



HAL
open science

The first data challenge of the french society of pathology: An international competition in 2020, a research tool in A.I. for the future?

Agathe Delaune, Séverine Valmary-Degano, Nicolas Loménie, Karim Zryouil, Nesrine Benyahia, Olivier Trassard, Virginie Eraville, Christine Bergeron, Mojgan Devouassoux-Shisheboran, Claire Glaser, et al.

► To cite this version:

Agathe Delaune, Séverine Valmary-Degano, Nicolas Loménie, Karim Zryouil, Nesrine Benyahia, et al.. The first data challenge of the french society of pathology: An international competition in 2020, a research tool in A.I. for the future?. *Annales de Pathologie*, 2022, 42 (2), pp.119-128. 10.1016/j.annpat.2021.10.002 . hal-04511964

HAL Id: hal-04511964

<https://hal.science/hal-04511964>

Submitted on 22 Jul 2024

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Titre : Le premier data challenge organisé par la Société Française de Pathologie: une compétition internationale en 2020, un outil de recherche en intelligence artificielle pour l'avenir ?

Title : The first data challenge of the french society of pathology : an international competition in 2020, a research tool in A.I. for the future ?

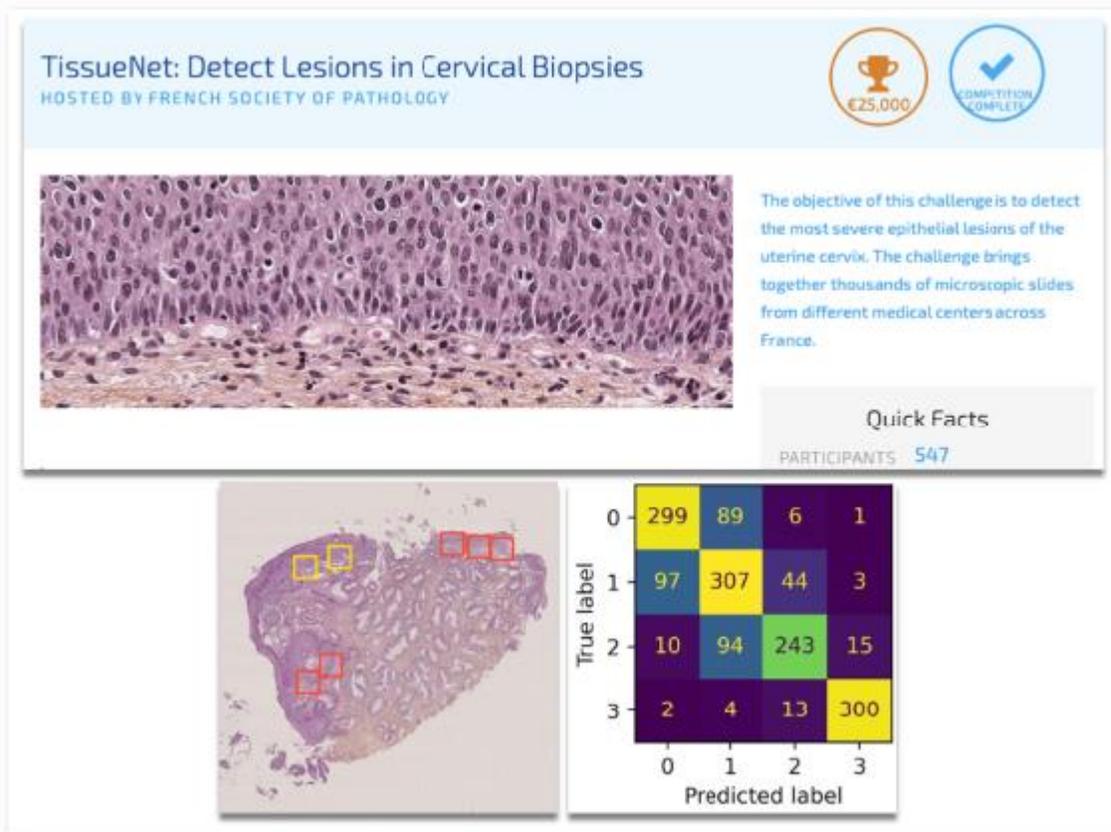
Auteurs :

Agathe Delaune, Séverine Valmary-Degano, Nicolas Loménie, Karim Zryouil, Nesrine Benyahia, Olivier Trassard, Virginie Eraville, Christine Bergeron, Mojgan Devouassoux-Shisheboran, Claire Glaser, Guillaume Bataillon, Emmanuel Bacry, Stéphanie Combes, Sophie Prevot, Philippe Bertheau

- Agathe Delaune, Plateforme de données de santé - Health Data Hub, 9 rue Georges Pitard 75015 Paris.
- Séverine Valmary-Degano, [Pathologie](#), Université Grenoble Alpes, Inserm U1209, CNRS UMR5309, Institute for Advanced Biosciences, CHU Grenoble, 38000, France.
- Nicolas Loménie, LIPADE, UFR Mathématiques-Informatiques, Université de Paris, 45 rue des Saints-Pères, 75006 Paris.
- Karim Zryouil, Capgemini Invent, 147 quai du Président Roosevelt, 92130 Issy-les-Moulineaux.
- Nesrine Benyahia, Dr Data, 81 Rue Réaumur, 75002 Paris.
- Olivier Trassard, Pathologie / Inserm , CHU Bicêtre, APHP, 78 Rue du Général Leclerc, 94270 Le Kremlin-Bicêtre.
- Virginie Eraville, Société Française de Pathologie.
- Christine Bergeron, CerbaPath, 16 bis rue Odessa, Paris 75014.
- Mojgan Devouassoux-Shisheboran, Centre de Pathologie Sud des Hospices Civils de Lyon, Centre Hospitalier Lyon Sud, 165 chemin du grand Revoyet 69495 Pierre Bénite Cedex.
- Claire Glaser, Pathologie, CHG Versailles, 177 Rue de Versailles, 78150 Le Chesnay-Rocquencourt.
- Guillaume Bataillon, Pathologie, Institut Curie, 26 r Ulm, 75005 Paris.
- Emmanuel Bacry, Plateforme de données de santé - Health Data Hub, 9 rue Georges Pitard 75015 Paris et CNRS, Université Paris Dauphine, PSL.
- Stéphanie Combes, Plateforme de données de santé - Health Data Hub, 9 rue Georges Pitard 75015 Paris.
- Sophie Prevot, Pathologie, CHU Bicêtre, APHP, 78 Rue du Général Leclerc, 94270 Le Kremlin-Bicêtre.
- Philippe Bertheau, Pathologie, CHU Saint-Louis, APHP, 1 avenue Claude Vellefaux 75010 Paris, et Université de Paris.

Auteur correspondant : Philippe Bertheau, 0142494135, philippe.bertheau@gmail.com

Graphical abstract



Résumé en français :

La Société Française de Pathologie (SFP) a organisé en 2020 son premier data challenge avec le soutien du Health Data Hub (HDH). La préparation de cet évènement a d'abord comporté une phase de recrutement de près de 5000 lames de biopsies du col utérin auprès de 20 centres volontaires. Après s'être assuré de l'absence d'opposition des patientes, les lames ont été anonymisées, numérisées, annotées par des pathologistes et finalement mises à disposition de compétiteurs du monde entier sur une plateforme dédiée. Les équipes en compétition devaient développer des algorithmes permettant de distinguer les 4 classes diagnostiques des lésions épithéliales du col utérin. Plus de 500 compétiteurs ont développé plusieurs dizaines d'algorithmes et les meilleurs d'entre eux ont atteint des scores proche de 95% de reconnaissance des lésions. Les 3 meilleures équipes se sont partagées 25000€ de prix à l'occasion d'une session pendant Carrefour Pathologie Digital 2020. Au-delà de cette compétition dont la partie finale s'est tenue sur une période de quelques semaines, la SFP et le HDH travaillent maintenant au maintien en ligne et en accès ouvert sur le long terme de cette collection afin de permettre à la communauté des data scientists et des pathologistes de poursuivre leurs progrès en intelligence artificielle dans ce champ thématique.

Abstract in english :

The french society of pathology (SFP) organized in 2020 its first data challenge with the help of Health Data Hub (HDH). The organisation of this event first consisted in recruiting almost 5000 slides of uterus cervical biopsies obtained in 20 pathology centers. After having made sure that patients did not refuse to include their slides in the project, the slides were anonymised, digitized and annotated by expert pathologists, and were finally uploaded on a data challenge platform for competitors all around the world. Competitors teams had to develop algorithms that could distinguish among four diagnostic classes in epithelial lesions of uterine cervix. Among many submissions by competitors, the best algorithms obtained an overall score close to 95%. The best 3 teams shared 25k€ prizes during a special session organised during the national congress of the SFP. The final part of the competition lasted only 6 weeks and the goal of SFP and HDH is now to allow for the collection to be published in open access. This final step will allow data scientists and pathologists to further develop artificial intelligence algorithms in this medical area.

Mots clefs en français

intelligence artificielle, data challenge, lames virtuelles, col utérin

Keywords in english

artificial intelligence, data challenge, whole slide images, uterine cervix

Texte

1/ Pourquoi un data challenge ?

Les champs d'application de l'intelligence artificielle (IA) s'étendent un peu plus chaque jour notamment dans le monde de la santé, domaine prometteur compte-tenu de la masse de données générées pour le diagnostic et le soin [1–3]. Que ce soit pour le suivi intelligent du patient par des dispositifs connectés, pour le choix du meilleur traitement chez un patient en fonction de plusieurs sources de données diagnostiques, ou pour l'optimisation des processus analytiques en biologie ou en imagerie, les applications se multiplient, portées par de multiples acteurs, académiques et/ou industriels. De nombreux dispositifs médicaux utilisant l'IA sont déjà approuvés aux Etats-Unis et en Europe [4].

Les avancées dans le domaine de l'image en général et dans celui de l'imagerie médicale en particulier [5] sont portées par la capacité des réseaux de neurones à classer des images après une phase d'apprentissage plus ou moins supervisée, mais parfois aussi sans apprentissage préalable, mettant en évidence des liens entre les images et le pronostic d'une maladie ou la réponse à un traitement. En pathologie, des applications prometteuses [6,7] et des recommandations sur l'IA en pathologie [8] commencent à voir le jour.

Développer un algorithme d'IA en santé nécessite une forte expertise informatique et statistique, mais aussi des données en quantité et de qualité, représentatives le plus possible de toutes les situations réelles [1]. L'accès à des données de santé représentatives et richement annotées est à l'heure actuelle le principal facteur limitant au développement d'algorithmes d'IA.

Un data challenge est une vraie opportunité pour réunir à la fois de multiples expertises en sciences des données mais aussi une vaste collection de données de qualité. D'une part, le fait d'organiser une compétition avec une récompense va mobiliser toute une communauté de développeurs et de data scientists. D'autre part, cette large mobilisation d'une communauté scientifique stimule en amont l'implication des producteurs de données. C'est la conjonction de ces deux dynamiques qui permettra d'obtenir les meilleurs algorithmes. Même si les algorithmes développés au cours d'un data challenge ont encore de nombreuses étapes à franchir avant une utilisation validée en pratique médicale, ils apportent déjà la première preuve de concept de leur intérêt potentiel.

Des data challenges autour des images en santé ont déjà été organisés ces dernières années. On peut citer ceux des radiologues français depuis 2018 [9,10], et les Camelyon 16 et 17 en pathologie [11]. Une revue fait le point sur les data challenges en pathologie et sur ce qu'apportent ces compétitions dans notre discipline [12].

Pour le data challenge 2020 organisé par la Société Française de Pathologie (SFP) et le Health Data Hub (HDH), la perspective d'une compétition internationale autour de Carrefour Pathologie 2020 a permis de mobiliser la communauté des pathologistes pour produire une importante base

de données de lésions du col utérin. Nous détaillons dans cet article la construction et le déroulement de cette compétition et nous en discutons les perspectives.

2/ La structure de gestion et le financement du data challenge

C'est à l'automne 2019 qu'ont eu lieu les premières réunions entre la SFP et le HDH concernant le projet de data challenge. Ce projet était innovant à la fois pour notre discipline mais aussi pour le HDH dont la création était récente. Le HDH est un groupement d'intérêt public créé par la Loi du 24 juillet 2019 relative à l'organisation et la transformation du système de santé. Il associe 56 parties prenantes, en grande majorité issues de la puissance publique (CNAM, CNRS, France Assos Santé...) et met en œuvre les grandes orientations stratégiques relatives au Système National des Données de Santé fixées par l'Etat et notamment le ministère des Solidarités et de la Santé. Le HDH est un service à destination de l'écosystème de santé et a pour objectif de promouvoir l'innovation en santé et l'accessibilité des données et des connaissances par le biais, entre autres, d'événements fédérateurs, comme l'organisation de data challenges.

La première étape du projet de data challenge a été de constituer un comité d'organisation et un conseil scientifique en décembre 2019 (Tableau 1). Ces deux comités étaient composés de pathologistes ayant différents modes d'exercice représentatifs de la discipline mais également de data scientists, de statisticiens, de juristes et de méthodologistes. Cette gouvernance multidisciplinaire a été déterminante pour la bonne avancée du projet.

Début 2020, le HDH a recruté un chef de projet dédié à l'organisation du data challenge, puis la SFP s'est dotée d'un responsable de la protection des données.

Outre la SFP pour l'organisation médicale, et le HDH pour le soutien logistique, informatique et juridique, le data challenge a reçu un soutien financier du Grand Défi du Conseil de l'Innovation « Amélioration des diagnostics médicaux par l'intelligence artificielle » [13]□.

3/ La question du data challenge 2020

. Choix de la question

La première mission du conseil scientifique a été de définir la question posée pour le challenge. Classiquement, notamment en imagerie radiologique, les questions se distribuent dans une des trois catégories suivantes : segmentation, détection ou classification.

Un algorithme de segmentation permet, sur une image, de distinguer une structure par rapport à une autre, par exemple sur une image radiologique le tissu musculaire par rapport au tissu adipeux, ou sur une lame histopathologique les cellules tumorales infiltrantes par rapport au stroma tumoral, ou bien la reconnaissance des mitoses dans une tumeur. Ce travail de segmentation automatique permet un gain de temps pour le radiologue ou le pathologiste pour l'analyse ultérieure des images.

Un algorithme de détection des lésions permet de mettre en évidence des lésions qui auraient pu passer inaperçues à l'observation humaine, telles que de faibles différences de contraste sur des clichés radiologiques, ou des micrométastases sur une lame d'histopathologie. L'algorithme a ici une valeur ajoutée pour diminuer le risque de faux négatif d'une observation visuelle seule.

Un algorithme de classification, en général développé à partir de réseaux de neurones profonds, permet de classer des images parmi plusieurs catégories diagnostiques ou pronostiques voire théranostiques. Il en est ainsi par exemple d'un algorithme qui permettra, sur une série de lames histopathologiques, de faire un pré-tri entre des lames probablement tumorales et des lames probablement non tumorales. Ce type d'algorithme de classification pourra faire gagner du temps au pathologiste et fiabiliser encore son diagnostic. Au-delà de l'aide fournie au pathologiste, d'autres algorithmes aideront à la prise en charge des patients à partir d'informations extraites de nos lames : il s'agira par exemple de classer des images de tumeurs en fonction d'une évolution clinique, ou en fonction d'une réponse à un traitement. Certains algorithmes de classification aideront alors à identifier de nouveaux facteurs pronostiques ou prédictifs encore méconnus. Cette mise en évidence de messages cachés dans les images médicales est un enjeu majeur des algorithmes d'IA dans les disciplines d'imagerie et en pathologie.

Pour cette première édition du data challenge SFP-HDH, le choix a été fait d'une question de classification permettant d'assister le pathologiste dans son diagnostic des lésions épithéliales du col utérin. Cette pathologie a été choisie parce qu'elle est d'une fréquence élevée en pratique diagnostique et n'a pas encore été étudiée sous l'angle d'une approche diagnostique par IA.

La question posée aux algorithmes pour ce data challenge était donc leur capacité à classer des lames de biopsie du col utérin dans une des quatre catégories de diagnostic de l'OMS [14] de la catégorie "normal" jusqu'au carcinome infiltrant, en passant par les lésions de bas grade ou de haut grade (Tableau 2). Les algorithmes devaient identifier sur chaque lame uniquement la classe diagnostique la plus grave, l'existence de lésions de classe moins grave sur la même lame n'étant pas à identifier dans le cadre de la compétition.

. Choix de la métrique et impact médical

Définir la question posée lors d'un data challenge impose de réfléchir à la manière dont les résultats des algorithmes seront évalués, à savoir comment seront mesurés les écarts entre les diagnostics proposés par l'algorithme et les diagnostics de référence des pathologistes ("vérité de terrain", voir plus bas). Pour chaque lame de la compétition, l'algorithme devait choisir une des 4 classes diagnostiques et ce choix était confronté avec la vérité de terrain, l'écart par rapport à cette référence étant plus ou moins pénalisé selon une grille qui a été discutée entre les pathologistes et les data scientists du projet. Plusieurs points de discussion ont été soulevés : devions-nous considérer plus grave de se tromper entre une classe 0 et 1 qu'entre une classe 2 et 3 ? souhaitons-nous pénaliser différemment un faux négatif et un faux positif ? quels étaient les risques que les algorithmes obtiennent de bons scores uniquement sur la base d'un calcul de probabilité et non sur l'analyse réelle des lames ? devions-nous nous préoccuper du caractère discriminant du mode de calcul et que faire en cas d'ex-aequo ?

Une grille de pondération des erreurs de l'algorithme (Tableau 3) a été utilisée pour construire les scores finaux des algorithmes à l'issue du data challenge et pour récompenser les meilleurs compétiteurs. Toutefois, d'autres systèmes de notation auraient pu être envisagés, tel qu'un système binaire avec par exemple une note maximale lorsque la bonne classe était identifiée et une pénalité maximale lorsque la bonne classe n'était pas identifiée. Ce type de grille de correction "correct vs incorrect" est beaucoup plus pénalisant qu'une grille pondérée car il ne

tolère aucun écart par rapport à la vérité de terrain. Le conseil scientifique a préféré retenir le principe d'une grille pondérée tolérant des écarts plus ou moins graves. Les coefficients de pondération de la grille retenue pour le data challenge auraient pu être différents selon le contexte médical dans lequel on décide de se placer, par exemple si on se situe dans un contexte de dépistage de masse des lésions du col utérin (quelques diagnostics en excès n'ont pas de conséquences graves) ou au contraire dans un contexte de diagnostic à visée thérapeutique (un seul diagnostic en excès peut avoir des conséquences graves). Les choix en matière de notation des algorithmes sont donc un élément important de discussion médicale en amont de la compétition, mais une fois que ce choix est fait et annoncé aux compétiteurs dans le règlement du data challenge, il faut s'y tenir.

4/ La collecte des cas du data challenge

. Appel à manifestation d'intérêt

Une fois la question du data challenge identifiée, l'étape suivante a été de rassembler suffisamment de données provenant elles-mêmes de suffisamment de structures afin de garantir une grande diversité d'images représentatives de la réalité de nos recrutements, pour ainsi augmenter la puissance des résultats des algorithmes et éviter des biais analytiques. Pour cela, un appel à manifestations d'intérêt a été diffusé par la messagerie de la SFP début mars 2020. Vingt structures réparties sur le territoire national ont répondu à cet appel, centres hospitalo-universitaires, centres hospitaliers généraux, centres de lutte contre le cancer, hôpitaux des armées, et laboratoires privés.

. Encadrement juridique et MR004

En parallèle, le délégué à la protection des données de la SFP, en anglais DPO pour Data Protection Officer, engageait la SFP devant la CNIL à suivre les recommandations de la méthodologie de référence (MR) 004. En bref, la MR-004 encadre les traitements de données à caractère personnel à des fins de recherche n'impliquant pas la personne humaine, en particulier les études portant sur la réutilisation de données. Parmi les différents engagements pris par la SFP, il y avait bien sûr l'information aux patientes, évoquée plus bas, et la dé-identification complète des données. Une analyse de risques de ré-identification des patientes a également été réalisée en collaboration avec le HDH.

. Les conventions

Plusieurs documents juridiques réalisés conjointement par la SFP et les juristes du HDH ont été envoyés aux pathologistes des centres inclueurs avec notamment des modèles de convention entre SFP et centres inclueurs, et le modèle de lettre d'information aux patientes. Une convention entre SFP et HDH a également été mise en place.

. Document d'information aux patientes

Ce document a été envoyé par voie postale à toutes les patientes qui disposaient de trois semaines pour y répondre si elles souhaitaient s'opposer à leur participation au projet. Parmi près de 5000 patientes contactées, moins d'une trentaine ont demandé à ne pas participer.

. Le travail des centres inclueurs

Le travail des centres inclueurs a pu démarrer dès la mi-mars 2020. Le premier confinement du printemps 2020 n'a pas trop pénalisé cette phase de recrutement grâce à l'implication de chacun dans les centres inclueurs. Comme quasiment aucune des lames du data challenge n'était déjà numérisée à l'origine, il a fallu désarchiver les lames physiquement. Les cas désarchivés avaient été choisis sur les systèmes informatiques des laboratoires pour respecter dans la mesure du possible un équilibre entre les 4 classes diagnostiques, tout en vérifiant la qualité des lames. Après s'être assuré de la non opposition des patientes, il a fallu effectuer l'anonymisation des lames à l'aide d'étiquettes d'anonymisation préparées par la SFP. Ces étiquettes portaient un numéro sans aucune clef de retour possible aux noms des patientes. Les lames anonymisées ont été numérisées par les centres disposant d'un scanner de lames ou par 2 centres référents (Bicêtre, Saint-Louis) pour les centres ne disposant pas d'un scanner. Le format de numérisation a été débattu et ce sont finalement les formats natifs des scanners qui ont été retenus. Toutefois, sur la plate-forme finale de compétition, les lames ont également été proposées dans un format homogénéisé de type tif. Une fois numérisées, les lames ont pu être uploadées par les centres sur un serveur sécurisé du HDH ou copiées sur un support de stockage acheminé physiquement jusqu'au HDH. Les lames virtuelles étaient uploadées sans aucun diagnostic associé, tous les diagnostics ayant été refaits ensuite par les experts pathologistes. Seule la notion du centre inclueur et la notion du scanner utilisé faisaient partie des métadonnées accompagnant chaque lame virtuelle au moment de l'upload vers le HDH.

Ces étapes d'inclusion des cas se sont étalées jusqu'au mois de juin 2020. La SFP a proposé à tous les centres inclueurs une indemnisation par lame désarchivée et par lame scannée.

5/ Annotation des cas par les experts

. Mise en ligne de cas sur la plate-forme Cytomine

Les lames virtuelles ont été déposées sur un serveur du HDH afin de pouvoir y être annotées par les experts pathologistes. Le logiciel de visualisation et annotation des lames virtuelles utilisé était le logiciel open-source Cytomine.org (<https://cytomine.be/>) qui a été installé sur les serveurs du HDH, après avoir fait l'objet de quelques développements par Cytomine.coop pour l'adapter au processus d'annotations par les pathologistes.

. Gestion du processus d'annotations

4934 lames ont été incluses et mises en ligne sur les serveurs du HDH, incluant 3709 biopsies du col utérin et 1225 conisations. Ces lames de conisation n'ont pas été incluses dans la compétition finale mais pourront être utilisées ultérieurement afin d'enrichir la banque d'images à disposition des chercheurs. Parmi les 3709 lames de biopsies, 2542 ont été relues par les 5 experts pathologistes (SP, CB, MD, GB, CG) qui ont indiqué la classe diagnostique la plus grave présente sur chaque lame.

Suite à cette phase d'annotation diagnostique, les 2542 lames ont été distribuées dans 3 groupes distincts : 1015 lames pour le jeu d'apprentissage, 513 lames pour le jeu de test et 1014 lames pour le jeu de validation finale. La répartition entre les différents ensembles d'apprentissage et de validation correspond aux standards des architectures neuronales profondes actuelles,

permettant aux algorithmes non seulement un apprentissage correct mais aussi une phase de validation finale suffisamment discriminante pour classer les algorithmes entre eux. Pour chaque jeu de lames, la proportion de chacune des quatre classes diagnostiques devait être aussi proche que possible de 25% afin d'équilibrer les classes et optimiser ainsi l'apprentissage des algorithmes.

Sur les 1015 lames du jeu d'apprentissage, les experts ont ensuite ajouté des annotations graphiques plus précises : des carrés de 300 microns de côté étaient déposés non seulement sur les lésions de classe diagnostique la plus grave, mais aussi sur des lésions de classe diagnostique moins grave si elles étaient présentes, des carrés de couleurs différentes étant utilisés pour distinguer les 4 classes diagnostiques (Figure 1). Un maximum de 10 carrés pour une même classe était déposé sur chaque lame, totalisant 5941 annotations graphiques sur ces 1015 lames.

Toutes les 2542 lames annotées pour la compétition ont été revues successivement par deux experts pathologistes. Quelques cas ont nécessité que les deux pathologistes s'accordent lors de réunions de consensus, conduisant à l'exclusion de certains cas jugés trop difficiles ou de qualité technique insuffisante.

Les 1015 lames finement annotées ont constitué le jeu de lames d'apprentissage du data challenge, les annotations graphiques déposées par les experts pouvant être utilisées comme références des 4 classes lésionnelles par les compétiteurs.

Les lames non annotées ont aussi été mises à disposition des compétiteurs comme données supplémentaires d'apprentissage s'ils le souhaitaient. Certaines lames ont été supprimées lors de la phase d'annotation car elles présentaient des défauts rendant leur analyse impossible

6/ La phase de compétition

. Export des cas auprès de Driven Data

Début septembre 2020, les lames virtuelles ont été exportées par File Transfer Protocol depuis le HDH vers les serveurs de Driven Data, la société hébergeant la compétition. Cela a représenté près de 5 To de données.

Les compétiteurs avaient pu se préinscrire dès le mois de juillet.

Pendant les quelques journées avant l'ouverture du site, de nombreuses vérifications ont été faites par Driven Data, le HDH et la SFP sur la qualité des données et leur accessibilité, afin que l'expérience des compétiteurs soit optimale.

La page web d'accueil du data challenge a été ouverte aux compétiteurs le 9 septembre 2020. Cette page et l'ensemble des caractéristiques du data challenge sont toujours accessibles sur ce lien: <https://www.drivendata.org/competitions/67/competition-cervical-biopsy/> (Figure 2).

. L'expérience des compétiteurs

Les 7 semaines de compétition qui ont suivi ont demandé aux compétiteurs de mobiliser très rapidement un ensemble de compétences et de ressources informatiques. Après le téléchargement des lames du jeu d'apprentissage, les compétiteurs devaient d'abord se familiariser avec ces images de taille très importante et avec les annotations déposées par les experts.

Les lames d'apprentissage ont été analysées et exploitées pour éduquer les algorithmes avec les ressources informatiques des compétiteurs. Ces algorithmes pouvaient ensuite être testés sur le jeu de 513 lames de test, permettant à chaque compétiteur d'avoir une idée de son degré d'avancement par rapport aux autres compétiteurs grâce à un classement sur un "leaderboard intermédiaire" disponible pendant toute la durée du challenge. Le classement final, affiché sur le "leaderboard final", était quant à lui calculé sur le jeu de 1014 lames de validation. Les compétiteurs soumettaient leur algorithme sur la plateforme Driven Data sous la forme d'un package et les calculs pour les jeux de test et de validation finale étaient effectués directement avec les processeurs graphiques (GPU, Graphics Processing Unit) de la plateforme.

. La clôture et les résultats

La compétition a été fermée le 29 octobre 2020. Les résultats bruts lame par lame ne sont pas donnés ici mais une synthèse pour les deux premiers compétiteurs est illustrée par leurs deux matrices de confusion Figure 3. La Figure 4 montre le haut du classement (leaderboard) avec les 4 premiers compétiteurs, venant respectivement, du 1er au 4ème, de France, du Kenya, du Canada et de Chine. Un bel exemple de compétition internationale ! Les scores obtenus avec la métrique officielle sont également indiqués Figure 4. Avec d'autres métriques, par exemple très stricte (métrique "correcte"), ou plus souple (métrique admettant une classe d'écart comme réponse exacte), les résultats auraient pu être un peu différents mais l'interprétation de scores qui "auraient été obtenus" est en fait très délicate, car si une autre métrique avait été choisie, il est très probable que les compétiteurs auraient optimisé leurs algorithmes pour cette autre métrique. Comparer des métriques après-coup n'a en fait pas de sens.

Une analyse plus poussée des méthodes utilisées par les compétiteurs et des résultats de leurs algorithmes est en cours et fera l'objet d'un article scientifique complémentaire.

Après la clôture de la compétition, Driven Data a contacté les gagnants en novembre 2020 afin d'obtenir une brève description de leur algorithme et leur proposer de participer à la session de remise des prix pendant le congrès Carrefour Pathologie 2020 Digital. Des prix d'un montant total de 25000 € étaient offerts aux trois meilleurs compétiteurs, à la condition que leur algorithme soit publié en open source.

. La remise des prix et la communication post challenge

Les 4 premiers compétiteurs au classement ont laissé sur le site du challenge quelques éléments décrivant leur solution, accessibles ici : <https://www.drivendata.co/blog/tissuenet-cervical-biopsies-winners/>

Le 1er au classement n'a pas souhaité recevoir le prix et les 2ème, 3ème et 4ème ont reçu respectivement 12000€, 8000€ et 5000€. Leurs algorithmes ont été ouverts en open source et déposés sur GitHub, ici : <https://github.com/drivendataorg/tissuenet-cervical-biopsies>

La remise des prix s'est déroulée en direct pendant Carrefour Pathologie Digital le 24 novembre 2020.

Voici les liens vers deux vidéos réalisées par le HDH et la SFP en octobre 2020 et vers les vidéos réalisées lors de Carrefour Pathologie sur les "coulisses du data challenge" et la remise des prix:

- Vidéo SFP-HDH (10mn) : <https://www.youtube.com/watch?v=Ue0Lt1RKaAE>

- Teaser SFP-HDH (2mn) : https://www.youtube.com/watch?v=rasEAOP_-WU
- Vidéos de Carrefour Pathologie 2020 : coulisses du data challenge (30mn) et session de remise des prix (55mn) : <https://www.youtube.com/watch?v=1fCQjb7oyP4&t=0s>

7/ Perspectives :

. Vers une collection de lames virtuelles ouvertes ?

Parmi les précédents data challenge en pathologie, le Camelyon 16 puis 17 [11] ont été des précurseurs par la qualité des données annotées (lames de cancers du sein et de ganglions métastatiques) et par le nombre de compétiteurs. Ces collections sont toujours ouvertes sur le site web du Camelyon, pour un usage maintenant pérenne dans l'intérêt commun. La SFP et le HDH travaillent également à une pérennisation de l'accès aux données du data challenge 2020. Le data challenge était à peine clôturé que déjà plusieurs compétiteurs sollicitaient le comité d'organisation pour connaître les modalités qui leur permettraient de poursuivre leurs travaux d'algorithmie sur ces données. L'accès à des données de santé de qualité et richement annotées reste en effet la principale difficulté rencontrée par les développeurs en IA. Les lames du data challenge ont fait l'objet d'un encadrement juridique dans le contexte du data challenge, chaque compétiteur s'engageant à n'utiliser les données que pour concourir au challenge et strictement pendant la période du challenge. Publier les résultats obtenus pendant le challenge est tout à fait possible, sous réserve d'y associer la SFP et le HDH.

Pour le long terme, la SFP et le HDH sont en train de compléter l'encadrement juridique des lames auprès des centres inclueurs afin d'aller autant que possible vers une collection pérenne et ouverte à l'usage de tous les médecins et chercheurs en IA. Cette étape supplémentaire passe par une étude confirmant que le risque de réidentification d'une patiente à partir d'une lame est quasi-inexistant. Cette étude maintenant finalisée, la SFP et le HDH reprennent contact avec les centres inclueurs pour avancer vers cette ouverture. Cette collection pourrait alors devenir une référence internationale en matière de lésions du col utérin, facilitant la mise au point d'algorithmes qui un jour pourront encore améliorer la qualité de la prise en charge des patientes.

. Vers d'autres data challenges de la SFP

Le data challenge 2020 a été une première expérience passionnante pour tous ceux qui y ont participé, une expérience aux multiples facettes que quelques chiffres peuvent résumer (Tableau 4).

Au moment où ces lignes sont écrites, la SFP a répondu à un nouvel appel à projets de data challenges publié par le HDH en mars 2021 et est en attente de la réponse. La SFP est donc prête à poursuivre cette aventure, et réfléchit aussi à s'associer à une autre discipline. Un challenge multidisciplinaire permet en effet de faire émerger encore davantage le potentiel de l'IA en matière d'analyses de données croisées et multi-échelles. Comme toute nouvelle percée technologique et médicale majeure, l'IA en santé a encore beaucoup de chemin à parcourir, mais son potentiel est certain, avec assurément de nombreuses découvertes à venir. Les pratiques des pathologistes vont suivre ces évolutions technologiques, mais c'est à eux de les apprivoiser et de participer à leurs développements, en y apportant leur indispensable expertise.

Remerciements

- Nos sincères remerciements à l'ensemble des patientes ayant accepté de participer à ce projet.
- A tous les compétiteurs, que l'on ne peut tous lister ici, seulement les 4 premiers : Saima Ben Hadj / Brice Tayart (Société Tribvn Healthcare, France), Raphael Kiminya (Kenya), Kirill Brodt (Canada), Feng Ming (Chine).
- Aux membres du Comité d'Organisation, du Conseil Scientifique et du Comité de relecture (voir Tableau 1)
- Aux personnels du Health Data Hub qui ont participé aux mises au point informatiques et aux discussions juridiques
- A Olivier Clatz pour ses conseils avisés et son soutien
- A la Société Française de Pathologie, son assistante, son CA et ses membres
- A Boris Hejblum, collègue de la Société Française de Statistique
- Aux pathologistes ayant contribué à inclure des cas : Louis Vaquier, Isabelle Goubin-Versini, Maguy Cherfan, Jean-Baptiste Souraud, Maxime Hamon, Claire Glaser, Marianne Ziol, Margot Bucau, Pierre Alexandre Just, Sophie Prévot, Françoise Memeteau, Cyprien Tilmant, Bertrand Hermineaud, Christine Bergeron, Françoise Thelu, Mojgan Devouassoux-Shisheboran, Claire Illac, Averous Gerlinde, Jean-Christophe Sabourin, Aziz Moutaz, Guillaume Bataillon, Eleni Nika, Delphine Loussouarn, Raphaël Bourgade, Alexandra Dervaux.
- Aux partenaires du projet: Grégoire Vincke et Renaud Hoyoux (Société Cytomine), Greg Lipstein, Robert Gibboni, Christine Chung, Emily Miller et Peter Bull (Société Driven Data), Charline Richard, Adrien Neyret et Delphine Baudouin (Société KIT)
- A la BPI et au Conseil de l'Innovation pour leur soutien financier

Références :

- [1] Kleppe A, Skrede O-J, De Raedt S, Liestøl K, Kerr DJ, Danielsen HE. Designing deep learning studies in cancer diagnostics. *Nat Rev Cancer* 2021. <https://doi.org/10.1038/s41568-020-00327-9>.
- [2] Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer* 2020. <https://doi.org/10.1038/s41416-020-01122-x>.
- [3] Kann BH, Hosny A, Aerts HJWL. Artificial intelligence for clinical oncology. *Cancer Cell* 2021. <https://doi.org/10.1016/j.ccell.2021.04.002>.
- [4] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Heal* 2021;7500:1–9. [https://doi.org/10.1016/s2589-7500\(20\)30292-2](https://doi.org/10.1016/s2589-7500(20)30292-2).
- [5] McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94. <https://doi.org/10.1038/s41586-019-1799-6>.
- [6] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology — new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019;16:703–15. <https://doi.org/10.1038/s41571-019-0252-y>.
- [7] Tizhoosh HR, Diamandis P, Campbell CJV, Safarpour A, Kalra S, Maleki D, et al. Searching Images for Consensus: Can AI Remove Observer Variability in Pathology? *Am J Pathol* 2021. <https://doi.org/10.1016/j.ajpath.2021.01.015>.
- [8] Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, Bachtiar V, et al. Artificial intelligence in digital pathology: a roadmap to routine use in clinical practice. *J Pathol* 2019;249:143–50. <https://doi.org/10.1002/path.5310>.
- [9] Lassau N, Bousaid I, Chouzenoux E, Lamarque JP, Charmettant B, Azoulay M, et al. Three artificial intelligence data challenges based on CT and MRI. *Diagn Interv Imaging* 2020;101. <https://doi.org/10.1016/j.diii.2020.03.006>.
- [10] Lassau N, Estienne T, de Vomecourt P, Azoulay M, Cagnol J, Garcia G, et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. *Diagn Interv Imaging* 2019;100:199–209. <https://doi.org/10.1016/j.diii.2019.02.001>.
- [11] Litjens G, Bandi P, Bejnordi BE, Geessink O, Balkenhol M, Bult P, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset. *Gigascience* 2018;7. <https://doi.org/10.1093/gigascience/giy065>.

- [12] Hartman D, Laak JWM Van Der, Gurcan M, Pantanowitz L. Value of public challenges for the development of pathology deep learning algorithms. J Pathol Inform 2020;11:7. https://doi.org/10.4103/jpi.jpi_64_19.
- [13] Conseil de l'Innovation : <https://www.gouvernement.fr/grand-defi-sante-ameliorer-les-diagnostics-medicaux-a-l-aide-de-l-intelligence-artificielle>.
- [14] Female Genital Tumours. WHO Classification of Tumours. 5th ed. 2020.

LIENS D'INTÉRÊTS

Les auteurs déclarent n'avoir aucun lien d'intérêt en rapport avec cet article.

SOUTIENS FINANCIERS

- BPI et Conseil de l'Innovation
- Health Data Hub

Tableaux

Tableau 1 : Les comités du data challenge

Table 1 : The data challenge committees

Comité d'organisation	<ul style="list-style-type: none"> ● SFP : Philippe Bertheau (responsable du CO), Sophie Prevot, Séverine Valmary-Degano, Jean-Christophe Sabourin, Philippe Camparo, Marie-Pierre Wissler ● CNPath : Jean-Yves Scoazec ● Société KIT (organisation Carrefour Pathologie) : Adrien Neyret, Charline Richard ● Health Data Hub : Stéphanie Combes, Karim Zryouil, Agathe Delaune ● Invité permanent : Olivier Clatz (Grand Défi «Amélioration des diagnostics médicaux par l'intelligence artificielle») ● Invités membres du CS : Nicolas Loménie, Boris Hejblum
Conseil scientifique	<p>Co-responsable pathologiste : Séverine Valmary-Degano Co-responsable informatique / IA / statistique / méthodologie : Nicolas Loménie</p> <ul style="list-style-type: none"> ● Pathologistes : Cécile Badoual, Guillaume Bataillon, Philippe Bertheau, Pierre Brousset, Julien Calderaro, Camille Franchet, Catherine Guettier, Valérie Paradis, Nicolas Poté, Sophie Prevot, Jean-Yves Scoazec, Séverine Valmary-Degano ● Informaticiens / IA / statisticiens / méthodologistes : Arnaud Abreu, Emmanuel Bacry, François-Xavier Frenois, Boris Hejblum, Nicolas Loménie, Eric Poullier, Olivier Trassard, Noureddine Zerhouni
Panel de relecteurs	Sophie Prévot (coordinatrice), Guillaume Bataillon, Christine Bergeron, Mojgan Devouassoux-Shisheboran, Claire Glaser

Tableau 2 : Classes diagnostiques référencées par l’OMS [14] □
Table 2 : Diagnostic classes based on the WHO classification [14] □

Classe 0	Non lésionnel – épithélium malpighien normal
Classe 1	Lésion malpighienne intraépithéliale de bas grade
Classe 2	Lésion malpighienne intraépithéliale de haut grade
Classe 3	Carcinome épidermoïde

Tableau 3 : Grille de pondération des erreurs utilisée pour le calcul final des scores
Table 3 : Matrix indicating the error weights used for the calculation of the final scores

	Prédiction Classe 0	Prédiction Classe 1	Prédiction Classe 2	Prédiction Classe 3
vérité de terrain				
Classe 0	0.0	0.1	0.7	1.0
Classe 1	0.1	0.0	0.3	0.7
Classe 2	0.7	0.3	0.0	0.3
Classe 3	1.0	0.7	0.3	0.0

Tableau 4 : Le data challenge en chiffres
Table 4 : The data challenge in numbers

compétiteurs	574
total lames numérisées (apprentissage, test, validation)	4934
To de données	5
annotations graphiques (sur lames d'apprentissage)	5941
courriers aux patientes	5010
courriels internes	plusieurs milliers
visio-conférences internes (heures)	~100
coût total estimé (k€) (hors relecteurs, non rémunérés)	150

Légendes des figures

Figure 1 :

Sur cette vue d'une lame virtuelle avec le viewer Cytomine sont déposées sept annotations graphiques sous la forme de carrés de 300µm de côté. La couleur de chaque carré est associée à une des 4 classes diagnostiques de lésion épithéliale du col utérin.

This whole slide image is displayed with the Cytomine viewer and seven square annotations are present, each colour indicating one of the four diagnostic classes of epithelial cervical lesions.

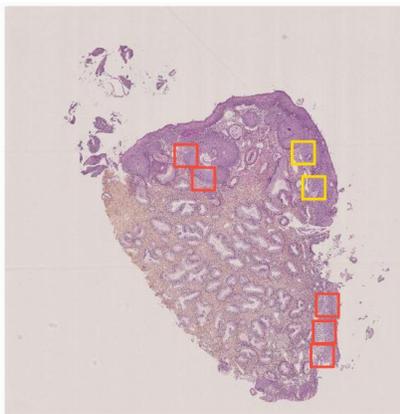


Figure 2 :

Page d'accueil du data challenge, ouverte aux compétiteurs du 9/9/2020 au 29/10/2020.

Data challenge home page, opened from 2020/09/09 until 2020/10/29.

TissueNet: Detect Lesions in Cervical Biopsies
HOSTED BY FRENCH SOCIETY OF PATHOLOGY

The objective of this challenge is to detect the most severe epithelial lesions of the uterine cervix. The challenge brings together thousands of microscopic slides from different medical centers across France.

Quick Facts

PARTICIPANTS	547
NO. OF ENTRIES	613
PRIZE	€25,000
WINNER	TriVn-Health, 12 PRAGUE TEAM

LEADERBOARD RESULTS

OFFICIAL RULES

SHARE THIS:
Facebook Twitter LinkedIn Email

The French Society of Pathology and the Health Data Hub are very enthusiastic about the data challenge results! They show that AI will sooner or later be part of the tools that pathologists use on a daily basis.

— Dr. Philippe Bertheau, President of the French Society of Pathology

Why

A biopsy is a sample of tissue examined at a microscopic level to diagnose cancer or signs of pre-cancer. Digital pathology has developed considerably over the past decade as it has become possible to work with digitized "whole slide images" (WSIs). These heavy image files contain all the information required to diagnose lesions as malignant or benign, yet present huge challenges to use effectively.

This challenge focused on epithelial lesions of the uterine cervix, and featured a unique collection of thousands of experts-labeled WSIs collected from medical centers across France. This is a sizable dataset (100GB) of extremely high resolution images. Given the scale of the dataset, handling the data efficiently is a critical problem to solve in the process of developing an accurate approach to diagnosis.

Figure 3 :

Matrices de confusion des deux premiers compétiteurs (1-Tribvn-Healthcare; 2-Karelds). La vérité de terrain est en lignes horizontales et la prédiction des algorithmes est en colonnes verticales. Une couleur de fond plus claire indique un effectif plus important.

Confusion matrices of the first two competitors (1-Tribvn-Healthcare; 2-Karelds). Ground truth is in rows, predicted data are in columns. A lighter background colour indicates a more important number of cases.

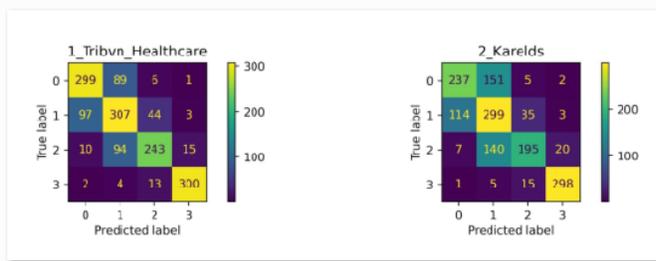


Figure 4 :

Vue du leaderboard final et des scores obtenus par les 4 premiers compétiteurs avec la métrique pondérée.

View of the final leaderboard with the best four competitors scores, calculated using the weighted metric.

