

Communication

Le rôle des Méga données dans l'évolution de la pratique médicale

MOTS-CLÉS : TRAITEMENT INFORMATIQUE MÉDICAL . PRATIQUE PROFESSIONNELLE. RECHERCHE

Role of Big Data in evolution of the medical practice

KEY-WORDS: MEDICAL INFORMATICS COMPUTING. PROFESSIONAL PRACTICE. RESEARCH

Jean-Patrick LAJONCHÈRE*

L'auteur déclare : être le directeur général du groupe hospitalier à Paris sans lien d'intérêt avec aucun industriel du secteur du numérique

RÉSUMÉ

Depuis cinq ans environ, sont annoncées des avancées médicales potentielles grâce à l'utilisation des « mégadonnées ». Ce terme, traduction de l'anglais « big data », rassemble la notion de stockage de données de masse mais également la conception d'algorithmes complexes en permettant l'exploitation.

Des travaux publiés, notamment en cardiologie, dermatologie et diabétologie, montrent la capacité d'expertise de l'exploitation de fichiers par intelligence artificielle avec un niveau aujourd'hui comparable à ceux d'experts de ces spécialités.

Les bénéfices potentiels concernent le renforcement de l'expertise et de l'aide au diagnostic pour la pratique médicale et de l'élargissement de la capacité d'analyse pour les chercheurs. Les patients trouveront des améliorations dans la prise en charge personnalisée de leur santé grâce à l'utilisation des objets connectés et en santé publique l'optimisation des systèmes de santé bénéficiera de l'exploitation des fichiers de l'Assurance maladie sur la base du service rendu au patient.

Des problèmes méthodologiques et techniques se posent encore et en terme d'éthique, le respect du secret médical et de liberté de l'exercice de la médecine seront interrogés par ces nouvelles technologies.

La voie étant tracée, il reste à nos équipes de chercheurs et de cliniciens à prendre toute leur place dans ce mouvement mondial, avec les forces que sont celles de notre pays.

SUMMARY

For about five years, medical advances have been announced thanks to the use of " Big Data". This term refers to the collection and storage of mass data, and also the design of complex algorithms allowing their exploitation.

Published works, notably in cardiology, dermatology and diabetology, show the expertise capacity of the exploitation of files by artificial intelligence with a level now comparable to those of experts of these specialties.

The potential benefits include strengthening expertise and diagnostic assistance for medical practice and expanding analytical capacity for researchers. Patients will find improvements in the personalized management of their health through the use of connected objects and, in public health, optimization of health systems will benefit from the use of health insurance files on the basis of the service given to the patient.

Methodological and technical problems still arise and in terms of ethics, respect of the medical secrecy and freedom of practice of medicine will be questioned by these new technologies.

The path being traced, it remains for our teams of researchers and clinicians to to play their role to the full in this global movement, with the forces that are those of our country.

* Directeur général de la Fondation Hôpital Saint-Joseph, 185 rue Raymond-Losserand 75014 Paris. E-mail : jlajonchere@hpsj.fr

Fédérateur de la famille « Mieux se soigner » auprès du ministre de l'Europe et des Affaires étrangères.

Tirés à part : Professeur Jean-Patrick LAJONCHÈRE, même adresse

Article reçu le 25 septembre 2017, accepté le 20 novembre 2017

INTRODUCTION

Le 22 mars 2012, le géant de l'informatique IBM conclut un partenariat avec l'hôpital newyorkais Memorial Sloan-Kettering, accord qui vise à adapter le logiciel d'intelligence artificielle « Watson » pour créer un outil d'aide à la décision médicale dans le diagnostic du cancer [1].

Google et Microsoft ne sont pas en reste : Google achète et développe Calico et Verily, deux sociétés présentes dans le secteur des biotechnologies et s'intéresse à tout objet qui peut capter des données de santé et se connecter (comme une lentille de contact pour diabétiques capable de mesurer la glycémie, développée avec le laboratoire Novartis) ; Microsoft a lancé Microsoft Health en 2014 afin de proposer une plateforme permettant à chacun de gérer sa santé ; Sanofi et Google créent en septembre 2016 un partenariat dans le traitement du diabète.

Les autres entreprises du secteur informatique et de l'intelligence artificielle s'intéressent de près au marché du numérique en santé. Celui-ci, évalué à 24,55 milliards de Dollars en 2020 (7,39 en 2016) constitue une cible de choix. En effet, les experts considèrent que cette évolution est notamment expliquée par le développement potentiel des EHR (Electronic health record) [2].

En France, les sociétés de conception de logiciel sont nombreuses, souvent de petite taille. Cependant, Dassault Systèmes émerge de ce paysage avec un potentiel de développement d'envergure internationale.

Dans le même temps, toute avancée en matière d'apport des « mégadonnées », en anglais « big data », en santé est abondamment commentée avec, quelquefois, une pression médiatique qui amplifie à l'excès leur apport réel.

Est-ce pour autant que l'utilisation et l'exploitation des mégadonnées, ne peuvent servir qu'à des fins commerciales ou de publication d'articles à sensation?

Dans ce contexte, les travaux de recherche conduits grâce à ces nouvelles techniques doivent-ils se développer et donner du sens à ces initiatives? Représentent-ils un intérêt pour les patients? L'utilisation des mégadonnées pour optimiser les systèmes de santé est-elle une utopie ou une solution permettant une importante avancée des connaissances?

En face de l'apport des mégadonnées, des questions se posent : le respect du secret médical, plus généralement des droits des patients est-il mis en risque ? Le rôle du médecin sera-t-il mis en cause par ces outils ou par l'utilisation qui en sera faite, y compris dans le contrôle de leur exercice ? Quelle qualité attribuer à l'utilisation de ces ensembles considérables de données dont la validité ne peut être garantie ?

Il n'est pas proposé ici de répondre exhaustivement à ces interrogations, mais d'apporter des éléments d'état des lieux et de compréhension.

LES MÉGADONNÉES

Le terme « Big data » semble avoir été utilisé pour la première fois en 1997 par Michael Cox et David Ellworth [3]. Il désigne un système rassemblant plus de mille milliards d'octets (tera octets soit 10^{12} octets), voire péta octets (10^{15} octets) ou plus qui peuvent comprendre données structurées, documents, images et sons en intégrant également les technologies spécifiques de recherche et de traitement de ces données.

À titre d'exemple, le fonds de la Grande bibliothèque de France, toutes sources de données confondues, est estimé à 700 tera octets.

En 2011, les données de santé stockées dans le monde étaient estimées à 10^{17} octets [3]. On parle aujourd'hui de 150×10^{18} octets.

Le magazine économique Forbes estime que le marché global de ce secteur va croître de 18,3 milliards de Dollars en 2014 à 92,2 en 2026, ce qui représente une croissance annuelle de 14,4% [4].

Le champ des mégadonnées ne se définit pas uniquement par la quantité stockée. Les anglo saxons utilisent l'acronyme de 5V pour en résumer l'étendue: volume, velocity, variety, veracity, value [5]. Les données doivent en effet être exploitables rapidement et pour certaines applications, en temps réel. Elles doivent être « vraies » : le chercheur ou l'exploitant doit avoir confiance en ces sources. La diversité de celles-ci et le volume des données constituent une réelle difficulté d'analyse mais en font une richesse potentielle.

Afin de traiter ces « data », se développent des programmes informatiques de plus en plus efficaces reposant sur l'« intelligence artificielle ». On distingue :

- le « machine learning », qui repose sur l'acquisition des données et l'élaboration d'algorithmes basés sur celles-ci, algorithmes qui peuvent déterminer les points communs entre deux populations en se basant sur des méthodes statistiques. L'exploitation des apprentissages sur de nouvelles données permet de traiter un grand nombre de variables et d'enregistrements ;

- le « deep learning » qui conduit l'apprentissage de la machine sans règle initiale, basée sur le concept des réseaux neuronaux. Ces méthodes reposent sur des modes d'analyse probabiliste, du type bayésien déjà utilisés, par exemple, dans les protocoles de recherche clinique pour en renforcer la puissance.

Les reconnaissances faciale et vocale, notamment, ont pour bases ces techniques de « deep learning ».

Ces deux approches de technologies algorithmiques sont indissociables des mégadonnées dont elles constituent un élément essentiel : le cœur de leur traitement.

Ainsi, derrière le terme de mégadonnées sont regroupées les notions de stockage de masse, de données issues de sources multiples, de « 5V », d'intelligence artificielle permettant leur exploitation, afin de répondre à une question posée sur un champ d'analyse bien déterminé.

LES APPORTS DES MÉGADONNÉES EN SANTÉ

« En santé, l'objectif du traitement des mégadonnées est [...] de fournir de nouvelles informations utiles à la société et aux malades en garantissant au citoyen, au chercheur la meilleure confidentialité et le meilleur accès aux données. » [6].

Aujourd'hui, quelques exemples tirés de la littérature permettent d'illustrer ce que peuvent apporter l'utilisation des mégadonnées et des logiciels d'intelligence artificielle :

- l'utilisation de quatre algorithmes dotés de la faculté d'auto apprentissage, a été comparée aux recommandations de l'American college of cardiology.(ACC) L'expérience porte sur 378 256 patients du Royaume-Uni victimes d'accidents

cardiovasculaires à partir de données de 2005. Les systèmes se sont entraînés sur 78% des données puis ont estimés sur les 22% restant les chances de survenue d'un incident cardiovasculaire. Ces résultats ont été comparés à la réalité sur les années suivant 2005. Les résultats prédictifs à dix ans obtenus par les algorithmes sont meilleurs que ceux utilisant les recommandations de l'ACC. Sur l'échantillon test de 83 000 patients, le meilleur algorithme détecte 7,6% d'alertes en plus et 1,6% de fausses alertes de moins que les recommandations. Selon les auteurs, 355 patients auraient ainsi pu être sauvés par l'utilisation des algorithmes [7] ;

- en dermatologie, l'analyse algorithmique de 2 032 cas a été comparée avec celle d'un groupe de 21 experts après entraînement de la machine sur une banque d'images. Les résultats obtenus sont comparables en termes de spécificité et de sensibilité, dans la majorité des cas meilleure pour le logiciel de « deep learning ». Ainsi, les perspectives d'utilisation de cet algorithme, couplé avec l'utilisation des smartphones, pour collecter les images d'anomalies cutanées, laisse entrevoir une extension importante de la prévention des cancers dermatologiques [8] ;
- dans le suivi du diabète, l'utilisation du « deep learning » pour détecter des rétinopathies diabétiques obtient d'ores et déjà des résultats exploitables en clinique. Un algorithme a été développé par Google en collaboration avec une équipe de l'université d'Austin (Texas) à partir d'un échantillon de test de 128 175 images de rétines de provenances différentes. La validation a été réalisée sur deux échantillons de 9 963 et 1 748 images. La spécificité sur le deuxième test a été de 96,2%, la sensibilité de 93,9%, comparables aux résultats des sept experts qui ont également interprété les images. Les auteurs mettent en évidence les limites du système au stade actuel, comme la difficulté d'utilisation de l'algorithme sur des images issues de matériels très différents et la supériorité d'experts très chevronnés capables de repérer d'infimes lésions que le programme ne détecte pas. Mais ils précisent également l'avantage de reproductibilité sur l'être humain, le logiciel ayant toujours la même performance. Même s'il reste à valider leur utilisation en routine et la faisabilité de l'application de ces algorithmes, la voie est tracée [9].
- depuis plusieurs années, Google Flu Trends (GFT) tente de prédire la dissémination de l'épidémie de grippe. Les résultats d'une étude réalisée sur huit pays d'Amérique du sud entre 2012 et 2014 ne sont pas complètement probants. Ils sont meilleurs au Mexique et les auteurs suggèrent que la corrélation est supérieure dans les pays plus tempérés que subtropicaux [10]. Le suivi du développement de l'épidémie de SIDA et les conséquences des actions de prévention et de traitement pourraient, selon des chercheurs, bénéficier de ces travaux en exploitant les données des réseaux sociaux en complément de celles des services de santé [11].

Cette liste ne se veut pas exhaustive. Elle montre cependant les potentiels de l'exploitation des mégadonnées, dans certains cas déjà équivalents aux performances de l'homme. Mais l'état d'avancement de chaque projet permet d'apprécier le chemin qui reste encore à parcourir jusqu'à l'utilisation en routine.

Mobiliser les chercheurs :

L'exploitation des données de santé à objectif de recherche repose en grande partie sur leur disponibilité et leur accessibilité, notion qualifiée d'« open data », qui vise à les partager, pour la santé, par des hôpitaux, des organismes de recherche, des administrations, des entreprises, des communautés de patients...

Les conditions d'accès, la structuration, les règles juridiques entre pays, rendent ce partage complexe, mais leur intérêt a été étudié et le bénéfice potentiel estimé, notamment en recherche [12]. En septembre 2016, l'éditorial du New England journal of medicine traite de l'implication de ce journal dans le développement de l'open data et incite les chercheurs à cette démarche de partage pour favoriser le développement de leurs travaux [13].

L'épidémiologie va devoir s'adapter très rapidement à l'utilisation des mégadonnées. Son enseignement doit rapidement en tenir compte afin de préparer les générations futures à ces évolutions [14]. L'émergence d'unités de recherche, en collaboration avec les sociétés développant les algorithmes est essentielle à cette évolution.

Harmoniser les approches juridiques entre les états :

Les bases de données en « Open Data » doivent être mises en places et managées par des autorités compétentes, publiques ou agences, afin d'en garantir la pérennité et d'en permettre l'utilisation sous conditions. A titre d'exemple, l'agence régionale de Santé d'Ile-de-France a élaboré une charte garantissant et précisant les conditions d'utilisation des données régionales en santé. Y sont précisées selon les types d'études, les partenaires concernés, les règles d'accès aux données et les conditions de publication. Le tout dans le respect des réglementations sur la protection des patients et de leur vie privée. Un comité scientifique et éthique régule l'accès à cette base de données.

Ce qui est réalisé à titre régional doit être validé, amplifié en France et partagé par d'autres états.

DOMAINES POTENTIELS D'APPLICATION :

Pour autant, à partir de ces exemples et des perspectives décrites dans la littérature, sans que des résultats probants ne soient encore obtenus, il est possible d'estimer d'ores et déjà les domaines où l'utilisation des mégadonnées et des algorithmes seront sources de progrès.

Pour la pratique médicale :

Un travail de recherche a tenté d'apprécier les bénéfices potentiels des analyses fondées sur 26 exemples d'exploitation des mégadonnées en santé pour des entreprises du médicament et du dispositif médical, en santé publique et en recherche clinique dans divers pays. Trois champs d'application et d'aide au médecin sont prévisibles [15] :

- le renforcement de la capacité à établir un pronostic dans le traitement du cancer notamment en enrichissant les scores actuels, limités par la capacité humaine d'analyse ;
- l'aide au diagnostic en imagerie et anatomie pathologie en permettant l'analyse des images en nombre toujours en augmentation ;
- l'amélioration de la fiabilité du diagnostic. Cette possibilité nécessitera du temps avant d'être opérationnelle, car il faut parvenir à exploiter des données non structurées pour bâtir, sur la base de travaux de recherche, les modèles d'analyse.

Les « assistants algorithmiques » [16] devraient fournir, en temps réel, les probabilités pour un patient, d'être atteint d'une ou de plusieurs pathologies, diminuant la marge d'erreurs potentielles, en intégrant de façon exhaustive l'état de l'art scientifique, où que l'on se trouve.

Que ce soit dans la façon d'acquérir les connaissances, que ce soit dans celles d'utiliser les outils d'aide au diagnostic existants ou développés grâce aux mégadonnées, la conception de l'acquisition du savoir médical en sera bousculée.

Pour la recherche :

Le jour où le débat sur l'ouverture des données des essais cliniques aura été tranché en faveur de celle-ci, comme c'est le cas pour les données des dossiers d'autorisation de mise sur le marché européenne, alors les chercheurs auront de nouveaux champs devant eux, comme l'analyse du rapport bénéfice/risque de médicaments inscrits dans différentes indications [17].

L'omni présence de l'analyse de données accélérera la modification des profils des équipes composant les unités de recherche. Les mathématiciens poursuivront leur entrée et les bio informaticiens y seront encore plus présents.

L'option prise par IBM de se positionner sur l'aide à l'inclusion de patients dans les protocoles de recherche constitue une application de routine dont les cliniciens chercheurs bénéficieront. « Watson for clinical trial matching » permet en effet de repérer les patients incluables dans les protocoles de recherche mis en place dans un hôpital donné. Les chiffres de la société donnent une réduction de 78% du temps médical nécessaire pour l'inclusion des patients.

Et, sans les rappeler, les diverses applications recensées supra illustrent la diversité des exemples de projets de recherche que les mégadonnées peuvent supporter.

Pour les établissements de santé :

Les enjeux du traitement des pathologies chroniques, du grand âge, mais aussi d'optimisation de la prise en charge non programmée rendent indispensables l'ouverture et le travail collaboratif avec tous les professionnels de santé non hospitaliers. En matière d'admission de patients en urgence, l'adaptation de la bonne attitude clinique, l'appréciation du parcours hospitalier optimal, doivent pouvoir reposer sur l'analyse en temps réel de la situation clinique du patient comparée à un échantillonnage de données. Deux programmes de recherche visant cette optimisation sont en cours par Kaiser permanent Northern California [18]. Au Portugal, les équipes du Sao Joao center hospital ont mis au point une surveillance de patients pour détecter en temps réel les résistances bactériennes. Elles annoncent être en mesure de prévoir 30% d'admission en réanimation huit jours en avance [19].

L'hôpital se trouvera au centre de la médecine dite « P4 » : prédictive, préventive, personnalisée et participative. En l'espèce, les diagnostics prédictifs réalisés grâce à la biologie moléculaire pourront être améliorés en intégrant l'analyse des facteurs environnementaux et sociétaux rendue possible par l'exploitation de données nombreuses et multi sources.

Il faut donc pour l'hôpital, achever sa transformation numérique et se doter d'équipes de développement afin d'être capable d'entrer dans l'échange de données. Les tendances

actuelles de fonctionnement collaboratif au sein d'un même territoire de santé accéléreront cette démarche.

Pour le patient et plus généralement le citoyen :

Toutes les avancées potentielles obtenues grâce à l'utilisation des mégadonnées sont à destination du patient. Mais l'arrivée dans le paysage technologique des objets connectés peut changer le rapport du patient avec sa santé car les informations transmises à partir de ces dispositifs de recueil et de transmission deviendront exploitables par les algorithmes.

Les objets connectés poursuivront inexorablement leur diffusion. Les estimations donnent 73 millions d'appareils de santé connectés dans le monde en 2016 et 161 millions en 2020 [20]. En Europe, s'ils sont utilisés à des fins médicales, ils ont d'ores et déjà l'obligation du marquage CE. Mais qu'en sera-t-il de ceux qui sont plus destinés à améliorer le « bien-être » et des applications smartphones diffusées à grande échelle ? Pourront-ils être utilisés dans les essais cliniques sans normalisation ?

Mieux prendre en charge sa santé, guidé par une expertise personnalisée est l'un des « grands rêves » des visionnaires des mégadonnées. Mais si cet objectif ambitieux est loin d'être atteint, la mise à disposition de données plus fiables et plus ciblées est à coup sûr source potentielle de meilleure responsabilisation de chacun. Les associations de patients devront s'inclure dans le dispositif pour encore mieux remplir leur rôle.

Reposant sur une meilleure connaissance des interactions, la prédiction et donc la prévention pourront mieux impliquer le patient qui aura les moyens d'infléchir par des attitudes adaptées le cours de sa santé [21].

Pour la santé publique :

L'enjeu se résume en trois objectifs : l'accès au système de santé, la qualité du service délivré et le coût. Souvent en compétition les uns avec les autres, ces trois facteurs doivent bénéficier d'autres méthodes d'analyse pour progresser ensemble.

Si l'on peut penser que l'exploitation des résultats générés par l'utilisation des mégadonnées puisse améliorer la prévention et la prédiction des maladies chroniques, alors il est indéniable que le bénéfice pour la santé publique sera réel, ne serait-ce que par la réduction potentielle et le ciblage des examens complémentaires. A ce jour, il n'y a pas d'estimation des éventuels bénéfices. C'est pourtant ce qui devrait incliner les responsables politiques à faire accélérer les travaux dans cette direction.

Les écarts dans les prises en charges, réels ou estimés pourront être expertisés grâce au support des mégadonnées. Il s'agit bien là d'analyser, sur des bases plus étendues qu'aujourd'hui, la performance de notre système de santé tous professionnels confondus.

En santé publique, aux USA, 5% des patients représentent 50% des dépenses du système de santé. Leur repérage et la mise en place d'un accompagnement adapté pourrait permettre des économies substantielles, Mais les tentatives n'ont pas toujours été couronnées de succès. Une intégration, grâce aux mégadonnées de facteurs sociaux, d'habitudes... pourrait conduire à adapter le suivi [18]. Ainsi, la « médecine du futur » considère, au-delà de la pathologie,

l'ensemble des variables biologiques, comportementales et environnementales permettant d'établir un diagnostic et de proposer un traitement adapté [22]. Dans cette perspective, Varian et Microsoft ont lancé en 2016 un partenariat pour affiner les phénotypes de cancers à travers l'extraction et la mesure de bio marqueurs.

La structuration des données de santé est une condition majeure de développement des travaux à partir des mégadonnées. En France, elle pourra se faire autour de la base du Système national d'information inter régimes d'Assurance maladie (SNIIRAM) qui permettra le développement d'une plateforme de recherche et d'essai des méthodes numériques de santé [23]. La définition de l'interopérabilité des diverses bases de données représente un enjeu fort [24]. La création de l'institut national des données de santé est en cours d'élaboration du nécessaire modèle économique du système national des données de santé (SNDS).

Dans le futur, le cout de remboursement du parcours reposera sur une approche « basée sur la valeur » [25] reposant sur l'évaluation de la qualité du service rendu. Mais la diversité des acteurs, des prises en charge et de leur chaînage est source de complexité. La détermination de la valeur doit donc s'apprécier en mettant en regard coûts et performance, entendue au sens de l'utilité pour les patients. L'analyse et la transparence des résultats seront des facteurs considérables d'amélioration. [26] Seules les mégadonnées et la capacité qu'elles offrent à traiter cette complexité peuvent y aider.

Enfin, l'évaluation des innovations technologiques et des médicaments au moment de la mise sur le marché et au cours de la durée de vie sera probablement facilitée, si l'on considère la taille du recueil des données, pour peu que soit résolu l'aspect qualitatif du recueil.

Des enjeux de positionnement dans le monde :

Le 15 juin 2017, le président de la République, Emmanuel Macron, a exprimé sa volonté de faire de la France un pays leader « dans la convergence entre le numérique et le healthcare ».

Pour y parvenir, il faut favoriser des initiatives collaboratives.

Dans cet esprit, le rapport « Cinq propositions pour la médecine du futur, un enjeu majeur pour la France » propose de créer des plateformes associant une offre technologique de pointe grâce à des plateaux techniques et des cliniciens testant les produits en condition de vie réelle [23]. Trois pistes sont données : plateforme collaborative, ingénierie de la cellule thérapeutique et des dispositifs médicaux à composante biologique, plateforme collaborative cardio métabolisme et plateforme collaborative pour les pathologies ostéo articulaires et l'orthopédie [24].

Par ailleurs, la stratégie nationale e-Santé, publiée par le ministère de la Santé le 4 juillet 2016 propose dans son axe 2 de renforcer l'appui aux projets numériques les plus innovants en organisant de nouveaux appels à projets. La puissance publique envisage de travailler à l'évaluation des objets connectés. Elle vise également à favoriser le développement des systèmes d'aide à la décision médicale. En tout quinze mesures dont la mise en œuvre, le calendrier et les conditions de mise en œuvre ne sont cependant pas décrites.

DISCUSSION

Le « médecin augmenté » :

L'aide à la décision par les mégadonnées ne doit pas faire oublier la place de la clinique dans le rôle du médecin, rôle que l'on espère voir renforcé par cette aide qui doit permettre le recentrage sur le conseil et le suivi du patient au sein d'un réseau mieux communiquant. Certains pourront s'interroger sur la place qu'il restera au professionnel aux côtés de la machine. Cette interrogation n'est pas sans fondement, sauf que la relation au patient et l'examen clinique constituent fondamentalement le cœur de l'exercice médical. Il faut probablement plus assimiler l'utilisation des mégadonnées à celle d'un stéthoscope ou de toute autre aide au diagnostic et se l'approprier, plutôt que d'y voir une intrusion dans la pratique clinique.

Le contrôle de ces nouvelles technologies, par qui ?

Quelle place devra avoir l'évaluation de ces technologies par les autorités sanitaires ? La FDA conclut en 2013 que les dispositifs qui guident une décision clinique doivent être évalués par elle [18]. Il paraît effectivement inconcevable que soient laissés sans évaluation l'ensemble des objets connectés quand ils concourent au diagnostic ou au soin, mais également les logiciels d'intelligence artificielle, dispositifs médicaux par destination.

La sécurité des données d'une part, celle de leur utilisation d'autre part, sont deux points majeurs d'attention. La protection contre les intrusions, virus... doit être considérée comme cruciale. Des hackers ont, par exemple, piraté noms et adresses de 80 millions de patients d'une assurance américaine, Anthem [26]. La collision entre ouverture, indispensable à la création des bases et à l'analyse des données et confidentialité, garantie donnée à tout citoyen doit constamment être appréciée par des autorités indépendantes soucieuses de supporter les évolutions technologiques.

Des aspects éthiques et déontologiques à traiter :

La sécurité vis-à-vis de l'utilisation qui pourrait amener, au travers d'analyses ou d'études à enfreindre le secret médical, est chose difficile à garantir sans management du système de gestion et d'accès aux bases de données. Cet aspect de management des bases constitue un point fondamental du dispositif. Mais l'utilisation des objets connectés peut également représenter un risque pour la préservation de la vie privée. Etre suivi en temps réel par une application santé par l'intermédiaire de ces objets pourrait, une fois le temps de l'engouement passé, devenir très intrusif si les limites à l'utilisation des éléments recueillis ne sont pas posées.

Collecter les données, les organiser, les stocker a un coût. Il faudra concevoir un modèle économique pour organiser le système et garantir l'accès à ces données dans des conditions financières et éthiques transparentes. Rien n'est encore établi dans cette direction, mais il est fort à prévoir que ce sujet émergera dès les premiers résultats probants de l'exploitation des mégadonnées en clinique.

Les bénéfices potentiels de la mise en place de l'analyse des parcours de soins des patients dans et hors hôpital concourant à analyser leurs performances ne sont probablement pas sans conséquences. En effet, l'exercice médical basé sur la liberté de prescription sera interrogé par

cette analyse. Les ordres professionnels devraient l'accompagner dans un souci collectif d'amélioration de notre système de santé, en anticipant les problèmes que cela pourrait poser dans l'exercice des professionnels. Ainsi, une fois de plus, quelle sera la place respective des recommandations des systèmes issus des mégadonnées et de la décision du médecin à qui doit, bien sûr revenir la décision?

Des questions méthodologiques et technologiques :

L'opacité des algorithmes créés par les logiciels de « deep learning » est peu compatible avec la validation scientifique des résultats. En effet, la pratique médicale repose sur la preuve. Les logiciels apprenants risquent de bouleverser cette méthode puisqu'ils reposent essentiellement sur des méthodes probabilistes et sur l'apprentissage permanent. C'est un obstacle à franchir pour les méthodologistes afin de valider les recherches sur les bases d'evidence based medicine.

Les problèmes techniques restent nombreux, comme par exemple sur le champ de l'imagerie : la traçabilité et la quantité de l'archivage, la nécessité de conserver les données brutes en vue de retraitement avec de nouvelles technologies... Mais les difficultés de mise en œuvre de protocoles d'échange entre les divers supports ne sont pas moindres. La rapidité d'analyse en application clinique ne va pas sans difficultés également.

CONCLUSION

En conclusion, si les travaux de recherche reposant sur l'utilisation des mégadonnées sont encore en émergence, nos équipes médicales et hospitalières doivent, supportées par les entreprises, se rendre capables d'intégrer ce nouveau champ d'analyse. C'est un enjeu de poursuite de l'excellence française en médecine. Des standards internationaux seront tôt ou tard établis. Ne pas en être acteur serait probablement grave pour l'indépendance de la pensée médicale française.

Nous avons les moyens de réussir :

- par le développement des travaux de recherche s'appuyant sur l'utilisation des mégadonnées, Il y a encore beaucoup de place. Des publications de haut niveau international sont au rendez-vous ;
- par l'appropriation des « mégadonnées » en recherche clinique, potentiellement plus rentable en terme de puissance d'essai, une fois résolu les problèmes de qualité ;
- par le lancement d'un appel d'offres type Programme Hospitalier de Recherche Clinique ou Programmes de Recherche sur la Performance des Soins interministériel, dans l'esprit du rapport cité supra, intégrant les industries du numérique pour financer au niveau national une action d'ampleur ;
- par une veille réglementaire permanente transversale pour garantir conjointement sécurité confidentialité et dynamisme.

L'histoire n'est pas, aujourd'hui, écrite. L'école de médecine française, nos hôpitaux, nos équipes universitaires, peuvent jouer encore pleinement leur rôle, pour peu que la dynamique

s'enclenche très rapidement. C'est une chance à saisir, au tournant de la médecine du XXI^{ème} siècle.

RÉFÉRENCES

- [1] [En ligne] Disponible sur : <http://www-03.ibm.com/press/us/en/pressrelease/37235.wss> (consulté le 6 février 2018)
- [2] Healthcare analytics Market worth 24.55 Billion USD by 2021, [En ligne] Disponible sur : <http://www.marketsandmarkets.com/Market-Reports/healthcare-data-analytics-market-905.html> (consulté le 6 février 2018)
- [3] Yichuang Wang, LeeAnn Kung, Terry Anthony Byrd, Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technological forecasting and social change 18447
- [4] [En ligne] Disponible sur : <http://www.forbes.com/sites/louiscolumbus/2016/08/20/roundup-of-analytics-big-data-bi-forecast-and-market-estimates-2016#514e69b86f21> (consulté le 6 février 2018)
- [5] Joonseok Kim, Big Data, Health Informatics, and the Fyture of cardiovascular medicine, Journal of the American college of cardiology Vol. 69, N° 7, 2017
- [6] Gilles Chatellier and all « Big data » et « open data » : quel accès pour la recherche ? Thérapie (2016) 71. 97-105
- [7] <http://HealthTechnology> DOI: 10.1126/science.aal1058
- [8] Andre Esteva and all, Dermatologist-level classification of skin cancer with deep neuronal networks, Nature 00 month 2017, doi:10.1038/nature21056
- [9] Varul Gulshan and all, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, JAMA, doi 10.100//jama.2016.17216. [En ligne] Disponible sur : <http://jama.jamanetwork.com/> (consulté le 20 novembre 2016)
- [10] Simon Polett and all, Evaluating Google flu trends in Latin America. Important lessons for the next phase of digital disease detection, clinical infection diseases, (2017) 64 (1): 34-41
- [11] Sean D. Young, A “Big Data” approach to HIV epidemiology and prevention, Preventive medicine (2014), <http://dx.doi.org/10.1016/j.prmed.2014.11.002>
- [12] Barry S. Rosenstein and all, How will big data improve clinical and basic research in radiation therapy? International Journal of Radiation oncology Biology Physics, 4 november 2015;
- [13] Jeffrey M. Drazen and all. The importance-and the complexities of data sharing New England journal of medicine September 21, 2016
- [14] Ross C. Brownson and all, Applied epidemiology and public health: are we training the future generations appropriately? Annals of epidemiology 27 (2017) 77-82

[15] Ziad Obermeyer, Ezekiel J. Emmanuel, Predicting the future-big data, machine learning, and clinical medicine, *The New England journal of medicine*, 375;13 NEJM.ORG, September 29, 2016

[16] [En ligne] Disponible sur : <http://www.lejdd.fr/Societe/Cedric-Villani-Le-big-data-va-booster-la-recherche-contre-le-cancer-845156> (consulté le 6 février 2018)

[17] Gilles Chatellier and all; “Big data” et “Open Data”: quel accès pour la recherché?, *Thérapie* (2016), 71, 97-105

[18] David W. Bates, and all, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, *Health Affairs*, 33, N°7 (2014): 1123-1131

[19] HVITAL Transforming hospital Big Data into real insights that empower clinical teams of timely act and save lives. Winner of HIMSS Elsevier Digital Healthcare Award 2016

[20] [En ligne] Disponible sur : <http://www.rflash.fr/donnees-massives-big-data--bouleversement-science-vie-et-medecine/article> (consulté le 6 février 2018)

[21] Mauricio Flores anf all, P4 medicine: how systems medicine will transform the healthcare sector and society, *Personalized medicine* (2013) 10 (6), 565, 576

[22] André Syrota, Olivier Charmeil, Cinq peopositions pour la médecine du futur, 26 avril 2017 page 27

[23] André Syrota, Olivier Charmeil, Cinq peopositions pour la médecine du futur, 26 avril 2017 pages 50 et 51

[24] André Syrota, Olivier Charmeil, Cinq peopositions pour la médecine du futur, 26 avril 2017 page 59

[25] Michael E. Porter, The strategy that will fix healthcare, *Harvard Business Revue*, October 2010

[26] Marc Fox, Ganesh Vaidyanathan, Impacts of healthcare big data: a framework with legal and ethical insights, *Issues in information systems*, volume 17, issue III, pp 1-10, 2016.