulièrement utile lorsque les bases de données exploitables sont peu nombreuses et que l'algorithme est destiné à analyser des images comportant d'importantes variations colorimétriques comme c'est le cas en anatomopathologie (122). L'augmentation de données permet de limiter la différence de perte observée entre le jeu d'entraînement et celui de validation (**annexe 7**), nous avons pu ainsi améliorer les performances et les capacités de généralisation de l'algorithme.

### 1.2.3. Ajout du classifieur

#### 1.2.3.1. Principe du classifieur et choix des paramètres d'inférence du détecteur

La détection des faux positifs est un des problèmes les plus difficiles à résoudre lors de la détection des mitoses. Pour contourner cette difficulté, une étape de classification a été introduite. Elle consiste à attribuer une classe aux objets ayant été localisés par le détecteur pour affiner les objets proposés au pathologiste. Le classifieur attribue un score de probabilité que chaque image détectée soit une mitose. Pour ce faire, un seuil est appliqué permettant d'obtenir un score binaire et donc la classe de l'objet. Dans notre cas, deux classes pouvaient être attribuées aux images analysées : mitose et artefact (ou "non-mitose").

Le réseau de neurones retenu par les *data scientists* de Primaia pour cette étape est le MobileNet. Il s'agit d'un algorithme de classification développé par Google requérant moins de calculs qu'un CNN classique tout en étant aussi performant (123). Il comporte 2 259 265 paramètres optimisables.

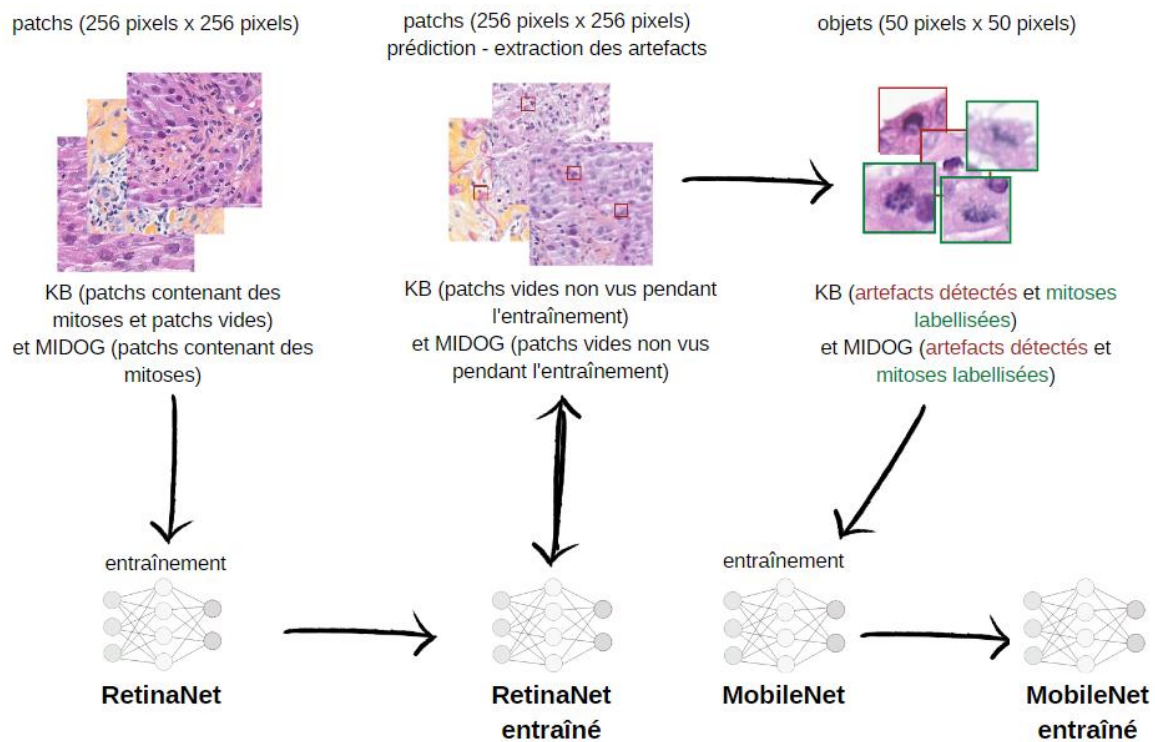
Ce classifieur a été pré-entraîné sur ImageNet (124) puis entraîné avec des artefacts détectés en première étape par le détecteur. Une fois le détecteur entraîné, l'inférence correspond à son exécution. La définition des paramètres d'inférence conditionne les données (mitoses et artefacts) qui sont présentées au classifieur. Ceux-ci sont définis à la fois pour entraîner le classifieur et pour la prédiction finale lors de son utilisation pratique.

Deux paramètres importants sont à faire varier :

- le nombre maximal de détection par image (NB\_DET)
- la probabilité minimale pour considérer une détection comme valide (MIN\_PROB)

Le but est donc de choisir un couple (NB\_DET/MIN\_PROB) qui obtienne la meilleure sensibilité pour un nombre modéré d'artefacts. Si la sensibilité n'est pas assez élevée, trop peu de mitoses seront envoyées à l'étape de classification. Au contraire, si trop d'artefacts sont détectés, l'entraînement du classifieur sera rendu compliqué par un fort déséquilibre entre les deux classes de données. Une fois ce couple défini par les *data scientists*, nous avons eu à notre disposition une quantité suffisante d'artefacts à proposer au classifieur. Les artefacts correspondaient à tous les objets détectés par le détecteur RetinaNet qui n'étaient pas des mitoses selon la labellisation des pathologistes (**figure 5**).

La répartition des mitoses et artefacts utilisés pour entraîner et tester le classifieur est présentée dans le **tableau III**.



**Figure 5 : cycle de la donnée lors de l'entraînement.** Les grands patches sont utilisés pour l'apprentissage du RetinaNet. Une fois entraîné, une passe de prédiction est faite sur de la donnée similaire mais non-vue par le modèle, afin d'extraire des artefacts. Les artefacts extraits, couplés aux mitoses labellisées, sont utilisés pour l'apprentissage du MobileNet.

**Tableau III : répartition des mitoses et artefacts utilisés pour entraîner et tester le classifieur**

		KB	MIDOG
Entraînement	Mitoses	1322	1260
	Artefacts	4814	3824
Validation	Mitoses	234	223
	Artefacts	850	675
Test	Mitoses	95	0
	Artefacts	688	0

### 1.2.3.2. Entraînement, test et métriques

L'entraînement du classifieur a permis de définir les paramètres optimaux de son fonctionnement au-delà desquels survenait un sur-apprentissage avec baisse préjudiciable de la précision malgré une bonne sensibilité (*recall*). Cet état a été obtenu au terme de 200 cycles d'apprentissage (**annexe 8**), chaque cycle induisant une modification des paramètres du réseau de neurones ; cela inclut le traitement de l'ensemble des données d'apprentissage, le calcul de l'erreur, la rétropropagation dans le réseau et la modification des facteurs de pondérations des neurones.

Les analyses suivantes ont porté sur les métriques de performance obtenues directement à partir du nombre de vrais positifs, faux positifs, faux négatifs déterminés par l'attribution des labels des images par les pathologistes. Du fait de l'utilisation de carrés ou de rectangles pour labelliser les mitoses, l'*Intersection over Union* (IoU) a été utilisée comme méthode de définition des vrais positifs, faux positifs et faux négatifs. Elle est définie comme l'aire commune entre la surface englobant la mitose prédite par l'algorithme et celle définie par le pathologiste divisée par l'aire globale que délimitent ces deux surfaces. Deux annotations étaient considérées comme faisant référence au même objet si l'IoU était supérieure à 0,4.

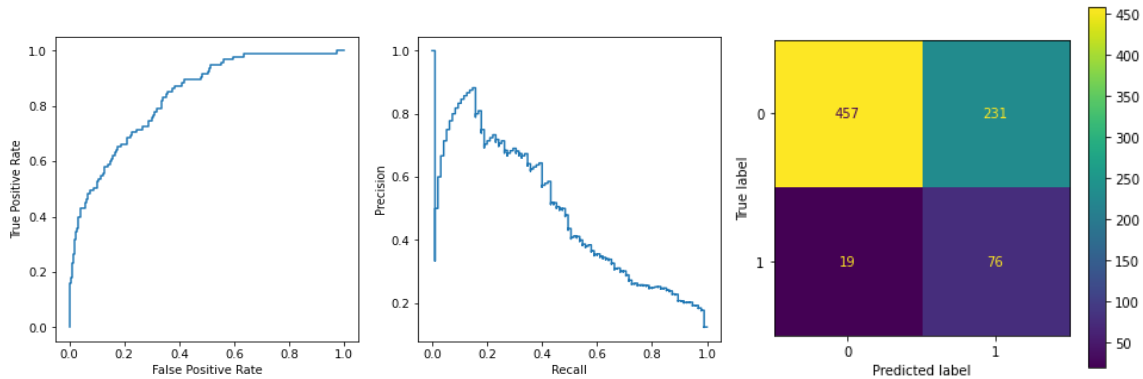
Les courbes ROC (*Receiver Operating Characteristic*) et PR (*Precision - Recall*) du classifieur sur les données de test avec ses poids définis au 200<sup>ème</sup> cycle d'apprentissage sont présentées en **figure 6**. Ces courbes en définissent les performances globales.

A gauche, la courbe ROC montre les taux de faux positifs et de vrais positifs. L'aire en dessous de cette courbe, utilisée comme un indicateur d'apprentissage, est de 0,84.

Au centre, la courbe PR montre la sensibilité (*recall*) et la précision. L'aire sous cette courbe est de 0,48. Chaque prédiction du classifieur a une valeur comprise entre 0 et 1. La valeur 0,1 a été spécifiée comme

étant le seuil à partir duquel un objet était défini comme une mitose afin d'obtenir une sensibilité (*recall*) de 80%.

La matrice de confusion présentée à droite de la **figure 6** a ainsi été obtenue présentant en haut à gauche, le nombre de vrais-négatifs ; en haut à droite, le nombre de faux-positifs ; en bas à gauche, le nombre de faux-négatifs et en bas à droite, le nombre de vrai-positifs.



**Figure 6 : courbe ROC (à gauche), courbe PR (au centre) et matrice de confusion (à droite) pour le MobileNet**

Pour cette sensibilité (*recall*) de 80%, la précision était de 24,8%, le F1-score de 37,8%. L'exactitude (*accuracy*) était de 68%, l'exactitude équilibrée correspondant à la moyenne des exactitudes de chaque classe était de 73%. Cette dernière valeur permet de mieux prendre en compte le déséquilibre des classes lorsqu'une classe est largement majoritaire par rapport à l'autre.

Le **tableau IV** résume les métriques obtenues en utilisant le détecteur seul et en combinant le détecteur avec le classifieur MobileNet. Elles ont été mesurées sur les patches de 256 x 256 pixels de l'ensemble de test. Le classifieur ayant une sensibilité de 80% mesurée sur les patches de 50 x 50 pixels en sortie du détecteur, l'association des deux algorithmes induit nécessairement une baisse de la sensibilité. Toutefois la performance globale du modèle mesurée par le F1-score s'en trouve améliorée.

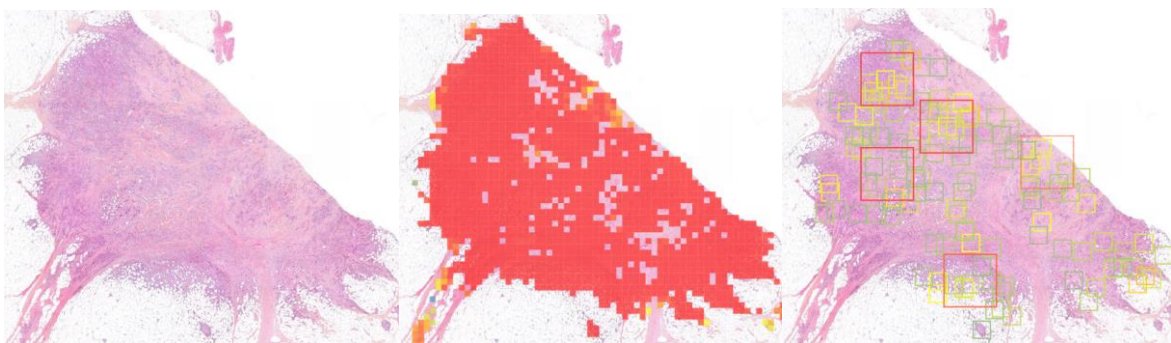
**Tableau IV : métriques obtenues en utilisant le détecteur seul et le détecteur combiné au classifieur**

	Sensibilité	Précision	F1-score
RetinaNet entraîné sur KB + MIDOG + augmentation de données	76%	12%	20.7%
Idem + classifieur MobileNet	61%	24.8%	35,2%

### 1.3. Présentation de l'interface de visualisation

Pour que le pathologiste puisse interagir avec les résultats de l'algorithme de détection de mitoses, une interface de visualisation nommée Cléo a été développée par la start-up sur la base de propositions et de tests effectués en amont par les pathologistes. Cette interface pourrait à l'avenir s'intégrer au sein des systèmes de gestion d'images employés en routine dans les services de pathologie.

L'algorithme de détection de mitoses est appliqué sur les zones ayant été préalablement considérées comme étant du carcinome infiltrant. Ces zones peuvent être visualisées par le pathologiste à l'aide d'une carte de fréquentation (*heatmap*) où le carcinome infiltrant apparaît en rouge et où les couleurs tendent vers le bleu au fur et à mesure que la probabilité de malignité du tissu diminue. Les mitoses détectées apparaissent sous la forme de carrés de petite taille dont la couleur varie avec l'indice de confiance avec lequel l'algorithme a considéré la cellule comme étant une mitose. Le pathologiste peut ainsi visualiser dès le faible grossissement les zones dans lesquelles la densité mitotique est à priori la plus forte. Ces zones de *hotspot* mitotiques sont mises en valeur par l'algorithme sous la forme de carrés de grande taille (cinq carrés de 1 mm<sup>2</sup> dans l'exemple présenté en **figure 7**).



**Figure 7 : échantillon de carcinome infiltrant (à gauche) ; détection du carcinome infiltrant (au centre) ; détection des mitoses et zones de *hotspot* mitotiques (à droite)**

Pour permettre au pathologiste d'apprécier la pertinence des résultats de l'algorithme dès le faible grossissement, un onglet latéral permet de visualiser sous la forme de vignettes les mitoses détectées. Cet outil met en surbrillance sur la lame une mitose sélectionnée dans l'onglet permettant ainsi de la localiser instantanément (**figure 8**).

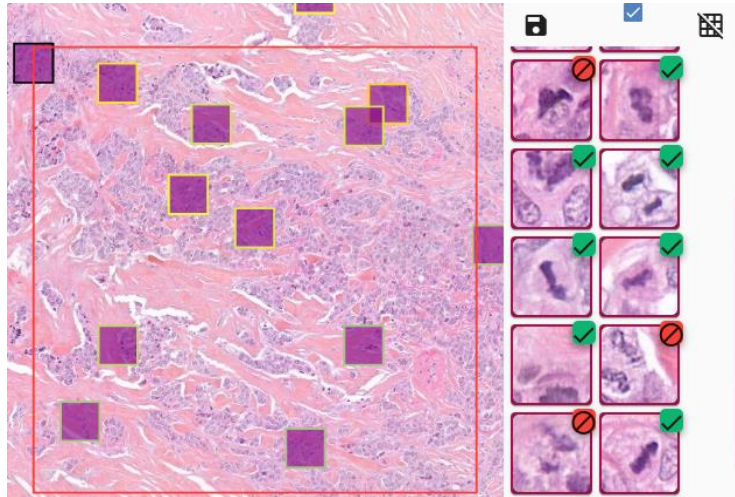


**Figure 8 : détection des mitoses (carrés de petite taille) ; zones de plus forte densité mitotiques proposées par l'algorithme (carrés de grande taille) ; visualisation des mitoses détectées (onglet latéral)**

Plusieurs problèmes se sont posés avec cette fonctionnalité :

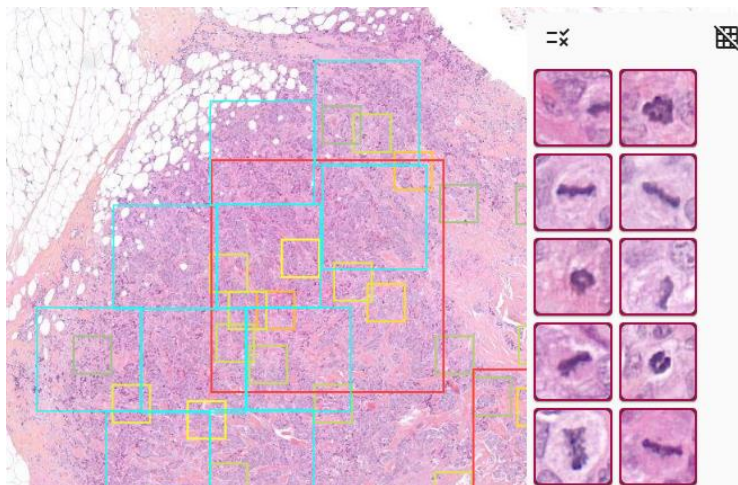
- Le premier est que l'algorithme affiche dans l'onglet les mitoses par ordre de probabilité quel que soit leur position sur la lame. Deux mitoses voisines sur l'onglet peuvent en réalité être éloignées l'une de l'autre sur la lame. Ainsi elles ne peuvent pas être prises en compte ensemble lors de l'évaluation du score mitotique devant être effectué sur dix champs contigus.
- Le second est que le nombre de mitoses détectées peut être élevé et plusieurs pages de vignettes sont à examiner par le pathologiste.
- Le troisième est le manque de spécificité de l'algorithme qui détecte inévitablement des faux positifs venant se mêler aux mitoses. Ainsi certaines mitoses authentiques sont parfois détectées avec un indice de confiance plus faible que certains faux positifs alors mis en avant par l'algorithme. Ces faux positifs sont également un problème car lors de l'analyse ils influent sur les zones de *hotspot* mitotiques proposées par l'algorithme.

Pour contourner la première difficulté, une fonctionnalité a été ajoutée permettant de filtrer les mitoses détectées par zone. Le pathologiste a ainsi la possibilité d'afficher exclusivement les mitoses détectées dans une zone de *hotspot* proposée par l'algorithme. Il est également possible de supprimer les faux positifs dans l'onglet permettant ainsi de réajuster les zones de *hotspot* (**figure 9**). Cela permet d'avoir une idée de la densité mitotique, de la pertinence des mitoses détectées zone par zone et d'estimer ainsi de façon adéquate quelle sera la zone de comptage la plus appropriée.



**Figure 9 : filtre des mitoses et suppression des faux positifs**

Le pathologiste peut alors déposer ses dix carrés correspondant aux dix champs contigus pour réaliser son comptage. Il est également possible d'enlever les résultats de l'algorithme pour faciliter la visualisation car le nombre de carrés affiché à l'écran peut devenir important (**figure 10**).



**Figure 10 : dix champs contigus déposés par le pathologiste (en bleu), zone de plus forte densité proposée par l'algorithme (carré rouge de grande taille) ; mitoses détectées (carrés de petite taille)**

## 2. Phase 2 : apport de l'algorithme pour le pathologiste dans la détection des mitoses

### 2.1. Matériel d'étude

Cette phase portait sur une série rétrospective de 50 carcinomes infiltrants du sein (28 biopsies et 22 pièces opératoires issues des hôpitaux Antoine Béclère et Bicêtre) pris en charge entre 2020 et 2021 dans le service d'anatomopathologie de Bicêtre. Pour chaque cas, une lame représentative de la tumeur a été sélectionnée.

Pour les pièces opératoires (tumorectomie ou mastectomie), les patientes n'avaient pas reçu de thérapie néo-adjuvante.

Pour cette étape, une comparaison de la détection des mitoses par quatre pathologistes avec et sans l'aide de l'algorithme a été effectuée sur une surface de 2 mm<sup>2</sup> préalablement délimitée par un cinquième pathologiste ne participant pas à l'étude.

Les prélèvements ont été pris en charge selon le protocole technique en usage dans le laboratoire décrit à la phase 1. L'ensemble des cas ont été numérisés sur le scanner de lames P1000. Les lames ont ensuite fait l'objet d'une pseudonymisation à l'aide d'un code alphanumérique conservé par le seul coordinateur de l'étude.

### 2.2. Déroulement de l'étude

Les quatre médecins pathologistes volontaires, de niveau d'expertise différents en pathologie mammaire, un PU-PH (pathologiste 1), deux assistants (pathologistes 2 et 3) et un interne (pathologiste 4) tous habitués à travailler sur lame numérique ont participé à l'étude après avoir été formés à l'utilisation de l'interface de visualisation et à l'outil d'aide à la détection de mitoses de la start-up. Les critères d'identification des mitoses étaient ceux utilisés dans l'étape de constitution de la base de données.

Ils ont dans un premier temps annoté les mitoses visibles sur les aires de 2 mm<sup>2</sup> prédéfinies. Seules les mitoses évidentes sur les lames numériques pour chacun des pathologistes ont été annotées.

Après une période de *wash-out* de trois semaines, les quatre pathologistes ont effectué une deuxième étape d'annotation sur les mêmes aires après que celles-ci aient été soumises au préalable à l'outil d'aide à la détection des mitoses. Les pathologistes visualisaient ainsi les mitoses détectées par l'algorithme. Chaque pathologiste a alors annoté les mitoses, qu'elles aient été ou non détectées par

l'algorithme. S'il estimait que l'algorithme s'était trompé, l'image détectée par celui-ci était ignorée du pathologiste.

Pour chaque étape, les mitoses détectées par chaque pathologiste étaient enregistrées sur des calques différenciés permettant d'établir le pourcentage de mitoses labellisées de façon commune par les pathologistes. Le nombre de mitoses de chaque pathologiste et le score correspondant (défini pour une surface de 2 mm<sup>2</sup> par l'OMS (1) : ≤ 7 mitoses : score 1, 8 à 14 mitoses : score 2, ≥ 15 mitoses : score 3) ont été reportés dans un tableau Excel.

### 3. Critères de jugements et analyse statistique

Pour évaluer l'apport de l'algorithme dans la détection des mitoses, les analyses statistiques ont porté sur:

- La reproductibilité intra et inter-observateur des pathologistes sur la surface prédéfinie avec et sans assistance de l'algorithme.
- La reproductibilité inter-observateur des pathologistes avec l'algorithme lui-même.

Pour la reproductibilité du compte mitotique (variable quantitative), le coefficient de corrélation intra-classe (CCI) a été calculé en utilisant un modèle aléatoire à deux facteurs portant sur l'accord absolu entre observateur à partir d'une mesure unique. Les recommandations de Koo et Li de 2016 (125), ont été considérées pour l'interprétation de l'accord : en dessous de 0,50 : faible ; entre 0,50 et 0,74 : moyen ; entre 0,75 et 0,90 : bon ; au-dessus de 0,90 : excellent.

Pour la reproductibilité du score mitotique (variable qualitative ordinale), le kappa de Cohen pondéré de façon linéaire pour chaque couple de pathologiste et le kappa de Fleiss pour l'ensemble des pathologistes ont été calculés. Les recommandations de McHugh de 2012 (126), ont été considérées pour l'interprétation de l'accord : en dessous de 0,20 : nul ; entre 0,21 et 0,39 : minimal ; entre 0,40 et 0,59 : faible ; entre 0,60 et 0,79 : modéré ; entre 0,80 et 0,90 : fort ; supérieur à 0,90 : presque parfait.

L'échelle de Koo et Li ainsi que celle de McHugh, plus strictes dans leur définition ont été préférées à l'échelle de Landis et Koch (127), décrite initialement pour l'interprétation du kappa ou celle de Cicchetti (128) décrite pour l'interprétation du CCI et utilisée notamment dans l'étude de Balkenhol *et al.* (118).

L'*Intersection over Union* (IoU) - définie précédemment en *section 1.2.3.2* - a été utilisée pour évaluer la concordance des pathologistes sur chaque mitose de façon individuelle. Une IoU supérieure ou égale

à 0,2 à été retenue pour considérer deux annotations par deux pathologistes différents comme se rapportant à la même figure mitotique.

Les CCI, kappa de Cohen et kappa de Fleiss ont été calculés à l'aide du package "irr" sur les logiciels R (version 4.0.4, cran, The R Foundation for Statistical Computing, <http://cran.r-project.org>) et R Studio (version 2022.02.2 Build 485 © 2009-2022 RStudio, PBC, <http://www.rstudio.com>).

## 4. Aspects éthiques et réglementaires

S'agissant d'une étude rétrospective portant uniquement sur des données recueillies habituellement pour les soins, ce travail de thèse a été effectué conformément aux dispositions du Code de la Santé Publique applicables aux recherches n'impliquant pas la personne humaine (Code de la Santé Publique - Article R1121-1 modifié par le Décret n° 2017-884 du 9 mai 2017) et ne relève donc pas de la compétence d'un Comité de Protection des Personnes. Il a obtenu l'avis favorable du Comité d'Expertise pour les Recherches, les Études et les Évaluations dans le domaine de la Santé ainsi que du comité d'éthique de la recherche des Hôpitaux Universitaires Henri-Mondor (n° d'avis 2022-134). Il a fait l'objet d'une soumission auprès de la Commission Nationale de l'Informatique et des Libertés sous la référence 2225201 v 0 et a été déclaré en conformité avec la méthodologie de référence MR-004. Les patientes impliquées ont été informées de la recherche via une note d'information distribuée par courrier postal avec possibilité de s'opposer à l'étude dans un délai de trois semaines à compter de la date d'envoi.

# Résultats

Remarque préliminaire : seuls ici sont présentés les résultats de la phase 2 de notre travail concernant la mise en application de l'outil développé au préalable. Ainsi que cela a été indiqué, les résultats portant sur les métriques de l'algorithme développé ont été présentés dans le chapitre Matériel et Méthodes à la section 1.2. *Entraînement et test de l'algorithme*.

## 1. Reproductibilité intra-observateur des pathologistes

Le nombre de mitoses de chaque pathologiste et le score correspondant pour les 50 cas avec et sans l'assistance de l'algorithme sont détaillés en annexe (**annexes 9 et 10**).

Les coefficients de corrélation intra-classe (CCI) de chaque pathologiste portant sur la reproductibilité intra-observateur du compte mitotique entre les deux lectures avec et sans assistance de l'algorithme sont présentés dans le **tableau V**.

**Tableau V : CCI pour l'évaluation du compte mitotique de chaque pathologiste entre leurs deux lectures**

Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
0,967	0,948	0,877	0,966
IC = [0,942 - 0,981]	IC = [0,911 - 0,970]	IC = [0,793 - 0,928]	IC = [0,941 - 0,981]

Les kappa de Cohen pondérés de façon linéaire de chaque pathologiste portant sur la reproductibilité intra-observateur du score mitotique correspondant sont présentés dans le **tableau VI**.

**Tableau VI : kappa pondéré linéaire pour la reproductibilité intra-observateur du score mitotique**

Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
0,872	0,795	0,697	0,869

Les matrices de confusion détaillant la répartition des scores mitotiques de chaque pathologiste entre la première et la seconde lecture sont disponibles en annexe (**annexe 11**). Il existait une modification du score mitotique entre les deux lectures avec et sans assistance de l'algorithme pour 6 ; 6 ; 8 et 4 cas sur 50 respectivement pour les pathologistes 1, 2, 3 et 4.

## 2. Reproductibilité inter-observateur des pathologistes avec et sans utilisation de l'algorithme

Les coefficients de corrélation intra-classe (CCI) portant sur la reproductibilité inter-observateur du compte mitotique pour chaque couple de pathologiste avec et sans assistance de l'algorithme sont présentés dans le **tableau VII**.

**Tableau VII : CCI pour l'évaluation du compte mitotique des pathologistes avec et sans l'algorithme**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
Pathologiste 1		0,484 IC = [0,122 - 0,707]	0,782 IC = [0,151 - 0,921]	0,707 IC = [0,166 - 0,878]
Pathologiste 2	0,642 IC = [0,176 - 0,833]		0,774 IC = [0,634 - 0,865]	0,798 IC = [0,671 - 0,880]
Pathologiste 3	0,666 IC = [0,091 - 0,861]	0,919 IC = [0,862 - 0,953]		0,897 IC = [0,825 - 0,940]
Pathologiste 4	0,718 IC = [0,157 - 0,886]	0,824 IC = [0,709 - 0,896]	0,946 IC = [0,907 - 0,969]	
<i>Au-dessus de la diagonale : lecture sans l'algorithme</i>				
<i>En dessous de la diagonale : lecture avec l'algorithme</i>				

Le CCI pour l'évaluation du compte mitotique par l'ensemble des pathologistes sans assistance de l'algorithme était de 0,716 IC = [0,531 - 0,833].

Le CCI pour l'évaluation du compte mitotique par l'ensemble des pathologistes avec assistance de l'algorithme était de 0,746 IC = [0,543 - 0,859].

Entre les pathologistes 2, 3 et 4 le CCI pour l'évaluation du compte mitotique sans assistance de l'algorithme était de 0,830 IC = [0,745 - 0,893]. Il était de 0,897 IC = [0,842 - 0,936] assisté de l'algorithme.

Les kappa de Cohen pondérés de façon linéaire portant sur la reproductibilité du score mitotique pour chaque couple de pathologistes avec et sans assistance de l'algorithme sont présentés dans le **tableau VIII**.

**Tableau VIII : kappa de Cohen pondérés linéaires portant sur la reproductibilité du score mitotique entre les pathologistes**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
Pathologiste 1		0,339	0,473	0,473
Pathologiste 2	0,496		0,655	0,655
Pathologiste 3	0,419	0,833		0,752
Pathologiste 4	0,447	0,738	0,890	
<i>Au-dessus de la diagonale : lecture sans l'algorithme</i>				
<i>En dessous de la diagonale : lecture avec l'algorithme</i>				

Le kappa de Fleiss portant sur la reproductibilité du score mitotique par l'ensemble des pathologistes était de 0,416 sans assistance de l'algorithme et de 0,512 avec assistance de l'algorithme.

Entre les pathologistes 2, 3 et 4, le kappa de Fleiss était de 0,594 sans utilisation de l'algorithme et de 0,763 avec utilisation de l'algorithme.

Les matrices de confusion détaillant la répartition des scores mitotiques des pathologistes entre eux lors de la lecture sans et avec algorithme sont disponibles en annexe (**annexe 12**).

### 3. Reproductibilité entre les pathologistes et l'algorithme

Les coefficients de corrélation intra-classe (CCI) pour l'évaluation du compte mitotique entre chaque pathologiste et l'algorithme sont présentés dans le **tableau IX**.

**Tableau IX : CCI pour l'évaluation du compte mitotique entre les pathologistes et l'algorithme**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
Algorithme	0,449 IC = [0,101 - 0,678]	0,674 IC = [0,488 - 0,801]	0,683 IC = [0,503 - 0,807]	0,769 IC = [0,627 - 0,862]

Les kappa de Cohen pondérés de façon linéaire pour l'évaluation du score mitotique entre chaque pathologiste et l'algorithme sont présentés dans le **tableau X**.

**Tableau X : kappa pondérés linéaires pour la reproductibilité du score mitotique entre les pathologistes et l'algorithme**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
Algorithme	0,424	0,583	0,624	0,687

Les matrices de confusion détaillant la répartition des scores mitotiques entre la lecture des pathologistes sans l'algorithme et l'algorithme lui-même sont disponibles en annexe (**annexe 13**).

#### 4. Concordance des pathologistes pour l'évaluation de chaque figure mitotique prise individuellement

Le nombre de mitoses annotées par chaque pathologiste ainsi que le nombre de mitoses en communs pour chaque couple de pathologistes sans et avec assistance de l'algorithme sont présentés dans les **tableaux XI et XII**.

**Tableau XI : nombre de mitoses annotées par chaque pathologiste et nombre de mitoses communes pour chaque couple de pathologistes sans assistance de l'algorithme**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4
Pathologiste 1	<b>693</b>	207	287	281
Pathologiste 2		<b>298</b>	150	145
Pathologiste 3			<b>341</b>	207
Pathologiste 4				<b>327</b>

**Tableau XII : nombre de mitoses annotées par chaque pathologiste et nombre de mitoses communes pour chaque couple de pathologistes avec assistance de l'algorithme**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4	Algorithme
Pathologiste 1	<b>675</b>	225	259	275	216
Pathologiste 2		<b>320</b>	156	147	133
Pathologiste 3			<b>297</b>	205	175
Pathologiste 4				<b>322</b>	156
Algorithme					<b>300</b>

Au cours de la première et de la seconde lecture, respectivement 856 et 821 mitoses ont été annotées par au moins un pathologiste. Les proportions de mitoses annotées par un ou plusieurs pathologistes par rapport à la totalité des mitoses détectées au cours de ces deux lectures sont présentées dans les **tableaux XIII et XIV**.

**Tableau XIII : proportion de mitoses annotées par un ou plusieurs pathologistes sans assistance de l'algorithme**

Critère	Nombre de mitoses annotées	Pourcentage par rapport à la totalité des mitoses détectées
Par tous les pathologistes	114	13%
Par au moins trois pathologistes	246	29%
Par au moins deux pathologistes	443	52%
Par le pathologiste 1 + un autre	419	49%
Par au moins un pathologiste	856	100%
Par un pathologiste seulement	413	48%

**Tableau XIV : proportion de mitoses annotées par un ou plusieurs pathologistes avec assistance de l'algorithme**

Critère	Nombre de mitoses annotées	Pourcentage par rapport à la totalité des mitoses détectées
Par tous les pathologistes	112	14%
Par au moins trois pathologistes	250	30%
Par au moins deux pathologistes	431	52%
Par le pathologiste 1 + un autre	402	49%
Par au moins un pathologiste	821	100%
Par un pathologiste seulement	449	55%

Les pourcentages de mitoses prises individuellement pour lesquelles chaque couple d'observateur (pathologiste ou algorithme) est d'accord sont présentés dans le **tableau XV**.

**Tableau XV : pourcentages de mitoses pour lesquelles chaque couple d'observateur (pathologiste ou algorithme) était d'accord**

	Pathologiste 1	Pathologiste 2	Pathologiste 3	Pathologiste 4	Algorithme
Pathologiste 1	<b>675</b>	33,3%	38,4%	40,7%	32,0%
Pathologiste 2	70,3%	<b>320</b>	48,8%	45,9%	41,6%
Pathologiste 3	87,2%	52,5%	<b>297</b>	69,0%	58,9%
Pathologiste 4	85,4%	45,7%	63,7%	<b>322</b>	48,4%
Algorithme	72,0%	44,3%	58,3%	52,0%	<b>300</b>

*Les chiffres en diagonale reprennent le nombre de mitoses détectées par chaque pathologiste avec assistance de l'algorithme et par l'algorithme lui-même. Les pourcentages de mitoses communes pour chaque couple d'observateur sont présentés et définis relativement à l'ensemble des mitoses détectées par l'observateur sur la ligne correspondante. Ainsi le pourcentage de mitoses communes à deux observateurs est différent selon qu'on le considère par rapport à l'ensemble des mitoses détectées par le premier ou le second.*

# Discussion

Les modalités de détection des mitoses sur lame numérique ne sont pas strictement identiques à celles sur lames de verre en microscopie optique conventionnelle. Dans le contexte du carcinome mammaire infiltrant, ces conditions constituent une nouvelle approche dans l'évaluation du grade d'Elston-Ellis pour laquelle il n'existe pas encore de consensus sur la méthode à adopter. Le choix de la zone de comptage, la définition de l'aire en mm<sup>2</sup> de la surface en fonction du moniteur ou des outils de mesure mis à la disposition du pathologiste et les critères d'identification morphologique des cellules en mitose sont autant d'éléments nouveaux dont l'approche diffère sur lame numérique.

Avec le déploiement des scanners de lames dans les laboratoires de pathologie, de plus en plus de pathologistes seront confrontés à ces problématiques.

Une des approches envisagées pour aider le pathologiste dans l'évaluation du score mitotique est l'utilisation d'outils d'intelligence artificielle.

Dans cette étude nous avons participé en collaboration avec une équipe de *data scientists* au développement et à la mise en application d'un outil d'aide à la détection de mitoses utilisable en pratique quotidienne par le pathologiste. Pour ce faire, l'équipe de Primaa est partie d'un modèle d'algorithme dont l'architecture et les paramètres ont fait la preuve de leur efficacité dans la reconnaissance d'image. Ce modèle a été entraîné et testé, puis nous avons évalué quel pouvait être son apport pour le pathologiste, dans un premier temps sur une surface prédéfinie avant de l'appliquer dans un second temps dans le cadre d'un test clinique sur lame entière dont la réalisation pratique est en cours.

## 1. Développement de l'algorithme et de l'interface de lecture

### 1.1. Étape de labellisation

Cette première étape de constitution d'une base de données de mitoses labellisées a été entreprise afin d'augmenter les chances de performance de l'algorithme développé car les bases de données de mitoses numérisées existantes le sont dans des conditions parfois éloignées de la routine diagnostique (36). La labellisation des mitoses soulève un paradoxe important en intelligence artificielle : nous créons nous-même la base d'apprentissage d'un algorithme pour lequel nous souhaitons qu'il nous apporte une aide. Pour obtenir un algorithme performant, il doit être entraîné sur des données robustes. Or, le manque de reproductibilité des pathologistes dans l'appréciation des mitoses en particulier sur lame

numérique fait qu'il peut être difficile d'obtenir un consensus pour certaines figures. Van Diest soulignait déjà en 1992 que ses critères mitotiques définis sur lame de verre pourraient être utilisés en vue de la reconnaissance des mitoses à l'aide de logiciels d'analyse d'image (103).

Néanmoins des études récentes ont montré que ces critères étaient insuffisants sur lame numérique (104,105).

Au cours de cette étape, le temps de labellisation initial par l'interne variait entre une et cinq heures par lame, ceci en fonction de la taille de l'échantillon, de la présence d'artefacts cytologiques, de la densité cellulaire de la tumeur et du nombre de mitoses présentes sur la lame. Pour limiter au maximum les erreurs de labellisation, nous avons effectué une relecture systématique des images avec analyse des figures douteuses sur lame de verre en cas de nécessité. Cette étape a de plus permis la poursuite de notre apprentissage et l'amélioration de notre reconnaissance des mitoses sur lame numérique en particulier quand les critères habituels étaient manquants. Ainsi notre perception des mitoses n'était plus la même au début et à la fin de la phase de labellisation. L'apprentissage des critères de définition des mitoses sur lame numérique tels que définis par l'équipe de Nottingham (105) constitue selon nous une démarche essentielle. En comparant les mitoses entre lame de verre et lame numérique, nous étions en effet arrivés à la fin de notre relecture aux mêmes conclusions que les auteurs en ce qui concerne les détails morphologiques des figures mitotiques.

En pratique quotidienne, il existe de réelles images sur lame de verre pour lesquelles nous ne pouvons pas trancher entre mitose et "non-mitose" sur lame numérique. Nous avons pris le parti de considérer les mitoses et "non-mitoses" douteuses sur lame numérique uniquement si elles étaient vérifiées sur lame de verre. L'idée était donc de constituer une base de données prenant en compte les contraintes de numérisation du laboratoire et de considérer les incertitudes de labellisation pour permettre aux algorithmes d'apprendre sur des données qui constituent la réalité de la pratique quotidienne. Nous avons labellisé au total 2259 mitoses et 4245 "non-mitoses" ce qui constitue l'une des plus importantes bases de données de ce type en pathologie humaine (129). Des bases de données plus importantes, constituées en utilisant des approches de labellisation différentes, existent en pathologie vétérinaire (130,131).

Avec le déploiement progressif de la pathologie numérique dans les laboratoires, la quantité de données disponible pour entraîner et tester les algorithmes augmentera dans les prochaines années. L'utilisation

de scanners de lames appliquant des protocoles de numérisation différents permettra de faciliter la généralisation des performances des modèles développés.

Il se pourrait qu'à l'avenir, un algorithme entraîné sur de multiples bases de données de mitoses de qualités de numérisations différentes, issues de scanners de lames différents et labellisées par des pathologistes de niveau d'expertise variés aboutisse à un "consensus global" qui permettrait de "lisser" les erreurs de labellisation présentes dans chacune des bases de données.

## 1.2. Métriques du modèle

Le modèle d'algorithme retenu par les *data scientists* était basé sur un détecteur de type RetinaNet suivi d'un classifieur de type MobileNet. Dans une démarche visant à généraliser les performances de l'algorithme, il était important que celui-ci ne soit pas uniquement entraîné sur des lames de Bicêtre. Le choix a donc été fait d'inclure des images de MIDOG et d'utiliser des techniques d'augmentation de données. Nous avons obtenu une sensibilité (*recall*) de 61%, une précision de 24,8% et un F1-score de 35,2% sur les lames entières que nous avons labellisées.

Ces performances peuvent paraître décevantes si on les compare à ce qui a pu être atteint au cours des différents challenges. A titre d'exemple, l'équipe gagnante du challenge MIDOG2021 (116) a obtenu une sensibilité (*recall*) de 74%, une précision et un F1-score de 75%. Toutefois plusieurs éléments relativisent le poids de cette différence. Ces métriques sont fortement impactées par le nombre de mitoses présentes sur les lames de test et dépendent de la surface analysée. Lors des challenges, les participants n'analysent pas des lames entières pour entraîner et tester leurs modèles d'identification des mitoses mais uniquement des aires de taille définie (2 mm<sup>2</sup> dans le cas de MIDOG). Dans cette situation, une grande partie des patchs analysés ne contient en réalité que pas ou peu de mitoses. Prenons l'exemple d'un patch où se trouvent dix mitoses. Si l'algorithme en détecte cinq et trouve également deux faux positifs, la sensibilité sera de  $5/10 = 50\%$ , et la précision de  $5/(2+5) = 71,4\%$ . Dans le cas d'un patch où se trouvent deux mitoses, si l'algorithme en trouve une et détecte deux faux positifs, la sensibilité sera de  $1/2 = 50\%$  et la précision de  $1/(1+2) = 33\%$ . De plus, les lames que nous avons utilisées n'ont pas fait l'objet d'une sélection sur la qualité de l'image. Des variations d'épaisseur de coupe ou de coloration sont donc présentes ainsi que cela s'observe dans le flux de routine d'un laboratoire ce qui complexifie encore la tâche de l'algorithme. Ces éléments expliquent pourquoi en utilisant des données de routine les résultats sont inférieurs à ce qui a pu être obtenu lors des challenges. Le challenge TUPAC16 a été conduit sur lame entière mais ne concernait que le score

mitotique sans labellisation des mitoses. L'algorithme le plus performant pour prédire le score mitotique avait obtenu un kappa de Cohen pondéré quadratique de 0,567 signifiant un accord modéré avec les pathologistes selon l'échelle de Landis et Koch mais faible selon l'échelle de McHugh. Entre le consensus et les prédictions de l'algorithme, il existait une discordance globale de 36% et une discordance de 11% entre score 1 et score 3 (113).

Un algorithme capable d'orienter le pathologiste vers la zone de plus forte densité mitotique doit pouvoir analyser la totalité de la surface tumorale présente sur la lame. Il n'existe pas à l'heure actuelle de consensus sur les performances minimales requises en termes de métriques pour pouvoir utiliser un algorithme. Ceci dépend en pratique de l'utilisation qui en est faite (enseignement, recherche, compétition, routine diagnostique, ...). Néanmoins, des métriques élevées sur une base de données dans le cadre d'un challenge ne sont pas une garantie d'efficacité du modèle en condition d'utilisation réelle (120).

### 1.3. Impact des paramètres de numérisation et généralisation des performances de l'algorithme

Entre la numérisation des lames utilisées pour entraîner et tester l'algorithme et la numérisation des cinquante lames utilisées pour la phase 2 de l'étude, des procédures d'entretien préventif hebdomadaires des scanners de lames ont été mises en place par le laboratoire, permettant d'éviter l'accumulation progressives d'altérations imperceptibles, en particulier de nature mécaniques survenant lors de l'utilisation quotidienne des scanners (plus de 1200 lames-jour) et de nature à altérer le rendu visuel de la lame numérisée. De plus, le respect strict des modalités de montage et de séchage des lames avant numérisation a été contrôlé.

Nous avons voulu explorer l'impact de ces procédures sur les performances de l'algorithme. Pour ce faire, quatre lames ont été re-numérisées et ré-analysées par l'algorithme. La comparaison entre les deux analyses est disponible en annexe (**annexe 14**).

Les nombres de vrais positifs, faux positifs et faux négatifs étaient comparables. Néanmoins, les faux positifs proposés étaient différents et certaines figures se situaient en dehors de la zone tumorale car la détection du carcinome en amont de l'algorithme de détection de mitose était parfois erronée.

Ces résultats posent le problème de la capacité des algorithmes à généraliser leurs performances. Même si les résultats en termes de métrique sont comparables, les paramètres de numérisation ont un impact direct sur les données présentées au pathologiste. Ces éléments soulignent que des contrôles

qualité réguliers seront nécessaires pour s'assurer que les performances de l'algorithme restent constantes au sein du laboratoire. Une étape de calibration de l'algorithme serait également indispensable en cas d'utilisation dans d'autres laboratoires. L'algorithme pourrait potentiellement être initialisé avec les lames du laboratoire précédent par des méthodes de *transfer learning* pour faciliter son apprentissage dans l'analyse des lames de nouveaux laboratoires. Toutefois dans la mesure où la grande majorité des laboratoires d'anatomopathologie n'ont pas encore effectué leur transition vers un flux de travail entièrement numérisé, cette étape n'est pas encore aisément généralisable.

#### 1.4. Interface de visualisation de l'algorithme

De bonnes performances dans le cadre d'un challenge avec des conditions spécifiques ne garantissent pas des performances similaires du modèle en pratique quotidienne et obtenir des performances sur lames entières comparables à celles des challenges est une tâche ardue. Pour essayer de résoudre cette difficulté, nous avons donc travaillé à une interface qui permettrait au pathologiste d'interagir avec les résultats en lui donnant la possibilité de visualiser dès le faible grossissement les mitoses détectées par l'algorithme. Par cette approche le pathologiste a ainsi la possibilité de rapidement évaluer les éventuels faux-positifs détectés.

Plusieurs pistes d'amélioration de l'interface à laquelle nous avons pour le moment abouti peuvent être suggérées. La première serait de permettre au pathologiste de sélectionner spécifiquement une mitose sur la lame pour la visualiser dans un onglet dédié sans être contraint de zoomer. En effet, il est actuellement seulement possible de faire le contraire, sélectionner une mitose dans l'onglet pour la localiser sur la lame.

La seconde amélioration serait de permettre au pathologiste de tracer directement son aire de comptage d'une surface et d'une forme définies pour ensuite faire apparaître directement dans l'onglet les mitoses détectées par l'algorithme présentes dans l'aire choisie. En effet il n'est actuellement possible de déposer que des carrés de 0,2 mm<sup>2</sup> et l'algorithme ne réajuste pas les données présentées dans l'onglet en fonction des carrés déposés par le pathologiste ce qui rend l'outil actuel utile pour choisir la zone de comptage mais peu ergonomique pour effectuer le comptage mitotique en tant que tel. Il est admis depuis la parution du dernier OMS que le comptage par unité de surface doit être privilégié par rapport au comptage par 10 champs à fort grossissement. Cette notion est encore plus vraie en pathologie numérique où la surface affichée varie fortement d'un moniteur à l'autre et n'est pas standardisée comme les champs au microscope (100). Parallèlement, des outils de mesure de surface sont

maintenant intégrés aux différents logiciels de visualisation et pourraient parfaitement à l'avenir se combiner avec des algorithmes de détection de mitoses.

## 2. Reproductibilité des pathologistes et apport de l'algorithme

Le second temps de ce travail a été l'étude de la reproductibilité des pathologistes lors de l'évaluation des mitoses dans une cohorte de 50 cas de carcinome mammaire infiltrant sur une surface de 2 mm<sup>2</sup> avec et sans utilisation de l'algorithme. Cela a également permis de pouvoir comparer la détection de chaque mitose de façon individuelle, chose impossible sur lame entière où la zone de comptage entre pathologiste peut être différente.

Nous avons considéré l'échelle de Koo et Li pour l'interprétation du coefficient de corrélation intra-classe (CCI) dans le cas du compte mitotique et celle de McHugh pour l'interprétation du coefficient kappa dans le cas du score correspondant.

Pour chaque pathologiste, la reproductibilité intra-observateur entre les deux lectures pour l'évaluation du compte mitotique était bonne pour un pathologiste et excellente pour les trois autres et celle pour l'évaluation du score mitotique correspondant était modérée à forte.

Concernant la reproductibilité inter-observateur, sauf entre les pathologistes 1 et 3 le CCI pour chaque couple de pathologiste était amélioré avec l'utilisation de l'algorithme mais la différence n'était pas significative.

Le CCI pour l'évaluation du compte mitotique par l'ensemble des pathologistes était de 0,716 IC = [0,531 - 0,833] sans assistance de l'algorithme et de 0,746 IC = [0,543 - 0,859] avec l'algorithme témoignant dans les deux cas d'une concordance moyenne à bonne. Le kappa de Fleiss portant sur la reproductibilité du score mitotique par l'ensemble des pathologistes était de 0,416 sans assistance de l'algorithme et de 0,512 avec assistance de l'algorithme témoignant dans les deux cas d'une faible concordance selon l'échelle de McHugh. Ces valeurs sont toutefois en cohérence avec l'étude de Rabe *et al.* (84). On peut cependant noter que les pathologistes 2, 3 et 4 étaient peu en accord avec le pathologiste 1 en ce qui concerne le nombre de mitoses détectées. Cela se traduit par un intervalle de confiance fortement élargi autour de la valeur du CCI par rapport aux trois autres pathologistes entre eux et une concordance plus faible sur le score mitotique correspondant mesurée par le kappa de Cohen. De ce point de vue, la concordance observée entre les pathologistes 2, 3 et 4 et l'algorithme était supérieure à celle observée avec le pathologiste 1. Néanmoins, la concordance entre ces trois

pathologistes était supérieure à celle observée entre eux et l'algorithme. Entre les pathologistes 2, 3 et 4 le CCI pour l'évaluation du compte mitotique sans assistance de l'algorithme était de 0,830 IC = [0,745 - 0,893] et de 0,897 IC = [0,842 - 0,936] assisté de l'algorithme témoignant dans les deux cas d'une bonne concordance. Le kappa de Fleiss était de 0,594 sans utilisation de l'algorithme et de 0,763 avec utilisation de l'algorithme avec un accord passant de faible à modéré. La concordance des pathologistes avec l'algorithme pour l'évaluation du compte mitotique était faible pour le pathologiste 1, moyenne pour le pathologiste 2 et 3 et bonne pour le pathologiste 4. La concordance des pathologistes avec l'algorithme pour l'évaluation du score mitotique correspondant était faible pour les pathologistes 1 et 2 et modérée pour les pathologistes 3 et 4.

En regardant les mitoses individuellement, nous avons constaté que pour chaque mitose labellisée par un pathologiste, seule la moitié d'entre elles étaient labellisées par un second pathologiste et 13% seulement ont obtenu l'approbation des trois autres médecins. Le **tableau XV** montre que les performances des pathologistes étaient comparable à l'algorithme pour reconnaître les mitoses des autres observateurs sur lame numérique. Ces résultats vont dans le même sens que l'étude de Veta *et al.* (117) où les désaccords des pathologistes sur chaque mitose prise individuellement ne les empêchaient pas de s'accorder sur le score mitotique. Toutefois dans leur étude, le meilleur accord observé entre un pathologiste et l'algorithme était inférieur au plus faible accord observé entre pathologiste soulignant que les pathologistes étaient davantage d'accords entre eux qu'avec l'algorithme. Nous avons pour notre part retrouvé des niveaux de concordance comparable en considérant l'ensemble des observateurs. Un élément pouvant expliquer cette différence est que dans leur étude, la base de données utilisée pour entraîner/tester l'algorithme et celle utilisée pour évaluer la reproductibilité des participants étaient issues de deux laboratoires différents. Bien que nos lames soient différentes en phase 1 et phase 2 de l'étude, elles étaient - sauf en ce qui concerne la base de données MIDOG - issues du même laboratoire. De plus, dans leur étude, les observateurs étaient tous trois des pathologistes experts en pathologie mammaire alors que dans la nôtre, les niveaux d'expertise étaient différents.

L'impact clinique des discordances peut être marqué quand le score diffère de deux points (1 *versus* 3) (85,113). De ce point de vue, l'utilisation de l'algorithme ne nous a pas permis de réduire cette discordance qui était respectivement de 6,3% et 6,6%, sans et avec l'algorithme. Elle était de 4,5%

entre les pathologistes et l'algorithme lui-même. En considérant les pathologistes 2, 3 et 4, cette discordance était de 2%, identique à celle observée avec les scores établis par l'algorithme.

Dans l'étude de Balkenhol *et al.* (118), ayant comparé le compte mitotique sur lame de verre par rapport à une zone de *hotspot* de 2 mm<sup>2</sup> prédéfinie par un algorithme, la reproductibilité des pathologistes était nettement améliorée dans la seconde situation. Ce résultat était plus ou moins attendu dans la mesure où dans cette situation la seconde lecture contraignait la zone de comptage par rapport à la lame de verre où les pathologistes étaient libres de faire leur compte où ils le souhaitaient. Il était en effet plus difficile de mettre en évidence une différence dans notre situation où la zone de comptage était identique lors des deux lectures. Néanmoins dans notre cas, les données présentées par l'algorithme étaient le seul facteur modifié entre les deux lectures. Dans l'étude de Balkenhol *et al.* seule la zone de *hotspot* était présentée et les mitoses individuelles détectées par l'algorithme n'étaient pas visibles par les pathologistes. Ceci explique que la reproductibilité intra-observateur des pathologistes dans notre travail soit nettement supérieure à celle observée dans leur étude et que notre reproductibilité inter-observateur ne soit par contre pas significativement augmentée en seconde lecture.

Toujours dans l'étude de Balkenhol *et al.* (118), concernant l'accord des deux pathologistes avec l'algorithme dans la zone définie de 2 mm<sup>2</sup>, le premier avait avec l'algorithme un kappa pondéré pour l'évaluation du score mitotique de 0,654 et un CCI pour le compte mitotique de 0,895 IC = [0,845 - 0,930]. Le second avait un kappa de 0,794 et un CCI de 0,888 IC = [0,783 - 0,936]. La concordance sur le compte mitotique était donc meilleure que ce que nous avons pu observer avec néanmoins une reproductibilité du score comparable. Les auteurs avaient également noté que dans le cas des tumeurs comportant peu de mitoses, l'algorithme avait défini la zone de *hotspot* à partir de détections erronées de faux-positifs, soulignant par là-même l'importance de permettre au pathologiste d'avoir un visuel sur les figures proposées par l'algorithme pour définir la zone de comptage. Ce regard critique est indispensable. En effet, dans le cadre de notre étude, certaines mitoses authentiques étaient proposées par l'algorithme avec un indice de confiance inférieur à certains faux-positifs.

En l'état actuel des connaissances, il apparaît peu envisageable d'obtenir un algorithme qui détecte toutes les mitoses d'une lame sans faire d'erreur. Les images douteuses, difficiles d'interprétation continueront d'exister indépendamment de l'utilisation de l'intelligence artificielle qui elle-même détectera invariablement des objets tendancieux et parfois difficiles à classer pour le pathologiste. La problématique sous-jacente étant celle du consensus ou de la "vérité" sur ce qui constitue une figure

mitotique sur lame numérique à l'heure où le gold standard pour l'évaluation des mitoses est encore l'utilisation de la lame de verre.

Dans notre étude nous n'avons pas établi de lecture consensus d'expert en amont pour l'évaluation des mitoses sur les aires de 2 mm<sup>2</sup> prédéfinies. Or une nouvelle méthode de mesure ou de comptage pour être considérée valide doit d'une part être reproductible et d'autre part fournir des valeurs dites "correctes". Il est donc nécessaire de comparer la mesure qu'elle propose à une mesure de référence. Néanmoins nous avons vu en introduction à quel point obtenir un consensus pour l'évaluation des mitoses peut être difficile et ceci est vérifié par les données des **tableaux XIII, XIV et XV**. Si on considère que la reproductibilité des pathologistes est améliorée par l'utilisation de l'algorithme, il est cependant impossible d'affirmer formellement qui du pathologiste 1 ou des trois autres, a ou ont établi les scores mitotiques les plus exacts et quelles sont les proportions de vrais positifs, faux positifs et faux négatifs de mitoses dans les évaluations de chacun.

Dans l'étude de Balkenhol *et al.* (118), la sélection des cas était stratifiée sur le score donné par un des deux pathologistes sur lame de verre avec une proportion égale des trois scores mitotiques. Les matrices de confusion montraient que dans leur cohorte, il existait une augmentation du nombre de scores 1 sur lame numérique par rapport à la lame de verre au dépend du nombre de scores 2. La proportion de scores 3 restait comparable.

Dans notre étude, la surreprésentation des scores 1 dans les évaluations des pathologistes 2, 3 et 4 par rapport au pathologiste 1 (**cf. annexe 11**) pourrait être liée soit à une sous-estimation du score mitotique sur lame numérique comme observé dans des études antérieures (81,86,105,106), soit à un biais de sélection des cas. Nous n'avons pas effectué de sélection stratifiée sur le score mitotique mais une sélection "temporelle" dans le cadre de l'enregistrement chronologique des cas dans le service. Bien que le but de notre étude était d'évaluer l'apport de l'algorithme et non de comparer lame numérique et lame de verre, il serait intéressant de comparer les scores obtenus par les quatre pathologistes à celui d'une lecture des cas au microscope.

On peut toutefois noter que dans l'hypothèse où les pathologistes de l'étude auraient sous-estimé les scores mitotiques, l'utilisation de l'algorithme en seconde lecture n'a pas permis de compenser cette sous-estimation.

### 3. Perspectives

L'apport de l'algorithme sur lame entière pour le pathologiste est en cours d'évaluation. Cette étude, phase 3 de notre travail porte sur 50 cas de carcinomes infiltrants du sein distincts des cas utilisés pour le test sur surface définie répartis en 25 biopsies et 25 pièces opératoires issues du flux de travail en routine du laboratoire d'anatomopathologie de l'hôpital de Bicêtre. Ces lames ont été lues par trois pathologistes experts afin d'établir un consensus sur le nombre de mitoses / 2 mm<sup>2</sup> et le score mitotique qui en découle pour chaque cas.

Deux pathologistes ont par ailleurs examiné ces 50 lames selon un schéma en *crossover* avec la moitié des cas analysés sans l'algorithme et l'autre moitié avec, puis inversion des cas lors d'une seconde lecture après une période de *wash-out* de quatre semaines.

Pour toutes les phases de l'étude, l'aire de comptage, les mitoses annotées et les comptes mitotiques de chaque pathologiste ont été enregistrés ou sont en cours d'enregistrement.

Les analyses des deux pathologistes seront comparées entre elles, avec celle consensuelle des experts et avec l'algorithme lui-même. Comme ce qui a été présenté dans ce travail, les résultats porteront sur la reproductibilité inter-observateur des pathologistes pour le compte mitotique et le score correspondant mais également pour le choix de la zone de comptage. La comparaison avec le consensus d'expert nous permettra de vérifier l'exactitude des analyses avec et sans l'algorithme.

Nous espérons ainsi apporter des éléments de réponse sur l'utilité du modèle développé et son interface pour assister le pathologiste dans son analyse des mitoses en routine sur lame numérique.

Une comparaison des scores avec une évaluation sur lame de verre pourrait également être envisagée dans la mesure où cette méthode reste pour l'heure la référence pour l'établissement du grade d'Elston-Ellis. A notre connaissance, ce travail est le premier à être concrètement engagé vers l'intégration d'un algorithme de détection de mitoses au flux de travail de routine du pathologiste.

Cet algorithme n'a pas vocation à se substituer à l'avis du pathologiste. Il ne peut faire que des propositions d'images que le pathologiste choisit de considérer ou non comme mitose.

Une des limites actuelles des algorithmes d'intelligence artificielle dans la reconnaissance des mitoses est leur manque de capacité à considérer les informations contextuelles. Ces données s'intègrent en effet de façon intuitive dans l'esprit du pathologiste lorsqu'il analyse ses lames et en pratique, deux cellules d'aspect comparable pourraient être considérées en mitoses ou en "non-mitoses" selon l'aspect global de la tumeur, des cellules adjacentes, de la présence de remaniements nécrotiques ou

d'artefacts. Néanmoins la question des capacités prédictives du modèle peut être soulevée. Dans notre étude, les comptes mitotiques et les scores des pathologistes 2, 3 et 4 étaient en effet très proches de ceux déterminés par l'algorithme (**annexes 9, 10 et 13**). De nombreuses études de validation sur des bases de données différentes seront nécessaires avant d'envisager la possibilité de déléguer l'établissement du score mitotique à un algorithme.

D'autres questions restent en suspens. Comment ce type d'outil serait-il accepté par la communauté des pathologistes? Comment serait-il concrètement intégré dans le flux de travail de nos laboratoires? Les analyses seraient-elles faites avant même que le pathologiste accède à la lame numérique par prescription automatique lors de l'enregistrement du cas dans le système de gestion de laboratoire? Où bien l'outil serait-il intégré comme une nouvelle fonctionnalité du système de gestion d'image utilisé, appliqué à la demande du pathologiste lors de la lecture des lames du cas? Comment le pathologiste s'appropriera-t-il les données présentées par l'algorithme et comment les intégrera-t-il dans son compte rendu? Quels seraient les contrôles qualité à mettre en place pour s'assurer que les performances de l'algorithme restent constantes? L'évaluation médico-économique de l'outil devra également être considérée.

Au-delà des problématiques de reproductibilité, l'utilité pratique de l'outil développé dans notre étude devra être prise en compte. L'apport de celui-ci en terme de gain de temps, d'ergonomie et de praticité pour le compte mitotique sur lame numérique devra faire l'objet d'études spécifiques.

Les résultats de notre étude et les questions évoquées pourraient s'intégrer dans une réflexion plus globale qui pourrait amener à modifier la méthode d'évaluation du score mitotique. Le nombre de mitoses requis par unité de surface pour définir chacun des trois scores pourrait également être amené à évoluer pour tenir compte des différences observables sur lame numérique et des modifications éventuelles induites par l'utilisation des outils d'intelligence artificielle. L'évolution des pratiques qui pourrait en découler nécessiterait des études supplémentaires à plus grande échelle et intégrant la survie des patientes pour valider la valeur pronostique de ces nouvelles méthodes d'analyse.

# Conclusion

De par sa valeur pronostique et son intégration dans le grade d'Elston-Ellis, l'évaluation du score mitotique est un élément indispensable dans la prise en charge de tout carcinome infiltrant du sein.

La perception morphologique des mitoses sur lame numérique n'est pas strictement superposable à celle sur lame de verre pouvant conduire à un risque potentiel de sous-estimation du score mitotique.

Néanmoins, et en tenant compte de ce risque potentiel, l'utilisation de l'intelligence artificielle constitue une approche possible pour assister le pathologiste dans la détection des mitoses sur lame numérique.

De nombreuses études ont montré la faisabilité de la détection de mitoses par des algorithmes mais dans des conditions spécifiques parfois éloignées de la routine diagnostique.

Dans cette étude, après avoir labellisé 2259 mitoses et 4245 "non-mitoses" - ce qui constitue l'une des plus importantes bases de données de ce type en pathologie humaine -, un algorithme de détection de mitoses et son interface utilisateur ont été développés.

Les résultats de notre évaluation de l'apport de cet outil pour le pathologiste sur une surface définie de 2 mm<sup>2</sup> ont montré que dans cette situation, la reproductibilité des pathologistes n'était pas significativement améliorée par l'utilisation de l'algorithme. La concordance des pathologistes entre eux pour l'évaluation du score mitotique était comparable à celle observée entre les pathologistes et l'algorithme. Les performances de l'algorithme étaient également comparables à celle des pathologistes pour reconnaître les mitoses identifiées par les autres observateurs sur lame numérique.

L'analyse de l'apport de cet outil sur lame entière - qui constitue la finalité de son utilisation - en termes de reproductibilité des pathologistes pour le compte mitotique, le score correspondant et le choix de la zone de comptage, par rapport à un consensus de trois experts, est en fin d'évaluation.

# Bibliographie

1. Organisation mondiale de la Santé. Cancer du sein [Internet]. 2021 [cité 23 août 2022]. Disponible sur: <https://www.who.int/fr/news-room/fact-sheets/detail/breast-cancer>
2. Le cancer du sein - Les cancers les plus fréquents [Internet]. [cité 18 août 2022]. Disponible sur: <https://www.e-cancer.fr/Professionnels-de-sante/Les-chiffres-du-cancer-en-France/Epidemiologie-des-cancers/Les-cancers-les-plus-frequents/Cancer-du-sein>
3. Organisation mondiale de la santé, éditeur. Breast tumours. 5th ed. Geneva: OMS; 2019. 368 p. (World health organization classification of tumours).
4. Allison KH. Prognostic and predictive parameters in breast pathology: a pathologist's primer. *Mod Pathol*. janv 2021;34(S1):94-106.
5. Ferreira R, Moon B, Humphries J, Sussman A, Saltz J, Miller R, et al. The Virtual Microscope. *Proc Conf Am Med Inform Assoc AMIA Fall Symp*. 1997;449-53.
6. Abels E, Pantanowitz L, Aeffner F, Zarella MD, van der Laak J, Bui MM, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol*. nov 2019;249(3):286-94.
7. Williams BJ, Treanor D. Practical guide to training and validation for primary diagnosis with digital pathology. *J Clin Pathol*. juill 2020;73(7):418-22.
8. Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: current status and future perspectives. *Histopathology*. juill 2012;61(1):1-9.
9. Vergier B, Guettier C. L'utilisation des lames virtuelles en pédagogie. *médecine/sciences*. nov 2012;28(11):986-9.
10. Bertheau P, Chabouis A, Fabiani B, Poullier É, Daniel C, Cucherousset J, et al. Télépathologie par lames virtuelles ou le diagnostic anatomo-pathologique en réseau numérique. *médecine/sciences*. nov 2012;28(11):983-5.
11. Laurent-Bellue A, Poullier E, Prevot S, Guettier C. [Telepathology for intra-operative frozen sections]. *Ann Pathol*. avr 2019;39(2):113-8.
12. Ameisen D, Le Naour G, Daniel C. Technologie des lames virtuelles: De la numérisation à la mise en ligne. *médecine/sciences*. nov 2012;28(11):977-82.
13. Snead DRJ, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*. juin 2016;68(7):1063-72.
14. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, et al. Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *Am J Surg Pathol*. janv 2018;42(1):39-52.
15. Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. *Mod Pathol Off J U S Can Acad Pathol Inc*. 2019;32(7):916-28.
16. Azam AS, Miligy IM, Kimani PKU, Maqbool H, Hewitt K, Rajpoot NM, et al. Diagnostic concordance and discordance in digital pathology: a systematic review and meta-analysis. *J Clin Pathol*. juill 2021;74(7):448-55.
17. Williams B, Hanby A, Millican-Slater R, Verghese E, Nijhawan A, Wilson I, et al. Digital pathology for primary diagnosis of screen-detected breast lesions - experimental data, validation and experience from four centres. *Histopathology*. juin 2020;76(7):968-75.
18. Al-Janabi S, Huisman A, Willems SM, Van Diest PJ. Digital slide images for primary diagnostics in breast pathology: a feasibility study. *Hum Pathol*. déc 2012;43(12):2318-25.
19. Williams BJ, Hanby A, Millican-Slater R, Nijhawan A, Verghese E, Treanor D. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. *Histopathology*. 2018;72(4):662-71.
20. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. nov 1958;65(6):386-408.
21. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep Learning and Its Applications in Biomedicine. *Genomics Proteomics Bioinformatics*. févr 2018;16(1):17-32.

22. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2 févr 2017;542(7639):115-8.
23. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 13 déc 2016;316(22):2402-10.
24. Anirudh R, Thiagarajan JJ, Bremer T, Kim H. Lung nodule detection using 3D convolutional neural networks trained on weakly labeled data. In: Tourassi GD, Armato SG, éditeurs. San Diego, California, United States; 2016 [cité 23 nov 2021]. p. 978532. Disponible sur: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2214876>
25. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep Learning to Improve Breast Cancer Detection on Screening Mammography. *Sci Rep*. 29 août 2019;9(1):12495.
26. Robertson S, Azizpour H, Smith K, Hartman J. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Transl Res*. avr 2018;194:19-35.
27. da Silva LM, Pereira EM, Salles PG, Godrich R, Ceballos R, Kunz JD, et al. Independent real-world application of a clinical-grade automated prostate cancer detection system. *J Pathol*. juin 2021;254(2):147-58.
28. Le Cun Y, Brizard C, Brizard C. Quand la machine apprend: la révolution des neurones artificiels et de l'apprentissage profond. Paris: Odile Jacob; 2019. 394 p.
29. Zemouri R, Devalland C, Valmary-Degano S, Zerhouni N. Intelligence artificielle : quel avenir en anatomie pathologique ? *Ann Pathol*. avr 2019;39(2):119-29.
30. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 28 mai 2015;521(7553):436-44.
31. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016;7(1):29.
32. Yousif M, van Diest PJ, Laurinavicius A, Rimm D, van der Laak J, Madabhushi A, et al. Artificial intelligence applied to breast pathology. *Virchows Arch Int J Pathol*. 18 nov 2021;
33. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. *J Pathol Inform*. 2019;10:9.
34. Laurinavicius A, Laurinaviciene A, Dasevicius D, Elie N, Plancoulaine B, Bor C, et al. Digital Image Analysis in Pathology: Benefits and Obligation. *Anal Cell Pathol*. 2012;35(2):75-8.
35. Park SH, Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology*. mars 2018;286(3):800-9.
36. Tizhoosh H, Pantanowitz L. Artificial intelligence and digital pathology: Challenges and opportunities. *J Pathol Inform*. 2018;9(1):38.
37. Cai D, Sun X, Zhou N, Han X, Yao J. Efficient Mitosis Detection in Breast Cancer Histology Images by RCNN. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) [Internet]. Venice, Italy: IEEE; 2019 [cité 2 déc 2021]. p. 919-22. Disponible sur: <https://ieeexplore.ieee.org/document/8759461/>
38. Riber-Hansen R, Vainer B, Steiniche T. Digital image analysis: a review of reproducibility, stability and basic requirements for optimal results: IMAGE ANALYSIS IN PATHOLOGY RESEARCH. *APMIS*. avr 2012;120(4):276-89.
39. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital Imaging in Pathology: Whole-Slide Imaging and Beyond. *Annu Rev Pathol Mech Dis*. 24 janv 2013;8(1):331-59.
40. Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, et al. BACH: Grand challenge on breast cancer histology images. *Med Image Anal*. août 2019;56:122-39.
41. Cruz-Roa A, Gilmore H, Basavanthally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci Rep*. juin 2017;7(1):46450.
42. Rakha EA, Ahmed MA, Aleskandarany MA, Hodi Z, Lee AHS, Pinder SE, et al. Diagnostic concordance of breast pathologists: lessons from the National Health Service Breast Screening Programme Pathology External Quality Assurance Scheme. *Histopathology*. 2017;70(4):632-42.
43. Rakha EA, Bennett RL, Coleman D, Pinder SE, Ellis IO. Review of the national external quality assessment (EQA) scheme for breast pathology in the UK. *J Clin Pathol*. 1 janv 2017;70(1):51-7.

44. Polónia A, Campelos S, Ribeiro A, Aymore I, Pinto D, Biskup-Fruzynska M, et al. Artificial Intelligence Improves the Accuracy in Histologic Classification of Breast Lesions. *Am J Clin Pathol.* 15 mars 2021;155(4):527-36.
45. Bejnordi BE, Zuidhof G, Balkenhol M, Hermsen M, Bult P, van Ginneken B, et al. Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *ArXiv170503678 Cs [Internet].* 10 mai 2017 [cité 11 déc 2021]; Disponible sur: <http://arxiv.org/abs/1705.03678>
46. Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermsen M, et al. From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE Trans Med Imaging.* févr 2019;38(2):550-60.
47. Steiner DF, MacDonald R, Liu Y, Truszowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol.* 2018;42(12):1636-46.
48. Andre F, Ismaila N, Henry NL, Somerfield MR, Bast RC, Barlow W, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: ASCO Clinical Practice Guideline Update-Integration of Results From TAILORx. *J Clin Oncol Off J Am Soc Clin Oncol.* 1 août 2019;37(22):1956-64.
49. Thomson TA, Hayes MM, Spinelli JJ, Hilland E, Sawrenko C, Phillips D, et al. HER-2/neu in breast cancer: interobserver variability and performance of immunohistochemistry with 4 antibodies compared with fluorescent *in situ* hybridization. *Mod Pathol Off J U S Can Acad Pathol Inc.* nov 2001;14(11):1079-86.
50. Mengel M, von Wasielewski R, Wiese B, Rüdiger T, Müller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J Pathol.* nov 2002;198(3):292-9.
51. Polley MYC, Leung SCY, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JMS, et al. An international study to increase concordance in Ki67 scoring. *Mod Pathol Off J U S Can Acad Pathol Inc.* juin 2015;28(6):778-86.
52. Leung SCY, Nielsen TO, Zabaglo LA, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. *Histopathology.* août 2019;75(2):225-35.
53. Tuominen VJ, Ruotoistenmäki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res BCR.* 2010;12(4):R56.
54. Tuominen VJ, Tolonen TT, Isola J. ImmunoMembrane: a publicly available web application for digital image analysis of HER2 immunohistochemistry. *Histopathology.* avr 2012;60(5):758-67.
55. Mungle T, Tewary S, Das DK, Arun I, Basak B, Agarwal S, et al. MRF-ANN: a machine learning approach for automated ER scoring of breast cancer immunohistochemical images. *J Microsc.* août 2017;267(2):117-29.
56. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. *JNCI J Natl Cancer Inst.* 1 juill 2021;113(7):808-19.
57. Rimm DL, Leung SCY, McShane LM, Bai Y, Bane AL, Bartlett JMS, et al. An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer. *Mod Pathol Off J U S Can Acad Pathol Inc.* janv 2019;32(1):59-69.
58. Vandenberghe ME, Scott MLJ, Scorer PW, Söderberg M, Balcerzak D, Barker C. Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer. *Sci Rep.* 5 avr 2017;7:45938.
59. Holten-Rossing H, Møller Talman ML, Kristensson M, Vainer B. Optimizing HER2 assessment in breast cancer: application of automated image analysis. *Breast Cancer Res Treat.* juill 2015;152(2):367-75.
60. Hartage R, Li A, Hammond S, Parwani A. A validation study of human epidermal growth factor receptor 2 immunohistochemistry digital imaging analysis and its correlation with human epidermal growth factor receptor 2 fluorescence *in situ* hybridization results in breast carcinoma. *J Pathol Inform.* 2020;11(1):2.
61. Acs B, Pelekanou V, Bai Y, Martinez-Morilla S, Toki M, Leung SCY, et al. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. *Lab Invest.* janv 2019;99(1):107-17.
62. Bui MM, Riben MW, Allison KH, Chlipala E, Colasacco C, Kahn AG, et al. Quantitative Image Analysis of Human Epidermal Growth Factor Receptor 2 Immunohistochemistry for Breast Cancer: Guideline From the College of American Pathologists. *Arch Pathol Lab Med.* 1 oct 2019;143(10):1180-95.

63. Whitney J, Corredor G, Janowczyk A, Ganesan S, Doyle S, Tomaszewski J, et al. Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. *BMC Cancer*. 30 mai 2018;18(1):610.
64. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *Npj Breast Cancer*. déc 2018;4(1):30.
65. Dodington DW, Lagree A, Tabbarah S, Mohebpour M, Sadeghi-Naini A, Tran WT, et al. Analysis of tumor nuclear features using artificial intelligence to predict response to neoadjuvant chemotherapy in high-risk breast cancer patients. *Breast Cancer Res Treat*. avr 2021;186(2):379-89.
66. Turkki R, Byckhov D, Lundin M, Isola J, Nordling S, Kovanen PE, et al. Breast cancer outcome prediction with tumour tissue images and machine learning. *Breast Cancer Res Treat*. août 2019;177(1):41-52.
67. Klimov S, Miligy IM, Gertych A, Jiang Y, Toss MS, Rida P, et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (DCIS) recurrence risk. *Breast Cancer Res BCR*. 29 juill 2019;21(1):83.
68. Greenough RB. Varying Degrees of Malignancy in Cancer of the Breast. *J Cancer Res*. 1 déc 1925;9(4):453-63.
69. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer*. sept 1957;11(3):359-77.
70. Elston CW, Gresham GA, Rao GS, Zebro T, Haybittle JL, Houghton J, et al. The cancer research campaign (King's/Cambridge trial for early breast cancer: clinico-pathological aspects. *Br J Cancer*. mai 1982;45(5):655-69.
71. Le Doussal V, Tubiana-Hulin M, Friedman S, Hacene K, Spyrtatos F, Brunet M. Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (SBR). An improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer*. 1 nov 1989;64(9):1914-21.
72. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. nov 1991;19(5):403-10.
73. Pereira H, Pinder SE, Sibbering DM, Galea MH, Elston CW, Blamey RW, et al. Pathological prognostic factors in breast cancer. IV: Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. *Histopathology*. sept 1995;27(3):219-26.
74. Amin MB, American Joint Committee on Cancer, American Cancer Society, éditeurs. *AJCC cancer staging manual*. Eight edition / editor-in-chief, Mahul B. Amin, MD, FCAP ; editors, Stephen B. Edge, MD, FACS [and 16 others] ; Donna M. Gress, RHIT, CTR-Technical editor ; Laura R. Meyer, CAPM-Managing editor. Chicago IL: American Joint Committee on Cancer, Springer; 2017. 1024 p.
75. Rakha EA, El-Sayed ME, Lee AHS, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 juill 2008;26(19):3153-8.
76. Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res BCR*. 2010;12(4):207.
77. van Dooijeweert C, Baas IO, Deckers IAG, Siesling S, van Diest PJ, van der Wall E. The increasing importance of histologic grading in tailoring adjuvant systemic therapy in 30,843 breast cancer patients. *Breast Cancer Res Treat*. juin 2021;187(2):577-86.
78. Dalle JR, Leow WK, Racoceanu D, Tutac AE, Putti TC. Automatic breast cancer grading of histopathological images. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf*. 2008;2008:3052-5.
79. Basavanahally A, Ganesan S, Feldman M, Shih N, Mies C, Tomaszewski J, et al. Multi-field-of-view framework for distinguishing tumor grade in ER+ breast cancer from entire histopathology slides. *IEEE Trans Biomed Eng*. août 2013;60(8):2089-99.
80. Genestie C, Zafrani B, Asselain B, Fourquet A, Rozan S, Validire P, et al. Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. *Anticancer Res*. févr 1998;18(1B):571-6.
81. Rakha EA, Aleskandarani M, Toss MS, Green AR, Ball G, Ellis IO, et al. Breast cancer histologic grading using digital microscopy: concordance and outcome association. *J Clin Pathol*. août 2018;71(8):680-6.
82. Donovan TA, Moore FM, Bertram CA, Luong R, Bolfa P, Klopffleisch R, et al. Mitotic Figures—Normal, Atypical, and Imposters: A Guide to Identification. *Vet Pathol*. mars 2021;58(2):243-57.

83. van Dooijeweert C, van Diest PJ, Ellis IO. Grading of invasive breast carcinoma: the way forward. *Virchows Arch* [Internet]. 1 juill 2021 [cité 17 nov 2021]; Disponible sur: <https://link.springer.com/10.1007/s00428-021-03141-2>
84. Rabe K, Snir OL, Bossuyt V, Harigopal M, Celli R, Reisenbichler ES. Interobserver variability in breast carcinoma grading results in prognostic stage differences. *Hum Pathol*. déc 2019;94:51-7.
85. Rakha EA, Aleskandarany MA, Toss MS, Mongan NP, ElSayed ME, Green AR, et al. Impact of breast cancer grade discordance on prediction of outcome. *Histopathology*. déc 2018;73(6):904-15.
86. Davidson TM, Rendi MH, Frederick PD, Onega T, Allison KH, Mercan E, et al. Breast cancer prognostic factors in the digital era: Comparison of Nottingham grade using whole slide images and glass slides. *J Pathol Inform*. 1 janv 2019;10(1):11.
87. Ginter PS, Idress R, D'Alfonso TM, Fineberg S, Jaffer S, Sattar AK, et al. Histologic grading of breast carcinoma: a multi-institution study of interobserver variation using virtual microscopy. *Mod Pathol*. avr 2021;34(4):701-9.
88. Maqlin P, Thamburaj R, Mammen JJ, Nagar AK. Automatic Detection of Tubules in Breast Histopathological Images. In: Bansal JC, Singh P, Deep K, Pant M, Nagar A, éditeurs. *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)* [Internet]. India: Springer India; 2013 [cité 23 nov 2021]. p. 311-21. Disponible sur: [http://link.springer.com/10.1007/978-81-322-1041-2\\_27](http://link.springer.com/10.1007/978-81-322-1041-2_27)
89. Nguyen K, Barnes M, Srinivas C, Chef'd'hotel C. Automatic glandular and tubule region segmentation in histological grading of breast cancer. In: Gurcan MN, Madabhushi A, éditeurs. *Orlando, Florida, United States; 2015* [cité 23 nov 2021]. p. 94200G. Disponible sur: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2082322>
90. Xu J, Janowczyk A, Chandran S, Madabhushi A. A high-throughput active contour scheme for segmentation of histopathological imagery. *Med Image Anal*. déc 2011;15(6):851-62.
91. Belsare AD, Mushrif MM, Pangarkar MA, Meshram N. Breast histopathology image segmentation using spatio-colour-texture based graph partition method. *J Microsc*. 2016;262(3):260-73.
92. Das A, Nair MS, Peter SD. Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review. *J Digit Imaging*. oct 2020;33(5):1091-121.
93. Naylor P, Lae M, Reyat F, Walter T. Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map. *IEEE Trans Med Imaging*. 2019;38(2):448-59.
94. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans Med Imaging*. juill 2017;36(7):1550-60.
95. Veta M, van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JPW. Automatic nuclei segmentation in H&E stained breast cancer histopathology images. *PloS One*. 2013;8(7):e70221.
96. Feng Y, Zhang L, Yi Z. Breast cancer cell nuclei classification in histopathology images using deep neural networks. *Int J Comput Assist Radiol Surg*. févr 2018;13(2):179-91.
97. Veta M, Kornegoor R, Huisman A, Verschuur-Maes AHJ, Viergever MA, Pluim JPW, et al. Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer. *Mod Pathol*. déc 2012;25(12):1559-65.
98. MITOS-ATYPIA-14 - Grand Challenge [Internet]. [grand-challenge.org](http://grand-challenge.org). [cité 18 août 2022]. Disponible sur: <https://mitos-atypia-14.grand-challenge.org/Results2/>
99. Contesso G, Mouriessse H, Friedman S, Genin J, Sarrazin D, Rouesse J. The importance of histologic grade in long-term prognosis of breast cancer: a study of 1,010 patients, uniformly treated at the Institut Gustave-Roussy. *J Clin Oncol Off J Am Soc Clin Oncol*. sept 1987;5(9):1378-86.
100. Ibrahim A, Lashen AG, Katayama A, Mihai R, Ball G, Toss MS, et al. Defining the area of mitoses counting in invasive breast cancer using whole slide image. *Mod Pathol*. 11 déc 2021;1-10.
101. Kaman EJ, Smeulders AW, Verbeek PW, Young IT, Baak JP. Image processing for mitoses in sections of breast cancer: a feasibility study. *Cytometry*. mai 1984;5(3):244-9.
102. ten Kate TK, Beliën JA, Smeulders AW, Baak JP. Method for counting mitoses by image processing in Feulgen stained breast cancer sections. *Cytometry*. 1993;14(3):241-50.

103. van Diest PJ, Baak JPA, Matze-Cok P, Wisse-Brekemans ECM, van Galen CM, Kurver PHJ, et al. Reproducibility of mitosis counting in 2,469 breast cancer specimens: Results from the Multicenter Morphometric Mammary Carcinoma Project. *Hum Pathol.* juin 1992;23(6):603-7.
104. Ibrahim A, Lashen A, Toss M, Mihai R, Rakha E. Assessment of mitotic activity in breast cancer: revisited in the digital pathology era. *J Clin Pathol.* 23 sept 2021;jclinpath-2021-207742.
105. Lashen A, Ibrahim A, Katayama A, Ball G, Mihai R, Toss M, et al. Visual assessment of mitotic figures in breast cancer: a comparative study between light microscopy and whole slide images. *Histopathology.* 28 août 2021;
106. Al-Janabi S, van Slooten HJ, Visser M, van der Ploeg T, van Diest PJ, Jiwa M. Evaluation of mitotic activity index in breast cancer using whole slide digital images. *PLoS One.* 2013;8(12):e82576.
107. Kim D, Pantanowitz L, Schüffler P, Yarlagadda DK, Ardon O, Reuter V, et al. (Re) Defining the high-power field for digital pathology. *J Pathol Inform.* 2020;11(1):33.
108. Hartman D, Laak JeroenAWMVD, Gurcan M, Pantanowitz L. Value of public challenges for the development of pathology deep learning algorithms. *J Pathol Inform.* 2020;11(1):7.
109. Jiménez G, Racoceanu D. Deep Learning for Semantic Segmentation vs. Classification in Computational Pathology: Application to Mitosis Analysis in Breast Cancer Grading. *Front Bioeng Biotechnol.* 21 juin 2019;7:145.
110. Nofallah S, Mehta S, Mercan E, Knezevich S, May CJ, Weaver D, et al. Machine learning techniques for mitoses classification. *Comput Med Imaging Graph Off J Comput Med Imaging Soc.* janv 2021;87:101832.
111. Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, et al. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *J Pathol Inform.* 2013;4(1):8.
112. Veta M, van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med Image Anal.* févr 2015;20(1):237-48.
113. Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, et al. Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Med Image Anal.* 2019;54:111-21.
114. Zerhouni E, Lanyi D, Viana M, Gabrani M. Wide residual networks for mitosis detection. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) [Internet]. Melbourne, Australia: IEEE; 2017 [cité 3 déc 2021]. p. 924-8. Disponible sur: <http://ieeexplore.ieee.org/document/7950667/>
115. Aubreville M, Bertram C, Veta M, Klopfleisch R, Stathonikos N, Breininger K, et al. Mitosis Domain Generalization Challenge. 2 mars 2021 [cité 3 déc 2021]; Disponible sur: <https://zenodo.org/record/4573978>
116. Aubreville M, Stathonikos N, Bertram CA, Klopfleisch R, ter Hoeve N, Ciompi F, et al. Mitosis domain generalization in histopathology images -- The MIDOG challenge [Internet]. arXiv; 2022 [cité 30 juill 2022]. Disponible sur: <http://arxiv.org/abs/2204.03742>
117. Veta M, Diest PJ van, Jiwa M, Al-Janabi S, Pluim JPW. Mitosis Counting in Breast Cancer: Object- Level Interobserver Agreement and Comparison to an Automatic Method. *PLOS ONE.* 16 août 2016;11(8):e0161286.
118. Balkenhol MCA, Tellez D, Vreuls W, Clahsen PC, Pinckaers H, Ciompi F, et al. Deep learning assisted mitotic counting for breast cancer. *Lab Invest J Tech Methods Pathol.* nov 2019;99(11):1596-606.
119. Pantanowitz L, Hartman D, Qi Y, Cho EY, Suh B, Paeng K, et al. Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagn Pathol.* 4 juill 2020;15(1):80.
120. Bertram CA, Veta M, Marzahl C, Stathonikos N, Maier A, Klopfleisch R, et al. Are pathologist-defined labels reproducible? Comparison of the TUPAC16 mitotic figure dataset with an alternative set of labels. *ArXiv200705351 Cs Eess Q-Bio.* 2020;12446:204-13.
121. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection [Internet]. arXiv; 2018 [cité 1 juill 2022]. Disponible sur: <http://arxiv.org/abs/1708.02002>
122. Tellez D, Litjens G, Bandi P, Bulten W, Bokhorst JM, Ciompi F, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal.* déc 2019;58:101544.
123. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [Internet]. arXiv; 2017 [cité 1 juill 2022]. Disponible sur: <http://arxiv.org/abs/1704.04861>
124. ImageNet [Internet]. [cité 31 août 2022]. Disponible sur: <https://www.image-net.org/index.php>
125. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* juin 2016;15(2):155-63.
126. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica.* 2012;22(3):276-82.

127. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. mars 1977;33(1):159-74.
128. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. déc 1994;6(4):284-90.
129. Çayır S, Solmaz G, Kusetogullari H, Tokat F, Bozaba E, Karakaya S, et al. MITNET: a novel dataset and a two-stage deep learning approach for mitosis recognition in whole slide images of breast cancer tissue. *Neural Comput Appl* [Internet]. 4 juin 2022 [cité 27 août 2022]; Disponible sur: <https://link.springer.com/10.1007/s00521-022-07441-9>
130. Aubreville M, Bertram CA, Marzahl C, Gurtner C, Dettwiler M, Schmidt A, et al. Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region. *Sci Rep*. déc 2020;10(1):16447.
131. Bertram CA, Aubreville M, Marzahl C, Maier A, Klopffleisch R. A large-scale dataset for mitotic figure assessment on whole slide images of canine cutaneous mast cell tumor. *Sci Data*. 21 nov 2019;6(1):274.

# Annexes

Annexe 1 : détail des cas labellisés sur Cytomine en phase 1 .....	77
Annexe 2 : protocoles de numérisation en usage en routine dans le service d'anatomopathologie de l'hôpital Bicêtre (scanners de lames P250 et P1000 3DHISTECH) .....	77
Annexe 3 : répartition des mitoses et "non-mitoses" labellisées en phase 1 pour chaque cas – utilisation des cas dans le processus d'entraînement/validation et de test de l'algorithme .....	78
Annexe 4 : <i>bounding-box</i> (carré vert) autour d'une mitose .....	79
Annexe 5 : minimisation de la fonction de perte lors de l'entraînement et la validation du détecteur ..	79
Annexe 6 : exemples d'augmentations de données pour deux images différentes .....	80
Annexe 7 : minimisation de la fonction de perte pour l'entraînement et la validation du détecteur avec et sans augmentation de données .....	81
Annexe 8 : minimisation de la fonction de perte (en haut), évolution du <i>recall</i> (au milieu) et de la précision (en bas) du classifieur MobileNet.....	82
Annexe 9 : nombre de mitoses et score mitotique des quatre pathologistes pour les 50 cas vus sans l'algorithme .....	83
Annexe 10 : nombre de mitoses et score mitotique des quatre pathologistes pour les 50 cas vus avec l'algorithme .....	84
Annexe 11 : matrices de confusion détaillant la répartition des scores mitotiques de chaque pathologiste avec et sans l'algorithme.....	85
Annexe 12 : matrices de confusion détaillant la répartition des scores mitotiques des pathologistes entre eux lors de la première et de la seconde lecture .....	86
Annexe 13 : matrices de confusion détaillant la répartition des scores mitotiques entre la première lecture des pathologistes et l'algorithme .....	87
Annexe 14 : impact de la renumérisation des lames sur la détection des mitoses par l'algorithme .....	88

### Annexe 1 : détail des cas labellisés sur Cytomine en phase 1

	Nombre de lames	Type de prélèvement	Age de la patiente	Type histologique	Grade (A/N/M)*	Présence de CCIS	Présence d'emboles	RO	RP	HER2	Ki67
Cas 1	3	Tumorectomie	66	Carcinome infiltrant de type NST	1 (2/2/1)	oui	non	positif	négatif	0	10%
Cas 2	3	Tumorectomie	63	Carcinome infiltrant de type NST	1 (2/2/1)	non	non	positif	positif	1+	20%
Cas 3	2	Tumorectomie	67	Carcinome infiltrant de type NST	1 (1/1/1)	non	non	positif	positif	0	20%
Cas 4	1	Tumorectomie	78	Carcinome infiltrant de type NST	2 (3/2/1)	non	non	positif	positif	1+	15%
Cas 5	1	Biopsie	40	Carcinome infiltrant de type NST	3 (3/2/3)	non	non	négatif	négatif	0	60%
Cas 6	3	Biopsie	86	Carcinome infiltrant de type NST	3 (3/3/3)	non	non	négatif	négatif	0	60%
Cas 7	1	Tumorectomie	44	Carcinome infiltrant de type NST	3 (3/2/3)	non	non	positif	positif	1+	30%
Cas 8	4	Tumorectomie	59	Carcinome infiltrant de type NST	3 (3/3/3)	non	non	négatif	négatif	0	25%
Cas 9	7	Biopsie	69	Carcinome infiltrant de type NST	2 (3/3/1)	oui	non	positif	positif	0	15%
Cas 10	6	Biopsie	60	Carcinome infiltrant de type NST	2 (3/2/2)	non	non	positif	négatif	2+ FISH amplifié	50%
Cas 11	1	Biopsie	43	Carcinome infiltrant de type NST à stroma riche en lymphocytes	3 (3/3/3)	non	non	négatif	négatif	0	80%

\*(Architecture/Noyaux/Mitoses)

### Annexe 2 : protocoles de numérisation en usage en routine dans le service d'anatomopathologie de l'hôpital Bicêtre (scanners de lames P250 et P1000 3DHISTECH)

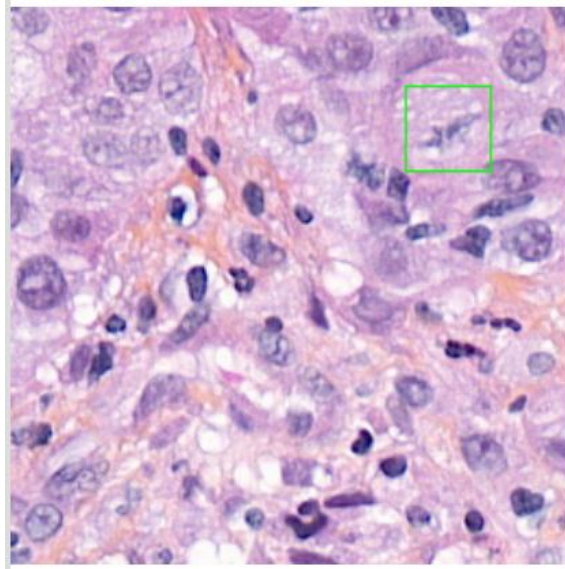
Name	20xAuto	40xAuto
Output resolution	41x Native	82x Native
Limits of the range	1750 - 1850	1750 - 1850
Focus distance	5 FOVs	10 FOVs
Illumination mode	BrightField	BrightField
Specimen threshold	220	205
Multilayer mode	Single layer	Single layer
Scanning mode	Flash	Flash
Color matrix	sRGB standard	sRGB standard
Image quality	good	good
Compression	JPG	JPG
Bit depth	8-bit	8-bit

**Annexe 3 : répartition des mitoses et "non-mitoses" labellisées en phase 1 pour chaque cas – utilisation des cas dans le processus d'entraînement/validation et de test de l'algorithme**

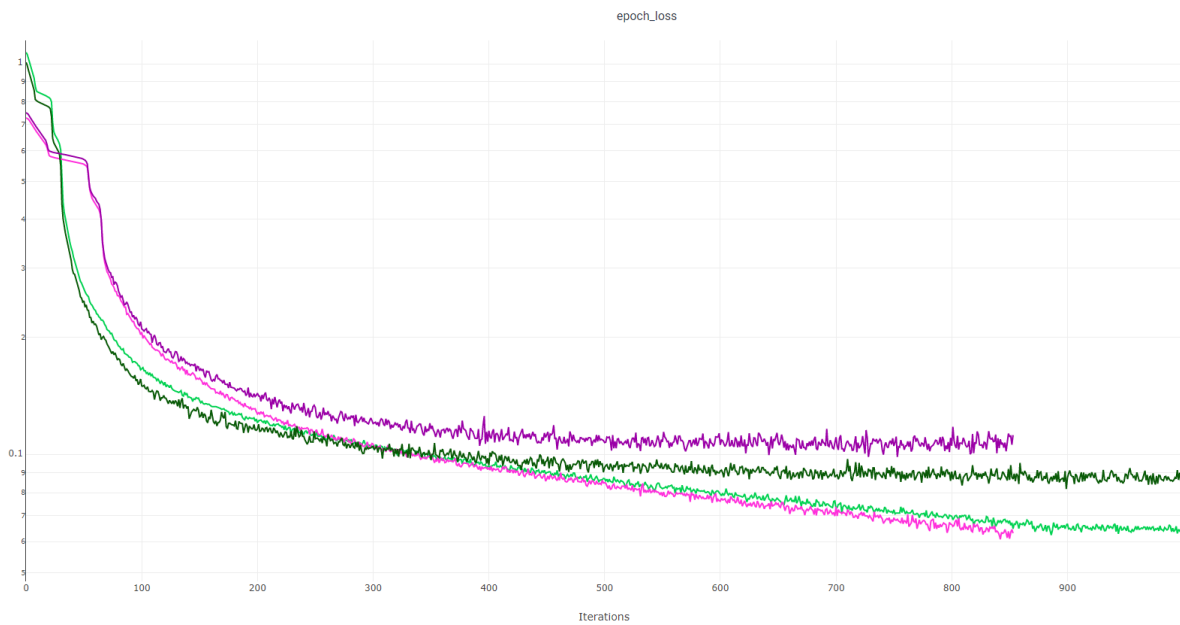
Cas	Lames	Mitoses à la fin de la relecture	"Non-mitoses" à la fin de la relecture	Total final	Mitoses reclassées en "non-mitoses" à la relecture	Mitoses exclues à la relecture	Utilisation du cas*
Cas 1	3	15	78	93	14	1	Test
Cas 2	3	37	328	365	33	3	Test
Cas 3	2	16	176	192	44	0	Test
Cas 4	1	24	222	246	17	1	Test
Cas 5	1	380	514	894	181	0	Non utilisée
Cas 6	3	799	770	1569	192	125	Entraînement / Validation
Cas 7	1	56	163	219	0	6	Test
Cas 8	4	655	752	1407	85	25	Entraînement / Validation
Cas 9	7	14	721	735	7	6	Test
Cas 10	6	102	439	541	9	69	Entraînement / Validation
Cas 11	1	161	82	243	17	19	Non utilisée
<b>Total</b>	<b>32</b>	<b>2259</b>	<b>4245</b>	<b>6504</b>	<b>599</b>	<b>255</b>	

\*Utilisation des cas par la start-up pour les étapes d'entraînement / validation et de test de l'algorithme. Deux lames n'ont pas été utilisées (cas 5, cas 11) du fait d'artefacts techniques rendant leur exploitation difficile par les *data scientists* dans le processus d'entraînement / validation et de test du modèle. Le nombre de mitoses utilisées pour la phase de test de l'algorithme dans notre travail (121) diffère du nombre total de mitoses à la fin de la relecture (162) car la base de donnée utilisée par les *data scientists* a été figée un peu avant la toute fin de la labellisation.

#### Annexe 4 : *bounding-box* (carré vert) autour d'une mitose



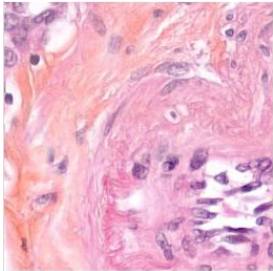
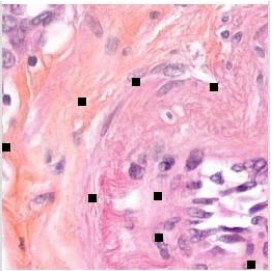
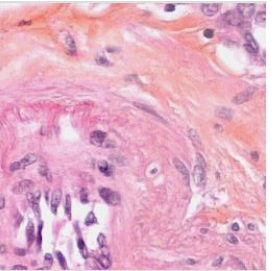
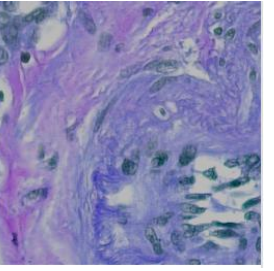
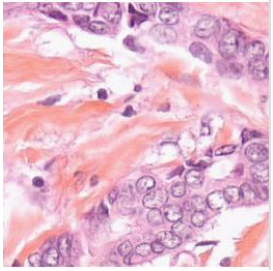
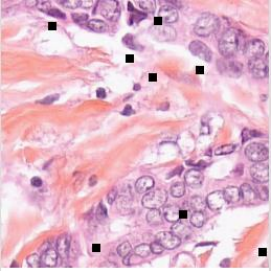
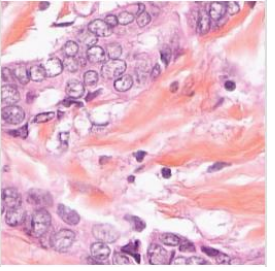
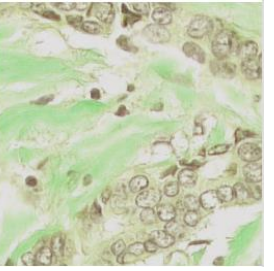
#### Annexe 5 : minimisation de la fonction de perte lors de l'entraînement et la validation du détecteur



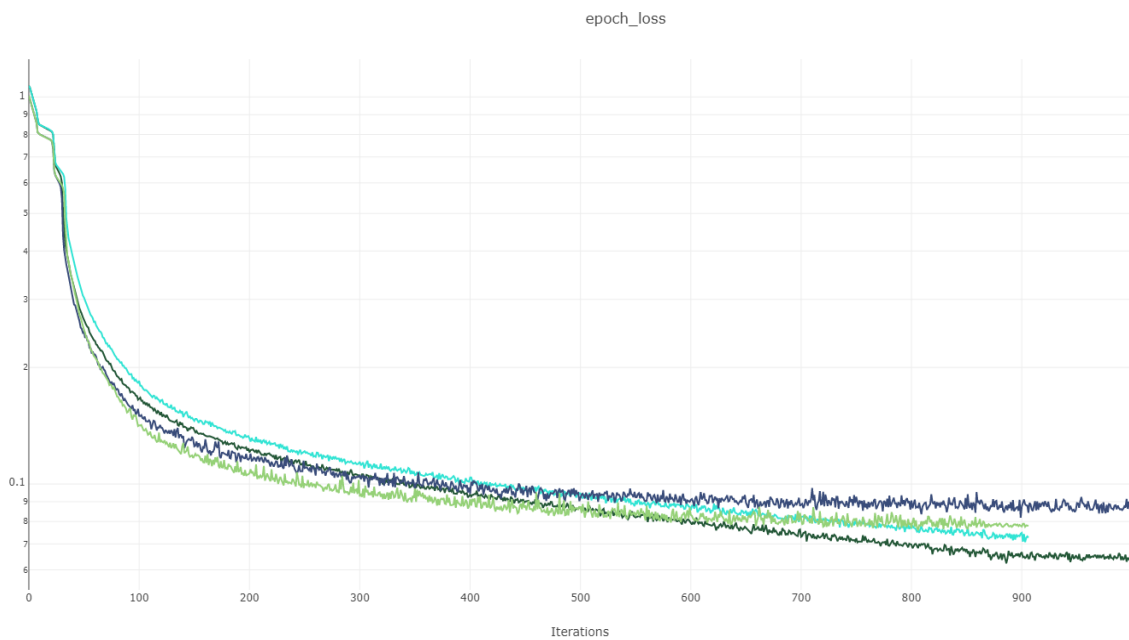
- █ perte (*train*) - entraîné avec des données du KB
- █ perte (*val*) - entraîné avec des données du KB
- █ perte (*train*) - entraîné avec des données du KB et de MIDOG
- █ perte (*val*) - entraîné avec des données du KB et de MIDOG

La minimisation de la fonction de perte entre les deux modalités d'entraînement rend compte de la différence entre la labellisation et la détection de l'algorithme. Plus elle est faible, moins il y a d'erreurs. Pour chaque modèle, la perte sur la donnée d'entraînement et la donnée de validation (donnée similaire à l'entraînement mais sur laquelle le réseau n'apprend pas) a été mesurée. Plus les courbes d'entraînement et de validation sont proches, plus l'algorithme est généralisable.

Annexe 6 : exemples d'augmentations de données pour deux images différentes

image	<i>cutout</i>	<i>flip</i>	couleurs
			
			

## Annexe 7 : minimisation de la fonction de perte pour l'entraînement et la validation du détecteur avec et sans augmentation de données

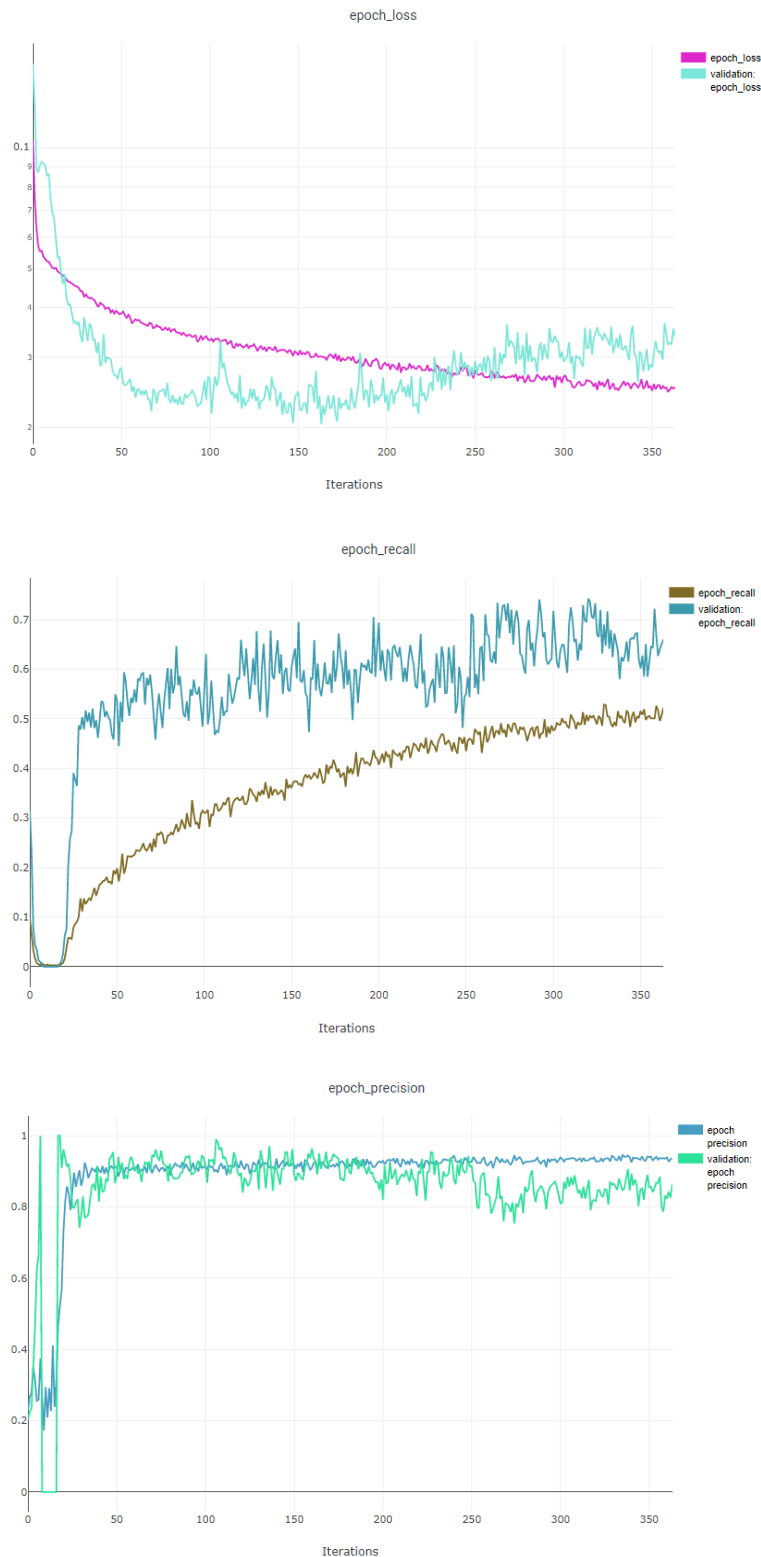


- perte (*train*) - entraîné avec des données du KB et de MIDOG
- perte (*val*) - entraîné avec des données du KB et de MIDOG
- perte (*train*) - entraîné avec des données du KB et de MIDOG + augmentation de données
- perte (*val*) - entraîné avec des données du KB et de MIDOG + augmentation de données

Deux points peuvent être soulignés à partir de ce graphique :

- Les courbes de perte d'entraînement et de validation correspondant au modèle sans augmentation de données sont plus éloignées l'une de l'autre. Dans ce cas, le modèle a sur-appris. Il est bon sur les données lors de son entraînement, mais moins sur les données de validation, pourtant similaires et issues des mêmes lames. Il n'a pas appris à généraliser. Lorsqu'on ajoute de l'augmentation de données, ces deux courbes se rapprochent : le modèle a généralisé.
- En plus d'avoir généralisé, le modèle a mieux appris avec l'augmentation de données : sa courbe de perte de validation est en dessous de celle attribuée au modèle sans augmentation de données.

## Annexe 8 : minimisation de la fonction de perte (en haut), évolution du *recall* (au milieu) et de la précision (en bas) du classifieur MobileNet



Les courbes modélisant la fonction de perte ainsi que les courbes d'évolution de la sensibilité (*recall*) et de la précision pour les jeux de données d'entraînement et de validation au cours de l'entraînement du MobileNet montrent un changement de comportement de l'algorithme à la 200<sup>ème</sup> epoch. C'est à ce point qu'a commencé à apparaître le sur-apprentissage du classifieur (la courbe d'entraînement continuait de diminuer tandis que la courbe de validation commençait à augmenter). C'est également à ce point que la sensibilité (*recall*) augmentait mais la précision diminuait (le F1-score n'était plus amélioré).

**Annexe 9 : nombre de mitoses et score mitotique des quatre pathologistes pour les 50 cas vus sans l'algorithme**

	Sans l'algorithme							
	Pathologiste 1		Pathologiste 2		Pathologiste 3		Pathologiste 4	
	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )
Cas 1	1	1	3	1	0	1	0	1
Cas 2	0	1	2	1	0	1	0	1
Cas 3	16	3	11	2	6	1	8	2
Cas 4	8	2	4	1	2	1	5	1
Cas 5	8	2	5	1	3	1	3	1
Cas 6	46	3	19	3	31	3	31	3
Cas 7	66	3	27	3	54	3	29	3
Cas 8	9	2	1	1	0	1	0	1
Cas 9	15	3	8	2	8	2	2	1
Cas 10	31	3	5	1	14	2	13	2
Cas 11	26	3	4	1	8	2	8	2
Cas 12	2	1	1	1	2	1	0	1
Cas 13	4	1	3	1	0	1	1	1
Cas 14	22	3	12	2	16	3	3	1
Cas 15	26	3	7	1	9	2	11	2
Cas 16	23	3	1	1	4	1	5	1
Cas 17	36	3	38	3	30	3	28	3
Cas 18	13	2	3	1	6	1	0	1
Cas 19	25	3	12	2	14	2	15	3
Cas 20	2	1	3	1	1	1	1	1
Cas 21	2	1	2	1	0	1	0	1
Cas 22	10	2	4	1	6	1	2	1
Cas 23	13	2	6	1	4	1	4	1
Cas 24	15	3	3	1	5	1	7	1
Cas 25	5	1	2	1	0	1	3	1
Cas 26	27	3	7	1	17	3	17	3
Cas 27	10	2	7	1	9	2	11	2
Cas 28	0	1	3	1	0	1	1	1
Cas 29	0	1	1	1	0	1	0	1
Cas 30	14	2	4	1	4	1	8	2
Cas 31	1	1	2	1	0	1	0	1
Cas 32	10	2	4	1	2	1	4	1
Cas 33	9	2	1	1	1	1	4	1
Cas 34	4	1	4	1	3	1	2	1
Cas 35	7	1	1	1	1	1	2	1
Cas 36	3	1	2	1	1	1	2	1
Cas 37	7	1	3	1	3	1	6	1
Cas 38	2	1	4	1	1	1	1	1
Cas 39	18	3	13	2	3	1	9	2
Cas 40	6	1	3	1	2	1	2	1
Cas 41	3	1	2	1	0	1	0	1
Cas 42	3	1	0	1	0	1	1	1
Cas 43	3	1	1	1	0	1	1	1
Cas 44	3	1	3	1	0	1	0	1
Cas 45	2	1	2	1	0	1	0	1
Cas 46	40	3	17	3	21	3	25	3
Cas 47	16	3	3	1	10	2	6	1
Cas 48	69	3	17	3	38	3	40	3
Cas 49	0	1	1	1	0	1	0	1
Cas 50	11	2	7	1	2	1	6	1

**Annexe 10 : nombre de mitoses et score mitotique des quatre pathologistes pour les 50 cas vus avec l'algorithme**

	Avec l'algorithme								Algorithme	
	Pathologiste 1		Pathologiste 2		Pathologiste 3		Pathologiste 4		Nombre de mitoses	Score (2 mm <sup>2</sup> )
	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )	Nombre de mitoses	Score (2 mm <sup>2</sup> )		
Cas 1	0	1	0	1	0	1	0	1	9	2
Cas 2	0	1	0	1	0	1	0	1	1	1
Cas 3	21	3	13	2	6	1	7	1	9	2
Cas 4	8	2	5	1	3	1	6	1	6	1
Cas 5	7	1	4	1	2	1	5	1	4	1
Cas 6	48	3	24	3	29	3	37	3	26	3
Cas 7	59	3	26	3	25	3	31	3	20	3
Cas 8	7	1	3	1	0	1	0	1	2	1
Cas 9	19	3	10	2	5	1	5	1	8	2
Cas 10	30	3	7	1	6	1	14	2	15	3
Cas 11	27	3	4	1	6	1	5	1	5	1
Cas 12	4	1	1	1	0	1	0	1	3	1
Cas 13	4	1	1	1	0	1	1	1	3	1
Cas 14	28	3	15	3	7	1	3	1	4	1
Cas 15	24	3	11	2	14	2	13	2	15	3
Cas 16	21	3	5	1	3	1	5	1	7	1
Cas 17	38	3	35	3	26	3	15	3	15	3
Cas 18	16	3	5	1	4	1	2	1	4	1
Cas 19	26	3	15	3	15	3	15	3	13	2
Cas 20	7	1	5	1	1	1	3	1	5	1
Cas 21	1	1	2	1	1	1	1	1	1	1
Cas 22	13	2	3	1	7	1	3	1	9	2
Cas 23	11	2	4	1	4	1	5	1	7	1
Cas 24	11	2	6	1	5	1	6	1	6	1
Cas 25	4	1	1	1	0	1	0	1	1	1
Cas 26	29	3	11	2	12	2	18	3	9	2
Cas 27	10	2	8	2	9	2	11	2	13	2
Cas 28	4	1	1	1	0	1	0	1	1	1
Cas 29	1	1	0	1	0	1	1	1	1	1
Cas 30	14	2	5	1	5	1	6	1	9	2
Cas 31	2	1	2	1	0	1	0	1	1	1
Cas 32	10	2	2	1	2	1	2	1	2	1
Cas 33	8	2	0	1	2	1	2	1	5	1
Cas 34	5	1	1	1	2	1	1	1	4	1
Cas 35	5	1	2	1	2	1	1	1	5	1
Cas 36	1	1	1	1	1	1	0	1	3	1
Cas 37	7	1	7	1	6	1	7	1	7	1
Cas 38	1	1	3	1	0	1	0	1	0	1
Cas 39	11	2	15	3	10	2	7	1	0	1
Cas 40	6	1	4	1	2	1	3	1	3	1
Cas 41	2	1	1	1	0	1	1	1	0	1
Cas 42	1	1	0	1	0	1	0	1	0	1
Cas 43	1	1	1	1	0	1	0	1	4	1
Cas 44	7	1	0	1	0	1	0	1	0	1
Cas 45	0	1	0	1	0	1	0	1	0	1
Cas 46	37	3	20	3	23	3	27	3	12	2
Cas 47	10	2	4	1	4	1	7	1	2	1
Cas 48	50	3	25	3	38	3	41	3	19	3
Cas 49	0	1	0	1	0	1	0	1	0	1
Cas 50	8	2	6	1	5	1	5	1	0	1

**Annexe 11 : matrices de confusion détaillant la répartition des scores mitotiques de chaque pathologiste avec et sans l'algorithme**

		Pathologiste 1 avec l'algorithme			
		Score 1	Score 2	Score 3	Total
<b>Pathologiste 1 sans l'algorithme</b>	<b>Score 1</b>	22	2	0	24
	<b>Score 2</b>	0	8	3	11
	<b>Score 3</b>	0	1	14	15
	<b>Total</b>	22	11	17	50
		Pathologiste 2 avec l'algorithme			
		Score 1	Score 2	Score 3	Total
<b>Pathologiste 2 sans l'algorithme</b>	<b>Score 1</b>	37	3	0	40
	<b>Score 2</b>	0	2	3	5
	<b>Score 3</b>	0	0	5	5
	<b>Total</b>	37	5	8	50
		Pathologiste 3 avec l'algorithme			
		Score 1	Score 2	Score 3	Total
<b>Pathologiste 3 sans l'algorithme</b>	<b>Score 1</b>	35	1	0	36
	<b>Score 2</b>	4	2	1	7
	<b>Score 3</b>	1	1	5	7
	<b>Total</b>	40	4	6	50
		Pathologiste 4 avec l'algorithme			
		Score 1	Score 2	Score 3	Total
<b>Pathologiste 4 sans l'algorithme</b>	<b>Score 1</b>	36	0	0	36
	<b>Score 2</b>	4	3	0	7
	<b>Score 3</b>	0	0	7	7
	<b>Total</b>	40	3	7	50



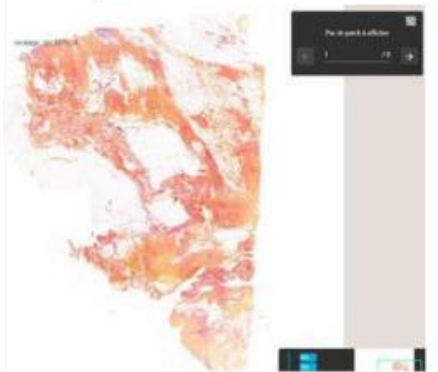

**Annexe 12 : matrices de confusion détaillant la répartition des scores mitotiques des pathologistes entre eux lors de la première et de la seconde lecture**

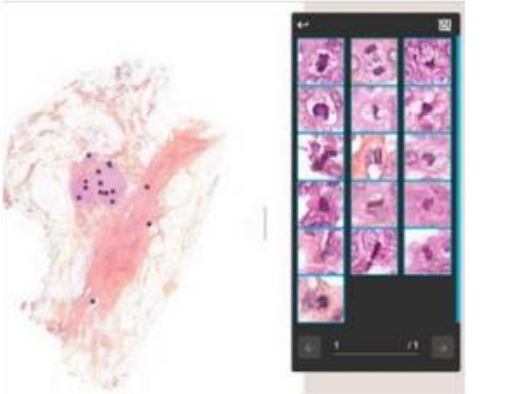
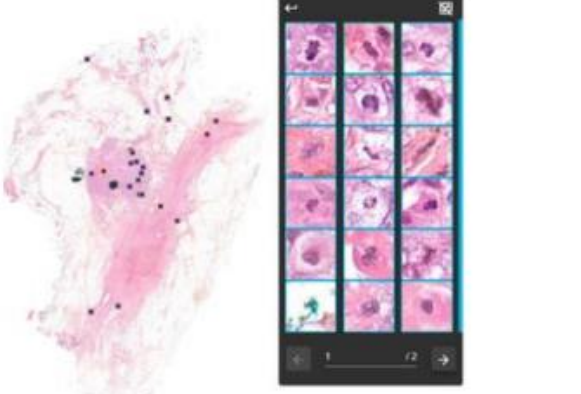


		Sans utilisation de l'algorithme				Avec utilisation de l'algorithme			
		Pathologiste 1							
		Score 1	Score 2	Score 3	Total	Score 1	Score 2	Score 3	Total
Pathologiste 2	Score 1	22	11	7	40	24	9	4	37
	Score 2	0	0	5	5	0	1	4	5
	Score 3	0	0	5	5	0	1	7	8
	Total	22	11	17	50	24	11	15	50
		Pathologiste 1							
Pathologiste 3	Score 1	22	10	4	36	24	9	7	40
	Score 2	0	1	6	7	0	2	2	4
	Score 3	0	0	7	7	0	0	6	6
	Total	22	11	17	50	24	11	15	50
		Pathologiste 1							
Pathologiste 4	Score 1	22	9	5	36	24	10	6	40
	Score 2	0	2	5	7	0	1	2	3
	Score 3	0	0	7	7	0	0	7	7
	Total	22	11	17	50	24	11	15	50
		Pathologiste 2							
Pathologiste 3	Score 1	34	2	0	36	37	2	1	40
	Score 2	5	2	0	7	0	3	1	4
	Score 3	1	1	5	7	0	0	6	6
	Total	40	5	5	50	37	5	8	50
		Pathologiste 2							
Pathologiste 4	Score 1	34	2	0	36	36	2	2	40
	Score 2	5	2	0	7	1	2	0	3
	Score 3	1	1	5	7	0	1	6	7
	Total	40	5	5	50	37	5	8	50
		Pathologiste 3							
Pathologiste 4	Score 1	33	2	1	36	39	1	0	40
	Score 2	3	4	0	7	1	2	0	3
	Score 3	0	1	6	7	0	1	6	7
	Total	36	7	7	50	40	4	6	50

**Annexe 13 : matrices de confusion détaillant la répartition des scores mitotiques entre la première lecture des pathologistes et l'algorithme**

		Algorithme			
		Score 1	Score 2	Score 3	Total
Pathologiste 1	Score 1	21	1	0	22
	Score 2	8	3	0	11
	Score 3	6	5	6	17
	Total	35	9	6	50
		Algorithme			
		Score 1	Score 2	Score 3	Total
Pathologiste 2	Score 1	33	5	2	40
	Score 2	2	3	0	5
	Score 3	0	1	4	5
	Total	35	9	6	50
		Algorithme			
		Score 1	Score 2	Score 3	Total
Pathologiste 3	Score 1	32	4	0	36
	Score 2	2	3	2	7
	Score 3	1	2	4	7
	Total	35	9	6	50
		Algorithme			
		Score 1	Score 2	Score 3	Total
Pathologiste 4	Score 1	33	3	0	36
	Score 2	2	3	2	7
	Score 3	0	3	4	7
	Total	35	9	6	50

**Annexe 14 : impact de la renumérisation des lames sur la détection des mitoses par l'algorithme**

	old numerisation	new numerisation
<i>slides</i>		
<i>metrics</i>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 5,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 20.0,   "precision": 2.857142857142857,    "true_positives": 1,   "false_positives": 34,   "false_negatives": 4 } </pre>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 5,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 20.0,   "precision": 3.3333333333333335,    "true_positives": 1,   "false_positives": 29,   "false_negatives": 4 } </pre>
<i>slides</i>		
<i>metrics</i>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": false,   "max_mitotic_zone_in_top_10": false,   "rank_most_mitotic_zone": NaN,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": NaN,   "precision": NaN,    "true_positives": 0,   "false_positives": 0,   "false_negatives": 0 } </pre>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 1,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": NaN,   "precision": NaN,    "true_positives": 0,   "false_positives": 11,   "false_negatives": 0 } </pre>

<i>slides</i>		
<i>metrics</i>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 1,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 25.0,   "precision": 6.25,    "true_positives": 1,   "false_positives": 15,   "false_negatives": 3 } </pre>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 1,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 0.0,   "precision": 0.0,    "true_positives": 0,   "false_positives": 23,   "false_negatives": 4 } </pre>
<i>slides</i>		
<i>metrics</i>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 2,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 42.10526315789473,   "precision": 20.0,    "true_positives": 8,   "false_positives": 32,   "false_negatives": 11 } </pre>	<pre> "mitosis": {   "max_mitotic_zone_in_top_5": true,   "max_mitotic_zone_in_top_10": true,   "rank_most_mitotic_zone": 1,    "grade_label": 1,   "grade_analyse": 1,    "sensitivity": 52.63157894736842,   "precision": 21.73913043478261,    "true_positives": 10,   "false_positives": 36,   "false_negatives": 9 } </pre>

## RESUME

**Introduction :** le grade d'Elston-Ellis, dont l'évaluation du score mitotique est un des éléments, est un facteur pronostique majeur du carcinome mammaire infiltrant. La reproductibilité de son évaluation reste néanmoins discutée. En pathologie numérique, la sous-estimation du score mitotique rapportée pourrait être évitée par l'utilisation d'outils d'intelligence artificielle. Notre objectif était de développer un algorithme de détection de mitoses et son interface utilisateur, puis de tester leur apport pour l'évaluation du score mitotique et son impact sur la reproductibilité intra- et inter-observateur des pathologistes.

**Matériel et Méthodes :** deux bases de données ont été utilisées, une de 2259 mitoses labellisées à partir de 32 lames de carcinome mammaire infiltrant et celle du challenge MIDOG2021. L'algorithme comprend la détection des objets pouvant correspondre à des mitoses, puis la classification des objets (mitose/"non-mitose"). Sur l'interface développée, les mitoses retenues et les zones de plus forte densité mitotique sont annotées facilitant le comptage et la détermination du score. Pour évaluer cet outil, quatre pathologistes ont réalisé un compte mitotique sans puis avec son assistance, avec *wash-out* de 3 semaines, sur une surface de 2 mm<sup>2</sup> prédéfinie sur 50 lames de carcinome mammaire infiltrant.

**Résultats :** le coefficient de corrélation intra-classe (CCI) pour le compte mitotique des pathologistes était de 0,716 IC [0,531 - 0,833] sans et de 0,746 IC [0,543 - 0,859] avec assistance de l'algorithme. Entre les pathologistes, le kappa de Cohen pour la reproductibilité du score mitotique variait entre 0,339 et 0,752 sans et entre 0,496 et 0,890 avec assistance de l'algorithme. Le CCI pour le compte mitotique entre pathologistes et algorithme variait entre 0,449 IC [0,101 - 0,678] et 0,769 IC [0,627 - 0,862]. Le kappa de Cohen pour la reproductibilité du score mitotique entre chaque pathologiste et l'algorithme variait entre 0,424 et 0,687. Cinquante-deux pour cent des mitoses annotées l'ont été par au moins deux pathologistes et seuls 14% ont fait l'objet d'un consensus par les quatre observateurs. Les performances de l'algorithme pour reconnaître les mitoses détectées par les quatre pathologistes étaient comparables à celles des pathologistes entre eux.

**Discussion :** la reproductibilité des pathologistes dans l'évaluation du score mitotique sur une surface définie n'était pas significativement augmentée par l'utilisation de l'algorithme. Les performances de l'algorithme pour reconnaître les figures mitotiques étaient comparables à celles des pathologistes et pour trois d'entre eux, les comptes mitotiques et les scores correspondant étaient comparables à ceux prédits par l'algorithme.

**Conclusion et Perspectives :** les performances de l'algorithme de détection de mitoses développé permettent d'envisager son intégration en routine dans le flux de travail du pathologiste. Cette intégration sera précédée de l'évaluation de son apport sur lame entière concernant le compte mitotique, le score correspondant et le choix de la zone de comptage avec analyse de la reproductibilité des pathologistes.

**Mots clés :** carcinome mammaire infiltrant, pathologie numérique, lame entière numérisée, intelligence artificielle, apprentissage profond, mitoses, score mitotique, reproductibilité

## SUMMARY

**Introduction :** Elston-Ellis grade, which the mitotic score assessment is one of the elements, is a major prognostic factor for invasive breast carcinoma. The reproducibility of its evaluation remains discussed. In digital pathology, the underestimation of the reported mitotic score could be avoided by the use of artificial intelligence tools. Our goal was to develop a mitosis detection algorithm and its user interface, and then test their contribution to the evaluation of the mitotic score and its impact on the intra- and inter-observer reproducibility of the evaluation.

**Materials and Methods :** two databases were used, one of 2259 mitoses labeled from 32 slides of invasive breast carcinoma and that of the MIDOG2021 challenge. The algorithm includes the detection of objects that may correspond to mitoses, then the classification of objects (mitosis/non-mitosis). On the developed interface, retained mitoses and areas of higher mitotic density are annotated facilitating counting and scoring determination. To evaluate this tool, four pathologists performed a mitotic count without then with his assistance, with wash-out of 3 weeks, on 2 mm<sup>2</sup> predefined area on 50 slides of invasive breast carcinoma.

**Results :** the intraclass correlation coefficient (ICC) for the mitotic count of pathologists was 0.716 CI [0.531 - 0.833] without and 0.746 CI [0.543 - 0.859] with algorithm assistance. Between pathologists, Cohen's kappa for mitotic score reproducibility ranged from 0.339 to 0.752 without and from 0.496 to 0.890 with algorithm assistance. The ICC for the mitotic count between pathologists and algorithm ranged from 0.449 CI [0.101 - 0.678] to 0.769 CI [0.627 - 0.862]. Cohen's kappa for reproducibility of the mitotic score between each pathologist and the algorithm ranged from 0.424 to 0.687. Fifty-two percent of mitoses annotated were annotated by at least two pathologists and only 14% were agreed upon by all four observers. The performance of the algorithm to recognize mitosis detected by the four pathologists was comparable to that of the pathologists among themselves.

**Discussion :** the reproducibility of pathologists in the evaluation of the mitotic score on a defined area was not significantly increased by the use of the algorithm. The algorithm's performance in recognizing mitotic figures was comparable to that of pathologists and for three of them, the mitotic counts and corresponding scores were comparable to those predicted by the algorithm.

**Conclusion and Outlook :** the performance of the developed mitosis detection algorithm makes it possible to consider its routine integration into the pathologist's workflow. This integration will be preceded by the evaluation of its application on whole slide image concerning the mitotic count, the corresponding score and the choice of the counting area with analysis of the reproducibility of the pathologists

**Keywords :** invasive breast carcinoma, digital pathology, whole slide imaging, artificial intelligence, deep learning, mitosis, mitotic score, reproducibility