

L'intelligence artificielle et le diagnostic

Des outils pour la pratique

ARTIFICIELLE, COLLECTIVE, DE NOUVELLES INTELLIGENCES AU SERVICE DU DIAGNOSTIC MÉDICAL ?



Maxime Demoly

PDG et co-fondateur d'AdviceMedica

Brève histoire de l'intelligence artificielle en médecine

Les premières applications mettant en œuvre des programmes d'intelligence artificielle (IA) dans le domaine de la médecine sont les "systèmes experts" des années 1970 et 1980. Dans leur architecture, ces systèmes cherchent à imiter les mécanismes cognitifs d'experts humains en combinant un algorithme de simulation de raisonnements déductifs et une base de données de faits et de règles propres à la spécialité médicale. Le premier outil du genre, **Mycin** [1] est spécialisé dans l'identification d'infections bactériennes. Vient ensuite **Caduceus** dont l'ambition est de couvrir toute la médecine interne. Bien que leur domaine d'application soit restreint, ces systèmes étaient performants, mais souffraient de problèmes d'intégration technique à une époque où l'informatique était naissante et où les prémices d'internet étaient à peine en place. Ils ont cependant entériné sinon l'usage, du moins la promesse, de l'intelligence artificielle appliquée à la recherche de diagnostic clinique.



Zoom sur Watson

Au début des années 2010, IBM (International Business Machine Corporation) participe au jeu télévisé *Jeopardy!* — dans lequel les candidats doivent trouver les questions correspondant aux réponses données — et l'emporte face au champion en titre, avec l'outil d'intelligence artificielle **Watson**. Fort de ce succès, IBM annonce son ambition de **révolutionner la médecine**. L'objectif poursuivi est d'apporter aux oncologues une aide au diagnostic et au choix d'un traitement adapté pour leurs patients, et donc de guérir le cancer.

Watson intègre la première génération d'algorithmes de traitement du langage naturel. Une question en entrée est analysée et transformée en un ensemble de requêtes qui sont adressées à une vaste base de connaissances afin d'établir des réponses candidates. Ces réponses sont ensuite évaluées, au regard de la pertinence des preuves qui les sous-tendent, grâce à une seconde base de données.

Les moyens financiers mis en œuvre pour acquérir les bases de connaissances sont importants. De même, IBM consacre de vastes moyens

marketing pour diffuser des annonces publiques concernant l'efficacité anticipée du produit **Watson Health**. Néanmoins, si la technologie fonctionnait [2], elle n'avait en fait que peu d'incidence sur les résultats cliniques là où elle était mise en place. **Elle avait tendance à expliquer aux experts ce qu'ils savaient déjà et à ajouter une contrainte administrative alors que les médecins auraient préféré passer plus de temps avec les patients qu'à saisir des données dans un système.** L'objectif de révolutionner le domaine de la santé n'a pas été atteint et **Watson Health** a fini par être vendu au début de l'année 2022.

Ces premières applications cherchaient en quelque sorte à remplacer le médecin, que ce soit pour orienter les patients dans leur parcours de soin, ou apporter des solutions de diagnostic dans des endroits reculés. D'autres outils ont par la suite été pensés pour accompagner le médecin dans sa pratique.

L'après Watson

C'est en particulier le cas des systèmes utilisant des algorithmes de vision, permettant d'analyser et de classer des images, mis en lumière en 2012 à l'occasion du concours de reconnaissance **ImageNet** qu'ils dominent largement. L'utilisation de ces algorithmes marque l'avènement de la mise en pratique concrète du *deep learning* [3].

Qu'elles soient produites par radiographie, scintigraphie, IRM, échographie, ou flux vidéo, les **images médicales** sont nécessaires pour visualiser les organes internes et en détecter les anomalies. Parmi les applications classiques, on note par exemple la recherche de tumeurs cancéreuses ou l'identification du type de maladie pulmonaire interstitielle. Ces outils fournissent une

aide précieuse aux médecins spécialistes leur permettant de ne pas passer à côté d'un diagnostic délicat, en particulier lors de périodes d'exercices prolongées.

Dans leur conception, ces algorithmes requièrent un soin tout particulier à la qualité des données initiales utilisées pour l'entraînement à la tâche de classification. Ces données doivent être parfaitement annotées et représenter fidèlement les situations rencontrées dans la réalité. Cette classe d'algorithme, souvent qualifiée de boîte noire, est réputée pour son caractère opaque, et il est difficile d'en expliquer le résultat, même en ayant accès à toutes les étapes intermédiaires permettant d'y arriver. Dans le domaine médical, où des images complexes sont fréquemment utilisées pour diagnostiquer des maladies, il est crucial d'avoir un système fiable et efficace, qui puisse également expliquer aux médecins experts les décisions prises.

L'intelligence collective et ses applications en médecine

En 1906, Francis Galton réalise une expérience qui est considérée comme un exemple classique d'estimation d'une grandeur en statistique [4]. Il se rend à une foire agricole où les visiteurs sont invités à participer à un concours consistant à estimer le poids d'un bœuf exposé, et demande à pouvoir analyser les résultats. Galton réunit les estimations des participants et montre que leur moyenne est très proche du poids réel du bœuf. Il constate avec surprise que la précision de l'estimation augmente avec le nombre de participants. Cette expérience introduit le concept d'**intelligence collective**, dans

lequel la synthèse des informations provenant d'un groupe se révèle plus pertinente que l'information fournie par chaque individu.

Si l'exercice de la médecine pousse spontanément à l'échange de pratiques entre professionnels, il est tentant de questionner la démarche permettant au groupe d'être plus précis que chacun de ses membres dans une discipline qui est autrement plus complexe que ne l'est l'estimation ponctuelle d'une variable. **Plusieurs méthodes mettant en application l'intelligence collective dans le domaine médical sont régulièrement utilisées.**

Échange de pratiques

Concernant l'échange de pratiques, la méthode la plus répandue est celle des réunions de concertation pluridisciplinaires (RCP). Initialement pensées pour le cancer, elles regroupent différents professionnels de santé autour du traitement d'un patient. Ces réunions permettent de discuter des cas complexes, de partager les compétences de chacun, et de prendre collectivement des décisions sur la meilleure prise en charge pour le patient, en se fondant sur les connaissances les plus récentes et les meilleures pratiques en matière de traitement. Les professionnels impliqués dans les RCP peuvent inclure des oncologues, des chirurgiens, des radiothérapeutes, des pathologistes, des infirmières spécialisées en oncologie, des psychologues et d'autres professionnels de la santé impliqués dans le traitement du cancer.

La méthode Delphi [5] a cherché à davantage cadrer l'échange et vise à obtenir un consensus dans un contexte impliquant un panel

d'experts, comme c'est le cas en médecine. La méthode fonctionne par tour de consultation et s'appuie sur des questionnaires qui sont distribués aux experts. Les participants commencent par une revue de la littérature autour du sujet d'intérêt. Ils construisent alors un premier questionnaire et expriment leur degré d'accord avec les propositions et en formulent de nouvelles qui seront soumises à l'avis du groupe lors du tour suivant. Les réponses sont analysées afin d'identifier les points de vue majoritaires et minoritaires. Le questionnaire est révisé à chaque tour jusqu'à ce qu'un consensus soit atteint ou que les divergences d'opinions persistent.

A-t-on raison de chercher un consensus entre pairs, le praticien n'a-t-il pas à lui seul toutes les cartes en main ? Quelques études récentes ont analysé la performance d'une approche mettant en œuvre l'intelligence collective pour améliorer la précision des diagnostics cliniques.

Revue de littérature

Dans la première [6], des radiologues doivent prendre la décision de rappeler ou non une patiente, pour des examens complémentaires, suite à une mammographie pour un dépistage du cancer du sein. Les décisions de l'ensemble des médecins sont réunies et agrégées selon différents modes, par majorité simple ou pondérée, de façon à faire émerger la décision du groupe. L'étude montre qu'utiliser la décision du groupe plutôt que celle d'un seul radiologue permet d'augmenter le taux de vrais positifs – rappeler une patiente ayant le cancer – et faire diminuer celui de faux positifs – rappeler une patiente n'ayant pas le cancer. **Le groupe est donc plus**

précis que ses membres les plus performants, même sur des problèmes cognitifs aussi complexes.

Une seconde étude [7] porte sur les conditions qui permettent au groupe d'être plus efficace que ses membres. Les chercheurs se sont appuyés sur une base de données en vie réelle qui a permis à 140 médecins de réaliser 20 000 diagnostics de cancer de la peau et du sein. **Les résultats montrent que pour que la synthèse des diagnostics soit plus précise que celui du meilleur médecin, il faut que les membres du groupe possèdent des capacités cliniques assez similaires.**

Enfin, une dernière étude [8] quantifie l'augmentation de la précision du diagnostic pour un groupe. Elle porte sur un ensemble de 2 069 médecins et 1 572 cas de douleur abdominale et de fièvre. La précision clinique passe de 62,5 % pour un seul médecin, à 85,6 % pour un groupe de neuf médecins, soit une différence de 23 % sur l'ensemble des cas – 17,3 % de différence pour les cas de douleur abdominale et 29,8 % pour les cas de fièvre.

Les approches exploitant l'intelligence collective offrent un bénéfice réel même pour des tâches complexes. Contrairement aux outils d'intelligence artificielle, elles permettent au médecin de comprendre les étapes du raisonnement clinique et ne pas endosser une décision qui n'est pas la sienne. Il garde ainsi la main sur son patient et conserve son autonomie.

Comment combiner intelligence artificielle et intelligence collective ?

D'une part les technologies d'intelligence artificielle sont performantes, mais donnent encore trop peu d'explications sur les décisions prises. D'autre part, avoir recours à l'intelligence collective augmente significativement la précision des diagnostics. Pourquoi ne pas faire converger les deux approches ?

L'outil **AdviceMedica** [9], développé avec le CHU de Montpellier, **permet aux médecins inscrits de demander des avis cliniques à des communautés de praticiens**, organisées par spécialité médicale ou par thématique. Lorsqu'un utilisateur soumet un cas clinique à l'application, **un algorithme de traitement du langage naturel l'analyse et propose la lecture des cas cliniques similaires**. Il s'agit de cas cliniques passés qui ont déjà été discutés et pour lesquels un diagnostic a été posé. L'intelligence artificielle sert à structurer la base de données selon un critère de proximité sémantique et clinique. **AdviceMedica** compte plus de 5 000 utilisateurs de toutes les spécialités médicales qui résolvent quelques milliers de cas cliniques par an. Au-delà des échanges quotidiens, qui participent à la formation continue et permettent de décloisonner les spécialités médicales, l'application s'est illustrée par des services rendus, notamment au sein de la spécialité d'allergologie.

De façon générale, pour qu'une application soit utilisée par le médecin dans sa pratique quotidienne, il ne suffit pas qu'elle rende le service qu'on lui demande, il faut également que le système soit utilisable sans surcharge cognitive. Avec un design inspiré des **listes de diffusion**, les utilisateurs d'**AdviceMedica** reçoivent chaque matin un compte rendu des échanges

sur leurs réseaux **sous forme de fil d'actualité**. Ils interagissent activement avec ce compte rendu et choisissent les cas cliniques qu'ils souhaitent lire. L'application s'utilise aussi naturellement que l'on consulte ses emails.

Un exemple d'application

À la fin de l'année 2020, lorsque les premiers cas d'allergie au vaccin Pfizer ont été signalés, les agences de santé ont rapidement interdit la vaccination chez les patients allergiques. Cette décision, motivée par la suspicion d'allergie au polyéthylène glycol, aurait privé de vaccination un tiers de la population. La communauté des allergologues français a rapidement édité et diffusé des recommandations, et les praticiens ont présenté leurs cas cliniques sur **AdviceMedica**, conseillés par les experts des centres de référence au fur et à mesure que les données cliniques se constituaient. Des centaines de cas sont documentés. **AdviceMedica** a permis, pour chaque cas, décrit de donner des solutions personnalisées, de diffuser les recommandations en cours, et de conclure à l'absence de contreindication allergologique à vaccination.

Certains réseaux sont organisés par thématiques, ce qui leur permet de remonter rapidement des alertes de pharmacovigilance, comme ça

a été le cas en allergie de contact. La plateforme **AdviceMedica** a permis aux médecins d'être avertis de l'émergence de nouveaux allergènes et des risques associés. Parmi ces nouvelles sources d'exposition, on note par exemple l'isobornyl acrylate utilisé dans la colle des capteurs de glycémie [10] ou les isothiazolinones du slime, pâte à malaxer élastique prisee des enfants et adolescents, dont le signalement a abouti à une étude de toxicovigilance complète.

Conclusion sur les dernières avancées en intelligence artificielle

Le modèle de langage **ChatGPT**, développé par OpenAI, occupe le devant de la scène depuis fin novembre 2022 et impressionne par sa capacité à exécuter un grand nombre de tâches de langage naturel. Contrairement aux algorithmes de *deep learning*, conçus pour classer des données souvent sous forme d'images, les modèles de langage sont entraînés pour prédire le mot suivant une séquence donnée en fonction du contexte. Avec un volume de données d'entraînement suffisant, ces modèles de langage sont capables de générer des suites de mots cohérentes, qui ne sont pas présentes dans leur base d'entraînement. Ils présentent de

très bonnes capacités explicatives et de raisonnement déductif, ce qui peut représenter une aubaine pour pallier l'opacité des algorithmes de *deep learning*, une aide pour les étudiants en médecine ou, au contraire, une béquille si importante qu'ils ne développeront jamais un raisonnement clinique critique, ni la capacité à produire des idées originales et à les étayer par des preuves [11].

Le pari d'**AdviceMedica** a été de **mettre les médecins experts de leur discipline au centre de l'application** et de **ne pas tenter de générer une réponse clinique à leur place**. Si de tels algorithmes [12] sont capables de passer des concours de médecine, il paraît peu probable qu'ils puissent assimiler l'expérience pratique du médecin qui nécessite une action dans le monde réel et reste à ce jour l'apanage de l'humain. Encore une fois, c'est la combinaison de l'intelligence artificielle et de l'intelligence collective qui pourra offrir la meilleure solution aux médecins. ■

✱ Maxime Demoly est PDG d'**AdviceMedica**.

Mots-clés :

Intelligence artificielle, Systèmes experts, *Deep learning*, Algorithmes de vision, Intelligence collective, Diagnostic clinique, Modèles de langage

Bibliographie

- Shortliffe E. Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. Symposium on Computer Applications in Medical Care 1977 ; 66-9.
- Aggarwal M, Madhukar M. IBM's Watson Analytics for Health Care. Cloud Computing Systems and Applications in Healthcare 2017.
- Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evolutionary Intelligence 2022 ; 15 : 1-22.
- Galton F. Vox Populi. Nature 1907 ; 450-1.
- Elisabeth L. À la recherche d'un consensus professionnel, la méthode Delphi. Sages-Femmes 2021 ; 20 : 52-4.
- Wolf M, Krause J, Carney PA et al. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. PLoS One 2015 ; 10 : e0134269.
- Kurvers RH, Herzog SM, Hertwig R et al. Boosting medical diagnostics by pooling independent judgments. Proc Natl Acad Sci U S A 2016 ; 113 : 8777-82.
- Barnett ML, Boddupalli D, Nundy S, Bates DW. Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians. JAMA Netw Open 2019 ; 2 : e190096.
- www.advicemedica.com.
- Nakara K, Bocquet A, Dupuis C et al. Eczémas de contact récurrents aux capteurs de glycémie : mécanismes, allergènes en cause et conséquences. Annales de Dermatologie et de Vénérologie 2021 ; 1 : A116.
- Arif TB, Munaf U, Ul-Haque I. The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Med Educ Online 2023 ; 28 : 2181052.
- Kung TH, Cheatham M, Medenilla A et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 ; 2 : e0000198.

UN OUTIL POUR REPÉRER LES PRESCRIPTIONS POTENTIELLEMENT INAPPROPRIÉES D'ANTIBIOTIQUES

 La rédaction

La problématique des prescriptions médicamenteuses potentiellement inappropriées est prépondérante en population âgée. Or, parmi les classes médicamenteuses concernées, on retrouve plus particulièrement les antibiotiques. Indépendamment du risque iatrogénique, ces cas participent à aggraver l'antibiorésistance. Outre la sensibilisation et la formation des professionnels de santé, d'autres actions

peuvent être intéressantes pour accompagner les prescripteurs telles que le développement d'outils numériques d'aide au ciblage des ordonnances concernées. Bien que des listes de prescription inappropriée de médicaments existent, ou que la pharmacie clinique fasse déjà manuellement, pour les antibiotiques, ce genre d'évaluation, un outil informatique pourrait faciliter la pratique.

Pour développer un outil de détection de prescriptions potentiellement inappropriées, il faut tout d'abord présenter les critères qui permettent de la définir. Ce travail permettra ainsi d'établir des règles de reconnaissance des cas de prescriptions potentiellement inappropriées, que l'outil observera afin de repérer une situation potentielle. En revanche, infirmer un tel cas de figure nécessitera forcément l'avis final d'un expert.

Le bon usage des antibiotiques est ainsi régi par des critères explicites et implicites.

Les critères implicites représentent une évaluation fondée sur un jugement expert sur la qualité de la prescription. Elle va tenir compte d'un ensemble de données concernant le contexte précis d'un patient, et ce, à un moment donné. « Par exemple, lors d'une prescription d'antibiotiques pour un patient, l'évaluation implicite aura pour rôle de juger, pour un patient donné, si cet antibiotique correspond bien à une qualité de prescription au vu de

ses antécédents, de son terrain, des autres traitements associés, des résultats microbiologiques, ou encore d'autres paramètres biologiques qui vont permettre de dire si c'est efficace ou pas efficace », précise le Dr Nicolas Baclet, infectiologue impliqué dans un projet de développement d'un outil d'évaluation. Cette approche se fonde donc sur un jugement, soit celui d'experts au cas par cas, soit d'échelles.

L'approche explicite, elle, est fondée sur des critères que l'on va pré-établir, soit à partir d'avis expert, soit sur la base de données de la littérature. Ces critères permettent de déterminer, indépendamment du contexte d'un patient, si telle prescription va être potentiellement inappropriée, et ce, dans quel contexte et quelles que soient les données précises du patient. Le Dr Nicolas Baclet détaille ainsi le travail qu'il mène en collaboration avec le GinGer : « Concernant les antibiotiques, je m'intéresse particulièrement à cette approche explicite au sujet des antibiotiques parce qu'il n'y a actuellement pas de référentiel validé pour définir ce

qu'est une prescription potentiellement inappropriée sur une base d'approche explicite. Mon travail de recherche était donc d'établir une première liste de critères explicites de prescriptions inappropriées d'antibiotiques. »

L'idée serait ainsi d'utiliser des outils informatiques pour détecter ces situations.

On apprendra à l'outil des règles simples de reconnaissance de situations problématiques à partir des recommandations données par des experts. L'outil pourra se présenter sous forme d'application par exemple, capable de passer au crible un grand volume de prescriptions.

Il pourrait être utilisé soit pour détecter *a posteriori*, afin d'avoir des données d'évaluation des pratiques et revenir ainsi vers les prescripteurs pour leur faire un bilan ou leur permettre de questionner leur pratique par rapport à d'autres.

Ou en temps réel, par exemple pour la réévaluation des traitements dans les 48-72 heures, une fois l'antibiothérapie commencée.

On peut donc imaginer qu'un outil de détection pourrait permettre aux infectiologues de mieux cibler leurs actions de bon usage pour aider les prescripteurs et évaluer en direct si on est en effet dans une situation de prescription inappropriée.

« Ces outils ne suffiront pas à confirmer l'alerte, seul le diagnostic

médical ou pharmaceutique peut la confirmer. Ce genre d'outils est destiné au niveau expert, par exemple une équipe d'infectiologie qui a besoin d'une plus-value pour essayer de faire le tri parmi un grand nombre de patients et cibler directement les patients. », ajoute le Dr Baclet. ■

✖ Le Dr Nicolas Baclet est à l'Université de Lille, CHU Lille, ULR 2694 - METRICS : Éva-

luation des technologies de santé et des pratiques médicales et Hôpitaux de l'Université Catholique de Lille, Unité transversale d'infectiologie, Hôpital St Vincent de Paul, Lille.

Mots-clés :

Antibiotiques,
Antibiorésistance, IA,
Outil de détection,
Prescription potentiellement
inappropriée

Pour en savoir plus

- Baclet N, Calafiore M, Fregnac C et al. Explicit definitions of potentially inappropriate prescriptions of antibiotics in hospitalized older patients. *Infect Dis Now* 2022 ; 52 : 214-22.
- Baclet N, Ficheur G, Alfandari S et al. Explicit definitions of potentially inappropriate

prescriptions of antibiotics in older patients: a compilation derived from a systematic review. *Int J Antimicrob Agents* 2017 ; 50 : 640-8.

- www.pro.univ-lille.fr/nicolas-baclet/axes-de-recherche
- www.infectiologie.com/fr/infectio-geriatrie-ginger.html

RENDEZ-VOUS DE L'INDUSTRIE

SANTÉ NUMÉRIQUE

Bien vieillir à domicile

Le groupe Alogia (société à mission) a annoncé le 23 septembre 2023 la sortie de RevealCare, un algorithme de prédiction des besoins des seniors, notamment concernant leur accès aux commerces, services, structures de soins, etc. Le logiciel développé se servira des données concernant l'âge, le sexe et l'adresse des usagers, ainsi que des

données publiques sur l'environnement géographique, afin de « prédire les besoins et classer les bénéficiaires senior par ordre de priorité d'intervention ». Cette technologie s'adresse aux acteurs de la Silver economy. ●

→ OR d'après le communiqué de presse d'Alogia Groupe du 23/09/23.

PHARMACOVIGILANCE

OSTÉOPOROSE

Risque sévère d'hypocalcémie chez les patients dialysés ou souffrant d'insuffisance rénale sévère sous dénosumab

La Food and Drug Administration (FDA) a alerté sur des cas d'hypocalcémie sous dénosumab chez des patients souffrant d'insuffisance rénale sévère ou sous dialyse (le dénosumab est prescrit pour les patientes post-ménopausiques et les hommes à un haut risque de fracture). Le risque sévère inclut celui d'hospitalisation et

de décès. La FDA recommande ainsi aux professionnels de santé de compléter en calcium et en vitamine D ces patients à risque et de réaliser un suivi biologique fréquent. ●

→ OR d'après un communiqué de la FDA du 22 novembre 2022.