

[Point de vue]

Projets d'intelligence artificielle à l'échelle d'un établissement de santé : l'exemple du centre Léon Bérard

Pierre-Étienne Heudel, Thierry Durand, Jean-Yves Blay

DANS REVUE FRANÇAISE DES AFFAIRES SOCIALES 2017/4, PAGES 133 À 140
ÉDITIONS DREES MINISTÈRE DE LA SANTÉ

ISSN 0035-2985

DOI 10.3917/rfas.174.0133

Article disponible en ligne à l'adresse

<https://www.cairn.info/revue-francaise-des-affaires-sociales-2017-4-page-133.htm>



CAIRN.INFO
MATIÈRES À RÉFLEXION



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...

Flashez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.

Distribution électronique Cairn.info pour DREES Ministère de la santé.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

[POINT DE VUE]

Projets d'intelligence artificielle à l'échelle d'un établissement de santé : l'exemple du centre Léon Bérard

Dr Pierre-Étienne Heudel, Thierry Durand et Pr Jean-Yves Blay³

Introduction

Le big data en santé représente une innovation technologique de rupture qui devra nous permettre d'aller vers cette médecine des « 4 P » (Hood et Galas, 2008). Cette médecine qui précise, qui prévient, qui fait participer et surtout qui personnalise. Il s'agit bien d'un changement de paradigme, permis par la révolution numérique que nous vivons, avec cette transformation complète d'un système influençant toute notre société. Les nouvelles technologies informatiques favorisent le stockage de données quelles que soient leurs formes. Mais le big data n'est pas juste un stockage d'informations, c'est une nouvelle discipline se situant au croisement de plusieurs domaines technologiques, statistiques et de compétences métiers spécifiques. Cette approche répond à l'explosion des données, structurées ou non, observée dans le domaine de la santé entre autres. Ces capacités technologiques permettent de capter ces informations numériques, quels que soient leurs natures et leurs volumes, de les traiter à très grande vitesse et de les rendre ainsi exploitables. La recherche scientifique puis la médecine de demain vont connaître de profonds changements. De nouveaux domaines de recherche, paraissant aujourd'hui trop complexes à explorer à cause du volume d'informations à traiter ou de leur répartition spatiale (différents lieux de stockage) et temporelle (à différents moments), vont s'ouvrir. La qualité des conclusions des différentes études devrait s'améliorer avec la preuve par les données, permettant ainsi une meilleure compréhension des maladies et de leurs facteurs de risques, points cruciaux, notamment en cancérologie où la prévention et le dépistage précoce sont des éléments fondamentaux. L'enjeu de l'intelligence artificielle en médecine clinique n'est pas uniquement de modéliser des risques ou de programmer des algorithmes procurant un résultat à partir des données saisies, mais bien de représenter un outil d'aide à la décision pour les cliniciens dans leur travail au quotidien.

Une fois présentée ces quelques généralités, comment doit se positionner un établissement de santé comme le centre de lutte contre le cancer Léon Bérard de Lyon ? Comme tous les centres de lutte contre le cancer (CLCC), il s'agit d'un établissement de santé privé à but non lucratif et de caractère hospitalo-universitaire. À ce titre, la recherche et l'innovation sont des points fondamentaux dans les différents projets médico-scientifiques depuis plusieurs décennies. C'est probablement pour cette raison que le dossier patient informatisé (DPI) unique pour

3. L'ordre de présentation a été défini par les auteurs.

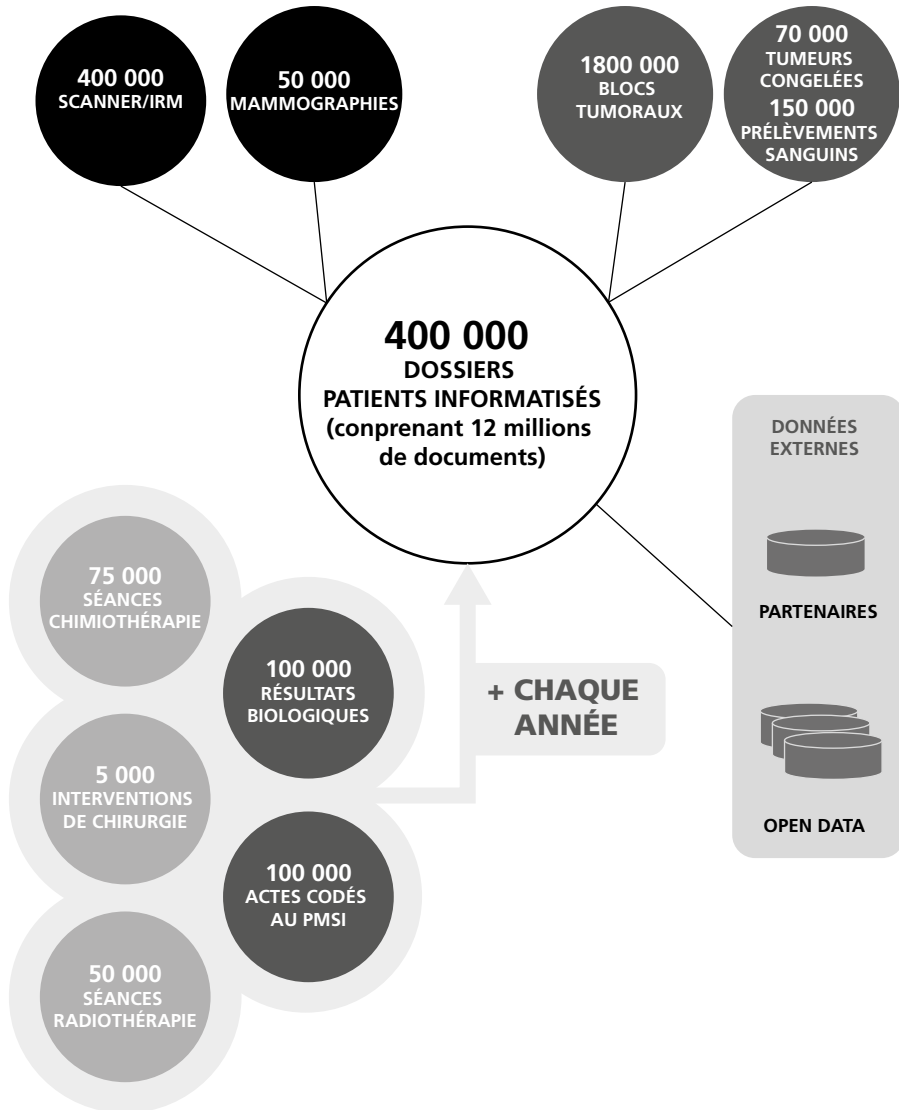
chaque patient est le dossier source au centre Léon Bérard (CLB). Les premières saisies de données médicales dans le système informatique datent de janvier 1993. Les grands gisements d'information, et donc de données, ont été rapidement informatisés : plateaux techniques, observations médicales, comptes rendus biologiques... Il aura fallu attendre juillet 2002 pour qu'officiallement le dossier informatique devienne le dossier patient de référence. Aujourd'hui, le métier d'archiviste a disparu de notre établissement. Cette informatisation complète des dossiers médicaux permet au CLB d'être un des premiers acteurs à une échelle nationale en ce qui concerne la dématérialisation dans les nouveaux modes de communication entre professionnels de santé et avec les patients via le portail patient myCLB. Lorsque l'on envisage une arrivée massive des objets connectés, on comprend aisément l'enjeu d'une communication sécurisée entre le patient et l'établissement hospitalier et surtout d'une intégration automatique de ces données dans le DPI du patient. Le volume des principales sources de données du centre Léon Bérard (figure 1) est nécessaire mais clairement insuffisant si l'on veut monter des projets de recherche utilisant les technologies d'intelligence artificielle. En simplifiant, pour faire du big data en santé clinique, il est nécessaire :

- d'avoir un établissement ayant des données de santé numérisées depuis de nombreuses années ;
- d'être capable de réunir une équipe projet avec des compétences très larges, bien au-delà des médecins et des informaticiens hospitaliers, intégrant idéalement une équipe de recherche s'intéressant à la thématique du projet ;
- d'avoir une forte compétence technologique, apportée par les industriels à ce jour.

Comme la figure 1 le montre, les données conservées dans un établissement de santé comme les autres sont très hétérogènes par leurs formats, leurs sources et leurs granularités. Une fois ce constat réalisé, il faut envisager dans un premier temps un outil permettant d'exploiter le gisement principal que représentent les dossiers médicaux informatisés. Dans un second temps, il faut ensuite envisager des projets de recherche permettant de croiser ces données médicales avec d'autres sources en cherchant à établir des corrélations entre des sources d'information disparates sans hypothèses formellement prédéfinies. Cette démarche inductive, appelée aussi « *data driven* », est également un nouveau mode de réflexion que doivent assimiler les médecins et les chercheurs formés à la recherche déductive, c'est-à-dire issue d'une hypothèse *a priori* (Provost et Fawcett, 2013).

Nous allons détailler, en premier lieu, le programme ConSoRe qui représente la première brique du projet big data du centre Léon Bérard. Ensuite, nous évoquerons quelques projets utilisant l'intelligence artificielle dans l'objectif d'aider le clinicien dans son activité quotidienne.

FIGURE 1
Données du centre Léon Bérard (PMSI : Programme de médicalisation des systèmes d'information)



ConSoRe

Afin de pouvoir répondre aux besoins d'utiliser les données de santé, Unicancer, qui regroupe les vingt CLCC français, a développé le programme ConSoRe (Continuum Soins Recherche) (Heudel, 2016). Il s'agit d'un outil de *data mining* qui permet de trouver facilement des données, structurées ou non, dans l'ensemble des dossiers patients informatisés. L'enjeu technologique est majeur car 80 % de données présentes dans un DPI sont présentes sous une forme textuelle dans les comptes rendus de consultations, de séjours... Développé par les équipes des quatre centres pilotes (Georges-François Leclerc à Dijon, Léon Bérard à Lyon, Institut du cancer à Montpellier et Institut Curie à Paris) et en collaboration avec l'industriel Sword, ConSoRe s'appuie sur les technologies Intel® et Luxid® Annotation Server d'Expert System, ainsi que sur de nombreuses autres technologies, *open source* ou non.

L'agrégation de données issues des différentes sources intégrant le système d'information hospitalier, permet à ConSoRe de modéliser l'histoire néoplasique du patient. Cela signifie que chaque document présent dans un DPI sera relié à un événement précis de l'évolution néoplasique (par exemple, une prise en charge initiale, une récurrence métastatique, un second cancer...). ConSoRe permet alors d'établir des requêtes extrêmement précises, et de manière simultanée, concernant les caractéristiques des patients (données démographiques), les cancers (localisation, période de diagnostic) et les traitements (chirurgie, chimiothérapie, radiothérapie).

Cette modélisation de l'histoire du cancer va permettre de structurer cette information de façon automatique pour en extraire des données et mettre à jour automatiquement des bases de données, aujourd'hui remplies manuellement et de manière rétrospective. Cette solution devra aider les médecins et les chercheurs des CLCC à identifier des pathologies ou des situations particulières. Grâce à l'architecture des données fédérées, les utilisateurs peuvent commencer par interroger les bases de données de leur propre CLCC, avant d'effectuer, si nécessaire, une recherche simultanée dans celles des autres centres. La recherche locale propose une liste de sujets correspondant aux critères, avec un résumé des informations du patient, ainsi qu'une liste de termes suggérés pour affiner la recherche. Les recherches concernant plusieurs CLCC indiquent le nombre de patients correspondant aux critères et la structure où ils sont traités. Aucune donnée protégée n'est divulguée.

I5ADORA

Le projet I5ADORA a pour but l'identification des seconds cancers primitifs et leurs facteurs de risques, notamment environnementaux, chez tous les patients suivis au centre Léon Bérard. Si nous nous intéressons aux patients atteints de deux cancers, c'est que près de 20 % des patients guéris d'un premier cancer

seront atteints par un second cancer primitif. Lorsque l'Institut du cancer estime à 3 millions le nombre de Français guéris d'un cancer en 2008, cela veut donc dire qu'environ 500 000 personnes seront touchées par un second cancer (Collectif, 2013). L'incidence du cancer augmentant avec les chances de guérison, la problématique des seconds cancers devient un problème de santé publique et la meilleure connaissance de leurs facteurs de risques un enjeu important. À ce jour, l'étude des facteurs de risques de seconds cancers primitifs s'est trop souvent limitée aux types de cancers, aux types de traitement reçus pour le premier cancer et à quelques facteurs de risques connus et prédéfinis (tabac, alcool, obésité) inscrits dans chaque dossier médical (Travis *et al.*, 2013). Le problème d'un tel sujet, c'est sa complexité avec l'étude de grandes populations hétérogènes et des suivis longitudinaux. On comprend très vite l'importance d'une approche multidisciplinaire et l'apport de technologies puissantes et innovantes pour croiser des données médicales avec des données environnementales ou d'exposition professionnelle. C'est par une meilleure compréhension de ces seconds cancers et de leurs facteurs de risques que l'on pourra envisager des stratégies individuelles de surveillance et des programmes personnalisés de prévention permise par des analyses prédictives.

Cette fois encore, ce projet est le fruit d'un travail collaboratif entre les équipes médicales et informatiques du CLB, l'unité de recherche Cancer et environnement pour son expertise sur les facteurs de risques environnementaux et les cancers professionnels, et l'industriel Sword Insight, qui fournit la conception de la plateforme technique et son expertise sur le brassage et l'investigation des big data. Sur cette thématique environnementale et à partir de cette collaboration, d'autres projets ont déjà été envisagés, notamment sur les facteurs de risques des sarcomes et des lymphomes.

Analyse automatique de l'image

La généralisation de l'accès internet à haut débit, l'utilisation importante des smartphones et des réseaux sociaux ont engendré en continu la création, la production et la diffusion des images. L'évolution technologique permet désormais d'analyser automatiquement ces images pour en extraire de l'information (Esteva *et al.*, 2017 ; Jalalian *et al.*, 2017 ; Vereb, 2014). Au sein d'un établissement de santé, l'imagerie est probablement le gisement de données qui est le moins utilisé au quotidien alors que le volume que représentent les scanners, les IRM ou encore les lames virtuelles d'anatomo-pathologie est très important. Or, on observe le développement d'analyses automatiques d'images médicales numériques par ordinateur, notamment pour détecter et caractériser des images anormales ou potentiellement anormales. La diversité des appareils d'imagerie, la complexité d'interprétation de ces images ainsi que leur multiplicité génèrent une charge de travail importante pour ceux qui doivent les interpréter. Afin de traiter cette masse

d'informations, l'objectif est d'élaborer un système d'aide au diagnostic, fondé sur l'analyse conjointe, c'est-à-dire la comparaison d'images médicales permettant de détecter des évolutions néoplasiques, ou des tissus cancéreux dans un ensemble donné, plutôt que de tenter de caractériser, avec un très fort a priori, le type de tissu. Cette analyse produite automatiquement a pour but de faciliter le diagnostic posé par le médecin, de le rendre le plus objectif et le plus fiable possible en intégrant au mieux les sources d'information disponibles. Pour essayer d'optimiser au mieux cette source d'images numérisées, plusieurs projets sont en cours de réflexion en partenariat avec de jeunes entreprises spécialisées dans ce domaine. Le premier projet se concentre sur l'analyse automatique des lames virtuelles d'anatomo-pathologie des cancers rares, c'est-à-dire pour des situations où le diagnostic est délicat à faire. Le second projet est orienté autour de la mammographie afin d'essayer de mieux prédire l'évolution d'images douteuses (classées ACR 3 [American College of Radiology]) et de mieux définir les indications de biopsies mammaires pour les images suspectes de cancer (classées ACR 4 et 5). Ces deux projets qui associent compétences médicales, informatiques et mathématiques ont là encore pour but d'aider les médecins au quotidien.

Génomique

Depuis environ une décennie, la cancérologie connaît une évolution rapide des concepts et des pratiques par l'intermédiaire du développement des connaissances scientifiques, notamment de la biologie moléculaire. Le séquençage à haut débit permet de séquencer l'intégralité du génome des cellules cancéreuses afin de mettre en évidence des mutations moléculaires responsables du développement du cancer afin d'envisager un traitement ciblant spécifiquement cette anomalie (Tredan *et al.*, 2017). L'objectif est de permettre aux médecins de disposer du profil génomique de chaque tumeur, de manière suffisamment fine et rapide pour l'intégrer à leur décision thérapeutique ou pour proposer à des patients des essais cliniques en vue de développer un traitement ciblé. Cependant la très grande variabilité des altérations génétiques (cible directe d'une drogue, activatrice d'un oncogène, altération de sensibilité, amplifications focales...) et leur évolution perpétuelle rendent difficiles la connaissance exhaustive de ces anomalies et de leurs corrélations possibles avec une ou des thérapies ciblées. Nous nous sommes rapprochés de plusieurs industriels pour envisager un partenariat pour l'utilisation d'outils permettant de repérer les variants anormaux tumoraux à partir de base de données scientifiques mises à jour en continu. L'objectif est ainsi de définir les possibilités thérapeutiques ciblées soit dans le cadre d'une autorisation de mise sur le marché, soit pour des molécules en essais thérapeutiques.

Les droits du patient

Tous ces travaux doivent se faire dans un total respect des droits du patient. Le centre Léon Bérard a rédigé un document d'information à l'attention des patients, validé par la commission des usagers de l'établissement. Nous tenons à expliquer aux patients les conditions d'usage de leurs données avec notamment les principes d'anonymisation que nous appliquons pour les travaux avec des équipes de recherche, la possibilité de refuser cette exploitation, et surtout les lignes blanches au niveau des partenariats qui ne seront jamais franchies.

Respecter les droits du patient, c'est aussi appliquer les règles de la commission nationale de l'informatique et des libertés (CNIL) qui, tant qu'elles ne sont pas modifiées¹, restent notre base de travail. Les données ne sortent donc jamais de l'établissement de santé, ce qui oblige les équipes de recherche et les jeunes entreprises à venir travailler dans nos murs. Cette règle est à prendre en compte lorsque le CLB doit travailler avec plusieurs établissements, comme dans le projet ConSoRe, lorsque la possibilité de connaître le nombre de patients dans chaque centre avec un profil donné a été développée, sans que les données de ces patients ne circulent.

Conclusion

Voici une rapide synthèse de nos différents projets issus d'une réflexion lancée il y a deux ans environ. Malgré des collaborations étroites entre les équipes du centre Léon Bérard, des équipes de recherches et des industriels impliqués, l'avancée reste limitée au vu de l'énergie dépensée, faute d'un canal de financement approprié pour ce type de projet d'innovation de rupture. Malgré cet obstacle, notre motivation reste intacte et notre réflexion évolue également avec nos besoins. C'est pourquoi nous devons multiplier les collaborations avec, par exemple, l'École normale supérieure dans le domaine des mathématiques, avec le cluster I-Care pour la dissémination des connaissances et des outils, avec les sciences humaines pour l'évaluation sociologique. Les modalités de valorisation financière de ces projets doivent être réfléchies dès le départ car le big data représente un relais de croissance économique pour ceux qui feront les bons choix. Enfin, ces projets doivent être construits pour être réutilisés, c'est le sens même de la dimension participative du big data.

1. Le nouveau règlement européen pour la protection des données est en place et s'imposera à tous à partir de mai 2018. La CNIL prévoit pour la fin 2017 une implémentation nationale, notamment pour les données de santé.

Références bibliographiques

Collectif (2013), *Identifier et prévenir les risques de second cancer primitif chez l'adulte, collection état des lieux et des connaissances*, ouvrage collectif édité par l'Institut national du cancer, Boulogne-Billancourt.

Esteva A., Kuprel B., Novoa R. A. et al. (2017), « Dermatologist-level Classification of Skin Cancer with Deep Neural Networks », *Nature*, 2 février, vol. 542, n° 7639, p. 115-118, doi : 10.1038/nature21056.

Heudel P., Livartowski A., Arveux P. et al. (2016), « The ConSoRe project supports the implementation of big data in oncology », *Bulletin du Cancer*, novembre, vol. 103, n° 11, p. 949-950, doi : 10.1016/j.bulcan.2016.10.001.

Hood L. et Galas D. (2008), « P4 Medicine: Personalized, Predictive, Preventive, Participatory: A Change of View that Changes Everything: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association », [en ligne] www.cra.org, <http://cra.org/ccc/resources/ccc-led-whitepapers/>, consulté le 19 octobre 2017.

Jalalian A., Mashohor S., Mahmud R. et al. (2017), « Foundation and Methodologies in Computer-aided Diagnosis Systems for Breast Cancer Detection », *EXCLI Journal*, février, vol. 20, n° 16, p. 113-137, doi : 10.17179/excli2016-701.

Provost F. et Fawcett T. (2013), « Data Science and its Relationship to Big Data and Data-Driven » *Decision Making*, février, vol. 1, n° 1, p. 51-59, <https://doi.org/10.1089/big.2013.1508>.

Tredan O., Corset V., Wang Q. et al. (2017), « Routine Molecular Screening of Advanced Refractory Cancer Patients: an Analysis of the first 2490 Patients of the ProfILER study », *Journal of Clinical Oncology*, vol. 35, suppl, abstr. LBA100.

Travis L. B., Demark Wahnefried W., Allan J. M. et al. (2013), « Aetiology, Genetics and Prevention of Secondary Neoplasms in Adult Cancer Survivors », *Nature Reviews. Clinical Oncology*, mai, vol. 10, n° 5, p. 289-301, doi :10.1038/nrclinonc.2013.41

Vereb G. (2014), « Advances in Automated Image Analysis », *Cytometry Part A.: the Journal of the International Society for Analytical Cytology*, may, vol. 85, n° 6, p. 478-479.