

Université Paris Descartes

École Doctorale de Santé Publique

Pierre Louis

INSERM UMR 1138 – équipe 22

Prédiction par Deep Learning de la réponse complète après radiochimiothérapie pré-opératoire du cancer du rectum localement avancé

Par Jean-Emmanuel Bibault

Thèse de doctorat de Santé Publique, spécialité Informatique biomédicale

Dirigée par Anita Burgun et Philippe Giraud

Présentée et soutenue publiquement le 21 Septembre 2018

Devant un jury composé de :

M. CUGGIA Marc, PU-PH, rapporteur

M. VERA Pierre, PU-PH, rapporteur

M. HENNEQUIN Christophe, PU-PH, examinateur

M. SAUT Olivier, HDR, examinateur

Mme BURGUN Anita, PU-PH, co-directrice de thèse

M GIRAUD Philippe, PU-PH, co-directeur de thèse

A ma femme, ma fille, ma mère et mon frère

Remerciements

Je tiens à remercier Anita Burgun et Philippe Giraud de m'avoir encadré pour cette thèse et de m'avoir accordé leur confiance. Votre aide, vos conseils, vos connaissances et votre enthousiasme pour la recherche m'ont beaucoup apporté tout au long de l'élaboration de ce travail.

Mes remerciements vont également aux membres du jury. Merci à mes rapporteurs, les Professeurs Marc Cuggia et Pierre Vera, de me faire l'honneur d'évaluer ce travail. Merci au Professeur Christophe Hennequin et au Dr Olivier Saut d'avoir accepté de participer au jury de cette thèse. Je tiens à leur exprimer ma reconnaissance pour l'intérêt qu'ils portent à ce travail.

Je souhaite remercier Bastien Rance pour l'aide qu'il m'a apportée au début de cette thèse dans la sélection des critères d'évaluation et le plan de travail. Je tiens également à remercier Eric Zapletal qui a travaillé avec moi sur l'intégration des données de radiothérapie de l'Hôpital Européen Georges Pompidou dans l'entrepôt de données cliniques. Je remercie Julien Taïeb, Anne Berger, Stanislas Chaussade, Romain Coriat, Bertrand Dousset et Bernard Nordlinger de m'avoir permis d'utiliser les données des patients que nous avons conjointement traités au sein des différents services. Ce travail illustre également l'importance de la multidisciplinarité dans la prise en charge en oncologie.

Je remercie également les patients eux-mêmes qui participent, par l'utilisation de leur donnée, à la recherche médicale.

Ce travail n'aurait pu se concrétiser sans le soutien de Martin Housset, puis de Catherine Durdux, qui m'ont permis de consacrer le temps nécessaire à la bonne conduite de ces recherches, en parallèle de mes activités cliniques, sans jamais interférer et en m'assurant de parfaites conditions de travail.

Enfin, je remercie chaleureusement ma famille et mes amis pour leur soutien et leur encouragement.

Table des matières

Remerciements	3
Table des matières.....	4
Résumé	7
Abstract.....	8
Productions scientifiques liées à la thèse	9
Liste des principales abréviations.....	11
Chapitre 1 - Introduction.....	12
1.1 <i>Contexte</i>	12
1.2 <i>Prise en charge standard du cancer du rectum localement avancé</i>	13
1.3 <i>Données nécessaires pour la prédiction de la réponse thérapeutique après radiochimiothérapie pré-opératoire</i>	15
Chapitre 2 – Sources des données utilisées dans le modèle.....	17
2.1 <i>Données cliniques et biologiques</i>	17
2.2 <i>Données radiomics</i>	18
2.3 <i>Données de radiothérapie</i>	20
2.3.1 La planification du traitement	20
2.3.2 La réalisation de la radiothérapie.....	21
2.4 <i>Prérequis à l'intégration des données de radiothérapie</i>	22
2.4.1 Ontologies de l'anatomie humaine.....	23
2.4.2 Unified Medical Language System	23
2.4.3 Ontologies en radiothérapie.....	24
2.4.4 Radiation Oncology Structure Ontology.....	24
2.5 <i>Article 1 : Labeling for Big Data in radiation oncology: The Radiation Oncology Structures Ontology</i>	25

2.6	<i>Perspectives</i>	38
Chapitre 3 - Méthodes de Machine Learning appliquées en oncologie radiothérapie 39		
3.1	<i>Méthodes d'intelligence artificielle en oncologie radiothérapie</i>	39
3.1.1	Machines à vecteur de support (Support Vector Machine, SVM)	42
3.1.2	Réseaux neuronaux (Artificial Neural Network, ANN)	42
3.1.3	Apprentissage profond (Deep Learning, DL)	44
3.1.4	Apprentissage supervisé (Supervised learning, SL) ou non-supervisé (Unsupervised Learning, UL)	45
3.2	<i>Systèmes d'aide à la décision</i>	45
3.3	<i>Article 2 : Big Data and machine learning in radiation oncology: state of the art and future prospects</i>	46
Chapitre 4 - Création d'un modèle prédictif par Deep Learning		
4.1	<i>Introduction</i>	55
4.2	<i>Méthodes</i>	55
4.2.1	Description des données	55
4.2.2	Extraction, contrôle qualité et filtrage des variables de radiomics	56
4.2.3	Création du réseau neuronal profond, entraînement et validation	57
4.3	<i>Résultats</i>	59
4.3.1	Variables radiomics	59
4.3.2	Prédiction de la réponse complète	61
4.4	<i>Article 3 : Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer</i>	64
Chapitre 5 - Discussion		
5.1	<i>Comparaison aux autres modèles prédictifs par machine learning</i>	73
5.2	<i>Intelligence Artificielle et médecine</i>	76
5.2.1	Un processus entamé il y a 40 ans	76
5.2.2	Deep Learning et perception	77
5.2.3	Le futur rôle du médecin	78
Chapitre 6 - Conclusion		
		79

Bibliographie.....	80
Annexes.....	94
<i>Annexe 1 – Liste des variables radiomics extraites par IBEX.....</i>	<i>94</i>
<i>Annexe 2 – Code python pour créer un DNN sous TensorFlow</i>	<i>96</i>

Résumé

L'utilisation de systèmes informatiques pour formaliser, organiser et planifier le traitement des patients a abouti à la création et à l'accumulation de quantité importante de données. Ces informations comprennent des caractéristiques démographiques, socio-économiques, cliniques, biologiques, d'imagerie, et, de plus en plus, génomiques. La médecine et sa pratique, fondées sur la sémiologie et la physiopathologie, vont être profondément transformées par ce phénomène. La complexité et la quantité des informations à intégrer pour prendre une décision médicale pourrait dépasser rapidement les capacités humaines. Les techniques d'intelligence artificielle pourraient assister le médecin et augmenter ses capacités prédictives et décisionnelles.

La première partie de ce travail présente les types de données désormais accessibles en routine en oncologie radiothérapie. Elle détaille les données nécessaires à la création d'un modèle prédictif. Nous explorons comment exploiter les données spécifiques à la radiothérapie et présentons le travail d'homogénéisation et de conceptualisation qui a été réalisé sur ces données, notamment via la création d'une ontologie, dans le but de les intégrer à un entrepôt de données.

La deuxième partie explore différentes méthodes de machine learning : k-NN, SVM, ANN et sa variante, le Deep Learning. Leurs avantages et inconvénients respectifs sont évalués avant de présenter les études ayant déjà utilisé ces méthodes dans le cadre de la radiothérapie.

La troisième partie présente la création d'un modèle prédictif de la réponse complète à la radiochimiothérapie (RTCT) pré-opératoire dans le cancer du rectum localement avancé. Cette preuve de concept utilise des sources de données hétérogènes et un réseau neuronal profond dans le but d'identifier les patients en réponse complète après RTCT qui pourraient ne pas nécessiter de traitement chirurgical radical.

Cet exemple, qui pourrait en pratique être intégré aux logiciels de radiothérapie déjà existant, utilise les données collectées en routine et illustre parfaitement le potentiel des approches de prédiction par IA pour la personnalisation des soins.

Mots-Clés : oncologie ; radiothérapie ; prédiction ; Deep Learning

Abstract

The use of Electronic Health Records is generating vast amount of data. They include demographic, socio-economic, clinical, biological, imaging and genomic features. Medicine, which relied on semiotics and physiopathology, will be permanently disrupted by this phenomenon. The complexity and volume of data that need to be analyzed to guide treatment decision will soon overcome the human cognitive abilities. Artificial Intelligence methods could be used to assist the physicians and guide decision-making.

The first part of this work presents the different types of data routinely generated in oncology, which should be considered for modelling a prediction. We also explore which specific data is created in radiation oncology and explain how it can be integrated in a clinical data warehouse through the use of an ontology we created.

The second part reports on several types of machine learning methods: k-NN, SVM, ANN and Deep Learning. Their respective advantages and pitfalls are evaluated. The studies using these methods in the field of radiation oncology are also referenced.

The third part details the creation of a model predicting pathologic complete response after neoadjuvant chemoradiation for locally-advanced rectal cancer. This proof-of-concept study uses heterogeneous sources of data and a Deep Neural Network in order to find out which patient could potentially avoid radical surgical treatment, in order to significantly reduce the overall adverse effects of the treatment.

This example, which could easily be integrated within the existing treatment planning systems, uses routine health data and illustrates the potential of this kind of approach for treatment personalization.

Key words: oncology ; radiation oncology ; prediction ; Deep Learning

Productions scientifiques liées à la thèse

Articles publiés directement liés à la thèse

1. Bibault J-E, Giraud P, Burgun A. 2016. « Big Data and Machine Learning in Radiation Oncology: State of the Art and Future Prospects ». *Cancer Letters*, 382(1):110-117
2. Bibault J-E, Burgun A, Giraud P. 2017. « Artificial intelligence applied to radiation oncology ». *Cancer Radiothérapie*, 21 (3): 239-43.
3. Bibault J-E, Zapletal E, Rance B, Giraud P, Burgun A. 2018. « Labeling for Big Data in Radiation Oncology: The Radiation Oncology Structures Ontology ». *PloS One* 13 (1): e0191263.
4. Zapletal E, Bibault J-E, Giraud P, Burgun A. 2018. « Integrating Multimodal Radiation Therapy Data Into i2b2 ». *Appl Clin Inform*. 2018 Apr;9(2):377-390. doi: 10.1055/s-0038-1651497.
5. Bibault J-E, Giraud P, Durdux C, Taïeb J, Berger A, Coriat R, Chaussade S, Dousset B, Nordlinger B, Burgun A. « Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer » – *Scientific Reports* volume 8, Article number: 12611 (2018)

Articles publiés indirectement liés à la thèse

1. Bibault J-E, Tinhofer I. 2017. « The role of Next-Generation Sequencing in tumoral radiosensitivity prediction », *Clinical and Translational Radiation Oncology*, 16–20.
2. Foy J-P, Durdux C, Giraud P, Bibault J-E. 2018. « RE: The Rise of Radiomics and Implications for Oncologic Management », *Journal of the National Cancer Institute*

Chapitre de livre

1. Bibault J-E, Giraud P, Burgun A. « Big data in radiation oncology: opportunities and challenges ». A paraître en 2019 dans *Big Data in radiation oncology*, éditions Taylor & Francis Book

Communications orales

1. Bibault J-E, Giraud P, Burgun A, « Prédiction par Deep Learning de la réponse complète après radiochimiothérapie néoadjuvante dans le cancer du rectum localement avancé », *29^{ème} Congrès de la Société Française d'Oncologie Radiothérapie (SFRO)*, Paris, Octobre 2018

Communications affichées

1. Bibault J-E, Giraud P, Burgun A, « Deep Learning predicts complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer », *37th Annual Meeting of the European Society for Radiotherapy and Oncology*, Barcelone, Avril 2018

Conférences invitées

1. Bibault J-E, « Comment les algorithmes ont révolutionné leur métier », *Conférence Le Monde O21*, Mars 2017, Paris
2. Bibault J-E, « Deep Learning in Medicine – state of the art », *La Paillasse Innovation in Healthcare Conference*, Mars 2017, Paris
3. Bibault J-E, A. Burgun, P. Giraud, « Applied artificial Intelligence in radiation oncology », *Artificial intelligence for medical imaging and precision medicine Conference*, Mai 2017, Institut Curie, Paris
4. Bibault J-E, A. Burgun, P. Giraud, « Radiomics biomarkers for prediction in radiation oncology » *Healthcare data integration methodologies, France Life Imaging Conference*, Septembre 2017, Orsay
5. Bibault J-E, A. Burgun, P. Giraud, « Predicting Treatment response with heterogeneous data », *Les Rencontres de la Cancérologie Française*, Novembre 2017, Paris

Liste des principales abréviations

ANN : Artificial Neural Network

APHP : Assistance Publique - Hôpitaux de Paris

API : Application Programming Interface

AUC : Area Under the Curve

CDW : Clinical Data Warehouse (Entrepôt de données cliniques)

CT : Computed Tomography

CTV : Clinical Target Volume

DL : Deep Learning

DT : Decision Tree

EHR : Electronic Health Record

GPU : Graphics Processing Unit

GTV : Gross Tumor Volume

HEGP : Hôpital Européen Georges Pompidou

DMI : Dossier Médical Informatisé

DNN : Deep Neural Network

HDV : Histogramme Dose Volume

IA : Intelligence Artificielle

IC 95% : Intervalle de confiance à 95%

IRM : Imagerie par Résonance Magnétique

kNN : k-Nearest Neighbor

ML : Machine Learning

OAR : Organe à risque

P : Degré de signification d'un test d'hypothèse statistique

PTV : Planning Treatment Volume

RCMI : Radiothérapie Conformationnelle avec Modulation d'Intensité

RL : Régression Logistique

ROC : Received Operating Characteristic

ROI : Region of Interest

ROS : Radiation Oncology Structures

RTCT : Radiochimiothérapie

SVM : Support Vector Machine

Chapitre 1 - Introduction

1.1 Contexte

L'augmentation significative des paramètres cliniques et biologiques à intégrer dans un essai thérapeutique, à l'heure de la médecine de précision, rend la réalisation d'essais dédiés impossible (C. Chen et al. 2015). Mener des études cliniques de phase 3 en radiothérapie, comparant différents schémas de traitement ou différentes techniques (typiquement un traitement en technique conformationnelle ou avec modulation d'intensité) peut se révéler difficile, voire impossible dans certains cas, soit parce que la technique ancienne n'est plus disponible, soit parce qu'il ne serait pas éthique de proposer une essai randomisé comparant ces techniques. Enfin, il existe une différence significative entre les populations de patients inclus dans les essais cliniques et les patients rencontrés dans la pratique clinique (Geifman et Butte 2016). Par exemple, les patients âgés, pour lesquels presque aucun essai dédié n'est disponible, constituent pourtant une part importante de la population à traiter.

De nouvelles approches sont donc nécessaires afin de générer des études de bon niveau de preuve et l'une de ces approches est la constitution et l'exploitation de bases de données massives détaillées, intégrant un grand nombre de paramètres (Skripcak et al. 2014). Au sein de ces données, la biologie moléculaire et la génomique sont appelées à jouer un rôle prépondérant. La prédiction de la radiosensibilité des tissus sains et des tumeurs devra intégrer l'ensemble de l'analyse du génome des patients (Georgakilas et al. 2015). La complexité des modèles prédictifs obtenus rendra impossible la simple analyse par un humain, puisque l'on considère que les capacités d'intégration cognitives humaines permettent de prendre en compte jusque cinq facteurs mais qu'en 2020, une décision médicale pourra reposer sur 10000 paramètres à la fois pour un unique patient (Abernethy et al. 2010).

Alors que le prix du séquençage génétique a continuellement baissé en parallèle du prix de la puissance de calcul informatique (Mardis 2011; Metzker 2010), le seul facteur nous empêchant de générer ces modèles prédictifs de précision est l'existence de large cohortes de patients dont le profil phénotypique précis serait connu. La généralisation du dossier médical informatisé (DMI) pourrait apporter une

solution à ce problème, avec l'utilisation des « omics », de large quantité de données génétiques, ou d'imagerie. La science des données (« Data Science »), qui correspond à un sous-domaine de la bioinformatique, concerne l'ensemble des moyens permettant d'exploiter ces données et de générer une prédiction afin de guider les traitements (Kohane, Drazen, et Campion 2012).

Après des étapes de préparation et de sélection des données, l'utilisation d'algorithmes d'intelligence artificielle permet déjà de construire des modèles prédictifs en médecine. Ce travail visait à exploiter des données médicales d'origine hétérogène pour créer un profil complet des patients et exploiter ce profil par des méthodes de machine learning (ou apprentissage statistique) pour prédire l'efficacité thérapeutique. L'utilisation du Deep Learning, en association avec la radiomics, n'ayant pas fait l'objet d'études déjà publiées dans ce contexte, nous avons pour but de proposer une preuve de concept de la faisabilité de la combinaison de ces deux méthodes innovantes pour prédire la réponse complète après radiochimiothérapie pré-opératoire du cancer du rectum localement avancé.

1.2 Prise en charge standard du cancer du rectum localement avancé

On estime qu'à partir en 2020, 45000 adénocarcinomes du rectum seront diagnostiqués chaque année en France. Ces carcinomes peuvent être traités par une résection du rectum seul (TME, total mesorectal excision) lorsque l'extension est limitée au mésorectum. Leur pronostic est aujourd'hui meilleur que les cancers du côlon avec une survie à cinq ans approchant les 90% (Park et al. 2014) et un taux de récurrences inférieur à 5% (Sebag-Montefiore et al. 2009). Environ 8% des patients présentent des cancers localement avancés avec une tumeur s'étendant au fascia mésorectal (Ptok et al. 2013).

Au-delà du stade tumoral, la marge de résection circonférentielle (CRM, circumferential resection margin), qui décrit la distance la plus courte entre la tumeur et le fascia mésorectal, a une grande importance sur les plans thérapeutique et pronostique. Un traitement néoadjuvant est recommandé pour toutes les tumeurs de stade cT4 (dépassement du fascia mésorectal) ou cT3 avec une CRM compromise

< 2 mm. Presque tous les cancers du bas rectum cT3 ont une CRM compromise, car le mésorectum devient très fin près de l'anus.

Une fois la décision d'une thérapie néoadjuvante prise, la radiothérapie peut être effectuée selon un schéma dit « long », 50,4 Gy (28 fractions de 1,8 Gy) en association avec de la Capecitabine les jours de séances. Quatre à six semaines après la thérapie néoadjuvante terminée, un nouveau bilan radiologique (restaging) par IRM pelvienne et CT-scan thoraco-abdominal sera réalisé (National Comprehensive Cancer Network s. d.). Il donnera des renseignements sur la réponse thérapeutique et sur l'étendue tumorale pour la planification de la chirurgie. Le patient est ensuite opéré par résection radicale du mésorectum (figure 1).

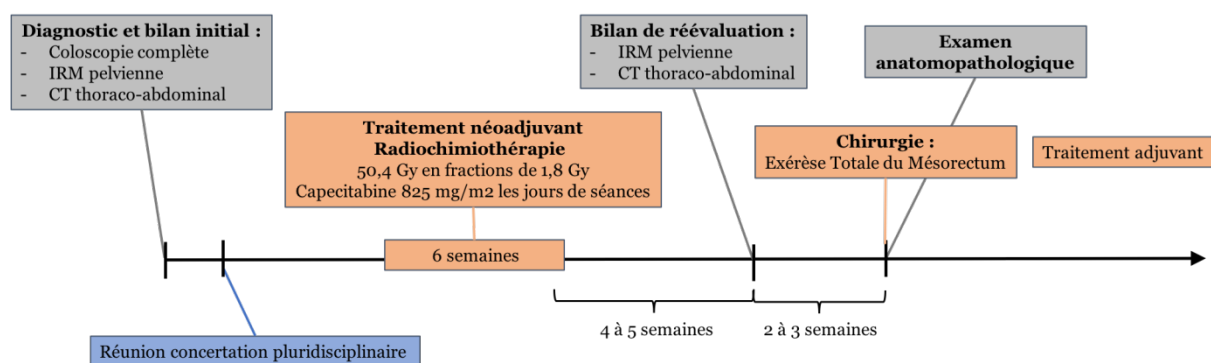


Figure 1. Schéma global de la prise en charge d'un adénocarcinome localement avancé du rectum

Dans 20 à 30% des cas, aucun résidu tumoral n'est retrouvé sur la pièce opératoire (Maas et al. 2010). La difficulté est de prédire quel patient sera en réponse histologique complète (pathologic Complete Response, pCR) sans faire la résection radicale du rectum. En effet pour ce groupe de patients, des approches conservatrices (watch-and-wait ou résection transanale) ont montré de bons résultats à long terme, sans avoir les effets secondaires lourds d'une chirurgie radicale (Habr-Gama et al. 2004, 2014; Appelt et al. 2015; Renehan et al. 2016; Martens et al. 2016).

1.3 Données nécessaires pour la prédiction de la réponse thérapeutique après radiochimiothérapie pré-opératoire

La quantité et la qualité des données sont primordiales afin de pouvoir construire un modèle prédictif efficace. Lambin et al. ont ainsi défini les types de données à utiliser (Lambin, van Stiphout, et al. 2013) comme suit :

- Cliniques (performance status, histologie, biologie, questionnaires patients)
- Thérapeutiques (distribution temporelle et spatiale de la dose)
- Imagerie (stade, taille tumorale, données métaboliques, radiomics)
- Biologie moléculaire (radiosensibilité intrinsèque, hypoxie, index de prolifération et plus globalement génomique).

Un dossier médical complet représenterait, au bilan initial d'un patient, environ 7 Go de données, avec les données génomiques correspondant à environ 70% du total (tableau 1).

Type	Format	Taille
Clinique	Texte	10 Mo
Résultats biologiques	Chiffres	1 Mo
Administrative	Codes ICD-10	1 Mo
Imagerie	DICOM	450 Mo
Radiothérapie (dosimétrie et record-and-verify)	DICOM, RT-DICOM	500 Mo
Génomique	BAM : Position, base, qualité	6 Go
Total		7,9 Go

Tableau 1. Types de données d'un dossier médical digital et taille approximative pour un seul patient.

Ces données doivent être anonymisées et intégrées dans un système unifié, indépendant du système de soin. Des procédures adaptées d'accès sécurisé et authentifié, au sein de l'hôpital et depuis l'extérieur, doivent être mises en place pour

faciliter leur utilisation à des fins de recherche. Des solutions d'intégration existent déjà et sont utilisées dans plusieurs hôpitaux (Canuel et al. 2015). Malgré les avancées technologiques, certains auteurs estiment que la quantité de données générées dépassera bientôt les capacités de stockage des institutions. Une solution à ce problème pourrait être de gérer ces données comme les dossiers médicaux classiques, en archivant les dossiers anciens sur des espaces de stockage externalisés (Huser et Cimino 2015).

Enfin, au-delà des données générées durant le bilan thérapeutique et pendant le traitement, le suivi des patients devra pouvoir être intégré de façon dynamique, y compris en exploitant des questionnaires au patient (patient-reported outcome) en ligne ou mobile (Denis et al. 2015; Falchook et al. 2016).

Chapitre 2 – Sources des données utilisées dans le modèle

2.1 Données cliniques et biologiques

Depuis son ouverture en 2001, l'hôpital Européen Georges Pompidou a adopté le dossier médical informatisé (Electronic Health Records, EHR). La faisabilité de l'utilisation secondaire des données afin de mener des recherches cliniques a déjà été démontrée (Danciu et al. 2014). L'HEGP a créé son entrepôt de données (Clinical Data Warehouse, CDW) en 2008, en utilisant la plateforme i2b2 (Zapletal et al. 2010a; Jannot et al. 2017). Depuis, plusieurs sources de données ont été progressivement intégrées (figure 2) :

- Données démographiques (sexe, âge, status familial, status vital, ...)
- Résultats biologiques et histologiques
- Prescriptions médicamenteuses et validations pharmaceutiques
- Diagnostics
- Procédures
- Parcours patient
- Formulaires structurés de l'EHR
- Comptes-rendus d'imagerie
- Paramètres cliniques (fréquence cardiaque, température, saturation en oxygène)
- Prescriptions de chimiothérapie

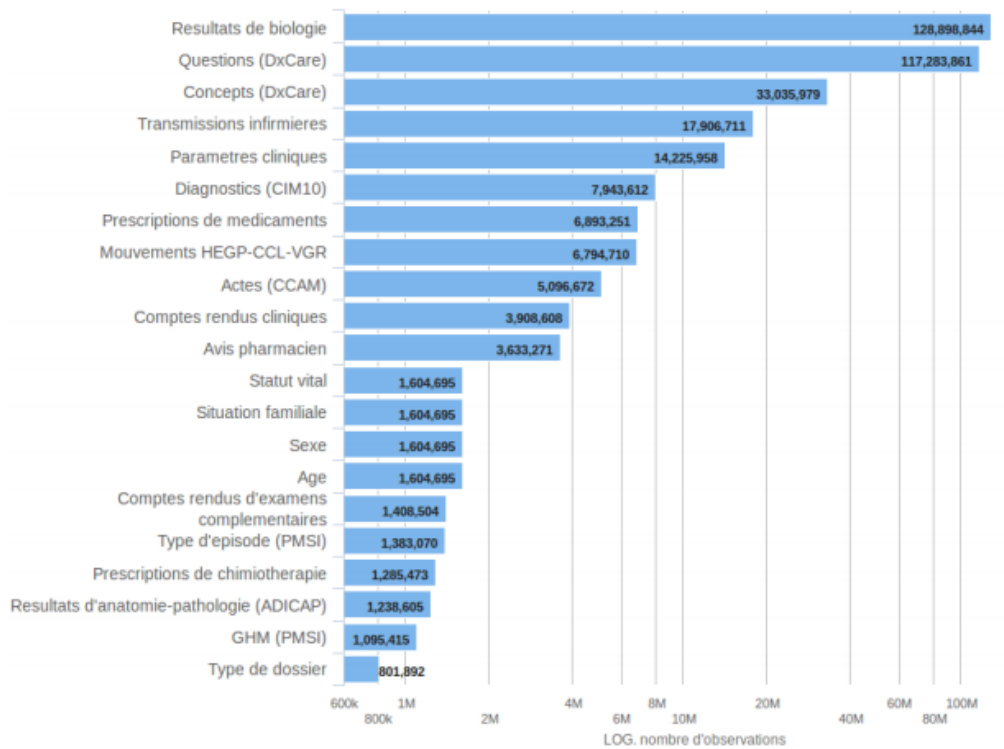


Figure 2. Nombre d’observations par catégorie de données dans le CDW de l’HEGP au 31 Janvier 2017

Les données liées à la prise en charge de patients dans les services d’oncologie de l’HEGP ont un intérêt particulier, notamment dans le cadre de projets du CARPEM (Cancer Research and PErsonalized Medicine), l’un des 8 SIRICS (Site de Recherche Intégré sur le Cancer) labellisé en France par l’Institut Nationale du Cancer. Cependant, les données spécifiques, générées par le service de radiothérapie, n’étaient pas intégrées au CDW.

2.2 Données radiomics

La radiomics est un ensemble de techniques statistiques qui permettent d’extraire à haut débit de grande quantité d’information quantitative à partir des imageries morphologiques (scanner et IRM) et métaboliques (TEP) (Lambin et al. 2012). Les images deviennent ainsi des données (Gillies, Kinahan, et Hricak 2016; Lambin et al. 2017). Il ne s’agit plus de décrire les caractéristiques d’un volume (forme, taille, prise de contraste), mais d’en extraire des valeurs numériques.

Ces techniques peuvent être appliquées sur les examens réalisés en routine en radiologie, en radiothérapie et en médecine nucléaire. Après l’acquisition d’images,

les volumes d'intérêt sont segmentés et les variables descriptives sont ensuite extraites avec un logiciel dédié. Plusieurs logiciels sont disponibles : en accès libre (Lifex, IBEX) ou payant (OncoRadiomics). Une des limites de la radiomics est l'absence de standardisation et d'assurance qualité de ces logiciels, ce qui entraîne la variabilité des résultats obtenus d'une extraction réalisée à partir des mêmes images par différents logiciels (Gan J et al. 2016; Emaminejad et al. 2018).

Les variables correspondent à des densités Hounsfield sur un scanner, une intensité en IRM ou la SUV en TEP. Elles peuvent être de plusieurs ordres statistiques :

- Premier ordre :
 - Histogramme
 - Médiane, maximum, minimum
 - Randomness (entropy)
 - Asymétrie
 - Kurtosis (flatness)
- Deuxième ordre :
 - Gray Level Co-Occurrence Matrix
 - Gray Level Size Zone Matrix (GLSZM)
 - Gray Level Run Length Matrix (GLRLM)
 - Neighbouring Gray Tone Difference Matrix (NGTDM)
 - Gray Level Dependence Matrix (GLDM)

Ces descripteurs peuvent être extraits sur un plan de coupe (2D) ou sur une imagerie volumétrique (3D). Il est donc possible d'extraire à partir de l'imagerie d'un patient plusieurs centaines voire milliers de descripteurs. Le rôle de ces descripteurs, ou de leur association, peut ensuite être exploré afin de définir une signature qui pourrait être utilisée comme biomarqueur pronostique ou prédictif de la maladie. Ce type de méthode est désormais largement utilisé en cancérologie dans de nombreuses études (Lambin et al. 2017). Les méthodes explorant les associations linéaires ou

non-linéaires entre ces descripteurs et la variable d'intérêt (survie, contrôle local) sont variables et incluent les tests statistiques classiques ou l'apprentissage machine (Parmar, Grossmann, Bussink, et al. 2015). L'ensemble de la méthodologie est schématisé en figure 3.

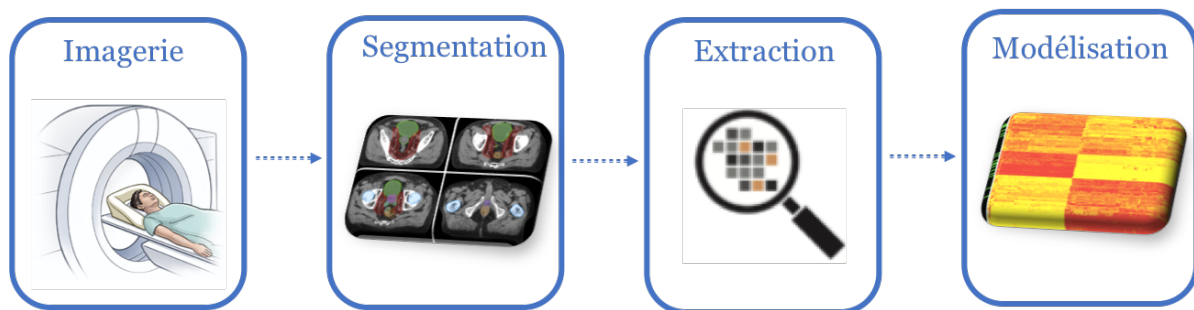


Figure 3. Processus de modélisation par radiomique

2.3 Données de radiothérapie

Au-delà des données classiquement incorporées dans un modèle prédictif (antécédents, symptômes, résultats de biologie et d'imagerie), l'utilisation des caractéristiques précises des traitements réalisés en médecine pourrait être envisagée. La radiothérapie est l'une des rares disciplines médicales où le traitement administré est documenté de façon précise et complète. L'ensemble des séances étant administrées via un système informatique qui enregistre en permanence les événements, il est possible d'extraire et d'analyser la pertinence de ces informations pour effectuer une prédiction.

Deux sources d'informations peuvent être exploitées : la planification et la réalisation de la radiothérapie

2.3.1 La planification du traitement

Avant de débuter sa radiothérapie, tout patient bénéficie d'un scanner. L'imagerie tridimensionnelle ainsi obtenue est ensuite importée sur le système de planification. Chaque cible (la tumeur) et chaque organe est alors délinéé

manuellement par l'oncologue (figure 4) dans un espace en trois dimensions (X, Y, Z). Une dosimétrie est ensuite réalisée pour administrer la dose prescrite sur la tumeur, sans dépasser certains seuils de dose sur les organes autour (on parle de contraintes de dose). Ainsi, pour une même maladie, le traitement planifié différera significativement selon l'anatomie du patient. Cette variabilité se retrouve dans les plans de traitement et les histogrammes dose volumes, qui sont enregistrés numériquement dans le système pour tout patient.



Figure 4. Délimitation d'une tumeur du rectum sur scanner (A) et IRM fusionnée (B)

2.3.2 La réalisation de la radiothérapie

Une fois le traitement planifié, le patient peut débuter les séances. Classiquement, un traitement est réalisé sur un accélérateur de particule comportant une table sur laquelle est positionné le patient et un bras dans lequel se trouve l'accélérateur (figure 5). Un faisceau d'électrons est dirigé sur une cible de tungstène, ce qui génère des photons de haute énergie. La trajectoire de ces photons sera modifiée et contrôlée pour la façonner et traiter la tumeur. Le patient vient ainsi tous les jours de la semaine pour réaliser une séance d'une durée approximative de 10 à 15 minutes.

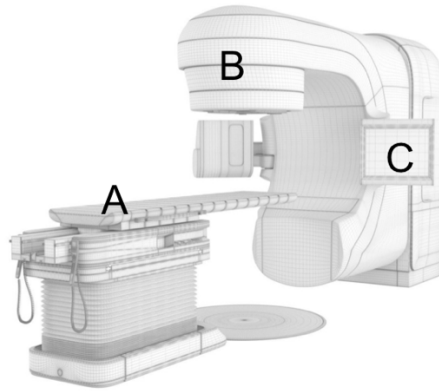


Figure 5. Accélérateur linéaire de particules. A : table de traitement ; B : Tête de l'accélérateur ; C : Système d'imagerie embarquée

Le positionnement du patient sous la machine fait l'objet d'un contrôle quotidien qui peut occasionner des corrections dans les trois axes de l'espace (+/- les rotations) afin d'être sûr d'irradier la bonne zone. L'ensemble de ces paramètres (date, heure des séances, décalage table) sont enregistrés de façon automatique dans un système informatique appelé « Record and Verify » qui permet de contrôler et de quantifier la correspondance entre le traitement planifié et celui effectivement réalisé. Cette particularité rend l'exploitation des données en radiothérapie particulièrement intéressante.

2.4 Prérequis à l'intégration des données de radiothérapie

L'intégration des données de radiothérapie dans le CDW de l'HEGP nécessite au préalable d'uniformiser l'ensemble des informations. Si la date, l'heure et la durée sont déjà stockées de façon homogène dans le Record and Verify, les noms des structures anatomiques qui ont été délinéées pour l'ensemble des patients sont très hétérogènes. En effet, les nombreuses personnes intervenant au fil de la préparation du traitement n'utilisent pas la même nomenclature. Cela rend impossible l'intégration et l'exploitation des données anatomiques et dosimétriques, pourtant primordiales. Dans un premier temps, il a donc été nécessaires d'extraire les noms des structures délinéées sur l'ensemble des patients traités dans le service de radiothérapie de l'HEGP afin de les homogénéiser et de les intégrer dans le CDW. Nous avons donc cherché à utiliser des outils et des ontologies déjà existants pour ce faire.

Une ontologie est une unité structurée des termes et concepts propres à un domaine. Elle comprend également les relations existantes entre ces concepts. Elle repose sur un formalisme logique, les logiques de description. En imagerie et en anatomie, plusieurs ontologies existaient préalablement à ce travail.

2.4.1 Ontologies de l'anatomie humaine

La FMA (The Foundational Model of Anatomy) a été développée par l'Université de Washington afin de décrire l'anatomie humaine de façon exhaustive (Noy et al. s. d.). Elle se compose de 4 parties reliées :

- 1) *Taxonomie anatomique (Anatomy taxonomy, At)*, qui classifie les entités anatomiques en fonction de leur caractéristique commune et de leurs différences
- 2) *Abstraction des structures anatomiques (Anatomical Structural Abstraction, ASA)*, qui spécifie les relations spatiales qui existent entre les entités de At
- 3) *Abstraction des transformations anatomiques (Anatomical Transformation Abstraction, ATA)* qui comprend les transformations des entités au cours du développement pré et postnatal
- 4) Métasavoir (Metaknowledge, Mk) qui inclut les principes, règles et définitions régissant les relations entre les entités.

La FMA contient 75000 classes et plus de 120000 termes, avec plus de 168 relations différentes. Même si elle est très détaillée, elle n'est pas adaptée au domaine de la radioanatomie qui correspond à une description par coupe dans un espace en trois dimensions. Ainsi, la Société Américaine de Radiology (RSNA) a développé une ontologie spécifique, RadLex, qui incorpore l'ensemble des informations anatomiques nécessaires à la radiologie (Rubin 2008). Par extension, RadLex a aussi été utilisée pour annoter les examens de médecine nucléaire tels que la Tomographie par Emission de Positons (Hwang et al. 2017).

2.4.2 Unified Medical Language System

La FMA est une des ontologies intégrées dans le Unified Medical Language System (UMLS). Ce système, créé en 1986, constitue une structure commune permettant de faire correspondre les termes de différentes terminologies ayant la même

signification. Il a été créé afin de notamment pouvoir unifier les concepts utilisés en biomédecine dans différentes base de données et systèmes d'informations. Le métathésaurus, qui constitue la base de l'UMLS, comporte un million de concepts biomédicaux et cinq millions de termes, parmi lesquels on retrouve l'ICD-10 (codes de diagnostics), la FMA, le SNOMED-CT (ontologies en radiologie) et le DSM-IV (codes des pathologies en psychiatrie).

2.4.3 Ontologies en radiothérapie

L'absence d'ontologie dédiée à la radiothérapie a nécessité la création par Dekker et al de la Radiation Oncology Ontology (ROO) (« Radiation Oncology Ontology - Summary | NCBO BioPortal » s. d.), qui réutilise les entités de la FMA et ajoute certains concepts absents, propres à la radiothérapie, tels que les volumes cibles (GTV, CTV, PTV). Pourtant, la ROO n'est plus mise à jour depuis Juillet 2015 et n'inclut pas les derniers concepts récemment apparus en radiothérapie ou les nomenclatures et conventions publiées. Ainsi les aires ganglionnaires cervicales, qui doivent être délinéées par l'oncologue radiothérapeute pour tout patient traité pour un cancer des voies aéro-digestives supérieures, ne sont pas détaillées. Il existe pourtant un atlas spécifique à la radiothérapie, non présente dans la FMA ou RadLex, décrivant précisément ces aires (Grégoire et al. 2014; Rusch et al. 2009). Il est dès lors impossible d'explorer le rôle prédictif de niveau de dose de radiothérapie administrée sur chaque aire si chaque aire n'est pas clairement individualisée dans l'ontologie servant à intégrer les données de radiothérapie.

2.4.4 Radiation Oncology Structure Ontology

Nous avons donc choisi de développer et de rendre publique une ontologie s'appuyant sur le standard de la FMA et intégrant l'ensemble des derniers concepts et recommandations en radiothérapie. Afin d'obtenir un haut niveau de détails, ce qui permettra de faire des analyses fines, chaque classe est divisée de façon hiérarchique en sous-classes. Les aires ganglionnaires cervicales sont donc divisées selon leur latéralité droite/gauche, puis leur zone (de I à XIII). Pour la création de l'ontologie, nous avons adopté une méthodologie de bas en haut, pour définir les concepts puis de haute en bas pour les rassembler et les hiérarchiser en classes. Nous avons ainsi extrait tous les noms de structures délinéées existants dans notre système de

planification grâce à l'API (Application Programming Interface) de VARIAN (ARIA, Varian, Palo Alto, Californie, Etats-Unis). Il existait au total 20758 noms de structures différents qui ont été rassemblés en 417 classes dans notre ontologie. Nous avons utilisé le logiciel open source Protégé 5.2.0, pour créer l'ontologie (Musen 2015, s. d.) et l'avons ensuite distribuée avec une licence Apache 2 sur BioPortal. L'ensemble de l'ontologie a été mappée sur l'UMLS.

2.5 Article 1 : *Labeling for Big Data in radiation oncology: The Radiation Oncology Structures Ontology*

2.6 Perspectives

L'ontologie que nous avons créée est utilisée pour intégrer l'ensemble des données dosimétriques du service de radiothérapie dans l'entrepôt de données cliniques (Clinical Data warehouse, CDW) de l'HEGP (figure 6).

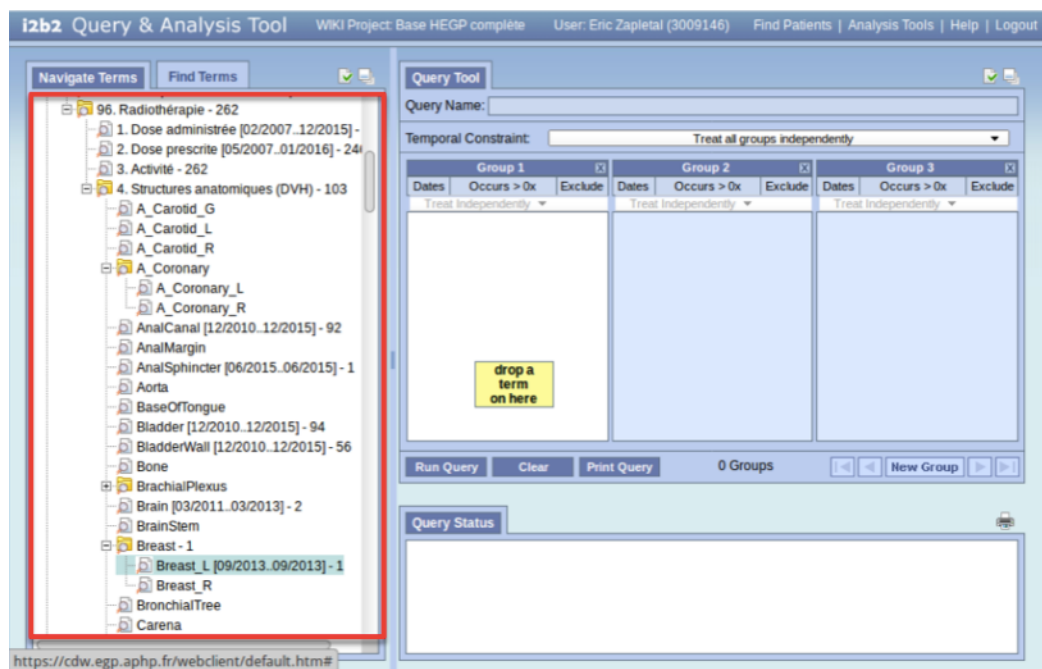


Figure 6. Interface du CDW sous i2b2

Il sera ainsi possible de faire une requête sur tous les patients traités dans le service. Un exemple d'applications concrète pourrait consister à croiser les patients traités à l'HEGP pour un infarctus du myocarde avec ceux traités par radiothérapie au niveau du thorax pour rechercher la dose cardiaque reçue.

A plus long terme, l'ontologie ROS pourrait être utilisée pour intégrer les données de l'ensemble des services de radiothérapie de l'AP-HP dans l'entrepôt de données qui rassemble déjà les données médicales de 6,5 millions de patients, traités dans les 39 hôpitaux de l'Assistance Publique. L'harmonisation des données grâce à notre ontologie permettra de disposer d'une base de données de qualité pour mener des études et développer des algorithmes d'IA en oncologie radiothérapie.

Chapitre 3 - Méthodes de Machine Learning appliquées en oncologie radiothérapie

3.1 Méthodes d'intelligence artificielle en oncologie radiothérapie

La création de modèle prédictif se déroule en deux étapes : la qualification et la validation. Un premier ensemble de données est utilisé pour déterminer un ou plusieurs facteurs prédictifs. Lorsque ces facteurs ont été définis, ils sont évalués et validés sur un second ensemble de données différent du premier. L'étape suivante consiste à utiliser le modèle ainsi créé pour guider les traitements et améliorer le devenir des patients. Si de nombreuses études ont été menées pour créer des modèles pour plusieurs localisations en radiothérapie, aucune étude n'a évalué ces modèles de façon prospective avec comme critères de jugement l'amélioration de la survie, du contrôle local ou de la qualité de vie.

Kang et al. ont proposé plusieurs principes à respecter pour la modélisation prédictive en radiothérapie (Kang et al. 2015) :

- Intégrer des facteurs dosimétriques et non-dosimétriques,
- Sélectionner et nettoyer manuellement les données avant de procéder à l'analyse automatisée,
- Sélectionner la meilleure méthode d'analyse, adaptée à l'objectif du modèle et aux données de départ,
- Prendre en compte l'interdépendance des facteurs explorés,
- Valider la généralisation du modèle avec un ensemble de données externes,
- Évaluer plusieurs modèles et comparer leur résultat à d'éventuels autres modèles connus.

Les méthodes d'apprentissage automatique suivantes ont été utilisées en oncologie :

- Arbres de décisions (*Decision Tree*, DT) (Quinlan 1986) où un algorithme simple crée des classes mutuellement exclusives en répondant à des questions prédéfinies
- Classification naïve bayésienne (*Naive Bayes*, NB) (Langley, Iba, et Thompson 1992; Langley et Sage 1994), qui permet de classer des variables selon un modèle probabiliste
- K plus proches voisins (*k-Nearest Neighbor*, *k*-NN) (Patrick et Fischer III 1970), où une entité est classifiée selon les caractéristiques de son plus proche voisin
- Machines à vecteur de support (*Support Vector Machine*, SVM) (Vapnik 1982), où un algorithme classifiera des données en fonction d'un seuil linéaire
- Réseaux neuronaux (*Artificial Neural Network*, ANN) (Rumelhart et McClelland 1986), qui utilisent des modèles inspirés des réseaux biologiques
- Apprentissage profond (*Deep Learning*, DL) (LeCun, Bengio, et Hinton 2015), une évolution des réseaux neuronaux qui comportent de multiples couches et sous-couches de neurones.

Le choix entre ces différentes méthodes doit être guidé par les avantages et inconvénients de chacune de ces méthodes (tableau 2). Nous détaillerons ci-dessous les trois méthodes qui ont été utilisées afin de créer des modèles prédictifs en radiothérapie.

Algorithme	Avantages	Limites
Arbres de décision	• facile à comprendre • rapide à réaliser	• uniquement pour des classes mutuellement exclusives • modèle dépend de l'ordre dans lequel les classes sont intégrées • risque de créer des arbres très complexes, peu utilisables
Classification naïve bayésienne	• facile à comprendre • rapide à réaliser • pas d'effet d'ordre sur l'entraînement	• les variables doivent être indépendantes • les variables numériques doivent suivre une distribution normale • uniquement pour des classes mutuellement exclusives • modèles peu précis
K plus proches voisins	• rapide et simple • tolérance au bruit et aux données manquantes • peut être utilisé pour la régression et la classification	• toutes les variables ont la même pertinence dans la modélisation • plus on intègre de variables, plus la puissance de calcul nécessaire est importante
Machines à vecteur de support	• créent des modèles robustes • limitent le risque d'erreur • modélisation de relations non-linéaires	• entraînement lent • risque slow training • risque de surapprentissage (le modèle créé ne s'applique qu'à un set de données, celui avec lequel on l'a entraîné) • le modèle créé peut être difficile à comprendre (effet de « boîte noire »)
Réseaux neuronaux Apprentissage profond	• tolérance au bruit et aux données manquantes • peut être utilisé pour la régression et la classification • peut être facilement mis à jour avec de nouvelles données	• risque de surapprentissage • le modèle créé peut être difficile à comprendre • nécessite d'ajuster le modèle pour optimiser les résultats

Tableau 2. Avantages et limites de différents algorithmes d'apprentissage automatique

3.1.1 Machines à vecteur de support (Support Vector Machine, SVM)

Les SVM permettent de classifier un nombre élevé de variables de façon non linéaire. La séparation des variables est réalisée avec une marge maximale à l'aide de fonctions de similarité (ou kernels) qui sont utilisées pour transformer les données et choisir les vecteurs de support. L'algorithme doit être entraîné sur un premier ensemble de patients pour déterminer ces vecteurs, qui seront ensuite utilisés pour classifier d'autres patients dont le devenir est inconnu.

Les SVMs ont déjà été utilisés en radiothérapie pour prédire le risque de pneumopathie radique après radiothérapie conformationnelle (S. Chen, Zhou, Yin, et al. 2007), le contrôle local (Klement et al. 2014) ou à distance (Zhou et al. 2016) après radiothérapie stéréotaxique de carcinomes bronchiques et la radiosensibilité des cancers de l'œsophage (Hayashida et al. 2005). Si ces études restent innovantes, les méthodologies restent encore très hétérogènes dans leur qualité, puisque seule l'étude la plus récente de Zhou et al. a comparé différents algorithmes d'apprentissage automatique (régression logistique, réseaux neuronaux et machines à vecteur de support) pour ne garder que celui donnant les meilleurs résultats (SVM) (Zhou et al. 2016).

Idéalement, au-delà de comparer les algorithmes d'IA entre eux, leurs résultats devraient également être comparés à des méthodes de statistique classique, ce qui n'est quasiment jamais fait dans les études publiées. Seule l'étude de Yahya et al., portant sur la prédiction des symptômes urinaires après radiothérapie prostatique, a comparé différentes méthodes, dont les SVMs et la régression logistique. Elle a montré que la régression adaptative (Multivariate Adaptive Regression Splines, MARS) avait les meilleures performances prédictives, même si elles restaient modestes ($AUC=0.649 \pm 0.074$ pour la dysurie de grade ≥ 1).

3.1.2 Réseaux neuronaux (Artificial Neural Network, ANN)

Les réseaux neuronaux sont constitués de couches de neurones successives. On attribue à chaque neurone une pondération qui définit son importance dans l'algorithme final. Chaque couche reçoit les données de la couche supérieure et les passent à la couche suivante. Lorsqu'on utilise un réseau neuronal, on attribue

d'abord des poids aléatoires à chaque neurone et l'on ajuste ces poids en fonction des résultats (par exemple une classification) obtenus après chaque calcul itératif. Les ANNs ont été plus souvent utilisés en médecine et en radiothérapie que les SVM. Une étude a par exemple utilisé un réseau neuronal à trois couches pour prédire la survie globale de patients traités par radiothérapie ou radiochimiothérapie pour un carcinome ORL (Bryce et al. 1998). Ce réseau intégrait 14 variables et a été entraîné pendant un millier d'itérations. Ce réseau a permis d'obtenir des résultats de prédiction meilleurs en intégrant un plus grand nombre de variables qu'une régression logistique classique (3 vs 5 variables, $AUC=0.67\pm0.05$ vs 0.78 ± 0.05 pour la LR et l'ANN respectivement). Une étude plus récente a utilisé l'apprentissage automatique appliqué à la radiomics dans les cancers ORL (Parmar, Grossmann, Rietveld, et al. 2015). Une première cohorte de 110 patients a été utilisée pour entraîner l'ANN puis une seconde cohorte de 95 patients a servi à la validation du modèle. Quatre cent quarante variables radiomics ont été incluses et 11 méthodes d'apprentissage automatiques ont été explorées. Le réseau neuronal ainsi créé a permis de réaliser des prédictions de bonne qualité ($AUC=0,62$).

Gulliford et al. ont utilisé une méthodologie similaire pour prédire la récurrence biologique et la toxicité après radiothérapie pour un cancer de prostate (Gulliford et al. 2004). Leur étude est une des seules à inclure, au-delà des données cliniques habituelles, les données dosimétriques précises. Ils ont ainsi intégré les histogrammes dose-volume et ont entraîné trois ANNs pour prédire la nocturie, les rectorragies et l'évolution du PSA. La sensibilité et la spécificité restent cependant médiocres (55%), mais des études plus récentes les ont significativement améliorées (jusque 68% dans le modèle prédictif des rectorragies de Tomatis et al.) (Pella et al. 2011; Tomatis et al. 2012; Coates, Souhami, et El Naqa 2016), sans atteindre des performances les rendant utilisables en clinique.

En radiothérapie thoracique, plusieurs études ont exploré l'intérêt des ANNs dans la prédiction de la pneumopathie radique (S. Chen, Zhou, Zhang, et al. 2007; Su et al. 2005). Dans l'étude de Chen et al., les variables suivantes ont été incluses dans un ANN : la V16, la dose moyenne aux poumons, la FEV1, la DLCO et l'administration préalable d'une chimiothérapie. Chaque variable a ensuite été retirée du modèle pour évaluer sa pertinence. Au total, chacune d'entre elle, exceptée la FEV1 et

l'administration préalable de chimiothérapie, était nécessaire à une prédiction optimale. Le nombre relativement faible de variables intégrées doit cependant pondérer l'interprétation des résultats obtenus.

Enfin, dans le cancer du col de l'utérus, une étude a utilisé un ANN pour prédire la survie après radiothérapie exclusive (Ochi et al. 2002). Centre trente-quatre patientes ont été incluses par l'intégration de ces paramètres : âge, état général, taux d'hémoglobine, protéinémie, stade FIGO, réponse histologique sur des biopsies périodiques pendant le traitement. L'efficacité de la prédiction a été évaluée en calculant l'aire sous la courbe ROC (AUC). L'AUC du modèle complet était de 0,7782 +/- 0,0105, ce qui représente une prédiction fiable.

3.1.3 Apprentissage profond (Deep Learning, DL)

L'apprentissage profond est une variante des ANNs. Alors qu'un réseau neuronal comporte une à trois couches de neurones et nécessite d'être entraîné (apprentissage supervisé), un modèle de DL peut comporter un nombre bien plus élevé de couches et peut réaliser de l'apprentissage non-supervisé. L'apprentissage profond a principalement été utilisé dans le champ de la détection de lésions cutanées (Esteva et al. 2015), pour classifier des lésions de rétinopathie diabétique (Gulshan et al. 2016) ou en imagerie médicale pour la détection ou la segmentation automatique (Hua et al. 2015; Guo, Gao, et Shen 2015) et n'a pas encore été utilisé dans des études de population en radiothérapie. Il demande une puissance de calcul plus importante qu'un ANN classique, mais la diminution du coût de la puissance de calcul et l'utilisation de méthodes de calcul déportées sur les processeurs graphiques (Graphics Processing Unit, GPU), vont probablement entraîner une démocratisation de cette méthode.

3.1.4 Apprentissage supervisé (Supervised learning, SL) ou non-supervisé (Unsupervised Learning, UL)

Le but de l'apprentissage supervisé est de prédire une variable de sortie connue. Il doit d'abord être entraîné sur un ensemble de données où les caractéristiques initiales et finales sont connues afin de créer un modèle qui classe le plus fidèlement possible ces données d'entraînement. Le modèle sera ensuite généralisé pour prédire les données finales (comme la survie globale, le contrôle local, ou la toxicité) à partir des données d'entrées de n'importe quel patient. Dans l'apprentissage non-supervisé, l'ensemble de données explorées n'est pas labélisé et c'est l'algorithme qui apprend et découvre des motifs de corrélation, sans entraînement initial et sans intervention humaine. Cette méthode permet de détecter des facteurs pronostics ou prédictifs qui n'auraient pas pu être déterminés par un raisonnement humain. En médecine, l'UL pourrait par exemple mettre en évidence des mécanismes physiopathologiques inconnus. L'inconvénient de l'UL est qu'il peut générer un modèle « boîte-noire », impossible à comprendre par un humain. A ce jour, toutes les études d'apprentissage automatique en radiothérapie sont de type supervisé (Cruz et Wishart 2006; Kourou et al. 2015). La mise à disposition gratuite de logiciel d'UL devrait rendre leur utilisation plus simple (« TensorFlow -- an Open Source Software Library for Machine Intelligence » s. d.; « Torch | Scientific computing for LuaJIT. » s. d.).

3.2 Systèmes d'aide à la décision

En oncologie, l'utilisation du machine learning devrait permettre de mieux personnaliser les traitements. Cette personnalisation repose sur le principe de prédiction : si l'on est capable, à partir du profil précis du patient, de connaître son devenir, on peut adapter au mieux la prise en charge afin d'optimiser d'une part l'efficacité du traitement et de diminuer d'autre part les effets secondaires. Avant que cette vision ne devienne une réalité, il sera nécessaire de développer des logiciels intuitifs utilisables à l'hôpital par les médecins.

Deux entreprises exploitent des données médicales grâce à l'intelligence artificielle : DeepMind (Londres, Royaume-Uni), une entreprise d'IA britannique rachetée en 2014 par Google (Mountain View, Californie, Etats-Unis), a développé la

première intelligence artificielle capable de battre un humain au jeu de Go (Silver et al. 2016). En février 2016, une branche DeepMind Health a été créée. Le National Health Service a donné accès à 1,6 million de dossiers de patients pour entraîner leurs algorithmes d'apprentissage automatiques (Hawkes 2016). Par ailleurs, IBM Watson Health (Cambridge, Massachusetts, Etats-Unis), une sous-division d'IBM, qui exploite le système d'IA Watson, a développé un partenariat avec le Memorial Sloan Kettering Cancer Center (MSKCC) de New-York pour créer et exploiter un système d'aide à la prise de décision en oncologie (« Watson Oncology » s. d.).

Ces systèmes sont appelés à se généraliser dans les 10 à 15 années à venir. Ils devront reposer sur des modèles prédictifs solides et dynamiques, mises à jour en permanence à chaque nouveau patient traité (système de santé apprenant, « learning health system ») (Lambin, van Stiphout, et al. 2013; Lambin, Roelofs, et al. 2013). Il sera nécessaire de réaliser des études à haut niveau de preuve pour valider les outils d'aide à la décision ou de diagnostic issus de l'intelligence artificielle avant qu'ils n'entrent dans le champ des dispositifs médicaux.

3.3 Article 2 : Big Data and machine learning in radiation oncology: state of the art and future prospects

Chapitre 4 - Création d'un modèle prédictif par Deep Learning

4.1 Introduction

Le but de cette étude preuve de concept était d'utiliser des données hétérogènes pour créer un modèle prédictif de la réponse à la radiochimiothérapie pré-opératoire dans le cancer du rectum. Les données cliniques (âge, sexe, taille de la tumeur, stade TNM), biologiques (numération formule sanguine avant traitement), d'imagerie (variables radiomics extraites du scanner de dosimétrie), de radiothérapie (dose totale au GTV et au CTV, dose par fraction, étalement et d'anatomopathologie (évaluation de la réponse thérapeutique sur la pièce d'exérèse après la chirurgie) ont été utilisées pour créer un profil patient complet. Ces données ont ensuite été filtrées afin de réduire la dimension du modèle et le risque d'overfitting avant d'entraîner un réseau neuronal profond. Cette étude a été approuvée par le CPP Ile-de-France II comité #00001072, référence #CDW_2015_0024.

4.2 Méthodes

4.2.1 Description des données

Les variables ont été extraites du CDW de l'HEGP, principalement des comptes-rendus initiaux des oncologues radiothérapeutes. Les patients nous ont été adressés de trois hôpitaux universitaires : l'Hôpital Européen Georges Pompidou (HEGP), l'Hôpital Cochin (HC) et l'Hôpital Ambroise Paré (HAP). Les informations manquantes, des patients traités à l'Hôpital Cochin (HC) ou à l'Hôpital Ambroise Paré (HAP) ont été collectées manuellement à partir de documents externes non numérisés. Les patients atteints de tumeurs T2 à T4 No-N1 traités entre juin 2010 et octobre 2016 par radiochimiothérapie néoadjuvante (45 à 50,4 Gy) avec de la Capécitabine (800 mg/m² matin et soir) en concomitant ont été inclus dans cette étude. Tous les patients avaient eu une IRM pelvienne et un scanner thoraco-abdomino-pelvien ou une TEP pour le bilan d'extension. La chirurgie par résection complète du mesorectum a été réalisée 6 à 10 semaines après la radiochimiothérapie. Quatre-vingt-quinze patients (HEGP: n=35, 37%; HC: n=23, 24%; HAP: n=37, 39%)

ont été inclus dans l'analyse finale. Le suivi médian était de 16 mois (3-65). La taille tumorale médiane était de 48 mm (15-130). Il y avait 9 T2 (9%), 75 T3 (79%) et 11 T4 (12%). Dix-neuf patients étaient No (20%). Les taux d'hémoglobine, de neutrophiles et de lymphocytes initiaux étaient respectivement de 13,6 g/dl (9.7-17.5), 1734/mm³ (336-3760) et 4050/mm³ (1100-11160). La dose totale médiane au GTV était de 50,4 Gy (45-50,4) et de 45 Gy au CTV. La dose par fraction médiane était de 2 Gy (1.8-2.25 Gy). L'étalement médian était de 39 jours (32-69). Le temps écoulé médian entre la radiochimiothérapie et la chirurgie était 9 semaines (4-11). Ving-deux patients étaient en réponse pathologique complète (pCR - 23%). Quarante-deux patients ont reçu une chimiothérapie adjuvante (44%). Deux patients ont présenté une récurrence locale (2,1%). Sept patients ont présenté une récurrence à distance (7,3%). Tous ces patients appartenaient au groupe de patients qui n'étaient pas en réponse complète. De toutes ces variables, seul le stade T était corrélé à la pCR (Test du Chi-2, p=0,036).

4.2.2 Extraction, contrôle qualité et filtrage des variables de radiomics

Les imageries de planification des patients ont été extraites du PACS de l'HEGP. Les scanners ont été réalisés sur un appareil General Electric Light Speed scanner (Boston, Massachusetts, USA) de la façon suivante : acquisition hélicale continue de coupes de 1,25 mm, matrice de 512x512x512, 120 kV, mAs > 350, vitesse: 3.75, mode: 0.75, avec injection de produit de contraste).

Le GTV de la tumeur rectale a été contourné par deux experts sur le logiciel Eclipse V.13 (Varian, Palo Alto, California, USA) avant d'être exporté pour l'extraction radiomics. Le logiciel IBEX (MD Anderson, Texas, USA) a été utilisé pour l'extraction des variables radiomics (Zhang et al. 2015) à partir des deux GTV ainsi obtenus.

Les variables suivantes ont été extraites (annexe 1): Shape, Intensity Direct, Gray Level Co-Occurrence Matrix (GLCM) 25 (2D) and 3 (3D), Neighbor Intensity Difference (NID) 3 et 25, Gray Level Run Length Matrix 25. Aucun preprocessing n'a été utilisé. En tout, 319700 variables ont été extraites (1683 pour chaque patient). Le jeu de données comprenant les variables décrites en 4.2.1 et 4.2.2 a été séparé en cinq afin de réaliser les 5 validations croisées.

Afin d'évaluer la robustesse des variables extraites, les corrélations intra-classes (ICC) entre les variables radiomics extraites du contour 1 et celles extraites du

contour 2 ont été calculées sur le jeu d'apprentissage (Grove et al. 2015; Parmar et al. 2014) avec le paquets ICC pour R.

$$ICC = \frac{MSr - MSw}{MSr + (k - 1)MSw}$$

où MSR = moyenne quadratique de chaque rangée, MSW = moyenne quadratique de la variance résiduelle, k = nombre d'observateurs, n = nombre de sujets

L'ICC varie entre 0 et 1. Les variables robustes devaient avoir une ICC supérieure à 0,8 (seuil retenu dans la majorité des publications de radiomics). Après ce premier filtrage, nous avons sélectionné les variables corrélées à la réponse complète avec un test de Wilcoxon, tel que décrit dans une publication évaluant différentes méthodes de sélection et de modélisation radiomics par machine learning (Parmar et al. 2014). Cette méthode a la meilleure performance pronostique avec une stabilité élevée. Un clustering non supervisé a ensuite été réalisé pour vérifier l'existence de groupes de patients répondeur et non répondeur individualisables sur les variables radiomics. Toutes ces analyses ont été réalisées avec R version 3.4.2 (R Core Team 2013) avec les paquets ggplot2 (Wickham 2009), survival (Therneau et al 2009) et survminer (Kassambara et al. 2017).

4.2.3 Création du réseau neuronal profond, entraînement et validation

Un réseau neuronal profond (Deep Neural Network, DNN) a été créé en utilisant le DNNClassifier Estimator de TensorFlow, un logiciel open-source (v1.3, Google, Mountain View, California, USA) en code informatique python (Annexe 2).

Il n'existe pas de méthodologie permettant d'établir à priori les paramètres optimaux d'un réseau neuronal. Une combinaison optimale de taille de batch, de profondeur de couches et de nombres de neurones par couche a donc été déterminée empiriquement en testant différentes possibilités. Le réseau final était un compromis entre les performances de prédiction et le coût de calcul et incluait trois couches de 10, 20 et 10 neurones chacune.

La fonction d'activation ReLu (Rectified Linear Unit) (Nair et Hinton 2010) a été utilisée. Cette fonction mathématique est appliquée aux données de sortie d'un neurone afin de déterminer si le neurone doit être activé dans le réseau (c'est-à-dire si le résultat du calcul dépasse un certain seuil de stimulation). Ainsi, si le neurone reçoit une valeur négative d'un autre neurone, la fonction retourne une valeur nulle. Si le neurone reçoit une valeur positive, la fonction passe cette valeur au neurone suivant. Elle peut être écrite :

$$f(x) = \max(0, x)$$

Un optimiseur Adagrad (Duchi, Hazan, et Singer 2011) a été utilisé afin de faire converger le réseau neuronal vers son état lui permettant la meilleure prédiction.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \varepsilon}} * g_t$$

Où θ sont les paramètres, t est le pas de temps, η est le taux d'apprentissage, g_t est le gradient, G_t est une matrice de la somme quadratique des gradients jusqu'au pas de temps t et ε est une variable de lissage permettant d'éviter de diviser par 0.

Cet optimiseur garde en mémoire les gradients (la différence entre la prédiction et la réalité), afin d'adapter de façon dynamique le taux d'apprentissage (qui correspond à la vitesse à laquelle le réseau oublie et apprend sur les données). Si le taux d'apprentissage du réseau est trop bas, son entraînement prendra plus de temps et s'il est trop élevé, le réseau ne réalisera pas de prédiction adaptée.

Afin d'éviter le surapprentissage (overfitting), un nombre d'epoch (nombre d'itération de cycles d'apprentissage) réduit a été choisi (3000 étapes, un epoch). Un nombre supérieur d'epoch aurait entraîné l'incapacité du réseau à réaliser une prédiction sur d'autres patients : la spécificité, sensibilité et la précision obtenue sur le jeu de données d'apprentissage aurait été excellentes, mais mauvaises sur le jeu de données de validation.

Les données ont été séparées en cinq jeux de données afin de réaliser 5 validations croisées. Dans chacun de ces jeux de données, la répartition des patients répondeurs/non-répondeurs était identique (23% de répondeur et 77% de non-répondeur).

Les calculs ont été faits sur une station Linux Ubuntu 17.04 avec un processeur Quad Core 2.8 Ghz Intel Core i7-770HQ et un GPU GeForce GTX1060. Les résultats ont été visualisés sur TensorBoard (Google, Mountain View, California, USA).

Afin de comparer les performances du réseau profond, un modèle de régression logistique (RL) très simple a été créé avec uniquement le stade T comme variable de prédiction avec le paquet R glmnet (Friedman et al. 2017). Une machine à vecteur de support (SVM) avec l'ensemble des variables utilisées dans le DNN a également été créé pour comparer les performances de prédiction avec Sci-Kit Learn, un framework de machine learning en libre accès. Les courbes AUC des trois différents modèles ont été comparées avec le test de DeLong (DeLong, DeLong, et Clarke-Pearson 1988) grâce au paquet R pROC (Robin et al. 2018).

4.3 Résultats

4.3.1 Variables radiomics

Cent vingt-quatre variables avaient une ICC supérieure à 0,8 (7,3%) dans les catégories suivantes :

- Texture :
 - Grey-Level Co-Occurrence Matrix en 2D et 3D,
 - Grey-Level Run Length Matrix,
 - IntensityDirect
 - IntensityHistogram
- Forme (shape).

Parmi ces variables, les variables radiomics robustes ($ICC > 0,8$) ont été filtrées par un test de Wilcoxon afin de garder les variables corrélées à la réponse complète ($p < 0,05$).

Au total, 28 (22%) variables robustes et corrélées ont été retenues pour le modèle final dans trois catégories de texture :

- Gray-Level Co-occurrence Matrix 2D
- Gray-Level Co-occurrence Matrix 3D
- IntensityDirect.

La figure 7 montre le clustering non supervisé de ces variables, qui retrouve bien un profil répondeur et non répondeur.

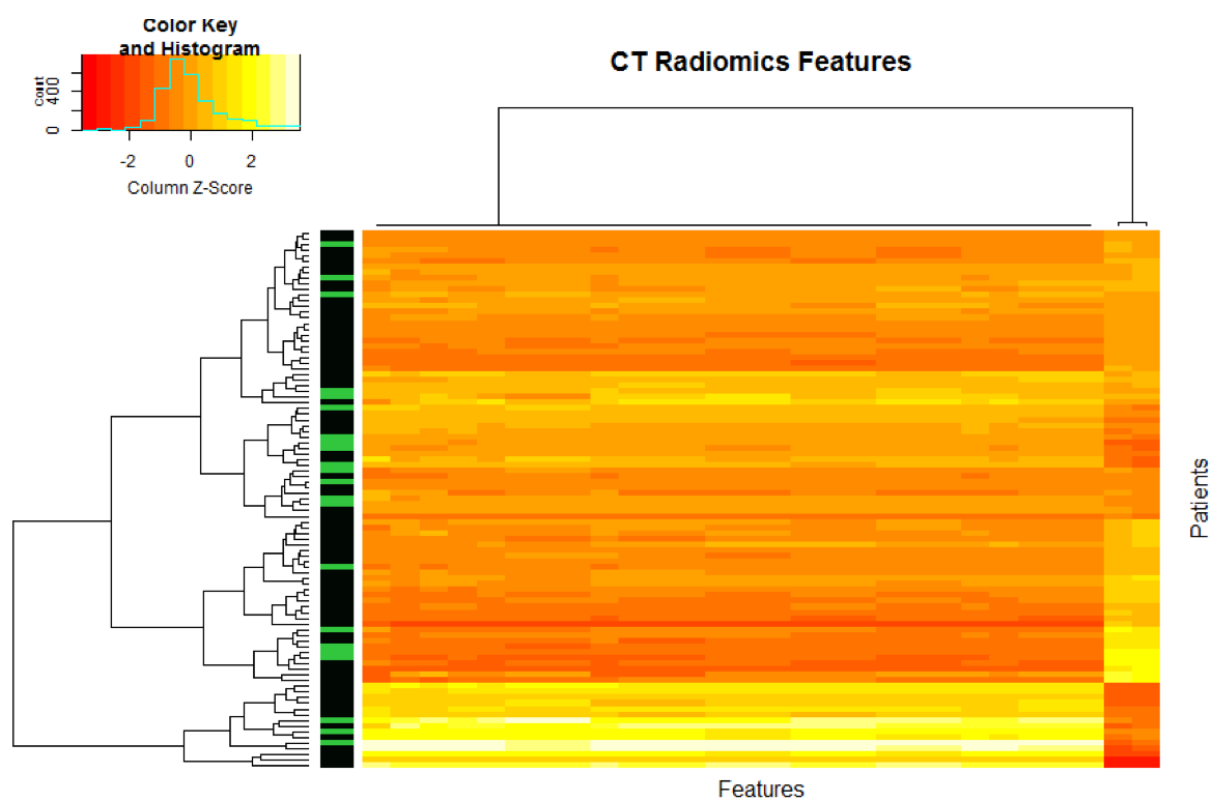


Figure 7. Clustering non supervisé des profils radiomics. Deux phénotypes sont retrouvés : le premier groupe comprend la grande majorité des patients en réponse complète (86%). Le second groupe ne comprend que 14% de patients en réponse complète.

4.3.2 Prédiction de la réponse complète

Le modèle simple LR utilisant uniquement le stade T a une précision de 69,5% (IC 95% = 59,2% - 78,51%). Sa sensibilité est de 34,78% (IC 95% = 16,38% - 57,27%) et sa spécificité de 80,56% (IC 95% = 69,53% - 88,94%). L'AUC pour ce modèle était de 0,59 (IC 95% = 0,458 - 0,686).

Dans le modèle DNN, 29 variables ont finalement été incluses (le stade T et 28 variables radiomics). La prédiction moyenne sur les 5 apprentissages/validations réalisés était de 80% (IC 95% = 70,54% - 87,51%). La sensibilité était de 68,2% (IC 95% = 45,13% - 86,14%), sa spécificité de 83,56% (IC 95% = 73,05% - 91,21%). L'AUC moyenne était de 0,72 (IC 95% = 0.65 - 0.87).

Dans le modèle SVM, la précision était de 71,58% (IC 95% = 61,4%-80,36%). La sensibilité était de 45,45% (IC 95% = 24,39%-67,79%), sa spécificité de 79,45% (IC 95% = 68,38% - 88,02%). L'AUC moyenne était de 0,62 ((IC 95% = 0.51 - 0.74).

Les différentes courbes ROC sont en figure 8. Un test de DeLong (DeLong, DeLong, et Clarke-Pearson 1988) a été réalisé afin de déterminer s'il existait une différence statistiquement significative entre le modèle LR et DNN d'une part et le modèle DNN et SVM d'autre part avec le paquet R pROC (Robin et al. 2018) :

- La comparaison du modèle LR au modèle DNN a montré une différence significative ($p=0,0004$, test de DeLong)
- La comparaison du modèle SVM au modèle DNN a également retrouvé une différence significative ($p=0,004$, test de DeLong).

Le Deep Learning a donc, sur cette population, un avantage statistiquement significatif sur un modèle plus simple réalisé en régression logistique avec le stade tumorale uniquement ou même un modèle de machine learning réalisé avec une machine à vecteur de support sur l'ensemble des 29 variables.

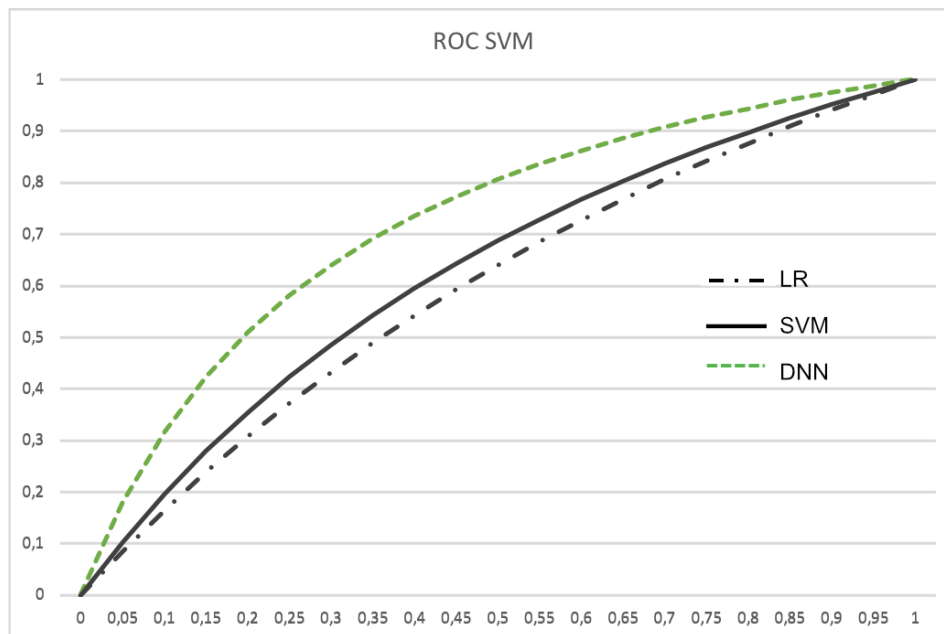


Figure 8. Courbes ROC des trois modèles.

Le tableau de contingence de chaque modèle est montré en tableau 3.

	Prédiction		Total
	pCR: n (%de la réalité)	Non pCR: n (% de la réalité)	
Réalité	LR: 8 (36%)	LR: 14 (64%)	22
	DNN: 15 (68%)	DNN: 7 (32%)	
	SVM: 10 (45,4%)	SVM: 12 (54,5%)	
	LR: 15 (21%)	LR: 58 (79%)	73
	DNN: 12 (16%)	DNN: 61 (84%)	
	SVM: 15 (21%)	SVM: 58 (79%)	
Total	LR: 23	LR: 72	95
	DNN: 27	DNN: 68	
	SVM: 25	SVM: 70	

Tableau 3. Tableau de contingence : Baseline LR: Régression Linéaire sur le stade T ; DNN: Deep Neural Network avec les 29 variables ; SVM: Support Vector Machine Model avec les mêmes 29 variables.

4.4 Article 3 : Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer

Chapitre 5 - Discussion

5.1 Comparaison aux autres modèles prédictifs par machine learning

Les méthodes de radiomics et de machine learning ont récemment été utilisées afin de prédire la réponse thérapeutique après radiochimiothérapie pré-opératoire dans le cancer du rectum.

La première étude a été publiée par Nie et al en 2016 : ils ont utilisé les imageries IRM multiparamétrique préthérapeutique de 48 patients traités par radiochimiothérapie pré-opératoire (Nie et al. 2016). Des séquences IRM T1, T2, de diffusion et de perfusion ont été réalisées. Cent trois variables radiomics volumétriques ont été extraites pour chaque patient. Un réseau neuronal à trois couches de type perceptron a été utilisé pour prédire les patients qui auront une réponse complète et les patients ayant une bonne réponse des autres. Une validation interne a été réalisée en randomisant quatre fois les jeux de données d'apprentissage et de validation sur les 48 patients inclus. La qualité des variables radiomics n'a pas été évaluée par les méthodes conventionnellement recommandées (extraction sur plusieurs segmentations réalisées par différentes personnes et calcul des indices de corrélation). Au total, l'AUC du réseau était comprise entre 0,54 et 0,73 pour la prédiction de la réponse complète et 0,71 et 0,79 pour la bonne réponse. Les autres indices de performance tels que la précision, les taux de vrais et de faux positifs et négatifs, les matrices de confusion, ne sont pas rapportés dans l'étude. Au total, cette première étude reste préliminaire et ne répond pas aux critères de qualité habituellement décrits (Lambin et al. 2017). Le faible nombre de patients, le caractère unicentrique, l'absence de contrôle qualité des variables radiomics extraites, l'absence de validation sur une cohorte de patients externes rendent l'interprétation de l'étude difficile.

Une seconde étude publiée en 2017 par Liu et al a recruté 222 patients (Liu et al. 2017). Une imagerie pré-opératoire IRM T2 et de diffusion a été utilisée pour l'extraction des variables. Deux mille deux cent cinquante-deux caractéristiques radiomics ont été extraites pour chaque patient. Après sélection des 30 caractéristiques les mieux corrélées à la réponse complète, une machine à support de vecteur a été utilisée pour prédire les patients qui seront en réponse complète à l'issue de la radiochimiothérapie. La machine a été entraînée sur 152 patients et validées sur 70 autres. Le modèle a également intégré la taille tumorale en plus des données radiomics.

L'AUC du modèle était de 0,9756 (intervalle de confiance à 95% = 0,9185-0,9711). Dans cette étude, la robustesse des caractéristiques radiomics extraites était bien évaluée en calculant leur variabilité sur deux segmentations réalisées par deux personnes différentes. Le nombre de patients plus important et la validation sur une cohorte indépendante de patients font de cette étude une approche plus robuste que la précédente. Cependant, l'utilisation de données IRM plutôt que scanner rend sa reproductibilité difficile dans d'autres services.

L'utilisation de données radiomics extraites à partir de scanner étant plus facile et reproductible (Hu et al. 2016), nous proposons dans cette étude preuve de concept d'extraire les variables radiomics directement sur le scanner réalisé pour calculer la dosimétrie en radiothérapie et d'utiliser ces données dans un modèle hétérogène explorant également l'intérêt des caractéristiques cliniques et thérapeutiques pour la prédiction.

Notre étude a plusieurs limites : alors que les meilleurs résultats du Deep Learning sont obtenus avec de grand corpus d'images standardisées (130 000 images dans la publication sur les lésions cutanées d'Esteva et al), notre étude preuve de concept porte sur des effectifs limités : le faible nombre de patients ne permet pas de considérer que le modèle sera applicable à d'autres populations, même si les patients proviennent de différents hôpitaux. La qualité de la prédiction avec le DNN est légèrement meilleure que celle faite avec un SVM plus simple, mais elle reste insuffisante si l'on souhaite l'utiliser pour l'aide à la décision en pratique clinique.

Par rapport aux études précédemment citées, notre étude présente l'avantage d'utiliser les données de patients provenant de différents hôpitaux mais traités selon le même protocole, ce qui n'est pas le cas des autres études qui sont monocentriques. Le second intérêt de notre approche est l'utilisation de l'imagerie scanner réalisée en routine pour planifier le traitement de radiothérapie, qui est plus facilement exploitable pour l'extraction radiomics que l'IRM

La diffusion de notre ontologie ROS et le partage de nos méthodes d'intégration des données de radiothérapie dans les entrepôts de données cliniques (en particulier i2b2) seront des facteurs majeurs pour conduire des études multicentriques incluant un grand nombre de dossiers patients. Cette dynamique, promue au sein de l'AP-HP, est destinée à s'étendre à d'autres sites au niveau national et international.

Des améliorations pourraient être réalisées en augmentant le nombre de patients et en intégrant l'imagerie IRM et TEP en plus du scanner. L'augmentation du nombre de variables du modèle le rendra cependant toujours plus difficile à utiliser en routine. De plus, la validation externe d'un modèle toujours plus complexe présente des difficultés méthodologiques : il faudrait par exemple que les protocoles d'imagerie soient alors les mêmes dans les différents services, ce qui n'est à l'heure actuelle pas le cas. La prochaine étape pourrait consister à ne plus réaliser d'extraction radiomics, mais à entraîner un réseau neuronal profond directement sur les images natives au format DICOM. Cette méthode permettrait probablement de s'affranchir des problématiques de standardisation du processus d'extraction et d'exploiter pleinement les avantages du Deep Learning, qui est particulièrement adapté à l'analyse d'image. Cette approche aurait cependant des besoins en puissance informatique et en temps machine supérieurs.

L'ontologie et le modèle de données que nous avons développés ne sont pas restreints au cancer colo-rectal mais s'appliquent à tout type de tumeur et toute localisation. Nous avons ainsi lancé trois projets de recherche translationnelle utilisant des méthodologies similaires :

- Dans le domaine du cancer du poumon, nous utiliserons une approche de radiomics et de machine learning pour prédire les patients ayant une bonne efficacité de la radiothérapie stéréotaxique sur une cohorte multicentrique de patients traités à l'HEGP et au CHU de Bordeaux.
- Dans le domaine des cancers tête et cou, les données des patients inclus dans l'étude prospective ART-ORL, provenant de 14 centres français, vont être analysées de la même façon afin de prédire le contrôle local et la survie globale.
- Dans le domaine du cancer du pancréas, une collaboration avec le service de radiothérapie de l'Hôpital Tenon (APHP), va permettre d'analyser les données des patients inclus dans l'essai prospectif LAP07 de radiochimiothérapie afin de prédire les patients qui peuvent bénéficier de ce traitement.

Enfin, une collaboration avec le laboratoire d'intelligence artificielle appliquée à la médecine de Stanford a vu le jour dans le cadre d'un postdoctorat qui sera réalisé en 2019 sur la même thématique.

5.2 Intelligence Artificielle et médecine

5.2.1 Un processus entamé il y a 40 ans

L'automatisation de la pratique médicale a fait l'objet d'études dès les années 1970. Le système MYCIN a été créé il y a un peu plus de quarante ans à Stanford et était fondé sur l'application de connaissances (knowledge-based) : il permettait d'identifier l'origine d'une infection à partir de 600 règles. Il reposait sur une programmation conditionnelle de type if/then, ou if/else. A l'époque déjà, les performances du système avaient été comparées à celles de 5 infectiologues spécialistes et retrouvait un taux de plan de traitement correct de 65% contre 42,5% à 62,5% pour le plan proposé par les médecins (Shortliffe 1977). Plus tard, en 1982, un système plus complet, reposant sur un principe similaire, appelé INTERNIST-1/CADUCEUS avait été créé à Pittsburgh et pouvait diagnostiquer 1000 maladies différentes (Myers, Pople, et Miller 1982; Banks 1986). Ces systèmes, bien que performants, n'ont jamais été utilisés dans la pratique clinique. Les raisons de ces échecs sont multiples, mais comportent notamment la difficulté et le temps nécessaire pour entrer manuellement l'ensemble des données qui permettaient de guider la décision du logiciel et l'absence de systèmes informatiques adaptés dans la plupart des hôpitaux de l'époque. Un autre facteur majeur expliquant l'échec de ces systèmes correspond au fait que la pratique de la médecine ne peut pas se résumer à l'application d'un savoir, aussi immense soit-il, et que la majorité des décisions médicales repose sur la reconnaissance instinctuelle d'un motif clinique et l'application d'une thérapeutique basée sur l'expérience du clinicien.

Les obstacles techniques que nous avons évoqués, liés au besoin d'entrer des quantités de données importantes, sont désormais plus facilement surmontables avec les logiciels médicaux et les entrepôts de données. Avec l'avènement des techniques de machine learning, la création d'un système d'aide à la décision ne repose plus sur l'écriture fastidieuse d'un programme d'afférences conditionnelles. Au-delà de la création du programme qui apprend à la machine à apprendre, il n'y a pas d'intervention humaine. Ces possibilités ont un intérêt particulier dans les domaines de reconnaissance et de perception, tels que l'imagerie, l'anatomopathologie, l'ophtalmologie et la dermatologie.

5.2.2 Deep Learning et perception

Le Deep Learning est un sous-domaine des réseaux neuronaux. D'abord considéré comme un domaine à la marge des mathématiques et du machine learning, très peu populaire, il s'est progressivement imposé (LeCun, Bengio, et Hinton 2015). En 2010, une base de données de plusieurs millions d'images, appelé ImageNet a été créé par l'équipe de Fei-Fei Li, à Stanford. Chaque année, un challenge, le LSVRC (Large Scale Visual Recognition Challenge) est organisé afin d'évaluer la précision d'algorithmes de reconnaissance visuelle de différentes équipes (Russakovsky et al. 2015). Durant les deux premières années du challenge, les taux d'erreur étaient situés entre 26 et 28%. Mais en 2012, une équipe de Toronto, menée par Geoff Hinton, a soumis un algorithme avec un taux d'erreur de 16%, ce qui représentait une véritable révolution à l'époque. Le réseau AlexNet, (Krizhevsky, Sutskever, et Hinton 2012) posait les bases du Deep Learning, avec l'utilisation de couches de dropout (élimination de neurones récurrents), afin de réduire les risques d'overfitting (apprentissage trop adapté à une population, non généralisable), et de la fonction d'activation ReLu (rectified linear activation Unit), qui sont désormais communs dans ce type de réseaux. Le Deep Learning est depuis utilisé dans de nombreux exemples célèbres, dont le jeu de Go (Silver et al. 2016)

En médecine, ce type d'approche a commencé à être utilisé dans des projets d'envergure dès 2016. Une première étude, publiée dans Nature, a notamment comparé les performances d'un réseau neuronal (Inception v3) à celles d'un groupe de 21 dermatologues pour la classification et le diagnostic de lésions cutanées. Une technique de transfer learning a été utilisée : Inception V3 a ainsi été entraîné sur la base de données Imagenet avant d'être de nouveau entraîné sur une seconde base de 129450 images. Les performances d'Inception v3 étaient comparables à celles des dermatologues (Esteva et al. 2017). De la même façon, le Deep Learning a été utilisé pour classifier les stades de rétinopathie diabétique (Gulshan et al. 2016), diagnostiquer une maladie un an à l'avance (Miotto et al. 2016), prédire le décès à 5 ans sur un simple scanner (Oakden-Rayner et al. 2017) ou identifier des métastases ganglionnaires (Ehteshami Bejnordi et al. 2017).

Il n'est pas possible actuellement d'expliquer comment et pourquoi un réseau neuronal réalise une prédiction ou une classification. Cet effet « boîte noire » est donc une limite dont il faudra tenir compte en médecine. L'absence « d'explicabilité » des

résultats obtenus rendent leur contrôle et leur certification difficile, voire impossible. Etant donné l'importance des décisions qui pourraient être prises en médecine à partir de ces résultats, une réflexion éthique doit être menée afin de déterminer si nous sommes prêts en tant que médecins et patients à accepter ces dispositifs.

5.2.3 Le futur rôle du médecin

On peut schématiquement considérer que le Deep Learning peut servir à deux grands types d'interventions en médecine : réaliser des tâches qu'un médecin sait faire (diagnostic sur images) mais aussi réaliser des tâches inaccessibles au médecin (prédire le devenir d'un patient). Dans notre étude, la tâche remplie par le réseau neuronal profond n'est pas réalisable par un médecin. Il est donc d'autant plus difficile de valider les résultats de la méthode, qui reste une boîte noire, impossible à comprendre pour un humain. Il est donc important, particulièrement dans le domaine médical, de valider ces approches avant qu'elles ne soient utilisées pour traiter des patients. Les méthodologies standard de validation d'un médicament ou d'un dispositif médical, par la réalisation d'études de phase III randomisées, ne sont peut-être pas adaptées à l'IA. L'algorithme, pour rester pertinent, doit en effet continuer à apprendre en permanence à partir de nouveaux patients. A partir du moment où les résultats d'un algorithme ne sont jamais figés, comment en contrôler la pertinence ?

Au-delà de cette question, quelle sera la place du médecin avec l'émergence de l'IA ? Geoff Hinton, l'un des inventeurs du Deep Learning, a notamment déclaré : « Nous devrions arrêter de former des radiologues dès maintenant. Il est absolument évident que d'ici cinq ans, le Deep Learning fera mieux que les radiologues » (Machine Learning and Market for Intelligence Conference 2016). Ces méthodologies ne sont probablement pas appelées à remplacer le médecin à court et moyen terme, mais plus probablement à augmenter ses capacités de décision. L'IA sera, comme le stéthoscope, un outil à la disposition du médecin. Son rôle à plus long terme reste beaucoup plus flou, si l'on considère le faible coût d'un algorithme et sa grande précision dans certains domaines, parmi lesquels on retrouve effectivement l'imagerie et l'anatomopathologie. Une réflexion éthique doit donc être menée dès aujourd'hui sur les conséquences de ces méthodes sur la pratique de la médecine (Cabitza, Rasoini, et Gensini 2017).

Chapitre 6 - Conclusion

Les systèmes informatiques qui régissent la pratique médicale génèrent des données très précises sur les patients. Ces données de nature hétérogène comprennent des informations démographiques, socio-économiques, cliniques, biologiques, d'imagerie et génomiques. L'utilisation de ces données nécessite la mise en place de processus automatisés d'uniformisation et d'intégration dans des entrepôts de données cliniques. Les données spécifiques à la radiothérapie sont très riches et de bonne qualité. Elles sont principalement issues de la cartographie de la dose de rayonnements reçues par la tumeur et les organes l'entourant. Elles proviennent également de systèmes d'enregistrement prospectifs des traitements réalisés. Ces données sont rarement intégrées aux entrepôts de données cliniques et ne sont donc pas exploitées dans les modèles prédictifs.

Nous avons donc développé une méthodologie permettant d'extraire, d'uniformiser et d'intégrer ces données dans le CDW de l'HEGP, en créant notamment une ontologie complète qui sera utilisée à l'échelle de l'APHP. Cette ontologie de 400 concepts formalise les structures délinéées en radiothérapie pour la planification du traitement en respectant les recommandations et les nomenclatures anatomiques existantes. Ce type d'approche permettra à terme d'envisager des études multicentriques à grande échelle.

L'utilisation du machine learning permet d'exploiter ces données médicales afin de réaliser un diagnostic ou une prédiction. Dans notre étude « preuve de concept », nous avons créé un modèle prédictif par Deep Learning sur une cohorte de patients atteints de cancer du rectum localement avancés. Ce modèle permet d'identifier quels patients seront en réponse complète après radiochimiothérapie, sans avoir besoin de les opérer. Ce type d'approche pourrait être utilisé pour personnaliser les traitements et diminuer leur séquelle. La validation de ces algorithmes sera cependant une étape importante et difficile avant qu'ils ne puissent être utilisés en routine clinique.

Ces systèmes d'IA, lorsqu'ils seront rigoureusement validés, seront d'abord de nouveaux outils à la disposition du médecin. A plus long terme, leur place sera probablement centrale dans la prise en charge des patients.

Bibliographie

Abernethy, Amy P., Lynn M. Etheredge, Patricia A. Ganz, Paul Wallace, Robert R. German, Chalapathy Neti, Peter B. Bach, et Sharon B. Murphy. 2010. « Rapid-Learning System for Cancer Care ». *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 28 (27): 4268-74. <https://doi.org/10.1200/JCO.2010.28.5478>.

Aerts, Hugo J. W. L. 2016. « The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review ». *JAMA Oncology* 2 (12): 1636-42. <https://doi.org/10.1001/jamaoncol.2016.2631>.

Appelt, Ane L., John Pløen, Henrik Harling, Frank S. Jensen, Lars H. Jensen, Jens C. R. Jørgensen, Jan Lindebjerg, Søren R. Rafaelsen, et Anders Jakobsen. 2015. « High-Dose Chemoradiotherapy and Watchful Waiting for Distal Rectal Cancer: A Prospective Observational Study ». *The Lancet. Oncology* 16 (8): 919-27. [https://doi.org/10.1016/S1470-2045\(15\)00120-5](https://doi.org/10.1016/S1470-2045(15)00120-5).

« Applying machine learning to radiotherapy planning for head & neck cancer ». s. d. DeepMind. Consulté le 17 mai 2017. <https://deepmind.com/blog/applying-machine-learning-radiotherapy-planning-head-neck-cancer/>.

Banks, G. 1986. « Artificial Intelligence in Medical Diagnosis: The INTERNIST/CADUCEUS Approach ». *Critical Reviews in Medical Informatics* 1 (1): 23-54.

Beam, Andrew L., et Isaac S. Kohane. 2016. « Translating Artificial Intelligence Into Clinical Care ». *JAMA* 316 (22): 2368-69. <https://doi.org/10.1001/jama.2016.17217>.

Bhangu, Aneel, Gina Brown, R. J. Nicholls, John Wong, Ara Darzi, et Paris Tekkis. 2013. « Survival Outcome of Local Excision versus Radical Resection of Colon or Rectal Carcinoma: A Surveillance, Epidemiology, and End Results (SEER) Population-Based Study ». *Annals of Surgery* 258 (4): 563-69; discussion 569-571. <https://doi.org/10.1097/SLA.0b013e3182a4e85a>.

Bibault, Jean-Emmanuel. (2017) 2017. *DeepRadiomicsRectalCancer: Code, parameters and scripts used for our study predicting pathologic complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer*. <https://github.com/jebibault/DeepRadiomicsRectalCancer>.

———. s. d. « Radiation Oncology Structures Ontology - Summary | NCBO BioPortal ».

<http://bioportal.bioontology.org/ontologies/ROS/?p=summary>.

Bibault, Jean-Emmanuel, Philippe Giraud, et Anita Burgun. 2016. « Big Data and Machine Learning in Radiation Oncology: State of the Art and Future Prospects ». *Cancer Letters* 382 (1): 110-17. <https://doi.org/10.1016/j.canlet.2016.05.033>.

Bibault, Jean-Emmanuel, Eric Zapletal, Bastien Rance, Philippe Giraud, et Anita Burgun. 2018. « Labeling for Big Data in Radiation Oncology: The Radiation Oncology Structures Ontology ». *PloS One* 13 (1): e0191263. <https://doi.org/10.1371/journal.pone.0191263>.

Bryce, T. J., M. W. Dewhirst, C. E. Floyd, V. Hars, et D. M. Brizel. 1998. « Artificial Neural Network Model of Survival in Patients Treated with Irradiation with and without Concurrent Chemotherapy for Advanced Carcinoma of the Head and Neck ». *International Journal of Radiation Oncology, Biology, Physics* 41 (2): 339-45.

Cabitza, Federico, Raffaele Rasoini, et Gian Franco Gensini. 2017. « Unintended Consequences of Machine Learning in Medicine ». *JAMA* 318 (6): 517-18. <https://doi.org/10.1001/jama.2017.7797>.

Canuel, Vincent, Bastien Rance, Paul Avillach, Patrice Degoulet, et Anita Burgun. 2015. « Translational Research Platforms Integrating Clinical and Omics Data: A Review of Publicly Available Solutions ». *Briefings in Bioinformatics* 16 (2): 280-90. <https://doi.org/10.1093/bib/bbu006>.

Chen, Chengshui, Mingyan He, Yichun Zhu, Lin Shi, et Xiangdong Wang. 2015. « Five Critical Elements to Ensure the Precision Medicine ». *Cancer Metastasis Reviews* 34 (2): 313-18. <https://doi.org/10.1007/s10555-015-9555-3>.

Chen, Shifeng, Sumin Zhou, Fang-Fang Yin, Lawrence B. Marks, et Shiva K. Das. 2007. « Investigation of the Support Vector Machine Algorithm to Predict Lung Radiation-Induced Pneumonitis ». *Medical Physics* 34 (10): 3808-14.

Chen, Shifeng, Sumin Zhou, Junan Zhang, Fang-Fang Yin, Lawrence B. Marks, et Shiva K. Das. 2007. « A Neural Network Model to Predict Lung Radiation-Induced Pneumonitis ». *Medical Physics* 34 (9): 3420-27.

Coates, James, Luis Souhami, et Issam El Naqa. 2016. « Big Data Analytics for Prostate Radiotherapy ». *Frontiers in Oncology* 6: 149. <https://doi.org/10.3389/fonc.2016.00149>.

Cruz, Joseph A., et David S. Wishart. 2006. « Applications of Machine Learning in Cancer Prediction and Prognosis ». *Cancer Informatics* 2: 59-77.

Danciu, Ioana, James D. Cowan, Melissa Basford, Xiaoming Wang, Alexander Saip, Susan Osgood, Jana Shirey-Rice, Jacqueline Kirby, et Paul A. Harris. 2014. « Secondary Use of Clinical Data: the Vanderbilt Approach ». *Journal of biomedical informatics* 52 (décembre): 28-35. <https://doi.org/10.1016/j.jbi.2014.02.003>.

DeLong, E. R., D. M. DeLong, et D. L. Clarke-Pearson. 1988. « Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach ». *Biometrics* 44 (3): 837-45.

Denis, Fabrice, Senna Yossi, Anne-Lise Septans, Alexandre Charron, Eric Voog, Olivier Dupuis, Gérard Ganem, Yoann Pointreau, et Christophe Letellier. 2015. « Improving Survival in Patients Treated for a Lung Cancer Using Self-Evaluated Symptoms Reported Through a Web Application ». *American Journal of Clinical Oncology*, mars. <https://doi.org/10.1097/COC.0000000000000189>.

« DNA Sequencing Costs ». s. d. Consulté le 12 mars 2016. <http://www.genome.gov/sequencingcosts/>.

Duchi, John, Elad Hazan, et Yoram Singer. 2011. « Adaptive Subgradient Methods for Online Learning and Stochastic Optimization ». *Journal of Machine Learning Research* 12 (Jul): 2121-59.

Ehteshami Bejnordi, Babak, Mitko Veta, Paul Johannes van Diest, Bram van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen A. W. M. van der Laak, et al. 2017. « Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer ». *JAMA* 318 (22): 2199-2210. <https://doi.org/10.1001/jama.2017.14585>.

Emaminejad, Nastaran, Muhammad Wahi-Anwar, John Hoffman, Grace H. Kim, Matthew S. Brown, et Michael McNitt-Gray. 2018. « The effects of variations in parameters and algorithm choices on calculated radiomics feature values: initial investigations and comparisons to feature variability across CT image acquisition conditions ». In *Medical Imaging 2018: Computer-Aided Diagnosis*, 10575:105753W. International Society for Optics and Photonics. <https://doi.org/10.1117/12.2293864>.

Escudié, Jean-Baptiste, Anne-Sophie Jannot, Eric Zapletal, Sarah Cohen, Georgia Malamut, Anita Burgun, et Bastien Rance. 2015. « Reviewing 741 Patients Records in Two Hours with FASTVISU ». *AMIA ... Annual Symposium Proceedings. AMIA Symposium* 2015: 553-59.

Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, et Sebastian Thrun. 2017. « Dermatologist-Level Classification of Skin Cancer

with Deep Neural Networks ». *Nature* 542 (7639): 115-18.
<https://doi.org/10.1038/nature21056>.

Falchook, Aaron D., Gregg Tracton, Lori Stravers, Mary E. Fleming, Anna C. Snavely, Jeanne F. Noe, David N. Hayes, et al. 2016. « Use of Mobile Device Technology to Continuously Collect Patient-Reported Symptoms during Radiotherapy for Head and Neck Cancer: A Prospective Feasibility Study ». *Advances in Radiation Oncology*, février. <https://doi.org/10.1016/j.adro.2016.02.001>.

Friedman, Jerome, Trevor Hastie, Noah Simon, Junyang Qian, et Rob Tibshirani. 2017. *glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models* (version 2.0-13). <https://CRAN.R-project.org/package=glmnet>.

Gan, J., J. Wang, H. Zhong, R. Luo, Z. Zhou, P. Hu, L. Shen, et Z. Zhang. 2016. « MO-DE-207B-09: A Consistent Test for Radiomics Softwares ». *Medical Physics* 43 (6): 3706. <https://doi.org/10.1118/1.4957258>.

Gan J, Wang J, Zhong H, Luo R, Zhou Z, Hu P, Shen L, et Zhang Z. 2016. « MO-DE-207B-09: A Consistent Test for Radiomics Softwares ». *Medical Physics* 43 (6Part30): 3706-3706. <https://doi.org/10.1118/1.4957258>.

Geifman, Nophar, et Atul J. Butte. 2016. « Do Cancer Clinical Trial Populations Truly Represent Cancer Patients? A Comparison of Open Clinical Trials to the Cancer Genome Atlas. » *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 21: 309-20.

Georgakilas, Alexandros G., Athanasia Pavlopoulou, Maria Louka, Zacharenia Nikitaki, Constantinos E. Vorgias, Pantelis G. Bagos, et Ioannis Michalopoulos. 2015. « Emerging Molecular Networks Common in Ionizing Radiation, Immune and Inflammatory Responses by Employing Bioinformatics Approaches ». *Cancer Letters* 368 (2): 164-72. <https://doi.org/10.1016/j.canlet.2015.03.021>.

Gillies, Robert J., Paul E. Kinahan, et Hedvig Hricak. 2016. « Radiomics: Images Are More than Pictures, They Are Data ». *Radiology* 278 (2): 563-77. <https://doi.org/10.1148/radiol.2015151169>.

Grégoire, Vincent, Kian Ang, Wilfried Budach, Cai Grau, Marc Hamoir, Johannes A. Langendijk, Anne Lee, et al. 2014. « Delineation of the Neck Node Levels for Head and Neck Tumors: A 2013 Update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG Consensus Guidelines ». *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 110 (1): 172-81. <https://doi.org/10.1016/j.radonc.2013.10.010>.

Grove, Olya, Anders E. Berglund, Matthew B. Schabath, Hugo J. W. L. Aerts, Andre Dekker, Hua Wang, Emmanuel Rios Velazquez, et al. 2015. « Quantitative Computed Tomographic Descriptors Associate Tumor Shape Complexity and Intratumor Heterogeneity with Prognosis in Lung Adenocarcinoma ». *PloS One* 10 (3): e0118261. <https://doi.org/10.1371/journal.pone.0118261>.

Gulliford, Sarah L., Steve Webb, Carl G. Rowbottom, David W. Corne, et David P. Dearnaley. 2004. « Use of Artificial Neural Networks to Predict Biological Outcomes for Patients Receiving Radical Radiotherapy of the Prostate ». *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 71 (1): 3-12. <https://doi.org/10.1016/j.radonc.2003.03.001>.

Gulshan, Varun, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, et al. 2016. « Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs ». *JAMA* 316 (22): 2402-10. <https://doi.org/10.1001/jama.2016.17216>.

Guo, Yanrong, Yaozong Gao, et Dinggang Shen. 2015. « Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching ». *IEEE Transactions on Medical Imaging*, décembre. <https://doi.org/10.1109/TMI.2015.2508280>.

Habr-Gama, Angelita, Joaquim Gama-Rodrigues, Guilherme P. São Julião, Igor Proscurshim, Charles Sabbagh, Patricio B. Lynn, et Rodrigo O. Perez. 2014. « Local Recurrence after Complete Clinical Response and Watch and Wait in Rectal Cancer after Neoadjuvant Chemoradiation: Impact of Salvage Therapy on Local Disease Control ». *International Journal of Radiation Oncology, Biology, Physics* 88 (4): 822-28. <https://doi.org/10.1016/j.ijrobp.2013.12.012>.

Habr-Gama, Angelita, Rodrigo Oliva Perez, Wladimir Nadalin, Jorge Sabbaga, Ulysses Ribeiro, Afonso Henrique Silva e Sousa, Fábio Guilherme Campos, Desidério Roberto Kiss, et Joaquim Gama-Rodrigues. 2004. « Operative versus Nonoperative Treatment for Stage o Distal Rectal Cancer Following Chemoradiation Therapy: Long-Term Results ». *Annals of Surgery* 240 (4): 711-17; discussion 717-718.

Hawkes, Nigel. 2016. « NHS Data Sharing Deal with Google Prompts Concern ». *BMJ (Clinical Research Ed.)* 353: i2573.

Hayashida, Yasuharu, Kazufumi Honda, Yoshiaki Osaka, Tomohiko Hara, Tomoko Umaki, Akihiko Tsuchida, Tatsuya Aoki, Setsuo Hirohashi, et Tesshi Yamada. 2005. « Possible Prediction of Chemoradiosensitivity of Esophageal Cancer by Serum Protein

Profiling ». *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 11 (22): 8042-47. <https://doi.org/10.1158/1078-0432.CCR-05-0656>.

Hu, Panpan, Jiazhou Wang, Haoyu Zhong, Zhen Zhou, Lijun Shen, Weigang Hu, et Zhen Zhang. 2016. « Reproducibility with Repeat CT in Radiomics Study for Rectal Cancer ». *Oncotarget* 7 (44): 71440-46. <https://doi.org/10.18632/oncotarget.12199>.

Hua, Kai-Lung, Che-Hao Hsu, Shintami Chusnul Hidayati, Wen-Huang Cheng, et Yu-Jen Chen. 2015. « Computer-Aided Classification of Lung Nodules on Computed Tomography Images via Deep Learning Technique ». *OncoTargets and Therapy* 8: 2015-22. <https://doi.org/10.2147/OTT.S80733>.

Huser, Vojtech, et James J. Cimino. 2015. « Impending Challenges for the Use of Big Data ». *International Journal of Radiation Oncology*Biophysics*, novembre. <https://doi.org/10.1016/j.ijrobp.2015.10.060>.

Hwang, Kyung Hoon, Haejun Lee, Geon Koh, Debra Willrett, et Daniel L. Rubin. 2017. « Building and Querying RDF/OWL Database of Semantically Annotated Nuclear Medicine Images ». *Journal of Digital Imaging* 30 (1): 4-10. <https://doi.org/10.1007/s10278-016-9916-7>.

Jannot, Anne-Sophie, Eric Zapletal, Paul Avillach, Marie-France Mamzer, Anita Burgun, et Patrice Degoulet. 2017. « The Georges Pompidou University Hospital Clinical Data Warehouse: A 8-years follow-up experience ». *International Journal of Medical Informatics* 102 (juin): 21-28. <https://doi.org/10.1016/j.ijmedinf.2017.02.006>.

Kang, John, Russell Schwartz, John Flickinger, et Sushil Beriwal. 2015. « Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective ». *International Journal of Radiation Oncology, Biology, Physics* 93 (5): 1127-35. <https://doi.org/10.1016/j.ijrobp.2015.07.2286>.

Kassambara, Alboukadel, Marcin Kosinski, Przemyslaw Biecek, et Scheipl Fabian. 2017. *survminer: Drawing Survival Curves using «ggplot2»* (version 0.4.1). <https://cran.r-project.org/web/packages/survminer/index.html>.

Klement, Rainer J., Michael Allgäuer, Steffen Appold, Karin Dieckmann, Iris Ernst, Ute Ganswindt, Richard Holy, et al. 2014. « Support Vector Machine-Based Prediction of Local Tumor Control after Stereotactic Body Radiation Therapy for Early-Stage Non-Small Cell Lung Cancer ». *International Journal of Radiation Oncology, Biology, Physics* 88 (3): 732-38. <https://doi.org/10.1016/j.ijrobp.2013.11.216>.

Kohane, Isaac S., Jeffrey M. Drazen, et Edward W. Campion. 2012. « A Glimpse of the next 100 Years in Medicine ». *The New England Journal of Medicine* 367 (26): 2538-39. <https://doi.org/10.1056/NEJMe1213371>.

Kourou, Konstantina, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, et Dimitrios I. Fotiadis. 2015. « Machine Learning Applications in Cancer Prognosis and Prediction ». *Computational and Structural Biotechnology Journal* 13: 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>.

Krizhevsky, Alex, Ilya Sutskever, et Geoffrey E. Hinton. 2012. « Imagenet classification with deep convolutional neural networks ». In *Advances in neural information processing systems*, 1097–1105.

Kumar, Virendra, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven A. Eschrich, Matthew B. Schabath, Kenneth Forster, et al. 2012. « Radiomics: The Process and the Challenges ». *Magnetic Resonance Imaging* 30 (9): 1234-48. <https://doi.org/10.1016/j.mri.2012.06.010>.

Lambin, Philippe, Ralph T. H. Leijenaar, Timo M. Deist, Jurgen Peerlings, Evelyn E. C. de Jong, Janita van Timmeren, Sebastian Sanduleanu, et al. 2017. « Radiomics: The Bridge between Medical Imaging and Personalized Medicine ». *Nature Reviews. Clinical Oncology* 14 (12): 749-62. <https://doi.org/10.1038/nrclinonc.2017.141>.

Lambin, Philippe, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G. P. M. van Stiphout, Patrick Granton, Catharina M. L. Zegers, et al. 2012. « Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis ». *European Journal of Cancer (Oxford, England: 1990)* 48 (4): 441-46. <https://doi.org/10.1016/j.ejca.2011.11.036>.

Lambin, Philippe, Erik Roelofs, Bart Reymen, Emmanuel Rios Velazquez, Jeroen Buijsen, Catharina M. L. Zegers, Sara Carvalho, et al. 2013. « “Rapid Learning Health Care in Oncology” - an Approach towards Decision Support Systems Enabling Customised Radiotherapy’ ». *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 109 (1): 159-64. <https://doi.org/10.1016/j.radonc.2013.07.007>.

Lambin, Philippe, Ruud G. P. M. van Stiphout, Maud H. W. Starman, Emmanuel Rios-Velazquez, Georgi Nalbantov, Hugo J. W. L. Aerts, Erik Roelofs, et al. 2013. « Predicting Outcomes in Radiation Oncology--Multifactorial Decision Support Systems ». *Nature Reviews. Clinical Oncology* 10 (1): 27-40. <https://doi.org/10.1038/nrclinonc.2012.196>.

Langley, Pat, Wayne Iba, et Kevin Thompson. 1992. « An analysis of Bayesian classifiers ». In *Aaai*, 90:223–228.

Langley, Pat, et Stephanie Sage. 1994. « Induction of selective Bayesian classifiers ». In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, 399–406. Morgan Kaufmann Publishers Inc. <http://dl.acm.org/citation.cfm?id=2074445>.

LeCun, Yann, Yoshua Bengio, et Geoffrey Hinton. 2015. « Deep Learning ». *Nature* 521 (7553): 436–44. <https://doi.org/10.1038/nature14539>.

Liu, Zhenyu, Xiao-Yan Zhang, Yan-Jie Shi, Lin Wang, Hai-Tao Zhu, Zhenchao Tang, Shuo Wang, Xiao-Ting Li, Jie Tian, et Ying-Shi Sun. 2017. « Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer ». *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, septembre. <https://doi.org/10.1158/1078-0432.CCR-17-1038>.

Lovinfosse, Pierre, Marc Polus, Daniel Van Daele, Philippe Martinive, Frédéric Daenen, Mathieu Hatt, Dimitris Visvikis, et al. 2017. « FDG PET/CT Radiomics for Predicting the Outcome of Locally Advanced Rectal Cancer ». *European Journal of Nuclear Medicine and Molecular Imaging*, octobre. <https://doi.org/10.1007/s00259-017-3855-5>.

Maas, Monique, Patty J. Nelemans, Vincenzo Valentini, Prajnan Das, Claus Rödel, Li-Jen Kuo, Felipe A. Calvo, et al. 2010. « Long-Term Outcome in Patients with a Pathological Complete Response after Chemoradiation for Rectal Cancer: A Pooled Analysis of Individual Patient Data ». *The Lancet. Oncology* 11 (9): 835–44. [https://doi.org/10.1016/S1470-2045\(10\)70172-8](https://doi.org/10.1016/S1470-2045(10)70172-8).

Machine Learning in Radiation Oncology - Theory and | Issam El Naqa | Springer. s. d. Consulté le 27 novembre 2017. [//www.springer.com/gp/book/9783319183046](http://www.springer.com/gp/book/9783319183046).

Mardis, Elaine R. 2011. « A Decade's Perspective on DNA Sequencing Technology ». *Nature* 470 (7333): 198–203. <https://doi.org/10.1038/nature09796>.

Martens, Milou H., Monique Maas, Luc A. Heijnen, Doenja M. J. Lambregts, Jeroen W. A. Leijtens, Laurents P. S. Stassen, Stephanie O. Breukink, et al. 2016. « Long-Term Outcome of an Organ Preservation Program After Neoadjuvant Treatment for Rectal Cancer ». *Journal of the National Cancer Institute* 108 (12). <https://doi.org/10.1093/jnci/djw171>.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig

Citro, Greg S. Corrado, et al. 2015. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. <https://www.tensorflow.org/>.

Metzker, Michael L. 2010. « Sequencing Technologies - the next Generation ». *Nature Reviews. Genetics* 11 (1): 31-46. <https://doi.org/10.1038/nrg2626>.

Miotto, Riccardo, Li Li, Brian A. Kidd, et Joel T. Dudley. 2016. « Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records ». *Scientific Reports* 6 (mai): 26094. <https://doi.org/10.1038/srep26094>.

Musen, Mark A. 2015. « The Protégé Project: A Look Back and a Look Forward ». *AI matters* 1 (4): 4-12. <https://doi.org/10.1145/2757001.2757003>.

———. s. d. « protégé ». Consulté le 15 avril 2017. <http://protege.stanford.edu/>.

Myers, J.D., H.E. Pople, et R.A. Miller. 1982. « CADUCEUS: A Computerized Diagnostic Consultation System in Internal Medicine ». *Proceedings of the Annual Symposium on Computer Application in Medical Care*, novembre, 44-47.

Nair, Vinod, et Geoffrey Hinton. 2010. « Rectified Linear Units Improve Restricted Boltzmann Machines ». *Proceedings of the 27th International Conference on Machine Learning*, 2010.

National Comprehensive Cancer Network. s. d. « NCCN Clinical Practice Guidelines in Oncology ». Consulté le 27 novembre 2017. https://www.nccn.org/professionals/physician_gls/default.aspx#site.

Nie, Ke, Liming Shi, Qin Chen, Xi Hu, Salma K. Jabbour, Ning Yue, Tianye Niu, et Xiaonan Sun. 2016. « Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI ». *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 22 (21): 5256-64. <https://doi.org/10.1158/1078-0432.CCR-15-2997>.

Noy, Natalya F., Mark A. Musen, José L.V. Mejino, et Cornelius Rosse. s. d. « Pushing the envelope: challenges in a frame-based representation of human anatomy ». Consulté le 25 avril 2017. <http://www.sciencedirect.com/science/article/pii/S0169023X03001253>.

Oakden-Rayner, Luke, Gustavo Carneiro, Taryn Bessen, Jacinto C. Nascimento, Andrew P. Bradley, et Lyle J. Palmer. 2017. « Precision Radiology: Predicting Longevity Using Feature Engineering and Deep Learning Methods in a Radiomics Framework ». *Scientific Reports* 7 (1): 1648. <https://doi.org/10.1038/s41598-017-01931-w>.

Obermeyer, Ziad, et Thomas H. Lee. 2017. « Lost in Thought - The Limits of the Human Mind and the Future of Medicine ». *The New England Journal of Medicine* 377 (13): 1209-11. <https://doi.org/10.1056/NEJMp1705348>.

Ochi, Takashi, Kenya Murase, Takashi Fujii, Masashi Kawamura, et Junpei Ikezoe. 2002. « Survival Prediction Using Artificial Neural Networks in Patients with Uterine Cervical Cancer Treated by Radiation Therapy Alone ». *International Journal of Clinical Oncology* 7 (5): 294-300. <https://doi.org/10.1007/s101470200043>.

Park, Jong Seob, Jung Wook Huh, Yoon Ah Park, Yong Beom Cho, Seong Hyeon Yun, Hee Cheol Kim, Woo Yong Lee, et Ho-Kyung Chun. 2014. « A Circumferential Resection Margin of 1 Mm Is a Negative Prognostic Factor in Rectal Cancer Patients with and without Neoadjuvant Chemoradiotherapy ». *Diseases of the Colon and Rectum* 57 (8): 933-40. <https://doi.org/10.1097/DCR.0000000000000171>.

Parmar, Chintan, Patrick Grossmann, Johan Bussink, Philippe Lambin, et Hugo J. W. L. Aerts. 2015. « Machine Learning Methods for Quantitative Radiomic Biomarkers ». *Scientific Reports* 5 (août): 13087. <https://doi.org/10.1038/srep13087>.

Parmar, Chintan, Patrick Grossmann, Derek Rietveld, Michelle M. Rietbergen, Philippe Lambin, et Hugo J. W. L. Aerts. 2015. « Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer ». *Frontiers in Oncology* 5: 272. <https://doi.org/10.3389/fonc.2015.00272>.

Parmar, Chintan, Emmanuel Rios Velazquez, Ralph Leijenaar, Mohammed Jermoumi, Sara Carvalho, Raymond H. Mak, Sushmita Mitra, et al. 2014. « Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation ». *PloS One* 9 (7): e102107. <https://doi.org/10.1371/journal.pone.0102107>.

Patrick, E. A., et F. P. Fischer III. 1970. « A generalized k-nearest neighbor rule ». *Information and Control* 16 (2): 128-52. [https://doi.org/10.1016/S0019-9958\(70\)90081-1](https://doi.org/10.1016/S0019-9958(70)90081-1).

Pella, Andrea, Raffaella Cambria, Marco Riboldi, Barbara Alicja Jereczek-Fossa, Cristiana Fodor, Dario Zerini, Ahmad Esmaili Torshabi, et al. 2011. « Use of Machine Learning Methods for Prediction of Acute Toxicity in Organs at Risk Following Prostate Radiotherapy ». *Medical Physics* 38 (6): 2859-67.

Ptok, Henry, Reinhard Ruppert, Joachim Stassburg, Christoph Andreas Maurer, Katja Oberholzer, Theodor Junginger, Susanne Merkel, et Paul Hermanek. 2013. « Pretherapeutic MRI for Decision-Making Regarding Selective Neoadjuvant Radiochemotherapy for Rectal Carcinoma: Interim Analysis of a Multicentric

Prospective Observational Study ». *Journal of Magnetic Resonance Imaging: JMRI* 37 (5): 1122-28. <https://doi.org/10.1002/jmri.23917>.

Quinlan, J. Ross. 1986. « Induction of decision trees ». *Machine learning* 1 (1): 81–106.

R Core Team. 2013. « R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/> ».

« Radiation Oncology Ontology - Summary | NCBO BioPortal ». s. d. Consulté le 7 mars 2016. <http://bioportal.bioontology.org/ontologies/ROO>.

Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, et al. 2017. « CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning », novembre. <https://arxiv.org/abs/1711.05225>.

Rance, Bastien, Vincent Canuel, Hector Countouris, Pierre Laurent-Puig, et Anita Burgun. 2016. « Integrating Heterogeneous Biomedical Data for Cancer Research: The CARPEM Infrastructure ». *Applied Clinical Informatics* 7 (2): 260-74. <https://doi.org/10.4338/ACI-2015-09-RA-0125>.

Renehan, Andrew G., Lee Malcomson, Richard Emsley, Simon Gollins, Andrew Maw, Arthur Sun Myint, Paul S. Rooney, et al. 2016. « Watch-and-Wait Approach versus Surgical Resection after Chemoradiotherapy for Patients with Rectal Cancer (the OnCoRe Project): A Propensity-Score Matched Cohort Analysis ». *The Lancet. Oncology* 17 (2): 174-83. [https://doi.org/10.1016/S1470-2045\(15\)00467-2](https://doi.org/10.1016/S1470-2045(15)00467-2).

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, Markus Müller, et Stefan Siegert (Fast DeLong code). 2018. *pROC: Display and Analyze ROC Curves* (version 1.12.1). <https://CRAN.R-project.org/package=pROC>.

Rubin, Daniel L. 2008. « Creating and Curating a Terminology for Radiology: Ontology Modeling and Analysis ». *Journal of Digital Imaging* 21 (4): 355-62. <https://doi.org/10.1007/s10278-007-9073-0>.

Rumelhart, David Everett, et James McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press. Cambridge.

Rusch, Valerie W., Hisao Asamura, Hirokazu Watanabe, Dorothy J. Giroux, Ramon Rami-Porta, Peter Goldstraw, et Members of IASLC Staging Committee. 2009. « The IASLC Lung Cancer Staging Project: A Proposal for a New International Lymph Node Map in the Forthcoming Seventh Edition of the TNM Classification for Lung Cancer ».

Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer 4 (5): 568-77.
<https://doi.org/10.1097/JTO.obo13e3181a0d82e>.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2015. « ImageNet Large Scale Visual Recognition Challenge ». *International Journal of Computer Vision* 115 (3): 211-52.
<https://doi.org/10.1007/s11263-015-0816-y>.

« scikit-learn: machine learning in Python — scikit-learn 0.19.1 documentation ». s. d. Consulté le 27 avril 2018. <http://scikit-learn.org/stable/>.

Sebag-Montefiore, David, Richard J. Stephens, Robert Steele, John Monson, Robert Grieve, Subhash Khanna, Phil Quirke, et al. 2009. « Preoperative Radiotherapy versus Selective Postoperative Chemoradiotherapy in Patients with Rectal Cancer (MRC CR07 and NCIC-CTG Co16): A Multicentre, Randomised Trial ». *Lancet (London, England)* 373 (9666): 811-20. [https://doi.org/10.1016/S0140-6736\(09\)60484-0](https://doi.org/10.1016/S0140-6736(09)60484-0).

Shortliffe, Edward H. 1977. « Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases ». *Proceedings of the Annual Symposium on Computer Application in Medical Care*, octobre, 66-69.

Shrout, P. E., et J. L. Fleiss. 1979. « Intraclass Correlations: Uses in Assessing Rater Reliability ». *Psychological Bulletin* 86 (2): 420-28.

Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. 2016. « Mastering the Game of Go with Deep Neural Networks and Tree Search ». *Nature* 529 (7587): 484-89.
<https://doi.org/10.1038/nature16961>.

Skripcak, Tomas, Claus Belka, Walter Bosch, Carsten Brink, Thomas Brunner, Volker Budach, Daniel Büttner, et al. 2014. « Creating a Data Exchange Strategy for Radiotherapy Research: Towards Federated Databases and Anonymised Public Datasets ». *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 113 (3): 303-9.
<https://doi.org/10.1016/j.radonc.2014.10.001>.

Su, Min, Moyed Miften, Curtis Whiddon, Xuejun Sun, Kim Light, et Lawrence Marks. 2005. « An Artificial Neural Network for Predicting the Incidence of Radiation Pneumonitis ». *Medical Physics* 32 (2): 318-25.

« TensorFlow -- an Open Source Software Library for Machine Intelligence ». s. d. Consulté le 12 mars 2016. <https://www.tensorflow.org/>.

Therneau, Terry M., et Thomas Lumley (original S.->R port and maintainer until 2009). 2017. *survival: Survival Analysis* (version 2.41-3). <https://cran.r-project.org/web/packages/survival/index.html>.

Tomatis, S., T. Rancati, C. Fiorino, V. Vavassori, G. Fellin, E. Cagna, F. A. Mauro, et al. 2012. « Late Rectal Bleeding after 3D-CRT for Prostate Cancer: Development of a Neural-Network-Based Predictive Model ». *Physics in Medicine and Biology* 57 (5): 1399-1412. <https://doi.org/10.1088/0031-9155/57/5/1399>.

« Torch | Scientific computing for LuaJIT. » s. d. Consulté le 7 juillet 2016. <http://torch.ch/>.

Vapnik, Vladimir. 1982. *Estimation of Dependences Based on Empirical Data*. Springer Verlag. New-York.

Varian Incorporated. s. d. « Varian Developers Forum ». CodePlex. Consulté le 15 avril 2017. <https://variandeveloper.codeplex.com/Wikipage?ProjectName=variandeveloper>.

« Watson Oncology ». s. d. Memorial Sloan Kettering Cancer Center. Consulté le 10 mars 2016. <https://www.mskcc.org/about/innovative-collaborations/watson-oncology>.

Wickham, H. 2009. « ggplot2: Elegant Graphics for Data Analysis. » Springer-Verlag New York.

Wolak, Matthew. 2015. *ICC: Facilitating Estimation of the Intraclass Correlation Coefficient* (version 2.3.0). <https://cran.r-project.org/web/packages/ICC/index.html>.

Zapletal, Eric, Nicolas Rodon, Natalia Grabar, et Patrice Degoulet. 2010a. « Methodology of Integration of a Clinical Data Warehouse with a Clinical Information System: The HEGP Case ». *Studies in Health Technology and Informatics* 160 (Pt 1): 193-97.

———. 2010b. « Methodology of Integration of a Clinical Data Warehouse with a Clinical Information System: The HEGP Case ». *Studies in Health Technology and Informatics* 160 (Pt 1): 193-97.

Zhang, Lifei, David V. Fried, Xenia J. Fave, Luke A. Hunter, Jinzhong Yang, et Laurence E. Court. 2015. « IBEX: An Open Infrastructure Software Platform to Facilitate Collaborative Work in Radiomics ». *Medical Physics* 42 (3): 1341-53. <https://doi.org/10.1118/1.4908210>.

Zhou, Zhiguo, Michael Folkert, Nathan Cannon, Puneeth Iyengar, Kenneth Westover, Yuanyuan Zhang, Hak Choy, et al. 2016. « Predicting Distant Failure in Early Stage

NSCLC Treated with SBRT Using Clinical Parameters ». *Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology* 119 (3): 501-4. <https://doi.org/10.1016/j.radonc.2016.04.029>.

Annexes

Annexe 1 – Liste des variables radiomics extraites par IBEX

Category	Feature algorithm
Shape	SurfaceArea
	SurfaceAreaDensity
	Mass
	Convex
	ConvexHullVolume
	ConvexHullVolume3D
	MeanBreath
	Orientation
	Roundness
	NumberOfObjects
	NumberOfVoxel
	VoxelSize
Intensity Direct	Skewness
	Range
	Percentile
	Quantile
	InterQuartileRange
	GlobalEntropy
	GlobalUniformity
	GlobalMax
	GlobalMin
	GlobalMean
	GlobalMedian
	GlobalStd
	MeanAbsoluteDeviation
	MedianAbsoluteDeviation
	LocalEntropy/Range/StdMax

	LocalEntropy/Range/StdMin
	LocalEntropy/Range/StdMean
	LocalEntropy/Range/StdMedian
	LocalEntropy/Range/StdStd
Gray Level Cooccurrence Matrix 25	Contrast
Gray Level Cooccurrence Matrix 3	Correlation
	Energy
	Homogeneity
Neighbor Intensity Difference 25	Busyness
Neighbor Intensity Difference 3	Coarseness
	Complexity
	Contrast
	TextureStrength
Gray Level Run Length Matrix 25	GrayLevelNonuniformity
Gray Level Run Length Matrix 3	HighGrayLevelRunEmphasis
	LongRunEmphasissis
	LongRunHighGrayLevelEmphasis
	LongRunLowGrayLevelEmphasis
	LowGrayLevelRunEmphasis
	RunLengthNonuniformity
	RunPercentage
	ShortRunEmphasissis
	ShortRunHighGrayLevelEmphasis
	ShortRunLowGrayLevelEmphasis

Annexe 2 – Code python pour créer un DNN sous TensorFlow

```
from __future__ import absolute_import
from __future__ import division
from __future__ import print_function

import os
import urllib
import numpy as np
import tensorflow as tf

# Data sets
Train = "Train.csv"
Test = "Test.csv"

def main():
    # Load datasets
    training_set =
    tf.contrib.learn.datasets.base.load_csv_with_header(
        filename=Train,
        target_dtype=np.int,
        features_dtype=np.float32)
    test_set =
    tf.contrib.learn.datasets.base.load_csv_with_header(
        filename=Test,
        target_dtype=np.int,
        features_dtype=np.float32)

    # Specify that all features have real-value data
    feature_columns = [tf.feature_column.numeric_column("x",
    shape=[28])]

    # Build 3 layer DNN with 10, 20, 10 units respectively
    classifier =
    tf.estimator.DNNClassifier(feature_columns=feature_columns,

    hidden_units=[10,20,10],
    n_classes=2,

    model_dir="/tmp/RadiomicsICC_model6")
```

```

# Define the training inputs
train_input_fn = tf.estimator.inputs.numpy_input_fn(
    x={"x": np.array(training_set.data)},
    y=np.array(training_set.target),
    num_epochs=None,
    shuffle=True)

# Train model
classifier.train(input_fn=train_input_fn, steps=1000)

# Define the test inputs
test_input_fn = tf.estimator.inputs.numpy_input_fn(
    x={"x": np.array(test_set.data)},
    y=np.array(test_set.target),
    num_epochs=3,
    shuffle=False)

# Evaluate model
accuracy_score =
classifier.evaluate(input_fn=test_input_fn) ["accuracy"]

print("\nAccuracy: {0:f}\n".format(accuracy_score))

if __name__ == "__main__":
    main()

```