

■ Le dossier – IA, télémédecine et avenir

L'intelligence artificielle va-t-elle remplacer les ophtalmologistes ?

RÉSUMÉ : Les modèles de prédiction clinique intégrant des techniques modernes d'intelligence artificielle (IA), telles que le *deep learning*, ont permis d'obtenir une précision de diagnostic à l'échelle humaine grâce à une série de techniques de classification par imagerie. L'ophtalmologie a été à la pointe de ces développements avec le tout premier modèle d'IA autonome approuvé par la FDA en 2018 pour la détection de la rétinopathie diabétique.

Le déploiement de ces modèles pose cependant un certain nombre de défis. La généralisation de leur utilisation pourra en particulier contraster avec un éventail très ciblé des indications utiles. L'IA est susceptible de transformer les modes traditionnels de prestation de soins et cette transformation se fera sans aucun doute en conjonction avec les progrès des approches numériques de la santé, telles que la télémédecine. Ces technologies ne remplaceront pas les ophtalmologistes, mais les cliniciens devront s'adapter à l'évolution du paysage des soins numériques pour tirer le meilleur parti de ces avantages potentiels.



S. K. WAGNER
Moorfields Eye Hospital, LONDRES,
Royaume-Uni.

■ En préambule

L'explosion quantitative des données numériques, associée à une amélioration sans précédent des ressources de calcul permises par des techniques informatiques en plein essor, a permis d'ouvrir la médecine à la recherche en intelligence artificielle (IA). Les succès retentissants de l'IA médicale pour utiliser ce qu'il est coutume d'appeler les "*big data*" ont surtout été médiatisés au cours de la dernière décennie, mais on peut rappeler ici que le terme IA a été inventé pour la première fois lors d'un atelier d'été au Dartmouth College en 1956 pour désigner un comportement intelligent observé au niveau de machines ou de logiciels.

Actuellement, la compréhension des nuances de la recherche en IA nécessite l'utilisation de deux autres termes clés : l'apprentissage machine (*machine learning*) et un sous-domaine particulier appelé apprentissage profond (ou

deep learning et ce terme sera conservé dans cet article [NDT]). Alors que les algorithmes informatiques traditionnels nécessitent la composition détaillée de règles sous la forme d'une recette, l'apprentissage machine implique l'exécution d'une tâche donnée sans programmation explicite mais plutôt par "expérience" acquise lors de l'intégration de nombreuses données avec un processus d'apprentissage.

Le *deep learning* est l'une des techniques d'apprentissage machine basées sur l'architecture des réseaux de neurones biologiques et qui est à l'origine de nombreuses études en IA aujourd'hui à la une des journaux. Bien qu'il persiste des controverses autour de l'avantage du *deep learning* par rapport aux anciennes techniques de modélisation statistique comme la régression logistique, le *deep learning* excelle probablement dans la cartographie des types de données en haute dimension, telles que les images médicales.

Ces dernières années, les performances des modèles d'IA médicale ont atteint des niveaux impressionnants, en particulier dans le domaine de la classification des images : la catégorisation d'une image donnée, qui indique généralement le diagnostic, l'urgence de la prise en charge ou le pronostic. La précision diagnostique des modèles d'IA a été comparable et, dans certains cas, a même dépassé les performances des humains dans des domaines allant de la mammographie à l'interprétation des radiographies pulmonaires. Des résultats similaires ont été observés en ophtalmologie concernant le diagnostic de la rétinopathie diabétique (RD) à partir de photographies de la rétine ou l'urgence de la prise en charge des patients basée sur les images d'OCT. Ces résultats, correspondant à des tâches de classification, ont suscité l'inquiétude de nombreux professionnels de la santé quant à la possibilité du remplacement des praticiens par de tels processus automatisés.

En effet, les analyses des effectifs, comme le rapport McKinsey 2017, suggèrent que, d'ici 2030, entre 75 et 375 millions de travailleurs seront amenés à changer de profession [1]. Geoffrey Hinton, que beaucoup considèrent comme le fondateur du *deep learning* moderne, a annoncé la réorientation imminente des domaines médicaux fortement dépendants de la classification des images : *“il est tout à fait évident que nous devrions arrêter de former des radiologues”*. Alors que l'ophtalmologie est devenue l'une des spécialités qui bénéficie le plus des données d'imagerie à haute résolution acquises en routine, cette déclaration devrait-elle être modifiée pour y inclure les ophtalmologistes ? Il ne fait aucun doute que l'IA va transformer l'avenir de la prestation de soins, transformation rendue possible par la télémédecine et d'autres applications numériques dans le domaine de la santé. Cependant, cette évolution prévisible doit être relativisée en faisant intervenir plusieurs notions comme la complexité des tâches à accomplir ou l'environnement du déploiement des modèles d'IA.

Les résultats des études en IA médicale peuvent être difficiles à interpréter étant donné l'hétérogénéité de la nomenclature, de la présentation des résultats et des procédures de déclaration incomplètes. Les lignes directrices récentes et émergentes, telles que l'extension CONSORT-AI, visent à homogénéiser la nomenclature pour rendre ces études comparables [2]. En outre, alors qu'un certain nombre d'études prospectives évaluant leur utilisation dans un cadre clinique commencent à voir le jour, la grande majorité des recherches implique une évaluation de la précision du diagnostic dans les données collectées en routine, qui peuvent ne pas être représentatives de la population dans laquelle ces modèles seront déployés en pratique clinique courante.

Le débat sur l'éventuel remplacement des médecins par l'IA est un sujet fréquent et controversé qui concerne des domaines allant de la relation médecin-patient à la précision du diagnostic, en passant par les économies potentielles. Certaines de ces questions, qui concernent notamment l'ophtalmologie, sont abordées ici en mettant l'accent sur la classification des images et la rétinopathie diabétique à titre d'exemple.

■ Tempérer les attentes

Avril 2018 a été un moment clé concernant le développement de l'IA en ophtalmologie, avec l'approbation par la *Food and Drug Administration* (FDA) aux États-Unis du tout premier dispositif médical autonome d'IA concernant la rétinopathie diabétique. En 2020, l'application a été reconnue par l'*American Diabetes Association* pour le dépistage de la RD [3].

Aux États-Unis, on considère que la moitié environ des patients atteints de diabète n'a pas les moyens de consulter un ophtalmologiste tous les ans. L'équipe de l'IDx-DR a mis au point un modèle de *deep learning* pour le dépistage auto-

matisé de la rétinopathie diabétique, avec lequel un médecin généraliste peut envoyer des photographies de la rétine sur un serveur pour solliciter une interprétation. La FDA a accordé cette autorisation sur la base d'une étude de précision du diagnostic démontrant une sensibilité de 87,2 % (IC 95 % : 81,8-91,2 %) et une spécificité de 90,7 % (IC 95 % : 88,3-92,7 %) [4].

Si l'IDx-DR apporte certainement une solution innovante à un besoin non satisfait de dépistage de la RD aux États-Unis, certaines considérations clés soulignent l'importance de reconnaître l'un des principaux obstacles à l'intégration de l'IA dans les soins de santé : le caractère “ciblé”.

L'IA est en effet plus performante lorsqu'elle se concentre sur une tâche unique et ciblée. Dans cet exemple, les indications d'utilisation de l'IDx-DR précisent que le dispositif est destiné aux patients présentant *“une rétinopathie diabétique modérée ou plus avancée”*, que les adultes âgés d'au moins 22 ans ne doivent pas présenter de symptômes visuels et, surtout, qu'ils ne doivent jamais avoir été diagnostiqués comme porteurs d'une rétinopathie diabétique. La présence d'un seul microanévrisme lors de la dernière visite invaliderait donc l'éligibilité d'un patient à l'utilisation de l'IDx-DR. L'appareil utilisé doit être une caméra rétinienne spécifique, le Topcon TRC-NW400 [5]. Enfin, l'IDx-DR n'est pas destiné à détecter les maladies oculaires concomitantes et les fabricants recommandent d'ailleurs, étant donné le risque accru de glaucome chez les patients atteints de diabète, que ces derniers continuent à consulter un ophtalmologiste pour le dépistage du glaucome. L'intégration de ce dispositif dans les systèmes de soins devient donc moins évidente quand les patients peuvent avoir un examen régulier avec des ophtalmologistes ou même des optométristes qui fournissent une évaluation ophtalmique générale complète concernant une série

Le dossier – IA, télémedecine et avenir

de pathologies oculaires, de la cataracte à la rétinopathie diabétique en passant par le glaucome.

Même lorsque les modèles d'IA médicale démontrent ou dépassent une précision "à l'échelle humaine", il se peut que l'on ne puisse pas extrapoler une efficacité significative de ces modèles s'ils sont déployés en pratique clinique courante.

Parmi les premières publications à fort impact qui ont mis en évidence la puissance de l'IA figure un modèle de classification pour la détection de la RD et de l'œdème maculaire diabétique par la photographie de la rétine rapporté par Google en 2016 [6]. Dans le cadre de la détection de la RD, le modèle a obtenu une aire sous la courbe (ASC) de 0,991 (IC 95 % : 0,988-0,993) dans un ensemble de données, un résultat indéniablement impressionnant. Cependant, au début de l'année, un rapport publié sur l'expérience d'utilisation de cette technologie dans une série de cliniques en Thaïlande a révélé des difficultés dans la logistique du déploiement [7]. Dans de nombreux modèles d'IA médicale, le seuil d'un résultat positif est ajusté de telle sorte que la sensibilité est optimisée, au prix d'un plus grand nombre de faux résultats positifs. En conséquence, les patients peuvent avoir besoin de tests de suivi supplémentaires. Dans l'évaluation du modèle de Google, les chercheurs ont noté que les infirmières signalaient souvent aux patients la nécessité de se rendre dans un hôpital plus éloigné pour une évaluation complémentaire en cas de résultat positif. Dans une clinique, cela a été considéré comme l'un des facteurs responsables du fait que seulement 50 % des patients consentent à une classification basée sur l'IA. Un tel problème a probablement des ramifications plus importantes pour les pays à revenus faible et moyen, mais peut poser des problèmes pour tout établissement de soins de proximité, où un résultat de dépistage positif implique un déplacement onéreux vers un établissement régional.

Les critères de qualité des images constituent un autre facteur limitant. Dans le rapport original, comme dans de nombreuses autres études sur l'IA en ophtalmologie, les images ont été évaluées par des ophtalmologistes pour déterminer si elles pouvaient être classées et seules celles jugées adéquates, bonnes et excellentes ont été incluses. Dans l'étude du système de Google, des infirmières ont déclaré que des images qu'elles jugeaient lisibles par l'homme étaient rejetées par le système, ce qui représentait une certaine frustration.

Enfin, si une connexion Internet à haut débit est habituellement disponible dans les établissements de soins des pays à revenu élevé, ce n'est pas toujours le cas sur l'ensemble de la planète. Un test a été réalisé en Thaïlande, où une clinique a subi une perte de connexion de deux heures. Le fait qu'un service entier, comme le dépistage de la rétinopathie diabétique, puisse dépendre de la disponibilité d'une connexion internet fiable peut impliquer le réaménagement des infrastructures de certaines institutions.

Un argument habituel en faveur du remplacement des cliniciens par des outils d'IA est la perception d'économies de coûts qui pourraient résulter de la mise en œuvre d'un système automatisé. L'analyse comparative pertinente peut cependant être difficile. Au niveau le plus simple, on peut comparer le nombre de patients évalués dans une période donnée par un clinicien vs un modèle d'IA, mais cela ne tient pas compte des coûts indirects liés aux résultats faussement positifs, de la mise en place d'une infrastructure et du contrôle de qualité du modèle.

À Singapour, les patients atteints de diabète sont évalués dans le cadre du *Singapore Integrated Diabetic Retinopathy Programme* (SIDRP), un système de stockage et de transmission basé sur la téléophtalmologie dans lequel des évaluateurs qualifiés, plutôt que des ophtalmologistes, évaluent les images. Une analyse de modélisation écono-

mique d'un modèle de détection de la rétinopathie diabétique, mis au point par un groupe de Singapour, montre une économie de 15 \$ par an et par patient lorsque le modèle d'IA est utilisé de manière semi-automatisée (62 \$) par rapport à la classification humaine traditionnelle (77 \$) [8]. Une comparaison directe avec les ophtalmologistes n'a pas eu lieu mais, auparavant, le SIDRP s'est avéré rentable par rapport à l'évaluation par un médecin généraliste (et non par un ophtalmologiste). Aux États-Unis, l'évaluation économique d'un modèle d'IA pour la classification des RD n'a pas été jugée rentable par rapport à l'évaluation traditionnelle par un professionnel des soins oculaires dans le scénario de base (10,85 \$ contre 8,20 \$ pour le diabète de type 2), mais elle a excellé dans les scénarios où l'adhésion au dépistage était de 23 % ou plus [3]. La rentabilité avérée des systèmes autonomes n'est donc pas évidente, bien que ce soit l'un des avantages potentiels les plus fréquemment cités de l'intégration de l'IA dans les systèmes de soins de santé.

Intelligence artificielle ou augmentation de l'intelligence ?

C'est surtout à partir des années 1950 que l'armée américaine a commencé à explorer les moyens d'améliorer "l'efficacité intellectuelle de chaque être humain". Ce concept, appelé augmentation de l'intelligence (AI), a été développé en tandem avec l'évolution des systèmes d'IA où les performances des cliniciens peuvent être augmentées à l'aide de techniques de modélisation modernes.

Pour de nombreux chercheurs en IA médicale, plutôt que de remplacer les cliniciens, l'IA devrait plutôt être utilisée pour améliorer les modèles plus traditionnels de prestations de soins de santé. L'un des rares essais contrôlés randomisés en double aveugle dans le domaine de l'apprentissage médical approfondi a été publié au début de cette

année dans le domaine de la gastro-entérologie. Après des résultats prometteurs dans des essais pilotes, Wang *et al.* ont évalué l'incorporation d'un système d'alarme basé sur le *deep learning* pour la détection de polypes pendant la coloscopie. En bref, le système évalue de manière synchrone les données vidéo entrantes pendant la coloscopie en émettant une boîte d'alerte visuelle et une alarme sonore correspondante lorsqu'un adénome est détecté. Le taux de détection des adénomes était plus élevé chez les patients pour lesquels le modèle AI était utilisé (34 contre 28 % sans l'AI), la plupart des adénomes reclassés présentant une taille plus petite, une forme plate ou étant cachés dans les plis du colon.

Cependant, toutes les expériences de collaboration entre l'homme et l'IA de ce genre n'ont pas été positives. Alors que d'autres rapports ont montré que les performances humaines peuvent effectivement être améliorées par l'IA, en particulier pour les personnes moins expérimentées, les cliniciens peuvent également être faussement rassurés par les décisions de l'IA. Un rapport distinct dans le domaine de la dermatologie, dans lequel les chercheurs ont délibérément modifié le résultat d'un modèle d'IA pour poser un faux diagnostic, a montré que même des cliniciens expérimentés peuvent être nettement moins performants dans ces scénarios [9].

L'amélioration des capacités humaines par l'IA n'a pas encore été évaluée de manière systématique en ophtalmologie, mais on pourrait envisager un scénario dans lequel l'interprétation de l'OCT maculaire par les ophtalmologistes pourrait être améliorée par l'inclusion d'un modèle d'IA pour la détection des caractéristiques de la maladie, conformément aux cartes de segmentation vues dans les travaux du Moorfields Eye Hospital et de DeepMind [10]. L'une des pistes potentielles pour l'IA en ophtalmologie est la personnalisation de la gestion de la dégénérescence maculaire liée à l'âge

(DMLA) en OCT en utilisant la détection de biomarqueurs basée sur le *deep learning*. La reconnaissance de la valeur pronostique de distributions spécifiques du liquide intrarétinien par rapport au liquide sous-rétinien, qui ont des implications pronostiques différentes, peut être effectuée par des méthodes d'IA et peut permettre aux ophtalmologistes, par exemple, d'adapter les intervalles de traitement par injections intravitréennes à chaque patient [11].

On peut également considérer l'inverse de la question, savoir si l'IA remplacera les ophtalmologistes : l'IA ajoutera-t-elle de nouveaux rôles à l'ophtalmologiste ou à l'optométriste ? Beaucoup ont été surpris par les résultats du Google Brain 2018, dont l'objectif initial était d'extraire les facteurs de risque cardiovasculaire des photographies de la rétine [12]. Leur modèle de *deep learning* a prédit l'âge du patient à partir d'une seule photographie de la rétine avec une erreur absolue moyenne de 3,26 ans ! Compte tenu des associations connues entre la tortuosité vasculaire de la rétine et le calibre, ce résultat n'est peut-être pas très surprenant, mais le modèle a également été capable de classer le sexe biologique avec une précision remarquable (ASC de 0,97). Si les ophtalmologistes et les médecins non ophtalmologistes sont conscients du vieil adage selon lequel l'œil est une fenêtre sur l'âme, l'IA pourrait permettre d'illustrer cette interaction de manière plus puissante et plus quantitative.

À partir de 2020, des modèles de *deep learning* ont été rapportés pour la stratification du risque de maladie cardiovasculaire et la détection de divers troubles systémiques, allant de l'insuffisance rénale chronique à l'anémie [13, 14]. Cependant, l'interprétation de ces images peut être difficile en présence de comorbidités (RD et maladies cardiovasculaires) et de maladies non liées (détection de l'anémie chez les personnes atteintes de glaucome primaire à angle ouvert). Il est donc possible que l'IA débloque une nouvelle catégorie

potentielle de répartition des patients et d'indications diverses nécessitant une évaluation ophtalmique.

■ Conclusion

L'impact des technologies de l'IA dans le domaine de la santé se concrétise avec les résultats des premières études comparatives prospectives et l'obtention d'autorisations de mise sur le marché. L'efficacité des modèles dépendra de la tâche particulière et du scénario de déploiement probable, mais la détermination de la valeur de l'IA dans l'évaluation et la gestion des patients exige beaucoup plus que de simplement atteindre ou dépasser la précision diagnostique des humains.

De nombreux aspects n'ont pas été pris en compte ici : qu'en est-il du rôle de l'IA dans l'anamnèse ou la chirurgie ? L'IA pourra-t-elle un jour imiter les éléments humains cruciaux de l'interaction entre le patient et le médecin, tels que le développement de rapports et l'interaction ? Les ophtalmologistes devront-ils superviser les résultats de l'IA et quelles en seront les implications juridiques ? Bien que l'IA soit très prometteuse pour la démocratisation et l'amélioration de l'efficacité des soins de santé, de nombreuses questions restent encore sans réponse. Les ophtalmologistes ne seront pas remplacés par l'IA, mais il est probable que nos méthodes traditionnelles de prestations de soins seront transformées, non seulement par l'IA mais aussi par la maturation d'autres solutions numériques de santé, comme la télémédecine.

Traduit de l'anglais par le Dr Thomas Desmettre.

BIBLIOGRAPHIE

- McKinsey Institute. Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages. 2017. Disponible sur www.mckinsey.com/featured-insights/future-of-work/

Le dossier – IA, télémedecine et avenir

- jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages. Accédé le 25/11/2020.
2. LIU X, CRUZ RIVERA S, MOHER D *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*, 2020;26:1364-1374.
 3. WOLF RM, CHANNA R, ABRAMOFF MD *et al.* Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmol*, 2020;138:1063-1069.
 4. ABRAMOFF MD, LAVIN PT, BIRCH M *et al.* Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*, 2018;1:39.
 5. Digital Diagnosis: IDx-DR indication for use: dxs.ai/products/idx-dr/indications-for-use/
 6. GULSHAN V, PENG L, CORAM M *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 2016;316:2402-2410.
 7. BEEDE E, BAYLOR E, HERSCH F *et al.* A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Honolulu, HI, USA: Association for Computing Machinery, 2020:1-12.
 8. XIE Y, NGUYEN QD, HAMZAH H *et al.* Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*, 2020;2:e240-e249.
 9. TSCHANDL P, RINNER C, APALLA Z *et al.* Human-computer collaboration for skin cancer recognition. *Nat Med*, 2020;26:1229-1234.
 10. DE FAUW J, LEDSAM JR, ROMERA-PAREDES B *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*, 2018;24:1342-1350.
 11. MORAES G, FU DJ, WILSON M *et al.* Quantitative analysis of OCT for neovascular age-related macular degeneration using deep learning. *Ophthalmology*, 2020 [online ahead of print].
 12. POPLIN R, VARADARAJAN AV, BLUMER K *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*, 2018;2:158-164.
 13. MITANI A, HUANG A, VENUGOPALAN S *et al.* Detection of anaemia from retinal fundus images via deep learning. *Nat Biomed Eng*, 2020;4:18-27.
 14. SABANAYAGAM C, XU D, TING DSW *et al.* A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health*, 2020;2:e295-e302.

L'auteur a déclaré ne pas avoir de conflits d'intérêts concernant les données publiées dans cet article.