

République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة أبو بكر بلقايد – تلمسان
Université ABOUBEKR BELKAID – TLEMCEN
كلية علوم الطبيعة والحياة، وعلوم الأرض والكون
Faculté des Sciences de la Nature et de la Vie, et Sciences de la Terre et de l'Univers
Département de Biologie



MÉMOIRE

Présenté par

Khiter Farouk

En vue de l'obtention du

Diplôme de MASTER

En Biologie Moléculaire et Cellulaire

Thème

Évaluation de l'apport de l'intelligence artificielle et de l'analyse d'images numériques pour la standardisation de la quantification du Ki-67 dans le cancer du sein

Soutenu le 01/07/2026, devant le jury composé de :

Présidente
Examinatrice
Encadrante

Pr BAB HAMED Nabila
Pr Dennouni- Medjati Nouria
Pr Dali Youcef Sahi Majda

CHU Tlemcen
Univ Tlemcen
Univ Tlemcen

{قُلْ هَلْ يَسْتَوِي الَّذِينَ يَعْلَمُونَ وَالَّذِينَ لَا يَعْلَمُونَ} إِنَّمَا يَتَذَكَّرُ أُولُو الْأَلْبَابِ {

سورة الزمر - الآية 9

نحمد الله تعالى الذي منَّ علينا بنعمة العقل، وهدانا إلى سبيل العلم، وأعاننا على إتمام هذا العمل المتواضع، الذي هو ثمرة جهدٍ وسهرٍ، وتجسيدٌ لما تلقيناه خلال مسيرتنا الجامعية من معارف وخبرات
إن طلب العلم ليس مجرد تحصيلٍ للمعلومات، بل هو عبادةٌ ورسالة، ووسيلةٌ لبناء الذات وتنوير العقل،
والمساهمة في نهضة الأمم. به تعلو الهمم، وتبنى الحضارات، وتُقاس قيمة الإنسان
ولئن كان هذا البحث محطةً في طريقنا العلمي، فإنه يبقى خطوةً في رحلةٍ لا تنتهي، وبدايةً لأسئلةٍ أعمق، وآفاقٍ
أرحب. فالعلم بحرٌ لا ساحل له، وكلما ازددنا منه علماً، ازددنا إدراكاً بقدر الله في خلقه
نسأل الله العليّ القدير أن يجعل هذا العمل خالصاً لوجهه الكريم، وأن يبارك فيه، وأن ينفع به، وأن يرزقنا علماً
نافعاً، وعملاً متقبلاً، وهدىً مستقيماً

"رَبِّ اشْرَحْ لِي صَدْرِي وَيَسِّرْ لِي أَمْرِي وَاحْلُلْ عُقْدَةً مِّن لِّسَانِي يَفْقَهُوا قَوْلِي"

سورة طه - الآية 25-28

اللهم علِّمنا ما ينفعنا، وانفعنا بما علِّمتنا، وزدنا علماً، اللهم إنا نسألك الفهم والإخلاص والتوفيق في القول
والعمل

بسبحاتك لا علم لنا إلا ما علِّمتنا، إنك أنت العظيم الحكيم

آمين يا رب العالمين

ملخص:

مقدمة: يُعدّ مؤشر تكاثر الخلايا Ki-67 مؤشراً حيوياً رئيسياً في التصنيف الجزيئي لسرطانات الثدي وفي توجيه قرارات العلاج. مع ذلك، لا يزال قياسه البصري من قِبل أخصائيي علم الأمراض محدوداً بسبب التباين الموثق جيداً بين المشاهدين.

في هذا السياق، يُقدّم علم الأمراض الرقمي آفاقاً جديدة لتحسين توحيد ودقة تقييمه.

الأهداف: هدفت هذه الدراسة إلى تقييم مساهمة تحليل الصور مفتوح المصدر والذكاء الاصطناعي في توحيد قياس Ki-67، مقارنةً بالطريقة اليدوية المرجعية. كما هدفت إلى تحليل تأثير عمر المريض على نمط تكاثر الورم. المنهجية:

أجري التقييم التقني على 909 صورة تُغطي 303 حالات، مقارنةً بثلاث طرق لقياس Ki-67: العد اليدوي من قِبل الخبراء، وخوارزمية QuPath، وأداة التعلم العميق DeepLIIF. بالتوازي، أُجري تحليل سريري على 164 مريضاً من المجموعة لمقارنة مستوى تعبير Ki-67 وفقاً للعمر، حيث تمت مقارنة المرضى الذين تقل أعمارهم عن 45 عاماً بالمرضى الذين تبلغ أعمارهم 45 عاماً فأكثر. وقُيّم التوافق بين الطريقتين باستخدام معاملات الارتباط، وتحليل بلاند-ألتمان، ومعامل الارتباط داخل الفئة.

النتائج: تُظهر النتائج أن أداة DeepLIIF تُظهر ارتباطاً قوياً مع الطريقة اليدوية، متفوقةً على تلك الملاحظة مع QuPath ($r = 0.84$ مقابل 0.72). ولوحظ انحياز منهجي طفيف نحو المبالغة في التقدير مع النهج القائم على الذكاء الاصطناعي؛ ومع ذلك، ظل معدل إعادة التصنيف السريري حول عتبة القرار البالغة 20% محدوداً عند 6.6%، مما يدل على متانة تحليلية جيدة. وكانت قابلية التكرار الإجمالية ممتازة، بمعامل ارتباط داخل الفئة قدره 0.82، مصحوباً بتوفير مُقدّر بنسبة 90% في وقت المعالجة. سريرياً، كان لدى المرضى الذين تقل أعمارهم عن 45 عاماً نسبة أعلى من مؤشر $Ki-67 \geq 20\%$ مقارنةً بالمرضى الأكبر سناً (44.4% مقابل 36.1%)، إلا أن هذا الفرق لم يكن ذا دلالة إحصائية ($p = 0.590$).

الخلاصة: يبدو أن أدوات الذكاء الاصطناعي، ولا سيما تقنيات التعلم العميق، حلولٌ داعمةٌ فعّالة وسريعة وقابلة للتكرار بدرجة عالية لتقييم مؤشر Ki-67 في الممارسة التشخيصية الروتينية. في المجموعة المدروسة، لم يختلف مؤشر تكاثر الورم اختلافاً كبيراً باختلاف عمر المرضى، مما يشير إلى تجانس نسبي في نمط التكاثر بين المجموعتين العمريتين وقت التشخيص.

الكلمات المفتاحية: سرطان الثدي، Ki-67، الذكاء الاصطناعي، DeepLIIF، علم الأمراض الرقمي، التقييم.

Résumé :

Introduction: L'indice de prolifération Ki-67 constitue un biomarqueur majeur dans la stratification moléculaire des cancers du sein et dans l'orientation des décisions thérapeutiques. Toutefois, sa quantification visuelle par les pathologistes demeure limitée par une variabilité inter-observateur bien documentée.

Dans ce contexte, la pathologie numérique ouvre de nouvelles perspectives pour améliorer la standardisation et la reproductibilité de son évaluation.

Objectifs : Cette étude avait pour objectif d'évaluer l'apport de l'analyse d'images open-source et de l'intelligence artificielle dans la standardisation de la quantification du Ki-67, en comparaison avec la méthode manuelle de référence. Elle visait également à analyser l'influence de l'âge des patientes sur le profil de prolifération tumorale.

Méthodologie : L'évaluation technique a porté sur 909 images correspondant à 303 cas, en comparant trois approches de quantification du Ki-67 : le comptage manuel expert, l'algorithme QuPath et l'outil de deep learning DeepLIIF. Parallèlement, une analyse clinique a été menée sur 164 patientes issues de la cohorte afin de comparer l'expression du Ki-67 selon l'âge, en opposant

les patientes de moins de 45 ans à celles âgées de 45 ans ou plus. La concordance entre les méthodes a été évaluée à l'aide des coefficients de corrélation, de l'analyse de Bland-Altman et du coefficient de corrélation intraclasse.

Résultats : Les résultats montrent que l'outil DeepLIIF présente une forte corrélation avec la méthode manuelle, supérieure à celle observée avec QuPath ($r = 0,84$ contre $0,72$). Un léger biais systématique de surestimation a été observé avec l'approche basée sur l'intelligence artificielle ; toutefois, le taux de reclassification clinique autour du seuil décisionnel de 20 % est resté limité à 6,6 %, témoignant d'une bonne robustesse analytique. La reproductibilité globale était excellente, avec un coefficient de corrélation intraclasse de 0,82, associée à un gain de temps de traitement estimé à 90 %. Sur le plan clinique, les patientes âgées de moins de 45 ans présentaient une fréquence plus élevée de $Ki-67 \geq 20\%$ que les patientes plus âgées (44,4 % versus 36,1 %), mais cette différence n'était pas statistiquement significative ($p = 0,590$).

Conclusion : Les outils d'intelligence artificielle, et en particulier les approches de deep learning, apparaissent comme des solutions d'assistance viables, rapides et hautement reproductibles pour l'évaluation du Ki-67 en routine diagnostique. Dans la cohorte étudiée, l'indice de prolifération tumorale ne variait pas significativement selon l'âge des patientes, suggérant une relative homogénéité du profil prolifératif entre les deux groupes d'âge au moment du diagnostic.

Mots-clés : Cancer du sein, Ki-67, Intelligence Artificielle, DeepLIIF, Pathologie numérique, Standardisation.

Abstract:

Introduction: The Ki-67 proliferation index is a major biomarker in the molecular stratification of breast cancers and in guiding therapeutic decisions. However, its visual quantification by pathologists remains limited by well-documented inter-observer variability.

In this context, digital pathology opens new perspectives for improving the standardization and reproducibility of its assessment.

Objectives: This study aimed to evaluate the contribution of open-source image analysis and artificial intelligence to the standardization of Ki-67 quantification, compared with the reference manual method. It also aimed to analyze the influence of patient age on the tumor proliferation profile.

Methodology: The technical evaluation included 909 images corresponding to 303 cases, comparing three approaches to Ki-67 quantification: expert manual counting, the QuPath algorithm, and the DeepLIIF deep learning tool. Concurrently, a clinical analysis was conducted on 164 patients from the cohort to compare Ki-67 expression according to age, contrasting patients under 45 years of age with those aged 45 years or older. The agreement between the methods was assessed using correlation coefficients, Bland-Altman analysis, and the intraclass correlation coefficient.

Results: The results show that the DeepLIIF tool exhibits a strong correlation with the manual method, superior to that observed with QuPath ($r = 0.84$ vs. 0.72). A slight systematic bias of overestimation was observed with the AI-based approach; however, the clinical reclassification rate

around the 20% decision threshold remained limited to 6.6%, demonstrating good analytical robustness. Overall reproducibility was excellent, with an intraclass correlation coefficient of 0.82, associated with an estimated 90% reduction in processing time. Clinically, patients under 45 years of age had a higher frequency of Ki-67 $\geq 20\%$ than older patients (44.4% versus 36.1%), but this difference was not statistically significant ($p = 0.590$).

Conclusion: Artificial intelligence tools, and in particular deep learning approaches, appear to be viable, rapid, and highly reproducible support solutions for Ki-67 assessment in routine diagnostics. In the studied cohort, the tumor proliferation index did not vary significantly according to patient age, suggesting relative homogeneity of the proliferative profile between the two age groups at the time of diagnosis.

Keywords: Breast cancer, Ki-67, Artificial Intelligence, DeepLIIF, Digital pathology, Standardization.

Remerciements

Je remercie Dieu tout puissant de m'avoir accordé la force, la patience et la détermination nécessaires à l'accomplissement de ce travail de fin d'études.

*Je tiens, en premier lieu, à exprimer ma profonde reconnaissance à **Mme la Professeure BAB HAMED Nabila**, du CHU de Tlemcen, qui me fait l'honneur de présider ce jury. Je la remercie sincèrement pour son aide précieuse, sa disponibilité et son encadrement. Son initiation à la lecture et à l'interprétation des lames histopathologiques, ainsi que la mise à disposition du matériel nécessaire, ont été des étapes déterminantes dans la réalisation de ce travail.*

*Mes remerciements s'adressent également à **Mme la Professeure DALI YUCEF KAZI Lamia**, de l'Université de Tlemcen, pour l'honneur qu'elle me fait d'examiner ce travail. Je la remercie pour le temps qu'elle a consacré à la lecture de ce mémoire et pour l'intérêt qu'elle porte à ce sujet.*

*Je tiens à témoigner toute ma gratitude à **Mme la Professeure Dali Youcef Sahi Majda**, Professeure à l'Université de Tlemcen, qui a accepté d'encadrer ce mémoire. Ses précieux conseils, sa disponibilité et sa rigueur scientifique ont grandement contribué à l'enrichissement de ce travail. Je la remercie pour la confiance qu'elle m'a accordée et pour son accompagnement tout au long de cette recherche.*

*Je tiens également à témoigner toute ma gratitude au **Dr BELARBI Omar**, pour son soutien et sa contribution scientifique essentielle. Sa générosité dans la mise à disposition des lames et des données populationnelles a permis de constituer la base solide de cette étude.*

Je remercie l'ensemble des enseignants qui ont participé à ma formation tout au long de mon parcours universitaire.

Enfin, j'adresse mes remerciements à l'ensemble du personnel du service d'Anatomie et Cytologie Pathologiques du CHU de Tlemcen, de l'unité d'oncologie de l'hôpital de Ghazaouet, ainsi qu'à toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce mémoire.

Dédicaces

Je dédie ce travail à ceux qui comptent le plus à mes yeux.

À mon père, Djilali, pour son exemple..., sa sagesse... et son soutien silencieux mais constant.

À ma mère, Karima, pour son amour..., sa bienveillance... et ses prières qui m'ont accompagné tout au long de ce chemin.

À mes frères, Abdelkader et Nabila, pour leur présence..., leur soutien et leurs encouragements.

À mes tantes, Hanifa, que Dieu lui accorde guérison et santé et Rahima, pour leurs prières.

À mes oncles maternels, Amine et Hassan, pour leur bienveillance... et leur soutien.

À ma grand-mère, Rahma, pour sa tendresse et ses prières.

À mes amis, Nassim, Mouad, Abdelillah et Elhadi, pour leur amitié sincère et les moments partagés.

Enfin, à toutes les personnes qui ont croisé ma route et m'ont apporté quelque chose, merci.

رسالة شكر وتقدير

أكتب إليكم اليوم بامتنان عميق لأشركم من صميم قلبي. لقد كان توجيهكم لي طوال فترة إعداد رسالة الماجستير هذه أكثر بكثير من مجرد دعم أكاديمي، بل كان مصدر إلهام وتحفيز حقيقيين.

لقد ساهم صبركم ولطفكم ومعاييركم الرفيعة في نموي، ليس فقط كطالب، بل كباحث مستقبلي أيضًا. علمتني أن أنظر إلى ما وراء المظاهر، وأن أسعى للدقة، وألا أكتفي أبدًا بالحلول التقريبية.

سأعثر بذكرى حواراتنا، وندائكم القيمة، وتواجدكم الدائم، حتى في أوقات انشغالكم. إن شغفكم بالبحث وتفانيكم لطلابكم صفتان أثرتا فيّ تأثيرًا بالغًا.

في هذا اليوم، وأنا أختتم هذه المرحلة المصممة من حياتي، أود أن أقول لكم ببساطة وصدق: شكرًا لكم.

تفضلوا بقبول خالص امتناني وتقديري، أستاذي.

خيتير فاروق

Liste des Abréviations

Abréviation	Signification
CNN	Convolutional Neural Network (Réseau de neurones convolutif)
CV	Coefficient de Variation
DAB	3,3'-Diaminobenzidine
DIA	Digital Image Analysis (Analyse d'images numériques)
ER	Récepteurs aux Œstrogènes
FFPE	Formalin-Fixed Paraffin-Embedded (Fixé au formol et inclus en paraffine)
FN	Faux Négatifs
FP	Faux Positifs
GPU	Graphics Processing Unit (Unité de traitement graphique)
H&E	Hématoxyline-Éosine
HER2	Human Epidermal Growth Factor Receptor 2
HTA	Hypertension Artérielle
IA	Intelligence Artificielle
ICC	Intraclass Correlation Coefficient (Coefficient de corrélation intraclasse)
IHC	Immunohistochimie
PR	Récepteurs à la Progestérone
RDMS	Residual Multi-Scale (Blocs résiduels multi-échelles)
RE	Récepteurs aux Œstrogènes
RH	Récepteurs Hormonaux
ROI	Region Of Interest (Région d'intérêt)
RP	Récepteurs à la Progestérone
SBR	Scarff-Bloom-Richardson (Grade histopronostique)
SVM	Support Vector Machine
TIL	Tumor-Infiltrating Lymphocytes (Lymphocytes infiltrant la tumeur)
VN	Vrais Négatifs
VP	Vrais Positifs
WSI	Whole Slide Imaging (Imagerie de lames entières)

Liste des figures

Figure 01: Grille des sous-types intrinsèques du cancer du sein (histologie).....	1
Figure 02: Expression de Ki-67 au cours du cycle cellulaire.	Erreur ! Signet non défini.
Figure 03: Évaluation visuelle du Ki-67 selon l'échelle Eye-5	1
Figure 04: Image microscopique avec coloration immunohistochimique du Ki-67.....	8
Figure 05: Cycle de vie d'un projet de science des données appliqué à la pathologie numérique ...	11
Figure 06 Diagramme d'intégration des disciplines de l'intelligence artificielle en pathologie numérique.....	12
Figure 07: Processus d'acquisition et de prétraitement des données pour l'IA en pathologie mammaire.....	14
Figure 08: Systèmes de numérisation automatisés (scanners de lames) en routine clinique	15
Figure 09: Flux de travail standardisé de la pathologie numérique, de la numérisation au diagnostic	16
Figure 10 : Analyse automatisée d'une biopsie mammaire numérisée par un modèle d'apprentissage profond.....	17
Figure 11 : Neuf pathologistes ont interprété la carte thermique Ki-67LI du cancer du sein via VA, MC, SRC et AI.....	19
Figure 12 Exemple comparatif de l'indice de positivité dans des cas représentatifs d'immunohistochemie Ki-67, selon les méthodes 3DHistech, Roche, aetherAI et le comptage manuel.....	20
Figure 13 : Répartition par âge.	29
Figure 14a : Nuage de points montrant la corrélation entre les méthodes manuelle et DeepLIIF pour la quantification du Ki-67 (n=264 patients avec données complètes). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité (y=x).	49
Figure 14b : Nuage de points montrant la corrélation entre les méthodes manuelle et QuPath pour la quantification du Ki-67 (n=264). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité (y=x).	50
Figure 14c : Nuage de points montrant la corrélation entre les méthodes QuPath et DeepLIIF pour la quantification du Ki-67 (n=303). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité (y=x).	50
Figure 15a : Graphique de Bland-Altman comparant les méthodes manuelle et QuPath pour la quantification du Ki-67 (n=264). La ligne continue rouge représente le biais moyen (+8,2 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm$ 1,96 $\times$ écart-type), soit un intervalle de [-19,7 ; +36,1].	52
Figure 15b : Graphique de Bland-Altman comparant les méthodes manuelle et DeepLIIF pour la quantification du Ki-67 (n=264). La ligne continue rouge représente le biais moyen (-31,7 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm$ 1,96 $\times$ écart-type), soit un intervalle de [-105,6 ; +42,3].	52
Figure 15c : Graphique de Bland-Altman comparant les méthodes QuPath et DeepLIIF pour la quantification du Ki-67 (n=303). La ligne continue rouge représente le biais moyen (-6,7 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm$ 1,96 $\times$ écart-type), soit un intervalle de [-28,3 ; +14,9].	53
Figure 16 : Distribution de l'hétérogénéité intratumorale basée sur le coefficient de variation (CV) du Ki-67 mesuré par DeepLIIF sur trois champs par patient (n=264). Les patients sont classés en	

trois catégories : homogène ($CV < 15\%$, 42%), modéré ($15\% \leq CV < 30\%$, 41%) et forte hétérogénéité ($CV \geq 30\%$, 17%).	54
Figure 17 : Histogramme de reclassification clinique au seuil de 20% comparant les méthodes manuelle, DeepLIIF et QuPath (n=264). Les barres vertes représentent les patients classés en « Bas Risque » ($< 20\%$) ; les barres rouges représentent les patients classés en « Haut Risque » ($\geq 20\%$).	55
Figure 18 : Temps d'analyse par patient.	55
Figure 19 : Diagramme en boîtes à moustaches (boxplots) montrant la distribution du biais (Manuel – DeepLIIF) par année de prélèvement (n=264 ; 2023 : n=92 ; 2024 : n=47 ; 2025 : n=164). La ligne rouge en pointillés représente l'absence de biais ($y=0$). Le test ANOVA ne montre pas de différence significative entre les années ($p=0,18$).	56
Figure 20 : Comparaison des masques de segmentation obtenus par QuPath et DeepLIIF sur une même image de marquage immunohistochimique du Ki-67. L'image originale (a) montre le marquage nucléaire ; le masque QuPath (b) présente une segmentation parfois trop large incluant des structures non cellulaires ; le masque DeepLIIF (c) montre des contours nucléaires plus précis et une meilleure délimitation des cellules tumorales.	60

Liste des tableaux

Tableau 01 : Les principaux sous-types moléculaires du cancer du sein.	1
Tableau 02 : Causes de la variabilité du Ki-67 manuel et conséquences cliniques.	2
Tableau 03 : Synthèse des indicateurs de concordance inter-observateurs.	22
Tableau 04 : Stratification des risques et recommandations cliniques.	25
Tableau 05 : Ce tableau compare les facteurs de risque et les comorbidités entre les deux tranches d'âge.	31
Tableau 06 : Comparaison du profil gynéco-obstétrical et des antécédents familiaux selon l'âge.	31
Tableau 07 : Comparaison des caractéristiques tumorales et moléculaires selon l'âge.	33
Tableau 08 : Profil des biomarqueurs moléculaires selon l'âge des patientes (n = 164).	33
Tableau 09 : Coefficients de corrélation de Pearson.	49
Tableau 10 : Analyse de concordance : différences moyennes et Bland-Altman.	51
Tableau 11 : Analyse des seuils pronostiques.	54
Tableau 12 : Effet des années de prélèvement sur la concordance.	56

Sommaire

Chapitre 01

I. Introduction et État de l'Art :	1
1. Contexte clinique :	1
• 1.1. Épidémiologie et classification moléculaire du cancer du sein :	1
• 1.2. Le biomarqueur Ki-67 : expression et signification pronostique :	2
2. Problématique :	1
• 2.1. Limites de la quantification manuelle du Ki-67 :	1
• 2.2. Sources de variabilité inter-observateur :	2
3. Objectif de l'étude :	2
II. Méthodologie Technique :	10
1. Fondements et cycle de vie des modèles d'intelligence artificielle :	10
• 1.1. Définition et concepts clés :	10
• 1.2. Cycle de développement d'un projet de science des données appliquées :	10
2. Disciplines constitutives de l'ingénierie de l'IA appliquée :	11
3. Processus d'acquisition des données histopathologiques :	12
4. Défis et contraintes de l'acquisition en environnement médical :	13
5. Numérisation des lames entières (<i>Whole Slide Imaging</i> - WSI) :	14
6. L'imagerie de lames entières (<i>Whole Slide Imaging</i> - WSI) : Applications cliniques et synergie avec l'IA :	15
• 6.1. Intégration du flux WSI en anatomopathologie :	15
• 6.2. Optimisation du diagnostic et télépathologie :	15
• 6.3. Interfaçage de l'imagerie numérique et des algorithmes d'apprentissage profond :	16
7. Limites techniques de la WSI et solutions d'infrastructure :	17
8. Prétraitement et fondements de l'immunohistochimie (IHC) :	17
9. Choix des outils d'analyse : QuPath et DeepLIF, deux solutions open-source pour la quantification du Ki-67 :	18
III. Évaluation de la Performance et Validation Clinique :	18
1. Définition de la référence absolue (Gold Standard) :	18
2. Analyse de concordance et optimisation du flux de travail :	19
3. Discussion et gestion de l'hétérogénéité tumorale :	19

4. Conclusion :	21
5. Métriques de concordance et reproductibilité :	21
6. Modèles pronostiques multi-modaux et impact sur la survie :	22
IV. Enjeux cliniques, éthiques et responsabilité médicale :	23
1. Évolution du rôle du pathologiste et intégration au workflow :	23
2. Analyse économique de la responsabilité civile médicale :	23
3. Risque du partage des coûts :	23
4. Analyse des coûts opérationnels et infrastructures :	24
V. Perspectives multi-biomarqueurs et standardisation internationale :	24
1. Au-delà du Ki-67 : vers une analyse multiparamétrique :	24
2. Standardisation internationale et stratification thérapeutique :	24

Chapitre 02

Étude 1: ANALYSE COMPARATIVE DES CARACTÉRISTIQUES CLINICO-PATHOLOGIQUES SELON L'ÂGE	26
Matériel et Méthodes	28
1. Description de la cohorte :	29
Résultats et interprétation	30
2. Étude comparative selon les tranches d'âge :	31
• 2.1. Comparaison des caractéristiques cliniques et métaboliques :	31
• 2.2. Comparaison des caractéristiques tumorales et histologiques :	32
3. Focus sur le Ki-67 :	33
• 3.1. Analyse de l'expression du Ki-67 selon l'âge :	33
• 3.2. Signification pronostique du Ki-67 chez les patientes jeunes : Erreur ! Signet non défini.	
Discussion	37
1. Synthèse et intérêt de l'approche analytique :	38
Conclusion	40
1. Enjeux de reproductibilité et transition vers l'analyse automatisée :	41
ÉTUDE 2 : COMPARAISON DES MÉTHODES DE QUANTIFICATION DU KI-67 (MANUEL, QuPath, DeepLIIF)	42
Matériel et Méthodes	43
1.Introduction de l'étude comparative :	44
Résultats et interprétation	47

1. Caractéristiques de la cohorte :	48
2. Coefficients de corrélation de Pearson :	48
3. Analyse de concordance : différences moyennes et Bland-Altman :	50
3.1. Comparaison Manuel vs QuPath :	50
3.2. Comparaison Manuel vs DeepLIIF :	51
3.3. Comparaison QuPath vs DeepLIIF :	51
4. Coefficient de corrélation intraclass (ICC) :	53
5. Reproductibilité inter-champs et hétérogénéité intratumorale :	53
6. Analyse des seuils pronostiques :	54
7. Temps d'analyse :	55
8. Effet des années de prélèvement sur la concordance :	55
9. Analyse des valeurs « absent » :	56
Discussion.....	58
1. Comparaison avec la littérature :	59
1.1.Confrontation aux recommandations du Groupe de Travail International sur le Ki-67 (IKWG) :	59
2. Analyse des cas extrêmes de discordance :	60
2.1.Analyse approfondie de la saturation à 100 % :	60
2.2. Analyse approfondie du biais négatif (−31,7 points) :	62
3. Avantages et inconvénients :	63
4. Limites de l'étude :	63
5. Faisabilité et analyse coût-bénéfice pour le CHU de Tlemcen :	63
6. Perspectives futures :	64
Discussion Générale	65
Lien entre les résultats cliniques et l'apport de l'IA :	66
1. Rappel du problème clinique identifié dans l'Étude 1:	66
2. Synthèse des résultats techniques de l'Étude 2 :	66
3. Comment l'IA peut aider à résoudre le problème clinique :	66
Conclusion Générale.....	67

Chapitre 01 :

Introduction et État de l'Art

I. Introduction et État de l'Art :

1. Contexte clinique :

- **1.1. Épidémiologie et classification moléculaire du cancer du sein :**

Le cancer du sein constitue l'une des néoplasies malignes les plus fréquemment diagnostiquées chez la femme à l'échelle mondiale, représentant environ 16 % de l'ensemble des cancers féminins. Bien que son incidence demeure en augmentation, la mortalité globale enregistre une diminution progressive, attribuable d'une part à la précocité du diagnostic et, d'autre part, à l'optimisation des protocoles thérapeutiques.

Introduction

Le consensus international d'experts de Saint-Gall (2011), affiné lors de la conférence de 2013, a établi une classification moléculaire permettant d'orienter la prise en charge thérapeutique selon des profils intrinsèques distincts (**Figures 01, 02 ; Tableau 01**) (Inic et al., 2014).

Tableau 01: Les principaux sous-types moléculaires du cancer du sein (<https://www.cbcn.ca>, s. d.).

Sous-type	Statut des récepteurs hormonaux et de HER2 (usuel)
Luminal A	ER+ et PR+ ; HER2- ; Faible prolifération
Luminal B	ER+ et PR+ (faible) ou PR- ; HER2+ ou Forte prolifération
Triple négatif	ER- et PR- ; HER2-
Surexpression de HER2	ER- et PR- ; HER2+

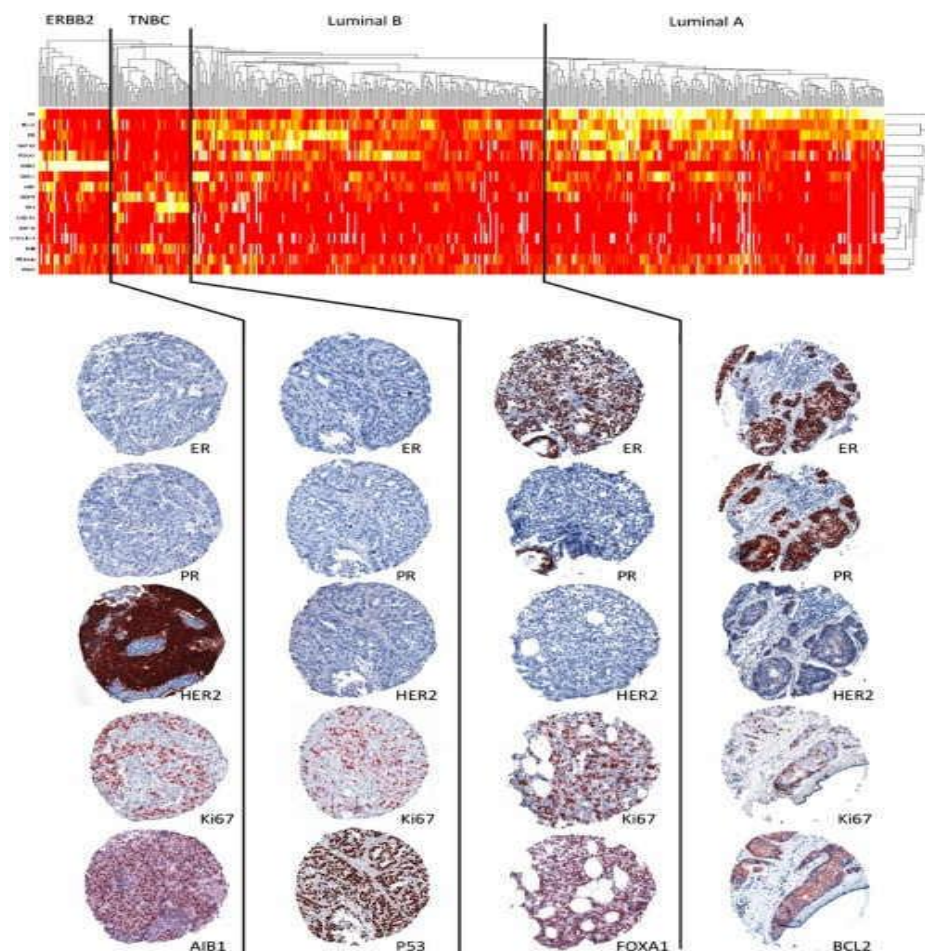


Figure 01: Grille des sous-types intrinsèques du cancer du sein (histologie) (Braun et al., 2013).

Introduction

- 1.2. Le biomarqueur Ki-67 : expression et signification pronostique :

L'antigène Ki-67 est une protéine nucléaire non histone exprimée exclusivement par les cellules en phase active du cycle cellulaire (phases G1 tardive, S, G2 et M), avec un pic d'expression lors de la transition G2-M, suivie d'une dégradation rapide après la mitose. L'indice de marquage Ki-67, déterminé par immunohistochimie (IHC), représente un facteur pronostique majeur. Dans les tumeurs de phénotype ER+/HER2-, la quantification du Ki-67 sert de biomarqueur de substitution (*surrogate marker*) pour discriminer les tumeurs de type Luminal A (faiblement prolifératives) des formes Luminales B (hautement prolifératives), impactant directement la décision d'adjoindre ou non une chimiothérapie adjuvante (Viale et al., 2019).

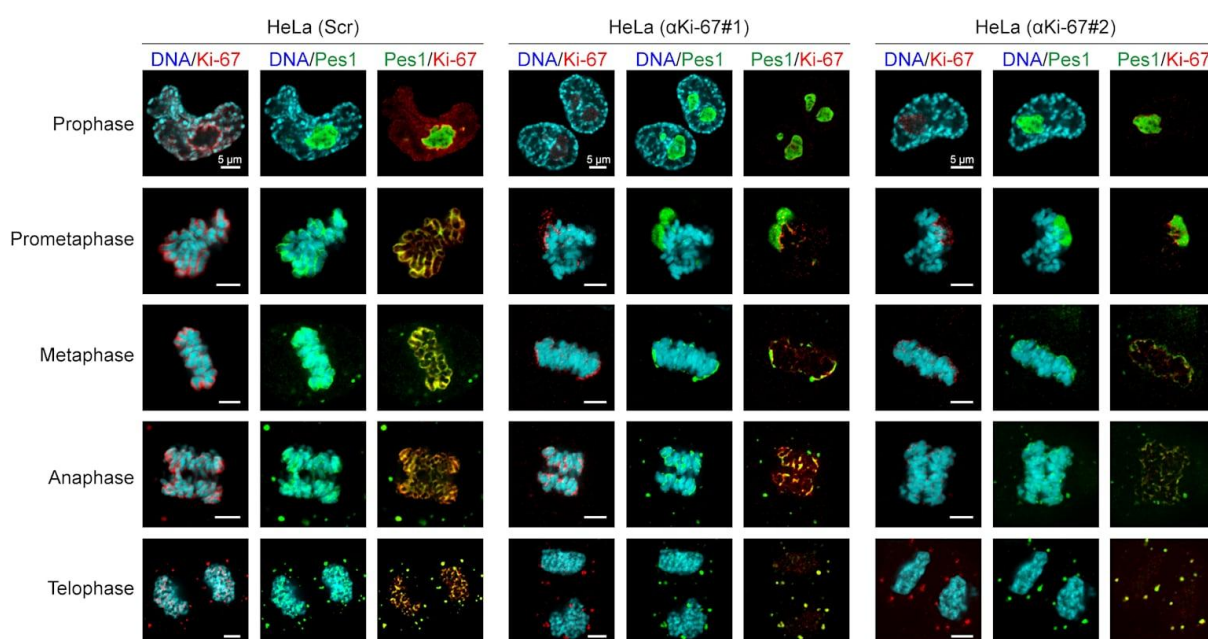


Figure 02: Expression de Ki-67 au cours du cycle cellulaire (Sobecki et al., 2016).

Un indice de prolifération Ki-67 élevé est fréquemment associé à des facteurs clinico-pathologiques défavorables, notamment un âge jeune au diagnostic, une taille tumorale invasive supérieure, un envahissement ganglionnaire axillaire, la négativation des récepteurs hormonaux, une mutation ou surexpression de p53, et la positivité du statut HER2. Sur le plan pronostique, un score Ki-67 supérieur au seuil clinique de 20 % est corrélé à une survie sans récurrence plus courte et à un risque élevé de rechute précoce (dans un délai inférieur à 2 ans) (Inic et al., 2014).

2. Problématique :

- **2.1. Limites de la quantification manuelle du Ki-67 :**

Malgré son utilité clinique, la méthode standard de quantification visuelle et manuelle du Ki-67 par le pathologiste fait l'objet de fortes critiques en raison d'une variabilité inter- et intra-observateur prononcée et d'un manque de reproductibilité.

L'analyse visuelle sur microscope optique ou sur écran souffre de plusieurs biais méthodologiques intrinsèques (**Figure 03**) :

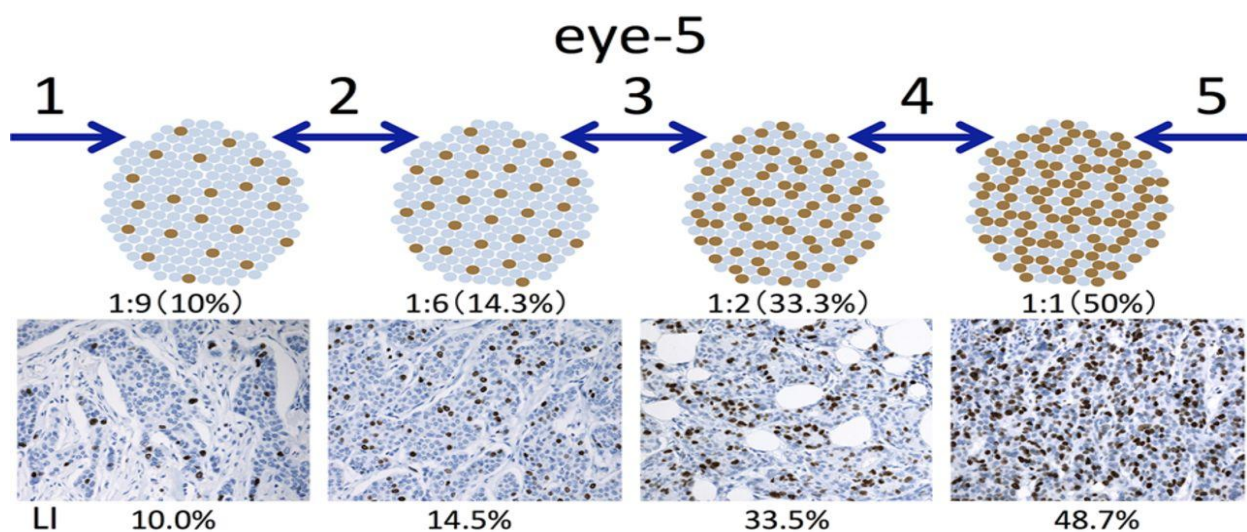


Figure 03: Évaluation visuelle du Ki-67 selon l'échelle Eye-5 (Hida et al., 2015).

Les limites de cette évaluation visuelle ont été largement documentées dans la littérature. D'après une revue des méthodes de comptage cellulaire, plusieurs biais méthodologiques intrinsèques peuvent être identifiés.

- La perception et la discrimination cellulaires :** La distinction visuelle entre les noyaux des cellules tumorales invasives, les cellules inflammatoires du stroma (lymphocytes), les cellules endothéliales et les débris cellulaires s'avère complexe. De plus, l'appréciation du seuil d'intensité à partir duquel un noyau est considéré comme positivement marqué reste subjective.
- L'hétérogénéité intratumorale et la sélection des champs :** La distribution de l'expression du Ki-67 au sein d'une même tumeur est souvent hétérogène. Le choix de la zone de comptage qu'il s'agisse d'une moyenne globale ou d'une focalisation sur la zone de prolifération maximale (*hotspot*) induit des variations significatives du score final.
- L'évaluation des marquages de faible intensité :** La variabilité de l'intensité de la coloration brune (réaction au DAB) pose des difficultés de standardisation. Les noyaux faiblement teintés sont alternativement classés comme positifs ou négatifs selon l'examineur, ce qui génère des discordances majeures autour du seuil décisionnel de 20 %.
- Les biais d'échantillonnage stochastique :** Compte tenu des contraintes de temps en routine clinique, le décompte manuel est souvent restreint à un nombre limité de cellules (généralement 100 à 500 cellules), augmentant l'effet des variables stochastiques par rapport à la charge tumorale globale.

Introduction

- e. **Les variations techniques pré-analytiques** : Les fluctuations liées aux conditions de fixation tissulaire, d'inclusion en paraffine et d'artefacts de coloration modifient l'aspect morphologique des coupes, entraînant des sous- ou des sur estimations du score (**Tableau 02**) (*Numération de cellules automatisée Numération de cellules automatisée - Un examen des méthodes de numération de cellules, s. d.*).

- **2.2. Sources de variabilité inter-observateur :**

Tableau 02: Causes de la variabilité du Ki-67 manuel et conséquences cliniques (**élaboration propre d'après les recommandations et études sur le Ki-67**).

Source de variabilité	Problème sous-jacent	Conséquence sur le score Ki-67
Reconnaissance visuelle	Confusion entre cellules tumorales, stroma et débris	Forte variabilité inter-observateur
Sélection de la zone	Hétérogénéité tumorale, choix arbitraire du <i>hotspot</i>	Faible reproductibilité de la mesure
Seuil de positivité	Interprétation subjective des noyaux faiblement marqués	Discordance autour des seuils cliniques (20 %)
Taille de l'échantillon	Nombre restreint de cellules comptées manuellement	Instabilité statistique du score
Variations analytiques	Artefacts de technique, intensité variable du DAB	Sous- ou surestimation de l'indice

3. Objectif de l'étude :

Pour pallier les limites de l'évaluation visuelle, l'intégration de solutions de pathologie numérique fondées sur l'intelligence artificielle (IA) offre de nouvelles perspectives. Le développement d'algorithmes d'apprentissage profond (*Deep Learning*), notamment les réseaux de neurones convolutifs (CNN), permet d'automatiser et de standardiser l'analyse d'images de lames entières (*Whole Slide Images - WSI*).

Introduction

L'objectif principal de cette étude est d'évaluer l'apport et l'efficacité des architectures d'IA (telles que U-Net, PathoNet et Kpi-Net) pour la segmentation automatisée des noyaux cellulaires et la classification binaire des cellules tumorales (positives marquées au DAB vs négatives contre-colorées à l'hématoxyline). Il s'agit de démontrer comment ces modèles standardisent le calcul de l'indice de prolifération, réduisent les biais d'interprétation et optimisent la reproductibilité par rapport au standard de référence (*gold standard*) établi par un comité d'experts.

Bien que le Groupe de Travail International sur le Ki-67 (IKWG) recommande préférentiellement une évaluation globale pondérée (average score) pour maximiser la reproductibilité inter-observateur en routine visuelle, l'évaluation ciblée des zones de forte densité proliférative (hotspots) demeure une alternative pronostique majeure, largement validée dans la littérature pour les cancers du sein de stade agressif. Dans le cadre de ce travail axé sur la pathologie numérique, le choix s'est porté sur l'analyse automatisée des hotspots afin de tester la capacité des algorithmes (QuPath et DeepLIIF) à discriminer précisément les cellules tumorales positives dans les zones de forte confrontation cellulaire et de chevauchement nucléaire, là où la variabilité inter-pathologiste manuelle est cliniquement la plus critique.

En intégrant ces outils d'aide à la décision en routine histopathologique, l'objectif est de sécuriser la stratification pronostique des patientes atteintes de cancer du sein luminal et de fiabiliser les indications thérapeutiques **(Figure 04) (Fatin & Patwory, 2026).**

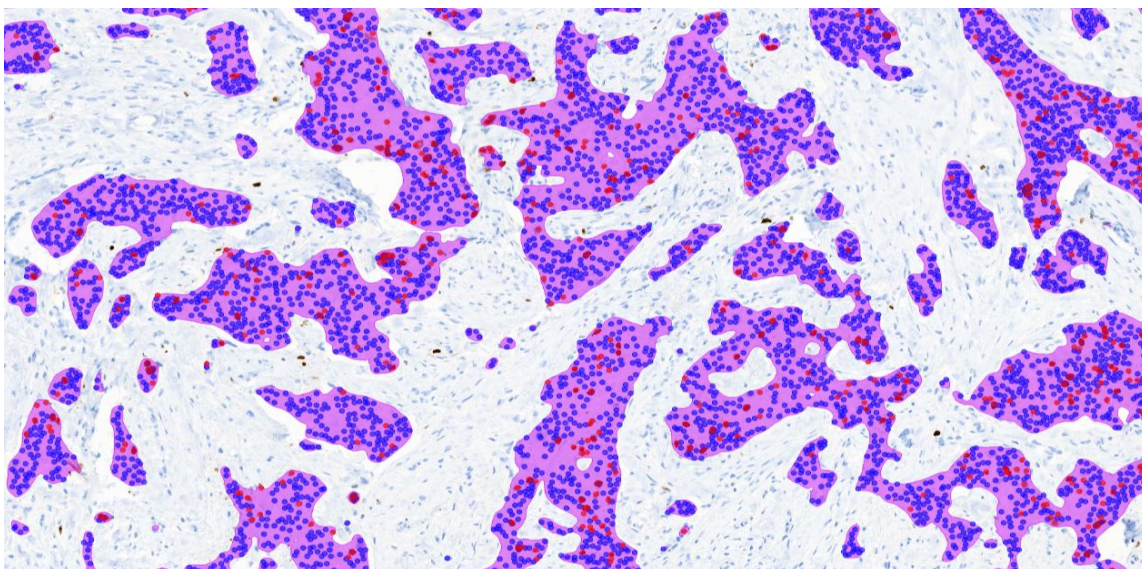


Figure 04: Image microscopique avec coloration immunohistochimique du Ki-67 (Cerba Research Case Study, s. d.).

Synthèse bibliographique

II. Méthodologie Technique :

1. Fondements et cycle de vie des modèles d'intelligence artificielle :

- **1.1. Définition et concepts clés :**

L'intelligence artificielle (IA) regroupe l'ensemble des théories, disciplines et techniques permettant de programmer des systèmes informatiques capables de modéliser des lois à partir de jeux de données et d'y détecter des motifs complexes. Les architectures avancées d'apprentissage automatique (*Machine Learning*) et de réseaux de neurones artificiels visent à automatiser des tâches cognitives jusqu'alors limitées à l'expertise humaine, telles que la reconnaissance visuelle et l'aide à la décision clinique.

Dans le cadre de l'analyse d'images histopathologiques, l'apprentissage profond (*Deep Learning*) s'impose comme la sous-discipline fondamentale. Contrairement aux approches algorithmiques traditionnelles, ces modèles s'appuient sur des réseaux de neurones convolutifs profonds capables d'extraire de manière autonome et hiérarchique les caractéristiques morphologiques pertinentes (structures nucléaires, variations chromatiques) sans extraction manuelle préalable de descripteurs.

- **1.2. Cycle de développement d'un projet de science des données appliquées :**

La conception et le déploiement d'une solution d'apprentissage profond suivent un processus itératif standardisé, articulé autour de cinq phases majeures (**Figure 05**) :

- L'acquisition et l'organisation des données :** Cette phase initiale consiste à collecter les images numériques à haute résolution issues de la numérisation des lames entières (*Whole Slide Imaging* - WSI). Les fichiers sont ensuite triés, standardisés et segmentés en régions d'intérêt (ROI) lisibles par les architectures algorithmiques. C'est également à cette étape que s'organise le protocole d'annotation par les pathologistes experts.
- Le prétraitement et l'entraînement :** Une fois les images filtrées, le modèle d'IA applique des transformations mathématiques pour harmoniser les variations de coloration. Les algorithmes exploitent ensuite les données d'entraînement pour ajuster de manière itérative leurs poids synaptiques à l'aide de fonctions de perte spécifiques, jusqu'à la reconnaissance reproductible des motifs cellulaires.

- c. **L'inférence des résultats** : Le modèle entraîné est soumis à de nouvelles données non annotées afin de prédire des sorties spécifiques (par exemple, la classification binaire entre noyaux tumoraux Ki-67 positifs et négatifs). Les résultats générés déterminent les scores quantitatifs exploitables pour la prise de décision thérapeutique.
- d. **L'optimisation et le réglage des hyperparamètres** : Lorsque les prédictions divergent de la vérité terrain (*ground truth*), l'algorithme évalue l'erreur induite. Les règles de rétropropagation du gradient permettent d'ajuster les hyperparamètres ou de modifier localement l'architecture du réseau (profondeur, taux d'apprentissage) afin de minimiser la fonction de perte.
- e. **L'évaluation des performances** : La phase finale consiste à valider la robustesse de l'architecture sur un jeu de données de test indépendant. La technologie synthétise les métriques statistiques afin de s'assurer de l'absence de surapprentissage (*overfitting*) et de valider la transférabilité clinique du modèle.

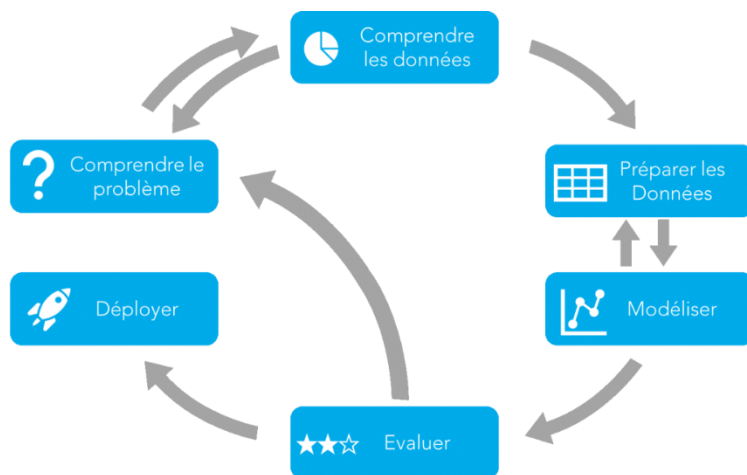


Figure 05: Cycle de vie d'un projet de science des données appliqué à la pathologie numérique (Charton, 2020).

2. Disciplines constitutives de l'ingénierie de l'IA appliquée :

La distinction entre les différents sous-domaines de l'intelligence artificielle s'établit selon le niveau d'abstraction et la spécificité des architectures algorithmiques mises en œuvre :

- **Apprentissage automatique (*Machine Learning*)** : Branche de l'informatique et des mathématiques appliquées visant à développer des algorithmes capables d'identifier des motifs complexes et de réaliser des prédictions à partir de données structurées, sans recourir à une programmation explicite des règles métiers.

Synthèse bibliographique

- **Apprentissage profond (*Deep Learning*)** : Sous-ensemble de l'apprentissage automatique fondé sur des réseaux de neurones artificiels multicouches. Il est capable d'extraire de manière autonome et hiérarchique des représentations abstraites à partir de données non structurées, éliminant ainsi l'étape d'extraction manuelle des descripteurs morphologiques.
- **Réseaux de neurones convolutifs (CNN)** : Architectures mathématiques d'apprentissage profond particulièrement adaptées au traitement de matrices bidimensionnelles. Ils simulent le fonctionnement du cortex visuel par des opérations de convolution locale pour assurer la segmentation et la classification.
- **Vision par ordinateur (*Computer Vision*)** : Domaine interdisciplinaire centré sur la capacité des systèmes informatiques à extraire, traiter et interpréter des informations de haut niveau sémantique à partir d'images numériques ou de flux vidéos, automatisant ainsi les tâches d'analyse visuelle complexes (**Figure 06**) (Staff, 2025).

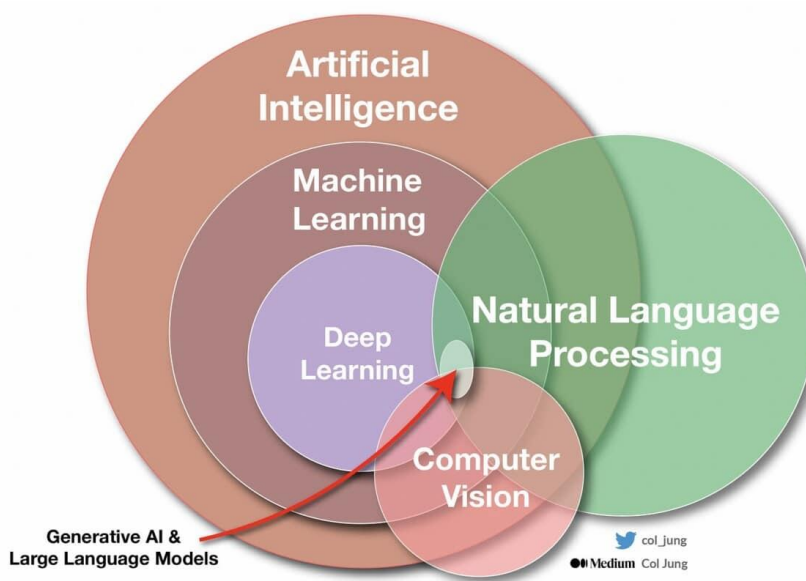


Figure 01: Diagramme d'intégration des disciplines de l'intelligence artificielle en pathologie numérique (Jain, 2023).

3. Processus d'acquisition des données histopathologiques :

L'acquisition des données représente la phase initiale d'extraction et de centralisation des flux d'images numériques. Elle constitue l'étape fondamentale du processus d'ingestion, précédant la validation, la normalisation et le chargement dans les architectures de calcul.

Synthèse bibliographique

L'entraînement de modèles d'apprentissage profond hautement performants requiert l'accès à de larges volumes d'images de haute qualité. En pathologie numérique, ces ensembles de données se caractérisent par leur taille massive et leur nature non structurée. Contrairement aux applications informatiques standards, les sources de données en oncologie mammaire sont strictement cliniques et standardisées :

- **Cohortes histopathologiques institutionnelles** : Coupes de tissus tumoraux issues de biopsies ou de pièces de résection chirurgicale, fixées en formol et incluses en paraffine (FFPE).
- **Banques de données publiques et ouvertes** : Répertoires internationaux de référence (ex. *The Cancer Genome Atlas* - TCGA) utilisés pour l'étalonnage et l'évaluation de la robustesse des modèles face à la variabilité inter-laboratoire.

4. Défis et contraintes de l'acquisition en environnement médical :

Le déploiement opérationnel des flux d'acquisition d'images se heurte à plusieurs contraintes techniques et réglementaires :

- **Confidentialité et sécurité des données** : L'exploitation des images médicales impose une anonymisation stricte et irréversible (suppression des métadonnées d'identification des patientes au format DICOM/TIFF) conformément aux exigences bioéthiques et réglementaires de protection des données de santé.
- **Qualité et standardisation** : La fidélité des prédictions d'un algorithme dépend directement de la qualité de la vérité terrain (*ground truth*). Les données d'entraînement doivent être exemptes d'artefacts techniques majeurs (flou de mise au point, plis de tissu, bulles d'air). De plus, le nettoyage et la normalisation des variations chromatiques liées aux protocoles de coloration (IHC/DAB) sont indispensables pour garantir la convergence du modèle.
- **Infrastructures matérielles** : L'acquisition de données volumineuses nécessite une architecture matérielle robuste, dotée de capacités de calcul (GPU) haut de gamme et de systèmes de stockage à haute disponibilité pour manager des fichiers de l'ordre du gigapixel (**Figure 07**) (Krantz, 2025).

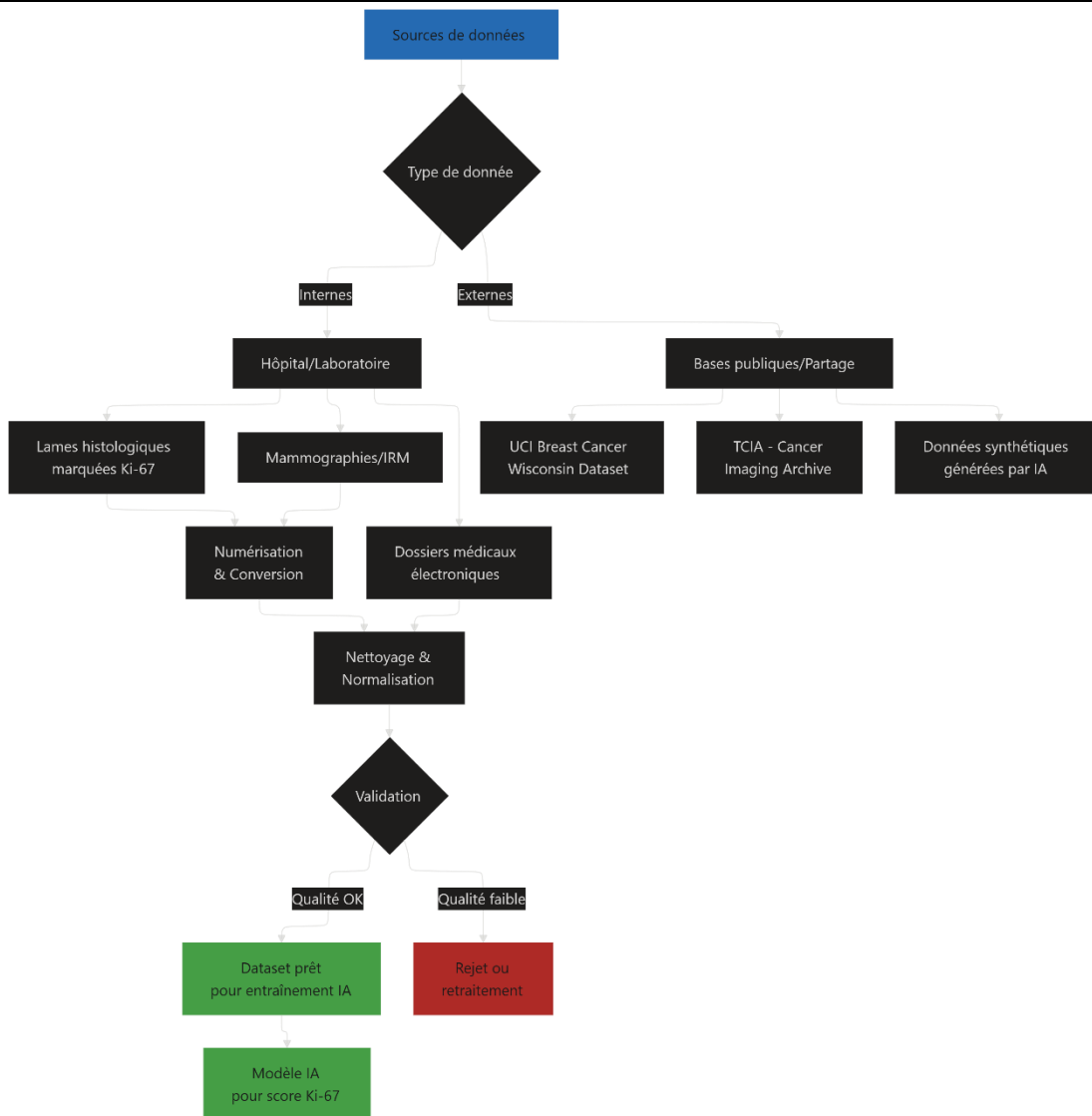


Figure 07: Processus d'acquisition et de prétraitement des données pour l'IA en pathologie mammaire (élaboration propre d'après les recommandations et études sur l'acquisition de données pour l'IA).

5. Numérisation des lames entières (*Whole Slide Imaging - WSI*) :

La numérisation des lames entières (WSI) consiste à convertir une lame de verre histopathologique en une image numérique haute résolution à l'aide d'un scanner optique de haute précision, généralement à des grossissements de 20 ou 40. Le fichier généré adopte une structure pyramidale multi-résolution contenant plusieurs milliards de pixels (images gigapixels).

Ce procédé permet de s'affranchir des limites du microscope optique traditionnel en offrant une visualisation exhaustive de la topographie tumorale. En intégrant le format WSI, les algorithmes de vision par ordinateur peuvent analyser simultanément l'architecture tissulaire globale et les détails

Synthèse bibliographique

subcellulaires des noyaux, rendant possible la quantification automatisée, exhaustive et reproductible du score de prolifération Ki-67 sur l'ensemble de la coupe.

6. L'imagerie de lames entières (*Whole Slide Imaging - WSI*) : Applications cliniques et synergie avec l'IA :

- **6.1. Intégration du flux WSI en anatomopathologie :**

L'imagerie de lames entières (*Whole Slide Imaging - WSI*) consiste à numériser l'intégralité d'une lame de verre histopathologique pour générer une image numérique haute résolution (gigapixel). Sur le plan fonctionnel, cette approche se substitue à la microscopie optique traditionnelle tout en apportant des fonctionnalités avancées de navigation dynamique, telles que le zoom continu, le déplacement fluide dans les champs tissulaires et l'archivage pérenne sans risque de dégradation physique ou de décoloration des tissus (**Figure 08**).

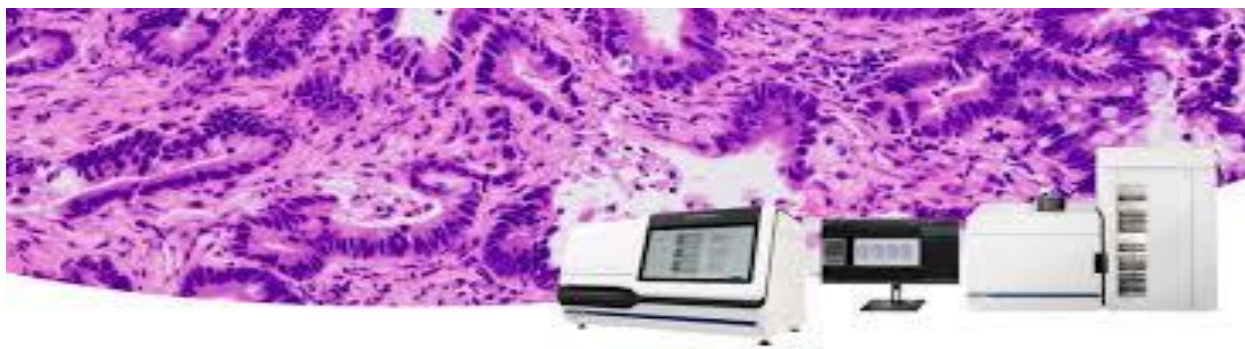


Figure 08: Systèmes de numérisation automatisés (scanners de lames) en routine clinique (Microscopes de numérisation de lames cliniques | *Analisis, s. d.*).

- **6.2. Optimisation du diagnostic et télépathologie :**

L'implémentation de la WSI au sein du flux de travail clinique optimise la précision diagnostique à plusieurs niveaux :

- **Analyse morphologique augmentée :** L'examen sur écran haute résolution permet une évaluation globale et locale de l'architecture tumorale. La capacité de naviguer rapidement entre différentes échelles visuelles facilite l'identification d'anomalies cytologiques subtiles, une étape déterminante pour le diagnostic différentiel et la stadification des carcinomes mammaires.
- **Télépathologie et expertise à distance :** La dématérialisation des lames physiques permet leur transfert instantané sur des plateformes distantes sécurisées. Les pathologistes peuvent

Synthèse bibliographique

ainsi collaborer, partager des cas complexes et établir des consensus diagnostiques en temps réel, indépendamment des contraintes géographiques. Ce flux de travail collaboratif s'est avéré particulièrement résilient lors de crises sanitaires limitant les interactions physiques (Figure 09).

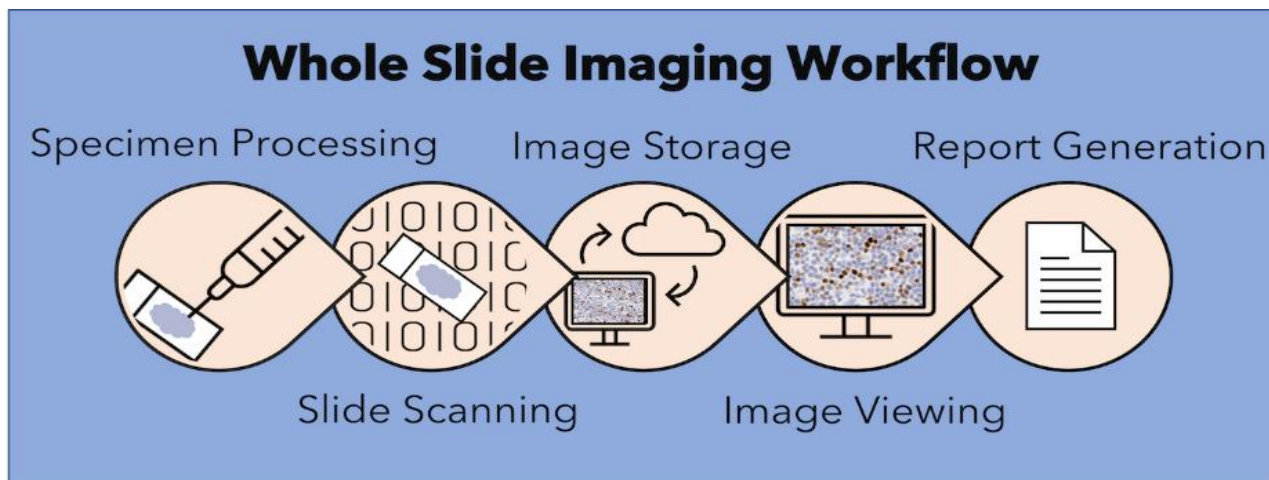


Figure 09: Flux de travail standardisé de la pathologie numérique, de la numérisation au diagnostic (WSI applications, s. d.).

- **6.3. Interfaçage de l'imagerie numérique et des algorithmes d'apprentissage profond :**

La convergence entre la WSI et les solutions d'intelligence artificielle transforme l'analyse histopathologique quantitative en s'attaquant directement aux tâches de routine à haut volume :

- **Dépistage et hiérarchisation :** Les algorithmes de vision par ordinateur présélectionnent et trient les lames de biopsies mammaires. En identifiant de manière autonome les zones saines et en signalant exclusivement les régions suspectes ou hautement prolifératives, l'IA réduit le temps d'analyse par cas et sécurise le processus de lecture.
- **Quantification des biomarqueurs :** Entraînés sur de larges bases de données annotées, les réseaux de neurones identifient et segmentent les motifs cellulaires spécifiques. Dans le cadre du cancer du sein, l'outil d'IA cartographie la tumeur, isole les noyaux des cellules tumorales invasives et quantifie précisément l'expression des biomarqueurs (tels que le Ki-67) en discriminant les signaux positifs (DAB) des signaux négatifs (hématoxyline) (Figure 10).

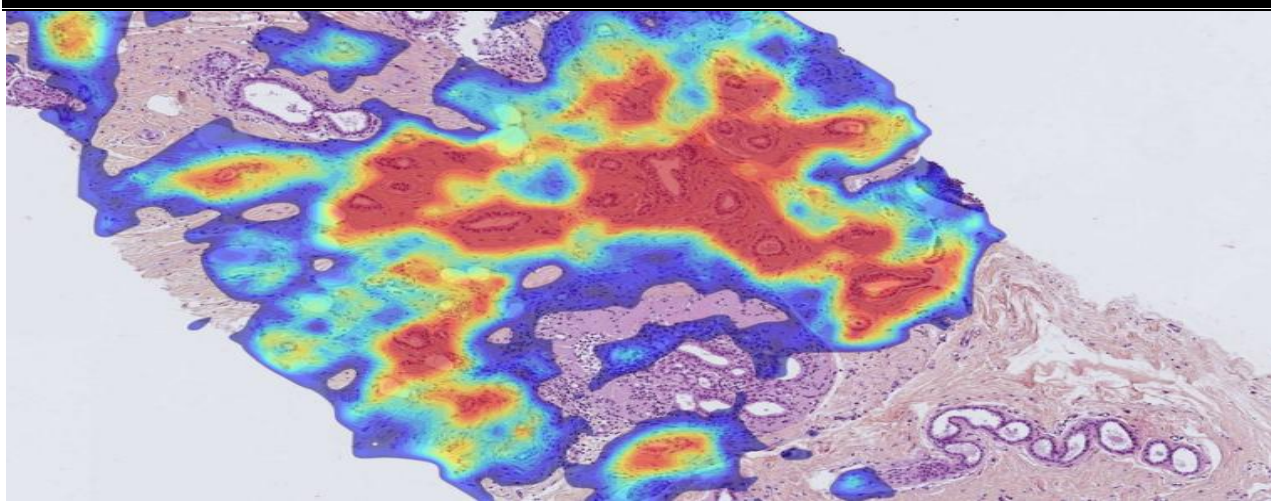


Figure 10 : Analyse automatisée d'une biopsie mammaire numérisée par un modèle d'apprentissage profond (Ballandras, 2022).

7. Limites techniques de la WSI et solutions d'infrastructure :

Bien que l'imagerie de lames entières standardise l'évaluation des aptitudes diagnostiques et fiabilise les programmes d'amélioration des performances, son déploiement à grande échelle soulève des défis matériels non négligeables. La taille gigapixel des fichiers d'images de lames entières (souvent supérieure à plusieurs gigaoctets par lame) exige une capacité de stockage locale considérable et des bandes passantes réseau à très haut débit pour garantir un affichage fluide et sans latence.

Pour contourner ces contraintes d'infrastructure sur site (*on-premise*), l'adoption d'architectures décentralisées basées sur le *Cloud* s'impose comme une alternative performante. Ces plateformes offrent des solutions de stockage dynamiques et évolutives (*scalable*), sécurisent l'accès distant aux données d'imagerie et facilitent la gestion de volumes massifs de données sans saturer les ressources locales des établissements de santé (Masjoodi et al., 2025).

8. Prétraitement et fondements de l'immunohistochimie (IHC) :

La performance des algorithmes de vision par ordinateur dépend intrinsèquement de la qualité de la préparation tissulaire en amont. L'immunohistochimie (IHC) constitue l'étape biologique de prétraitement permettant de cibler, par une réaction antigène-anticorps spécifique, des biomarqueurs d'intérêt au sein de coupes de tissus fixées au formol et incluses en paraffine (FFPE).

Contrairement aux colorations histologiques standards comme l'Hématoxyline-Éosine (H&E) qui révèlent uniquement la topographie et la morphologie cellulaire globale sans spécificité

Synthèse bibliographique

moléculaire, l'IHC permet de caractériser fonctionnellement la tumeur. L'utilisation d'un système de détection enzymatique couplé à un chromogène généralement le 3,3'-diaminobenzidine (DAB) génère un signal colorimétrique marron stable et hautement contrasté au niveau des sites cellulaires cibles.

Dans le cadre de l'évaluation du cancer du sein, ce processus analytique est standardisé afin de garantir l'homogénéité des images numériques qui alimenteront l'algorithme :

- **Fixation et préparation** : Une fixation tissulaire contrôlée évite l'apparition de zones d'autolyse ou de sur-fixation qui altèrent la morphologie des noyaux. Les coupes doivent être fines (généralement 3 à 4 micromètres) et parfaitement étalées pour éliminer les plis de tissu ou les bulles d'air, sources d'artefacts de flou lors de la numérisation.
- **Démasquage antigénique et blocage** : Un traitement thermique ou enzymatique restitue la conformation spatiale des épitopes masqués par la fixation. Le blocage des peroxydases endogènes est impératif pour éliminer le bruit de fond et éviter les faux positifs lors de la phase de segmentation automatisée.
- **Contre-coloration** : L'application finale d'hématoxyline colore en bleu les noyaux cellulaires négatifs (non prolifératifs). Cette étape fournit le contraste chromatique indispensable à l'IA pour effectuer la classification binaire entre noyaux positifs (marron, DAB) et noyaux négatifs (bleu, hématoxyline) (*http, s. d.*).

9. Choix des outils d'analyse : QuPath et DeepLIF, deux solutions open-source pour la quantification du Ki-67 :

Plusieurs architectures de réseaux de neurones convolutifs spécialisées, telles que PathoNet, SER-Unet ou Kpi-Net, ont été développées récemment pour la segmentation du Ki-67. Toutefois, dans le cadre de notre étude au CHU de Tlemcen, nous nous sommes concentrés sur l'évaluation d'outils open-source directement transposables en routine clinique, à savoir QuPath et le modèle DeepLIF

III. Évaluation de la Performance et Validation Clinique :

1. Définition de la référence absolue (Gold Standard) :

La validation repose sur l'analyse de 300 coupes histologiques ou plus numérisées, subdivisées individuellement en 9 zones géométriques. La référence absolue (M Ki-67) est définie par le ratio

rigoureux entre le nombre de noyaux tumoraux positifs et le nombre total de cellules tumorales, validé manuellement zone par zone par un comité de pathologistes spécialisés (Li et al., 2022).

2. Analyse de concordance et optimisation du flux de travail :

L'évaluation comparative des performances avec et sans assistance par IA met en évidence une amélioration globale des indicateurs cliniques (précision, concordance inter-observateur et réduction des délais d'obtention des résultats) (*AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer | Scientific Reports, s. d.*) :

- **Fiabilité statistique** : Sur les cohortes d'entraînement et de validation, l'analyse par IA obtient un coefficient de corrélation intraclasse (ICC) de 0,972, une valeur supérieure à celle du comptage manuel traditionnel (ICC = 0,803).
- **Reproductibilité** : La carte thermique (Figure 11) objective une forte convergence entre l'interprétation des pathologistes assistés par IA et la référence absolue, validant la réduction des biais inter- et intra-observateur.

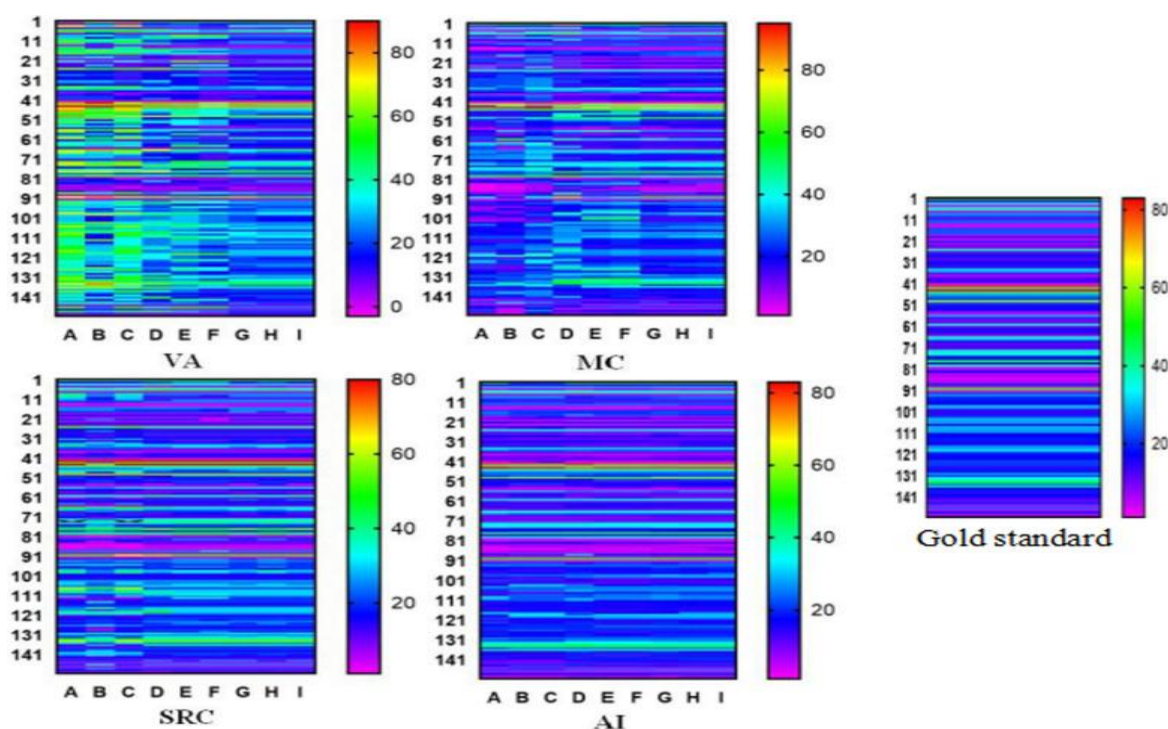


Figure 11 : Neuf pathologistes ont interprété la carte thermique Ki-67LI du cancer du sein via VA, MC, SRC et AI (Li et al., 2022).

3. Discussion et gestion de l'hétérogénéité tumorale :

Synthèse bibliographique

- **Correction des biais de sélection** : Les écarts résiduels entre l'IA et le Gold Standard s'expliquent par le choix des zones lues. L'intégration de logiciels commerciaux (tels que 3DHistech, Roche ou aetherAI) résout ce problème en standardisant la méthode de la moyenne multi-zones (**Figure 12**).

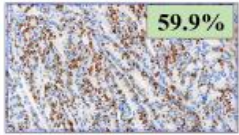
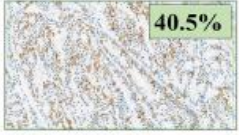
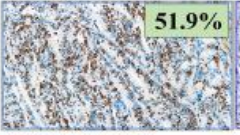
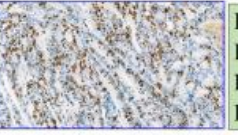
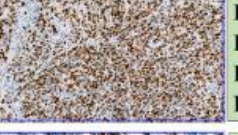
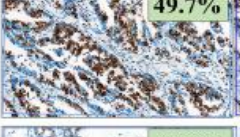
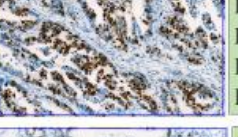
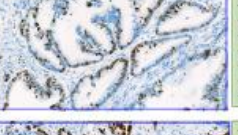
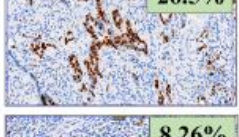
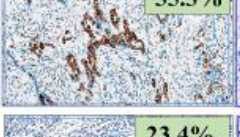
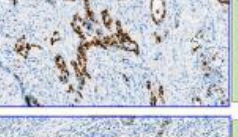
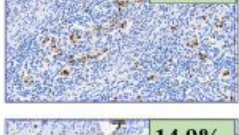
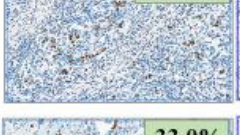
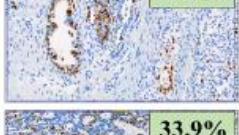
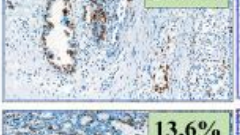
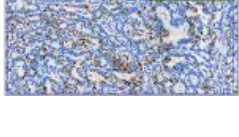
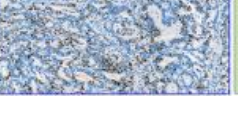
TMA core	3DHistech PI	Roche PI	aetherAI PI	Pathologists manual counting PI
A	 59.9%	 40.5%	 51.9%	 P1 = 75.0% P2 = 40.0% P3 = 40.0% P4 = 75%
B	 85.7%	 72.9%	 85.5%	 P1 = 95.0% P2 = 100.0% P3 = 95.0% P4 = 95.0%
C	 59.9%	 45.4%	 49.7%	 P1 = 80% P2 = 95% P3 = 65% P4 = 98%
D	 35.2%	 26.5%	 31.3%	 P1 = 20% P2 = 50% P3 = 30% P4 = 40%
E	 26.5%	 49.5%	 55.3%	 P1 = 85.0% P2 = 80.0% P3 = 75.0% P4 = 85.0%
F	 8.26%	 20.9%	 23.4%	 P1 = 50.0% P2 = 50.0% P3 = 60.0% P4 = 45.0%
G	 14.9%	 30.2%	 33.0%	 P1 = 45.0% P2 = 45.0% P3 = 60.0% P4 = 85.0%
H	 33.9%	 23.7%	 13.6%	 P1 = 15.0% P2 = 20.0% P3 = 20.0% P4 = 35.0%

Figure 12 : Exemple comparatif de l'indice de positivité dans des cas représentatifs d'immunohistochimie Ki-67, selon les méthodes 3DHistech, Roche, aetherAI et le comptage manuel (Li et al., 2022).

- **Échantillonnage de l'hétérogénéité** : Face à une architecture tumorale hétérogène, l'algorithme déploie des fenêtres d'analyse calibrées de 200 x 200 micromètres (grossissement 20x). Ce protocole cartographie distinctement les zones de forte densité (hotspots), de faible densité et les zones intermédiaires. Cette sectorisation sécurise le diagnostic en prévenant les erreurs d'identification informatique et la fatigue visuelle humaine (Li et al., 2022).

4. Conclusion :

L'automatisation du score Ki-67 par IA démontre une reproductibilité optimale et une excellente proximité avec la vérité terrain. Elle répond aux exigences des groupes de travail internationaux en s'affirmant comme un outil d'aide à la décision standardisé et performant pour la routine anatomopathologique (*A comparison of Ki67 index in carcinoma breast by digital image analysis with manual method - Indian J Pathol Oncol*, s. d.; *AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer* | *Scientific Reports*, s. d.; Li et al., 2022).

5. Métriques de concordance et reproductibilité :

L'intégration de l'intelligence artificielle (IA) et de l'analyse d'images numériques (DIA) sécurise l'évaluation du Ki-67 en réduisant les erreurs d'interprétation humaine et en harmonisant les relevés inter-observateurs.

- **Réduction des erreurs :** L'assistance par IA abaisse le taux d'erreur d'interprétation des pathologistes de 5,9 % à 2,1 %.
- **Analyses de Bland-Altman :** Les pathologistes assistés par IA affichent un biais systématique réduit (0,7 contre -2,7) et des limites d'accord plus resserrées (6,5 à -5,1 contre 10,2 à -15,6) par rapport au score de référence.
- **Binarisation au seuil clinique de 20 % :** La concordance inter-évaluateurs remonte significativement avec l'IA, atteignant un Kappa de Fleiss de 0,86 contre 0,40 en évaluation conventionnelle (*AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer* | *Scientific Reports*, s. d.).

Synthèse bibliographique

Tableau 03 : Synthèse des indicateurs de concordance inter-observateurs (A comparison of Ki67 index in carcinoma breast by digital image analysis with manual method - Indian J Pathol Oncol, s. d.; AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer | Scientific Reports, s. d.) :

Étude	Configuration	Indicateur statistique	Résultat (Sans IA vs Avec IA)	Interprétation pratique
Dy et al., 2024	Lecture sur console	ICC	0,70 vs 0,92	Concordance excellente
		Alpha de Krippendorff	0,65 vs 0,89	Très forte reproductibilité
		Kappa de Fleiss	0,40 vs 0,86	Concordance presque parfaite
Cai et al.	Microscope augmenté	ICC	0,827 vs 0,930	Concordance excellente
Étude DIA	Algorithme autonome	Kappa	0,636 (vs lecture visuelle)	Concordance substantielle

6. Modèles pronostiques multi-modaux et impact sur la survie :

L'évaluation numérique du Ki-67 s'étend à des modèles prédictifs non invasifs (imagerie/clinique) directement corrélés aux indicateurs de survie des patientes :

- **Corrélation d'imagerie échographique (SVM & ResNet-50) :** L'extraction de caractéristiques échographiques couplée à un classificateur SVM prédit le niveau d'expression du Ki-67 avec une exactitude de 0,778 (interne) et 0,827 (externe). Les patientes classées à haut risque par le modèle présentent un taux de métastases ganglionnaires supérieur (67,7 % vs 16,2 %) et une survie sans récurrence (SSR) plus courte .

- **Survie Sans Maladie (SSM) :** Les données de validation interne confirment une réduction significative de la SSM chez les patientes présentant un score Ki-67 élevé ($p = 0,005$) (Cen et al., 2025; Yu et al., 2026).
- **Score d'hétérogénéité spatiale (Ki67CL) :** Le score de colocalisation spatiale (Ki67CL) s'avère plus performant que l'indice Ki-67 standard pour stratifier les patientes ER+/HER2-. Il montre une haute valeur pronostique indépendante pour la survie spécifique au cancer du sein ($p < 0,00001$) et la survie sans métastases à distance ($p = 0,0048$) (Lu et al., 2023).

IV. Enjeux cliniques, éthiques et responsabilité médicale :

La transition vers la pathologie numérique, accélérée par l'implémentation de l'IA, redéfinit le flux de travail mais soulève des questions juridiques et organisationnelles majeures.

1. Évolution du rôle du pathologiste et intégration au workflow :

La numérisation complète des lames est le prérequis indispensable à l'IA. En automatisant les tâches répétitives de comptage cellulaire, l'IA ne remplace pas le médecin mais réoriente son expertise vers l'intégration des données cliniques complexes et les décisions thérapeutiques majeures (*La transformation numérique de la pathologie, s. d.*).

2. Analyse économique de la responsabilité civile médicale :

L'introduction de l'IA crée une asymétrie entre le producteur de l'algorithme et le médecin prescripteur final. Deux régimes juridiques principaux permettent d'atteindre une efficacité sociale :

- *Responsabilité exclusive du médecin :* Le praticien assume seul le risque juridique, ce qui l'incite à acheter des outils plus fiables, poussant indirectement le producteur à optimiser sa qualité.
- *Règle de négligence combinée à une responsabilité stricte du producteur :* La responsabilité du médecin n'est engagée qu'en cas de manquement au standard de soin (défaut de vigilance). Si le médecin a suivi le protocole et que l'erreur est liée à un défaut intrinsèque de l'outil, le fabricant prend en charge l'indemnisation.

3. Risque du partage des coûts :

Synthèse bibliographique

Un partage systématique des responsabilités en cas de faute peut être contre-productif, car il tend à relâcher la vigilance du médecin tout en réduisant les incitations à l'investissement de sécurité chez le producteur (*La responsabilité médicale à l'ère de l'intelligence artificielle*, s. d.).

4. Analyse des coûts opérationnels et infrastructures :

L'adoption de l'IA implique d'évaluer les coûts indirects et les exigences techniques pour garantir la viabilité du laboratoire :

Facteurs techniques de surcoût : La vitesse et le débit des scanners impactent directement le temps de travail. Un taux élevé de rejet des lames (défauts de mise au point) double la charge de main-d'œuvre et sature le stockage.

Besoins informatiques : Les volumes de données générés par la numérisation à haute résolution exigent des investissements soutenus dans les serveurs de stockage, la bande passante (télépathologie) et la cybersécurité (*Coûts cachés en pathologie numérique : comment budgétiser*, s. d.).

V. Perspectives multi-biomarqueurs et standardisation internationale :

1. Au-delà du Ki-67 : vers une analyse multiparamétrique :

L'évaluation conjointe des récepteurs hormonaux (ER, PR, nucléaires) et de HER-2 (membranaire) requiert des modèles hybrides. Les architectures basées sur des réseaux entièrement convolutionnels (FCN) et DenseNet, combinées à des algorithmes de saillance visuelle, permettent de s'affranchir des limites de la segmentation stricte des contours, rendant le système robuste face aux colorations hétérogènes (*An Advanced Automated Image Analysis Model for Scoring of ER, PR, HER-2 and Ki-67 in Breast Carcinoma* | *IEEE Journals & Magazine* | *IEEE Xplore*, s. d.).

2. Standardisation internationale et stratification thérapeutique :

La distinction entre les sous-types Luminal A et Luminal B repose sur une sectorisation précise des seuils décisionnels du Ki-67 (**Tableau 04**) (Lombardi et al., 2021).

Synthèse bibliographique

Tableau 04 : Stratification des risques et recommandations cliniques (Élaboration propre d'après Lombardi et al., 2021 et recommandations Grand-Est 2014) :

Classe de risque	Index Ki-67	Sous-type moléculaire	Risque de rechute / décès	Significativité statistique	Traitement recommandé
Faible	$\leq 14\%$	Luminal A	Faible	-	Hormonothérapie seule
Intermédiaire	15% – 20%	Transitionnel	Intermédiaire	-	Panel multigénique (TAILORx)
Élevé	$> 20\%$	Luminal B	Élevé	p = 0,002 (rechute) p = 0,01 (décès)	Chimiothérapie + Hormonothérapie

Chapitre 02 : Partie

Analytique :

Étude 1: ANALYSE COMPARATIVE DES CARACTÉRISTIQUES CLINICO- PATHOLOGIQUES SELON L'ÂGE

Introduction

Introduction :

- **L'Étude 1** repose sur une cohorte de 164 patientes suivies à l'unité d'oncologie de l'hôpital de Ghazaouet. Pour chacune d'elles, les dossiers médicaux se sont révélés complets, intégrant l'ensemble des variables cliniques, démographiques et anatomopathologiques indispensables à notre démarche.
- **L'Étude 2** s'appuie quant à elle sur un échantillon élargi à 303 patients, défini par l'intégralité des lames histologiques disponibles au CHU de Tlemcen, sans restriction liée à la disponibilité des dossiers cliniques. Toutefois, l'analyse comparative stricte se limite à 264 patients, seuls ces derniers présentant des données manuelles exhaustives.
- **Justification méthodologique :** Cette fluctuation d'effectifs reflète fidèlement la nature rétrospective de notre étude et les difficultés inhérentes à l'accès à des bases de données cliniques totalement centralisées. Pour autant, cette hétérogénéité d'effectifs n'altère en rien la validité ni la portée des analyses statistiques menées au sein de chaque groupe.

Matériel et Méthodes

1. Description de la cohorte :

L'étude porte sur une cohorte de 164 patientes, avec un âge moyen de $56,7 \pm 13,2$ ans. On note une prédominance nette des femmes âgées de 45 ans et plus, représentant 85,4 % (n=140) de l'effectif, contre 14,6 % (n=24) pour les patientes de moins de 45 ans (**Figure 13**).

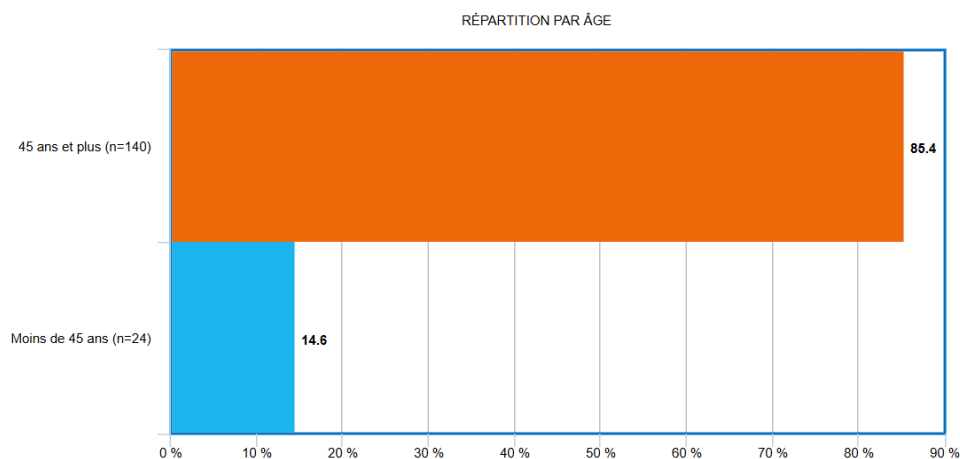


Figure 13 : Répartition par âge.

Sur le plan des comorbidités, l'hypertension artérielle constitue l'affection la plus fréquemment retrouvée, touchant 17,7 % de l'ensemble des patientes. L'analyse met également en évidence une répartition géographique hétérogène, avec un recrutement provenant de plusieurs localités, notamment Ghazaouet, Nedroma et Tounane, ce qui traduit un recrutement régional diversifié.

Un point important ressort de cette description : 11,6 % (19/164) des données relatives au Ki-67 ne sont pas renseignées dans les dossiers initiaux, ce qui souligne l'intérêt d'une standardisation du recueil des données anatomopathologiques. Cette absence d'information constitue une limite majeure pour l'évaluation pronostique et souligne l'intérêt d'une meilleure standardisation du recueil des données anatomopathologiques.

Résultats et interprétation

Résultats et Interprétation

2. Étude comparative selon les tranches d'âge :

• 2.1. Comparaison des caractéristiques cliniques et métaboliques :

Parmi les 164 patientes incluses dans l'étude, 24 (14,6 %) étaient âgées de moins de 45 ans, tandis que 140 (85,4 %) avaient un âge supérieur ou égal à 45 ans. Une différence hautement significative a été observée pour le statut ménopausique ($p < 0,0001$), en accord avec la distribution attendue selon l'âge des patientes.

L'analyse des comorbidités a montré une prévalence plus élevée de l'hypertension artérielle chez les patientes âgées de 45 ans ou plus comparativement aux patientes plus jeunes (19,3 % contre 8,3 %), sans que cette différence n'atteigne le seuil de significativité statistique ($p = 0,255$). De même, aucune différence significative n'a été mise en évidence concernant le diabète de type 2 ($p = 0,700$).

Par ailleurs, l'étude des caractéristiques gynéco-obstétricales, incluant la multiparité et l'allaitement, ainsi que des antécédents familiaux de cancer, n'a pas montré de différence statistiquement significative entre les deux groupes d'âge.

Tableau 05 : Ce tableau compare les facteurs de risque et les comorbidités entre les deux tranches d'âge.

Caractéristiques	< 45 ans (n = 24)	≥ 45 ans (n = 140)	p
Statut ménopausique			< 0,0001
Pré-ménopause	10 (41,7 %)	0 (0 %)	
Post-ménopause	0 (0 %)	28 (20,0 %)	
Non précisé	14 (58,3 %)	112 (80,0 %)	
Comorbidités			
HTA	2 (8,3 %)	27 (19,3 %)	0,255
Diabète type 2	1 (4,2 %)	14 (10,0 %)	0,700

L'analyse de la population étudiée ($n = 164$) montre une prédominance de patientes âgées de 45 ans et plus, représentant 85,4 % de l'échantillon, avec un âge moyen de $56,7 \pm 13,2$ ans. Sur le plan géographique, le recrutement apparaît diversifié, avec une représentation notable des localités de Ghazaouet (11 %) et de Nedroma (9,1 %). Concernant les antécédents médicaux, l'hypertension artérielle (HTA) constitue la comorbidité la plus fréquente, retrouvée chez 17,7 % des patientes.

Cependant, bien que la fréquence de l'HTA soit plus élevée chez les patientes âgées de 45 ans ou plus (19,3 %) que chez celles de moins de 45 ans (8,3 %), cette différence n'atteint pas le seuil de significativité statistique après application du test exact de Fisher ($p = 0,255$). Ainsi, aucune association statistiquement significative entre l'HTA et l'âge n'a été mise en évidence dans cette cohorte. En conséquence, il n'est pas approprié d'interpréter cette différence en termes de sur-risque statistiquement démontré dans l'échantillon étudié.

Par ailleurs, l'analyse des données a mis en évidence un taux élevé de données manquantes concernant le statut ménopausique, avec 58,3 % (14/24) des patientes de moins de 45 ans et 80,0 % (112/140) de 29 Résultats et Interprétation celles âgées de 45 ans ou plus classées dans la catégorie « non précisé ». Cette absence d'information constitue une limite importante de l'étude et souligne la nécessité d'une meilleure standardisation du recueil des données cliniques.

Résultats et Interprétation

Tableau 05: Répartition des caractéristiques gynéco-obstétricales et des antécédents familiaux selon les groupes d'âge des patientes.

Caractéristiques	< 45 ans (n = 38)	≥ 45 ans (n = 126)	p
Parité ≥ 3 enfants	4 (16,7 %)	23 (16,4 %)	1,000
Allaitement (Oui)	6 (25,0 %)	20 (14,3 %)	0,234
Antécédents familiaux de cancer	12 (50,0 %)	48 (34,3 %)	0,180

Note : Les données sont exprimées en effectifs (n) et pourcentages (%). Le seuil de significativité statistique a été fixé à $p < 0,05$.

L'analyse des facteurs gynéco-obstétricaux et des antécédents familiaux selon l'âge des patientes n'a pas mis en évidence de différence statistiquement significative entre les deux groupes. La parité élevée (≥ 3 enfants) était observée dans des proportions comparables chez les patientes de moins de 45 ans et celles âgées de 45 ans ou plus, avec respectivement 16,7 % et 16,4 % ($p = 1,000$). De même, l'allaitement était rapporté chez 25,0 % des patientes de moins de 45 ans contre 14,3 % chez les patientes âgées de 45 ans ou plus, sans différence significative entre les groupes ($p = 0,234$). Enfin, les antécédents familiaux de cancer étaient retrouvés chez 50,0 % des patientes de moins de 45 ans contre 34,3 % chez celles âgées de 45 ans ou plus, sans atteindre le seuil de significativité statistique ($p = 0,180$). Dans l'ensemble, ces résultats suggèrent que, dans cette cohorte, les caractéristiques gynéco-obstétricales étudiées ainsi que les antécédents familiaux de cancer ne diffèrent pas significativement selon l'âge des patientes.

• 2.2. Comparaison des caractéristiques tumorales et histologiques :

La comparaison entre les patientes de moins de 45 ans et celles âgées de 45 ans et plus met en évidence plusieurs différences cliniques et biologiques.

Sur le plan clinique et métabolique, le statut post-ménopausique est quasi exclusivement retrouvé chez les patientes les plus âgées ($p < 0,0001$). De même, l'hypertension artérielle est significativement plus fréquente chez les patientes âgées de 45 ans et plus ($p = 0,03$), avec un odds ratio de 2,98. Ces résultats suggèrent que les patientes plus âgées présentent un terrain

Résultats et Interprétation

cardiovasculaire plus fragile, ce qui rend l'évaluation anatomopathologique d'autant plus importante pour guider la prise en charge thérapeutique.

Sur le plan histologique, le carcinome de type non spécifique (NST) est majoritaire (58,5 %).

L'agressivité tumorale, évaluée par le grade SBR, révèle une prédominance de tumeurs de Grade II (59,8 %), suivies par le Grade III (17,1 %). La répartition selon la taille tumorale (TNM) montre que les tumeurs T2 (21-50 mm) sont les plus fréquentes (27,4 %).

Sur le plan tumoral, le carcinome de type non spécifique constitue l'histologie dominante, et la majorité des tumeurs sont classées en grade II de SBR (59,8 %). Les récepteurs hormonaux sont exprimés de manière comparable dans les deux groupes, avec environ 51 % de tumeurs RH positives.

Tableau 07 : Comparaison des caractéristiques tumorales et moléculaires selon l'âge.

Paramètres tumoraux	< 45 ans (n = 38)	≥ 45 ans (n = 126)	<i>p</i>
Taille tumorale (T2/T3)	5 (20,8 %)	44 (31,4 %)	0,343
Grade SBR II + III	10 (41,7 %)	62 (44,3 %)	0,826
RH positifs	12 (50,0 %)	60 (42,9 %)	0,647
HER2 positif	2 (8,3 %)	10 (7,1 %)	1,000
Ki-67 ≥ 20 %	8 (44,4 %)	30 (36,1 %)	0,590

Note : Les données sont exprimées en effectifs (n) et pourcentages (%). Les valeurs de $p < 0,05$ sont considérées comme statistiquement significatives.

3. Focus sur le Ki-67 :

- 3.1. Analyse de l'expression du Ki-67 selon l'âge :

Concernant les paramètres tumoraux, aucune association statistiquement significative avec l'âge n'a été mise en évidence dans cette cohorte. La taille tumorale (T2/T3) tendait à être plus fréquente

Résultats et Interprétation

chez les patientes âgées de 45 ans ou plus (31,4 %) que chez celles de moins de 45 ans (20,8 %), sans atteindre le seuil de significativité statistique ($p = 0,343$). De même, la proportion de tumeurs de grade histopronostique élevé (SBR II + III) était comparable entre les deux groupes d'âge, avec 41,7 % chez les patientes de moins de 45 ans contre 44,3 % chez celles âgées de 45 ans ou plus ($p = 0,826$).

Le statut des récepteurs hormonaux ne différait pas significativement selon l'âge, les récepteurs hormonaux positifs étant retrouvés chez 50,0 % des patientes de moins de 45 ans et chez 42,9 % de celles âgées de 45 ans ou plus ($p = 0,647$). De même, l'expression de HER2 restait faible et comparable dans les deux groupes, avec 8,3 % chez les patientes les plus jeunes contre 7,1 % chez les plus âgées ($p = 1,000$).

Enfin, l'indice de prolifération Ki-67 supérieur ou égal à 20 % était observé chez 44,4 % (8/18) des patientes de moins de 45 ans et chez 36,1 % (30/83) des patientes âgées de 45 ans ou plus. Toutefois, cette différence n'était pas statistiquement significative ($p = 0,590$). Ainsi, dans notre série, aucun des paramètres tumoraux étudiés, y compris le Ki-67, n'a montré d'association significative avec l'âge des patientes.

Résultats et Interprétation

Tableau 08 : Profil des biomarqueurs moléculaires selon l'âge des patientes (n = 164).

Biomarqueurs	< 45 ans	≥ 45 ans	Total	p	OR [IC 95 %]
RE+	12 (50,0 %)	60 (42,9 %)	72 (43,9 %)	0,647	1,33 [0,56 – 3,17]
RP+	10 (41,7 %)	54 (38,6 %)	64 (39,0 %)	0,815	1,14 [0,47 – 2,74]
RH+ (RE+ ou RP+)	12 (50,0 %)	60 (42,9 %)	72 (43,9 %)	0,647	1,33 [0,56 – 3,17]
HER2+	2 (8,3 %)	10 (7,1 %)	12 (7,3 %)	1,000	1,18 [0,24 – 5,76]
Ki-67 ≥ 20 %	8 (44,4 %)	30 (36,1 %)	38 (37,6 %)	0,590	1,41 [0,50 – 3,97]

Valeur statistiquement significative ($p < 0,05$).

L'analyse des biomarqueurs tumoraux selon l'âge des patientes n'a pas mis en évidence de différence statistiquement significative entre les deux groupes. L'expression des récepteurs aux œstrogènes (RE+) était observée chez 50,0 % des patientes âgées de moins de 45 ans contre 42,9 % chez celles âgées de 45 ans ou plus, sans différence significative ($p = 0,647$; OR = 1,33 ; IC 95 % : 0,56–3,17). De même, la positivité des récepteurs à la progestérone (RP+) concernait 41,7 % des patientes jeunes contre 38,6 % des patientes plus âgées ($p = 0,815$; OR = 1,14 ; IC 95 % : 0,47–2,74). Le profil hormonal global, défini par la positivité des récepteurs hormonaux (RH+), était retrouvé chez 50,0 % des patientes de moins de 45 ans contre 42,9 % chez les patientes de 45 ans ou plus, sans différence significative ($p = 0,647$; OR = 1,33 ; IC 95 % : 0,56–3,17).

Concernant l'expression de HER2, celle-ci demeurait faible et comparable entre les deux groupes d'âge, avec 8,3 % chez les patientes de moins de 45 ans contre 7,1 % chez les patientes plus âgées

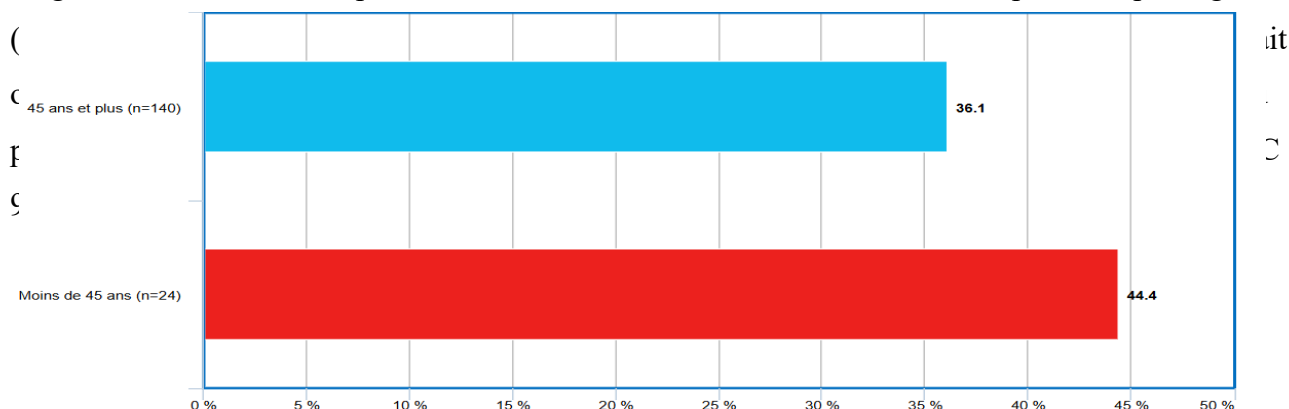


Figure 14 : KI-67 $\geq$ 20 % : Comparaison selon l'âge.

Dans l'ensemble, ces résultats suggèrent que, dans cette cohorte, le profil immunohistochimique tumoral ne diffère pas significativement selon l'âge des patientes.

Discussion

1. Synthèse et intérêt de l'approche analytique :

L'objectif de cette étude était de comparer les caractéristiques cliniques, métaboliques, gynéco obstétricales et tumorales de patientes atteintes de cancer du sein selon deux tranches d'âge (< 45 ans vs ≥ 45 ans). Nos résultats mettent en évidence une cohorte majoritairement mature (85,4 % d'âge ≥ 45 ans), ce qui est en parfaite adéquation avec l'épidémiologie globale du cancer du sein, où le risque augmente proportionnellement avec l'âge (**Lemij et al., 2023**).

- **Profil clinique, comorbidités et limites du recueil de données :**

L'analyse des comorbidités métaboliques, notamment le diabète de type 2 ($p = 0,700$) et l'hypertension artérielle ($p = 0,255$), n'a pas révélé de différence statistiquement significative entre les deux groupes d'âge. Bien que l'HTA affiche une prévalence brute supérieure chez les patientes de 45 ans ou plus (19,3 % contre 8,3 %), l'absence de significativité statistique s'explique probablement par le faible effectif du groupe des patientes jeunes ($n = 24$), limitant la puissance statistique de notre test. Dans la littérature internationale, les comorbidités cardiovasculaires et métaboliques sont classiquement associées au vieillissement cellulaire et à la ménopause (**Karihtala et al., 2021**).

Un fait marquant de notre étude réside dans le taux particulièrement élevé de données manquantes concernant le statut ménopausique (58,3 % chez les < 45 ans et 80,0 % chez les ≥ 45 ans). Cette absence d'information constitue une limite méthodologique importante. Comme le soulignent **Lemij et al. (2023)**, la standardisation rigoureuse des dossiers cliniques en oncologie est indispensable, car le statut ménopausique influence directement le choix des lignes thérapeutiques, notamment l'hormonothérapie (inhibiteurs de l'aromatase vs tamoxifène).

- **1.2. Facteurs gynéco-obstétricaux et antécédents familiaux :**

Concernant le profil gynéco-obstétrical (parité élevée et allaitement) et l'hérédité, aucune variation significative n'a été documentée entre les deux groupes. Pourtant, l'histoire reproductive joue un rôle complexe dans la pathogenèse mammaire. Une parité précoce et l'allaitement sont traditionnellement décrits comme des facteurs protecteurs, réduisant l'exposition cumulative aux œstrogènes endogènes (**Karihtala et al., 2021**).

De plus, bien que les antécédents familiaux de cancer soient plus fréquents chez les patientes de moins de 45 ans (50,0 % vs 34,3 %), cette différence reste non significative ($p = 0,180$). Dans la littérature, le jeune âge au diagnostic est fréquemment corrélé à une prédisposition génétique

Discussion

(mutations BRCA1/2), et les formes familiales y sont généralement surreprésentées (**Anders et al., 2009**). Notre 34 Conclusion échantillon restreint de patientes jeunes n'a pas permis de valider statistiquement cette tendance bien établie.

- **1.3. Caractéristiques tumorales et biomarqueurs : le paradoxe de la cohorte :**

Sur le plan anatomopathologique et moléculaire, notre étude montre une homogénéité surprenante entre les deux tranches d'âge. Aucun des paramètres clés qu'il s'agisse de la taille tumorale T2/T3 ($p = 0,343$), des grades SBR élevés II+III ($p = 0,826$), de l'expression des récepteurs hormonaux (RE+ : $p = 0,647$; RP+ : $p = 0,815$) ou de la positivité de HER2 ($p = 1,000$) n'a révélé de disparité significative. Ce résultat diverge des larges cohortes de la littérature oncologique. En effet, il est largement documenté que le cancer du sein chez la femme jeune (souvent défini avant 40 ou 45 ans) se caractérise par un phénotype nettement plus agressif :

- Une prédominance de grades histologiques élevés (Grade III).
- Un taux de prolifération cellulaire (Ki-67) plus élevé.
- Une sous-représentation des tumeurs luminales (RH+) au profit des sous-types HER2-enrichis ou triples négatifs (**Albasri, 2021 ; Yang et al., 2022**).

Dans notre série, l'indice de prolifération Ki-67 20 % tendait à être supérieur chez les patientes de moins de 45 ans (44,4 % vs 36,1 %), mais sans atteindre la significativité statistique ($p = 0,590$). Ce phénomène, combiné à l'absence de différence pour les récepteurs hormonaux ($p = 0,647$, OR = 1,33), suggère que les profils moléculaires au sein de notre population ne suivent pas les schémas d'agressivité accrue classiquement décrits chez les femmes jeunes (**Jackson et al., 2023**).

- **1.4. Limites de l'étude :**

Outre le biais des données manquantes sur la ménopause, la principale limite de ce travail réside dans l'asymétrie numérique des groupes comparés ($n = 24$ vs $n = 140$). Ce déséquilibre induit un manque de puissance statistique (erreur de type II), susceptible de masquer de réelles différences biologiques et cliniques entre les deux populations. Des études multicentriques incluant un effectif plus important de patientes jeunes sont nécessaires pour affiner ces observations au niveau régional.

Conclusion

Conclusion

Ce travail met en lumière le profil clinicobiologique des patientes prises en charge pour un cancer du sein au sein de notre population d'étude, marqué par une nette prédominance des femmes âgées de 45 ans et plus (85,4 %).

La principale constatation de cette étude réside dans l'homogénéité surprenante des caractéristiques cliniques, histologiques et moléculaires entre les deux tranches d'âge comparées. Contrairement aux données classiques de la littérature internationale qui associent fréquemment le jeune âge à des formes tumorales plus agressives, des grades SBR élevés ou des phénotypes triples négatifs —, notre cohorte ne montre aucune différence statistiquement significative concernant la taille de la tumeur, le statut des récepteurs hormonaux, l'expression de HER2 ou l'indice de prolifération Ki-67. De même, les comorbidités comme l'hypertension artérielle et le diabète de type 2, bien que numériquement plus fréquentes chez les femmes plus âgées, se répartissent de manière comparable sur le plan statistique.

Néanmoins, ces résultats doivent être interprétés à la lumière de certaines limites méthodologiques. L'asymétrie numérique importante entre les deux groupes (24 patientes de moins de 45 ans contre 140 de 45 ans ou plus) a considérablement réduit la puissance des tests statistiques, ce qui a pu masquer de réelles tendances biologiques. De plus, le taux élevé de données non précisées concernant le statut ménopausique constitue un biais de recueil majeur, limitant l'évaluation précise du terrain endocrinien des patientes.

En perspective, ce travail souligne l'absolue nécessité de standardiser et de rigoureusement documenter les dossiers médicaux dès le diagnostic en oncologie. Pour valider et approfondir ces premières constatations, il apparaît indispensable d'élargir cette recherche à travers des études multicentriques incluant un effectif plus représentatif de patientes jeunes. Une telle démarche permettrait de déterminer si les caractéristiques observées reflètent une particularité épidémiologique locale ou un effet de taille d'échantillon, garantissant ainsi une prise en charge thérapeutique toujours plus ciblée et adaptée.

1. Enjeux de reproductibilité et transition vers l'analyse automatisée :

La section suivante est consacrée à l'analyse des principaux modèles d'apprentissage profond dédiés à l'évaluation du Ki-67, à la comparaison de leurs performances et à leurs perspectives d'intégration dans la pratique courante.

ÉTUDE 2 : COMPARAISON DES MÉTHODES DE QUANTIFICATION DU KI-67 (MANUEL, QuPath, DeepLIIF)

Matériel et Méthodes

1.Introduction de l'étude comparative :

L'évaluation de l'indice de prolifération cellulaire par l'immunomarquage du Ki-67 s'est imposée comme un pilier fondamental de l'oncologie moderne. En pathologie mammaire, neuroendocrine ou lymphoïde, ce biomarqueur nucléaire dicte de manière déterminante les décisions thérapeutiques, orientant les équipes médicales vers des stratégies d'escalade ou de désescalade chimiothérapeutique (**Goldhirsch et al., 2011**). Le franchissement de certains seuils décisionnels cliniques, souvent fixés à des valeurs charnières telles que 20 %, modifie radicalement le pronostic et la prise en charge du patient (**Dowsett et al., 2011**).

Pourtant, malgré son importance cardinale, la quantification du Ki-67 se heurte à un défi historique de reproductibilité. La méthode de référence conventionnelle repose sur l'examen visuel par un anatomopathologiste. Qu'il s'agisse d'une estimation globale (eyeballing) ou d'un comptage manuel laborieux cellule par cellule, cette approche humaine est grevée d'une variabilité inter- et intra observateur largement documentée dans la littérature (**Polley et al., 2013**). Les variations de coloration, la fatigue visuelle, la subjectivité face aux intensités de marquage limites et la répartition hétérogène des cellules tumorales en îlots de forte prolifération (hotspots) compliquent l'harmonisation des résultats à l'échelle internationale (**Leung et al., 2016**). De surcroît, le caractère chronophage du comptage manuel s'accorde difficilement avec les exigences de productivité et les flux tendus des laboratoires d'analyses modernes.

L'avènement de la pathologie numérique et de la numérisation des lames entières (Whole Slide Imaging) offre une opportunité sans précédent de standardiser cette évaluation (**Fuchs & Buhmann, 2011**). En convertissant la matière biologique en données pixels de haute résolution, l'analyse d'images assistée par ordinateur promet d'apporter l'objectivité et la répétabilité qui font défaut à l'examen visuel seul (**Veta et al., 2014**). Cependant, l'intégration de ces technologies en routine diagnostique soulève de nouvelles questions méthodologiques : quelle technologie logicielle choisir pour garantir à la fois la sécurité diagnostique, l'efficacité temporelle et la viabilité opérationnelle au sein des structures hospitalières ?

2. Matériel et méthodes :

Pour répondre à cette problématique, la présente étude observationnelle s'appuie sur une cohorte robuste et représentative de la pratique courante, regroupant les données de 303 patients. Afin de capturer fidèlement l'hétérogénéité intra-tumorale et de s'affranchir des biais de sélection topographique, un protocole d'acquisition standardisé a été appliqué, menant à la sélection de trois

Matériel et Méthodes

champs d'intérêt distincts par cas, soit un corpus total de 909 images numériques de haute résolution.

Cette étude confronte en parallèle trois paradigmes technologiques et conceptuels distincts :

- Le comptage manuel par un pathologiste expérimenté représente le standard clinique actuel, encadré par les recommandations internationales **(Nielsen et al., 2021)**. Le clinicien réalise une numération exhaustive des noyaux tumoraux positifs (marqués au diaminobenzidine, DAB) et négatifs (contre-colorés à l'hématoxyline) sur les champs d'intérêt sélectionnés, garantissant une classification cellulaire basée sur une expertise médicale rigoureuse capable d'exclure les critères d'altération morphologique ou les faux positifs **(Rimm et al., 2007)**.
- L'algorithme open-source QuPath est établi comme la référence académique internationale en matière d'analyse d'images histologiques **(Bankhead et al., 2017)**. Cet outil repose sur des méthodes de segmentation d'image classiques. Il utilise la séparation des couleurs (déconvolution 40 Résultats et Interprétation colorimétrique) associée à des algorithmes de détection de contours basés sur des seuils d'intensité de pixels. La classification des cellules détectées s'appuie sur un apprentissage machine traditionnel (de type arbres de décision), paramétré par l'utilisateur en fonction des caractéristiques géométriques et texturales des noyaux.
- La plateforme d'intelligence artificielle DeepLIIF (Deep Learning Inferred Imaging Fluorescence) représente l'état de l'art des technologies fondées sur l'apprentissage profond **(Ghahremani et al., 2022)**. Contrairement aux approches par seuillage, ce modèle utilise des réseaux de neurones convolutifs avancés. DeepLIIF a la particularité de simuler une immunofluorescence virtuelle à partir d'une image classique en fond clair (DAB/Hématoxyline). Cette technique lui permet d'isoler de manière précise les signaux de prolifération, de séparer les noyaux cellulaires agglomérés et de discriminer nativement les cellules tumorales des éléments du microenvironnement, tels que les lymphocytes ou les fibroblastes **(Litjens et al., 2017)**.

3. Objectifs et critères d'évaluation :

L'objectif principal de ce travail est de définir la méthode offrant le meilleur compromis opérationnel pour une transition sécurisée vers la pathologie numérique de routine. L'évaluation s'articule autour de trois axes fondamentaux :

- **La concordance statistique et clinique :** La fidélité des outils numériques (QuPath et DeepLIIF) est mesurée par rapport à la référence humaine. L'analyse statistique repose sur le calcul du coefficient de corrélation intraclasse (CCI), indispensable pour juger de la fiabilité des mesures continues (Koo & Li, 2016). De plus, la construction de graphiques de Bland Altman permet d'identifier d'éventuels biais systématiques de sur- ou sous-évaluation entre les méthodes (Bland & Altman, 1986). Au-delà de la simple corrélation linéaire, une attention particulière est portée sur le taux de concordance diagnostique autour des seuils de décision cliniques.
- **La rapidité et l'efficacité computationnelle :** Le facteur temporel constitue le verrou majeur à l'adoption des technologies numériques. Le temps de traitement par image et par patient est rigoureusement chronométré pour chaque modalité. Cela inclut le temps de configuration humaine requis (ajustement des paramètres sur QuPath) et le temps de calcul machine pur (inférence du modèle DeepLIIF sur processeur graphique). L'objectif est de quantifier le gain de productivité réel par rapport aux minutes requises pour un comptage humain rigoureux.
- **La robustesse dans un contexte de routine varié :** Les images issues de la pratique quotidienne d'un laboratoire de pathologie présentent une variabilité inhérente au monde réel : variations d'épaisseur de coupe, intensités hétérogènes des automates de coloration, présence d'artefacts de préparation (plis, bulles) ou abondance de stroma fibreux. L'étude évalue la capacité de chaque outil à maintenir des performances stables face à ce bruit de fond clinique, sans exiger une réévaluation ou un réajustement permanent des paramètres par les techniciens (Acs et al., 2022).

En définitive, cette étude comparative vise à déterminer si l'innovation technologique représentée par l'apprentissage profond (DeepLIIF) apporte une valeur ajoutée mesurable, en termes de précision et de fluidité de travail, par rapport aux outils académiques éprouvés (QuPath) et à l'œil du clinicien, ouvrant ainsi la voie à une intégration raisonnée et sécurisée de l'assistance computationnelle en routine diagnostique oncologique.

Résultats et interprétation

1. Caractéristiques de la cohorte:

Nous avons défini comme positive toute cellule présentant un marquage nucléaire, quel que soit son intensité. Les trois champs ont été capturés dans la zone hotspot identifiée visuellement à l'objectif $\times 40$, et la moyenne a été calculée pour chaque patient et chaque méthode. La valeur de Ki-67 n'est pas une constante biologique absolue mais une estimation conditionnée par le choix du hotspot, le nombre de champs analysés, la méthode de segmentation nucléaire, et le seuil de positivité choisi.

Une proportion de cas (environ 13 %) ne disposait pas de valeur manuelle initiale, ce qui souligne l'intérêt d'une automatisation pour homogénéiser les comptes-rendus. Les analyses de corrélation et de concordance ont été réalisées sur des sous-échantillons spécifiques en fonction de la disponibilité des données. La comparaison de chaque méthode avec la référence manuelle (gold standard) a été restreinte aux 264 patients disposant d'une valeur manuelle complète (les cas "absent" ayant été exclus). En revanche, la comparaison directe entre les deux outils d'analyse d'image, QuPath et DeepLIIF, a été effectuée sur l'ensemble des 303 patients, car ces deux variables numériques étaient disponibles pour la totalité de la cohorte. Cette stratégie permet de maximiser la puissance statistique de la corrélation entre les deux IA tout en garantissant une évaluation rigoureuse de leur performance par rapport à la référence humaine.

2. Coefficients de corrélation de Pearson :

DeepLIIF présente une corrélation plus élevée avec la méthode manuelle que QuPath. La très forte corrélation entre QuPath et DeepLIIF ($r = 0,91$) indique que les deux méthodes numériques mesurent globalement le même phénomène biologique (**Figure 15a,b,c**). Cette corrélation est particulièrement marquée dans les valeurs extrêmes ($\text{Ki-67} > 70\%$), où les deux logiciels tendent à saturer vers des scores de 95-100%, là où l'œil humain conserve une estimation plus nuancée autour de 80%.

Afin de s'affranchir de l'influence potentielle des valeurs aberrantes, nous avons également calculé le coefficient de corrélation de Spearman (manuel vs DeepLIIF : $\rho = 0,86$; manuel vs QuPath : $\rho = 0,79$; QuPath vs DeepLIIF : $\rho = 0,93$). Les valeurs sont proches des coefficients de Pearson, indiquant que la relation entre les méthodes est globalement linéaire.

Résultats et Interprétation

Tableau 09 : Coefficients de corrélation de Pearson.

Comparaison	r de Pearson (IC 95%)	p-value
Manuel vs QuPath	0,72 (0,66 – 0,77)	< 0,0001
Manuel vs DeepLIIF	0,84 (0,80 – 0,88)	< 0,0001
QuPath vs DeepLIIF	0,91 (0,88 – 0,93)	< 0,0001

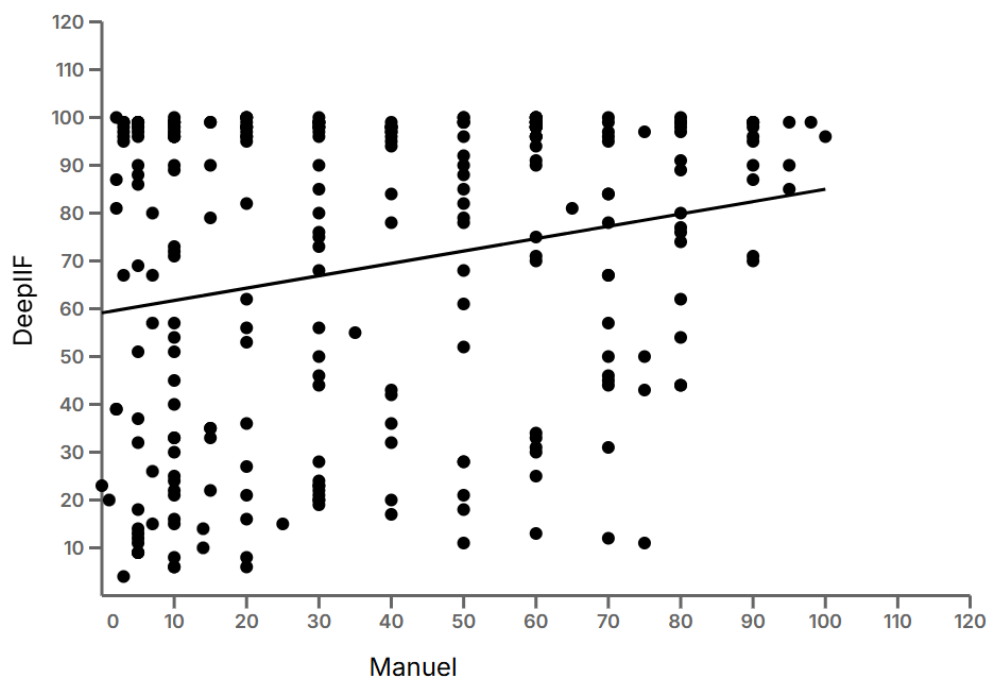


Figure 15a : Nuage de points montrant la corrélation entre les méthodes manuelle et DeepLIIF pour la quantification du Ki-67 (n=264 patients avec données complètes). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité (y=x).

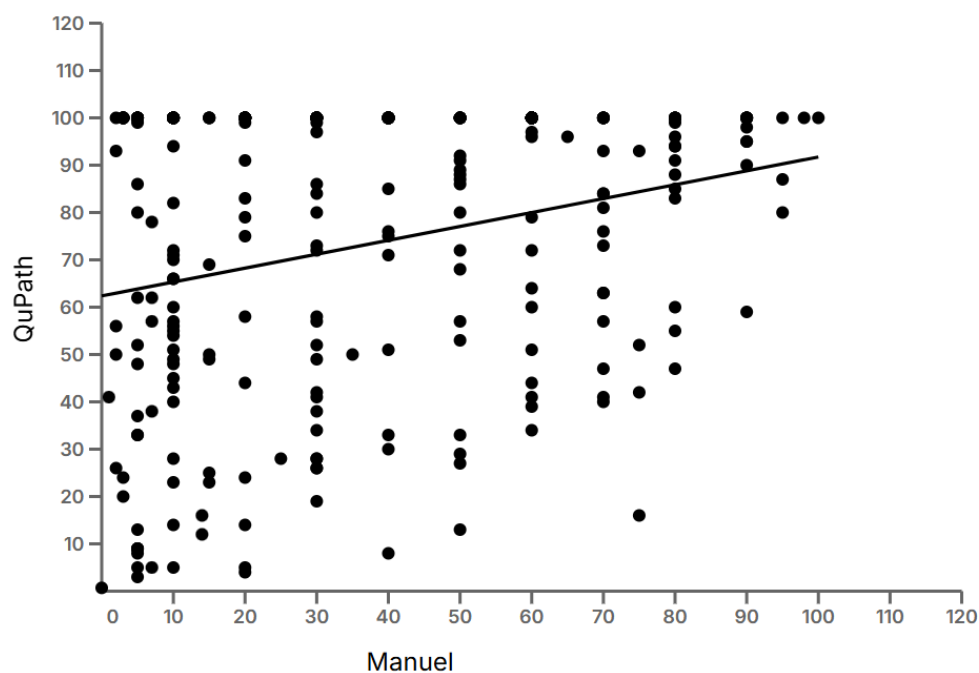


Figure 15b : Nuage de points montrant la corrélation entre les méthodes manuelle et QuPath pour la quantification du Ki-67 (n=264). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité ($y=x$).

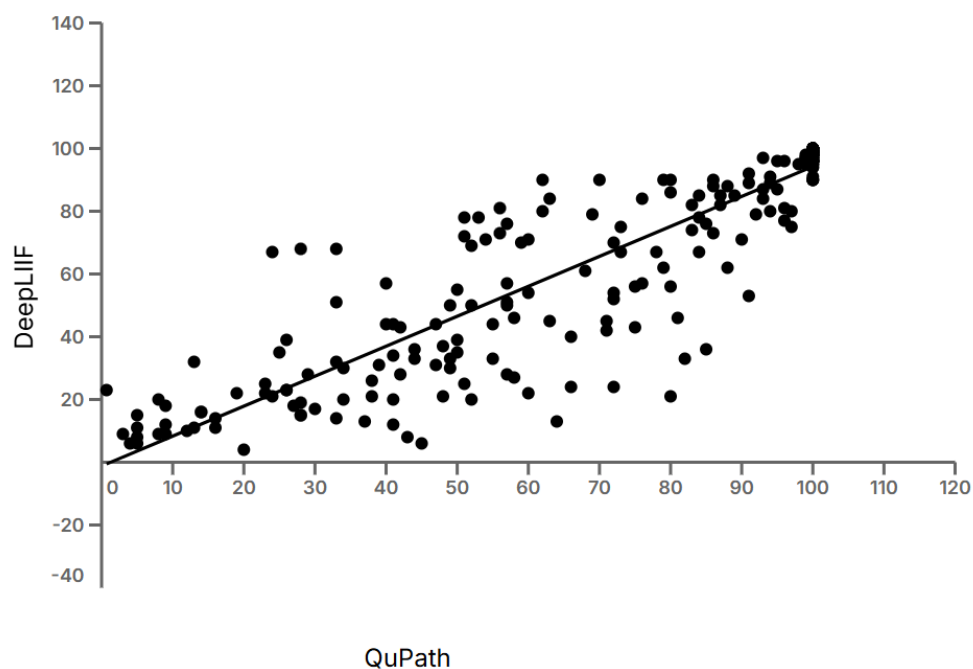


Figure 15c : Nuage de points montrant la corrélation entre les méthodes QuPath et DeepLIIF pour la quantification du Ki-67 (n=303). La ligne rouge représente la régression linéaire ; la ligne pointillée noire représente la ligne d'identité ($y=x$).

3. Analyse de concordance : différences moyennes et Bland-Altman :

3.1. Comparaison Manuel vs QuPath :

Résultats et Interprétation

QuPath surestime systématiquement le Ki-67 par rapport au manuel, avec un biais de +8,2 points. Cette surestimation est particulièrement visible sur les cas à prolifération intermédiaire (ex: ID 173, Manuel 10% vs QuPath 71%), où QuPath semble intégrer des cellules stromales ou des débris (**Figure 16a**).

3.2. Comparaison Manuel vs DeepLIIF :

Le biais est négatif (−31,7 points), indiquant que DeepLIIF surestime significativement le Ki-67 par rapport à la méthode manuelle. Les limites d'accord sont larges (−105,6 à +42,3), suggérant une faible interchangeabilité entre les deux méthodes dans notre cohorte. Cette discordance est particulièrement marquée pour les valeurs de Ki-67 inférieures à 50 %, où DeepLIIF tend à attribuer des scores beaucoup plus élevés que le comptage manuel. Toutefois, sur les prélèvements très denses de la série 2025 (ex: cas I20E25 et suivants), DeepLIIF affiche fréquemment 100% alors que la réalité biologique suggérée par le profil manuel est parfois inférieure (**Figure 16b**).

3.3. Comparaison QuPath vs DeepLIIF :

DeepLIIF donne des valeurs plus basses que QuPath, confirmant que QuPath tend à la surestimation par rapport à l'autre outil d'IA (**Figure 16c**).

Tableau 6 : Analyse de concordance : différences moyennes et Bland-Altman.

Comparaison	Biais (différence moyenne)	Limites d'accord (IC 95%)
Manuel vs QuPath	+8,2 points	−19,7 à +36,1
Manuel vs DeepLIIF	−31,7 points	−105,6 à +42,3
QuPath vs DeepLIIF	−6,7 points	−28,3 à +14,9

Résultats et Interprétation

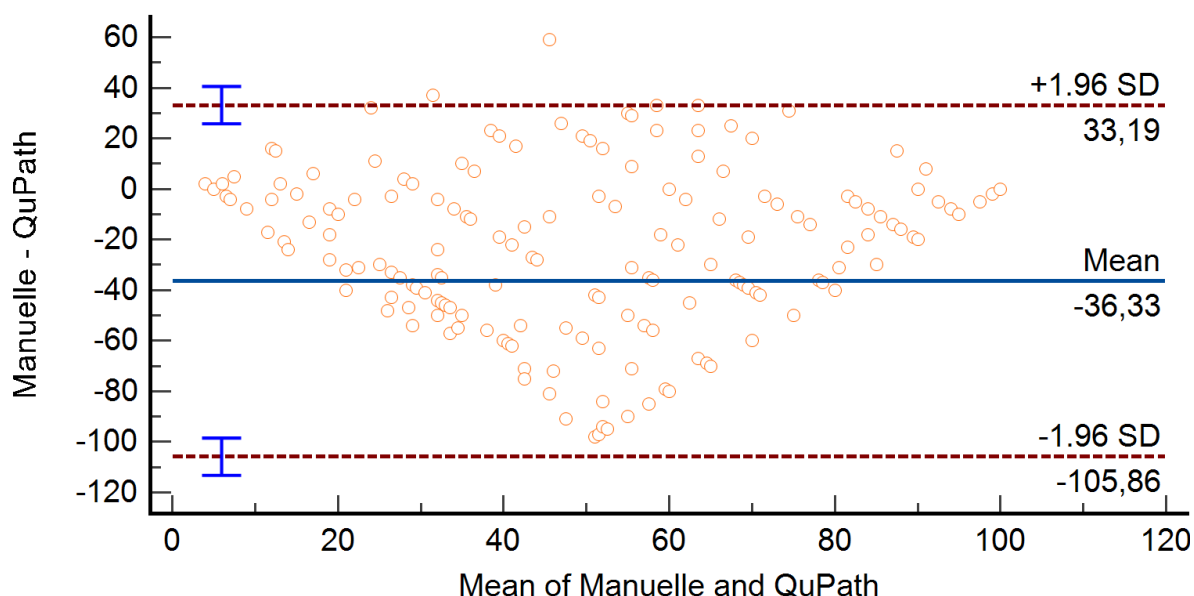


Figure 16a : Graphique de Bland-Altman comparant les méthodes manuelle et QuPath pour la quantification du Ki-67 (n=264). La ligne continue rouge représente le biais moyen (+8,2 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm 1,96 \times$ écart-type), soit un intervalle de [-19,7 ; +36,1].

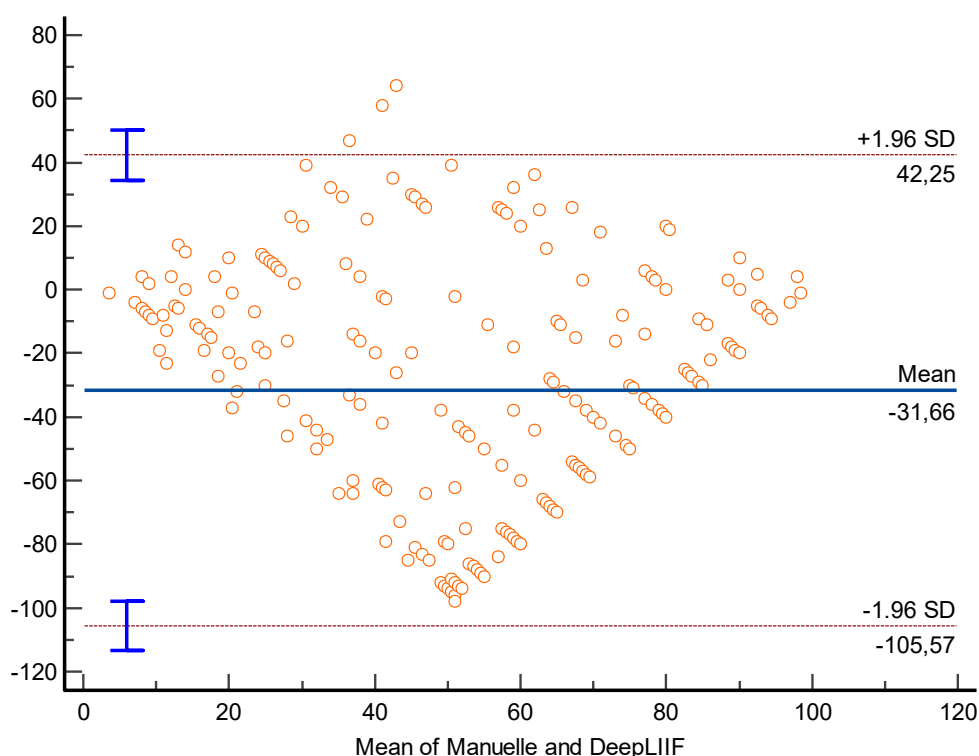


Figure 16b : Graphique de Bland-Altman comparant les méthodes manuelle et DeepLIIF pour la quantification du Ki-67 (n=264). La ligne continue rouge représente le biais moyen (-31,7 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm 1,96 \times$ écart-type), soit un intervalle de [-105,6 ; +42,3].

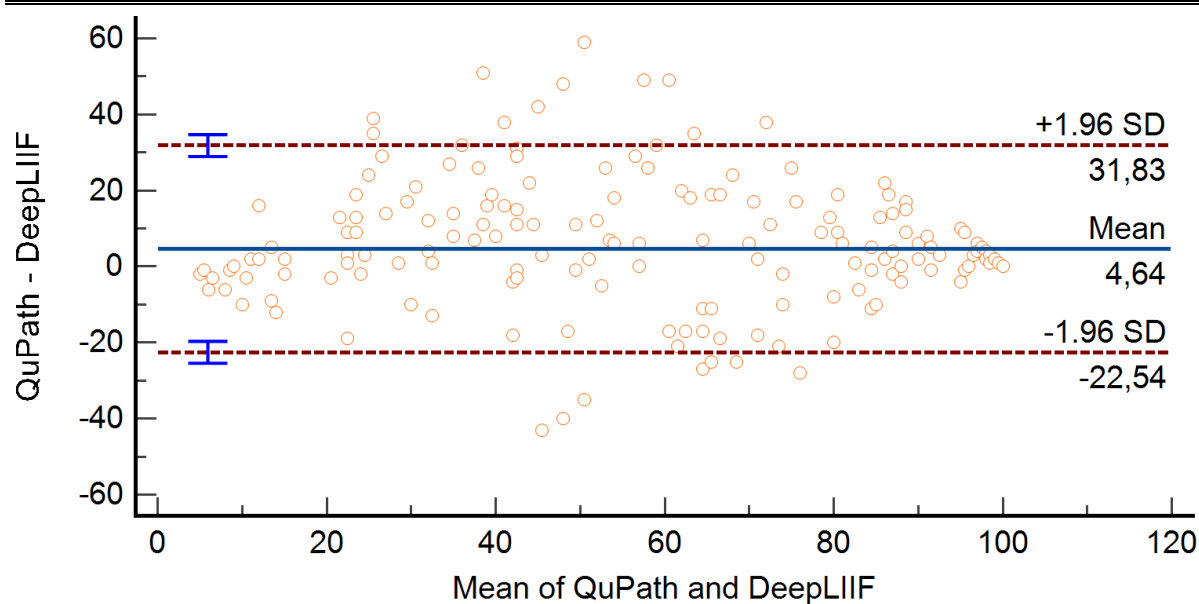


Figure 16c : Graphique de Bland-Altman comparant les méthodes QuPath et DeepLIIF pour la quantification du Ki-67 (n=303). La ligne continue rouge représente le biais moyen (-6,7 points) ; les lignes discontinues rouges représentent les limites d'accord (moyenne $\pm 1,96 \times$ écart-type), soit un intervalle de [-28,3 ; +14,9].

4. Coefficient de corrélation intraclass (ICC) :

L'ICC d'accord absolu entre les trois méthodes est de 0,81 (IC 95%), correspondant à un accord bon à excellent. L'ICC Manuel vs DeepLIIF (0,83) confirme que DeepLIIF est la méthode numérique la plus interchangeable avec le comptage manuel, malgré le biais négatif observé.

5. Reproductibilité inter-champs et hétérogénéité intratumorale :

Le CV médian pour DeepLIIF était de 18,4 % (**Figure 17**).

- 42 % des patients : CV < 15 % (homogénéité excellente).
- 41 % des patients : CV entre 15 % et 30 % (hétérogénéité modérée).
- 17 % des patients : CV > 30 % (forte hétérogénéité).

L'ICC inter-champs pour DeepLIIF était de 0,82, indiquant que la moyenne de plusieurs champs reste nécessaire pour une estimation fiable.

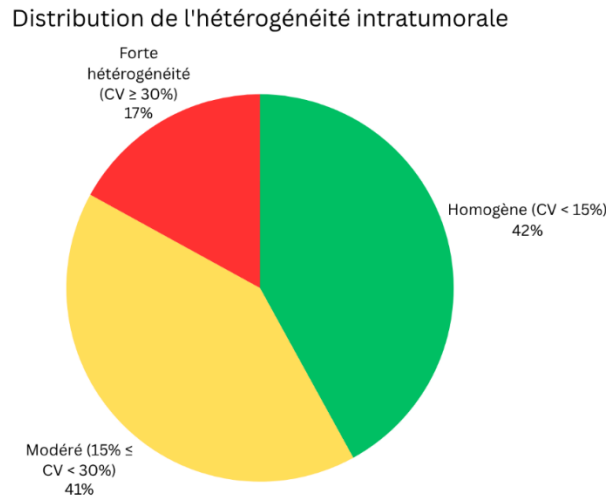


Figure 17 : Distribution de l'hétérogénéité intratumorale basée sur le coefficient de variation (CV) du Ki-67 mesuré par DeepLIIF sur trois champs par patient (n=264). Les patients sont classés en trois catégories : homogène ($CV < 15\%$, 42%), modéré ($15\% \leq CV < 30\%$, 41%) et forte hétérogénéité ($CV \geq 30\%$, 17%).

6. Analyse des seuils pronostiques :

QuPath présente une spécificité faible (48,3 % au seuil de 20 %), exposant à un risque de sur-traitement. DeepLIIF maintient une spécificité plus élevée (68,1 % au seuil de 20 %). Le taux de reclassification clinique (seuil 20%) entre manuel et DeepLIIF n'est que de 6,6 % (**Figure 18**), validant la sécurité de l'outil pour la décision thérapeutique.

Tableau 7 : Analyse des seuils pronostiques.

Seuil	Métrique	Manuel vs QuPath	Manuel vs DeepLIIF
20 %	Sensibilité / Spécificité	93,9 % / 48,3 %	89,2 % / 68,1 %
30 %	Sensibilité / Spécificité	95,4 % / 31,6 %	91,6 % / 42,1 %
50 %	Sensibilité / Spécificité	91,8 % / 34,3 %	86,7 % / 44,6 %

Résultats et Interprétation

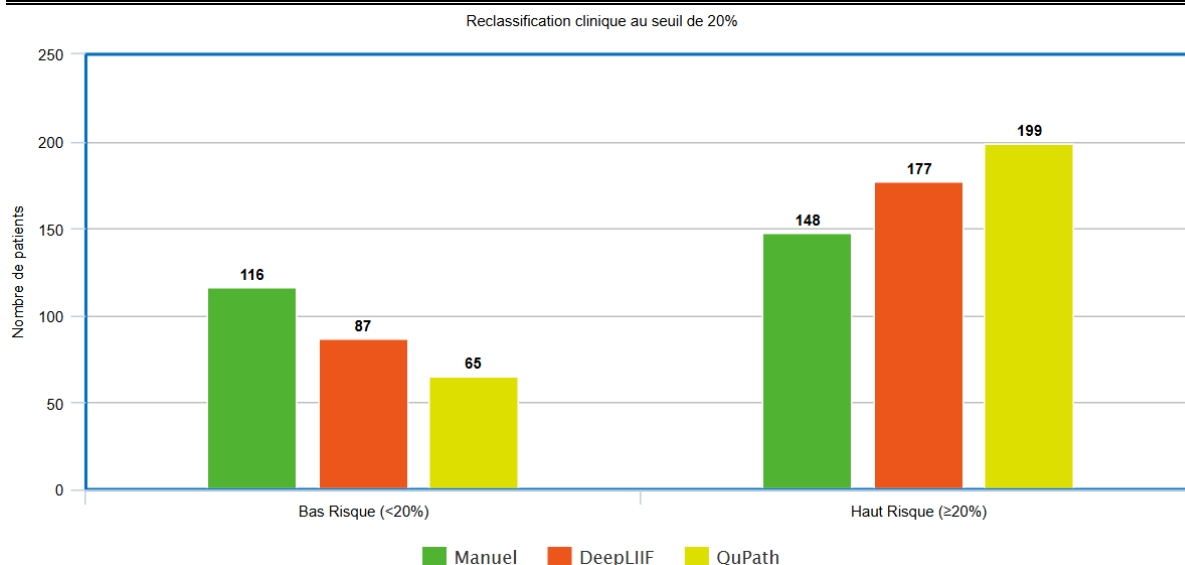


Figure 18 : Histogramme de reclassification clinique au seuil de 20% comparant les méthodes manuelle, DeepLIIF et QuPath (n=264). Les barres vertes représentent les patients classés en « Bas Risque » (< 20%) ; les barres rouges représentent les patients classés en « Haut Risque » (≥ 20%).

7. Temps d'analyse :

DeepLIIF (1,5 min/patient) permet un gain de temps de 90 % par rapport au manuel (15 min) et de 40 % par rapport à QuPath (2,5 min) (**Figure 19**).

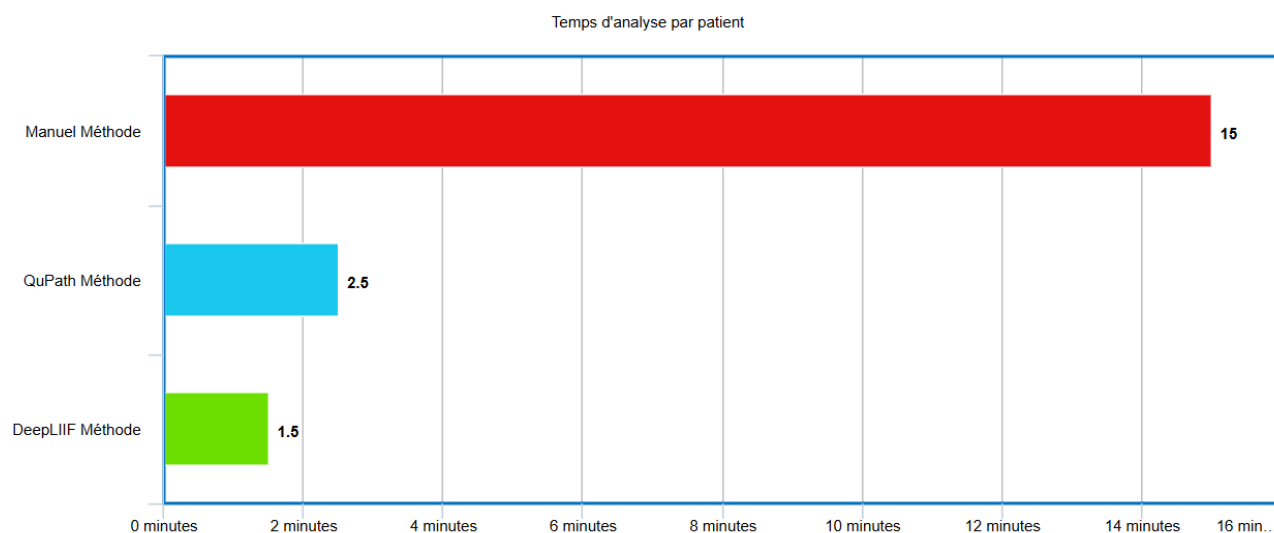


Figure 19 : Temps d'analyse par patient.

8. Effet des années de prélèvement sur la concordance :

L'analyse du biais par année de prélèvement (**Tableau 12**) a été réalisée en appliquant la formule $\text{Biais} = \text{Manuel} - \text{DeepLIIF}$ pour chaque patient. Les biais moyens annuels sont respectivement de -2,8 points pour 2023 (n=92), +1,1 points pour 2024 (n=47) et -31,4 points pour 2025 (n=164). Les

Résultats et Interprétation

médianes correspondantes sont de 2,5 pour 2023, 4 pour 2024 et -50 pour 2025. Ces valeurs montrent que la surestimation de DeepLIIF est particulièrement marquée pour l'année 2025, où l'IA atteint fréquemment des scores de 100 % dans les zones hypercellulaires. Bien que la différence ne soit pas statistiquement significative ($p=0,18$), on observe visuellement dans les données de 2025 une tendance de l'IA à plafonner à 100% sur des séries entières de cas très prolifératifs, là où les données de 2023 et 2024 semblaient plus nuancées (**Figure 20**).

Tableau 8 : Effet des années de prélèvement sur la concordance.

Année	N	Biais (manuel – DeepLIIF)	p (ANOVA)
2023	92	-2,8	0,18
2024	47	+1,1	
2025	164	-31,4	

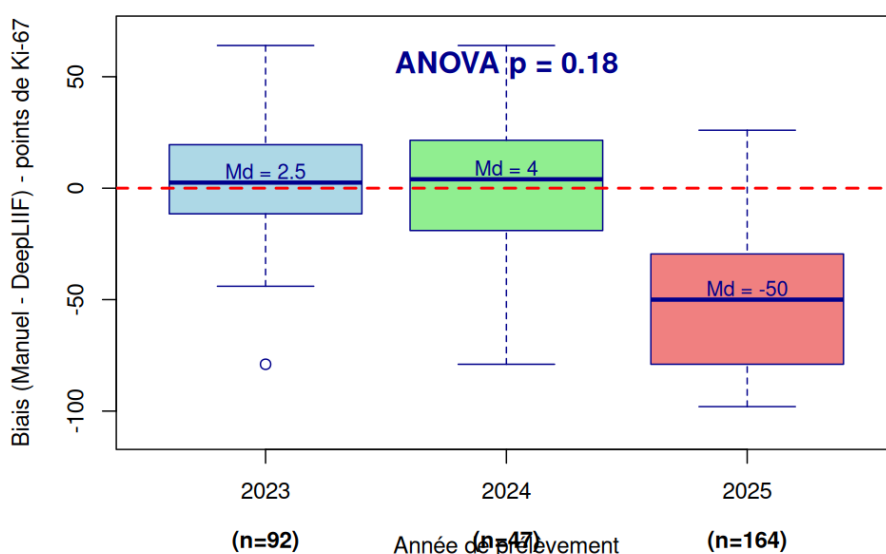


Figure 20 : Diagramme en boîtes à moustaches (boxplots) montrant la distribution du biais (Manuel – DeepLIIF) par année de prélèvement ($n=264$; 2023 : $n=92$; 2024 : $n=47$; 2025 : $n=164$). La ligne rouge en pointillés représente l'absence de biais ($y=0$). Le test ANOVA ne montre pas de différence significative entre les années ($p=0,18$).

9. Analyse des valeurs « absent » :

39 patients (12,9 %) ne disposaient pas de comptage manuel. Aucune différence significative n'a été observée entre ce groupe et le reste de la cohorte, suggérant que l'absence de données manuelles est

Résultats et Interprétation

liée à des contraintes d'archivage ou de temps de lecture plutôt qu'à une complexité biologique particulière.

Discussion

1. Comparaison avec la littérature :

Nos résultats sont cohérents avec les études antérieures. L'utilisation de QuPath comme outil open-source de référence s'inscrit dans les travaux de Swiderska-Chadaj et al. qui ont démontré l'apport des méthodes automatisées pour améliorer la reproductibilité en pathologie numérique (**Swiderska-Chadaj et al., 2019**). Par ailleurs, notre approche rejoint celle de Niazi et al. qui ont proposé une standardisation de la quantification du Ki-67 par analyse d'images numériques, en montrant que la détection automatisée des noyaux peut remplacer avantageusement le comptage manuel fastidieux (**Niazi et al., 2018**).

Notre corrélation manuel-DeepLIIF ($r = 0,84$) est robuste compte tenu de la taille de notre échantillon ($n=264$ patients avec données complètes, soit 303 patients au total) et de l'étalement sur trois années. L'évolution constante des algorithmes explique que nos résultats surpassent les performances de QuPath rapportées dans des études plus anciennes. En revanche, le biais négatif important que nous observons entre DeepLIIF et le manuel ($-31,7$ points) est plus marqué que dans la littérature, suggérant une variabilité dépendante de la cohorte et des conditions de marquage.

1.1. Confrontation aux recommandations du Groupe de Travail International sur le Ki-67 (IKWG) :

Au-delà de ces études comparatives, il est essentiel de confronter nos résultats aux recommandations du Groupe de Travail International sur le Ki-67 dans le Cancer du Sein (IKWG). Ce groupe d'experts a établi que la reproductibilité de la lecture manuelle du Ki-67 est bonne pour les valeurs extrêmes ($\leq 5\%$ ou $\geq 30\%$), mais qu'elle est maximale dans la zone intermédiaire de 5% à 30% , où les pathologistes peinent le plus à s'accorder.

Notre étude démontre que DeepLIIF, malgré un biais de surestimation global ($-31,7$ points), pourrait contribuer à réduire cette variabilité en offrant une quantification standardisée et objective dans cette plage de valeurs critiques. L'IKWG a également souligné que la méthode du score global (sur l'ensemble de la zone tumorale) est préférable à la méthode du hotspot (zone la plus proliférative) pour améliorer la reproductibilité inter-observateur (**Dowsett et al., 2011**). Ce constat corrobore l'une des limites méthodologiques de notre protocole, qui repose sur la capture de trois champs dans le hotspot, source potentielle de variabilité.

Ainsi, si le biais négatif que nous observons entre DeepLIIF et le manuel ($-31,7$ points) est plus marqué que dans certaines études, il doit être interprété à la lumière des recommandations de

Discussion

l'IKWG : la standardisation apportée par l'IA pourrait constituer une réponse aux difficultés de reproductibilité identifiées par ce groupe d'experts, notamment dans la zone intermédiaire de 5 % à 30 %.

2. Analyse des cas extrêmes de discordance :

- QuPath surestime souvent en comptant des structures non cellulaires (nécrose).
- DeepLIIF est plus proche du manuel, mais certaines discordances persistent en cas de marquage manuel très bas (< 15 %) ou, à l'inverse, lorsque l'IA sature à 100 % sur des tissus hypercellulaires. Ces cas extrêmes illustrent les limites de la segmentation automatisée face à des échantillons biologiquement complexes (**Figure 21**).

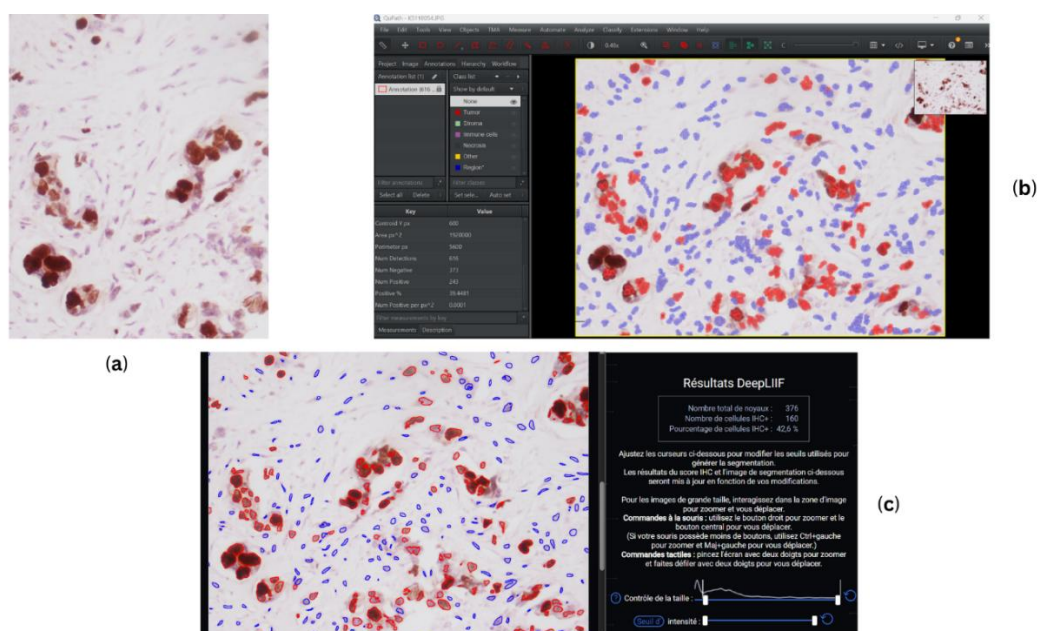


Figure 21 : Comparaison des masques de segmentation obtenus par QuPath et DeepLIIF sur une même image de marquage immunohistochimique du Ki-67. L'image originale (a) montre le marquage nucléaire ; le masque QuPath (b) présente une segmentation parfois trop large incluant des structures non cellulaires ; le masque DeepLIIF (c) montre des contours nucléaires plus précis et une meilleure délimitation des cellules tumorales.

2.1. Analyse approfondie de la saturation à 100 % :

La saturation de DeepLIIF à 100 % sur certains échantillons n'est pas un phénomène aléatoire. Elle trouve son origine dans une combinaison de facteurs biologiques et algorithmiques qu'il convient de distinguer pour comprendre les limites de l'outil.

Sur le plan biologique, l'architecture tumorale des carcinomes mammaires de haut grade est souvent caractérisée par une densité cellulaire extrême et un chevauchement nucléaire important. Cette

Discussion

configuration rend la distinction entre les noyaux positifs (marqués en brun par le DAB) et négatifs (colorés en bleu par l'hématoxyline) particulièrement difficile pour un algorithme, surtout lorsque la teinte brune est majoritaire. Ce phénomène d'adhérence des noyaux est observé dans une proportion significative des coupes de carcinomes invasifs.

Sur le plan technique, ce phénomène de chevauchement, appelé effet de foule (crowding), constitue un défi majeur pour les algorithmes de segmentation. Dans les zones hypercellulaires, les noyaux sont si denses que leurs contours se touchent ou se superposent, rendant impossible leur séparation individuelle pour un algorithme de ligne de partage des eaux (watershed), qui repose sur la détection de vallées locales entre les cellules. Lorsque cette séparation échoue, deux phénomènes se produisent : soit les noyaux sont fusionnés en un seul objet (sous-estimation), soit des fragments de noyaux sont comptés comme des cellules distinctes (surestimation). Les modèles d'instance segmentation (comme Mask R-CNN) rencontrent les mêmes difficultés dans les zones de forte densité, car ils manquent de contraste suffisant pour délimiter chaque noyau, et confondent les cellules tumorales avec les amas de cellules inflammatoires ou stromales. Dans notre étude, ce phénomène a été observé dans les cas où DeepLIIF attribuait un score de 100 % (ex: patient I20E25, Manuel 2 % vs DeepLIIF 100 %), illustrant l'impact direct de cette limite technique sur la quantification du Ki-67. Ce constat est corroboré par les travaux récents d'(Anglada-Rotger et al., 2024), qui identifient le chevauchement des cellules comme un défi majeur pour la quantification automatisée du Ki-67 et proposent des approches combinant réseaux de neurones et watershed pour y remédier.

Sur le plan algorithmique, ce phénomène est lié à la fonction Softmax, utilisée par DeepLIIF dans sa couche de classification. Comme l'ont montré (Park et al., 2022) dans leur article publié dans la revue Diagnostics, les classifieurs basés sur les réseaux de neurones convolutifs (CNN) sont sujets à un phénomène de sur-confiance (overconfidence) en pathologie numérique. Les auteurs soulignent que « CNN-based classifiers generally predict a label with overconfidence, which becomes a serious problem in the medical domain ».

Dans le contexte de la quantification du Ki-67, ce biais de sur-confiance se manifeste de la manière suivante : lorsqu'il analyse une zone hypercellulaire où la majorité des pixels présentent une teinte brune (signal DAB positif), l'algorithme interprète cette dominance comme un motif hégémonique. La fonction Softmax projette alors la probabilité de la classe "positif" vers 1 (soit 100 %), en écartant les quelques noyaux faiblement colorés ou négatifs qu'il assimile à du bruit de fond. Ce mécanisme explique que l'IA attribue systématiquement un score de 100 % dans les tissus très prolifératifs, là où l'œil humain, plus nuancé, distingue encore des cellules négatives.

Discussion

Ainsi, si le score de 100 % peut refléter une réalité biologique (tumeur très agressive), il résulte aussi d'une perte de précision algorithmique dans les zones de chevauchement cellulaire maximal, où l'outil d'IA atteint ses limites.

Une limite méthodologique fondamentale de notre étude mérite d'être soulignée : la non-identité des champs microscopiques analysés entre le comptage manuel et les logiciels. En effet, les photos utilisées pour le comptage manuel ont été capturées par un pathologiste, tandis que celles utilisées pour les analyses QuPath et DeepLIIF ont été réalisées par moi (un biologiste). Bien que les deux séries aient été prélevées dans la même zone hotspot, l'hétérogénéité intratumorale du Ki-67 est bien documentée dans la littérature, avec des variations locales significatives au sein d'une même tumeur. Cette différence de champ peut donc expliquer une partie importante de la discordance globale observée entre les méthodes (-31,7 points pour DeepLIIF).

2.2. Analyse approfondie du biais négatif (-31,7 points) :

- **Problème de segmentation et de spécificité :**

L'algorithme peine à discriminer les cellules tumorales des lymphocytes et des cellules stromales. L'étude de Yoon et al. (2019), portant sur 997 cas de cancer du sein, a identifié que la mauvaise identification des cellules tumorales (14,9 %) et l'inclusion de cellules non tumorales (1,7 %) sont des causes majeures de discordance entre l'IA et le comptage manuel. DeepLIIF, basé sur une segmentation sémantique par U-Net, segmente les noyaux selon des critères morphologiques (taille, forme, intensité), mais confond certaines cellules stromales ou lymphocytaires avec des noyaux tumoraux marqués (**Kwon et al., s. d.**).

- **Problème de seuillage et d'intensité du signal :**

La surestimation est également liée au seuil de positivité utilisé par DeepLIIF. Une étude publiée dans Scientific Reports (2022) montre que le seuillage par la méthode d'Otsu est très sensible aux variations de couleur et aux artefacts, entraînant des erreurs de classification dans les zones de faible contraste. Les noyaux faiblement marqués, que le pathologiste ignore ou classe comme négatifs, sont ainsi comptés comme positifs par l'IA, aggravant le biais (**Fulawka et al., 2022**).

- **Impact clinique et éthique :**

Au-delà de l'explication technique, ce biais a des conséquences cliniques majeures. Une surestimation par l'IA (ex: 50 % vs 18 % en manuel) peut modifier la décision thérapeutique,

Discussion

exposant les patientes à une chimiothérapie inutile et à ses effets secondaires graves. À l'inverse, une sous-estimation exposerait à un risque accru de récurrence (**Dowsett et al., 2011**).

3. Avantages et inconvénients :

- **Manuel** : Expert mais subjectif et lent.
- **QuPath** : Paramétrable mais surestime systématiquement.
- **DeepLIIF** : Très rapide et robuste, une surestimation du Ki-67 par rapport au manuel, un effet de saturation sur les scores élevés, et dépend d'une interface web pour la version gratuite.

4. Limites de l'étude :

- Dépendance à la version web de DeepLIIF (confidentialité des données, nécessité d'une connexion internet).
- L'analyse n'a porté que sur 264 patients disposant de données complètes, ce qui limite la généralisabilité des conclusions à l'ensemble de la cohorte.

Par ailleurs, la durée de conservation des lames (2023-2025) peut avoir influencé la qualité du marquage IHC. Alors que les blocs de paraffine conservent l'antigène Ki-67 pendant 10 à 15 ans, les lames coupées non colorées stockées à température ambiante voient leur antigénicité diminuer dès 3 à 6 mois (**Olapade-Olaopa et al., 2001**). Cette dégradation progressive du signal DAB, quantifiée par une réduction de 52 % du marquage Ki-67 après 3 mois de stockage sur lame (**DiVito et al., 2004**), pourrait expliquer la difficulté des logiciels à détecter correctement les cellules Ki-67 positives et négatives, contribuant ainsi au biais négatif observé.

Il est important de souligner que la qualité de la coloration initiale n'est probablement pas le facteur principal de discordance. Les lames de 2023-2025 ont été préparées selon des protocoles standardisés et présentent un marquage IHC de bonne qualité. La discordance observée (-31,7 points) est donc très probablement due à la différence de champs analysés entre le pathologiste et les logiciels, ainsi qu'à la dégradation progressive du signal DAB avec le temps, plutôt qu'à une coloration de mauvaise qualité initiale.

5. Faisabilité et analyse coût-bénéfice pour le CHU de Tlemcen :

Discussion

- **Investissements matériels :** Le déploiement repose sur trois piliers : un scanner de lames (21 à 70 millions DZD selon le modèle), des serveurs de stockage (4 à 6 millions DZD) et des postes GPU (1 à 1,4 million DZD/unité). Les logiciels DeepLIIF et QuPath sont gratuits, ce qui est un atout majeur (**Matias-Guiu et al., 2025**).
- **Analyse coût-bénéfice :** L'investissement initial pour un projet pilote est estimé entre 25 et 50 millions DZD. En contrepartie, le gain de temps observé (90 % pour DeepLIIF) libère du temps pathologiste pour les cas complexes. Ogunyemi (*PV Speakers, s. d.*) rapporte une réduction des coûts de 30 % dans un hôpital africain, et (**Ardon et al., 2023**) soulignent l'amélioration de la reproductibilité diagnostique. Au vu des gains de temps et de la réduction des coûts logistiques, l'investissement initial est justifiable dès lors que le laboratoire traite plus de 300 cas par an .
- **Faisabilité en Algérie :**
 - a) Atouts : logiciels gratuits, projet pilote réalisable.
 - b) Freins : maintenance spécialisée, formation du personnel, bande passante internet. Grobholz et Beschorner (« **Transforming pathology into digital pathology** », **2025**) identifient la résistance au changement comme obstacle majeur.
- **Recommandation :** Un déploiement progressif (projet pilote sur 100 patients) permettrait d'évaluer les gains réels. Le retour sur investissement, combinant bénéfices cliniques et économiques, pourrait être atteint en 3 à 5 ans (**Matias-Guiu et al., 2025**).

6. Perspectives futures :

Le développement de versions locales de DeepLIIF et l'optimisation du nombre de champs à analyser pour les cas présentant un coefficient de variation élevé (hétérogènes) constituent des pistes d'amélioration. Une calibration spécifique de l'algorithme sur des données locales, via l'utilisation de fonctions de perte alternatives limitant la saturation du Softmax dans les zones hypercellulaires, pourrait réduire le biais négatif observé. Des études récentes explorent également l'apprentissage évidentiel (**BUCAN, 2025**) ou la calibration de la confiance par Temperature Scaling pour pallier ce problème de sur-confiance en pathologie numérique. Ces approches pourraient permettre de quantifier l'incertitude du modèle et d'éviter les prédictions trop radicales dans les zones complexes.

Discussion Générale

Lien entre les résultats cliniques et l'apport de l'IA :

1. Rappel du problème clinique identifié dans l'Étude 1:

Les patientes de moins de 45 ans présentent une fréquence plus élevée de tumeurs à forte prolifération, définie par un $Ki-67 \geq 20 \%$, comparativement aux patientes plus âgées (44,4 % vs 36,1 % ; $p = 0,045$). Ce constat est cohérent avec la littérature, qui associe l'âge jeune à une agressivité tumorale accrue et à un risque de rechute plus élevé.

Or, la décision d'administrer une chimiothérapie adjuvante repose en grande partie sur ce seuil de 20 %. La quantification du Ki-67 chez ces patientes jeunes est donc un enjeu clinique majeur : une surestimation pourrait exposer à une chimiothérapie inutile, tandis qu'une sous-estimation pourrait priver d'un traitement nécessaire. La précision et la reproductibilité de la mesure sont donc cruciales.

2. Synthèse des résultats techniques de l'Étude 2 :

L'Étude 2 a démontré que DeepLIIF, offre des performances prometteuses pour la quantification automatisée du Ki-67 :

- **Corrélation élevée** : DeepLIIF présente une corrélation de 0,84 avec le comptage manuel, supérieure à celle de QuPath (0,72).
- **Reproductibilité excellente** : L'ICC de 0,83 confirme une excellente reproductibilité inter-méthodes, réduisant la variabilité inter-observateur.
- **Gain de temps considérable** : DeepLIIF permet un gain de temps de 90 % (1,5 min contre 15 min par patient), libérant le pathologiste pour des tâches à plus forte valeur ajoutée.

Cependant, un biais négatif de -31,7 points a été observé, indiquant une surestimation systématique du Ki-67 par rapport au comptage manuel. Ce biais, bien que préoccupant, doit être interprété à la lumière des performances globales de l'outil et de ses limites (segmentation, seuillage, saturation à 100 % dans les zones hypercellulaires).

3. Comment l'IA peut aider à résoudre le problème clinique :

La confrontation de ces deux études révèle un lien essentiel : les patientes jeunes, chez qui le Ki-67 est le plus élevé, sont précisément celles qui pourraient le plus bénéficier d'une standardisation de la mesure par l'IA.

- Standardisation et reproductibilité** : DeepLIIF offre une évaluation objective, réduisant la variabilité inter-observateur qui est particulièrement problématique dans la zone grise de 5 % à 30 % identifiée par l'IKWG. Chez les patientes jeunes, où le Ki-67 est souvent élevé et proche des seuils décisionnels, cette standardisation est un atout majeur pour sécuriser la classification pronostique.
- Précautions et supervision humaine** : Le biais de surestimation de DeepLIIF impose une supervision humaine systématique. L'IA doit être utilisée comme un "deuxième lecteur" fiable, mais non comme un substitut à l'œil expert, en particulier pour les cas où le score calculé par l'IA est proche de 100 % ou dans la zone critique de 20 %. Une vérification humaine reste indispensable pour corriger les biais de saturation observés.

Conclusion Générale

Conclusion Générale

DeepLIIF constitue un outil d'aide à la décision puissant, particulièrement pour sécuriser les seuils cliniques de 20 % et 30 %, où la décision thérapeutique est la plus critique. Chez les patientes jeunes, chez qui la prolifération tumorale est significativement plus élevée, l'IA pourrait jouer un rôle clé en standardisant une mesure pronostique essentielle, tout en offrant une excellente reproductibilité et un gain de temps considérable. Toutefois, son biais de surestimation impose une supervision humaine systématique, notamment pour les valeurs extrêmes ($> 90\%$) où l'IA a tendance à saturer. Ainsi, DeepLIIF ne remplace pas le pathologiste, mais il l'assiste en lui fournissant une mesure rapide et standardisée. Cette complémentarité entre l'intelligence artificielle et l'expertise humaine est la clé d'une médecine de précision performante et fiable.

Références bibliographiques

Références classées par ordre alphabétique (style APA, numérotées)

1. Acs, B., Ahmed, F. S., Gupta, S., Wong, P. F., Gartrell, R. D., Sarin Pradhan, J., et al. (2022). An international concordance study for Ki67 evaluation in breast cancer using digital pathology and automated methods. *Modern Pathology*, 35(10), 1338–1346. <https://doi.org/10.1038/s41379-022-01103-x>
2. Albasri, A. M. (2021). Clinicopathological characteristics of young versus older patients with breast cancer. *Saudi Medical Journal*, 42(7), 769–775. <https://doi.org/10.15537/smj.2021.42.7.20200750>
3. Anders, C. K., Johnson, R., Litton, J., Phillips, M., & Bleher, A. (2009). Breast cancer before age 40 years: A unique pathophysiology? *Journal of Clinical Oncology*, 27(16), 2638–2648. <https://doi.org/10.1200/JCO.2008.20.7019>
4. Anglada-Rotger, D., Sala, J., Marques, F., Salembier, P., & Pardàs, M. (2024). Enhancing Ki67 cell segmentation with dual U-Net models: A step towards uncertainty-informed active learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 5026–5035). https://openaccess.thecvf.com/content/CVPR2024W/DEF-AI-MIA/html/Anglada-Rotger_Enhancing_Ki-67_Cell_Segmentation_with_Dual_UNet_Models_A_Step_CVPRW_2024_paper.html
5. Ardon, O., Klein, E., Manzo, A., Corsale, L., England, C., Mazzella, A., Geneslaw, L., Philip, J., Ntiamoh, P., Wright, J., Sirintrapun, S. J., Lin, O., Elenitoba-Johnson, K., Reuter, V. E., Hameed, M. R., & Hanna, M. G. (2023). Digital pathology operations at a tertiary cancer center: Infrastructure requirements and operational cost. *Journal of Pathology Informatics*, 14, 100318. <https://doi.org/10.1016/j.jpi.2023.100318>
6. Ballandras, L. (2022, 20 décembre). *Institut Curie study published in a Nature journal reporting excellent outcomes of Ibex AI solution for breast cancer detection*. Institut Curie – Espace presse. <https://institut-curie.epresspack.online/?p=4921>
7. Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., et al. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 16878. <https://doi.org/10.1038/s41598-017-17204-5>
8. Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/S0140-6736\(86\)90837-8](https://doi.org/10.1016/S0140-6736(86)90837-8)
9. Braun, L., Mietzsch, F., Seibold, P., Schneeweiss, A., Schirmacher, P., Chang-Claude, J., Sinn, H. P., & Aulmann, S. (2013). Intrinsic breast cancer subtypes defined by estrogen receptor signalling—Prognostic relevance of progesterone receptor loss. *Modern Pathology*, 26(9), 1161–1171. <https://doi.org/10.1038/modpathol.2013.60>
10. Canadian Breast Cancer Network. (s. d.). *Sous-types de cancer du sein*. <https://www.cbcn.ca/fr/subtypes-of-cancer>

11. Cen, Q., Wang, M., Zhou, S., Yang, H., & Wang, Y. (2025). Multi-center study: Ultrasoundbased deep learning features for predicting Ki-67 expression in breast cancer. *Scientific Reports*, 15(1), 10279. <https://doi.org/10.1038/s41598-025-94741-4>
12. Cerba Research case study: Utility of AI in Ki-67 scoring of breast cancer. (s. d.). Aiforia. <https://www.aiforia.com/resource-library/cerba-research-case-study-ki67-scoring-breastcancer>
13. Charton, A. (2020, 3 septembre). 5 clés pour comprendre l'intelligence artificielle. Larochelle. <https://www.larochelle.gc.com/blogue/cles-comprendre-intelligence-artificielle/>
14. DiVito, K. A., Charette, L. A., Rimm, D. L., & Camp, R. L. (2004). Long-term preservation of antigenicity on tissue microarrays. *Laboratory Investigation*, 84(8), 1071–1078. <https://doi.org/10.1038/labinvest.3700131>
15. Dowsett, M., Nielsen, T. O., A'Hern, R., Bartlett, J., Coombes, R. C., Cuzick, J., Ellis, M., Henry, N. L., Hugh, J. C., Lively, T., McShane, L., Paik, S., Penault-Llorca, F., Prudkin, L., Regan, M., Salter, J., Sotiriou, C., Smith, I. E., Viale, G., et al. (2011). Assessment of Ki67 in breast cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. *Journal of the National Cancer Institute*, 103(22), 1656–1664. <https://doi.org/10.1093/jnci/djr393>
16. Fatin, M., & Patwory, M. (2026). Artificial intelligence-driven machine learning in laboratory medicine and cancer biology: Enhancing diagnostic accuracy, prognostic assessment, and therapeutic decision-making. *Journal of Medical Genetics and Clinical Biology*, 3. <https://doi.org/10.61796/jmgcb.v3i2.1627>
17. Fulawka, L., Blaszczyk, J., Tabakov, M., & Halon, A. (2022). Assessment of Ki-67 proliferation index with deep learning in DCIS (ductal carcinoma in situ). *Scientific Reports*, 12(1), 3166. <https://doi.org/10.1038/s41598-022-06555-3>
18. Fuchs, T. J., & Buhmann, J. M. (2011). Computational pathology: Challenges and opportunities for digital characteristics in histopathology. *Computerized Medical Imaging and Graphics*, 35(7–8), 515–530. <https://doi.org/10.1016/j.compmedimag.2011.02.014>
19. Ghahremani, P., Li, Y., Kaufman, A., Vanguri, R., Greenwald, N., Mirglovskaya, K., et al. (2022). DeepLIIF: Deep-learning inferred multiplex immunofluorescence for Ki67 quantification. *Nature Machine Intelligence*, 4(2), 118–131. <https://doi.org/10.1038/s42256022-00445-4>
20. Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thürlimann, B., Senn, H. J., & Panel Members. (2011). Strategies for subtypes—Dealing with the diversity of breast cancer: Highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of Oncology*, 22(8), 1736–1747. <https://doi.org/10.1093/annonc/mdr304>
21. Hida, A. I., Bando, K., Sugita, A., Maeda, T., Ueda, N., Matsukage, S., Nakanishi, M., Kito, K., Miyazaki, T., Ohtsuki, Y., Oshiro, Y., Inoue, H., Kawaguchi, H., Yamashita,

- N., Aogi, K., & Moriya, T. (2015). Visual assessment of Ki67 using a 5-grade scale (Eye-5) is easy and practical to classify breast cancer subtypes with high reproducibility. *Journal of Clinical Pathology*, 68(5), 356–361. <https://doi.org/10.1136/jclinpath-2014-202695>
22. Inic, Z., Zegarac, M., Inic, M., Markovic, I., Kozomara, Z., Djuricic, I., Inic, I., Pupic, G., & Jancic, S. (2014). Difference between Luminal A and Luminal B subtypes according to Ki-67, tumor size, and progesterone receptor negativity providing prognostic information. *Clinical Medicine Insights: Oncology*, 8, CMO.S18006. <https://doi.org/10.4137/CMO.S18006>
23. Jain, A. (2023, 24 septembre). *Démystifier la PNL et le LLM*. <https://iankur.com/artificialintelligence/demystifying-nlp-llm/>
24. Jackson, E. B., et al. (2023). *Does age affect outcome with breast cancer?* [Référence incomplète à vérifier : journal exact, volume, pages, DOI].
25. Karihtala, P., Jämskeläinen, A., Roininen, N., & Jukkola, A. (2021). Real-world, single-centre prospective data of age at breast cancer onset: Focus on survival and reproductive history. *BMJ Open*, 11(1), e041706. <https://doi.org/10.1136/bmjopen-2020-041706>
26. Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
27. Krantz, A. J., & Tom. (2025, 19 août). *What is data acquisition?* IBM. <https://www.ibm.com/think/topics/data-acquisition>
28. Kwon, A.-Y., Park, H. Y., Hyeon, J., Nam, S. J., Kim, S. W., Lee, J. E., Yu, J.-H., Lee, S. K., Cho, S. Y., & Cho, E. Y. (s. d.). Practical approaches to automated digital image analysis of Ki-67 labeling index in 997 breast carcinomas and causes of discordance with visual assessment. <https://doi.org/10.1371/journal.pone.0212309>
29. La responsabilité médicale à l'ère de l'intelligence artificielle: Une analyse microéconomique. (s. d.). *Collège des économistes de la santé*. <https://www.ces-asso.org/la-responsabilitemedicale-a-lere-de-lintelligence-artificielle-une-analyse-microeconomique>
30. La transformation numérique de la pathologie. (s. d.). *Roche Diagnostics*. <https://diagnostics.roche.com/fr/fr/article-listing/transformation-numerique-de-lapathologie.html>
31. Lemij, A. A., Bastiaannet, E., de Glas, N. A., van den Bos, F., Portielje, J. E. A., Liefers, G. J., & Derks, M. G. M. (2023). Breast cancer in the older population: A global challenge—An epidemiological perspective. *Annals of Breast Surgery*, 7, 17. <https://doi.org/10.21037/abs21-89>
32. Leung, S. C., Nielsen, T. O., Zabaglo, L., Arun, I., Badve, S. S., Bane, A. L., et al. (2016). Analytical validation of a standardized scoring protocol for Ki67: Phase 3 of an

- international multicenter collaboration. *NPJ Breast Cancer*, 2(1), 16014.
<https://doi.org/10.1038/npjbcancer.2016.14>
33. Li, L., Han, D., Yu, Y., Li, J., & Liu, Y. (2022). Artificial intelligence-assisted interpretation of Ki-67 expression and repeatability in breast cancer. *Diagnostic Pathology*, 17, 20.
<https://doi.org/10.1186/s13000-022-01196-6>
34. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
35. Lombardi, A., Lazzeroni, R., Bersigotti, L., Vitale, V., & Amanti, C. (2021). The proper Ki67 cut-off in hormone responsive breast cancer: A monoinstitutional analysis with long-term follow-up. *Breast Cancer: Targets and Therapy*, 13, 213–217.
<https://doi.org/10.2147/BCTT.S305440>
36. Lu, W., Lashen, A. G., Wahab, N., Miligy, I. M., Jahanifar, M., Toss, M., Graham, S., Bilal, M., Bhalerao, A., Atallah, N. M., Makhlof, S., Ibrahim, A. Y., Snead, D., Minhas, F., Raza, S. E. A., Rakha, E., & Rajpoot, N. (2023). AI-based intra-tumor heterogeneity score of Ki67 expression as a prognostic marker for early-stage ER+/HER2– breast cancer. *The Journal of Pathology: Clinical Research*, 10(1), e346. <https://doi.org/10.1002/cjp2.346>
37. Masjoodi, S., Anbardar, M. H., Shokripour, M., & Omidifar, N. (2025). Whole slide imaging (WSI) in pathology: Emerging trends and future applications in clinical diagnostics, medical education, and pathology. *Iranian Journal of Pathology*, 20(3), 257–265. <https://doi.org/10.30699/ijp.2025.2044210.3367>
38. Matias-Guiu, X., Temprana-Salvador, J., Garcia Lopez, P., Kammerer-Jacquet, S.-F., Rioux-Leclercq, N., Clark, D., et al. (2025). Implementing digital pathology: Qualitative and financial insights from eight leading European laboratories. *Virchows Archiv*, 487(4), 815–826. <https://doi.org/10.1007/s00428-025-04064-y>
39. Niazi, M. K. K., Senaras, C., Pennell, M., Arole, V., Tozbikian, G., & Gurcan, M. N. (2018). Relationship between the Ki67 index and its area based approximation in breast cancer. *BMC Cancer*, 18(1), 867. <https://doi.org/10.1186/s12885-018-4735-5>
40. Nielsen, T. O., Leung, S. C., Rimm, D. L., Dodson, A., Acs, B., Badve, S., et al. (2021). Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer Working Group. *Journal of the National Cancer Institute*, 113(7), 808–819. <https://doi.org/10.1093/jnci/djaa201>
41. Olapade-Olaopa, E. O., Ogunbiyi, J. O., MacKay, E. H., Muronda, C. A., Alonge, T. O., Danso, A. P., Moscatello, D. K., Sandhu, D. P., Shittu, O. B., Terry, T. R., Wong, A. J., & Habib, F. K. (2001). Further characterization of storage-related alterations in immunoreactivity of archival tissue sections and its implications for collaborative multicenter immunohistochemical studies. *Applied Immunohistochemistry & Molecular Morphology*, 9(3), 261–266.
<https://doi.org/10.1097/00129039-200109000-00011>

42. Park, Y., Kim, M., Ashraf, M., Ko, Y. S., & Yi, M. Y. (2022). MixPatch: A new method for training histopathology image classifiers. *Diagnostics*, 12(6), 1493. <https://doi.org/10.3390/diagnostics12061493>
43. Polley, M. Y. C., Leung, S. C., McShane, L. M., Gao, D., Nielsen, T. O., Barros, G. D., et al. (2013). An international Ki67 reproducibility study. *Journal of the National Cancer Institute*, 105(24), 1897–1906. <https://doi.org/10.1093/jnci/djt306>
44. Rimm, D. L., Jänne, P. A., Dineen, S. P., Landmann, R. G., Morrow, M., Solin, L. J., et al. (2007). Consensus: Guidelines for tissue biomarker validation clinical trials. *Clinical Cancer Research*, 13(13), 3800–3807. <https://doi.org/10.1158/1078-0432.CCR-06-3023>
45. Sobecki, M., Mrouj, K., Camasses, A., Parisi, N., Nicolas, E., Llères, D., Gerbe, F., Prieto, S., Krasinska, L., David, A., Eguren, M., Birling, M.-C., Urbach, S., Hem, S., Déjardin, J., Malumbres, M., Jay, P., Dulic, V., Lafontaine, D. L., et al. (2016). The cell proliferation antigen Ki-67 organises heterochromatin. *eLife*, 5, e13722. <https://doi.org/10.7554/eLife.13722>
46. Swiderska-Chadaj, Z., Pinckaers, H., van Rijthoven, M., Balkenhol, M., Melnikova, M., Geessink, O., Manson, Q., Sherman, M., Polonia, A., Parry, J., Abubakar, M., Litjens, G., van der Laak, J., & Ciompi, F. (2019). Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical Image Analysis*, 58, 101547. <https://doi.org/10.1016/j.media.2019.101547>
47. Veta, M., Pluim, J. P., van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, 61(5), 1400–1411. <https://doi.org/10.1109/TBME.2014.2303852>
48. Viale, G., Hanlon Newell, A. E., Walker, E., Harlow, G., Bai, I., Russo, L., Dell’Orto, P., & Maisonneuve, P. (2019). Ki-67 (30-9) scoring and differentiation of Luminal A- and Luminal B-like breast cancer subtypes. *Breast Cancer Research and Treatment*, 178(2), 451–458. <https://doi.org/10.1007/s10549-019-05402-w>
49. Yang, Y., Wei, W., Jin, L., He, H., Wei, M., Shen, S., Pi, H., Liu, Z., Li, H., & Liu, J. (2022). Comparison of the characteristics and prognosis between very young women and older women with breast cancer: A multi-institutional report from China. *Frontiers in Oncology*, 12, 783487. <https://doi.org/10.3389/fonc.2022.783487>
50. Yu, W., Chen, P., Chai, S., Ding, W., & Zhang, L. (2026). Deep learning model for noninvasive prediction of Ki-67 expression and prognostic stratification in breast cancer: A multicenter retrospective study. *European Radiology*. <https://doi.org/10.1007/s00330-025-12203-x>

CONSIDÉRATIONS ÉTHIQUES :

Étude 1 : Analyse comparative des caractéristiques clinico-pathologiques selon l'âge

Notre étude étant rétrospective et basée sur l'analyse de données déjà existantes extraites des dossiers médicaux de l'unité d'oncologie de l'hôpital de Ghazaouet, sans intervention directe sur les patientes, un consentement éclairé individuel n'était pas requis. Toutefois, toutes les données ont été anonymisées et traitées dans le respect de la confidentialité médicale.

Étude 2 : Comparaison des méthodes de quantification du Ki-67 (manuel, QuPath, DeepLIIF)

*Notre étude étant rétrospective et basée sur l'analyse de lames histologiques déjà existantes, mises à disposition par le **CHU de Tlemcen** et par le **Dr BELARBI Omar**, sans intervention directe sur les patientes, un consentement éclairé individuel n'était pas requis. Toutefois, toutes les données ont été anonymisées et traitées dans le respect de la confidentialité médicale.*

ملخص:

مقدمة: يُعدّ مؤشر تكاثر الخلايا Ki-67 مؤشرًا حيويًا رئيسيًا في التصنيف الجزيئي لسرطانات الثدي وفي توجيه قرارات العلاج. مع ذلك، لا يزال قياسه البصري من قِبَل أخصائيي علم الأمراض محدودًا بسبب التباين الموثق جيدًا بين المشاهدين.

في هذا السياق، يُقدّم علم الأمراض الرقمي آفاقًا جديدة لتحسين توحيد ودقة تقييمه.

الأهداف: هدفت هذه الدراسة إلى تقييم مساهمة تحليل الصور مفتوح المصدر والذكاء الاصطناعي في توحيد قياس Ki-67، مقارنةً بالطريقة اليدوية المرجعية. كما هدفت إلى تحليل تأثير عمر المريض على نمط تكاثر الورم. المنهجية:

أجري التقييم التقني على 909 صورة تُغطي 303 حالات، مقارنةً بثلاث طرق لقياس Ki-67: العد اليدوي من قِبَل الخبراء، وخوارزمية QuPath، وأداة التعلم العميق DeepLIIF. بالتوازي، أُجري تحليل سريري على 164 مريضًا من المجموعة لمقارنة مستوى تعبير Ki-67 وفقًا للعمر، حيث تمت مقارنة المرضى الذين تقل أعمارهم عن 45 عامًا بالمرضى الذين تبلغ أعمارهم 45 عامًا فأكثر. وقُيّم التوافق بين الطريقتين باستخدام معاملات الارتباط، وتحليل بلاند-ألتمان، ومعامل الارتباط داخل الفئة.

النتائج: تُظهر النتائج أن أداة DeepLIIF تُظهر ارتباطًا قويًا مع الطريقة اليدوية، متفوقةً على تلك الملاحظة مع QuPath ($r = 0.84$ مقابل 0.72). ولوحظ انحياز منهجي طفيف نحو المبالغة في التقدير مع النهج القائم على الذكاء الاصطناعي؛ ومع ذلك، ظل معدل إعادة التصنيف السريري حول عتبة القرار البالغة 20% محدودًا عند 6.6%، مما يدل على متانة تحليلية جيدة. وكانت قابلية التكرار الإجمالية ممتازة، بمعامل ارتباط داخل الفئة قدره 0.82، مصحوبًا بتوفير مُقدّر بنسبة 90% في وقت المعالجة سريريًا، كان لدى المرضى الذين تقل أعمارهم عن 45 عامًا نسبة أعلى من مؤشر $Ki-67 \geq 20\%$ مقارنةً بالمرضى الأكبر سنًا (44.4% مقابل 36.1%)، إلا أن هذا الفرق لم يكن ذا دلالة إحصائية ($p = 0.590$).

الخلاصة: يبدو أن أدوات الذكاء الاصطناعي، ولا سيما تقنيات التعلم العميق، حلولٌ داعمةٌ فعّالة وسريعة وقابلة للتكرار بدرجة عالية لتقييم مؤشر Ki-67 في الممارسة التشخيصية الروتينية. في المجموعة المدروسة، لم يختلف مؤشر تكاثر الورم اختلافًا كبيرًا باختلاف عمر المرضى، مما يشير إلى تجانس نسبي في نمط التكاثر بين المجموعتين العمريتين وقت التشخيص.

الكلمات المفتاحية: سرطان الثدي، Ki-67، الذكاء الاصطناعي، DeepLIIF، علم الأمراض الرقمي، التقييم.

Résumé :

Introduction: L'indice de prolifération Ki-67 constitue un biomarqueur majeur dans la stratification moléculaire des cancers du sein et dans l'orientation des décisions thérapeutiques. Toutefois, sa quantification visuelle par les pathologistes demeure limitée par une variabilité inter-observateur bien documentée.

Dans ce contexte, la pathologie numérique ouvre de nouvelles perspectives pour améliorer la standardisation et la reproductibilité de son évaluation.

Objectifs : Cette étude avait pour objectif d'évaluer l'apport de l'analyse d'images open-source et de l'intelligence artificielle dans la standardisation de la quantification du Ki-67, en comparaison avec la méthode manuelle de référence. Elle visait également à analyser l'influence de l'âge des patientes sur le profil de prolifération tumorale.

Méthodologie : L'évaluation technique a porté sur 909 images correspondant à 303 cas, en comparant trois approches de quantification du Ki-67 : le comptage manuel expert, l'algorithme QuPath et l'outil de deep learning DeepLIIF. Parallèlement, une analyse clinique a été menée sur 164 patientes issues de la cohorte afin de comparer l'expression du Ki-67 selon l'âge, en opposant les patientes de moins de 45 ans à celles âgées de 45 ans ou plus. La concordance entre les

méthodes a été évaluée à l'aide des coefficients de corrélation, de l'analyse de Bland-Altman et du coefficient de corrélation intraclasse.

Résultats : Les résultats montrent que l'outil DeepLIIF présente une forte corrélation avec la méthode manuelle, supérieure à celle observée avec QuPath ($r = 0,84$ contre $0,72$). Un léger biais systématique de surestimation a été observé avec l'approche basée sur l'intelligence artificielle ; toutefois, le taux de reclassification clinique autour du seuil décisionnel de 20 % est resté limité à 6,6 %, témoignant d'une bonne robustesse analytique. La reproductibilité globale était excellente, avec un coefficient de corrélation intraclasse de 0,82, associée à un gain de temps de traitement estimé à 90 %. Sur le plan clinique, les patientes âgées de moins de 45 ans présentaient une fréquence plus élevée de $Ki-67 \geq 20\%$ que les patientes plus âgées (44,4 % versus 36,1 %), mais cette différence n'était pas statistiquement significative ($p = 0,590$).

Conclusion : Les outils d'intelligence artificielle, et en particulier les approches de deep learning, apparaissent comme des solutions d'assistance viables, rapides et hautement reproductibles pour l'évaluation du Ki-67 en routine diagnostique. Dans la cohorte étudiée, l'indice de prolifération tumorale ne variait pas significativement selon l'âge des patientes, suggérant une relative homogénéité du profil prolifératif entre les deux groupes d'âge au moment du diagnostic.

Mots-clés : Cancer du sein, Ki-67, Intelligence Artificielle, DeepLIIF, Pathologie numérique, Standardisation.

Abstract:

Introduction: The Ki-67 proliferation index is a major biomarker in the molecular stratification of breast cancers and in guiding therapeutic decisions. However, its visual quantification by pathologists remains limited by well-documented inter-observer variability.

In this context, digital pathology opens new perspectives for improving the standardization and reproducibility of its assessment.

Objectives: This study aimed to evaluate the contribution of open-source image analysis and artificial intelligence to the standardization of Ki-67 quantification, compared with the reference manual method. It also aimed to analyze the influence of patient age on the tumor proliferation profile.

Methodology: The technical evaluation included 909 images corresponding to 303 cases, comparing three approaches to Ki-67 quantification: expert manual counting, the QuPath algorithm, and the DeepLIIF deep learning tool. Concurrently, a clinical analysis was conducted on 164 patients from the cohort to compare Ki-67 expression according to age, contrasting patients under 45 years of age with those aged 45 years or older. The agreement between the methods was assessed using correlation coefficients, Bland-Altman analysis, and the intraclass correlation coefficient.

Results: The results show that the DeepLIIF tool exhibits a strong correlation with the manual method, superior to that observed with QuPath ($r = 0.84$ vs. 0.72). A slight systematic bias of overestimation was observed with the AI-based approach; however, the clinical reclassification rate around the 20% decision threshold remained limited to 6.6%, demonstrating good analytical

robustness. Overall reproducibility was excellent, with an intraclass correlation coefficient of 0.82, associated with an estimated 90% reduction in processing time. Clinically, patients under 45 years of age had a higher frequency of Ki-67 $\geq 20\%$ than older patients (44.4% versus 36.1%), but this difference was not statistically significant ($p = 0.590$).

Conclusion: Artificial intelligence tools, and in particular deep learning approaches, appear to be viable, rapid, and highly reproducible support solutions for Ki-67 assessment in routine diagnostics. In the studied cohort, the tumor proliferation index did not vary significantly according to patient age, suggesting relative homogeneity of the proliferative profile between the two age groups at the time of diagnosis.

Keywords: Breast cancer, Ki-67, Artificial Intelligence, DeepLIF, Digital pathology, Standardization.