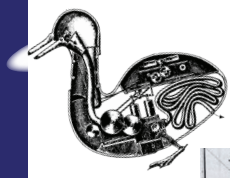
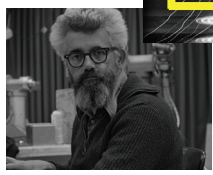
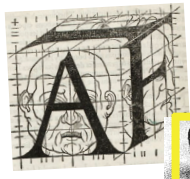
INTELLIGENCE & MALADIES ARTIFICIELLE DIGESTIVES

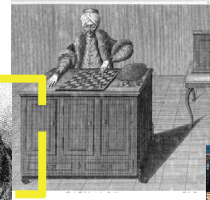


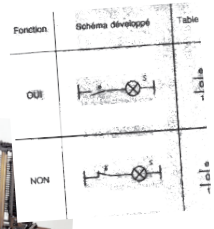
How Will The New 'Golem' Help Israel?

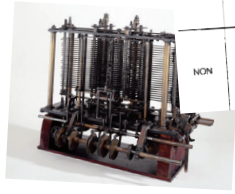
Giant Computer With Legendary Name Being Built At Weizmann Institute

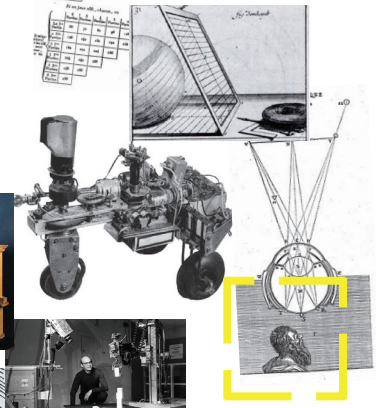












L'Institut des
Maladies de l'Appareil Digestif

INTELLIGENCE ET MALADIES ARTIFICIELLE DIGESTIVES

Coordonné par Emmanuel Coron
& Geoffroy Vanbiervliet

IMAD

Avec le concours de la fondation
SANTÉDIGE

Sous l'égide de la Société française
d'endoscopie digestive

IMAD

SOMMAIRE

Intelligence artificielle et santé : histoire et modèles 1
Romain Galmiche

Principe de fonctionnement et applications 15
Harold Mouchère

IA et pathologie digestive haute 24
Emmanuel Coron, Gabriel Rahmi

IA et pathologie du grêle 33
Xavier Dray, Aymeric Histace, Romain Leenhardt

IA et pathologie digestive colique 48
Aymeric Becq

IA et maladies chroniques digestives intestinales 57
Catherine Le Berre

IA et pathologie pancréatique 67
Geoffroy Vanbiervliet

À la frontière des autres disciplines : nouvelles technologies 77
Lucille Quénéhervé

À la frontière des autres disciplines : anatomopathologie 86
Raphaël Bourgade, Céline Bossard

Quelques perspectives 99
Robert Benamouzig

l'IMAD de Nantes s'associe avec grand intérêt à La Société française d'endoscopie dans cette initiative originale : publier le premier livre francophone sur l'intelligence artificielle dans les maladies digestives. À la veille de l'introduction de ces techniques novatrices dans la pratique courante, cet ouvrage répond à un besoin urgent de mises au point sur les données cliniques à disposition. Ne nous méprenons pas : l'intelligence artificielle marque un moment crucial de l'histoire de nos disciplines. Elle parachève les efforts entrepris depuis 15 ans pour inventer une nouvelle endoscopie fine, décisionnelle et proactive, en même temps qu'elle ouvre un nouveau chapitre en redéfinissant en profondeur le rôle du praticien.

Il importe que la communauté médicale ne soit pas seulement spectatrice de ce bouleversement technique. Parce qu'elles se situent à l'interface du complexe techno-industriel, des patients, des sphères politique et administrative, parce que leur rôle dans l'évaluation, mais aussi dans la définition et la mise au point de solutions transversales est incontournable, parce qu'enfin elles sont seules à même de comprendre le malade dans sa singularité, nos institutions ont un rôle à jouer dans la révolution qui s'annonce. Elles le joueront.

EMMANUEL CORON
Directeur de l'Institut des maladies digestives (IMAD)
du CHU de Nantes

IMAD

L'IA en santé : histoire et modèle

Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

1956	Conférence de Dartmouth	**1997**	Deep Blue bat Kasparov
1957	Création du Perceptron	**2011**	Watson au Jeopardy
1958	Invention du langage Llsp	**2011**	Watson aux Urgences (UK)
1972	Début du programme MYCIN	**2016**	AlphaGo champion
	Premier hiver de l'IA	**2019**	Prix Turing de Y. Lecun
1979	Arrivée du Neogognitron		

par Romain Galmiche
Éditeur indépendant

Délimiter le champ de l'intelligence artificielle en santé passe par la définition de ce qu'est l'intelligence artificielle (IA) elle même. Notre première expérience de celle-ci est d'abord celle d'un ensemble de techniques permettant l'agrégation et le traitement de données hétérogènes et/ou complexes. Les dispositifs en santé sont en effet relatifs à : 1) l'interprétation automatique d'informations, 2) la fusion des données, 3) la construction de représentations à partir d'informations. Cette première délimitation du champ de l'IA, pour opérative qu'elle soit, reste frustrante. Elle a pour mérite principal de grouper des dispositifs de traitement de l'image, de reconnaissance vocale et d'analyse statistique, voire d'aide à la décision, mais laisse de côté la diversité des processus à l'œuvre. L'effet « boîte noire » de l'intelligence artificielle est bien connu des praticiens. Il peut constituer un piège ou une facilité dans la mesure où il génère une confiance absolue dans la machine sans se soucier de ses limitations.

Notre seconde définition de l'IA, est donc, par-delà la diversité des techniques, celle d'un champ disciplinaire, situé à l'intersection de la logique formelle, des sciences cognitives et de l'informatique, visant à reproduire ou à simuler des opérations propres à l'intelligence humaine. Cette définition, évidemment plus satisfaisante, permet de mieux appréhender la singularité des méthodes et partis pris à l'œuvre dans cette science. Elle rend compte notamment de la complémentarité des approches dites cognitivistes et connexionnistes, qui renvoient à deux conceptions de l'esprit humain, l'une d'un système discret basé sur le traitement logique et modulaire d'un signal, l'autre d'un auto-ajustement effectué à partir d'une pondération [1].

Cette définition, pour satisfaisante et rigoureuse qu'elle soit, n'empêche pas une autre dimension du problème, celle d'une « technoscience » existant dans la société et suscitant attentes et projections. Le concept d'intelligence forte, opposée à l'IA faible actuelle, avait été amené par le philosophe Searle. Ce dernier concluait à l'impossibilité d'une telle IA, capable d'accéder à une vraie compréhension sémantique par simple manipulation de symboles. Depuis, le terme d'IA forte a été réinvesti d'un sens positif, que lui ont trouvé des essayistes persuadés de l'avènement d'une intelligence appelée à supplanter les capacités humaines. Cette dimension fantasmatique, pour accessoire qu'elle semble, ne saurait être ignorée en santé. La présence de thématiques connexes comme le transhumanisme, est là pour le rappeler [2].

Il semble donc opportun en un bref historique de faire le point sur les différentes facettes de l'IA, sur ce qu'elle permet, sur les modèles fondamentaux et expérimentaux mis à l'œuvre et sur les attentes qu'elle a suscitées.

L'IA ou l'histoire d'une émulation entre l'homme et la machine

« Cette bonne vieille intelligence artificielle »

L'intelligence artificielle comme discipline a un acte de naissance reconnu : la conférence de John MacCarthy tenue à Dartmouth en 1956, à laquelle assistent notamment Marvin Minsky et Herbert Simon [3-5]. C'est là que s'énonce un programme ambitieux et cohérent : confier à la machine un certain nombre de tâches cognitives dévolues à l'homme. Cette date précise n'empêche pas cette idée de s'insérer dans une généalogie déjà ancienne. On peut citer notamment Alan Turing et sa machine du même nom où, pour la première fois, est confiée à un « proto-ordinateur » la tâche d'infirmer ou de confirmer un théorème mathématique. Autre proposition de Turing, la reproduction de processus mentaux chez l'animal ou le jeune enfant, plus faciles à isoler et à comprendre que ceux de l'adulte, est destinée à un riche avenir…

L'intelligence artificielle dès ses débuts est aussi très vite en concurrence avec un concept voisin, celui de cybernétique : cette science, fondée par Norbert Wiener, et qui se définit comme la science de la régulation des systèmes complexes, et est à l'origine de la robotique. Encore aujourd'hui les contours des deux disciplines, AI et robotique, peuvent paraître flous. Ainsi, une voiture autonome inclut un dispositif d'intelligence artificielle. De même, l'apparition contemporaine de l' « Internet des objets » rend possible l'interopérabilité via IA de robots, notamment en santé : pace-maker, robots chirurgicaux et autres, seraient ainsi susceptibles d'envoyer et d'intégrer en temps réel des données complexes. Par ailleurs si, sur un plan plus théorique, la cybernétique médicale, axée sur l'idée de régulation interne, a permis l'émergence d'une « physiologie mathématique » annonciatrice du concept de biomarqueurs, l'intelligence artificielle en santé se consacre, elle, à l'identification fine de ces derniers [6]…

Dans un premier temps, la recherche en intelligence artificielle a d'abord pour objet de reproduire certains traits mentaux, qui nous permettent de penser et d'agir en un temps limité, avec des ressources finies et des informations imparfaites [3,4]. La première intelligence artificielle répond ainsi à une problématique heuristique : optimiser l'utilisation des algorithmes. On peut citer l'élagage alpha-béta de MacCarthy, méthode pour explorer un arbre décisionnel en éliminant a priori les branches suboptimales, qui est depuis longtemps utilisée dans les programmes d'échecs. De même l'algorithme A*, algorithme du plus court chemin, est à la base d'applications d'usage courant comme le GPS mais aussi à l'œuvre dans des techniques d'imagerie comme l'IRM. « La première idée centrale de la période pré-1962 a été de minimiser l'ampleur des essais et erreurs » (Marvin Minsky). Ce souci de diminuer le coût de fonctionnement des programmes est inséparable d'une exigence formelle. Un des premiers langages symboliques, Lisp, devient vite le langage de prédilection de l'IA, basé sur le lambda calcul, c'est-à dire sur une théorie mathématique dédiée à l'obtention de solutions avec des ressources limitées et à la catégorisation des problèmes selon des classes de difficulté. De fait, l'une des tâches les plus délicates est de modéliser une logique non monotone, c'est-à-dire susceptible de revenir sur un postulat d'entrée…

Cette approche formelle, axée sur la logique, est aussi complétée par l'utilisation des statistiques. Les réseaux bayésiens permettent notamment de représenter un ensemble de variables sous la forme d'un graphe [7]. Ils sont à la base de systèmes d'aide à la décision, puisqu'ils permettent à un agent de s'orienter au fil de nœuds logiques où est calculée la plausibilité d'une hypothèse en fonction d'une probabilité. Ce système, basé sur l'induction, permet de calculer des inférences, c'est-à-dire des propositions

tenues pour vraies en raison d'une proposition préalable. Les réseaux bayésiens sont une aide puissante au diagnostic médical, mais sont aussi utilisés en imagerie, puisque des modélisations du volume tumoral via TDM et PET Scan reposent sur l'algorithme FLAB *(Fuzzy Locally Adaptive Bayesian Segmentation)*, le modèle probabiliste bayésien permettant la délimitation de régions homogènes de la tumeur et de zones floues de transition entre ces régions [8,9].

Les réseaux neuronaux

L'effort logique cité et l'exploration des arborescences constitue ce que d'aucuns ont pu appeler la GOFAI (*Good Old Fashioned Artificial Intelligence*). Au tournant des années 60 une nouvelle piste est explorée, basée sur la reproduction des circuits neuronaux. Le neurone artificiel s'inspire notamment des travaux des neurologues Warren MacCulloch et Walter Pitts. Il comprend une entrée et une sortie : si le signal d'entrée atteint une valeur seuil, un signal de sortie est modulé en fonction d'un « poids synaptique » prédéterminé. La mise en réseau de ces neurones artificiels constitue une des principales branches de l'apprentissage-machine (Machine Learning). Suivant ce qu'on a appelé la « règle de Hebb », un réseau de neurones apprend en renforçant les connexions synaptiques stimulées, d'où le principe : « *cells that fire together, wire together* ». Le premier réseau de neurones est constitué par le psychologue Rosenblatt en 1957. Il s'agit du perceptron : constitué d'une couche de neurones, il constitue, au moins au départ, une machine de plus d'une tonne, subventionné par l'ARPA (Agence de la défense américaine), et qui opère comme un classificateur binaire, c'est-à-dire séparant deux classes (*figure 1*). On parle d'apprentissage supervisé, puisqu'est nécessaire un « entrainement » qui consiste à ajuster les poids jusqu'à ce que les données d'apprentissage rentrent dans la bonne classe. Cette méthode est prometteuse, notamment pour la reconnaissance d'images, puisqu'elle permet la généralisation de son utilisation, à condition que les cas présentés ne soient pas trop différents des données d'apprentissage.

Cette intelligence artificielle suscite donc dans les années 60 les projections les plus optimistes, voire les plus délirantes : reconnaissance de la parole, traductions instantanée et invincibilité aux échecs sont d'ores et déjà évoquées. Or, ces perspectives se heurtent au tournant des années 70 à un « hiver de l'intelligence artificielle » qui se traduit par une chute brutale des subventions. Est en cause sans doute l'irréalisme de certaines projections mais aussi un certain nombre de limitations. L'essor de la micro-informatique et des langages compilés relativise d'abord les prouesses une intelligence artificielle perçue comme coûteuse et tributaire d'un matériel dédié (les machines Lisp). Surtout, le livre de Minsky et Papert *Perceptron : an introduction to computationnal Geometry* dévoile les limites rédhibitoires du Perceptron. Est notam-

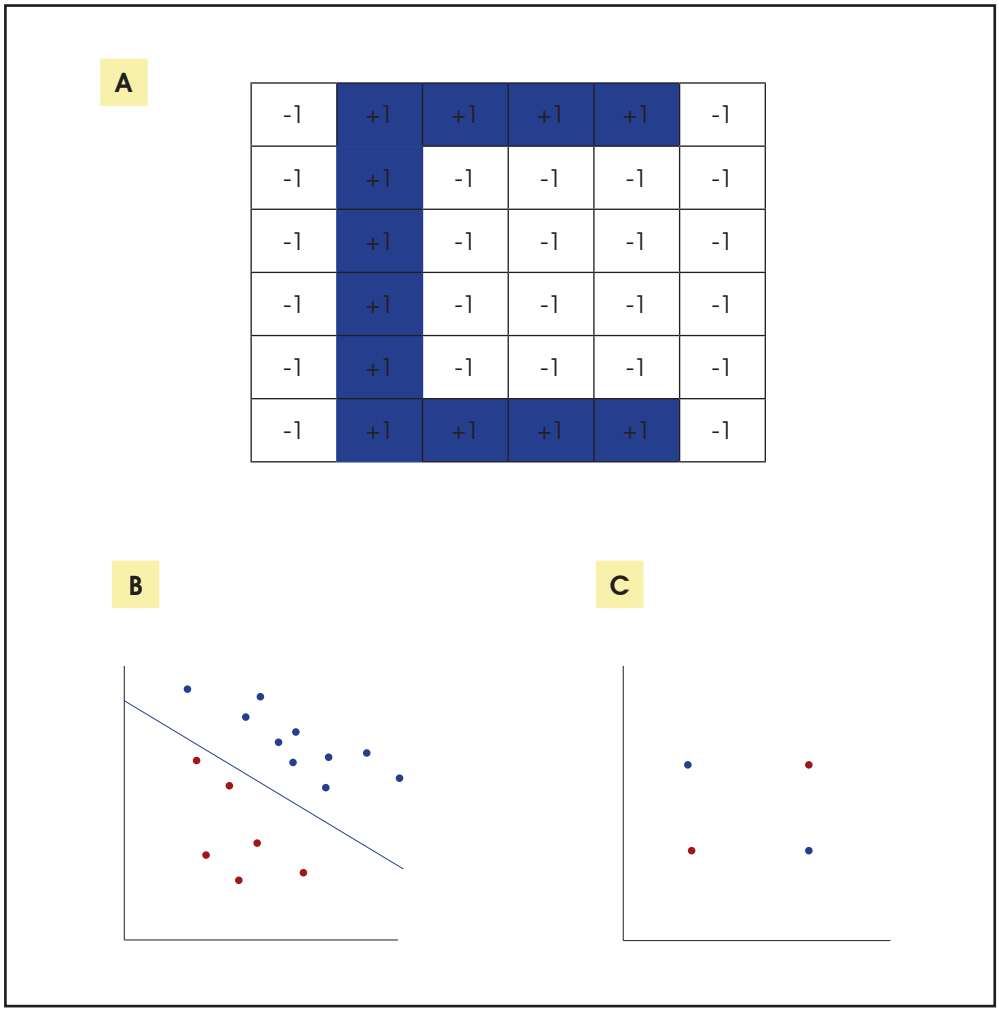


FIGURE 1. Fonctionnement et limite du perceptron. A) Patron destiné à déterminer les poids synaptiques d'un réseau de neurones dans un espace à deux dimensions. Ici, le pattern de reconnaissance visuelle est celui d'un C ; B) classification binaire opérée d'après une fonction linéaire ; C) classification non linéaire, relevant du Ou Exclusif en logique booléenne.

ment mise en lumière l'incapacité pour un réseau de neurones à une couche de traiter les fonctions non linéaires, c'est-à-dire qui relèvent en logique booléenne du OU exclusif (*figure 1C*) [3,4].

Cette impossibilité, qui sonne de prime abord le glas de l'intelligence artificielle est toutefois surmontée par les réseaux de neurones multi-couches. Ceux-ci ne sont pas limités par les fonctions non linéaires et offrent des perspectives nouvelles en termes de reconnaissance de formes. Le Neocognitron de Fukushima s'inspire ainsi dans les années 80 des travaux de neurosciences sur la rétine du chat en associant une couche de neurones simples à plusieurs couches de neurones complexes. Ces dernières permettent notamment de de maintenir la représentation de l'objet même s'il bouge dans le champ visuel. De fait, si dans le Neocognitron la première couche de neurones est entrainée comme un perceptron, les couches sous-jacentes sont, elles, entrainées à l'aveugle : l'apprentissage dans ce dernier cas est dit non supervisé. Aussi, les réseaux neuronaux multi-couches qui se développent dans les décennies suivantes mettent en place des procédures d'apprentissage plus ou moins autonomes, notamment d'apprentissage dit renforcé, l'on présente à la machine un motif (pattern) et où on la récompense lorsqu'elle parvient à le reconnaître. À ce stade, reste néanmoins entier le problème de l'extraction des caractéristiques. Enfin, l'étiquetage des données reste une composante et une limitation essentielles du Machine learning.

La problématique de l'apprentissage des réseaux multi-couches peut être reformulée en termes de diminution de la fonction de coût, qui traduit l'écart entre le comportement d'un modèle et le comportement désiré. Si la rétropropagation du gradient de cette fonction de coût, de la sortie vers l'entrée, par différentiation automatique, n'est pas spécifique au deep learning, elle en constitue néanmoins une caractéristique essentielle. Ainsi, le système peut, dans le cas d'un appentissage non supervisé, découvrir les interdépendances entre les variables d'entrée sans être entrainé pour une tâche particulière. Cette méthode d'apprentissage des réseaux multicouches a notamment, à partir des années 2000, donné des résultats spectaculaires en termes de reconnaissance d'images, bénéficiant notamment de l'essor de banques de données d'images (notamment ImagesNet, créée par des universitaires américains) et du développement des cartes graphiques apportant la puissance de calcul désirée. Les réseaux dits convolutifs permettent ainsi la reconnaissance d'un motif quelle que soit sa position dans la structure du réseau.

Les réseaux neuronaux profonds, après avoir été en butte dans un premier temps à un certain scepticisme, ont rencontré la faveur du public grâce à des réussites spectaculaires (AlphaGo, 2016). Néanmoins ils ne sauraient constituer la panacée. Un grand

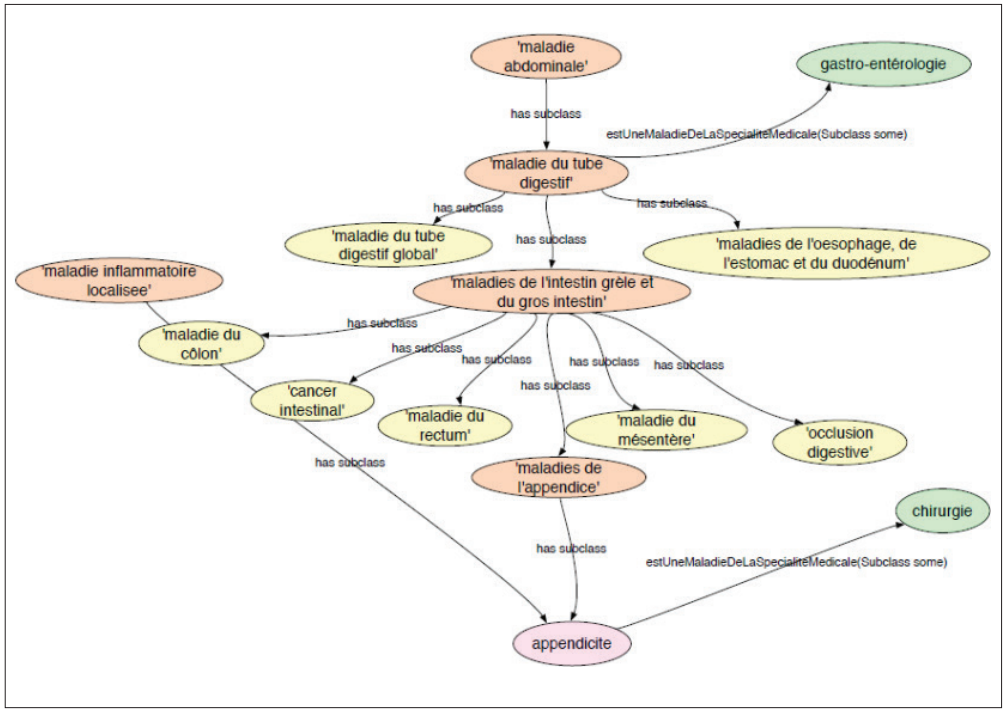


FIGURE 2. Ontologie convergeant vers un diagnostic d'appendicite dans le cadre du projet Lerudi. Dans le contexte de l'IA symbolique, les connaissances sur un domaine médical sont organisées dans une ontologie, c'est-à-dire un ensemble structuré de concepts et de leurs relations, qui décrit un domaine en respectant des principes des langages formels. Une ontologie fournit un modèle de données permettant de raisonner à l'intérieur du domaine en question.

nombre de systèmes sont composites et font appel autant aux ontologies complexes de l'IA formelle (*figure 2*) et aux modèles bayésiens qu'au deep learning. Ensuite, parce qu'une limitation forte du deep learning est un déficit d'explicabilité : l'observateur n'a pas accès à l'algorithme d'apprentissage. Enfin, si les réseaux profonds constituent un outil puissant pour déceler des corrélations, ils ne permettent pas de réellement saisir des causalités. « *L'IA actuelle n'a pas de sens commun, or le sens commun est essentiel. Il conditionne notre lien au monde. Il remplit les blancs et comble l'implicite* » (Yann Le Cun) [3].

Intelligence artificielle et santé, une relation déjà ancienne

L'introduction des techniques d'intelligence artificielle en santé ne saurait se limiter à un schéma linéaire Arrivée des dispositifs à maturité -> application en santé. Au contraire, dès les débuts de l'intelligence artificielle, la médecine apparaît comme un champ de prédilection de la recherche en IA, puisqu'elle correspond à un domaine à part de la « rationalité limitée » (Herbert Simon), c'est-à-dire un domaine où un agent prend une décision en fonctions d'informations limitées. Le glissement conceptuel qui s'opère alors entre rationalité comprise comme idéal et rationalité dite procédurale, où ce qui compte est la conformité du processus cognitif à un certain nombre de règles, est à l'œuvre dès la première application en santé, le MYCIN.

MYCIN et Watson : des modèles séminaux

Le MYCIN est un système expert, l'un des premiers du genre, développé à Stanford au début des années 70 [7]. Il consistait en une machine à inférences causales dotées de près de 600 règles qui lui permettaient de lister les causes probables d'une infection et de suggérer un traitement. Il a été conçu dans une double optique : 1) stimuler la recherche en intelligence artificielle, notamment la théorie des graphes, par l'émulation du raisonnement médical ; 2) offrir au médecin une aide effective au diagnostic basée sur des données solides. Le contexte d'alors est celui du développement des logiques algorithmiques en santé - qui permettent à l'exercice de la médecine de gagner en reproductibilité et de s'abstraire des biais d'interprétation - et des premières bases de données de publications médicales, avec l'apparition Pubmed… Le MYCIN reste un projet ambivalent destiné à la fois à imiter le médecin et à suppléer ses insuffisances. Il met en place des catégories qui perdurent encore aujourd'hui, notamment dans le domaine du design d'application où la réalisation coordonnée du système par un ingénieur et un spécialiste du champ concerné est pionnière. Néanmoins, cette application échoue à s'installer dans la pratique. Son efficacité et sa pertinence sont moins en cause que la sous-estimation du temps nécessaire pour mettre à jour la base de données. La principale utilité du MYCIN reste pédagogique : il demeure un outil d'apprentissage pour le médecin.

L'intelligence artificielle en santé n'échappe pas à l'hiver de IA et il faut attendre 2010 et l'arrivée de Watson (IBM) [5] pour qu'elle refasse réellement parler d'elle, selon un modèle tout à fait différent. Ici l'on n'a pas affaire à un système dédié à la sphère médicale, mais à un programme d'exploration du langage naturel (Deep QA) polyvalent et d'abord mis à l'œuvre dans un jeu télévisé, le Jeopardy, où le candidat doit répondre à une question de culture générale. Sa puissance de calcul lui permet de s'orienter dans le web sémantique pour trouver la bonne réponse. Cette capacité peut être transposée au domaine médical pour fournir une aide au diagnostic : Watson utilise le contexte d'un cas pour générer une liste de diagnostics. Là où le MYCIN procédait d'abord par induction, Watson se montre capable d'abduction : il formule des hypothèses d'après des corrélations avant de les vérifier sur des cas précis. Cette capacité de la machine la rapproche de la « sérendipité » humaine, la faculté de découvrir des schémas généraux à partir de données non structurées [8].

Une IA pour quelle médecine ?

Cette capacité nouvelle du deep learning à opérer des corrélations entre en résonance avec le souci d'une médecine 4p : prédictive, personnalisée, préventive et participative. Relève ainsi d'une médecine participative une application qui permet au patient de détecter seul la malignité éventuelle d'une lésion dermatologique avec une fiabilité supérieure à celle d'un expert. Des applications d'IA en santé sont susceptibles d'être implémentées sur des appareils grand public (téléphone, montre connectée, etc.) et leur caractérisation comme étant ou non des dispositifs médicaux à part entière est une question complexe : l'une des premières utilisations de Watson dans la sphère médicale concerne la prise en charge par un « bot », un assistant vocal, de l'accueil des patients aux urgences britanniques.

L'IA en santé va aussi dans le sens de la personnalisation des soins. Elle favorise l'agrégation de données complexes, notamment en termes d'imagerie, et permet au soignant d'avoir accès à une caractérisation fine et évolutive des pathologies. La notion de « patient numérique » (N. Ayache) a deux portées : 1) sur un plan clinique, la modélisation par l'IA permet un traitement individualisé et déterminé d'après une modélisation prédictive de la réaction du patient ; 2) après anonymisation et aplatissement des données, le regroupement au sein de bases de données populationnelles offre un intérêt majeur pour la recherche médicale.

L'avenir de la recherche s'écrit-il ***in silico*** ?

La mise en place en place de bases de données populationnelles renouvelle en effet les perspectives de la recherche, qu'elle soit clinique ou relative à la santé

publique. L'intelligence artificielle détermine des corrélations insoupçonnées entre phénotypes et génotypes, ou entre des populations et leur environnement. Elle peut ainsi mettre en évidence des signaux faibles, habituellement passés sous silence, favorisant ainsi une gestion proactive et coût-efficace des pathologies. De même dans l'évaluation pharmaceutique, le deep learning et l'analyse des données *in silico* offrent des solutions alternatives aux méthodes de référence (notamment la méthode de régression Cox), basées sur sur la comparaison d'une cohorte avec un groupe témoin. Avec les algorithmes SCCS (*Self Controlled Case Series*), l'évaluation des effets secondaires se fait désormais à l'échelle de l'individu lui-même, par mise en comparaison de la période où le patient a été sujet à un effet indésirable et celle qui la précède. Le travail d'homogénéisation de la cohorte ciblée est quant à lui réduit à son minimum. Cette méthode a notamment déterminé le retrait en France d'un antidiabétique, la pioglitazone.

De même, les études *in silico* sont appelées à être partie prenante non seulement de l'évaluation des solutions basées sur des données mais aussi de leur élaboration. Les GAN, ou *Generative Adversarial Networks* (réseaux antagonistes génératifs) permettent notamment de compléter un système de données lacunaires : un système génère les données manquantes dont un réseau « rival » évalue la pertinence. On touche là à un fonctionnement des réseaux qui se rapproche de la créativité humaine…

Aujourd'hui, toutefois l'analyse génétique des marqueurs cancérologiques offre un aperçu frappant de la complémentarité des apprentissages supervisés et non supervisés, de leurs forces et de leurs limites. Le séquençage génomique des cancers suppose en effet une puissance d'analyse et de stockage peu commune : il faut entre 10 gigaoctets et plusieurs téraoctets pour stocker un seul génome. Or, des modèles graphiques particuliers, les chaînes de Markov cachées, permettent d'inférer automatiquement l'annotation du génome, et notamment de déterminer les régions codantes. Elles constituent une forme d'apprentissage non supervisé qui permet de par sa flexibilité de reconstruire l'histoire évolutive d'un cancer en termes d'altérations génétiques. Il est de même possible d'identifier les processus impliqués, tels que tabac, soleil, etc. grâce à des techniques de factorisation de matrice, les mêmes que celles qui sont utilisées par les plateformes de vidéo à la demande pour personnaliser leur contenu. Cet apprentissage non supervisé n'empêche nullement le recours à l'apprentissage supervisé, par exemple pour identifier un profil type de récidive ou de non récidive. Cette approche classificatoire pâtit néanmoins du décalage entre le grand nombre de données par individus et le nombre limité des individus catalogués : les performances restent médiocres par rapport aux modèles basés sur les seules données cliniques.

IA et santé en France

Le rapport Villani, rendu public en 2018, est une date importante pour l'intelligence artificielle en santé en France. Son élaboration, dans un contexte de méfiance envers la robotique en général, répondait au souci d'un développement économique et universitaire harmonieux de la discipline. De même que la Cnil dans les années 70, l'insistance sur l'usage éthique des données ainsi que sur la transparence des algorithmes, voire sur l'usage de logiciels « open source », est censée vaincre la technophobie du public et offrir un cadre directeur en plein accord avec la RGPD (régime de protection des données) européenne. Deux réquisits s'imposent très vite :

- l'obligation de développer un tissu économique et industriel dense, capable de porter des initiatives pertinentes et soutenables à long terme ;
- la nécessité de regrouper des données de santé publiques déjà considérables au sein d'un système unique et centralisé (*tableau 1)*.

La logique *top down* de regroupement des données de santé et l'intégration poussée du secteur privé s'appuient notamment sur le modèle israélien, celui d'un écosystème riche et dynamique reposant sur une valorisation intelligente des data. Ce choix de la centralisation reste toutefois un choix ancré dans les pratiques et l'histoire françaises, à l'opposé de ce que l'Allemagne choisit au même moment, sans que l'intéropérabilité des deux voisins soit particulièrement à l'étude.

Certaines initiatives comme la création de l'institut PRAIRIE, axé sur le regroupement de plusieurs institutions universitaires prestigieuses, correspondent au souhait de développer des compétences pas toujours techniques mais aussi relatives aux soft skills nécessaires à l'ingénierie de lA. Cette approche intéresse tout particulièrement l'IA en santé où « l'interopérabilité humaine », le design d'application et l'expérience utilisateur sont des champs parfaitement légitimes.

Cependant, l'annonce la plus spectaculaire reste la création d'un National Health Data Hub, construit autour d'une gouvernance public/privé, qui est supposé être très vite opératoire et devenir d'utilité publique dès 2019. Certaines contraintes politico-légales s'avèrent toutefois vite insurmontables : les normes très rigoureuses de la Commission Informatique et libertés (Cnil) en termes de bruitage et d'anonymisation des données imposent le recours à un prestataire étranger, Microsoft, plutôt qu'à un partenaire français novice dans la gestion des données de santé. C'est ce recours même qui est sanctionné par un arrêté du Conseil d'Etat… Du fait de ces achoppements, la logique top down se voit vite redoublée par une logique *bottom up* de bases de données régionales, parmi lesquelles : le Ouest

Tableau 1. Principales bases de données en France et organismes constituant le Système national des données de santé (SNDS).

Base de données	Types de données	Organisme
SNIIRAM (Système national d'information interrégimes)	- 15 bases de données thématiques de données agrégées appelées datamarts orientés vers une finalité particulière : suivi des dépenses (Damir), analyse de l'offre de soins libérale, biologie, pharmacie, dispositifs médicaux, établissements privés ; - échantillon général des bénéficiaires (EGB) au 1/97e de la population protégée : l'EGB permet de réaliser des études longitudinales et d'analyser le parcours individuel de près de 660 000 bénéficiaires en ville et à l'hôpital ; - base de données individuelles des bénéficiaires (DCIR) pour réaliser des études sur la consommation des soins	Cnam
Cépi-DC	Statistique nationale des causes médicales de décès	Inserm/Insee
PMSI (Programme de médicalisation des systèmes d'information)	Recueil systématique et minimal d'informations administratives et médicales qui sont utilisées principalement pour le financement des établissements de santé (tarification à l'activité) et pour l'organisation de l'offre de soins (planification).	Hôpitaux

Data hub, à l'heure actuelle première base de données en santé en France qui émane du GSC HUGO et regroupe les CHU de Nantes, Angers et Ancenis ; l'EDS pour l'AP-HP.

Enfin, l'intelligence artificielle et son intégration dans la pratique médicale concernent aussi la Haute Autorité de Santé. La mise à disposition fin 2020 d'une grille d'analyse des algorithmes faisant appel à l'intelligence artificielle participe de la volonté d'orienter les fabricants en amont et de mettre en place un cadre descriptif normatif utile et pertinent. Si la première validation des produits est tributaire des marquages CE en vigueur, la HAS entend au travers de son évaluation des dispositifs médicaux aussi aider à la description du fonctionnement de ce type de technologie pour en améliorer la connaissance et rendre leur fonctionnement plus transparent.

L'auteur tient à remercier Harold Mouchère de ses précieuses remarques.

RÉFÉRENCES

1. Smolensky P.. IA connexionnisteIA cognitiviste et cerveau. In *: Introduction aux sciences cognitives,* ed. Daniel Andler. Paris : Gallimard, 1992
2. Ganascia JY. *L'intelligence artificielle.* Dominos. Paris : Flammarion 1993
3. Le Cun Y, Dehaene, Girardon J. *La Plus Belle Histoire de l'intelligence*. Paris, Odile Jacob, 2018.
4. Le Cun Y. *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*. Paris : Odile Jacob, 2019.
5. Jean A. Une brève introduction à l'intelligence artificielle. *Médecine sciences* ; 36 : 1059-67
6. Haiech J. Parcourir l'histoire de l'intelligence artificielle, pour mieux la définir et la comprendre. *Med Sciences* 2020 ; 36 : 919–23
7. Matuchansky. Intelligence clinique et intelligence artificielle : Une question de nuance. *Médecine/sciences* 2019, 35 : 797-803. Doi : 10.1051/medsci/2019158.
8. Korb KB, Nicholson A. *Bayesian artificial intelligence.* London : Chapman & Hall/CRC Press UK, 2004.
9. Negnevitsky M. *Artificial Intelligence: A Guide to Intelligent Systems,* Addison-Wesley, 2011.

10. Levas A, Bagchi S. Gondek D, Mueller ET. Watson: Beyond Jeopardy! *Artificial Intelligence* 2013 ;199-200 : 93-105

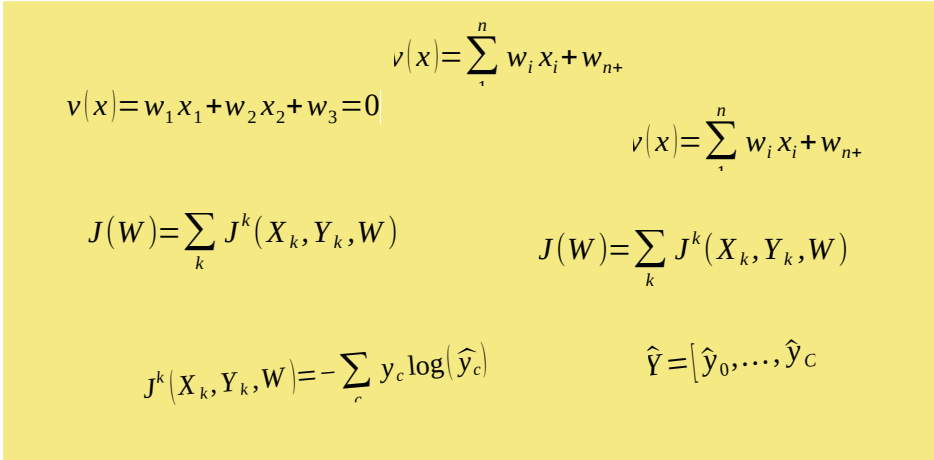
11. Raikomar A, Dean J, Kohane I. Machine Learning in Medicine. *NJOM* 2021 ; 380 : 1347-58

12. Nordlinger B, Villani C. *Santé et intelligence artificielle.* Paris : CNRS, 2018.

13. Ayache N. *Le Patient numérique.* Leçon inaugurale. Paris : Collège de France, 2013.

14. Cuggia M , Combes S. The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare. *Yearb Med Inform* 2019 ; 28: 195–202. doi: 10.1055/s-0039-1677917

Principes de fonctionnement et applications de l'IA



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Harold Mouchère
Université de Nantes,
Laboratoire des Sciences du Numérique de Nantes

L'« Intelligence Artificielle » est un terme qui recouvre de nombreux concepts. Ce chapitre se focalisera sur une partie de cette grande famille d'algorithmes et d'approches : les réseaux de neurones profonds. Le lecteur intéressé pourra se référer à des ouvrages dédiés comme [1].

Principe de base des réseaux profonds

Il existe de nombreuses architectures de réseaux de neurones, mais elles ont toutes en commun deux choses : leur brique de base est le neurone formel et l'algorithme de la descente de gradient est utilisé pour l'apprentissage. Une fois ces fondamentaux expliqués, nous définirons les architectures plus modernes : les réseaux convolutifs et les réseaux récurrents.

Du neurone isolé au perceptron multi-couches

Dès les années 50 les chercheurs en cybernétique tentaient de reproduire le comportement d'un neurone biologique. Le neurone artificiel moderne ressemble encore à celui proposé par Rosenblatt en 1957, schématisé dans la *figure 1*.

Le neurone formel est constitué de n entrées numérotées de *1* à *n*. Ces entrées peuvent être des caractéristiques extraites (automatiquement ou non) de la forme à reconnaître. Chaque entrée est pondérée par un poids w_i. Le « corps » du neurone réalise la combinaison linéaire de ces entrées pondérées par ces poids. Une entrée supplémentaire, toujours fixée à la valeur *1*, permet d'ajouter un biais à cette somme appelée fonction potentielle $v(x)=\sum_{1}^{n} w_i x_i + w_{n+1}$. Ce potentiel est ensuite passé dans une fonction d'activation *f* qui réalise la décision. Différentes fonctions peuvent être utilisées en fonction de l'usage du neurone formel.

Le neurone formel peut donc être résumé à une simple équation : $y=f(v(x))$. La *figure 2* donne une représentation de cette fonction pour un problème en 2 dimensions : (*x1,x2*) sont les deux caractéristiques et la ligne d'équation $v(x)=w_1x_1+w_2x_2+w_3=0$ représente la frontière de décision. La difficulté consiste à trouver un ensemble de coefficients (w_1,w_2,w_3) qui minimise le nombre d'erreurs (les points du mauvais côté de la droite). Pour cela nous utilisons des algorithmes d'apprentissage qui seront expliqués dans la section suivante.

Un neurone seul ne permet de résoudre que les problèmes linéairement séparables et à deux classes. Pour des problèmes plus complexes, l'idée est de projeter les données successivement dans des espaces de représentation où on espère que le problème devienne linéairement séparable. Chaque projection se fait grâce à une couche de neurones. Il s'agit alors d'un perceptron multi-couches (MLP [*Multi Layer Perceptron*]). Comme le montre la *figure 3*, un MLP est constitué d'une couche d'entrée (le vecteur X des entrées), d'une ou plusieurs couches cachées réalisant les changements d'espace, et enfin de la couche de sortie dite couche de décision. Ces différentes couches sont complètement connectées, chaque neurone est connecté à tous les neurones de la couche précédente avec un poids. La couche de décision finale peut être considérée comme un ensemble de séparateurs linéaires à 2 classes : chaque neurone tente de séparer sa classe de toutes les autres. Si la fonction d'activation utilisée dans les neurones des couches cachées est non linéaire, alors le nouvel espace de représentation, appelé espace latent, peut permettre à la couche suivante de résoudre le problème de façon linéaire. C'est ce changement d'espace de représentation des données qui fait la force des réseaux de neurones.

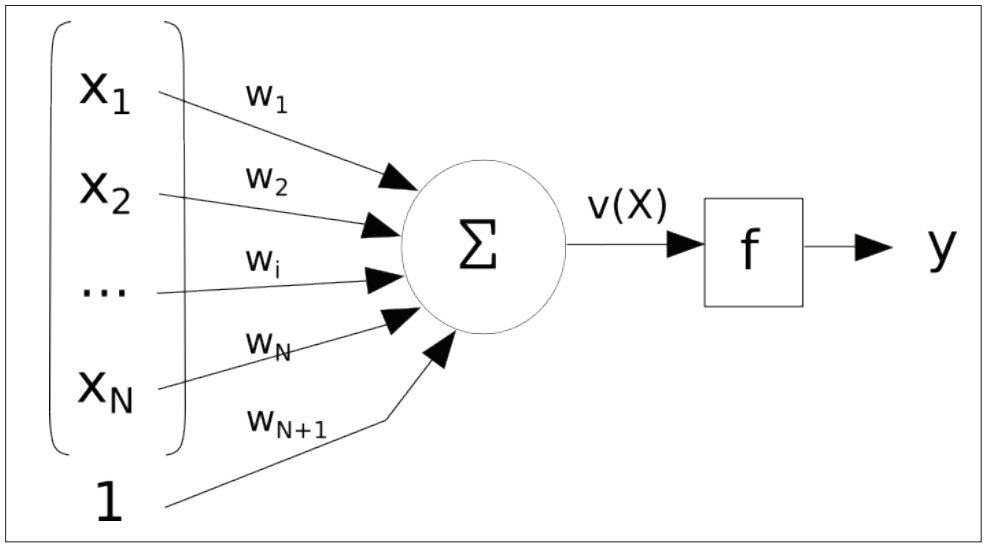


FIGURE 1. Le neurone formel : les entrées x_i sont pondérées par les poids w_i , ce potentiel $v(x)$ est passé à la fonction d'activation f qui donne la sortie y.

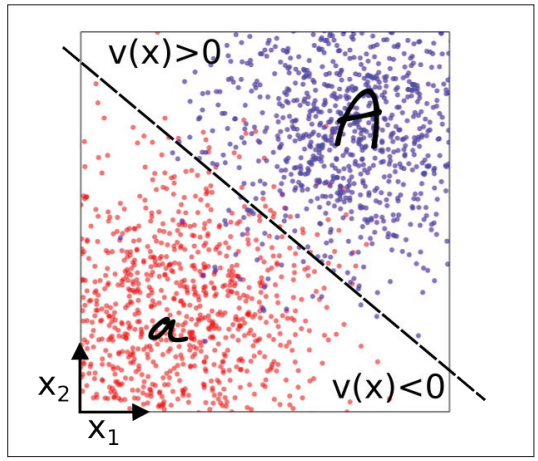


FIGURE 2. Un neurone seul peut être vu comme un séparateur linéaire : exemple en 2D pour un problème non linéairement séparable.

Cette architecture est limitée par sa complexité. En effet, le nombre de connexions (et donc de paramètres à apprendre) augmente de façon polynomiale avec le nombre de caractéristiques utilisées, le nombre de neurones dans les couches cachées et le nombre de sorties.

Apprentissage automatique

Les algorithmes permettant de trouver les meilleurs poids à partir d'exemples bien étiquetés sont dit d'apprentissage supervisé. Les couples données d'entrée et sorties attendues sont notés (X_k, Y_k) ; il s'agit de la base d'apprentissage. En utilisant aléatoirement les exemples un par un, on vient estimer progressivement les meilleurs poids, en suivant trois étapes :

- calculer la sortie du réseau de neurones pour les C classes: $\hat{Y}=[\hat{y}_0,\ldots,\hat{y}_C]$,
- estimer l'erreur commise par la configuration actuelle des poids W : $J(W)$,
- modifier les poids W pour réduire cette erreur.

L'algorithme utilisé pour réaliser cet apprentissage encore aujourd'hui est nommé Descente de gradient stochastique. Pour utiliser cet algorithme, il faut d'abord définir une fonction de coût $J(W)$ qui mesure la quantité d'erreur commise par le réseau pour l'ensemble des données :

$$J(W)=\sum_k J^k(X_k, Y_k, W)$$

La fonction de coût utilisée doit correspondre au problème qui cherche à être résolu. Pour des problématiques de régression une simple erreur quadratique moyenne est généralement utilisée. On peut aussi chercher à approcher une densité de probabilité avec la fonction de coût de Kullback-Leiber. Lorsqu'il s'agit de classification où les sorties $\hat{y}_c$ sont des probabilités sommant à 1, on utilise généralement une entropie croisée :

$$J^k(X_k, Y_k, W)=-\sum_c y_c \log(\hat{y}_c)$$

Il s'agit donc de trouver l'ensemble des poids W qui minimise la valeur de la fonction de coût choisie sur l'ensemble de données d'apprentissage. Pour cela nous modifions la valeur des poids du réseau dans le sens opposé au gradient du coût :

$$\Delta W_k=-\eta \frac{\partial J^k}{\partial W}$$

Ce calcul dépend de la fonction de coût utilisée, du choix de la fonction d'activation de chaque neurone, des connexions entre ces neurones et d'un paramètre appelé taux d'apprentissage. Si ces calculs et leur implémentation étaient complexes il y a quelques années, ils sont maintenant complètement transparents pour les utilisa-

teurs des bibliothèques logicielles récentes. Une des difficultés de cet apprentissage est appelée la dispersion du gradient (*gradient vanishing*). À chaque couche du réseau la quantité de gradient se disperse dans les neurones précédents. L'usage de fonctions d'activation linéaires par partie (*Rectified Linear Unit,* ReLU) permet de réduire cet effet.

L'apprentissage d'un réseau est donc un processus itératif et stochastique. À chaque modification des poids, l'erreur diminue pour les exemples utilisés à cette itération. À force de rencontrer des exemples différents, les poids tendent vers une configuration globalement intéressante. La fonction de coût étant une surface assez chaotique avec plusieurs minimums locaux, des stratégies d'apprentissage permettent d'explorer cet espace et d'augmenter les chances de trouver rapidement un minimum global. Une architecture suffisamment complexe permet toujours de diminuer l'erreur sur la base d'apprentissage. Pour éviter cette situation de surapprentissage, il est important de surveiller l'évolution de la fonction de coût sur une base de validation.

Réseaux convolutifs et réseaux récurrents

La complexité des couches complètement connectées rend impossible leur application directe sur des images qui sont des entrées de trop grande dimension. La solution démocratisée par LeCun en 1998 [2] consiste à utiliser un réseau à convolution qui réduit énormément le nombre de paramètres. En effet, comme le montre la *figure 4*, un noyau de convolution, qui peut être vu comme un petit perceptron, est appliqué sur une petite portion de l'image. Le nombre de poids nécessaire pour ce calcul local est réduit car le neurone n'est pas connecté à toute l'image. Ce noyau est appliqué à l'identique à l'ensemble de l'image en le glissant suivant les lignes et les colonnes (opération de convolution). Avec un seul noyau et donc un seul jeu réduit de poids, il est possible d'extraire une information locale sur toute l'image. On peut appliquer en parallèle plusieurs dizaines de noyaux de convolutions, chacun extrayant une caractéristique locale de l'image. Plusieurs couches de convolution sont ensuite empilées, chacune produisant un nouveau volume de caractéristiques utilisé comme entrée de la couche suivante. Ces cartes peuvent être réduites à certaines étapes en faisant des agrégations locales (*sum-pooling* ou *max-pooling*). Ainsi, les couches successives extraient une information de plus en plus haut niveau et de moins en moins complexe. Une couche finale complètement connectée permet d'utiliser ces caractéristiques pour la décision finale. L'apprentissage de ces réseaux convolutifs se fait exactement de la même manière qu'expliquée précédemment, mais ils tirent parti du partage des poids au sein des couches de convolution pour apprendre rapidement des structures locales. Leur implémentation étant aussi fortement parallélisable, l'usage des cartes GPU permet d'accélérer leur utilisation.

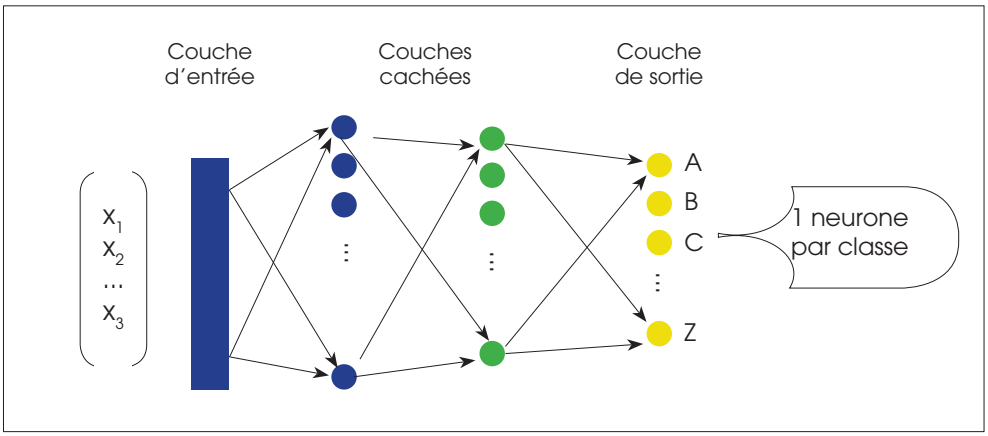


FIGURE 3. Le perceptron multicouches (MLP). Chaque neurone est connecté à tous les neurones de la couche précédente.

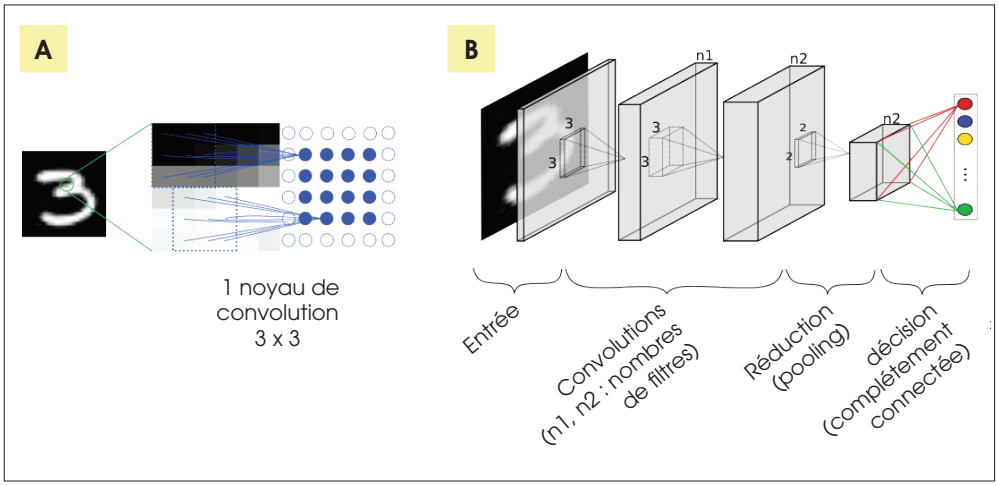


FIGURE 4. Réseau à convolutions. A) une convolution 3 x 3 appliquée à une image 2D. B) Réseau complet simplifié : une image 2D en entrée, 2 couches de convolution, 1 couche de *pooling* et la couche de décision complètement connectée.

Si les réseaux à convolution ont été créés pour traiter des entrées en 2D (images), ils ont aussi été étendus aux traitements des volumes (convolutions 3D) et des séquences (2D+t ou 3D+t). Néanmoins, l'analyse d'une séquence nécessite parfois de prendre en compte un contexte sur une longue période. Les réseaux récurrents permettent de propager une information tout au long d'une séquence pour construire ce type de contexte.

Un réseau récurrent utilise aussi le principe de la convolution car c'est le même neurone qui est appliqué tout au long du signal d'entrée. Mais en plus de ne voir qu'une partie de ce signal, l'activation du neurone est aussi calculée en utilisant la sortie du (ou des) neurones de l'étape précédente. Plusieurs formulations des neurones récurrents ont donc été proposés : GRU (*Gate Recurrent Unit*) et LSTM (*Long Short Term Memory*).

Les réseaux récurrents peuvent être utilisés pour différentes tâches, la classification de séquences étant la plus simple: le réseau est un simple extracteur de caractéristiques qui alimente une couche de décision finale complètement connectée. Les réseaux récurrents permettent aussi de produire une séquence de décisions pour, par exemple, des tâches « séquence à séquence » utilisées notamment en traitement automatique du langage naturel.

Exemple d'application en gastro-entérologie

Les réseaux profonds sont utilisés dans un grand nombre d'applications dans le domaine médical [3]. Nous avons choisi une application en gastro-entérologie [4] pour la détection et classification d'images pathologiques dans des vidéo-capsules endoscopiques sur des patients atteints de la maladie de Crohn.

Il a d'abord fallu collecter et annoter des images (63 patients, 3 498 images) pour des différents types de lésions (6 lésions) et des images non pathologiques. L'étiquetage de ces données est déjà une difficulté. Plusieurs lésions peuvent apparaître sur les images (1.2 étiquettes par image en moyenne). Les données ont été étiquetées par 3 experts et, sur 15 % des images, 1 des experts n'était pas d'accord (consensus par vote). La proportion des différentes classes est aussi un challenge car l'apprentissage étant stochastique, il fonctionne mieux lorsque les classes sont équilibrées (ce qui est rarement le cas). Après le consensus, 60 % des images sont non pathologiques et les lésions n'ont pas du tout la même fréquence (de 5,8 % pour les sténoses à 14,6 % pour les ulcérations de moins de 10mm). La base peut donc être utilisée pour de la détection (2 classes) ou pour de l'identification de lésion (7 classes).

Plusieurs architectures ont été testées parmi les solutions « sur étagère » : ResNet et VGG qui sont les réseaux les plus utilisées dans la commuté d'analyse d'images. Par

exemple VGG19 [5] prend en entrée une image 224 x 224 et est composé de 16 couches de convolution (de 64 à 512 noyaux de 3 x 3 pixels), intercalées avec 5 couches de type *max-pooling* pour agréer localement les informations extraites. La dernière couche est ensuite passée à travers trois couches complètement connectées, avec 4 096 neurones dans les 2 premières, la dernière effectuant la classification (les 2 ou 7 classes de notre problème). Il y a environ 140 millions de paramètres à apprendre.

L'évaluation a été faite en validation croisée pour ne pas être dépendante du découpage des données en bases d'apprentissage-validation-test. Pour le problème de détection des images pathologiques, le meilleur système (ResNet34) obtient une performance moyenne sur les 5 validations de 94,56 % et une aire sous la courbe ROC de 98,23 %.

Les défis de demain

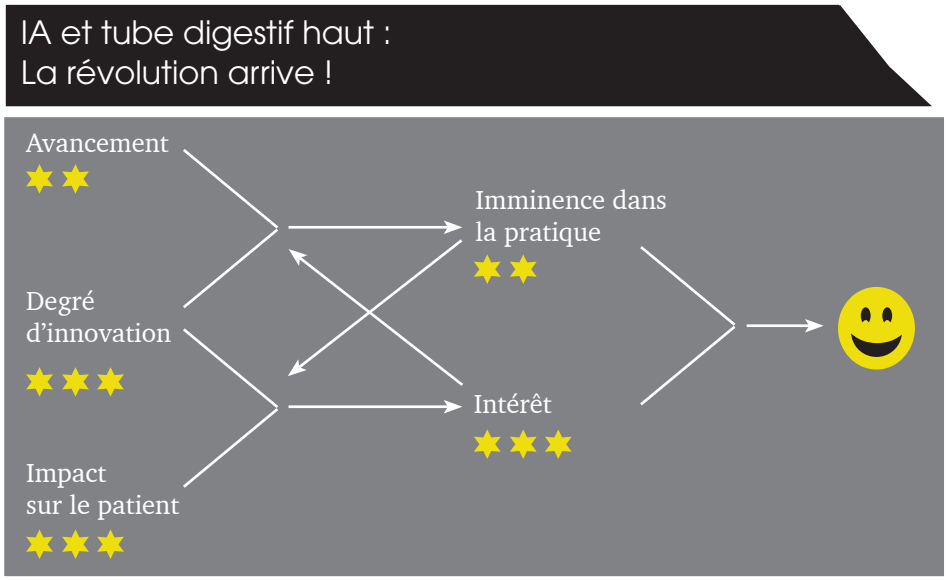
Pour l'instant, les approches qui ont fait leur preuve dans des domaines applicatifs variés sont les systèmes supervisés, c'est à dire utilisant des données étiquetées par des experts humains pour une tâche dédiée. La constitution de grandes bases de données (publiques ou privées) permet donc d'aborder différents applicatifs. Les réflexions en cours sur la propriété et la confidentialité des données, notamment médicales, sont donc des points cruciaux.

Parallèlement à ce courant, les chercheurs en IA essayent de réduire la quantité de données étiquetées nécessaires pour l'apprentissage. Deux solutions semblent se dégager depuis quelques années. Tout d'abord l'apprentissage non supervisé, ou faiblement supervisé, consiste à fournir au réseau des données brutes sans étiquette et à le laisser découvrir seul un espace latent de dimension réduite utile pour d'autres tâches. Le transfert de connaissance poursuit le même but : construire un espace de représentation en utilisant un problème similaire pour lequel beaucoup de données sont disponibles.

Avec la démocratisation de son usage, l'acceptabilité par les utilisateurs des décisions d'IA est un point incontournable, surtout dans le domaine médical. Ce problème est souvent abordé par la création d'une IA explicable, capable de convaincre l'utilisateur, expert ou non, de la pertinence de la décision. Une des pistes explorées pour les réseaux profonds, qui sont considérés comme des boîtes noires inexplicables, est l'usage d'un processus d'attention capable de délimiter dans le signal d'entrée quelle portion ou quelle structure a permis de prendre la décision.

RÉFÉRENCES

1. Goodfellow I, Bengio Y, Courville A. *Deep Learning*, MIT Press, 2016.
2. Lecun Y, Bottou L, Y. Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998 ; 86 : 2278-24, 1998, doi: 10.1109/5.726791.
3. Esteva A, Chou K, Yeung S *et al.* Deep learning-enabled medical computer vision. *npj Digit Med* 2021 ; 4, 5. doi : 10.1038/s41746-020-00376-2
4. Vallée R, De Maissin A, Coutrot A, Mouchère H, Bourreille A, Normand N. CrohnIPI: *An endoscopic image database for the evaluation of automatic Crohn's disease lesions recognition algorithms*. Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113171Q (28 February 2020); doi : 10.1117/12.2543584
5. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014. arXiv:1409-1556.



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Emmanuel Coron [1] & Gabriel Rahmi [2]

[1] *Institut des maladies de l'appareil digestif, CHU de Nantes ; université de Nantes*

[2] *Hôpital Georges Pompidou, AP-HP ; université Paris-Descartes*

De prime abord, l'IA semble plus avancée dans le grêle et le colon, pour différentes raisons. Ainsi, les bénéfices d'une reconnaissance automatisée des lésions en vidéo capsule du grêle sont évidents et ont permis la mise au point de travaux précurseurs. De même, dans le colon, les enjeux d'une détection accrue des polypes coliques afin de diminuer le risque de cancer colorectal et la différenciation des adénomes et des polypes hyperplasies permettant d'adopter une stratégie « réséquer et jeter » (« *resect and discard* ») avec un bénéfice coût/efficacité sont d'une importance majeure. Du fait de ces opportunités économiques, les lésions du grêle et du colon ont été les premières à lancer les chercheurs et les industriels sur la piste de l'IA en endoscopie. De plus, la relative homogénéité des lésions à reconnaitre a probablement facilité les résultats impressionnants des premières études publiées.

Pour autant, le tube digestif haut n'est pas le parent pauvre des développements en IA. En effet, là aussi les bénéfices attendus sont énormes. Les cancers digestifs hauts sont non seulement fréquents, mais également grevés d'un pronostic qui nécessite de dépister les lésions à un stade précoce afin d'envisager les meilleures chances de guérison, ou de survie prolongée, chez ces patients. Ainsi, détectés précocement, la survie de ces cancers digestifs hauts est de plus de 90% [2].

Comment intégrer le développement de l'IA dans un contexte de progrès rapide des techniques endoscopiques et dans la vraie vie ?

Les techniques endoscopiques évoluent rapidement et les générations successives de zooms optiques ou de chromoscopie digitale (*Blue Light Imaging* [BLI], *Linked-Color Imaging* [LCI, *Narrow Band Imaging* [NBI]) mettent en évidence des processus de carcinogénèse de plus en plus précoces. L'un des meilleurs exemples est la visualisation des boucles capillaires intrapapillaires (*Intrapapillary Capillary Loops* [IPCL]) dans la muqueuse épidermoïde œsophagienne, dont l'aspect se modifie au cours des différents stades de dysplasie et d'envahissement de la muqueuse [3]. Le défi devient alors, non plus de mettre en évidence ces anomalies, mais de pouvoir les interpréter avec une expertise suffisante, comme l'a récemment rappelé l'ESGE (*European Society of Digestive Endoscopy*) [4]. De même, les différents états précancéreux au sein de la muqueuse de Barrett sont parfaitement identifiables avec des zooms et des colorants optiques ou digitaux. Mais, compte tenu de la complexité à reconnaître les différents sous-types lésionnels, il devient évident que la difficulté n'est plus d'apprendre à maîtriser une technique comme la dissection sous-muqueuse, mais de sélectionner les lésions qui sont les meilleures candidates à ce traitement et à bien en délimiter les berges et le risque estimé d'infiltration profonde avant de se lancer dans le geste interventionnel. Sur ce plan, il existe une différence flagrante entre le monde occidental et les pays asiatiques, dans lesquels l'utilisation minutieuse de ces outils innovants de diagnostic optique a permis une diminution de la mortalité par cancers digestifs hauts [5-7].

La complexité d'apprentissage de la sémiologie endoscopique avec ces nouveaux outils ne doit pas faire oublier que, dans la majorité des centres dans le monde (la « vraie vie »), une approche « simple », en lumière blanche, par des endoscopistes non experts ou avec des endoscopes non munis de zooms est la règle. De plus, la variété de lésions « de base » que l'on peut rencontrer dans l'œsophage ou l'estomac rend plus difficile les reconnaissances automatiques par des programmes d'IA. En fait, l'enjeu est à la fois de ne pas manquer de lésions à risque avec une approche « de dépistage » mais également de faciliter la caractérisation des lésions complexes pouvant nécessiter un geste expert. L'IA a déjà démontré sa capacité à identifier des microstructures ou de

quantifier dans l'image des motifs architecturaux au niveau du pixel, indétectables par l'oeil humain. Cette supériorité est encore accentuée par la robustesse des systèmes informatiques, insensibles à la fatigue, au stress ou à la lassitude engendrée par la répétition des procédures.

Applications de l'IA en pathologie œsophagienne

Œsophage de Barrett

L'œsophage de Barrett ou endobrachyœsophage (EBO) présente un difficulté particulière. En effet, la dysplasie se présente soit comme une anomalie de relief, soit comme une zone plane avec une irrégularité architecturale de la muqueuse ou des microvaisseaux, ce qui rend très difficile le contourage précis des lésions. L'EBO étant rare voire absent dans les pays asiatiques, aucun groupe chinois ou japonais travaillant sur l'IA n'a encore eu l'opportunité d'y travailler. Cependant, comme souvent dans les travaux menés en endoscopie sur l'EBO, les Néerlandais ont eu un rôle de pionniers [9,10]. Le portail GastroNet, issu du consortium ARGOS, est une vaste database internationale qui a permis de regrouper et de trier 494 364 images endoscopiques, dont 1 544 spécifiquement caractérisées à partir d'EBO dysplasiques et non dyplasiques. Le système d'IA ensuite mis au point a été validé dans deux datasets différents constitués chacun de 160 patients. L'algorithme mis au point a été confronté à un panel de 53 endoscopistes internationaux, tous battus par l'IA dans une étude basée sur une analyse d'images évaluées rétrospectivement…[9]. Cet algorithme a ensuite été incorporé à un système de vidéoendoscopie en temps réel dans une étude pilote portant sur 20 patients (n = 10 EBO dysplasiques vs n = 10 EBO non dysplasiques). Seules les images en lumière blanche étaient utilisées pour comparer l'IA et l'oeil humain, selon un protocole standardisé d'exploration de l'EBO. En effet, depuis la partie distale jusqu'à la partie proximate de l'EBO, l'endoscope était stoppé tous les 2 cm, et 3 images de chacun de ces ni-veaux étaient enregistrées par l'endoscopiste et analysées par le système d'IA. Ceci permettait à la fois d'évaluer la reproductibilité des prédictions faite par l'IA pour un même segment, et de voir si une analyse combinant les différents niveaux permettait d'améliorer les performances du système. Outre l'identification d'une image contenant possiblement de la dysplasie, le système d'IA permettait de contourer la zone et ainsi d'aider l'endoscopiste à réaliser des biopsies ciblées ou une résection endoscopique dans le même temps. Le système avait 75 % de concordance par niveau entre ces 3 images. L'analyse par image montrait une performance diagnostique, une sensibilité, une spécificité, de 84 %, 76 % et 86 %, respectivement. Lorsque le système classait ces images avec un fort indice de confiance, ces chiffres se situaient aux alentours de 91%. Le système détectait 9 patient sur 10 classés par l'endoscopiste comme « EBO dyspla-

sique». Néanmoins, le spécimen de mucosectomie réalisé chez ce seul patient «manqué» par l'IA ne comportait que de la muqueuse de Barrett non dysplasique, reposant ainsi l'éternel problème du *gold standard*. Dans cette étude pilote, le système d'IA était très rapide, avec des temps respectivement de 0,2 s et 0,3 s pour identifier et contourer une lésion potentiellement dysplasique. Ce système s'avère donc extrêmement prometteur et en est cours d'évaluation par des études prospectives internationales *(figure 1)*.

Carcinome épidermoïde et IA

La détection du carcinome épidermoïde de l'œsophage (CEO) est un autre enjeu majeur, à la fois du fait de son incidence encore très élevée dans certains endroits du monde et de son pronostic sombre. L'étude de Guo *et al.* [11] s'est concentrée sur la capacité de l'IA combinée au *Narrow band imaging* (NBI) à détecter des CEO superficiels (*figure 2*). Ce système d'IA chinois a été entraîné à la reconnaissance des images d'intérêt à partir de 6 473 images de CEO et de muqueuse épidermoïde non dysplasiques. Le premier set de validation comportait des images de CEO (n=1 480) *versus* des images diverses (n = 5 191) de muqueuse normale, œsophagites, hétérotrophe de muqueuse gastrique, tumeur sous-muqueuse voire de varices œsophagiennes. La majorité des lésions était plane ou légèrement déprimée, mais de relativement grande taille (34 mm en moyenne). Dans ce dataset, la sensibilité était de 98 % et la spécificité de 95 %. Bien sûr, le biais potentiel porte sur le choix des images envoyées au système d'IA. C'est la raison pour laquelle un deuxième dataset portant sur des vidéos (n = 27 sans zoom et n = 20 avec zoom) était également testé. Avec ou sans zoom, la lésion était détectée avec une sensibilité de 100 % (par une ou plusieurs images). L'analyse par image montrait des performances nettement supérieures des images avec zoom (sensibilité 96,1 %) par rapport aux images sans zoom (sensibilité 60,8 %). Cette différence était expliquée par un risque accru d'artefacts de mouvements en l'absence de zoom selon les auteurs. Dans un dernier set de validation comportant 33 cas, la vidéo entière (c'est-à-dire non filtrée, durant en moyenne 80 secondes) était envoyée au système. Dans ce dernier set, la spécificité par image et par patient était de 99,9 % et de 90,9 %, respectivement. Cette étude est emblématique car, malgré les différents biais (étude monocentrique, sélection des images et des vidéos), elle s'intéressait à des lésions très hétérogènes en terme de présentation, à la fois en ce qui concernait les CEO et les lésions servant de contrôles. De plus, elle est l'une des premières à intégrer l'approche de chromoscopie combinée au zoom, ici pour l'analyse des IPCL. Elle démontre ainsi la capacité de l'IA à identifier facilement des motifs picturaux qui restent d'interprétation difficile, au moins pour les endoscopistes européens.

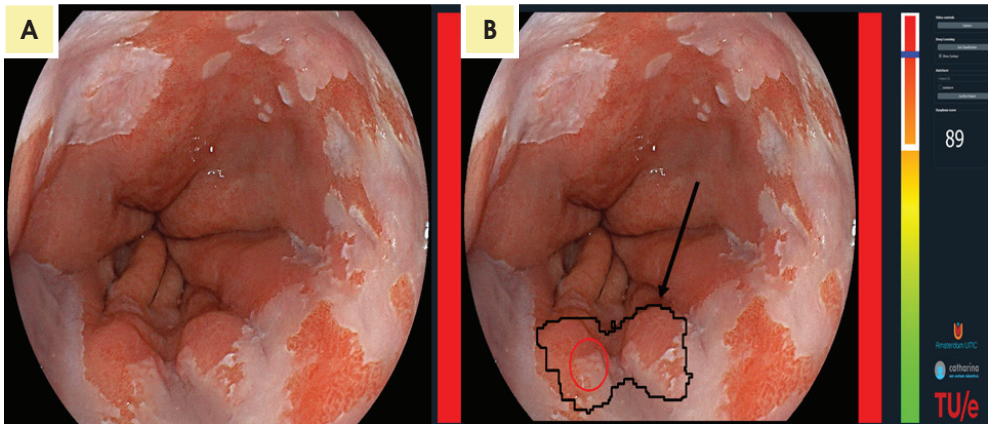


FIGURE 1. Endobrachyœsophage (EBO) examiné en lumière blanche, sans (A) puis avec apport d'intelligence artificielle (B). Le système d'IA montre la zone suspecte de dégénérescence (en noir) et l'endroit où cibler la biopsie de façon préférentielle (en rouge). La barre latérale estime l'indice de confiance du système d'IA dans la prédiction de néoplasie sur EBO (ici 89 %).

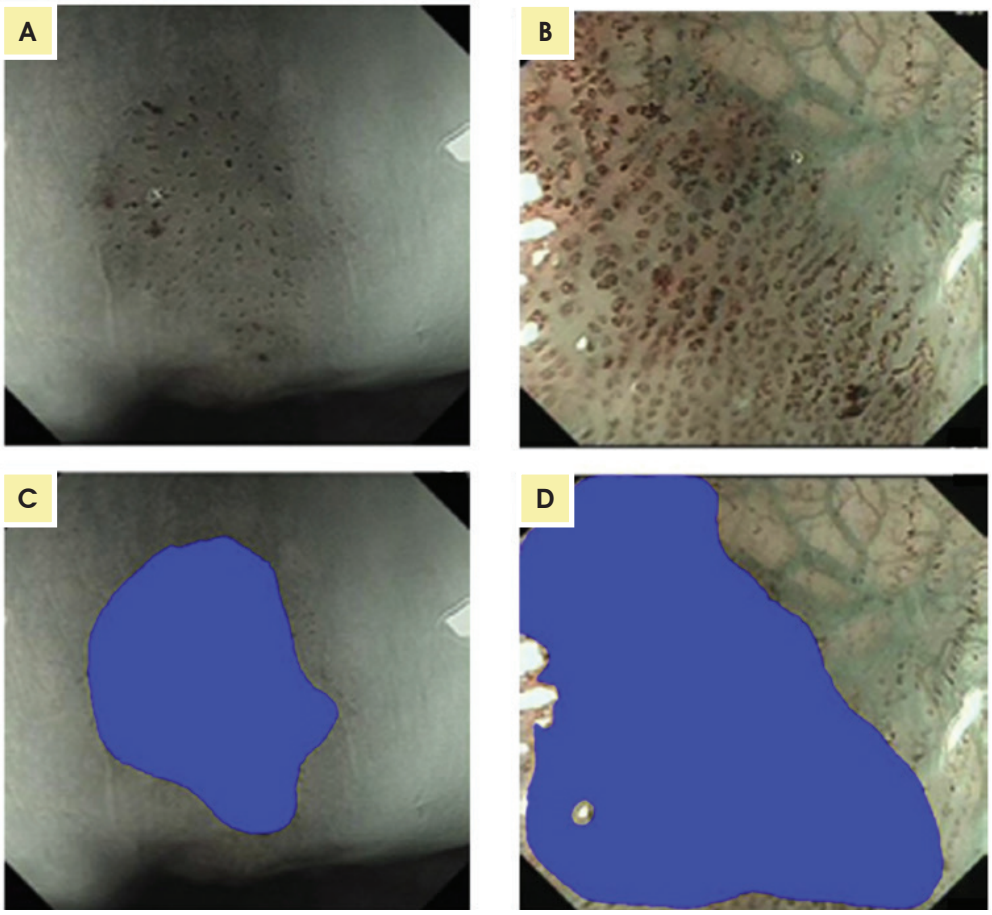


FIGURE 2. Carcinome épidermoïde œsophagien (CEO) examiné en lumière blanche, sans (A,B) puis avec apport d'intelligence artificielle (C,D). Le système d'IA montre la zone suspecte à biopsier ou réséquer (en bleu). L'évaluation en mode zoom (D) augmente encore les performances du système, et permettra probablement de prédire la profondeur d'infiltration de façon automatisée dans un futur proche.

Applications de l'IA en pathologie gastrique

Dans l'estomac, la majorité des études s'est pour l'instant concentrée sur la possibilité de détecter précocement des cancers en lumière blanche. Ainsi, l'étude chinoise multicentrique publiée par Luo *et al.* [12] s'est concentrée sur la mise au point d'un système appelé GRAIDS (*Gastrointestinal Artificial Intelligence Diagnostic System*). Dans cette étude, 1 036 496 images obtenues chez 84 424 patients ont été utilisées pour développer et tester GRAIDS. La phase initiale a consisté en un entrainement du système, puis une phase de vérification intrinsèque et de validation interne. La phase secondaire a permis de faire une première évaluation prospective dans un centre hospitalier national, puis une deuxième évaluation prospective dans cinq centres hospitaliers généraux, en comparant les performances de GRAIDS à celles d'endoscopies ayant trois niveaux d'expertise différents (expert, compétent, endoscopiste en formation). La performance diagnostique de GRAIDS était de 93 % dans le set « centre hospitalier national » et variait peu (92-98 %) dans les cinq centres hospitaliers généraux. Surtout, ce système d'IA avait une sensibilité équivalente à celle d'un expert (94 %) et supérieure aux endoscopistes compétents (86 %) et en formation (72 %). Surtout, plutôt que d'opposer l'IA et l'œil humain, le fait de combiner GRAIDS avec l'interprétation humaine permettait d'améliorer la sensibilité des endoscopistes pour la détection des cancers, de façon modeste (98 %) pour les experts et importante pour les endoscopistes compétents (98 %) et les endoscopistes en formation (96 %). Les faux positifs (mucus, anomalies de relief liées à des contractions gastriques) semblent faciles à identifier pour l'œil humain et donc n'induiraient que peu ou pas de biopsies excessives. Le système semble en outre assez rapide (analyse de 25 images par seconde en vidéo en temps réel) pour pouvoir être réellement utilisable en pratique clinique. Il est important de noter que ce système permettait également l'identification des cancers œsophagiens en lumière blanche, dont l'incidence en Europe est plus proche de l'incidence chinoise qu'en ce qui concerne les cancers gastriques, rendant donc séduisante cette approche dans une cohorte européenne.

Conclusion et perspectives

La littérature sur l'intelligence artificielle (IA) dans le tube digestif haut évolue rapidement et les données premières méta-analyses sont encourageantes, comme celle d'Arribas *et al.* regroupant 19 études avec 1116 patients (près de 24 000 images…) atteints de carcinome épidermoïde œsophagien (CEO), de cancer développé sur muqueuse de Barrett, ou de cancer gastrique [1]. En regroupant ces différents cancers, l'IA avait une sensibilité, spécificité, valeur prédictive positive et valeur prédictive négative de 90 %,

89 %, 87 % et 91 % pour la détection des lésions. Cela suggère un intérêt majeur pour la pratique clinique future, sachant qu'il s'agit essentiellement d'études pilotes et que les systèmes ne vont faire que s'améliorer avec le temps.

En pratique, la commercialisation de systèmes dédiés à la détection et bientôt la caractérisation des lésions du tube digestif haut est imminente. Il faudra cependant garder la tête froide et savoir que l'IA ne peut être qu'un compagnon de l'endoscopiste qui reste seul maître de la capacité à analyser minutieusement la muqueuse oesogastroduodénale, de faire des biopsies ciblées ou des résections, et qu'elle ne pourra jamais identifier des lésions atypiques ou des pièges diagnostiques comme le pourra l'expérience d'un bon endoscopiste. Il restera également à surmonter les obstacles éthiques, réglementaires et économiques à l'utilisation de l'IA en pratique de routine, et à continuer d'évaluer par des études méthodologiquement solides le réel apport de ces technologies.

RÉFÉRENCES

1. Arribas J, Antonelli G, Frazzoni L, Fuccio L, *et al.* Standalone performance of artificial intelligence for upper GI neoplasia: a meta-analysis. *Gut* 2020 : epub. doi : 10.1136/gutjnl-2020-321922.
2. Veitch AM, Uedo N, Yao K, East JE. Optimizing early upper gastrointestinal cancer detection at endoscopy. *Nat Rev Gastroenterol Hepatol* 2015 ; 12 : 660-7. doi : 10.1038/nrgastro.2015.128.
3. Inoue H, Kaga M, Ikeda H, *et al.* Magnification endoscopy in esophageal squamous cell carcinoma: a review of the intrapapil-lary capillary loop classification. *Ann Gastroenterol* 2015 ; 28 : 41-8.
4. Dekker E, Houwen BBSL, Puig I, *et al*. Curriculum for optical diagnosis training in Europe: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy* 2020 ; 52 : 899-923. doi : 10.1055/a-1231-5123
5. Chiu PWY, Uedo N, Singh R, et al. An Asian consensus on standards of diagnostic upper endoscopy for neoplasia. *Gut* 2019 ; 68 : 186-197. doi : 10.1136/gutjnl-2018-317111.

6. Hamashima C. Systematic Review Group and Guideline Development Group for Gastric Cancer Screening Guidelines. Update version of the Japanese Guidelines for Gastric Cancer Screening. *Jpn J Clin Oncol* 2018 ; 48 : 673-83. doi : 10.1093/jjco/hyy077

7. Jun JK, Choi KS, Lee HY, *et al*. Effectiveness of the Korean National Cancer Screening Program in Reducing Gastric Cancer Mortality. *Gastroenterology* 2017 ; 152 : 1319-328.e7. doi : 10.1053/j.gastro.2017.01.029.

8. de Groof AJ, Struyvenberg MR, van der Putten J, *et al*. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. *Gastroenterology* 2020 ; 158 : 915-29.e4. doi : 10.1053/j.gastro.2019.11.030

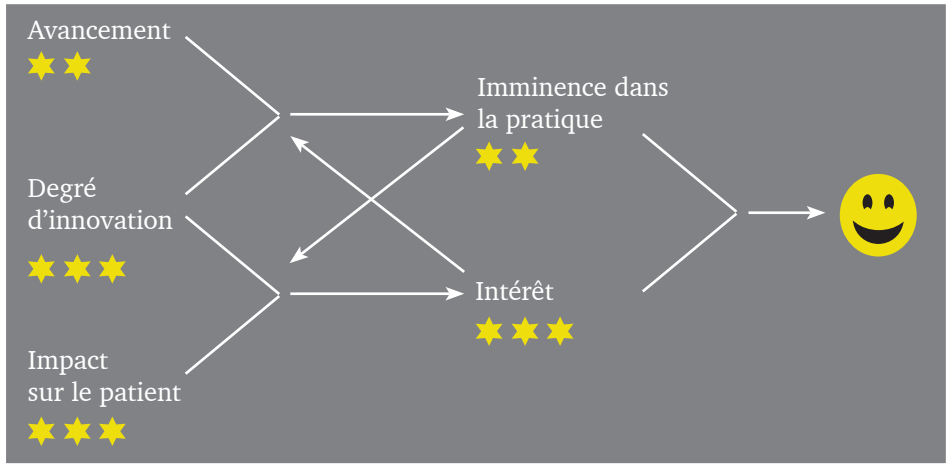
9. de Groof J, van der Sommen F, van der Putten J, *et al*. The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy. *United European Gastroenterol J* 2019 ; 7 : 538-47. doi : 10.1177/2050640619837443.

10. de Groof AJ, Struyvenberg MR, Fockens KN, *et al*. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest Endosc* 2020 ; 91 : 1242-50. doi : 10.1016/j.gie.2019.12.048.

11. Guo L, Xiao X, Wu C, *et al.* Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastroin-test Endosc* 2020 ; 91 : 41-51. doi : 10.1016/j.gie.2019.08.018

12. Luo H, Xu G, Li C, *et al*. Real-time artificial intelligence for detection of upper gastrointesti-nal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* 2019 ; 20 : 1645-54. doi : 10.1016/S1470-2045(19)30637-0

L'IA appliquée à la vidéocapsule endoscopique de l'intestin grêle : état de l'art et perspectives



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Xavier Dray [1,2], Aymeric Histace [2] et Romain Leenhardt [1,2]
[1] *Sorbonne Université, Centre d'Endoscopie Digestive, Hôpital Saint-Antoine, APHP, Paris*
[2] *ETIS UMR 8051 (CY Paris Cergy University, ENSEA, CNRS), Cergy*

Les indications de vidéocapsule endoscopique (VCE) de l'intestin grêle (IG) font l'objet de recommandations internationales pour l'exploration de saignements digestifs inexpliqués (SDI) et la suspicion de maladie de Crohn (MC). La VCE-IG trouve également des indications dans certains cas de polypose intestinale et de maladie coeliaque réfractaire. Les gastro-entérologues passent 30 à 120 minutes à examiner et interpréter des enregistrements de VCE-IG complets comprenant des dizaines de milliers d'images. Cette lecture est fastidieuse, souvent monotone et exigeante. Elle impose en effet au médecin des créneaux horaires dédiés et une pleine concentration. Récemment, Beg *et al.* ont montré que les performances diagnostiques d'un lecteur de VCE-IG diminuaient dès après le premier examen, alors même que des lecteurs expérimentés déclaraient une moyenne de 3,4 examens par session [1].

Il est devenu évident que des solutions d'intelligence artificielle (IA) vont avoir un impact majeur en médecine, notamment par leurs applications en imagerie et en endoscopie digestive, et particulièrement en VCE-IG compte tenu du caractère chronophage et fastidieux des lectures [2]. Ces solutions ont déjà pénétré le marché, avec à ce jour plusieurs solutions disponibles en coloscopie et en VCE-IG [2].

La revue présentée ici examine les données actuelles et les perspectives futures de l'IA en VCE-IG, ainsi que les obstacles potentiels à leur mise en œuvre complète.

Les spécificités de l'IA appliquée à la VCE

En endoscopie conventionnelle, l'opérateur manœuvre l'endoscope en fonction de ce qui est observé à l'écran. Dans ce cadre, synchrone de l'observation «humaine», les solutions d'IA fournissent une aide supplémentaire pour la détection et éventuellement pour la caractérisation des lésions. Le principal défi technologique de l'IA dans ce contexte est d'afficher les résultats avec précision et de manière synchrone, c'est-à-dire aussi rapidement que la fréquence de capture d'images (24-60 images par seconde).

L'IA appliquée à la VCE-IG est confrontée à un défi presque inverse. La lecture différée des images de VCE-IG par rapport à leur capture par le dispositif capsulaire est la règle, qu'elle soit réalisée par un humain ou par une solution d'IA. La contrainte de temps du traitement des images n'est donc pas le principal problème. En revanche, le transit de la capsule le long du tractus digestif reste hors du contrôle de l'opérateur ; dans un futur proche, lorsque l'IA sélectionnera un nombre limité d'images à analyser (disons 2 %) par l'endoscopiste qui lui fait confiance, les 98 % d'images restantes ne seront pas relues par l'opérateur. Au total, l'objectif principal de l'IA en VCE est actuellement d'atteindre une sensibilité très élevée pour la détection des anomalies tout en réduisant significativement le temps de lecture (en lien avec la spécificité, et donc directement avec le nombre d'images «faussement positives» que l'opérateur devra invalider). Cette quête d'une haute sensibilité repose essentiellement sur un apprentissage machine (*machine learning*, orienté par le choix d'experts humains), tandis que l'aide à la caractérisation repose sur une approche d'apprentissage profond (*deep learning*, sans intervention d'experts humains).

Détection des lésions et des anomalies

La détection des lésions et des anomalies fait l'objet du *tableau 1* et de la *figure 1*.

Sang

Avant l'apparition de l'IA, les évaluations d'un indicateur de présence de sang (*Suspected Blood Indicator* [SBI]) montraient qu'il était très sensible (96 %) mais peu spé-

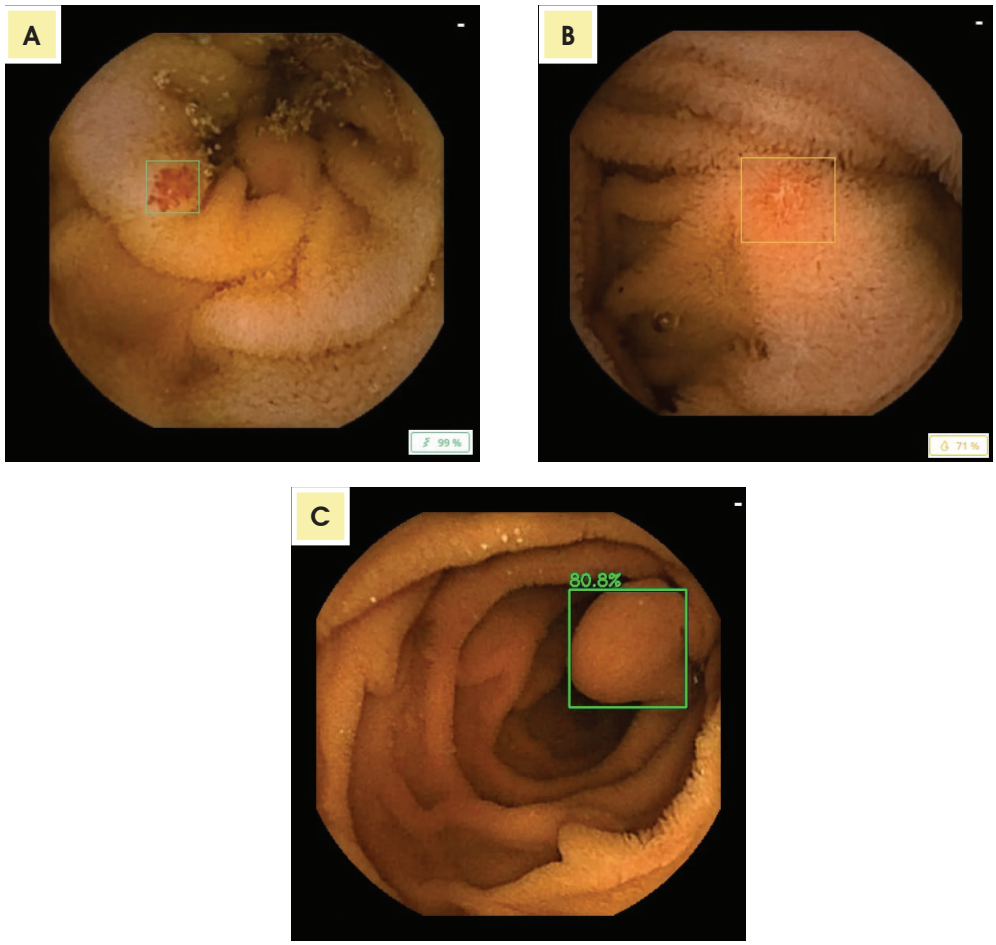


FIGURE 1. Exemples de lésions détectées par le système d'intelligence artificielle Axaro ® (Augmented Endoscopy ®, France) dédié à la lecture de vidéocapsule endoscopique. Le type de lésion détecté est symbolisé par un pictogramme (vasculaire, inflammatoire, tumoral...) et le degré de confiance (en pourcentage) du diagnostic apparaît en vignette. A) Angiectasie ; B) érosion aphtoïde ; C) tumeur sous-muqueuse.

Tableau 1. Principales études récentes sur l'intelligence artificielle pour la détection et/ou la caractérisation des lésions et des anomalies en endoscopie capsulaire.

1er Auteur	Cible	Nombre de patients / images	Design
Xing X, 2018	Sang	30 patients 1000 images	Rétrospective
Aoki T, 2020	Sang	41 patients 49 191 images	Rétrospective
Noya F, 2017	Angiectasies	36 patients 1 648 images	Rétrospective
Leenhardt R,2019	Angiectasies	408 patients 1 200 images	Rétrospective
Tsuboi A,2020	Angiectasies	48 patients 10 488 images	Rétrospective
Fan S, 2018	Ulcérations	144 patients 21 160 images	Rétrospective
Aoki T, 2019	Ulcérations	180 patients 15800 images 16 patients	Rétrospective
Wang S, 2019	Ulcérations	1504 patients 47 202 images	Rétrospective
Klang E, 2020	Ulcérations	49 patients 17 640 images	Rétrospective
Saito H, 2020	Lésions saillantes	385 patients 48 091 images	Rétrospective
Iakovidis D, 2014	Multiclasse	251 patients 1 370 images	Rétrospective
Ding Z, 2019	Multiclasse	6,970 patients 113 millions d'images	Rétrospective

* Temps de lecture moyen.

Entrainement / validation	Images/vidéo	Performances
Oui	Images fixes	Sensibilité : 98,5% Spécificité : 99,5 %
Oui	Images fixes	Sensibilité: 96,6% Spécificité : 99,9 %
Oui	Images fixes	Sensibilité : 89,5 % Spécificité : 96,8 %
Oui	Images fixes	Sensibilité : 100 % Spécificité : 96 %
Oui	Images fixes	Sensibilité : 98,8 % Spécificité : 98,4 %
Oui	Images fixes	Sensibilité : 96,8 % Spécificité : 94,8 %
Oui Non	Images fixes Vidéos	Sensibilité : 88,2 % Spécificité : 90,9 % Lecture 3 minutes *
Oui	Images fixes	Sensibilité: 89,7 % Spécificité: 90,5 %
Oui	Images fixes	Sensibilité : 96,8 % Spécificité : 96,6 %
Oui	Images fixes	Sensibilité : 90,7 % Spécificité : 79,8 %
Oui	Images fixes	Sensibilité : 94,0 % Spécificité : 95,4 %
Oui	Vidéos	Sensibilité: 99,9 % Spécificité : 97,0 % Lecture 6 minutes *

cifique (17 % à 65 %) [3]. Depuis lors, Xing *et al.* ont proposé une méthode d'apprentissage profond (Deep Learning, DL) fortement supervisé avec une sensibilité, une spécificité et une précision de plus de 98 % [4]. Plus récemment, Aoki *et al.* ont développé un algorithme, moins supervisé, basé sur les réseaux neuronaux (RN), à partir d'images fixes de VCE-IG avec une sensibilité de 96,6 %, une spécificité de 99,9 % et une précision de 99,9 % [5]. Bien que ces résultats semblent prometteurs, ils ont tous été obtenus à l'échelle d'images fixes, et on ne dispose pas encore d'évaluation à l'échelle vidéo, ni en terme de détection, ni en termes de caractérisation (sang frais, caillots ou méléna, traces ou hémorragie massive).

Angiectasies

Les angiectasies - ou angiodysplasies - sont les lésions les plus courantes vues en VCE-IG. Dans une étude de preuve de concept d'IA basée sur les RN, notre équipe a démontré une sensibilité de 100 % et une spécificité de 96 % dans leur détection, à partir d'images fixes [6]. En 2020, Tsuboi *et al.* ont élaboré un système IA sur une base de données de 189 vidéos VCE-IG, avec sensibilité de 98,8 % et une spécificité de 98,4 %. Ces résultats prometteurs exigent de vastes essais cliniques prospectifs [7].

Érosions, ulcérations et ulcères

La diversité de ces lésions en termes de profondeur, de taille, de formes et d'étiologie constitue un défi pour la recherche sur l'IA en VCE-IG. Les solutions d'IA basées sur les RN ont abouti à d'excellentes performances diagnostiques après apprentissage et évaluation sur images fixes [8]. Dans une série pilote à l'échelle vidéo, comprenant 16 VCE-IG avec 37 images de pertes de substance ainsi que 4 vidéos normales, les temps de lecture des experts avec/sans IA ont diminué de 12 à 3 minutes, avec des taux de détection inchangés (87 % et 84 %, respectivement) [9]. Dans l'ensemble, les solutions de détection des lésions ulcérées doivent encore être affinées et mieux évaluées à l'échelle des vidéos.

Lésions saillantes

L'équipe de Saito *et al.* a collecté 30 584 images de polypes/masses/tumeurs parmi 292 patients, pour développer une solution d'IA basée sur les RN. Le système, ensuite testé sur un lot indépendant d'images, a montré une sensibilité de 90,7 % et une spécificité de 79,8 % [10]. Bien que la preuve de concept soit maintenant proche, il n'existe aucune étude disponible à l'échelle vidéo sur la question critique de la détection automatisée des lésions saillantes. Par ailleurs, il sera difficile de caractériser des lésions « saillantes » variantes de la normale (plis, phlébectasies, lymphangiectasies, kystes chyleux, hyperplasies nodulaires lymphoïdes, glandes de Brünner...) de

diverses affections pertinentes, bénignes ou malignes, qu'elles soient épithéliales ou sous-muqueuses.

Détection multiclasse

En 2014, Iakovidis *et al.* ont proposé une méthodologie capable de détecter plusieurs types de lésions différentes notamment en classant des images anormales, non plus en basant l'apprentissage sur la reconnaissance de différents types de lésions… mais sur le simple fait de les distinguer d'images … normales. En d'autres termes, « ce qui n'est pas normal est, par définition, anormal » (au risque de ne pas être ensuite capable de correctement catégoriser les anomalies) (*figure 2*) [11]. Quelques années plus tard, Ding *et al.* réussissaient, en entraînant un tel modèle de DL sur plus de 100 millions d'images de VCE-IG collectées dans 77 centres chinois, à rejoindre les capacités de détection de lésions et de variants de la normales (lymphagiectasies, points rouges,…) d'un lecteur gastro-entérologue [12]. Le taux de faux positifs était seulement 3 %. Le fait le plus notable était est que le temps moyen de lecture par vidéo était de 97 min par lecture conventionnelle (un chiffre pour le moins surprenant…) et de 6 min avec l'aide de l'IA [12].

Caractérisation

Seules quelques contributions ont été rapportées dans la littérature en termes de caractérisation des lésions vues en VCE-IG par des solutions d'IA. Notre équipe a notamment montré qu'après la détection d'angiectasies, leur segmentation (détermination des limites anatomiques de la lésion) peut être réalisée par l'IA à l'échelle du pixel, ouvrant donc la voie à des extractions de caractéristiques telles que la taille, le type, la profondeur, des lésions détectées [6]. La principale limite de cette approche est qu'elle repose sur des bases de données annotées de très haute qualité, et des calculs lourds et longs.

Détection et caractérisation multiclasses, combinées

En combinant une approche «faiblement supervisée» pour la détection avec différents sous-réseaux neuronaux pour la classification, Otani *et al.* ont récemment fait état d'une solution avec des performances de détection de 0,928 pour les ulcérations, 0,884 pour les anomalies vasculaires et 0,902 pour les lésions saillantes, sur la base d'images fixes [13].

Évaluation de la qualité de la préparation intestinale

Bien qu'elle soit recommandée, l'évaluation par des gastroentérologues de la qualité de la préparation intestinale en VCE reste peu reproductible [14]. Deux algorithmes basés sur l'IA ont montré que l'IA peut conduire à une métrique de propreté robuste et fiable au niveau de l'image [14,15]. Une contribution récente de notre équipe a fourni

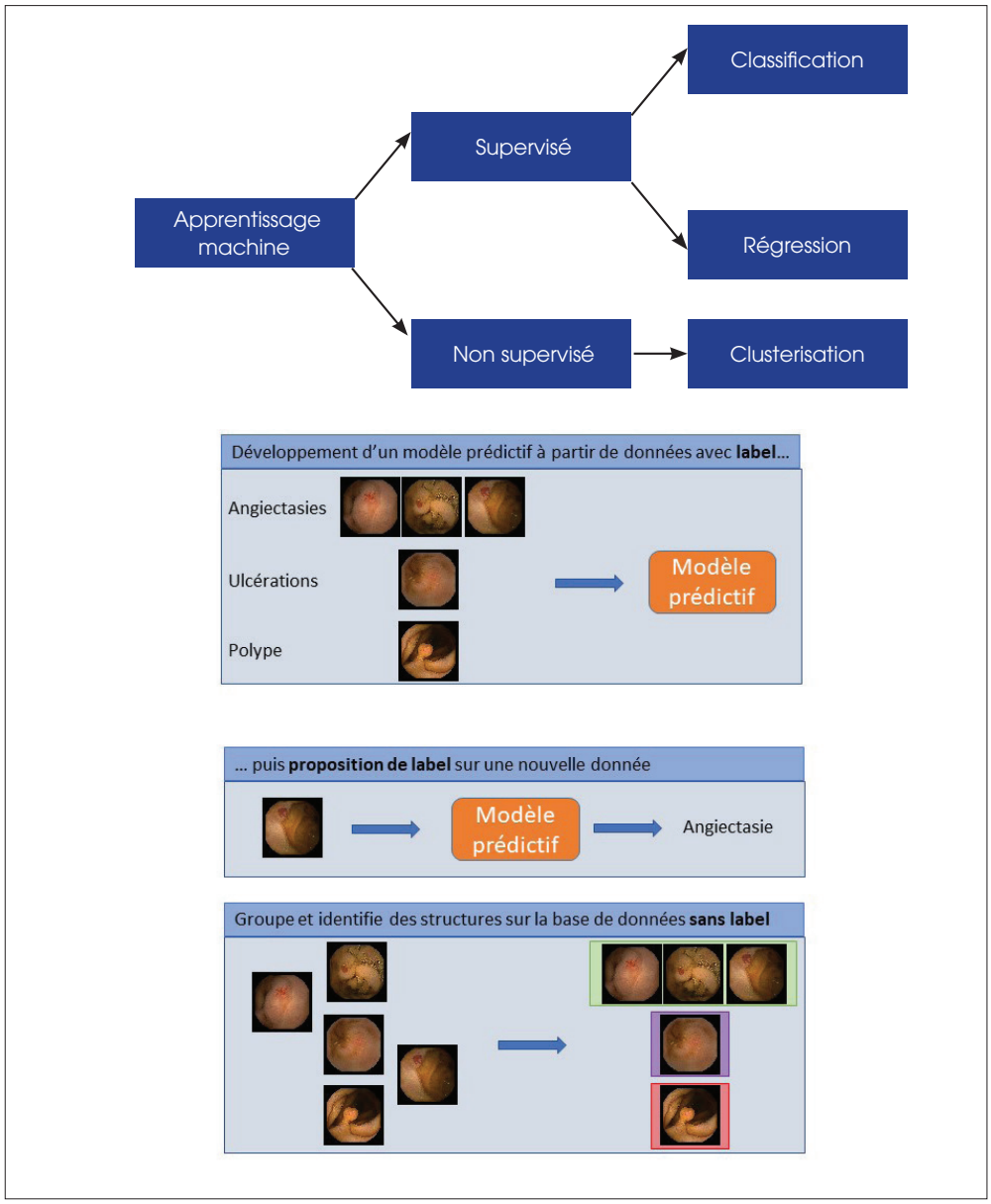


FIGURE 2. Principes de l'apprentissage machine supervisé et non supervisé.

la preuve de concept qu'une approche de DL a une sensibilité de l'ordre de 90% à l'échelle vidéo, en comparaison d'un consensus d'experts, pour reconnaître un examen mal préparé (*figure 3* et *tableau 2*) [15].

Industrialisation

Le premier pas en avant a été fait en 2019 par la firme Ankon ® (Chine), après que la grande étude multicentrique, pré-citée, réalisée avec leur dispositif ait confirmé qu'un algorithme basé sur les RN était capable de détecter des anomalies en de VCE-IG avec des performances équivalentes à celles de gastroentérologues, et de réduire considérablement les temps de lecture (à environ 6 minutes) [12]. La société Jinshan (qui commercialise la capsule Omom ®) a également mis sa solution sur le marché en Asie.

Limites actuelles

Il convient de reconnaître plusieurs limites importantes des études actuelles sur l'IA en endoscopie par VCE-IG. Tout d'abord, ces études sont rétrospectives. De plus, la plupart des ensembles de données pour la formation à l'IA comprennent des images fixes sélectionnées, avec un risque inhérent de surapprentissage. Par ailleurs, les études de validation externes sont rares (voire inexistantes), peut-être parce que les équipes de recherche utilisent des dispositifs, des architectures de réseaux et des images (avec des contrastes, des résolutions et des labels) différents. Enfin, certains aspects sont à peine abordés par les chercheurs en IA, comme la reconnaissance de repères anatomiques (pylore, valvule iléocæcale) ou l'évaluation de la taille des lésions.

Besoins prioritaires

À un niveau élevé de leur supervision, des RN devraient permettre à la technologie actuelle de relever le défi de la classification des lésions en termes de diagnostic d'abord, et d'imputabilité ensuite. Toutefois, ce développement achoppe sur le fait que la concordance diagnostique entre lecteurs de VCE-IG est médiocre, de l'ordre de 60 % [16], probablement du fait d'importantes variations entre lecteurs dans le vocabulaire utilisé, dans la caractérisation des lésions et dans l'interprétation qui en est faite selon le contexte clinique. En l'absence d'une interprétation consensuelle des résultats de VCE-IG dotée d'une reproductibilité élevée, tout apprentissage par des machines est illusoire. Un groupe européen a relevé ce défi en proposant un consensus d'experts sur la nomenclature, la description et l'imputabilité des lésions vasculaires, ulcéreuses et saillantes, selon les principales indications de la VCE-IG [17], pour servir le développement de bases de données solides et l'établissement d'une « vérité terrain » (*ground truth*) indispensable à l'apprentissage machine.

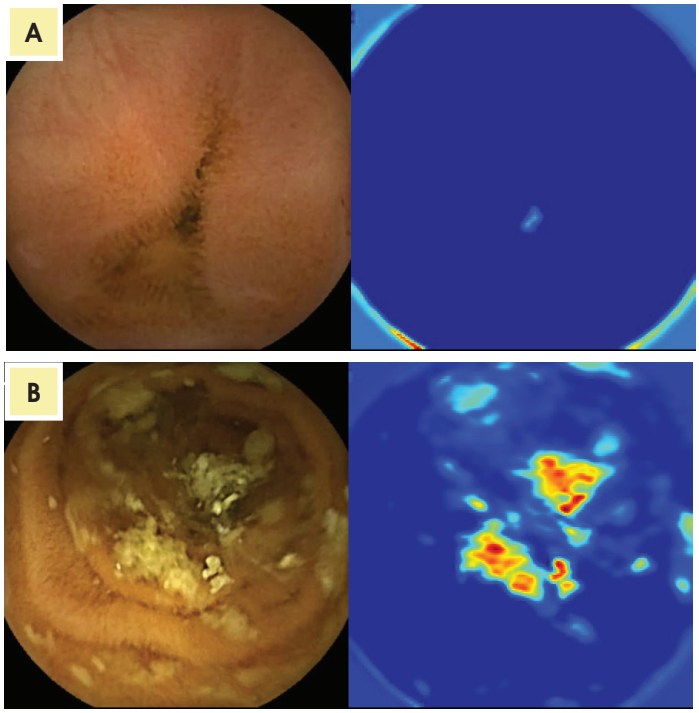


FIGURE 3. Représentation de l'évaluation d'images de vidéocapsules endoscopiques en termes de qualité de visualisation par le système d'intelligence artificielle Axaro ® (Augmented Endoscopy, France) : (a) adéquate (b) inadéquate. L'image native est à gauche, la "carte chaude" (heatmap) à droite où les zones bien visibles ont des couleurs plutôt "froides" (tonalités bleues) et les zones mal visibles des couleurs plutôt chaudes (tonalités rouges). Le résultat n'est cependant pas rendu sous forme de heatmaps (qui ont pour seul intérêt ici d'illustrer le propos) mais il est rendu en proportion d'images sales au sein de l'ensemble de la séquence vidéo de l'intestin grêle et pour chaque quartile de cette séquence. Un pourcentage inférieur à 21% semble lié à une préparation inadéquate à l'échelle vidéo.[15]

Une fois ces principes acquis, des bases de données de meilleure qualité pourront voir le jour avec la perspective d'y intégrer un mécanisme d'apprentissage actif permettant non seulement d'enrichir prospectivement ces banques mais également de signaler les erreurs de l'IA.

Freins potentiels

On peut citer brièvement quelques possibles freins au développement des solutions détaillées dans ce texte, en gardant à l'esprit les aspects éthiques, économiques et réglementaires de l'IA appliquée à l'endoscopie. Tout d'abord, il existe une large part d'incompréhension concernant les critères que l'IA retient pour accomplir telle ou telle tâche technique (détection, caractérisation...). Cette part d'inconnue ne facilitera pas l'adoption de l'IA en médecine en général, en endoscopie en particulier, et en VCE spécifiquement. La responsabilité de la lecture des VCE-IG continuera d'incomber aux médecins, qui pourtant n'auront vu qu'une faible proportion des images captées. Ces images seront archivées et donc réexaminables par des experts. D'ici à ce que les autorités de santé posent les principes de la régulation des solutions d'IA en médecine (qui ne sont ni des médicaments, ni des dispositifs), un certain flou exposera peut être le médecin et son institution du point de vue du droit, et par leur intermédiaire, ce flou risque de freiner l'adoption même (et donc le développement et les améliorations) de la technique. Et pourtant, ces améliorations ont une portée d'autant plus considérable qu'avec une démographie médicale chancelante, et particulièrement dans notre spécialité, le temps médical dédié à des tâches plus nobles a une valeur inestimable. Dans la même veine, on peut s'interroger sur le degré de confiance que les patients (et la société) accorderont (et donc autoriseront) à un tel système (sans filet de sécurité de relecture…), surtout dans ses premières années (et avec ses premières erreurs). Notre groupe a démarré des audits auprès de médecins et de patients sur ces questions. Elles ouvriront des réflexions éthiques qui sont incontournables. Enfin, la transposabilité des développements majeurs réalisés en Asie n'est pas forcément immédiate en Occident. Elle se double d'enjeux complexes de souveraineté numérique et économique.

Perspectives

On peut citer dans les perspectives de développement de l'IA en VCE-IG, l'application à la VCE de l'estomac, du côlon, et aux systèmes dits "panentériques" (capsules gastro-grêliques ou grêlocoliques). Si les lectures deviennent moins fastidieuses et plus automatisées, il est vraisemblable que des dispositifs avec de multiples têtes de lectures et/ou panoramiques deviennent la norme, non plus seulement pour les segments les plus larges du tube digestif (estomac, côlon) mais aussi pour l'intestin grêle. Enfin, avec

Tableau 2. Principales études récentes sur l'intelligence artificielle pour l'évaluation de la propreté de l'intestin grêle pendant l'endoscopie en capsule

1er Auteur	Vérité terrain / Critère de jugement	Dispositif	Entrainement	Test etvalidation	Performance sur la banque de validation
Noorda R, 2020	Échelle non validée, en 2 à 4 points, évaluée par 2 experts indépendants (sans adjudication)	Pillcam® SB3, Medtronic, USA	563 images (augmentées à 55293) à partir de 35 vidéos	854 images à partir de 30 vidéos	Images fixes
Leenhardt R, 2020	Échelle validée, en 2 à 10 points, évaluée par 3 experts indépendants (avec adjudication)	Pillcam® SB3, Medtronic, USA	600 images (augmentées à 3000) à partir de 30 vidéos	156 vidéos	Images fixes

ce développement, il est attendu que le marché des dispositifs capsulaires augmente, avec plus d'examens prescrits et interprétés. La réalisation d'examens à domicile a débuté, on ne peut exclure que demain nos patients téléchargent eux-mêmes les vidéos issues de leur tube digestif et les fassent interpréter en ligne et en direct par l'intermédiaire de plateformes sécurisées…

Conclusion

Vingt ans après son déploiement révolutionnaire en endoscopie digestive, la VCE est sur le point de vivre son âge d'or grâce au développement de l'IA. Des solutions de détection automatisées sont désormais disponibles dans le commerce. Des essais prospectifs contrôlés par des équipes indépendantes sont encore nécessaires avant que les médecins puissent se fier entièrement à ces solutions et réduire considérablement leur temps de lecture tout en étant assuré de leurs excellentes capacités de détection. Lorsque cette première étape critique sera franchie, la caractérisation automatisée (en termes de diagnostic et de pertinence) sera la deuxième. Le chemin est cependant beaucoup plus long, car les enjeux réglementaires et éthiques sont importants, et qu'en parallèle, la technologie de la VCE suit sa propre voie et ouvre de nouveaux champs d'application - examens panentériques, chromoendoscopie virtuelle, locomotion active, options thérapeutiques - et autant de sujets dans lesquels l'IA peut également jouer un rôle.

RÉFÉRENCES

1. Beg S, Card T, Sidhu R, Wronska E, Ragunath K. How many capsule endoscopy cases can be read before accuracy is affected? *Gut* 2018 ; 67 (S1) : A1–30.
2. Dray X, Iakovidis D, Houdeville C, Jover R, Diamantis D, Histace A, Koulaouzidis A. Artificial intelligence in small bowel capsule endoscopy: Current status, challenges, and future promise. *Journal of Gastroenterology and Hepatology* : 2021-01-15 , doi : 10.1111/jgh.15341
3. Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: a systematic review and meta-analysis. Expert Rev Gastroenterol. *Hepatol* 2017 ; 11 : 43–51
4. Xing X, Jia X, Meng MQ. Bleeding Detection in Wireless Capsule Endoscopy Image Video Using Superpixel-Color Histogram and a Subspace KNN Classifier. *Annu Int Conf IEEE Eng Med Biol Soc* 2018 : 1-4.
5. Aoki T, Yamada A, Kato Y, et al. Automatic detection of blood content in capsule endoscopy images based on a deep convolutional neural network. *J Gastroenterol Hepatol* 2020 ; 35 : 1196-200
6. Leenhardt R, Vasseur P, Li C, et al. A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019 ; 89 : 189-94
7. Tsuboi A, Oka S, Aoyama K et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. *Dig Endosc* 2020 ; 32 : 382-90.
8. Klang E, Barash Y, Margalit RY, et al. Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020 ; 91 :606-13
9. Aoki T, Yamada A, Aoyama K, et al. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. *Dig Endosc* 2020;32 : 585-91
10. Saito H, Aoki T, Aoyama K, et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2020;92 : 144-51

11. Iakovidis DK, Koulaouzidis A Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software. *Gastrointest Endosc* 2014 ; 80 : 877-8

12. Ding Z, Shi H, Zhang H, Meng L, Fan M, Han C, et al. Gastroenterologist-Level Identification of Small-Bowel Diseases and Normal Variants by Capsule Endoscopy Using a Deep-Learning Model. *Gastroenterology* 2019 ; 157 : 1044-1054

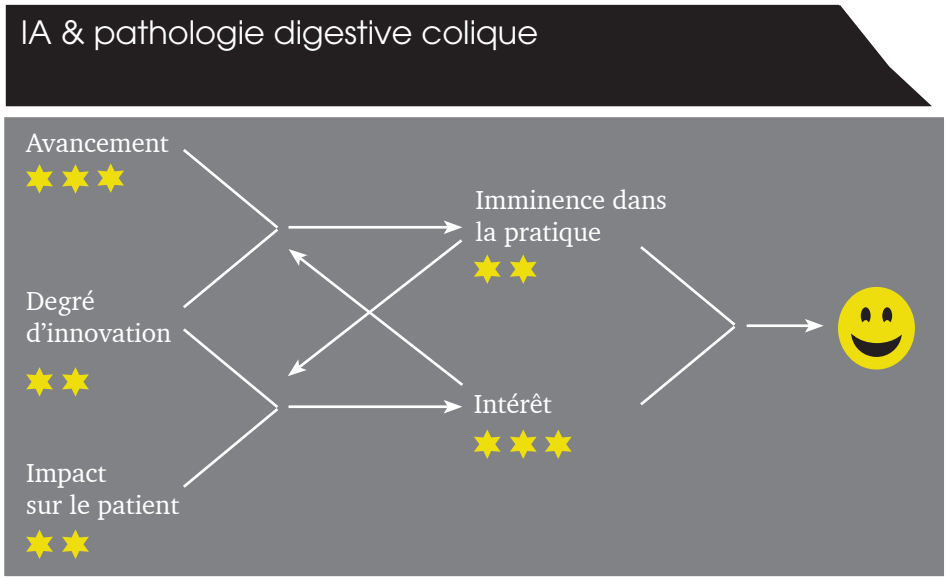
13. Otani K, Nakada A, Kurose Y et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network. *Endoscopy* 2020 ; 52 : 786-91. doi : 10.1055/a-1167-8157

14. Noorda R, Nevárez A, Colomer A, Beltrán VP, Naranjo V. Automatic evaluation of degree of cleanliness in capsule endoscopy based on a novel CNN architecture. *Sci Rep* 2020 ; 10 : 17706

15. Leenhardt R, Souchaud M, Houist G, et al. A Neural Network-based Algorithm for Assessing the Cleanliness of Small Bowel during Capsule Endoscopy. *Endoscopy* 2020 ; epub. doi: 10.1055/a-1301-3841.

16. Jang BI, Lee SH, Moon JS, Cheung DY, Lee IS, Kim JO, et al. Inter-observer agreement on the interpretation of capsule endoscopy findings based on capsule endoscopy structured terminology: a multicenter study by the Korean Gut Image Study Group. *Scand J Gastroenterol* 2010;45(3):370-4

17. Leenhardt R, Koulaouzidis A, McNamara D. A guide for assessing the clinical relevance of findings in small bowel capsule endoscopy: Analysis of 8064 answers of international experts to an illustrated script questionnaire. *Clin Research Hepatol Gastroenterol* (sous presse).



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Aymeric Becq
[1] *Centre d'endoscopie digestive, Hôpital Saint-Antoine, APHP, Paris*
[2] *Institut des systèmes intelligents et de robotique, Paris Jussieu*

La coloscopie est l'examen de référence pour le dépistage de polypes coliques précancéreux [1]. Cet examen de dépistage est efficace mais imparfait. En effet le taux de détection d'adénomes (TDA) est très variable (7-53 %) et le taux global de polypes non vus est de 22 % [2,3]. Cela pose le problème du cancer d'intervalle et de sa mortalité, directement lié au TDA [3]. L'enjeu face à ces limitations est de diminuer la variabilité inter-opérateurs et d'optimiser la prévention du cancer colorectal par coloscopie. L'intelligence artificielle (IA) présente dans ce contexte une possible solution. En effet, il y a deux points majeurs sur lesquels l'IA peut être utile ici :

- Aide pour la détection de polypes (*Computer-Aided Detection* [CADe]) Objectif : augmentation du TDA
- Aide pour la caractérisation des polypes (*Computer-Aided Diagnosis* [CADx]) Objectif : diminuer la résection inutile de polypes hyperplasiques (PH)

Les recommandations européennes de l'ESGE publiées en 2019 sont bien sûr prudentes mais mentionnent l'IA pour la détection et la caractérisation comme outil potentiel si de hautes précision et reproductibilité peuvent être démontrées dans des essais multicentriques de bonne qualité. Il est également fait mention des difficultés et risques que représente l'implémentation à large échelle de l'IA en coloscopie : baisse du niveau de l'endoscopiste, excès de confiance en l'IA, banques de données d'entrainement des algorithmes d'IA qui ne sont pas représentatives de la vérité du terrain, piratage [4]…

Détection de polypes

En ce qui concerne la détection de polypes dans le colon, ce qu'on demande à l'IA c'est d'être sensible, avec peu de faux positifs, et de fonctionner en temps réel. Afin de limiter au maximum le risque de ne pas détecter un polype, c'est-à-dire d'avoir une sensibilité la plus élevée possible, le seuil de détection par l'algorithme est diminué, quitte à perdre en spécificité avec plus de faux positifs. Cela est souhaitable si le gain en sensibilité est significatif sans augmenter trop le nombre de faux positifs.

L'utilisation de l'apprentissage profond (*deep learning*) tel que les réseaux neuronaux (*Convolutional Neural Networks* [CNN]) a été permis par l'augmentation de la puissance des processeurs permettant l'analyse des très nombreuses données nécessaires à l'élaboration des algorithmes d'IA. Ceux-ci sont entrainés à l'identification des caractéristiques de polypes vs non-polypes sur de larges bases de données, sans directives de la part des chercheurs.

Au cours des dernières années, plusieurs algorithmes d'IA ont vu le jour, permettant leur évaluation sur images, puis vidéos, puis *in vivo* en temps réel. Parmi les systèmes existants on peut signaler le GI-Genius (Medtronic), Wision AI (Shanghai Wision AI Co., Ltd.) et CAD-Eye (Fujifilm). Le principe d'utilisation se veut pragmatique et simple. Sur le modèle de « *plug and play* », l'appareil est branché et fonctionne directement une fois allumé. Deux signaux préviennent alors l'endoscopiste qu'un polype a été identifié : un signal visuel à l'écran, et un signal (facultatif) sonore. Lors d'une coloscopie, 25 à 60 images/secondes défilent à l'écran. Ainsi, afin d'être performante en temps réel, l'IA doit reconnaitre et afficher le signal en 1/25-1/60 de seconde (40 ms), ce qui est très rapide et nécessite des processeurs puissants.

Le *tableau 1* résume les études principales publiées à ce sujet aujourd'hui. La première étude *in vivo,* publiée en 2019, prospective, incluant 55 coloscopies tout venant montrait un TDA de 29,1 % avec l'IA contre 30,9% pour l'endoscopiste. Les TDA étaient donc comparables, signifiant que l'algorithme d'IA était perfectible, mais cette étude

Tableau 1. Principales études d'évaluation de l'IA pour la détection des polypes coliques.

Étude	Type d'étude	Nombre d'images de polypes	Mode d'image	Vidéo en temps réel	Évaluation des polypes diminutifs	Résultats
Urban et al. (2018)	RS	4088	LB	+	+	
Figueiredo et al (2018)	RS	1680	LB	-	-	Sensibilité de 99,7% Spécificité de 84.9% Précision de 84,9%
Wang et al (2018)	ECR	6749	LB	+	+	9% d'augmentation du TDA
Misawa et al (2019)	RS	87	LB	-	+	Sensibilité de 90 % Spécificité de 63,3% Précision de 76.5%
Matsui et al (2019)	MA	606	LB	+	+	Sensibilité de 92.3 % Spécificité de 89,8% Précision de 96 %
Lui et al (2019)	RS	12 654	NBI	+	+	Augmentation du TDA 39,2% vs 23,9%
Repici et al (2020)	ECR	X	LB	+	+	Augmentation du TDA 54,8% vs 40,4%

RS: rétrospectif, ECR : étude contrôlée randomisée, MA : méta-analyse, LB : lumière blanche, NBI : Narrow Band Imaging.

prouvait que la détection en temps réel était faisable [5]. La première étude prospective randomisée, par Wang *et al,* publiée en 2019, comparait les performances d'un système d'IA (Wision-AI) chez 1 130 patients avec indication à une coloscopie. Le TDA était supérieur avec l'IA (29 % vs 20 % ; p < 0,001). Ce taux était notamment supérieur pour les polypes diminutifs (70,6% vs 63,5 ; p < 0,001). Les limites de l'étude étaient le manque de généralisabilité (coloscopies tout venant, âge moyen de 50 ans seulement, étude réalisée dans un pays (Chine) où la prévalence de cancer colorectal est moins élevée). Par ailleurs, les polypes les plus détectés par l'IA étaient diminutifs et/ou de nature hyperplasique et qui interroge sur la plus-value de l'IA. La première étude réalisée avec un système d'IA ayant un marquage CE date de 2020. Il s'agissait d'une étude contrôlée randomisée, multicentrique, incluant 685 patients avec coloscopie de dépistage. Le TDA était de 54,8 % avec l'IA contre 40,4 % pour le groupe contrôle (RR 1,3 (IC 1,14-1,45)). L'IA permettait la détection de plus d'adénomes par coloscopie, et de plus d'adénomes ≤ 5 mm et entre 6 et 9 mm [6]. Enfin, une méta-analyse publiée en 2020 regroupait 6 études sur la détection par IA et démontrait une sensibilité globale de 95,0 % (IC 95 % : 91,0 %-97,0 %), et une spécificité de 88,0 % (IC 95 % : 58,0 %-99,0 %) [7].

Au total, en 2021, on peut considérer que le problème de la détection de polypes par l'IA est résolu. Les prochaines étapes sont, d'une part, la validation clinique avec la réalisation d'autres essais randomisés sur l'impact de l'IA sur le TDA, et d'autre part l'obtention de l'autorisation de mise sur le marché pour les différents systèmes.

Caractérisation des polypes

La caractérisation des polypes est un second objectif de l'IA, bien différent du premier, et qui ne cible pas les mêmes endoscopistes. En effets, les endoscopistes les plus expérimentés y verront un plus grand intérêt que celui de la détection. L'intérêt sera plus grand également selon le pays voir continent, les Japonais étant par exemple bien plus intéressés par l'aspect caractérisation.

L'ASGE (*American Society for Gastrointestinal Endoscopy)* Technology Committee a édicté puis validé des seuils de performance estimés acceptables pour l'utilisation de nouvelles technologies afin d'estimer en temps réel de l'histologie de polypes diminutifs lors de coloscopies : les critères PIVI (*Preservation and Incorporation of Valuable endoscopic Innovations*). Afin qu'une technologie puisse être utilisée pour guider la décision de ne pas réséquer des polypes suspects d'être hyperplasiques du recto-sigmoïde ≤ 5 mm, la valeur prédictive négative pour l'histologie type adénome devrait

être de 90 % ou plus [8]. Cette valeur seuil fait ainsi office de baromètre en ce qui concerne l'évaluation de l'IA pour la prédiction de l'histologie des polypes coliques.

Il existe deux sortes de systèmes. Soit, un système d'analyse d'image de coloscopie classique, soit un système d'analyse d'image zoomée (endocytoscopie). Parmi les systèmes existants on peut signaler le AI4GI, le CAD-Eye (Fujifilm) et le EndoBRAIN ® (endocytoscopie). Le *tableau 2* résume les études principales publiées à ce sujet aujourd'hui.

En ce qui concerne l'image de coloscopie, l'étude publiée en 2019 par Bryne *et al.* fait office de référence. Il s'agissait d'un système de deep learning réalisé à partir de plus de 60 000 images sélectionnées issues de 223 vidéos, en *Narrow Band Imaging* (NBI), de muqueuse normale et de polypes coliques en vision proche et très proche (technologie near focus, Olympus, Tokyo, Japon). Le système proposé affichait un degré de « crédibilité » de diagnostic de muqueuse normale, de polype hyperplasique (NICE 1) ou d'adénome (NICE 2). Sur un jeu indépendant de 125 séquences vidéo de polypes de 10 à 20 secondes, les performances du système étaient comparées à l'analyse anatomopathologique, et dépassaient les critères PIVI avec une sensibilité de 98%, une spécificité de 83 %, et une précision diagnostique de 94 % [9].

L'endocytoscopie permet l'analyse *in vivo* de la morphologie des cellules de la couche superficielle du tissu étudié et de leurs noyaux, *via* une ultra-magnification (x520), après coloration par du bleu de méthylène (1 %) et du crystal violet (0,05 %). La combinaison avec une coloration par NBI permet de plus l'observation des microvaisseaux. Un système appelé EndoBRAIN ® a été évalué dans une étude pilote publiée en 2015 sur images de polypes < 10mm, montrant une sensibilité de 92.0%, une spécificité de 79,5 % et une précision de 89,2 % pour la différenciation entre ceux qui étaient néoplasiques et non néoplasiques. Les performances, parfaitement reproductibles et avec un temps d'analyse de 0,3 seconde, étaient similaires à celles des experts et significativement meilleure que celles des endoscopistes en formation [10]. Ce système, fondé sur du « deep learning » sur près de 70 000 images a ensuite fait l'objet d'une étude multicentrique rétrospective de validation pour mise sur le marché au Japon, publiée en 2020 [11]. Les performances de l'IA étaient les suivantes. Images avec coloration : sensibilité de 96,9 %, spécificité de 100 % et précision de 98,0 %. Images avec NBI : sensibilité de 96,9 %, spécificité de 94,3 % et précision de 96,0 %. Les performances étaient supérieures à celles des endoscopistes en formation, et à celle des experts pour la sensibilité. La valeur prédictive négative pour les polypes diminutifs du rectosigmoïde était de 100 % indiquant que ceux de ces polypes suspects d'êtres des polypes hyperplasiques pourraient être laissés en place selon les seuils de

Tableau 2. Principales études d'évaluation de l'IA pour la caractérisation des polypes coliques.

Étude	Type d'étude	Nombre d'images de polypes	Mode d'image	Vidéo en temps réel	Évaluation des polypes diminutifs	Résultats
Mori et al. (2015)	RS	176	Endocytoscopie	-	-	Sensibilité 92 % Spécificité 79,5 % Précision 89,2 %
Misawa et al. (2016)	RS	100	Endocytoscopie avec NBI	-	-	Sensibilité 84,5 % Spécificité 97,6 % Précision 90,0 %
Kominami et al. (2016)	P	118	NBI + Zoom	+	+	Sensibilité 95,9 % Spécificité 93, 3 % Précision 94, 9%
Takeda et al. (2017)	RS	200	NBI + Zoom	-	-	Sensibilité 89.4% Spécificité 98.9% Précision 94.1%
Mori et al. (2018)	P	466	Endocytoscopie avec NBI et coloration	+	+	Taux de prédiction 98,1 %
Chen et al. (2019)	RS	284	NBI + Zoom	-	+	Sensibilité 96,3 % Spécificité 78,1 % Précision 90,1 %
Repici et al (2020)	RS	125	NBI	+	+	Sensibilité 98 % Spécificité 83 % Précision 94 %

RS: rétrospectif, P : prospectif , NBI : Narrow Band Imaging.

PIVI. Cette étude présentait néanmoins quelques limites, à savoir la toxicité du bleu de méthylène, le temps de la coloration au diagnostic (73 secondes). Les auteurs par ailleurs concluaient que l'endocystoscopie avec coloration virtuelle seule était acceptable mais que la coloration au bleu de méthylène pourrait aider à prédire le degré d'envahissement sous-muqueux. Cela reste cependant à évaluer [11].

Enfin, la méta-analyse publiée en 2020, mentionnée dans le paragraphe précédent, regroupait également 18 études sur la prédiction histologique par IA (dont trois prospectives), 7 680 images de polypes, et démontrait une sensibilité globale de 92.3% (IC 95 % : 88,8 %-94,9 %), et une spécificité de 89,8% (IC 95 % : 83,5 %-93,0 %) [7]. Il est à noter que parmi ces études, cinq concernaient des algorithmes fonctionnant sur le principe du deep learning, avec une sensibilité globale de 94,7 % et une spécificité de 91,7 %. Aussi, la sensibilité globale pour la caractérisation de polypes diminutifs était de 93,5 % et la spécificité de 90,8 %. Enfin, l'IA était globalement supérieure aux endoscopistes non experts [7].

Tâches combinées

Certains logiciels d'IA peuvent effectuer plusieurs tâches telles que la detection puis la caractérisation. C'est le cas par exemple du système CAD EYE (Fujifilm ©). Basé sur une technologie de deep learning, l'algorithme permet de traiter 60 images par secondes soit une analyse en temps réel. Un marquage CE a été obtenu en 2020. La détection est réalisée avec la chromo-endoscopie LCI ou en lumière blanche. La caractérisation est réalisée avec la chromo-endoscopie BLI, sans arrêt sur image et sans zoom. L'algorithme ne montre pas un niveau de confiance ou de précision. Pour les lésions néoplasiques, un cercle d'assistance visuelle s'affiche en jaune (adénomes et cancers). Pour les lésions hyperplasiques, un cercle d'assistance visuelle s'affiche en vert (polypes hyperplasiques et SSA/P sont inclus). Des études de validation sont en cours.

Un autre système intéressant est le système ENDO.ANGEL, développé par apprentissage profond (réseaux neuronaux). Il permet une aide à la détection et la caractérisation des polypes ainsi qu'un contrôle qualité en temps réel par : reconnaissance du caecum, monitoring de la vitesse/temps/processus de retrait et l'évaluation de la qualité de préparation (score de Boston). Le taux de détection d'adénome dans une étude randomisée incluant 704 patients était supérieur avec l'IA (odds ratio 2.30 (95% IC 1.40-3.77, p 0.001)). Le temps de retrait était significativement supérieur avec l'IA (6,38 min +/- 2,48 vs 4,76 ; in +/- 2,54 ; p < 0.001) [12]. En ce qui concerne l'évaluation de la propreté colique, une autre étude, sur image, a montré une précision de 93,33%, supérieure à celle des endoscopistes, dont une précision de 80 % sur les images avec bulles [13].

Conclusions

Aujourd'hui, on peut dire que l'IA est capable de détecter et caractériser des polypes coliques avec une haut degré de précision. La technologie est prête : les performances des équipements informatiques (ordinateurs, puissance de calcul) et logiciels (intelligence artificielle, réseaux neuronaux) ont permis un progrès notable dans ce domaine au cours de ces dernières années. Il existe aujourd'hui des algorithmes de détection validés en Europe et de caractérisation validés au Japon. Cela étant, la majorité des polypes détectés sont diminutifs, posant la question de l'intérêt clinique réel. Aussi, nous ne disposons actuellement que de peu de recul. Par ailleurs, les algorithmes souffrent de certaines limitations : la majorité des études publiées se sont basées sur des images (vs vidéo en temps réel) avec utilisation d'un nombre limité d'images de polypes pour l'entrainement de l'IA puis sa validation. En outre, ce sont des données humaines qui entrainent l'IA, avec des biais qui peuvent être transférés, avec comme conséquence une base de données qui n'est pas représentative de la population. L'algorithme résultant pourrait alors être biaisé. De la même manière, l'utilisation de coloscopies avec un TDA moyen ne permettra pas d'améliorer les performances de l'IA, qui gagnerait à être entrainée par des images de polypes non vus par l'endoscopiste.

In fine, des études supplémentaires sont nécessaires afin d'évaluer les conséquences sur le TDA et, à long terme, sur l'incidence du cancer colo-rectal, et d'évaluer la capacité de l'IA à détecter des polypes plans plus larges. Réponse dans un futur proche...

RÉFÉRENCES

1. Burt RW, Cannon JA, David DS, *et al.* Colorectal cancer screening. *J Natl Compr Cancer Netw JNCCN* 2013 ; 11 : 1538-75. doi : 10.6004/jnccn.2013.0180.
2. Corley DA, Jensen CD, Marks AR, *et al.* Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014 ; 370 : 1298-306. doi : 10.1056/NEJMoa1309086.
3. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006 ; 101 : 343-50. doi : 10.1111/j.1572-0241.2006.00390.x.
4. Bisschops R, East JE, Hassan C, *et al.* Advanced imaging for detection and differentiation of colorectal neoplasia: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* 2019 ; 51 : 1155-79. doi : 10.1055/a-1031-7657.

5. Klare P, Sander C, Prinzen M, *et al.* Automated polyp detection in the colorectum: a prospective study (with videos). *Gastrointest Endosc* 2019 ; 89: 576-82.e1. doi : 10.1016/j.gie.2018.09.042.

6. Repici A, Badalamenti M, Maselli R, *et al.* Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial. *Gastroenterology* 2020 ; 159 : 512-20.e7. doi : 10.1053/j.gastro.2020.04.062.

7. Lui TKL, Guo C-G, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. *Gastrointest Endosc* 2020 ; 92 : 11-22.e6. doi : 10.1016/j.gie.2020.02.033.

8. ASGE Technology Committee, Abu Dayyeh BK, Thosani N, *et al.* ASGE Technology Committee systematic review and meta-analysis assessing the ASGE PIVI thresholds for adopting real-time endoscopic assessment of the histology of diminutive colorectal polyps. *Gastrointest Endosc* 2015;81:502.e1-502.e16. doi : 10.1016/j.gie.2014.12.022.

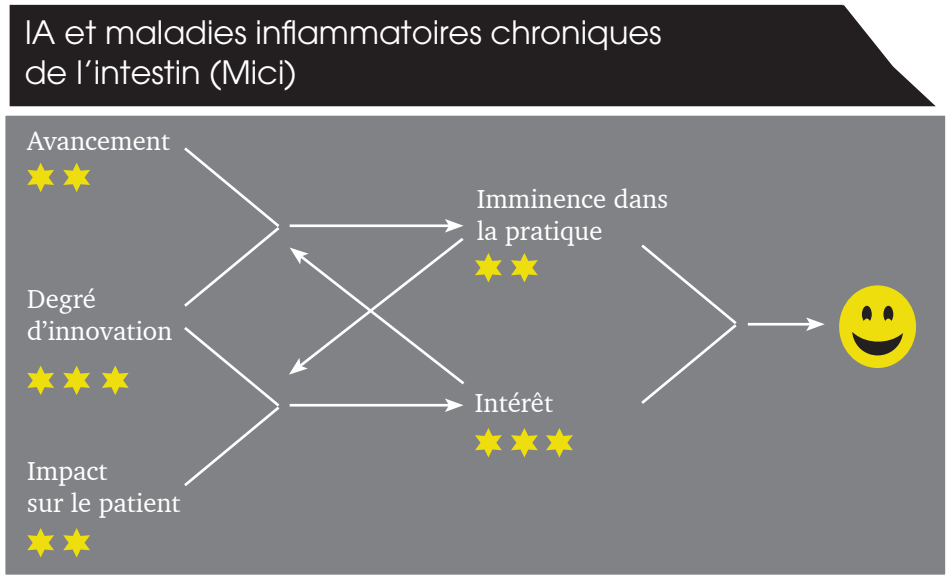
9. Byrne MF, Chapados N, Soudan F, *et al*. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut* 2019;68:94-100. doi : 10.1136/gutjnl-2017-314547.

10. Mori Y, Kudo S-E, Wakamura K, *et al.* Novel computer-aided diagnostic system for colorectal lesions by using endocytoscopy (with videos). *Gastrointest Endosc* 2015 ; 81 : 621-9. doi : 10.1016/j.gie.2014.09.008.

11. Kudo S-E, Misawa M, Mori Y, *et al.* Artificial Intelligence-assisted System Improves Endoscopic Identification of Colorectal Neoplasms. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc* 2020 ; 18 : 1874-81.e2. doi : 10.1016/j.cgh.2019.09.009.

12. Gong D, Wu L, Zhang J, *et al.* Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 2020 ; 5 : 352-61. doi : 10.1016/S2468-1253(19)30413-3.

13. Zhou J, Wu L, Wan X, *et al.* A novel artificial intelligence system for the assessment of bowel preparation (with video). *Gastrointest Endosc* 2020 ; 91 : 428-35.e2. doi : 10.1016/j.gie.2019.11.026.



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Catherine Le Berre
Institut des Maladies de l'Appareil Digestif,
CHU de Nantes - Hôtel Dieu, Nantes,

Les maladies inflammatoires chroniques intestinales (MICI) sont des pathologies dysimmunitaires chroniques du tube digestif, atteignant exclusivement le rectum et le colon pour la rectocolite hémorragique (RCH), et pouvant atteindre tout le tube digestif pour la maladie de Crohn (MC). Aucun « gold standard » n'existe à l'heure actuelle pour affirmer le diagnostic de MICI. Les recommandations européennes actualisées en 2019 précisent que la confirmation diagnostique repose sur un faisceau d'arguments cliniques, biologiques, endoscopiques, histologiques, voire radiologiques [1]. Plusieurs examens sont donc nécessaires au diagnostic initial d'une MICI. La coloscopie avec biopsies étagées multiples reste à l'heure actuelle l'examen de première intention. Une fibroscopie œsogastroduodénale est bien souvent réalisée dans le même temps anesthésique. Une exploration de l'intestin grêle

par entéro-IRM ou vidéocapsule endoscopique (VCE) est recommandée pour les patients chez qui il existe une forte suspicion clinique de MC malgré des explorations endoscopiques conventionnelles normales, et doit être systématique chez tout patient nouvellement diagnostiqué d'une MC [1].

Cependant, un nombre significatif de gastroentérologues (36,9 %) n'a pas d'accès facile à la VCE comme cela a été récemment démontré dans une enquête espagnole [2], et il est bien connu que la lecture d'une entéro-IRM est radiologue-dépendante avec une grande hétérogénéité de qualité d'interprétation selon les centres [3]. Pour ce qui est des examens endoscopiques conventionnels, il a été démontré une variabilité inter-observateurs non négligeable dans l'évaluation de la sévérité de l'atteinte inflammatoire [4]. Aussi, la prise en charge d'une MICI peut varier considérablement d'un patient à l'autre.

Dans les dix dernières années, les « big data » issues de larges essais cliniques et des dossiers médicaux informatisés, les biocollections, les banques de données « -omiques » (génomique, transcriptomique, métabolomique, protéomique) ont été utilisées de façon croissante en vue d'identifier des profils de patients cliniquement pertinents. Ainsi, en combinant une multitude d'informations, l'intelligence artificielle (IA) et les techniques de « machine learning » ont la capacité d'augmenter la performance diagnostique et de prédire l'évolution des MICI dont la physiopathologie complexe allie immunité, microbiote, génétique, et facteurs environnementaux, et ainsi d'uniformiser la prise en charge des patients atteints de ces maladies.

Applications de l'IA dans le diagnostic et l'évaluation de la sévérité des MICI

Diagnostic

Comme dit précédemment, l'évaluation de l'intestin grêle est souvent problématique chez les patients nouvellement diagnostiqués ou suspectés d'une MC. Pourtant, les lésions du grêle proximal sont associées à un risque significatif de sténoses et de chirurgies multiples, avec un impact pronostique et thérapeutique potentiellement important. Le temps nécessaire à la lecture d'une VCE (25-60 minutes) est une limite importante à la réalisation de cet examen, et une attention soutenue durant tout le temps de lecture est nécessaire en raison de la vitesse de défilement des images (15-20 images par seconde). De nombreuses équipes se sont donc attachées à développer des programmes de détection automatique des lésions associées à la MC en VCE, avec pour la plupart une précision diagnostique dépassant les 90 % [5–7]. Avec l'apparition du « deep learning » et de réseaux de

neurones beaucoup plus puissants, le prétraitement des données n'est quasiment plus utilisé au profit d'un apprentissage dit peu supervisé, où les données sont fournies à l'algorithme de calcul sans autre prétraitement qu'une annotation sur leur nature pathologique ou non. L'intérêt principal de ces réseaux de neurones est leur simplicité d'usage une fois développés et leur capacité d'auto-amélioration. L'équipe du CHU de Nantes en collaboration avec celle du LS2N a récemment développé son propre réseau neuronal entraîné sur une base de données personnelle nommée Crohn-IPI, composée de 1 628 images normales et 1 590 images de lésions de MC, obtenant une exactitude de 89,3 %, une sensibilité de 87,8 % et une spécificité de 90,7 % [8].

Des techniques de « machine learning » ont également été développées pour la lecture semi-automatique d'entéro-scanner ou -IRM, principalement basées sur l'épaisseur pariétale, la dilatation des anses grêliques, le diamètre luminal, et l'existence de sténoses, et ont prouvé leur efficacité pour identifier le dommage intestinal avec une précision similaire à celle de radiologues experts [9,10].

Les examens endoscopiques et morphologiques sont parfois mis en défaut pour distinguer les deux phénotypes de MICI. Pourtant, distinguer une MC d'une RCH a une grande importance dans le choix des traitements médicaux mais surtout des modalités chirurgicales, notamment en cas de nécessité de colectomie totale avec éventuelle anastomose iléo-anale. Les caractéristiques génétiques des patients pourraient être une aide à la distinction phénotypique. À l'heure actuelle, les études d'association pangénomique (GWAS) ont identifié plusieurs centaines de loci associés au risque de développer une MC ou une RCH [11], et une technique de hiérarchisation génique basée sur le « machine learning » a récemment permis d'extraire 67 nouveaux gènes candidats parmi une liste de 16 390 gènes [12]. Une équipe américaine a développé en 2018 un algorithme de probabilité diagnostique basé non plus uniquement sur les gènes présents mais sur les interactions géniques à l'échelle individuelle, permettant de distinguer les patients atteints d'une MC de ceux atteints d'une RCH avec une bonne efficacité [13].

Évaluation de la sévérité

Au vu d'une certaine subjectivité dans l'évaluation des scores de sévérité endoscopique dans la RCH (sous-score endoscopique de Mayo, UCEIS), de nombreuses équipes ont travaillé au développement d'algorithmes de « deep learning » visant une classification automatique de la sévérité des lésions décelées en endoscopie chez ces patients. Par exemple, une équipe japonaise a développé un réseau neuronal permettant d'identifier les patients en rémission endoscopique définie par

un Mayo 0 ou 0-1 avec une aire sous la courbe (AUROC) de 0,86 et 0,98, respectivement, avec une meilleure performance dans le rectum que dans le reste du colon [14]. Une équipe américaine a elle aussi développé son propre réseau neuronal permettant de distinguer les patients Mayo 0-1 (rémission endoscopique) des patients Mayo 2-3 (activité modérée à sévère) avec une AUROC de 0,966 [15]. Au-delà de distinguer les patients en rémission de ceux en activité, deux équipes ont chacune développé leur algorithme permettant de définir l'exact sous-score endoscopique de Mayo, avec une concordance lecture automatisée - humaine aux alentours de 70 % dans les deux cas [16,17].

Bien que toujours débattue, la cicatrisation histologique est de plus en plus recherchée après mise en place d'un traitement chez les patients atteints de RCH, imposant la réalisation de biopsies à chaque endoscopie de contrôle. Plusieurs équipes ont donc développé des algorithmes de « deep learning » visant à s'affranchir de l'analyse anatomopathologique en se basant uniquement sur les images d'endoscopie. Le réseau neuronal développé par Takenaka *et al.* permet par exemple d'identifier en temps réel les patients en rémission endoscopique (définie par un score UCEIS à 0) avec une précision de 90,1 % et ceux en rémission histologique (définie par un score de Geboes inférieur ou égal à 3) avec une précision de 92,9 % [18]. Une équipe belge a quant à elle développé un algorithme définissant, sur des images endoscopiques de patients atteints d'une RCH minime à modérée, un score de « densité de rouge » significativement corrélé à l'index histologique de Robarts (r = 0,74), au sous-score endoscopique de Mayo (r=0,76) et au score UCEIS (r = 0,74) [19]. Une autre équipe japonaise a utilisé la technique d'endocytoscopie pour développer un système d'aide au diagnostic en temps réel permettant de détecter une inflammation histologique persistante (définie par un score de Geboes supérieur ou égal à 3,1) avec une sensibilité de 74 %, une spécificité de 97 %, une précision de 91 %, et surtout une reproductibilité parfaite (100 %) [20].

Toutes les études précédemment citées décrivaient des réseaux neuronaux entraînés sur des images fixes capturées lors d'une endoscopie, mais une équipe américaine est récemment parvenue à développer un algorithme entraîné sur des vidéos entières d'endoscopie issues de l'essai de phase 2 évaluant le mirikizumab, un anti-IL23, chez les patients atteints de RCH modérée à sévère, et permettant de définir le sous-score endoscopique de Mayo avec une concordance machine-humain de 0,844, et le score UCEIS avec une concordance de 0,855 [21].

Les études pour l'évaluation automatique de la sévérité des lésions chez les patients atteints de MC sont plus rares. L'équipe de Kopylov a récemment développé un ré-

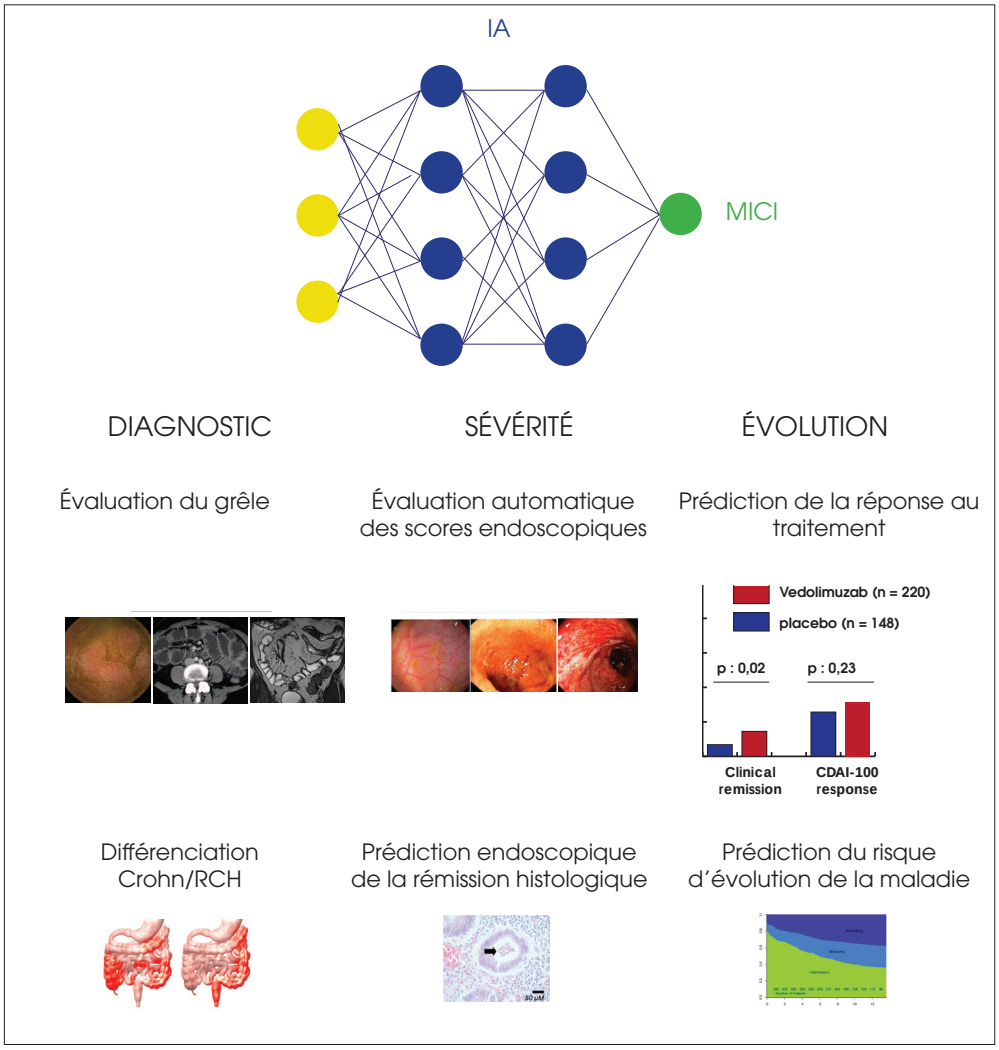


FIGURE 1. Intérêt de l'intelligence artificielle dans la prise en charge des MICI.

seau neuronal sur plus de 17 640 images de VCE issues de 49 patients atteints d'une MC ; 7391 images comportant au moins une ulcération muqueuse et 10 249 images normales. Les ulcères étaient classés en trois grades, de minime à sévère, par deux experts, avec une concordance inter-observateurs de 31 % seulement, mais atteignant tout de même 76 % pour la distinction des grades 1 et 3. Le réseau neuronal parvenait à distinguer les ulcères de grade 1 versus 3 avec une précision diagnostique de 0,91, les ulcères de grade 2 *versus* 3 avec une précision de 0,78, et les ulcères de grade 1 versus 2 avec une précision de 0,62, permettant d'envisager à terme une aide précieuse au diagnostic et au monitoring des patients atteints d'une MC du grêle [22].

Applications de l'IA dans la prédiction de la réponse au traitement et de la rechute

La meilleure compréhension de la physiopathologie des MICI a permis d'importantes avancées ces dernières décennies quant aux thérapies disponibles et aux objectifs thérapeutiques à atteindre dans les MICI. En effet, l'avènement des biothérapies (anti-TNF, anti-IL12/IL23, anti-intégrine) et des petites molécules (inhibiteurs de JAK) a révolutionné la prise en charge de ces maladies. Cependant, la réponse globale à chacun de ces traitements est inférieure à 50 %, et il est encore à l'heure actuelle impossible de prédire la réponse au traitement chez les patients atteints d'une MICI.

Des systèmes basés sur l'IA ont donc été développés pour l'analyse des indicateurs prédictifs de la réponse aux différents traitements disponibles dans les MICI. Waljee *et al.* ont par exemple développé un algorithme utilisant l'âge et des paramètres biologiques communs permettant de différencier les patients répondeurs des patients non répondeurs à un traitement par thiopurines avec une AUROC de 0,856, versus 0,594 en utilisant le dosage des 6-TGN [23]. La même équipe a utilisé les données des essais cliniques GEMINI, évaluant l'efficacité du vedolizumab chez les patients atteints d'une RCH et d'une MC, pour développer un modèle de « machine learning » permettant de prédire dès la semaine 6 quels patients seraient en rémission clinique sans corticoïdes à la semaine 52, avec une AUROC de 0,73 pour la RCH, et 0,75 pour la MC [24,25], permettant de sélectionner les patients chez qui poursuivre le vedolizumab malgré l'absence de bénéfices apparents à la semaine 6. Plus récemment, une analyse de facteurs multi-omiques a été réalisée pour identifier des biomarqueurs transcriptomiques et génomiques prédictifs de la réponse endoscopique à un traitement par ustekinumab avec une précision diagnostique allant jusqu'à 98 % [26]. Dans le cadre d'une colite aiguë grave, prédire la réponse au traitement est encore plus crucial car environ 15 % des patients nécessitent encore une colectomie en 2020, et un traitement médical inefficace peut entraîner un retard de prise en charge et une augmentation de la morbidité

et de la mortalité. Une étude française a développé un algorithme basé sur des profils d'expression de micro-ARN sur des biopsies coliques issues de patients atteints de RCH en colite aiguë grave, permettant d'identifier les répondeurs aux corticoïdes avec une précision de 93 %, à l'infliximab avec une précision de 84 % et à la ciclosporine avec une précision de 80 % [27]. Plus récemment, une équipe indienne a développé un réseau neuronal basé sur des paramètres clinico-biologiques simples, tels que le taux d'hémoglobine et l'albuminémie, permettant de prédire efficacement la réponse au traitement médicamenteux chez les patients atteints de colite aiguë grave [28].

L'IA peut aussi être utilisée pour prédire l'évolution des MICI. Un réseau neuronal analysant les biopsies issues de patients atteints d'une MC à un stade précoce permettait d'identifier ceux à risque de progression sténosante et/ou fistulisante et ceux à risque de chirurgie avec une précision de plus de 80 % [29]. L'équipe de Waljee a développé un autre modèle basé sur l'analyse des dossiers médicaux électroniques permettant de prédire les hospitalisations et le recours à la corticothérapie dans les 6 mois avec une AUROC de 0,87 [30].

Conclusion et perspectives

L'IA est une discipline grandissante qui a le potentiel de révolutionner la prise en charge des patients atteints de MICI. Les techniques de « machine learning » offrent la possibilité de synthétiser d'énormes quantités de données afin d'améliorer nos performances pour le diagnostic des MICI, d'homogénéiser les prises en charge en automatisant l'évaluation de la sévérité des lésions endoscopiques, et de prédire la réponse aux traitements et le cours évolutif de la maladie pour une médecine de plus en plus personnalisée. L'IA devrait aussi permettre d'améliorer la détection endoscopique des lésions de dysplasie non conventionnelle propres aux patients atteints d'une MICI, si difficiles à détecter à l'heure actuelle avec les techniques d'endoscopie standard. Certains obstacles restent encore à franchir avant l'implémentation de l'IA en pratique courante, mais il est certain que les techniques de « deep learning » nous accompagneront très bientôt pour la prise en charge au quotidien de nos patients atteints de MICI.

RÉFÉRENCES

1. Maaser C, Sturm A, Vavricka SR, *et al.* ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. *J Crohns Colitis* 2019 ; 13 : 144-64K.

2. Marín Jiménez I. Survey on the use of capsule endoscopy for patients with known or suspected Crohn's disease in Spain. *Rev Esp Enferm Dig* 2015 ; 107. http://online.reed.es/fichaArticulo.aspx?iarf=221687760-749233419277.

3. Taylor SA, Mallett S, Bhatnagar G, *et al.* Magnetic resonance enterography compared with ultrasonography in newly diagnosed and relapsing Crohn's disease patients: the METRIC diagnostic accuracy study. *Health Technol Assess* 2019 ; 23 : 1–162.

4. Daperno M, Comberlato M, Bossa F,*et al*. Inter-observer agreement in endoscopic scoring systems: preliminary report of an ongoing study from the Italian Group for Inflammatory Bowel Disease (IG-IBD). *Dig Liver Dis* 2014 ; 46 : 969–973.

5. Girgis HZ, Mitchell BR, Dassopoulos T, *et al.* An intelligent system to detect Crohn's disease inflammation in Wireless Capsule Endoscopy videos 2010; 1373–76.

6. Kumar R, Qian Zhao, Seshamani S, *et al.* Assessment of Crohn's Disease Lesions in Wireless Capsule Endoscopy Images. *IEEE Transactions on Biomedical Engineering* 2012 ; 59 : 355–62.

7. Klang E, Barash Y, Margalit RY,*et al.* Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy. *Gastrointest Endosc* 2020 ; 91 : 606-13.e2.

8. Vallée R, Coutrot A, Normand N, *et al*. *Accurate small bowel lesions detection in wireless capsule endoscopy images using deep recurrent attention neural network. IEEE 21st International Workshop on Multimedia Signal Processing (MMSP 2019)*. Kuala Lumpur (Malaisie), 2019. https://hal.archives-ouvertes.fr/hal-02296282

9. Mahapatra D, Vos FM, Buhmann JM. Active learning based segmentation of Crohns disease from abdominal MRI. *Comput Methods Programs Biomed*. 2016;128:75–85.

10. Stidham RW, Enchakalody B, Waljee AK, *et al.* Assessing Small Bowel Stricturing and Morphology in Crohn's Disease Using Semi-automated Image Analysis. *Inflamm Bowel Dis* 2020 ; 26 : 734–42.

11. Cleynen I, Boucher G, Jostins L, *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016 ; 387 : 156–67.

12. Isakov O, Dotan I, Ben-Shachar S. Machine Learning-Based Gene Prioritization Identifies Novel Candidate Risk Genes for Inflammatory Bowel Disease. *Inflamm Bowel Dis* 2017 ; 23:1516–23.

13. Han L, Maciejewski M, Brockel C, *et al.* A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease. *Bioinformatics* 2018 ; 34 : 985–93.

14. Ozawa T, Ishihara S, Fujishiro M, *et al.* Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointest Endosc* 2019 ; 89 : 416-21.e1.

15. Stidham RW, Liu W, Bishu S, *et al*. Performance of a Deep Learning Model vs Human Reviewers in Grading Endoscopic Disease Severity of Patients With Ulcerative Colitis. *JAMA Netw Open* 2019 ; 2 : e193963.

16. Yao H, Najarian K, Gryak J, *et al.* Fully automated endoscopic disease activity assessment in ulcerative colitis. *Gastrointest Endosc* 2020; epub. doi: 10.1016/j.gie.2020.08.011

17. Bhambhvani HP, Zamora A. Deep learning enabled classification of Mayo endoscopic subscore in patients with ulcerative colitis. *Eur J Gastroenterol Hepato*l 2020; epub. doi: 10.1097/MEG.0000000000001952

18. Takenaka K, Ohtsuka K, Fujii T, *et al.* Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology* 2020 ; 158 : 2150–57.

19. Bossuyt P, Nakase H, Vermeire S, *et al.* Automatic, computer-aided determination of endoscopic and histological inflammation in patients with mild to moderate ulcerative colitis based on red density. *Gut 2*020 ; 69 : 1778–86.

20. Maeda Y, Kudo S-E, Mori Y, *et al.* Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest Endosc*. 2019 ; 89 : 408–415.

21. Gottlieb K, Requa J, Karnes W, *et al.* Central Reading of Ulcerative Colitis Clinical Trial Videos Using Neural Networks. *Gastroenterology* 2020; epub.

22. Barash Y, Azaria L, Soffer S, *et al.* Ulcer severity grading in video capsule images of patients with Crohn's disease: an ordinal neural network solution. *Gastrointest Endosc* 2020; epub. doi : 10.1016/j.gie.2020.05.066.

23. Waljee AK, Joyce JC, Wang S, *et al.* Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. *Clin Gastroenterol Hepatol* 2010 ; 8 : 143–50.

24. Waljee AK, Liu B, Sauder K, *et al.* Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Aliment Pharmacol Ther* 2018 ; 47 : 763–72.

25. Waljee AK, Liu B, Sauder K, *et al.* Predicting Corticosteroid-Free Biologic Remission with Vedolizumab in Crohn's Disease. *Inflamm Bowel Dis* 2018 ; 24: 1185–92.

26. Verstockt B, Sudahakar P, Creyns B, *et al.* DOP70 An integrated multi-omics biomarker predicting endoscopic response in ustekinumab treated patients with Crohn's disease. *Journal of Crohn's and Colitis* 2019 ; 13 : S072-73.

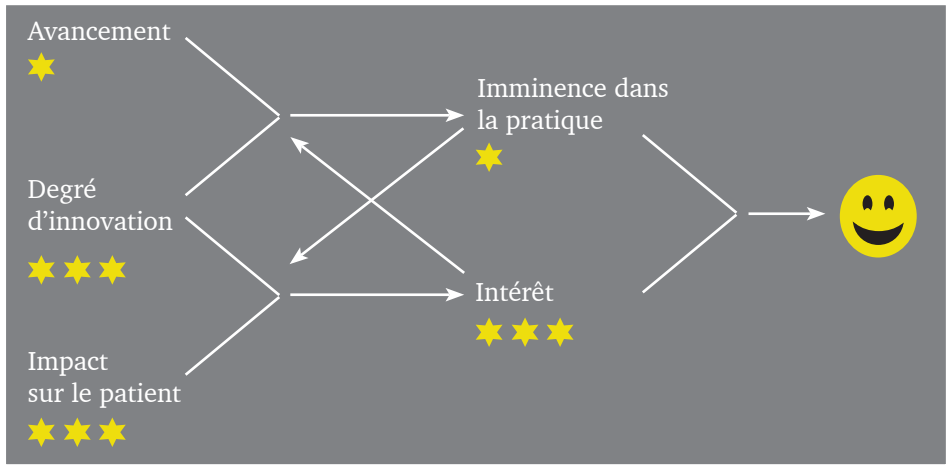
27. Morilla I, Uzzan M, Laharie D, *et al.* Colonic MicroRNA Profiles, Identified by a Deep Learning Algorithm, That Predict Responses to Therapy of Patients With Acute Severe Ulcerative Colitis. *Clin Gastroenterol Hepatol* 2019 ; 17 : 905–13.

28. Ghoshal UC, Rai S, Kulkarni A, *et al.* Prediction of outcome of treatment of acute severe ulcerative colitis using principal component analysis and artificial intelligence. *JGH Open* 2020 ; 4 : 889–97.

29. Klein A, Mazor Y, Karban A, *et al.* Early histological findings may predict the clinical phenotype in Crohn's colitis. *United European Gastroenterol J* 2017; 5: 694–701.

30. Waljee AK, Lipson R, Wiitala WL, *et al.* Predicting Hospitalization and Outpatient Corticosteroid Use in Inflammatory Bowel Disease Patients Using Machine Learning. *Inflamm Bowel Dis*. 2017;24:45–53.

IA et pathologie pancréatique



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Geoffroy Vanbiervliet
CHu de Nice

L'intelligence artificielle (IA) dans le domaine médical est à l'origine d'un engouement considérable. Ce phénomène est relativement aisé à comprendre : il s'agit d'un moyen simple (malgré ce que l'on pense) et efficace d'être plus précis, plus précoce dans son diagnostic, plus reproductif, et de gagner du temps ou plutôt d'offrir plus de temps humain au patient à l'inverse de ce qui pourrait être évoqué. Or les pathologies pancréatiques, par leur pronostic souvent et malheureusement réservé (comme pour l'adénocarcinome ductulaire), par les conséquences thérapeutiques notamment chirurgicales qu'elles impliquent, demandent une certaine agilité diagnostique encore insuffisante aujourd'hui. Si l'évaluation de l'IA dans le domaine des maladies du pancréas n'en est qu'à un stade débutant et balbutiant, il est fort à parier que la dextérité qu'elle va apporter au praticien dans les prochaines années dans le domaine

devrait la rendre inévitable et irremplaçable. Ce chapitre a pour but de présenter les premiers résultats concernant l'IA en pancréatologie et en particulier en évoquant l'utilisation de l'écho-endoscopie dont l'utilisation est incontournable dans le domaine.

Deep learning et imagerie pancrétique

L'évaluation de l'IA reste débutante pour toutes les techniques d'imagerie pancréatique. Pour exemple, l'utilisation du Deep Learning dans l'analyse des images pancréatiques obtenues par tomodensitométrie n'a été que récemment décrite. Une étude taïwanaise a ainsi obtenu une précision diagnostique par l'IA de 0,989 avec une aire sous la courbe (ROC) quasi parfaite de 0,999 (IC 95% : 0,998 à 1,000) dans l'adénocarcinome pancréatique [1]. Le système informatique présentait aussi une sensibilité significativement supérieure que le radiologue (0,983 *vs* 0,929, différence 0,054 [IC 95% - 0,011 à 0,098] ; p=0,014). Il en est de même pour l'IRM où les aires sous la courbe dans les études préliminaires semblent pour l'instant inférieures et fluctuent entre 0,7189 pour les maladies pancréatiques inflammatoires et 0,93019 pour le diagnostic de tumeur neuro-endocrine [2]. La corrélation avec le lecteur humain était par ailleurs excellente (coefficient κ de 0,8862 (IC 95% - 0,7759 à 0,9738)) dans cette même étude qui était multicentrique et rétrospective. Aucune étude visant l'exploration de l'IA en échographie percutanée pancréatique n'a été publiée à ce jour. En résumé, l'on peut dire que la littérature est prometteuse mais préliminaire concernant l'imagerie pancréatique non invasive en coupe.

Écho-endoscopie pancréatique et IA

Diagnostic de l'adénocarcinome

Une des problématiques en pancréatologie reste la différenciation entre tissu tumoral et pseudo-tumoral inflammatoire, voire normal *(tableau 1)*. Les diagnostics inappropriés de cancer sont encore à l'origine par excès de près de 7% de duodéno-pancréatectomies céphaliques malgré les techniques modernes d'imagerie [3]. De manière surprenante, les premières études ayant proposé l'IA comme outil permettant de différencier les composantes et caractéristiques du tissu pancréatique datent des années 2000. Trois études entre 2000 et 2008 ont ainsi été publiées, n'offrant qu'un niveau de preuve faible mais ouvrant la voie à l'utilisation de l'intelligence augmentée pour obtenir une aide dans le diagnostic précoce de l'adénocarcinome ductulaire pancréatique.

La publication historique en 2001 d'une équipe de la Mayo Clinic démontrait déjà toute la faisabilité d'un logiciel informatique capable de digitaliser chaque colonne de pixels par variation des niveaux de gris sur un simple cliché de masse pancréatique, scanné, obtenu à l'aide d'une sonde d'écho-endoscopie radiale mécanique [4]. Une

analyse histologique était disponible pour chaque lésion analysée. Alors que le logiciel ne bénéficiait que d'une seule image, générée par une sonde paraissant aujourd'hui obsolète, et sans avoir le support clinique ou iconographique complémentaire du patient, celui-ci présentait déjà des performances similaires à l'expert humain, parfaitement sensible (100%) mais encore peu spécifique (50 %).

C'est en 2008 que l'utilisation de programmations basées sur des réseaux de neurones artificiels a été décrite pour l'analyse des images pancréatiques obtenues par écho-endoscopie. Après avoir entraîné le système à l'aide d'un échantillon d'aires tissulaires soit normales, soit présentant des signes typiques de pancréatite chronique (lobularité, calcifications et trame hyper échogène) et enfin d'adénocarcinome, l'équipe de Das *et al.* [5] montrait sur un échantillon de validation que le logiciel obtenait une excellente aire sous la courbe de 0,93 (AUROC) en termes de précision diagnostique.

Ces mêmes techniques mathématiques basées sur des réseaux de neurones ont alors été utilisées pour l'analyse des histogrammes d'élastographie et la prise de contraste (SonoVue*) des syndromes de masse pancréatique sous écho-endoscopie [6-7]. Ils permettaient pour les deux méthodes largement utilisées en pratique courante, une amélioration de l'aire sous la courbe de la performance diagnostique : l'utilisation du logiciel entrainait pour l'analyse de l'élastographie tumorale une augmentation de 0,847 (IC 95 %; 0,721 à 0,972) à 0,957 pour distinguer une lésion bénigne vs. maligne et enfin à 0,965 pour distinguer une pancréatite pseudo-tumorale d'un adénocarcinome [6]. Un gain de sensibilité était également noté lorsque le logiciel était couplé à l'évaluation quantitative de la prise de contraste (84,8 % pour la ponction à l'aiguille fine *vs* 87,5% pour la seule prise de contraste *vs* 94,6 % pour la prise de contraste couplée à l'IA) [7]. Les limites communes de ces premières études restent des échantillons de validation de petite taille, une méthode de validation croisée interne uniquement, une absence de relecture et de comparaison avec les performances des experts « humains » et une preuve histologique inconstante de la nature des tissus analysés, notamment en cas de pancréatite chronique.

Plus récemment le concept de diagnostic écho-endoscopique de cancer pancréatique assisté par l'informatique (Endoscopic Ultrasound computer assisted diagnosis ou EUS-CAD) a pris son sens par l'intermédiaire de trois études. L'utilisation soit de réseaux de neurones artificiels soit des machines à vecteurs de support était décrite pour effectuer le diagnostic différentiel entre cancer pancréatique et pancréatite chronique par analyse de la texture des images écho-endoscopiques [8–10]. Les précisions diagnostiques de ces modèles mathématiques étaient alors élevées allant de 87 à 98 %, bien qu'une seule méthode de validation croisée interne y fût adoptée.

Tableau 1. Performances diagnostiques de l'IA et diagnostic d'adénocarcinome pancréatique en écho-endoscopie.

Auteur, année	N	Technique
Norton ID, 2001	35 (14 PC et 21 AC)	Écho-endoscopie radiale mécanique
Das A, 2008	56 (22 PN, 12 PC et 22 AC)	Écho-endoscopie radiale mécanique
Saftoiu A, 2008	68 (22 PN, 11 PC, 32 AC et 3 TNE)	Écho-endoscopie linéaire électronique et élastographie (histogramme)
Zhang MM, 2010	216 (153 AC, 43 PC, 20 PN)	Écho-endoscopie linéaire électronique
Zhu M, 2013	388 (262 AC et 126 PC)	Écho-endoscopie linéaire électronique
Saftoiu A, 2015	167 (112 AC et 55 PC)	Écho-endoscopie linéaire électronique et prise de contraste (SonoVue®) (Courbe temps-intensité)
Ozkan M, 2016	323 (202 AC et 130 PN)	Écho-endoscopie linéaire électronique
Tonozuka R, 2020	139 (76 AC, 34 PC, et 29 pancréas normaux)	Écho-endoscopie linéaire électronique
Marya NB, 2020	123 (60 AC, 28 PAI, 16 PC et 19 pancréas normaux)	Écho-endoscopie linéaire électronique

Précision diagnostique (AUROC)	Sensibilité (%)	Spécificité (%)
80	100	50
0,93	93 (IC 95% (89-97))	92 (IC 95% (88-96))
0,965	-	-
0,98	94,3	99,4
0,94	96,2	93,4
	94,6 (IC 95 % (88,2-97,8))	94,4 (IC 95% (83,9-98,6))
0,875	83,3	93,3
0,94	92,4	84,1
0.976	95 (IC 95% (91-98))	Images fixes

Finalement la notion de « Deep Learning » en écho-endoscopie pancréatique n'a vu le jour qu'en 2020 avec une étude prospective japonaise préliminaire sur 139 malades (76 avec adénocarcinome, 34 avec pancréatite chronique, et 29 avec pancréas normal) [11]. Sur 920 images utilisées pour l'entraînement et la validation du réseau convolutionnaire de neurones, seules 470 constituèrent l'échantillon de test (47 cas testés de manière séparée et indépendante). Les performances du système étaient comparées soit à l'histologie (en cas d'adénocarcinome), soit au diagnostic écho-endoscopique de pancréatite chronique selon la classification de Rosemont. Ainsi le logiciel présentait une sensibilité, une spécificité, une valeur prédictive positive et négative de 92,4%, 84,1%, 86,8%, et 90,7% respectivement. Ces résultats sont bien évidemment excellents, quasi identiques de prime abord à la meilleure de nos aiguilles coupantes utilisées pour obtenir du matériel histologique pancréatique sous écho-endoscopie.

De la même manière, l'IA apparaît très compétitive, voire meilleure que les experts humains dans le domaine. Une étude rétrospective récente sur une cohorte test de 123 patients confirmait d'une part que la corrélation inter-observateurs n'était que modérée (coefficient κ de 0,458) entre experts (au nombre de 7 dans ce travail) et montrait surtout que la machine présentait comparativement à l'homme une performance diagnostique significativement plus élevée pour les principales pathologies pancréatiques incluant l'adénocarcinome (75,6% *vs* 61,6% ; p=0,026) [12].

Diagnostic des lésions kystiques pancréatiques

Voilà un graal en écho-endoscopie dont la conquête est toujours galopante : déterminer de manière optimale le diagnostic des lésions kystiques mucineuses pancréatiques et leur potentiel dégénératif. Il n'en fallait pas plus pour lancer également l'EUS-CAD dans ce domaine. Une seule étude rétrospective est à ce jour publiée mais elle montre la voie [13]. Celle-ci basée sur une population de 206 malades dont le diagnostic de tumeur intra-pancréatique mucineuse papillaire (TIPMP) fut finalement établi par examen anatomo-pathologique de pièce opératoire. Parmi ces patients, une cohorte test de 50 malades (dont 23 avec cancer) et 3,970 images écho-endoscopiques de TIPMP fut analysée. L'aire sous la courbe ROC concernant la capacité de l'IA à effectuer un diagnostic correct de TIPMP maligne était de 0,98, significativement supérieure à la présence d'un nodule mural lors de l'examen (0,74, p=0,001). Non seulement les performances diagnostiques de la machine (94 %) étaient supérieures à celle de l'homme (diagnostic présomptif préopératoire de l'opérateur) (56 %), mais également à celle des indications retenues de chirurgie, relatives (40 %) et absolues (68 %), par les recommandations internationales.

Diagnostic de la pancréatite auto-immune

La pancréatite auto-immune (PAI) de type I se présente typiquement avec un ictère et un syndrome de masse pseudo-obstructif chez les personnes âgées de plus de 60 ans expliquant les difficultés de diagnostic. Dans ce contexte précis, l'utilisation de l'IA prend tout sens afin de distinguer une lésion tissulaire plutôt inflammatoire d'un authentique adénocarcinome. C'est ainsi que l'équipe de Michaël Levy de la Mayo clinic a rapporté également récemment son expérience rétrospective basée sur 4945 images et 1 852 vidéos d'écho-endoscopie collectées lors de l'exploration de 583 patients ayant présenté des PAI, des pancréatites chroniques, des adénocarcinomes ou bien encore des pancréas normaux [12].

Le système de réseau convolutionnaire neuronal était élaboré selon la même technique connue avec une assignation au hasard de cas pour l'entraînement et la validation de la machine (460 patients dont 118 (25,7%) avec PAI) et une cohorte de patients test (123 patients dont 28 (22,8%) avec PAI). La vitesse d'exécution du système était remarquable puisqu'elle atteignait 955 images par seconde, largement au-dessus des 30 requises pour effectuer une analyse en temps réel de l'acquisition écho-endoscopique actuelle.

Les performances diagnostiques obtenues étaient excellentes en particulier lorsque seules les vidéos étaient prises en compte pour l'analyse. Ainsi pour distinguer la PAI de la pancréatite chronique, la sensibilité et spécificité du système étaient de 94 % et 71 % respectivement avec une aire sous la courbe (ROC) de 0,892 (IC 95 % : 0,829 à 0,946) ; pour distinguer la PAI de l'adénocarcinome, la sensibilité et spécificité étaient de 90 % et 93 % respectivement avec une aire sous la courbe (ROC) de 0,963 (IC 95 % : 0,941 à 0,981). Enfin de la même manière, les performances du système furent comparées aux performances des experts humains pour distinguer PAI et adénocarcinome. Si la spécificité était à peu près identique avec un taux de 88,2% (IC 95% : 72,6% à 96,7%) pour l'IA *vs* 82,4% (76,9% à 87,0%) pour l'homme, la machine prenait assez largement l'avantage en termes de sensibilité (88,2% (IC 95% : 63,6% à 98,5%) *vs* 53,8% (IC 95% : 44,4% à 62,9%).

Apprentissage de techno-endoscopie pancréatique

Le domaine de la pédagogie et de l'apprentissage est également concerné depuis peu par l'IA. Nous savons l'écho-endoscopie difficile et longue à appréhender, à apprendre, la rendant très opérateur-dépendant *in fine* dans sa pratique. L'IA vient d'être proposée par une équipe chinoise comme un modèle de plateforme d'entraînement et de contrôle de qualité [14]. Six aires pancréatiques étaient identifiées et classifiées

au préalable au sein de 19,486 images. Une validation interne et externe était réalisée ainsi que des stations « tests » à l'aide de 396 clips vidéo. Une comparaison entre le modèle de prédiction obtenu et des experts en EUS, ainsi que l'amélioration de la vitesse d'apprentissage par une étude en *cross over* chez huit étudiants furent également évaluées.

Les résultats ainsi obtenus sont excellents puisque lors du test vidéo, la précision par image du modèle pour classifier correctement les aires était de 86,2 % avec un excellent coefficient de corrélation entre experts et machine supérieur à 0,8. Dans la reconnaissance des aires pancréatiques, le système permettait un gain significatif dans la vitesse d'apprentissage des étudiants avec une précision diagnostique passant de 0,672 à 0,784 (différence de 0,112 (IC 95 % : 0,058 à 1,663), p = 0,002).

Dans un monde où le « jamais la première fois sur l'homme » est en train de prévaloir avec une utilisation de plus en plus codifiée et démocratisée de la simulation et des modèles *ex vivo* et *in vivo* lors des premières étapes de l'apprentissage des étudiants, gageons qu'un modèle d'assistance de ce type pour les premiers passages chez l'homme devrait être d'une aide avantageuse. Il n'en reste pas moins que le compagnonnage garde tout son sens dans une discipline où l'œil et l'expérience humaine verbalisée ont démontré toute leur valeur.

Futur, espoir et limite

Ce chapitre situe probablement simplement et clairement quel potentiel immense l'IA présente dans le domaine de la pancréatologie et en particulier de son exploration par écho-endoscopie. Rapide et simple d'usage, nous pouvons légitimement dans un futur proche en attendre des performances diagnostiques toujours plus élevées avec des sensibilités déjà excellentes plaçant la technique au même niveau voire au-dessus de celles obtenues par une ponction écho-guidée avec aiguille coupante. Or, comment ne pas vouloir éviter toute comorbidité inutile à son patient avec une ponction inutile ? Comment ne pas vouloir éviter la variabilité inéluctable dans la vraie vie de l'interprétation anatomo-pathologique de prélèvements délicats à analyser ? comment ne pas vouloir ne plus être à la merci d'une technique nécessitant, malgré tout le bien que nous pensons d'elle, un haut degré d'expertise et de connaissance écho-anatomique, scientifique et endoscopique.

Plusieurs limites se dressent encore malgré tout sur ce chemin : pour approvisionner un système d'intelligence augmentée, il est nécessaire d'avoir des bases de données conséquentes et de collecter un grand nombre d'images, annotées, ce qui n'est pas toujours ef-

fectué aussi aisément en écho-endoscopie. Le retard pris dans ce domaine en 2021, notamment en France (et ailleurs) trouve probablement sa naissance dans des aspects éthiques et légaux parfois complexes. Mais celui-ci constitue un frein certain qui pourrait être levé par l'organisation de base de données médicales encryptées anonymes virtuellement et aisément disponibles, déterminant des ressources de recherche et d'exploitation inépuisable, comme le modèle Nord-américain de l'institut WestHealth.org le propose. Sortons enfin du suranné débat méthodologique rétrospectif/insuffisant *vs* prospectif/idéal : la qualité des études devraient croître inévitablement. Il restera en revanche la nécessité d'avoir un *gold standard* histologique de qualité, dont le manque constitue le reproche régulièrement fait aux études évaluant les explorations diagnostiques pancréatiques en écho-endoscopie.

La convergence de ces trois points explique qu'il faudra des études d'envergure nationale voire internationale, multicentriques, pour asseoir définitivement la technique au sein de notre arsenal diagnostique..

RÉFÉRENCES

1. Liu K-L, Wu T, Chen P-T, et al. Deep learning to distinguish pancreatic cancer tissue from non-cancerous pancreatic tissue: a retrospective study with cross-racial external validation. Lancet Digit Health 2020; 2: e303–e313. doi:10.1016/S2589-7500(20)30078-9
2. Gao X, Wang X. Performance of deep learning for differentiating pancreatic diseases on contrast-enhanced magnetic resonance imaging: A preliminary study. Diagn Interv Imaging 2020; 101: 91–100. doi:10.1016/j.diii.2019.07.002
3. Gerritsen A, Molenaar IQ, Bollen TL, et al. Preoperative characteristics of patients with presumed pancreatic cancer but ultimately benign disease: a multicenter series of 344 pancreatoduodenectomies. Ann Surg Oncol 2014; 21: 3999–4006. doi:10.1245/s10434-014-3810-7
4. Norton ID, Zheng Y, Wiersema MS, et al. Neural network analysis of EUS images to differentiate between pancreatic malignancy and pancreatitis. Gastrointest Endosc 2001; 54: 625–629. doi:10.1067/mge.2001.118644
5. Das A, Nguyen CC, Li F, et al. Digital image analysis of EUS images accurately differentiates pancreatic cancer from chronic pancreatitis and normal tissue. Gastrointest Endosc 2008; 67: 861–867. doi:10.1016/j.gie.2007.08.036

6. Sâftoiu A, Vilmann P, Gorunescu F, et al. Neural network analysis of dynamic sequences of EUS elastography used for the differential diagnosis of chronic pancreatitis and pancreatic cancer. Gastrointest Endosc 2008; 68: 1086–1094. doi:10.1016/j.gie.2008.04.031

7. Sâftoiu A, Vilmann P, Dietrich CF, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). Gastrointest Endosc 2015; 82: 59–69. doi:10.1016/j.gie.2014.11.040

8. Zhu M, Xu C, Yu J, et al. Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. PloS One 2013; 8: e63820. doi:10.1371/journal.pone.0063820

9. Zhang M-M, Yang H, Jin Z-D, et al. Differential diagnosis of pancreatic cancer from normal tissue with digital imaging processing and pattern recognition based on a support vector machine of EUS images. Gastrointest Endosc 2010; 72: 978–985. doi:10.1016/j.gie.2010.06.042

10. Ozkan M, Cakiroglu M, Kocaman O, et al. Age-based computer-aided diagnosis approach for pancreatic cancer on endoscopic ultrasound images. Endosc Ultrasound 2016; 5: 101–107. doi:10.4103/2303-9027.180473

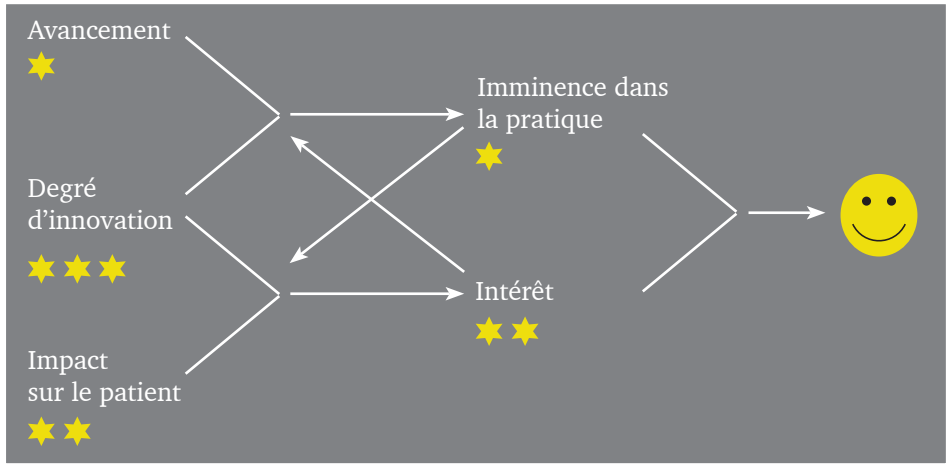
11. Tonozuka R, Itoi T, Nagata N, et al. Deep learning analysis for the detection of pancreatic cancer on endosonographic images: a pilot study. J Hepato-Biliary-Pancreat Sci 2020. doi:10.1002/jhbp.825

12. Marya NB, Powers PD, Chari ST, et al. Utilisation of artificial intelligence for the development of an EUS-convolutional neural network model trained to enhance the diagnosis of autoimmune pancreatitis. Gut 2020. doi:10.1136/gutjnl-2020-322821

13. Kuwahara T, Hara K, Mizuno N, et al. Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas. Clin Transl Gastroenterol 2019; 10: 1–8. doi:10.14309/ctg.0000000000000045

14. Zhang J, Zhu L, Yao L, et al. Deep learning-based pancreas segmentation and station recognition system in EUS: development and validation of a useful training tool (with video). Gastrointest Endosc 2020; 92: 874-885.e3. doi:10.1016/j.gie.2020.04.071

A la frontière des autres disciplines : Nouvelles technologies



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Lucille Quénéhervé
Service d'hépato-gastroentérologie, CHRU Brest La Cavale Blanche

L'APPORT DE L'INTELLIGENCE artificielle (IA) en endoscopie digestive est désormais si évident qu'il pourrait occulter les avancées des autres développements technologiques. Pourtant les techniques d'imagerie non basées sur la vidéo-endoscopie, en lumière blanche ou en chromoendoscopie virtuelle, bénéficient aussi du développement de solutions d'IA.

Certains auteurs considèrent que l'on peut distinguer globalement deux branches principales à l'IA dans le domaine médical : une branche virtuelle et une branche physique[1]. La branche virtuelle comprendrait l'imagerie médicale, l'aide au diagnostic, l'évaluation du pronostic, etc. alors que les différents types de robots appartiendraient à la branche physique. Les applications en technologie endoscopique innovante sont donc multiples et en constant développement.

Techniques d'imagerie innovantes et IA

Endomicroscopie confocale

L'endomicroscopie confocale (EMC) est une technique basée sur l'illumination du tissu par un laser permettant la détection de la fluorescence réfléchie par le tissu à travers un pinhole. Elle nécessite en général l'injection intraveineuse d'un produit de contraste fluorescent à la longueur d'onde d'intérêt. L'EMC en endoscopie digestive n'est plus une technologie nouvelle en soi mais cette technique qui permet de réaliser de véritables biopsies optiques en temps réel pourrait connaître un renouveau grâce à une lecture automatisée des images. En effet le développement de l'EMC est notamment limité par l'expertise nécessaire pour lire des images très différentes de celles qui sont obtenues en endoscopie en lumière blanche. Quelques études ont été publiées sur l'IA appliquée à l'EMC.

Tube digestif haut

L'endobrachyœsophage (EBO) en tant que pathologie pré-néoplasique de la muqueuse œsophagienne est un modèle de choix pour le développement de techniques d'imagerie. Or, pour construire et entraîner un système de diagnostic assisté par ordinateur fiable, un grand nombre de données étiquetées généralement « à la main » est souvent nécessaire. En effet ces données préanalysées permettent de sélectionner les sous-ensembles de caractéristiques optimales et de former des modèles de classification robustes par un apprentissage supervisé. L'accumulation d'un grand ensemble d'images d'EMC étiquetées est cependant longue et coûteuse alors qu'il est relativement facile de collecter un nombre important d'images non étiquetées lors de chaque session d'EMC. Une équipe américaine a donc proposé d'améliorer la classification des images d'EMC d'EBO en utilisant des images non étiquetées par le biais d'un apprentissage semi-supervisé [2]. Un réseau neuronal profond semi-supervisé basé sur les auto-codeurs convolutifs a donc été développé pour améliorer la classification de l'EBO. Cette équipe postule que si cette technique fonctionne pour l'analyse d'images d'EMC elle est susceptible de fonctionner pour des images d'OCT par exemple.

Tube digestif bas

L'amélioration du dépistage et de la prise en charge du cancer colorectal est également un sujet porteur dans le domaine des nouvelles technologies endoscopiques. Dans une étude roumaine, des images d'EMC de muqueuse colique saine et d'adénocarcinome colique ont été analysées par un système d'IA [3]. Le diagnostic assisté par ordinateur a été utilisé pour traiter les images et le modèle a été entraîné à l'aide d'un

réseau neuronal à deux couches pour diagnostiquer automatiquement les échantillons malins en fonction de sept paramètres testés.

Tomographie par cohérence optique

La tomographie en cohérence optique (OCT) est basée sur la technique de l'interférométrie. Un laser émet une onde lumineuse qui est divisée en deux bras, le bras échantillon dirigé vers le tissu biologique et le bras de référence dirigé vers un miroir. Lorsque les ondes lumineuses du bras échantillon et du bras de référence sont réfléchies vers le capteur, elles sont combinées ce qui crée des interférences, dont le profil est ensuite analysé pour déterminer la composition des milieux traversés selon le principe du délai d'écho. L'OCT conventionnel scanne à travers la surface du tissu un point à chaque profondeur, formant un scan longitudinal appelé scan d'amplitude ou A-scan, comme en échographie. En assemblant les différents A-scans réalisés lors du balayage du faisceau, on obtient une reconstruction en 2 dimensions appelée tomogramme ou B-scan. En empilant les B-scan, on obtient une image en 3 dimensions.

Tube digestif haut

L'endomicroscopie volumétrique par laser (*Volumetric Laser Endomicroscopy* [VLE]) est une technologie dérivée de l'OCT dans laquelle le laser, émis à l'extrémité d'un cathéter, est retiré automatiquement de façon longitudinale en effectuant une rotation. Le cathéter est inclus dans un ballon qui, passé par le canal opérateur de l'endoscope et gonflé au niveau de la jonction œso-gastrique, permet un balayage circonférentiel des couches de la paroi œsophagienne à une résolution proche du micron. Les appareils de VLE récents permettent également de réaliser un marquage au laser des zones d'intérêt ce qui permet de réaliser leur exérèse au cours du même geste. Cet appareil pourrait améliorer la détection des néoplasies sur EBO, cependant l'interprétation des images de VLE est longue et complexe, en raison de la grande quantité d'information visuelle en niveaux de gris.

Une équipe néerlandaise a démontré dans une première étude que des algorithmes entraînés à l'aide de paramètres inspirés de la clinique pouvaient permettre la détection assistée par ordinateur de lésions dysplasiques au sein d'EBO[4]. Cette étude se basait sur des images de VLE réalisées *ex vivo* en scannant des spécimens de résection qui avaient été corrélées à de l'histologie. Cependant, une modalité de détection assistée par ordinateur qui analyserait plusieurs coupes voisines pourrait améliorer la détection des néoplasies de l'EBO par rapport aux analyses automatisées d'une seule image. Cette même équipe a donc par la suite démontré la faisabilité de l'extraction automatique des données suivie d'une détection assistée par ordinateur en analysant

plusieurs coupes voisines, pour la détection de la néoplasie sur EBO [5]. Ces études ont cependant porté uniquement sur des banques d'images obtenues *ex vivo*.

Une autre étude a été réalisée[6], cette fois à partir d'images de VLE collectées *in vivo*, et a démontré l'efficacité d'un encodage par dimension principale (qui permet notamment de diviser un problème en plusieurs zones distinctes, ici en analysant chaque colonne (A-line) de l'image séparément) suivi d'une analyse utilisant le machine learning. Ces données encourageantes nécessitent bien sûr une validation clinique en temps réel.

Parallèlement, une équipe états-unienne a développé une technologie permettant d'améliorer les images produites en VLE en temps réel et surtout de repérer les zones d'intérêt grâce à l'IA [7] (*figure 1*). Ce repérage permet à l'endoscopiste de marquer ces lésions au laser et d'en réaliser l'exérèse, le tout au cours de la même endoscopie. Une étude prospective est en cours aux USA pour valider cette technique.

Tube digestif bas

Une équipe américaine a conçu un réseau neuronal convolutif dans le but de reconnaître le pattern muqueux du colon humain dans des images obtenues à l'aide d'une sonde d'OCT. Le réseau entraîné sur des images *ex vivo* de spécimens chirurgicaux a distingué avec succès les tissus colorectaux normaux et néoplasiques avec une aire sous la courbe ROC de 0,998 [8]. Les auteurs suggèrent que l'OCT avec reconnaissance de pattern pourrait donner un diagnostic assisté par ordinateur précis en temps réel au cours d'une coloscopie.

Imagerie hyperspectrale

L'imagerie hyperspectrale collecte et analyse des bandes étroites et continues de longueurs d'onde à travers le spectre électromagnétique, ne se limitant donc pas au spectre du visible. Cet outil pourrait être utilisé pour déterminer les marges tumorales pendant une chirurgie oncologique [9]. Au sein d'images de pièces de résections œsophagiennes, différentes zones de caractéristiques différentes ont ainsi été classifiées en zones cancéreuses ou non (*figure 2*). L'objectif des chercheurs est de développer une modalité per opératoire pour guider le chirurgien au cours de son geste. On peut imaginer que dans le futur de tels outils pourraient assister les opérateurs à déterminer les marges des lésions avant dissection sous muqueuse.

Endoscopes robotisés et IA

Plusieurs équipes développent des robots pour l'endoscopie souple dans le but de : 1) progresser dans le tube digestif en limitant les effets indésirables de l'examen ou

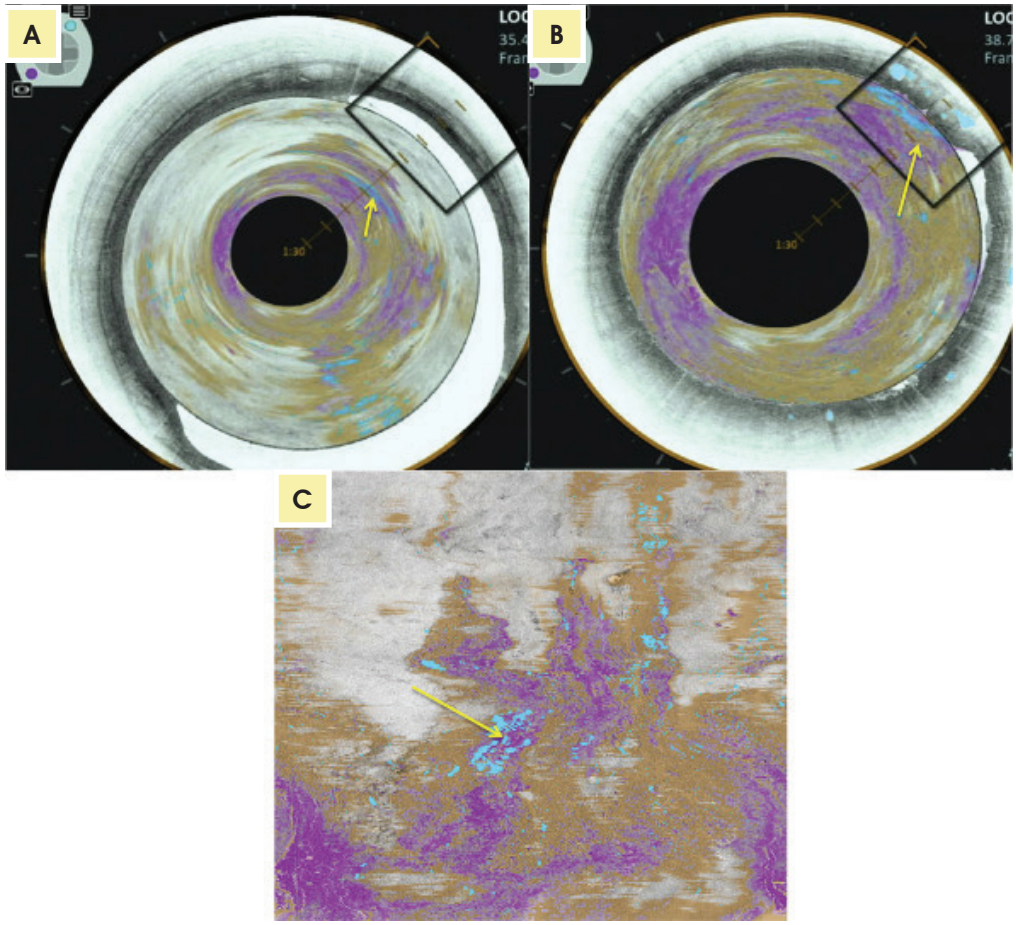


FIGURE 1. Endomicroscopie volumétrique par laser d'un endobrachyoesophage *in vivo* avec traitement de l'image par intelligence artificielle. Vue luminale en face d'une zone de chevauchement (flèche jaune) entre les trois caractéristiques de la dysplasie (orange : absence de stratification, bleu : structures glandulaires, rose : surface hyper-réfléchissante). (A) Vue de l'œsophage proximal. (B) Vue plus proche de la zone suspecte de dysplasie. La vue de face est également représentée (C). D'après [7].

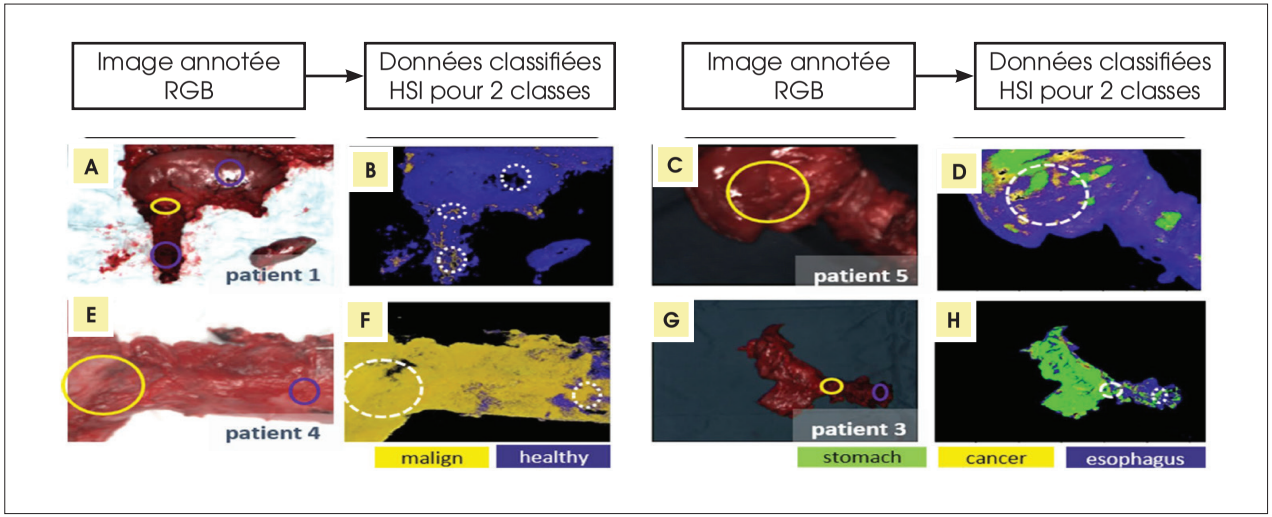


FIGURE 2. Imagerie hyperspectrale assistée par intelligence artificielle appliquée à la détermination des marges tumorales sur des pièces opératoires de cancer de l'œsophage. Approche à deux classes avec les photographies annotées (A, E) et les images hyperspectrales (HSI) classifiées (B, F) et sur la droite, l'approche à trois classes avec les photographies annotées (C, G) et les images hyperspectrales classifiées (D, H). D'après [9].

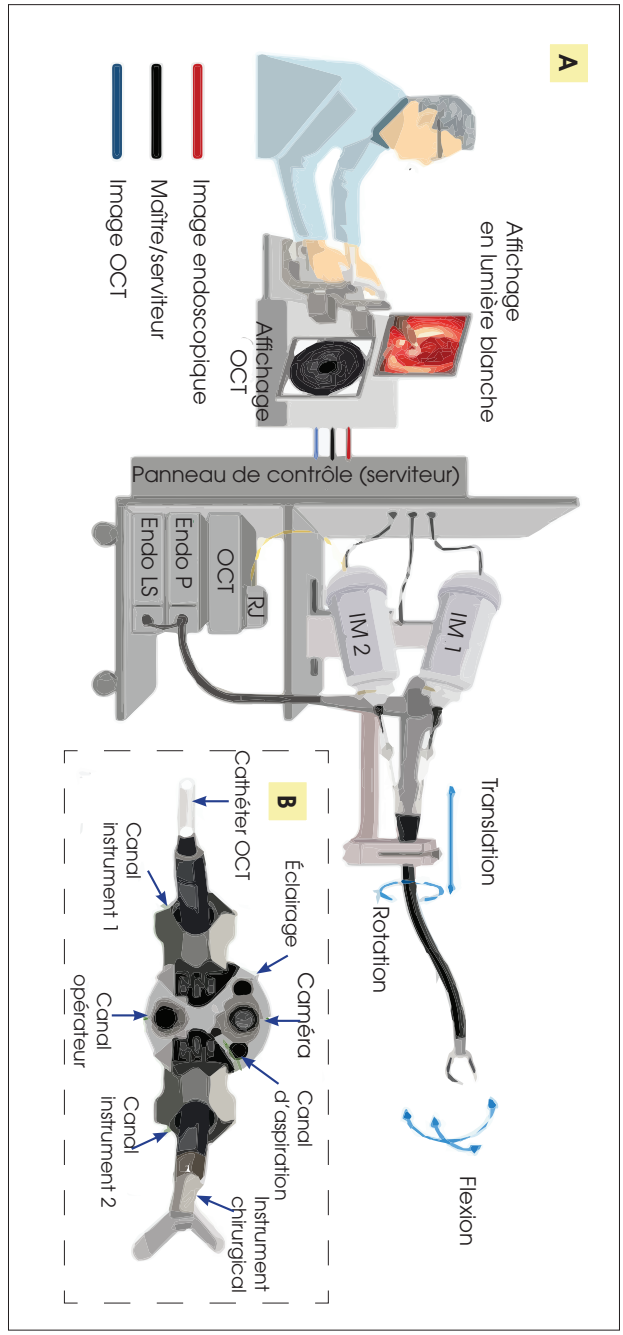


FIGURE 3. Module de tomographie par cohérence optique adapté sur un robot de type Anubiscope (Karl Storz). A) Schéma de l'endoscope interventionnel flexible robotisé avec le cathéter OCT (tomographie en cohérence optique) orientable fixé à un chariot « slave » connecté aux contrôleurs de l'utilisateur pour la téléopération du dispositif : actionneurs de l'instrument (IM1, IM2), actionnement du balayage de recul (RJ), système OCT (OCT), processeur de l'endoscope (Endo P), source de lumière de l'endoscope (Endo LS) ; B) vue frontale de l'extrémité distale de l'endoscope interventionnel flexible robotisé avec cathéter OCT orientable. D'après [13].

2) d'assister l'opérateur dans un geste technique, de résection par exemple [10–12]. Leur implémentation en pratique suppose que ces appareils soient capables de compléter le geste de l'opérateur et non pas seulement de l'égaler. Actuellement, les modèles destinés à réaliser de l'endoscopie « active » ne sont pas capables de progresser en autonomie complète et sont généralement contrôlés par l'opérateur, problème auquel l'IA pourra apporter des solutions grâce à l'interprétation des images en temps réel. D'autre part, les modèles permettant une triangulation n'ont pas été testés pour réséquer de grandes et difficiles lésions.

Combinant robotique et technique optique innovante, une équipe française a mis au point un système pour analyser la muqueuse colique en OCT tout en réalisant un geste [13]. Il s'agit d'une sonde d'OCT orientable qui est insérée dans l'un des 2 canaux instrumentaux d'un endoscope robotisé, l'autre canal permettant d'utiliser une pince chirurgicale par exemple (*figure 3*).

Conclusion

Les progrès de la vidéoendoscopie, *a fortiori* associés à ceux de l'IA, sont tels que les nouvelles technologies doivent aujourd'hui pouvoir résoudre des problèmes insolubles en endoscopie classique, ou dans des conditions moins invasives, afin de ne pas être abandonnées. Dans le cas de techniques d'imagerie, le recours à l'IA permet de diminuer l'expertise et le temps nécessaire à leur utilisation. En ce qui concerne les techniques robotiques, les applications de l'IA sont nombreuses, pour prédire le trajet du tube digestif ou les limites d'une lésion.

RÉFÉRENCES

1. Hamet P, Tremblay J. *Artificial intelligence in medicine.* Metabolism. 2017;69S:S36–S40.
2. Tong L, Wu H, Wang MD. CAESNet : Convolutional AutoEncoder based Semi-supervised Network for improving multiclass classification of endomicroscopic images. *J Am Med Inform Assoc.* 2019;26:1286-96
3. Ştefânescu D, Streba C, Cârtânâ ET, *et al.* Computer Aided Diagnosis for Confocal Laser Endomicroscopy in Advanced Colorectal Adenocarcinoma. *PLoS One.* 2016;11:e0154863.

4. Swager A-F, van der Sommen F, Klomp SR, *et al.* Computer-aided detection of early Barrett's neoplasia using volumetric laser endomicroscopy. *Gastrointest Endosc* 2017;86:839–846.

5. Struyvenberg MR, van der Sommen F, Swager AF, *et al.* Improved Barrett's neoplasia detection using computer-assisted multiframe analysis of volumetric laser endomicroscopy. *Dis Esophagus* 2020;33.

6. van der Putten J, Struyvenberg M, de Groof J, *et al*. Deep principal dimension encoding for the classification of early neoplasia in Barrett's Esophagus with volumetric laser endomicroscopy. *Comput Med Imaging Graph* 2020;80:101701.

7. Trindade AJ, McKinley MJ, Fan C, *et al.* Endoscopic Surveillance of Barrett's Esophagus Using Volumetric Laser Endomicroscopy With Artificial Intelligence Image Enhancement. *Gastroenterology* 2019;157:303–305.

8. Zeng Y, Xu S, Chapman WC, *et al.* Real-time colorectal cancer diagnosis using PR-OCT with deep learning. *Theranostics* 2020;10:2587–2596.

9. Maktabi M, Köhler H, Ivanova M, *et al.* Tissue classification of oncologic esophageal resectates based on hyperspectral data. *Int J Comput Assist Radiol Surg* 2019 ; 14 : 1651–61.

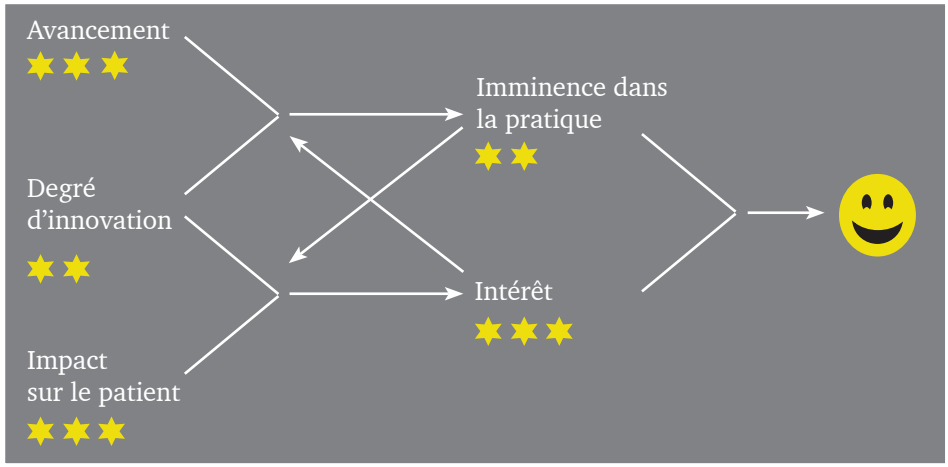
10. Li Z, Chiu PW-Y. Robotic Endoscopy. Visc Med. 2018;34:45–51.

11. Yeung BPM, Chiu PWY. Application of robotics in gastrointestinal endoscopy: A review. *World J Gastroenterol* 2016 ; 22 : 1811–25

12. Visconti TA de C, Otoch JP, Artifon EL de A. Robotic endoscopy. A review of the literature 1. *Acta Cir Bras* 2021 ; 35. : epub. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7184939/.

13. Mora OC, Zanne P, Zorn L, *et al.* Steerable OCT catheter for real-time assistance during teleoperated endoscopic treatment of colorectal cancer. *Biomed Opt Express* 2020;11:12 : 1231-43

A la frontière des autres disciplines : l'anatomie pathologique



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Céline Bossard, Raphaël Bourgade
Service d'anatomie patholoqique,
CHU de Nantes

Définition de la pathologie numérique et ouverture sur l'intelligence artificielle

Dans un contexte de démographie médicale défavorable, l'exercice et le temps médical du pathologiste sont de plus en plus contraints, notamment en raison de la complexité croissante de la prise en charge diagnostique du patient. Guidée par la médecine personnalisée, elle s'appuie sur un compte rendu anatomopathologique tenant compte de classifications diagnostiques de plus en plus précises et d'un nombre croissant de facteurs pronostiques et théranostiques. L'ensemble de ces caractéristiques font de l'anatomie pathologique (ACP) une discipline au cœur du parcours de soin, véritable pierre angulaire de la décision thérapeutique, notamment en oncologie.

Depuis les années 2000, l'émergence de scanners de lames nouvelle génération, capables de numériser entièrement (WSI [*Whole Slide Imaging*]) plusieurs centaines de lames en quelques heures, a définitivement introduit notre discipline dans l'ère du numérique. De fait, la pathologie numérique, également appelée microscopie virtuelle, nouvel outil pédagogique attractif pour les plus jeunes, va rendre possible le développement et le déploiement de l'intelligence artificielle (IA) en ACP, discipline qui y consacre d'ores et déjà des pans entiers de sa recherche, comme en témoigne le nombre exponentiel de publications scientifiques dans le domaine. L'ensemble de ces développements devrait conduire à une véritable révolution de la pratique quotidienne du pathologiste.

En effet, les réseaux de neurones artificiels dits à convolution (CNN), directement inspirés du cortex visuel des animaux, permettent l'extraction et l'analyse d'un nombre considérable d'informations contenues dans cette lame virtuelle (WSI), dont seule une infime partie est réellement exploitée par l'œil du pathologiste. Ces données massives contribuent non seulement à l'élaboration d'algorithmes diagnostiques (reconnaissance des tissus, des cellules...), mais également prédictifs aussi bien du statut mutationnel d'une tumeur ou de sa signature moléculaire, que de son évolution.

Caractéristiques et contraintes techniques des WSI

Si l'avenir semble prometteur, les applications de l'IA en ACP n'en sont finalement qu'à leurs débuts en raison de différentes contraintes et limites actuelles.

De par leur très haut niveau de résolution lié notamment à leur format pyramidal (*figure 1*), les WSI demeurent des données complexes à analyser. Une lame numérisée au grossissement x400 (résolution d'environ 0,2 µm/pixel), peut mesurer plus de 200 000 x 100 000 pixels (> 20 Gpx) et peser plusieurs gigaoctets dans un format non compressé [1]. Ces données, trop volumineuses pour être traitées telles quelles, nécessitent la réalisation d'une étape de prétraitement, consistant notamment en leur découpage en tuiles carrées mesurant traditionnellement entre 128 et 1 024 pixels de côté (*figure 2*) [2]. Ce tuilage permet en outre la suppression de tuiles vides, optimisant ainsi le volume du dataset.

Durant l'étape d'apprentissage supervisé, les CNN sont entraînés sur des images labellisées par des pathologistes experts et donc considérées comme « cas de références » (*ground truth*). Malheureusement, les banques d'images annotées demeurent encore trop rares et sont généralement de taille restreinte. En effet, en raison du coût nécessaire à sa mise en place, tenant compte notamment d'espaces de stockage informatique de très grande capacité, un faible nombre de laboratoires d'ACP

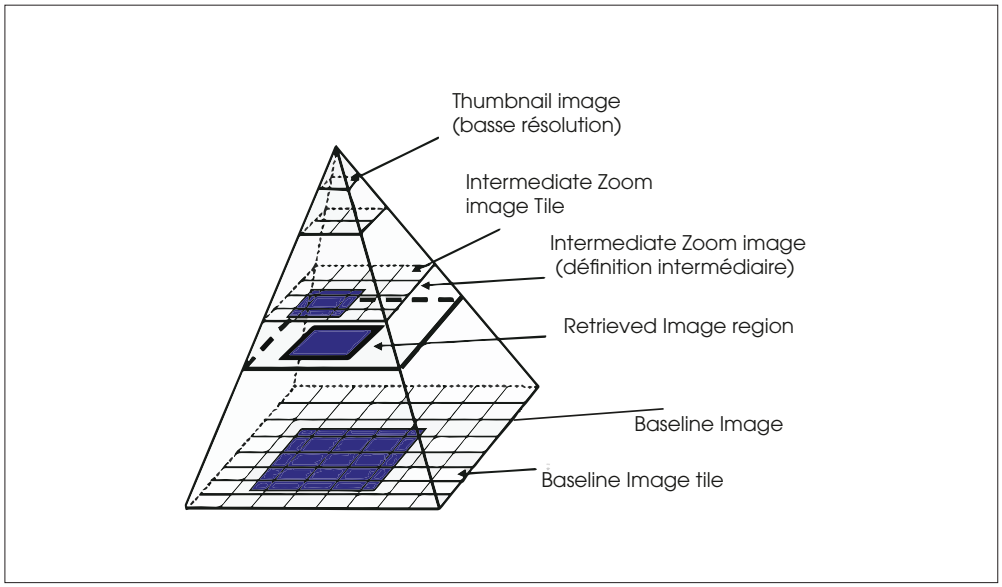


FIGURE 1. Représentation de l'architecture pyramidale d'une WSI (*Whole Slide Imaging*), adaptée de Lajara *et al.*[1]

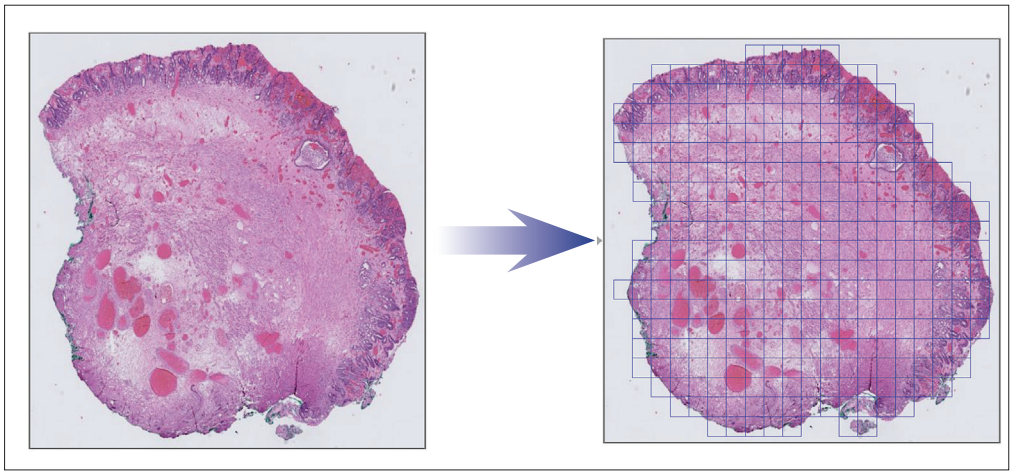


FIGURE 2. Tuilage d'une lame histopathologique, adaptée de la figure de Sali *et al.*[2].

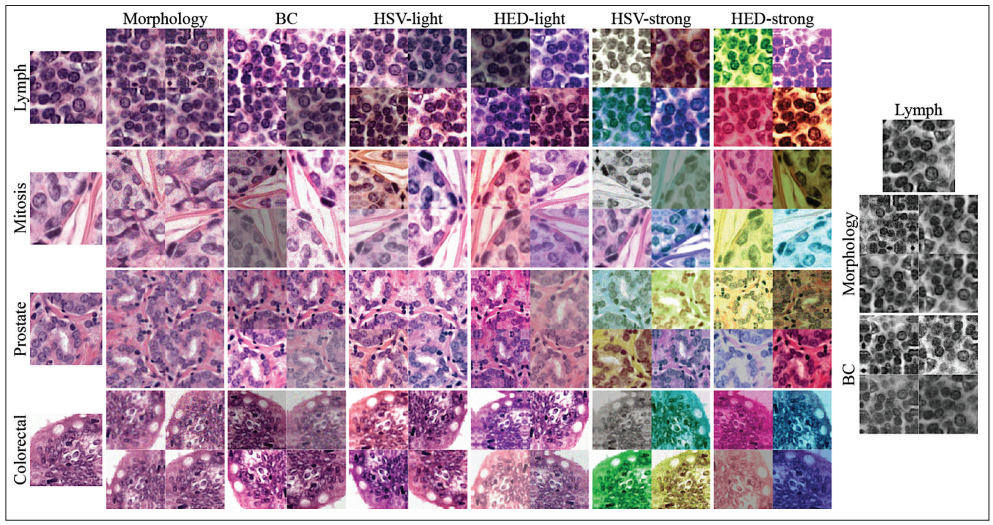


FIGURE 3. Résumé des principales techniques de data augmentation appliquées sur 4 tuiles. BC : luminosité et contraste. HSV : saturation et teinte. HED : Hématoxyline-éosine-DAB. Adaptée de Tellez *et al.*[3].

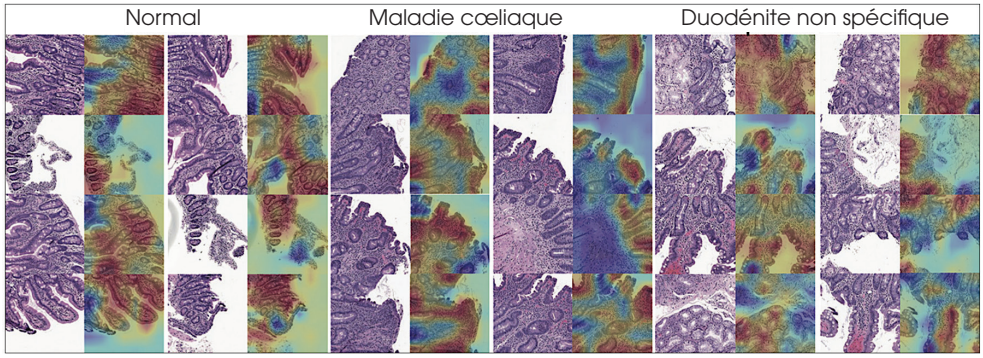


FIGURE 3. Heatmaps illustrant l'importance de chaque tuile lors de la classification de la WSI en « Normal », « Maladie cœliaque » et « Duodénite non spécifique ». Plus une tuile tend vers le rouge, plus son rôle a un impact sur la prédiction. Adaptée de Wei *et al.*[4].

sont à ce jour équipés en solutions de numérisation WSI haut débit, intégrées dans le workflow habituel du prélèvement. Or, la performance et la fiabilité des algorithmes, qualités indispensables à leurs validations et exploitations en routine clinique, augmentent avec le nombre et la qualité des images disponibles. Pour pallier cette limite actuelle et prévenir le surapprentissage des algorithmes, des techniques de data augmentation sont donc systématiquement appliquées, multipliant ainsi virtuellement le dataset d'entraînement. Ces techniques consistent généralement en la rotation, l'inversion, le grossissement ou encore la variation de teinte, saturation et luminosité des couleurs (*figure 3*) [3].

Enfin, l'étape de validation du modèle, reposant sur la mesure d'une métrique, témoin direct de sa performance, doit porter sur un dataset externe aussi varié que possible, afin de prendre en compte l'hétérogénéité de la phase pré-analytique (fixation, imprégnation, coupe, coloration, montage...) inhérente à chaque laboratoire d'ACP. Cette limite actuelle devrait vraisemblablement être surmontée dans un futur proche, grâce à l'automatisation du workflow technique (à ce jour manuel) du prélèvement déjà fixé et mis en cassette.

Applications de l'intelligence artificielle en pathologie digestive

Classification diagnostique

Pierre angulaire de l'existence du pathologiste, les classifications diagnostiques basées sur la reconnaissance tissulaire et/ou cellulaire sont au cœur des travaux de recherche en intelligence artificielle, notamment en oncologie digestive comme en témoigne une revue récente [4]. À titre d'exemple, O. Iizuka et son équipe se sont intéressés à la classification de tumeurs épithéliales gastriques et coliques à partir de biopsies chirurgicales. Ils ont ainsi mis au point un CNN entraîné sur plus de 4000 lames, à classer des lésions selon les labels « adénocarcinome », « adénome » et « lésion non néoplasique ». Les courbes ROC correspondantes ont été mesurées sur une cohorte de validation externe issue d'un autre centre, à > 0,972 pour l'adénocarcinome colique et > 0,966 pour l'adénocarcinome gastrique. De plus, afin d'évaluer au mieux la généralisabilité de leur modèle, mais également d'en définir les éventuelles limites, une seconde évaluation a été réalisée sur une cohorte du TCGA (*The Cancer Genome Atlas*), faite uniquement de pièces d'exérèses chirurgicales. Les courbes ROC sont alors mesurées à 0,982 (0,968-0,991 IC 95%) pour l'adénocarcinome colique et 0,924 (0,887-0,952 IC 95%) pour l'adénocarcinome gastrique, preuve de la capacité du modèle à généraliser sur des images de taille et donc de complexité différentes, sans y avoir été explicitement entraîné.

L'autre attente des pathologistes vis-à-vis de l'apport de l'IA dans leur pratique quotidienne, outre la performance et la fiabilité des algorithmes, est le temps médical nécessaire au diagnostic. Vingt-trois pathologistes dans cette même étude, ont ainsi été évalués sur la lecture de 45 lames d'adénocarcinome gastrique, avec un délai maximum de 30 secondes par cas. Avec une précision moyenne de 85,9 % de réponses correctes, la performance de l'intelligence humaine reste bien inférieure au modèle entraîné, à qui il faut entre 5 et 30 secondes pour établir un diagnostic avec une précision > 95 %.

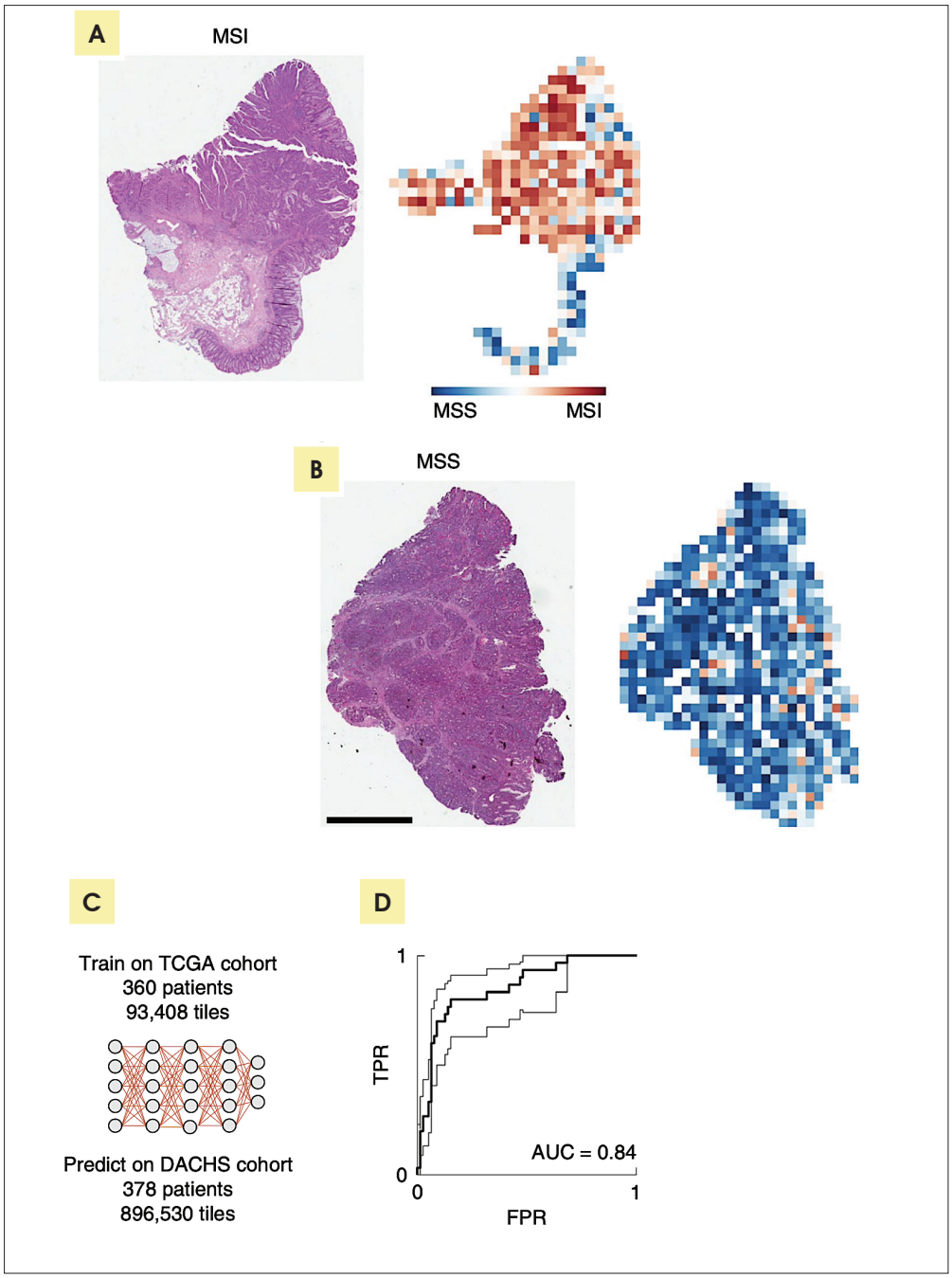
D'autres travaux ont porté sur l'œsophage de Barrett et la reconnaissance de lésions métaplasiques +/- dysplasiques ou adénocarcinomateuses issues de prélèvements biopsiques, sur un dataset constitué de plus de 500 lames, avec une précision moyenne de 0,83 (0,80-0,86 IC 95 %) [5]. A. Kiani, quant à lui, a étudié l'impact d'un modèle entraîné, sur la performance diagnostique de 11 pathologistes à distinguer des cholangiocarcinomes de carcinomes hépatocellulaires, prouvant la complémentarité d'une telle association [6].

Les classifications diagnostiques en deep learning sont également possibles en pathologie non tumorale. Des équipes ont entraîné des algorithmes à distinguer des lésions de duodénite non spécifique, de lésions de duodénite auto-immune liées à une maladie coeliaque sur des biopsies duodénales avec une précision > 95% [7,8]. Les heatmaps, représentations illustrant l'importance de chaque tuile dans la prédiction finale, sont représentées dans la *figure 4*.

Néanmoins, si ces classifications reposent essentiellement sur l'analyse globale du tissu, elles peuvent également se baser sur l'analyse cellulaire. C'est ce qu'a démontré Y. H. Chang en 2017 en corrélant des WSI de tumeurs du pancréas colorées à l'HES (lumière blanche) à leurs homologues colorées au DAPI (lumière fluorescente). Grâce à des techniques combinées de segmentation en machine learning et de classification en deep learning, il a pu mettre au point un modèle capable de distinguer les noyaux tumoraux, des noyaux non tumoraux avec une précision > 90 %.

Substitution aux techniques complémentaires : prédiction des anomalies et signatures moléculaires

Bien que très étudiée, la classification diagnostique à partir de l'analyse d'une WSI n'est cependant pas la seule application offerte par le deep learning. Des équipes ont en effet développé des algorithmes capables de prédire le statut mutationnel d'une tumeur ou son profil d'expression génique par la simple analyse d'une WSI colorée à l'HES. Cette approche, très prometteuse, permettrait de s'affranchir des techniques complémentaires, lourdes et chronophages, telles que l'immunohisto-


A
MSI
MSS
MSI
B
MSS
C
Train on TCGA cohort
360 patients
93,408 tiles
Predict on DACHS cohort
378 patients
896,530 tiles
D
1
TPR
0
0
1
FPR
AUC = 0.84

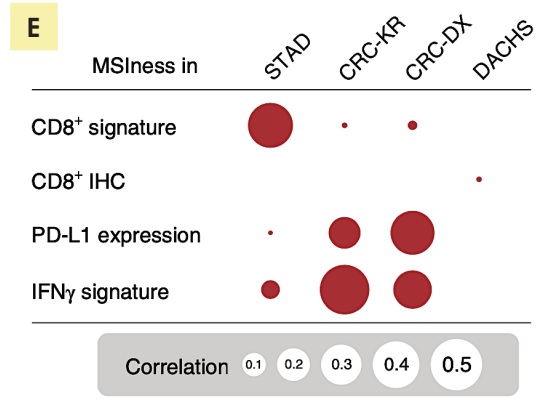


FIGURE 5. A) et B), Heatmaps cartographiant la répartition des tuiles prédites MSI (instabilité microsatellitaire) ou MSS (stabilité microsatellitaire) au sein d'un cancer colorectal ; C) le modèle a été entraîné sur une cohorte issue du TCGA, constituée de 360 patients atteints d'un cancer colorectal, puis déployé sur une cohorte indépendante de 378 patients ; D) courbe ROC mesurée à 0,84 (0,73-0,91 IC 95 %) sur la cohorte de validation externe. TPR : taux de vrais positifs (sensibilité), FPR : taux de faux positifs (1-spécificité) ; E) coefficient de Pearson mesurant la corrélation entre le ratio de tuiles prédites MSI (MSIness) et l'expression de marqueurs immunohistochimiques et transcriptomiques dans l'adénocarcinome gastrique (STAD) et le cancer colorectal (CRC-KR, CRC-DX & cohorte DACHS). Aadaptée de Kather *et al.*[11].

chimie ou encore la biologie moléculaire (NGS [*Next Generation Sequencing*], WES [*Whole Exome Sequencing*], etc.). Ces dernières étant indispensables à l'identification de marqueurs diagnostiques, pronostiques ou théranostiques, essentiels à la prise en charge thérapeutique du patient, à l'ère de la médecine personnalisée. De récentes publications ont notamment porté sur la détermination de la classification moléculaire consensuelle (CMS) du cancer colorectal, avec une courbe ROC évaluée à 0,84 [9]. De même, J. N. Kather et son équipe ont démontré que l'analyse en deep learning d'une WSI colorée à l'HES, pouvait être prédictive de certaines mutations (TP53, KRAS, BRAF…) et de signatures moléculaires dont la traduction phénotypique a été déterminée par l'IA [10]. L'utilisation de ces modèles en routine diagnostique permettrait ainsi la réalisation d'un screening à moindre coût, informant ainsi sur la pertinence ou non de réaliser une recherche moléculaire *ad hoc*.

J. N. Kather s'est également intéressé à la prédiction du statut microsatellitaire dans des adénocarcinomes gastriques et colorectaux avec une courbe ROC mesurée à 0,84 (0,72 - 0,92 IC 95 %) [11]. Afin de conforter leurs résultats, ils ont comparé la proportion de tuiles prédites MSI à l'expression de marqueurs connus pour être associés à ce statut (*figure 5*). La visualisation spatiale de l'hétérogénéité moléculaire mise en évidence par les heatmaps, participera probablement à une meilleure compréhension des mécanismes de carcinogénèse.

Prédiction de l'évolution de la maladie

Bien qu'aujourd'hui l'ensemble des marqueurs morphologiques (pTNM, mitoses, *budding*, différenciation, grade et emboles vasculo-nerveux), phénotypiques et moléculaires est indispensable à la détermination du pronostic, de nombreux travaux semblent montrer que l'intelligence artificielle permettrait de s'en affranchir. O-J. Skrede a ainsi mis au point un modèle pronostique évaluant la survie du patient, basé uniquement sur l'analyse de WSI d'adénocarcinomes colorectaux réséqués chirurgicalement [12]. Ce modèle, tenant compte d'une dizaine de CNN entraînés sur plus de 12 000 000 de tuiles issues des lames de 828 patients, a permis de différencier deux groupes de patients selon leur pronostic avec un hazard ratio à 3,84 (2,72-5,43 ; $p < 0,0001$). Ce dernier est finalement mesuré à 3,04 (2,07-4,47 ; $p < 0,0001$) après ajustement sur les facteurs pronostiques connus (pT, pN et emboles veino-lymphatiques), en analyse univariée.

D. Bychkov a réalisé un travail similaire, comparant trois approches différentes, destinées à différencier les patients de bon ou de mauvais pronostic à partir de l'analyse en deep learning de WSI de cancers colorectaux [13]. La première ap-

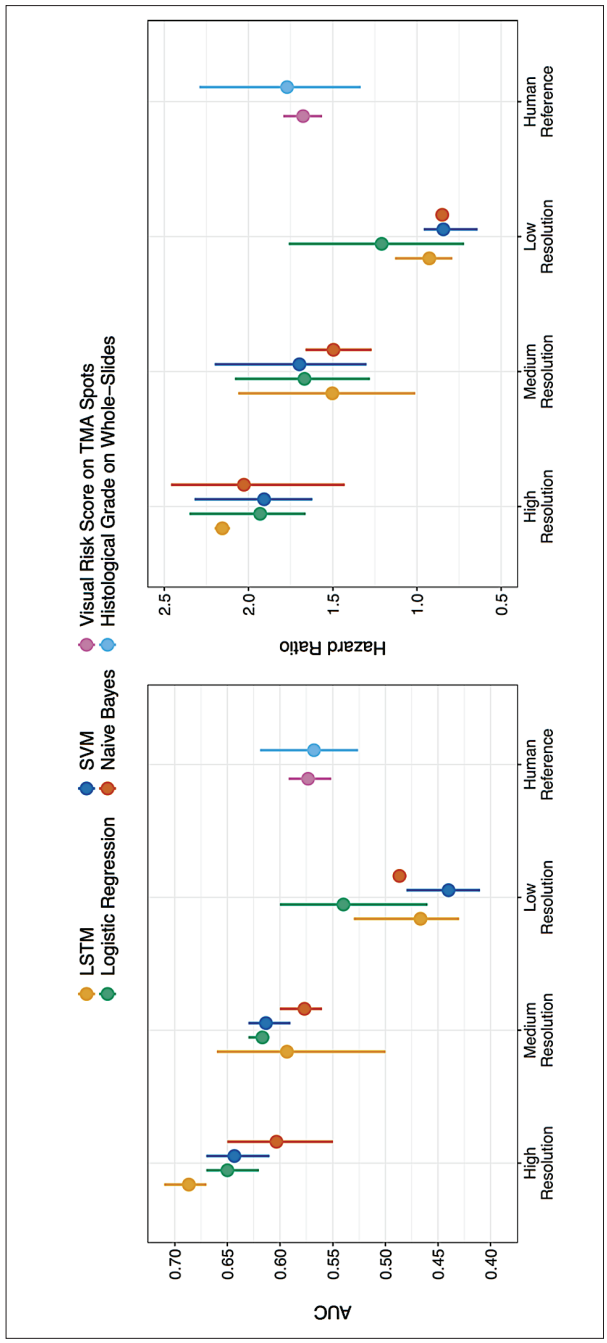


FIGURE 6. Représentation des mesures de l'aire sous courbe (AUC) et de l'hazard ratio, comparant la performance prédictive de 4 modèles automatiques avec celle de l'évaluation humaine. La barre verticale indique les variations possibles, tandis que le point symbolise la valeur moyenne. Ces mesures ont été réalisées sur des images de tissue microarrays à différentes résolutions (« high » correspond à une résolution de 0,22 μm/px, tandis que « medium » et « low » correspondent à des images réduites d'un facteur 4 et 16 respectivement. Adaptée de Bychkov *et al.* [13]

proche reposait sur l'estimation du grade histologique et de la classification de Duke par trois pathologistes expérimentés. La seconde approche, sur l'utilisation d'algorithmes traditionnels de machine learning, tels que la régression logistique (*logistic Regression*), la classification naïve bayésienne (*Naive Bayes*) ou encore les machines à vecteur de support (SVM). Enfin, la troisième approche reposait sur l'utilisation du deep learning avec un réseau de neurones récurrents (LSTM [Long Short-Term Memory]) entraîné sur des patterns extraits par un réseau de neurones dédié (VGG-16). Les performances de ces trois méthodes, mesurées à différentes résolutions et illustrées dans la *figure 6*, prouvent la supériorité des approches automatisées, lorsque la résolution est suffisante.

Conclusion et perspectives

La pathologie numérique exploitée par l'IA ouvre désormais un champ d'investigations sans limite pour une prise en charge optimisée et personnalisée du patient. L'intégration et le traitement en deep learning de l'ensemble des données morphologiques et phénotypiques, aux données cliniques, biologiques, radiologiques et moléculaires, va considérablement améliorer les approches diagnostiques, pronostiques et théranostiques, identifiant de nouveaux marqueurs et créant ainsi de nouvelles classifications.

Cependant, le développement de ces nouveaux outils génère immanquablement des questions éthiques. La fiabilité de ceux-ci sera un prérequis obligatoire et conditionnera la confiance qui leur sera accordée, par le pathologiste mais également par le patient. Cette fiabilité dépendra directement de la qualité des données ayant servi à leur entraînement, mais également à celle de leur validation avant leur déploiement en routine clinique. Leurs fondements ainsi que la base de leur raisonnement devront être compris et critiquables par le pathologiste, afin que ce dernier puisse construire l'information qu'il transmettra au clinicien, et in fine au patient.

Cette nouvelle ère de la pathologie digitale va donc profondément améliorer le travail du pathologiste, non pas remplacé, mais plus vraisemblablement assisté par l'IA. Ce dernier devra être un acteur central de cette transformation numérique, pour bénéficier ainsi des performances « augmentées« d'un troisième œil !

RÉFÉRENCES

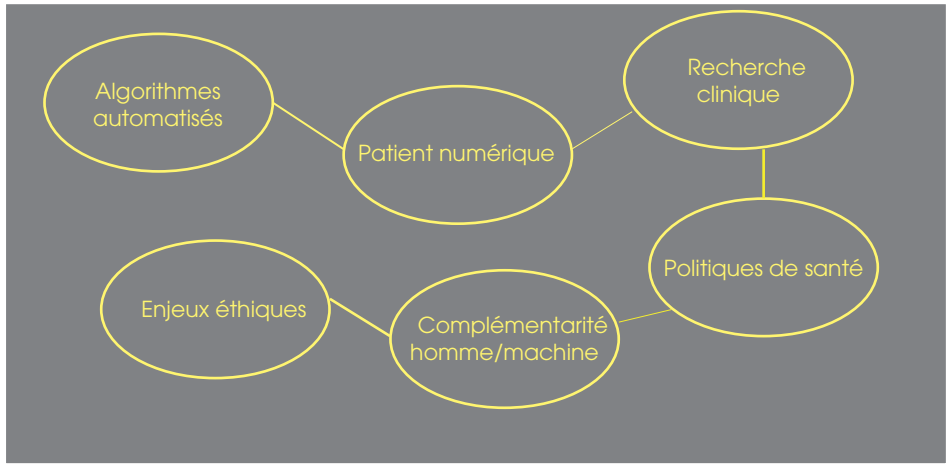
1. Lajara N, Espinosa-Aranda JL, Deniz O et al. Optimum web viewer application for DICOM whole slide image visualization in anatomical pathology. Computer Methods and Programs in Biomedicine 2019 ; 179 : 104983.
2. Sali R, Moradinasab N, Guleria S et al. Deep Learning for Whole-Slide Tissue Histopathology Classification: A Comparative Study in the Identification of Dysplastic and Non-Dysplastic Barrett's Esophagus. Journal of Personalized Medicine 2020 ; 10 : 141.
3. Tellez D, Litjens G, Bandi P et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis 2019 ; 58 : 101544.
4. Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 2020 : epub. doi : gutjnl-2020-322880.
5. Tomita N, Abdollahi B, Wei J et al. Attention-Based Deep Neural Networks for Detection of Cancerous and Precancerous Esophagus Tissue on Histopathological Slides. JAMA Netw Open 2019 ; 2 : e1914645.
6. Kiani A, Uyumazturk B, Rajpurkar P et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. npj Digit Med. déc 2020 ; 3 : 23.
7. Wei J, Wei J, Jackson C et al. Automated detection of celiac disease on duodenal biopsy slides: A deep learning approach. J Pathol Inform. 2019;10 : 7.
8. Srivastava A, Sengupta S, Kang S-J et al. Deep Learning for Detecting Diseases in Gastrointestinal Biopsy Images. In: 2019 Systems and Information Engineering Design Symposium (SIEDS). Charlottesville (Etats-Unis) : IEEE; 2019
9. Sirinukunwattana K, Domingo E, Richman SD et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut. 20 juill 2020;gutjnl-2019-319866.
10. 0. Kather JN, Heij LR, Grabsch HI et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer. août 2020;1 : 789-99

11. Kather JN, Pearson AT, Halama N et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. juill 2019;25(7):1054-6.

12. Skrede O-J, De Raedt S, Kleppe A et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. The Lancet. févr 2020;395(10221):350-60.

13. Bychkov D, Linder N, Turkki R et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep. déc 2018;8(1):3395.

Quelques perspectives dans le domaine de l'IA et impacts attendus



Représentation graphique d'après les indications de l'auteur à valeur purement indicative.

par Robert Benamouzig
Service de gastroentérologie et d'oncologie digestive,
Hôpital Avicenne AP-HP,
université Paris Nord- la Sorbonne

L'intelligence artificielle a déjà démontré son intérêt pratique pour la prédiction du diagnostic de lésions dermatologiques comme le mélanome ou la reconnaissance de certaines pathologies rétiniennes. Un système d'intelligence artificielle avec robot interactif a été déployé depuis cette année dans certains services d'urgences en Israël, aux USA et en Grande Bretagne pour aider au tri et fluidifier les flux de patients. L'intelligence artificielle est en train de se développer pour le traitement des données en imagerie radiologique et plus récemment en endoscopie et en anatomopathologie. Les précédents chapitres de cet ouvrage francophone pionnier sur l'intelligence artificielle dans notre discipline vous ont sans doute convaincu de l'apport majeur attendu de cette approche, bien sûr en endoscopie digestive mais aussi dans d'autres domaines. L'aide à la détection et à la caractérisation de lésions digestives encore débutantes sera bientôt incontournable.

Le champ de la santé est considéré aujourd'hui comme particulièrement concerné par le développement des algorithmes automatisés d'analyse des données digitales disponibles, c'est-à-dire ce qu'on nomme Intelligence artificielle. Ceci est lié à deux paramètres : l'importance des enjeux économiques et sociétaux dans ce domaine et la montée en puissance de la digitalisation des données de santé au cours des 15 dernières années. Cette digitalisation, qui était ainsi estimée à moins de 15%, vers le début des années 2000, est aujourd'hui supérieure à 95%. Une analyse récente du ministère de l'Économie a identifié l'intelligence artificielle en santé comme un enjeu économique stratégique pour la France avec un tissu économique déjà constitué autour d'environ 150 entreprises pionnières de taille variable [1]. Cet enjeu est l'objet d'une attention particulière avec des investissements publics comme ceux en cours pour la mise en place de la plateforme des données de santé dite PDS ou HDH pour « Health Data Hub » [2]. La PDS été créée pour faciliter le partage des données de santé issues de sources très variées afin de favoriser la recherche par un arrêté en date du 29 novembre 2019. Sa création suit les préconisations du rapport du député Cédric Villani intitulé « *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne* » de mars 2018. L'évolution du statut de cette structure, d'un groupement d'intérêt public avec une gouvernance complexe freinant son action vers celui de société par action simplifiée où l'État resterait majoritaire, est en cours de discussion au niveau parlementaire. Le partenariat privilégié de la PDS avec l'acteur économique privé majeur du secteur qui est Microsoft Heath implique l'exportation de données individuelles hors Union Européenne ce qui fait l'objet d'un débat juridique et de souveraineté. Par ailleurs, le développement de cette approche nécessitera aussi de maitriser les enjeux techniques d'interopérabilité des langages entre les très nombreux outils informatiques impliqués ainsi que ceux autour de la pérennité des données dans le temps.

L'intelligence artificielle permettra d'améliorer le diagnostic des pathologies avec de nouvelles classifications attendues reposant sur des données classiques « revues » notamment cliniques, radiologiques, biologiques, anatomopathologiques et moléculaires mais aussi sur des données nouvelles prenant en compte les résultats d'analyse, sans *a priori*, des big data disponibles rassemblées dans les bases de données médico-administratives [3].

Le regroupement de l'ensemble de ces données massives produites en santé a trouvé son aboutissement dans une notion globale qui est celle de patient « numérique » dont l'identité se manifestera par l'association de la totalité des données disponibles concernant une personne qu'il s'agisse de données cliniques, de données d'imagerie, de données biologiques notamment génétiques, métabolomiques, microbiotiques ainsi que de données

environnementales ou comportementales. L'analyse algorithmique supervisée ou non supervisée de l'ensemble de ces données permettra une meilleure approche diagnostique mais surtout une prédiction accrue du pronostic et des réponses thérapeutiques. Cette approche contribuera ainsi à l'essor de la médecine de précision. Ces associations de données permettront sans doute d'envisager des actions préventives dédiées.

La recherche clinique est aussi dès à présent concernée. Un système d'intelligence artificielle développé par Microsoft Heath pourra dès cette année permettre une interaction en temps réel entre les patients consultant le site clinicaltrial.gov qui regroupe sous forme synthétique l'essentiel des essais cliniques *via* un agent conversationnel ou chatbot qui permet le recueil d'informations par analyse de mots clés. Le chatbot guidera les patients qui le souhaiteraient vers un essai clinique pouvant les concerner en tenant compte des informations recueillies lors de l'échange électronique. Cette démocratisation de l'abord pourrait permettre une sensibilisation accrue des patients concernés et peut-être un recrutement facilité. D'autres chatbots ont été développés pour contribuer au meilleur suivi des patients pendant les essais cliniques. Ils permettront de diminuer le délai entre la survenue d'un possible effet secondaire et sa déclaration. Ils pourront aussi se substituer à certains carnets de suivi, ce qui devrait améliorer le suivi, l'observance aux traitements testés ainsi que la collecte de données, d'où une optimisation du temps de technicien de recherche nécessaire.

L'intelligence artificielle pourrait aussi être un instrument de gestion des politiques de santé au niveau d'un territoire. L'identification des parcours de soins, l'évaluation de leurs conséquences en termes prognostiques permettront de proposer des améliorations structurelles. Les comparaisons entre territoires permettront d'affiner l'offre de soins et de mieux gérer les ressources disponibles.

Ces développements impliquent les patients et plus largement les citoyens qui devront être informés de toute proposition diagnostique ou thérapeutique impliquant un système d'intelligence artificielle. Une information éclairée devra être envisagée, comme aujourd'hui nous le faisons lors du choix d'une thérapeutique, en présentant loyalement à la fois les effets bénéfiques attendus et les potentiels effets secondaires.

Une supervision humaine globale des systèmes d'intelligence artificielle est et restera nécessaire. Le maintien d'une expertise indépendante de très haut niveau de ces systèmes est indispensable pour assurer cette maitrise. Une labellisation, actuellement non encore anticipée, est à envisager rapidement sur le modèle de ce qui est réalisé pour les médicaments par l'Agence nationale de sécurité du médicament et des produits de santé (ANSM).

Le médecin restera indispensable pour la prise en charge du patient et ne sera jamais remplacé par une machine quels que soient son apport et son degré « d'intelligence ». La maitrise de la machine par le médecin doit être possible, et cet enjeu majeur actuellement difficilement atteignable doit être gardé en tête au moment du développement et de l'adoption d'un système. Parallèlement au déploiement de ces systèmes, un enseignement adapté doit dès à présent être envisagé tant en formation initiale qu'en formation continue. Cet ouvrage y contribue.

L'usage de l'IA en médecine une fois son utilité établie s'imposera comme naturel. Il impliquera des enjeux de formation aux technologies « primitives » pour les jeunes collègues « natifs » de ces nouvelles technologies. Comment former ces nouvelles générations à l'utilisation optimale de ces systèmes tout en maintenant une expertise intellectuelle humaine autonome hors système ? Ces questions peuvent sembler aujourd'hui superflues mais elles doivent être anticipées pour éviter des déconvenues majeures. Les systèmes aéronautiques sophistiqués actuellement disponibles n'évitent pas aux futurs pilotes l'apprentissage du vol à l'ancienne qui doit rester maitrisé.

Le temps médecin nécessaire impliquant la maitrise et l'usage de ces nouvelles technologies doit aussi être anticipé en particulier au moment d'une contraction démographique médicale notable. Actuellement, les médecins cliniciens passent en moyenne deux fois plus de temps à consulter et renseigner des systèmes informatiques que directement avec leurs patients. Cette évolution non anticipée et souvent méconnue est une cause notable de burn-out car ces tâches additionnelles sont souvent assurées pour une large part une fois la journée de consultation achevée (concept de « pyjama doctor »). Les systèmes d'IA devraient permettre une automatisation de certaines de ces tâches, notamment en matière de comptes rendus et de codage automatisés mais ce bénéfice attendu restera à démontrer.

Enfin ces systèmes ouvrent vers des problématiques nouvelles de propriété et/ou d'autorisation d'exploitation des données concernées avec des enjeux de droits de l'homme et de liberté individuelle. Le développement d'algorithmes mal compris ou non maitrisés peut devenir un problème éthique majeur dont le cadre juridique commence seulement à être saisi au niveau européen et français dans ses principes et reste à développer dans son application. Le sujet étant économiquement et socialement majeur, le développement de ces systèmes sera l'objet d'une compétition internationale avec des différentiels de valeurs entre différents groupes humains qu'il sera difficile de gérer.

Il nous appartiendra de développer nos connaissances et de maitriser ces enjeux pour permettre le développement le plus harmonieux possible de ces techniques au

service du progrès dans la prise en charge de nos patients. Cet ouvrage nous permet dès à présent de nous préparer à ces enjeux.

RÉFÉRENCES

1. *Panorama des startups santé françaises utilisant l'IA*. BPI, 2020. https://lehub.bpifrance.fr/panorama-startups-sante-francaises-ia/
2. *La Plateforme des données de santé (Health Data Hub)*. Cnil, 2021. https://www.cnil.fr/fr/la-plateforme-des-donnees-de-sante-health-data-hub
3. *[Grand Défi] Santé: améliorer les diagnostics médicaux à l'aide de l'intelligence artificielle*. Gouvernement français, 2021. https://www.gouvernement.fr/grand-defi-sante-ameliorer-les-diagnostics-medicaux-a-l-aide-de-l-intelligence-artificielle

SANTEDIGE, C'EST....

- une **FONDATION** d'entreprise, depuis 15 ans aux côtés de l'Institut des maladies de l'appareil digestif de Nantes (IMAD) ;

- une **MISSION**, accompagner l'IMAD et garantir l'excellence des soins, de la recherche et de l'enseignement ;

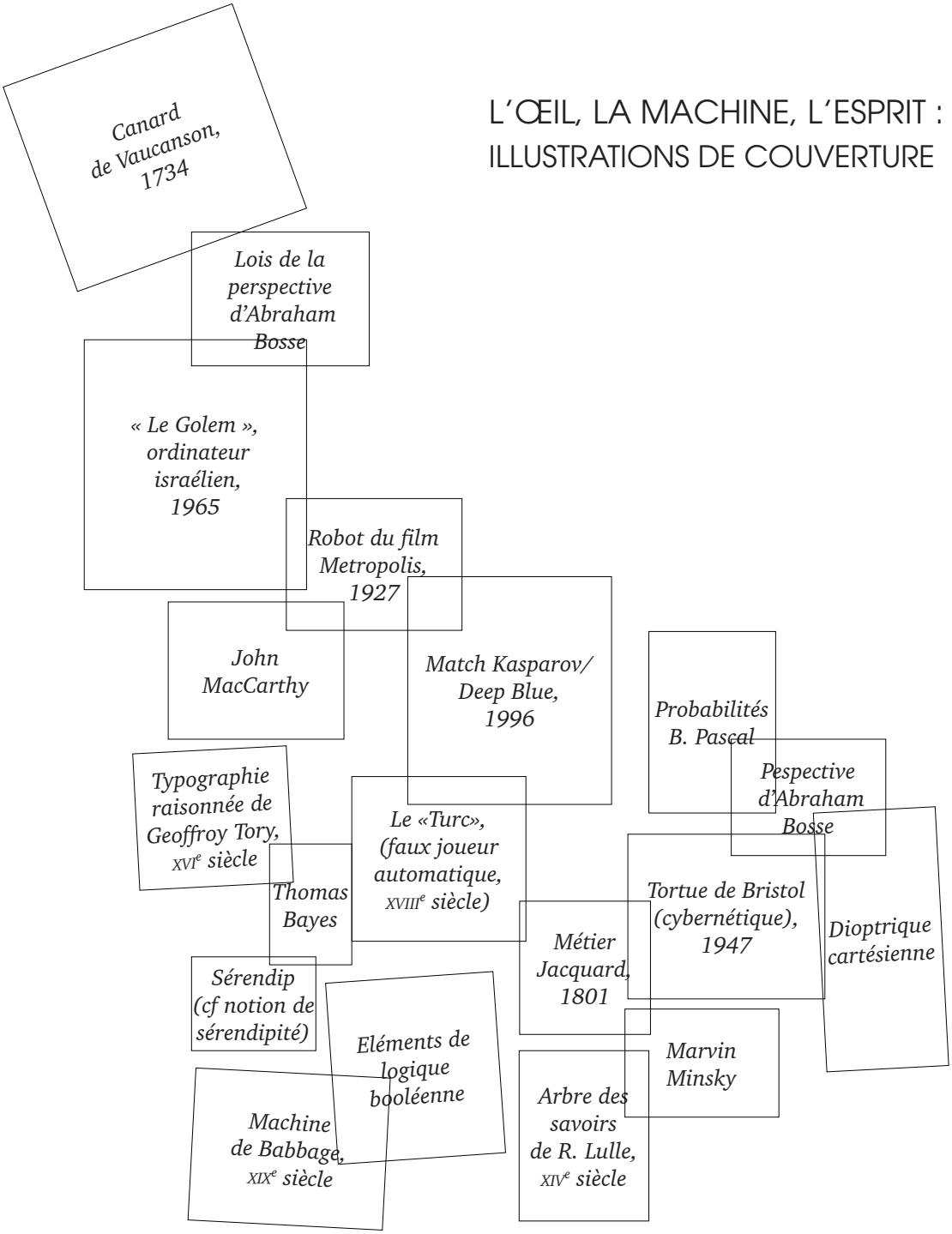
- une **AMBITION**, aider à mettre en place la médecine de demain, innovante, transparente et respectueuse du patient ;

- un **ENGAGEMENT** de dix mécènes issus ou non du monde médical au service de l'innovation scientifique ;

- une **CONVICTION** : les enjeux de santé peuvent et doivent mobiliser la société civile...

Santé Dige
FONDATION D'ENTREPRISE
POUR LA SANTÉ DIGESTIVE

L'ŒIL, LA MACHINE, L'ESPRIT : ILLUSTRATIONS DE COUVERTURE



Ce livre est le fruit d'une initiative
de l'Institut des maladies de l'appareil digestif de Nantes.

S'insérant dans le cadre d'une politique de partage
des connaissances scientifiques,
il est librement partageable à l'identique.

Il est édité par les éditions Le Grand Métier
et peut être cité comme suit :

Intelligence artificielle et maladies digestives.
Ed : Coron E, Vanbiervliet G. LGM : Rouen, 2021.