

Pour l'imagerie médicale, l'informatique n'est pas exactement une nouveauté. On peut même affirmer qu'il y a des algorithmes depuis qu'il y a de l'imagerie numérique car, sans eux, il serait impossible d'exploiter les signaux originaux fournis par les dispositifs, IRM et autres, qui échappent totalement à nos sens et avec lesquels sont construites les images du corps humain. Toutefois, depuis quatre à cinq ans, un nouveau pas a été franchi dans le traitement informatique des images médicales numériques avec l'avènement des algorithmes d'apprentissage profond, qui utilisent, avec succès, des masses de données considérables qu'aucun cerveau humain n'est capable d'appréhender en totalité.

Des exemples récents ont frappé les esprits. En dermatologie, un logiciel d'apprentissage profond, préalablement entraîné sur plus de 1 million d'images générales annotées, a été ajusté sur les images de 130 000 anomalies dermatologiques pour apprendre à distinguer automatiquement les mélanomes des lésions bénignes avec l'expertise d'un dermatologue. En radiologie, l'entreprise française Therapixel a soumis son logiciel aux 640 000 mammographies proposées lors d'une compétition mondiale et a remporté celle-ci en repérant mieux que tous ses concurrents les images suspectes. Ce programme se mesure aujourd'hui aux radiologues experts. En pneumologie, un logiciel développé par Google, entraîné sur des milliers de scanners et testé sur plus de 1 000 nouvelles images, est capable de détecter la présence de nodules suspects dans les scanners volumiques des poumons, et de les caractériser aussi bien qu'un professionnel.

— En bref —

> Algorithmes d'apprentissage profond et vastes bases de données d'images médicales offrent à l'IA une expertise comparable à celle de l'humain pour détecter certaines pathologies.

> Ce savoir-faire est néanmoins très spécialisé et coûteux à acquérir.

> La recherche en IA travaille aussi à établir des «jumeaux numériques» du patient, permettant un pronostic et une thérapie personnalisés.

> Un large partage des données et une formation de tous les professionnels de santé à l'IA sont nécessaires pour faire progresser la discipline.

Enfin, en ophtalmologie, d'autres logiciels d'apprentissage, entraînés sur plus de 130 000 images de la rétine, identifient les rétinopathies diabétiques avec la fiabilité d'un ophtalmologue. C'est le cas d'IDx-DR, approuvé en 2018 par la FDA, l'agence américaine du médicament, pour faire ce diagnostic automatiquement sans que l'image ne soit vue par un spécialiste. Depuis, plus d'une douzaine de systèmes d'imagerie assistés par l'IA ont reçu ce label.

UN DOUBLE NUMÉRIQUE POUR CHAQUE PATIENT

Bien que ces résultats soient spectaculaires, ils restent pour l'instant confinés à des tâches précises, relativement étroites, pour lesquelles d'immenses bases de données renseignées par des spécialistes sont disponibles. Or ces bases ne sont pas légion, leur acquisition et surtout leur analyse par des experts coûtent très cher (voir l'encadré page ci-contre). De plus, en cas de doute, les algorithmes d'apprentissage profond sont caractérisés par des millions de paramètres opaques à l'interprétation. Ils sont aussi sensibles aux attaques contradictoires. En 2019, une étude a révélé qu'ajoutées dans les images, de petites perturbations totalement invisibles à l'œil nu mais calculées très précisément bouleversaient les résultats (voir *Un seul pixel vous manque, et tout est désappris*, par J.-P. Delahaye, page 60). Dans ces conditions, une lésion de la peau initialement classée bénigne avec 100 % de confiance devient maligne avec le même taux de confiance, alors qu'à l'œil nu la différence est imperceptible. L'inverse est aussi envisageable,

et plus grave, puisqu'une base piratée pourrait conduire à l'absence de traitement de cancers.

Comment rendre les algorithmes d'IA plus robustes ? Une piste consiste à les combiner avec des modèles numériques du patient, ces derniers exploitant les connaissances générales que nous avons en anatomie et en physiologie, mais en se limitant à un nombre réduit de paramètres plus facilement interprétables. Plus précisément, des algorithmes et des modèles mathématiques, physiques, et biologiques du vivant permettent aujourd'hui de construire, à partir d'images médicales et d'autres données qui lui sont propres, une représentation numérique et personnalisée du patient, un «patient numérique personnalisé», «e-patient», ou «jumeau numérique». Les paramètres et les modèles de cette réplique sont ensuite utilisables par des algorithmes de médecine numérique conçus pour l'aide au diagnostic, au pronostic et à la thérapie.

Les modèles biophysiques du patient numérique aident aussi à enrichir les bases de données en simulant, par exemple, des maladies rares ou des cas peu ou mal représentés parmi les malades répertoriés. Des simulations intensives réalisées *a priori* sur des masses d'informations variées sont utiles pour accélérer les calculs scientifiques sur de nouveaux cas. Les approches statistiques de l'apprentissage automatique sont donc complémentaires d'approches plus déterministes.

Enfin, le cadre théorique du patient numérique et les méthodes modernes de traitement des mégadonnées favorisent l'intégration de données qu'on peut qualifier d'holistiques sur le patient, incluant non seulement ses images anatomiques et fonctionnelles, mais aussi des informations biologiques très complètes (génétiques, métabolomiques...), voire comportementales et même environnementales selon le style de vie (on parle alors d'«exposome»).

Analyser l'ensemble de ces renseignements très hétérogènes et de grandes dimensions est un défi scientifique, mais la recherche est très active sur ces sujets, notamment dans le cadre des 3IA (instituts interdisciplinaires d'intelligence artificielle) déployés en France. L'objectif est de faire progresser la compréhension des maladies et de mieux guider leur prise en charge, notamment le dépistage et le traitement de cancers comme ceux du poumon et du sein, ainsi que de maladies neurodégénératives.

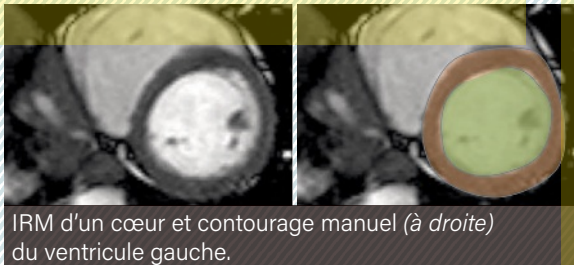
À terme, le champ d'application de l'intelligence artificielle en imagerie médicale est amené à s'étendre à toutes les disciplines médicales (radiologie, cardiologie, oncologie, neurologie, radiothérapie, chirurgie...). En conséquence, les médecins et, au-delà, tous les professionnels de santé, notamment les pharmaciens, le personnel médical... doivent être formés à l'IA, afin de

21

Quelques précautions s'imposent

Les réseaux de neurones utilisés en imagerie médicale ont besoin d'ingurgiter des dizaines, voire des centaines de milliers de scanners ou d'IRM préalablement préparés. Des dizaines de spécialistes doivent les annoter à la main, en précisant s'il y a ou non une

pathologie (annotation faible), voire en contournant les éléments pathologiques (annotation forte, comme ici un ventricule gauche). Cette phase de préparation, puis d'apprentissage est très coûteuse, car elle nécessite une main-d'œuvre hautement qualifiée et une forte puissance de calcul. Une fois correctement entraînés, les algorithmes détectent automatiquement des pathologies dans de nouvelles images... à condition de veiller à certaines précautions : s'assurer de la bonne représentativité des cas rares (qui par définition ne sont pas nombreux !), de l'absence de biais (genre, âge, ethnicité...), ou de la présence de sujets sains en nombre suffisant (ce qui pose un problème éthique lorsque la modalité d'imagerie est irradiante par exemple).



IRM d'un cœur et contourage manuel (à droite) du ventricule gauche.

Les algorithmes sont sensibles aux attaques contradictoires. Une perturbation peut faire passer une lésion classée 100 % bénigne en 100 % maligne

mieux comprendre son utilité et ses limites. Et déjà des 3IA développent des diplômes universitaires qui y sont consacrés (c'est le cas de celui de la Côte d'Azur, à Nice-Sophia Antipolis).

RASSEMBLER LES DONNÉES

22 Un partage plus large de l'information est aussi nécessaire. Au Royaume-Uni, la base de données UK-Biobank met à la disposition de la communauté scientifique l'ensemble des informations (images, analyses génétiques et paramètres environnementaux) acquises progressivement sur plus de 100 000 participants volontaires suivis dans le temps. Aux États-Unis, la base de données ADNI propose, pour une immense cohorte de patients Alzheimer, des images anonymisées de leur cerveau, ainsi que des données cliniques et génétiques.

La France se lance à son tour : Health Data Hub (HDH), Entrepôt des données de santé de l'AP-HP (EDS), France Life Imaging, Drim France IA, etc., autant d'initiatives pour rassembler des données de santé associant imagerie et informations complémentaires sur des patients, et pour les mettre à la disposition des chercheurs en préservant bien sûr la confidentialité et avec l'accord des participants. Des projets indispensables pour espérer des progrès, possiblement considérables, dans la recherche, et pour cela nous devons notamment veiller à ce que la réglementation, tout en protégeant suffisamment les patients, ne pénalise pas les chercheurs français et européens face au reste du monde.

L'un des moyens d'y parvenir consiste à développer l'apprentissage fédéré. Les données restent dans les centres qui les ont produites, et ce sont les algorithmes d'apprentissage qui sont exécutés sur

les ordinateurs de chaque centre et qui partagent entre eux les apprentissages locaux pour fournir un résultat global. Un tel projet d'apprentissage fédéré pour l'imagerie de médecine nucléaire est mené par nos 3IA en partenariat avec plusieurs centres de lutte contre le cancer, parmi lesquels le centre Antoine-Lacassagne à Nice, l'institut Curie, à Paris, et le centre Henri-Becquerel, à Rouen.

En définitive, intelligence artificielle, imagerie médicale et patient numérique forment un ensemble d'outils informatiques au service d'une médecine aux «4 P» : plus personnalisée, précise, prédictive et préventive. Ces outils numériques feraient gagner beaucoup de temps au thérapeute pour mieux prendre en charge son patient. Ils assisteraient également des généralistes en leur proposant d'orienter les patients, lorsque cela est nécessaire, vers un spécialiste. Ces nouveaux outils ont vocation à seconder le médecin, pas à le remplacer.

— L'auteur —

> **Nicholas Ayache**
est directeur de recherche à l'Inria et directeur scientifique de l'Institut interdisciplinaire d'intelligence artificielle (3IA) Côte d'Azur.

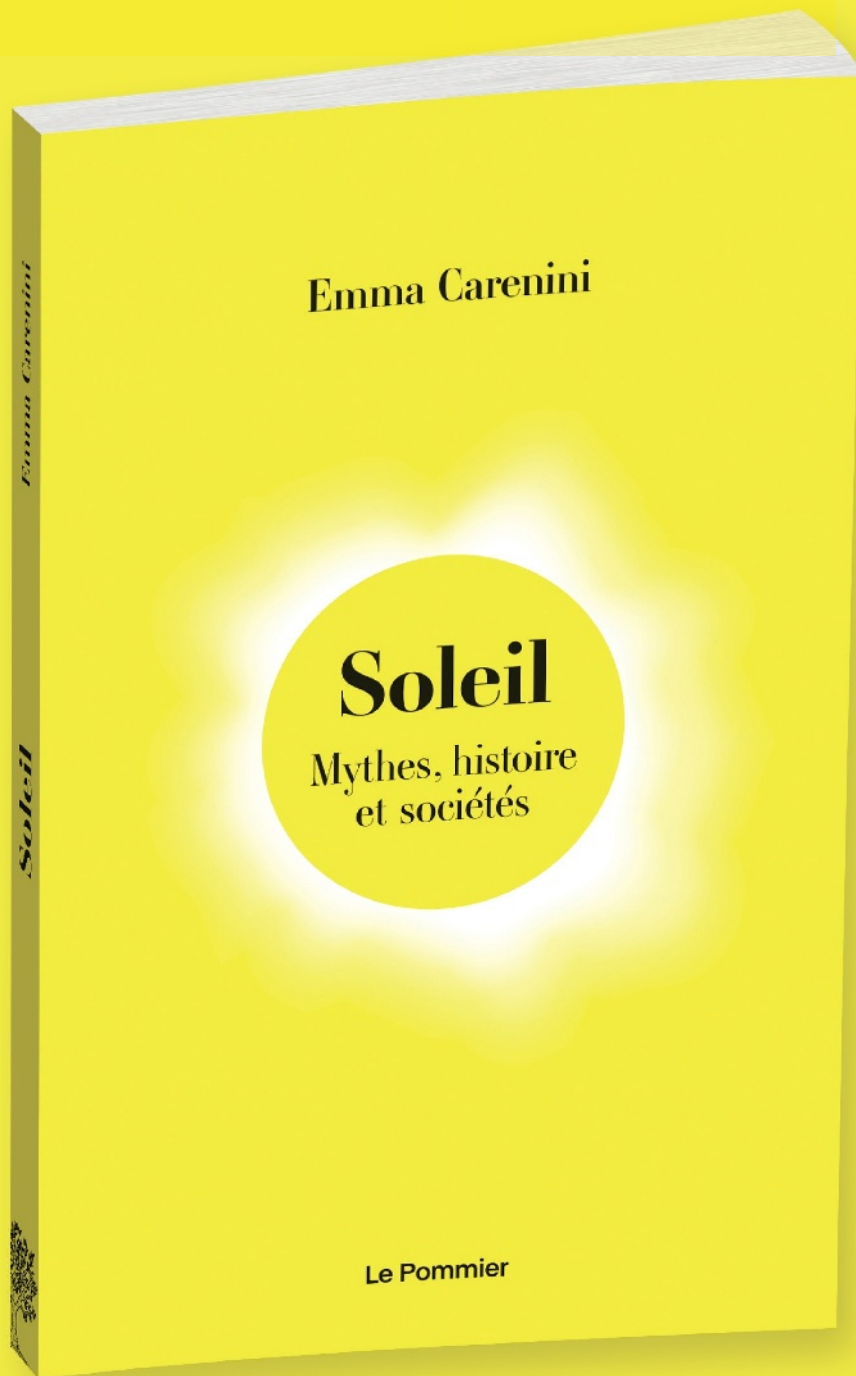
— À lire —

> **N. Ayache et al.**, *Nouvelle enquête sur l'intelligence artificielle*, Champs Actuel, Flammarion, 2020.

> **S. G. Finlayson et al.**, *Adversarial attacks on medical machine learning*, *Science*, 2019.

> **N. Ayache**, *Des images médicales au patient numérique, Leçon inaugurale du Collège de France, chaire annuelle 2014.*
<https://bit.ly/3pSrnnin>

Du cadran solaire à la fusion nucléaire



En librairie



“La traduction automatique doit son succès aux statistiques

24

Thierry Poibeau

est directeur de recherche
au CNRS et directeur adjoint
du laboratoire Lattice
(Langues, textes, traitements
informatiques et cognition).

