



HAL
open science

Évaluation de différentes approches d'analyse radiomique pour prédire le risque de récurrence chez des patients atteints de tumeurs rectales localement avancées

Hichem Tibermacine

► To cite this version:

Hichem Tibermacine. Évaluation de différentes approches d'analyse radiomique pour prédire le risque de récurrence chez des patients atteints de tumeurs rectales localement avancées. Human health and pathology. 2020. dumas-03140051

HAL Id: dumas-03140051

<https://dumas.ccsd.cnrs.fr/dumas-03140051>

Submitted on 12 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

UNIVERSITÉ DE MONTPELLIER
FACULTÉ DE MÉDECINE MONTPELLIER-NÎMES

THESE

Pour obtenir le titre de
DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement par
TIBERMACHINE Hichem

Le 09/03/2020

Évaluation de différentes approches d'analyse radiomique
pour prédire le risque de récurrence chez des patients atteints de
tumeurs rectales localement avancées

Directrice de thèse : **Madame la Docteure NOUGARET Stéphanie**
Praticien hospitalier – Radiologie

JURY

Président : **Monsieur le Professeur KOTZKI Olivier**
Professeur des universités - Médecine nucléaire

Assesseurs : **Monsieur le Professeur COLOMBO Pierre-Emmanuel**
Professeur des universités - Chirurgie oncologique
Madame la Docteure BOURGIER Céline
Maître de conférences des universités - Radiothérapie
Madame la Docteure NOUGARET Stéphanie
Praticien hospitalier – Radiologie

Invitée : **Madame la Professeure MILLET Ingrid**
Professeur des universités - Radiologie

UNIVERSITÉ DE MONTPELLIER
FACULTÉ DE MÉDECINE MONTPELLIER-NÎMES

THESE

Pour obtenir le titre de
DOCTEUR EN MÉDECINE

Présentée et soutenue publiquement par
TIBERMACHINE Hichem

Le 09/03/2020

Évaluation de différentes approches d'analyse radiomique
pour prédire le risque de récurrence chez des patients atteints de
tumeurs rectales localement avancées

Directrice de thèse : **Madame la Docteure NOUGARET Stéphanie**
Praticien hospitalier – Radiologie

JURY

Président : **Monsieur le Professeur KOTZKI Olivier**
Professeur des universités - Médecine nucléaire

Assesseurs : **Monsieur le Professeur COLOMBO Pierre-Emmanuel**
Professeur des universités - Chirurgie oncologique
Madame la Docteure BOURGIER Céline
Maître de conférences des universités - Radiothérapie
Madame la Docteure NOUGARET Stéphanie
Praticien hospitalier – Radiologie

Invitée : **Madame la Professeure MILLET Ingrid**
Professeur des universités - Radiologie



ANNEE UNIVERSITAIRE 2017 - 2018

PERSONNEL ENSEIGNANT

Professeurs Honoraires

ALLIEU Yves
ALRIC Robert
ARNAUD Bernard
ASTRUC Jacques
AUSSILLOUX Charles
AVEROUS Michel
AYRAL Guy
BAILLAT Xavier
BALDET Pierre
BALDY-MOULINIER Michel
BALMES Jean-Louis
BALMES Pierre
BANSARD Nicole
BAYLET René
BILLIARD Michel
BLARD Jean-Marie
BLAYAC Jean Pierre
BLOTMAN Francis
BONNEL François
BOUDET Charles
BOURGEOIS Jean-Marie
BRUEL Jean Michel
BUREAU Jean-Paul
BRUNEL Michel
CALLIS Albert
CANAUD Bernard
CASTELNAU Didier
CHAPTAL Paul-André
CIURANA Albert-Jean
CLOT Jacques
D'ATHIS Françoise
DEMAILLE Jacques
DESCOMPS Bernard

DIMEGLIO Alain
DU CAILAR Jacques
DUBOIS Jean Bernard
DUMAS Robert
DUMAZER Romain
ECHENNE Bernard
FABRE Serge
FREREBEAU Philippe
GALIFER René Benoît
GODLEWSKI Guilhem
GRASSET Daniel
GROLLEAU-RAOUX Robert
GUILHOU Jean-Jacques
HERTAULT Jean
HUMEAU Claude
JAFFIOL Claude
JANBON Charles
JANBON François
JARRY Daniel
JOYEUX Henri
LAFFARGUE François
LALLEMANT Jean Gabriel
LAMARQUE Jean-Louis
LAPEYRIE Henri
LESBROS Daniel
LOPEZ François Michel
LORIOT Jean
LOUBATIERES Marie Madeleine
MAGNAN DE BORNIER Bernard
MARY Henri
MATHIEU-DAUDE Pierre

MEYNADIER Jean
MICHEL François-Bernard
MICHEL Henri
MION Charles
MION Henri
MIRO Luis
NAVARRO Maurice
NAVRATIL Henri
OTHONIEL Jacques
PAGES Michel
PEGURET Claude
POUGET Régis
PUECH Paul
PUJOL Henri
PUJOL Rémy
RABISCHONG Pierre
RAMUZ Michel
RIEU Daniel
RIOUX Jean-Antoine
ROCHFORT Henri
ROUANET DE VIGNE
LAVIT Jean Pierre
SAINT AUBERT Bernard
SANCHO-GARNIER Héléène
SANY Jacques
SENAC Jean-Paul
SERRE Ariette
SIMON Lucien
SOLASSOL Claude
THEVENET André
VIDAL Jacques
VISIER Jean Pierre

Professeurs Emérites

ARTUS Jean-Claude
BLANC François
BOULENGER Jean-Philippe
BOURREL Gérard
BRINGER Jacques
CLAUSTRES Mireille
DAURES Jean-Pierre
DAUZAT Michel
DEDET Jean-Pierre
ELEDJAM Jean-Jacques
GUERRIER Bernard
JOURDAN Jacques
MAURY Michèle

MILLAT Bertrand
MARES Pierre
MONNIER Louis
PRAT Dominique
PRATLONG Francine
PREFAUT Christian
PUJOL Rémy
ROSSI Michel
SULTAN Charles
TOUCHON Jacques
VOISIN Michel
ZANCA Michel

Professeurs des Universités - Praticiens Hospitaliers

PU-PH de classe exceptionnelle

ALBAT Bernard - Chirurgie thoracique et cardiovasculaire
ALRIC Pierre - Chirurgie vasculaire ; médecine vasculaire (option chirurgie vasculaire)
BACCINO Eric - Médecine légale et droit de la santé
BASTIEN Patrick - Parasitologie et mycologie
BONAFE Alain - Radiologie et imagerie médicale
CAPDEVILA Xavier - Anesthésiologie-réanimation
COMBE Bernard - Rhumatologie
COSTA Pierre - Urologie
COTTALORDA Jérôme - Chirurgie infantile
COUBES Philippe - Neurochirurgie
CRAMPETTE Louis - Oto-rhino-laryngologie
CRISTOL Jean Paul - Biochimie et biologie moléculaire
DAVY Jean Marc - Cardiologie
DE LA COUSSAYE Jean Emmanuel - Anesthésiologie-réanimation
DELAPORTE Eric - Maladies infectieuses ; maladies tropicales
DE WAZIERES Benoît - Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
DOMERGUE Jacques - Chirurgie générale
DUFFAU Hugues - Neurochirurgie
DUJOLS Pierre - Biostatistiques, informatique médicale et technologies de la communication
ELIAOU Jean François - Immunologie
FABRE Jean Michel - Chirurgie générale
GUILLOT Bernard - Dermato-vénérologie
HAMAMAH Samir - Biologie et Médecine du développement et de la reproduction ; gynécologie médicale
HEDON Bernard - Gynécologie-obstétrique ; gynécologie médicale
HERISSON Christian - Médecine physique et de réadaptation
JABER Samir - Anesthésiologie-réanimation
JEANDEL Claude - Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
JONQUET Olivier - Réanimation ; médecine d'urgence
JORGENSEN Christian - Thérapeutique ; médecine d'urgence ; addictologie
KOTZKI Pierre Olivier - Biophysique et médecine nucléaire
LANDAIS Paul - Epidémiologie, Economie de la santé et Prévention
LARREY Dominique - Gastroentérologie ; hépatologie ; addictologie
LEFRANT Jean-Yves - Anesthésiologie-réanimation
LE QUELLEC Alain - Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
MARTY-ANE Charles - Chirurgie thoracique et cardiovasculaire
MAUDELONDE Thierry - Biologie cellulaire
MERCIER Jacques - Physiologie
MESSNER Patrick - Cardiologie
MOURAD Georges - Néphrologie
PELLISSIER Jacques - Médecine physique et de réadaptation
RENARD Eric - Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
REYNES Jacques - Maladies infectieuses, maladies tropicales
RIBSTEIN Jean - Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
RIPART Jacques - Anesthésiologie-réanimation
ROUANET Philippe - Cancérologie ; radiothérapie
SCHVED Jean François - Hématologie ; Transfusion
TAOUREL Patrice - Radiologie et imagerie médicale
UZIEL Alain - Oto-rhino-laryngologie
VANDE PERRE Philippe - Bactériologie-virologie ; hygiène hospitalière

YCHOU Marc-Cancérologie ; radiothérapie

PU-PH de 1^{re} classe

AGUILAR MARTINEZ Patricia-Hématologie ; transfusion
AVIGNON Antoine-Nutrition
AZRIA David -Cancérologie ; radiothérapie
BAGHDADLI Amaria-Pédopsychiatrie ; addictologie
BEREGI Jean-Paul-Radiologie et imagerie médicale
BLAIN Hubert-Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
BLANC Pierre-Gastroentérologie ; hépatologie ; addictologie
BORIE Frédéric-Chirurgie digestive
BOULOT Pierre-Gynécologie-obstétrique ; gynécologie médicale
CAMBONIE Gilles -Pédiatrie
CAMU William-Neurologie
CANOVAS François-Anatomie
CARTRON Guillaume-Hématologie ; transfusion
CHAMMAS Michel-Chirurgie orthopédique et traumatologique
COLSON Pascal-Anesthésiologie-réanimation
CORBEAU Pierre-Immunologie
COSTES Valérie-Anatomie et cytologie pathologiques
COURTET Philippe-Psychiatrie d'adultes ; addictologie
CYTEVAL Catherine-Radiologie et imagerie médicale
DADURE Christophe-Anesthésiologie-réanimation
DAUVILLIERS Yves-Physiologie
DE TAYRAC Renaud-Gynécologie-obstétrique, gynécologie médicale
DEMARIA Roland-Chirurgie thoracique et cardio-vasculaire
DEMOLY Pascal-Pneumologie ; addictologie
DEREURE Olivier-Dermatologie - vénéréologie
DROUPY Stéphane -Urologie
DUCROS Anne-Neurologie -
FRAPIER Jean-Marc-Chirurgie thoracique et cardiovasculaire
KLOUCHE Kada-Réanimation ; médecine d'urgence
KOENIG Michel-Génétique moléculaire
LABAUGE Pierre- Neurologie
LAFFONT Isabelle-Médecine physique et de réadaptation
LAVABRE-BERTRAND Thierry-Cytologie et histologie
LECLERCQ Florence-Cardiologie
LEHMANN Sylvain-Biochimie et biologie moléculaire
LUMBROSO Serge-Biochimie et Biologie moléculaire
MARIANO-GOULART Denis-Biophysique et médecine nucléaire
MATECKI Stéfan -Physiologie
MEUNIER Laurent-Dermato-vénéréologie
MONDAIN Michel-Oto-rhino-laryngologie
MORIN Denis-Pédiatrie
NAVARRO Francis-Chirurgie générale
PAGEAUX Georges-Philippe-Gastroentérologie ; hépatologie ; addictologie
PETIT Pierre-Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
PERNEY Pascal-Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
PUJOL Jean Louis-Pneumologie ; addictologie
PUJOL Pascal-Biologie cellulaire
PURPER-OUAKIL Diane-Pédopsychiatrie ; addictologie
QUERE Isabelle-Chirurgie vasculaire ; médecine vasculaire (option médecine vasculaire)
SOTTO Albert-Maladies infectieuses ; maladies tropicales
TOUITOU Isabelle-Génétique
TRAN Tu-Anh-Pédiatrie
VERN HET Hélène-Radiologie et imagerie médicale

PU-PH de 2ème classe

ASSENAT Eric-Gastroentérologie ; hépatologie ; addictologie
BERTHET Jean-Philippe-Chirurgie thoracique et cardiovasculaire
BOURDIN Arnaud-Pneumologie ; addictologie
CANAUD Ludovic-Chirurgie vasculaire ; Médecine Vasculaire
CAPDEVIELLE Delphine-Psychiatrie d'Adultes ; addictologie
CAPTIER Guillaume-Anatomie
CAYLA Guillaume-Cardiologie
CHANQUES Gérard-Anesthésiologie-réanimation
COLOMBO Pierre-Emmanuel-Cancérologie ; radiothérapie
COSTALAT Vincent-Radiologie et imagerie médicale
COULET Bertrand-Chirurgie orthopédique et traumatologique
CUVILLON Philippe-Anesthésiologie-réanimation
DAIEN Vincent-Ophtalmologie
DE VOS John-Cytologie et histologie
DORANDEU Anne-Médecine légale -
DUPEYRON Arnaud-Médecine physique et de réadaptation
FESLER Pierre-Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
GARREL Renaud-Oto-rhino-laryngologie
GAUJOUX Viala Cécile-Rhumatologie
GENEVIEVE David-Génétique
GODREUIL Sylvain-Bactériologie-virologie ; hygiène hospitalière
GUILLAUME Sébastien-Urgences et Post urgences psychiatriques -
GUILPAIN Philippe-Médecine Interne, gériatrie et biologie du vieillissement, addictologie
GUIU Boris-Radiologie et imagerie médicale
HAYOT Maurice-Physiologie
HOUEDE Nadine-Cancérologie ; radiothérapie
JACOT William-Cancérologie ; Radiothérapie
JUNG Boris-Réanimation ; médecine d'urgence
KALFA Nicolas-Chirurgie infantile
KOUYOUMDJIAN Pascal-Chirurgie orthopédique et traumatologique
LACHAUD Laurence-Parasitologie et mycologie
LALLEMANT Benjamin-Oto-rhino-laryngologie
LAVIGNE Jean-Philippe-Bactériologie-virologie ; hygiène hospitalière
LE MOING Vincent-Maladies infectieuses ; maladies tropicales
LETOUZEY Vincent-Gynécologie-obstétrique ; gynécologie médicale
LOPEZ CASTROMAN Jorge-Psychiatrie d'Adultes ; addictologie
LUKAS Cédric-Rhumatologie
MAURY Philippe-Chirurgie orthopédique et traumatologique
MILLET Ingrid-Radiologie et imagerie médicale
MORANNE Olivier-Néphrologie
MOREL Jacques -Rhumatologie
NAGOT Nicolas-Biostatistiques, informatique médicale et technologies de la communication
NOCCA David-Chirurgie digestive
PANARO Fabrizio-Chirurgie générale
PARIS Françoise-Biologie et médecine du développement et de la reproduction ; gynécologie médicale
PASQUIE Jean-Luc-Cardiologie
PEREZ MARTIN Antonia-Physiologie
POUDEROUX Philippe-Gastroentérologie ; hépatologie ; addictologie
PRUDHOMME Michel-Anatomie
RIGAU Valérie-Anatomie et cytologie pathologiques
RIVIER François-Pédiatrie
ROGER Pascal-Anatomie et cytologie pathologiques
ROSSI Jean François-Hématologie ; transfusion
ROUBILLE François-Cardiologie

SEBBANE Mustapha-Anesthésiologie-réanimation
SEGNARBIEUX François-Neurochirurgie
SIRVENT Nicolas-Pédiatrie
SOLASSOL Jérôme-Biologie cellulaire
SULTAN Ariane-Nutrition
THOUVENOT Éric-Neurologie
THURET Rodolphe-Urologie
VENAIL Frédéric-Oto-rhino-laryngologie
VILLAIN Max-Ophthalmologie
VINCENT Denis -Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
VINCENT Thierry-Immunologie
WOJTUSCISZYN Anne-Endocrinologie-diabétologie-nutrition

PROFESSEURS DES UNIVERSITES

1^{re} classe :

COLINGE Jacques - Cancérologie, Signalisation cellulaire et systèmes complexes

2^{ème} classe :

LAOUDJ CHENIVESSE Dalila - Biochimie et biologie moléculaire

VISIER Laurent - Sociologie, démographie

PROFESSEURS DES UNIVERSITES - Médecine générale

1^{re} classe :

LAMBERT Philippe

2^{ème} classe :

AMOUYAL Michel

PROFESSEURS ASSOCIES - Médecine Générale

DAVID Michel

RAMBAUD Jacques

PROFESSEUR ASSOCIE - Médecine

BESSIS Didier - Dermato-vénérologie

PERRIGAULT Pierre-François - Anesthésiologie-réanimation ; médecine d'urgence

ROUBERTIE Agathe – Pédiatrie

Maîtres de Conférences des Universités - Praticiens Hospitaliers

MCU-PH Hors classe

CACHEUX-RATABOUL Valère-Génétique

CARRIERE Christian-Bactériologie-virologie ; hygiène hospitalière

CHARACHON Sylvie-Bactériologie-virologie ; hygiène hospitalière

FABBRO-PERAY Pascale-Epidémiologie, économie de la santé et prévention

HILLAIRE-BUYS Dominique-Pharmacologie fondamentale ; pharmacologie clinique ; addictologie

PELLESTOR Franck-Cytologie et histologie

PUJOL Joseph-Anatomie

RAMOS Jeanne-Anatomie et cytologie pathologiques

RICHARD Bruno-Thérapeutique ; addictologie

RISPAIL Philippe-Parasitologie et mycologie

SEGONDY Michel-Bactériologie-virologie ; hygiène hospitalière

STOEBNER Pierre-Dermato-vénérologie

MCU-PH de 1^{re} classe

ALLARDET-SERVENT Annick-Bactériologie-virologie ; hygiène hospitalière

BADIOU Stéphanie-Biochimie et biologie moléculaire
BOUDOUSQ Vincent-Biophysique et médecine nucléaire
BOULLE Nathalie-Biologie cellulaire
BOURGIER Céline-Cancérologie ; Radiothérapie
BRET Caroline-Hématologie biologique
COSSEE Mireille-Génétique Moléculaire
GABELLE DELOUSTAL Audrey-Neurologie
GIANSILY-BLAIZOT Muriel-Hématologie ; transfusion
GIRARDET-BESSIS Anne-Biochimie et biologie moléculaire
LAVIGNE Géraldine-Hématologie ; transfusion
LE QUINTREC Moglie-Néphrologie
MATHIEU Olivier-Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
MENJOT de CHAMPFLEUR Nicolas-Neuroradiologie
MOUZAT Kévin-Biochimie et biologie moléculaire
PANABIERES Catherine-Biologie cellulaire
PHILIBERT Pascal-Biologie et médecine du développement et de la reproduction
RAVEL Christophe - Parasitologie et mycologie
SCHUSTER-BECK Iris-Physiologie
STERKERS Yvon-Parasitologie et mycologie
TUAILLON Edouard-Bactériologie-virologie ; hygiène hospitalière
YACHOUH Jacques-Chirurgie maxillo-faciale et stomatologie

MCU-PH de 2^{ème} classe

BERTRAND Martin-Anatomie
BRUN Michel-Bactériologie-virologie ; hygiène hospitalière
DU THANH Aurélie-Dermato-vénérologie
GALANAUD Jean Philippe-Médecine Vasculaire
GOUZI Farès-Physiologie
JEZIORSKI Eric-Pédiatrie
KUSTER Nils-Biochimie et biologie moléculaire
LESAGE François-Xavier-Médecine et Santé au Travail
MAKINSON Alain-Maladies infectieuses, Maladies tropicales
MURA Thibault-Biostatistiques, informatique médicale et technologies de la communication
OLIE Emilie-Psychiatrie d'adultes ; addictologie
THEVENIN-RENE Céline-Immunologie

MAITRES DE CONFERENCES DES UNIVERSITES - Médecine Générale

COSTA David
FOLCO-LOGNOS Béatrice

MAITRES DE CONFERENCES ASSOCIES - Médecine Générale

CLARY Bernard
GARCIA Marc
MILLION Elodie
PAVAGEAU Sylvain
REBOUL Marie-Catherine
SEGURET Pierre

MAITRES DE CONFERENCES DES UNIVERSITES

Maitres de Conférences hors classe

BADIA Eric - Sciences biologiques fondamentales et cliniques

Maitres de Conférences de classe normale

BECAMEL Carine - Neurosciences
BERNEX Florence - Physiologie
CHAUMONT-DUBEL Séverine - Sciences du médicament et des autres produits de santé
CHAZAL Nathalie - Biologie cellulaire

DELABY Constance - Biochimie et biologie moléculaire
GUGLIELMI Laurence - Sciences biologiques fondamentales et cliniques
HENRY Laurent - Sciences biologiques fondamentales et cliniques
LADRET Véronique - Mathématiques appliquées et applications des mathématiques
LAINE Sébastien - Sciences du Médicament et autres produits de santé
LE GALLIC Lionel - Sciences du médicament et autres produits de santé
LOZZA Catherine - Sciences physico-chimiques et technologies pharmaceutiques
MAIMOUN Laurent - Sciences physico-chimiques et ingénierie appliquée à la santé
MOREAUX Jérôme - Science biologiques, fondamentales et cliniques
MORITZ-GASSER Sylvie - Neurosciences
MOUTOT Gilles - Philosophie
PASSERIEUX Emilie - Physiologie
RAMIREZ Jean-Marie - Histologie
TAULAN Magali - Biologie Cellulaire

PRATICIENS HOSPITALIERS UNIVERSITAIRES

CLAIRE DAIEN-Rhumatologie
BASTIDE Sophie-Epidémiologie, économie de la santé et prévention
FAILLIE Jean-Luc-
Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
GATINOIS Vincent-Histologie, embryologie et cytogénétique
HERLIN Christian -Chirurgie plastique ; reconstructrice et esthétique ; brûlologie
HERRERO Astrid-Chirurgie générale
PANTEL Alix-Bactériologie-virologie ; hygiène hospitalière
PERS Yves-Marie-Thérapeutique, médecine d'urgence ; addictologie
PINETON DE CHAMBRUN Guillaume-Gastroentérologie ; hépatologie ; addictologie
TORRE Antoine-Gynécologie-obstétrique ; gynécologie médicale

Remerciements

Au président et aux membres du jury :

Madame la Docteure NOUGARET Stéphanie

Praticien Hospitalier - Radiologie

Je te remercie d'avoir accepté de diriger ma thèse, de ton aide précieuse et de ta disponibilité durant la réalisation du travail et la rédaction de l'article.

Merci également pour tous tes enseignements sur la radiologie et en particulier sur l'IRM rectale.

Monsieur le Professeur KOTZKI Olivier

Professeur des universités - Médecine nucléaire

Je vous remercie d'avoir accepté d'évaluer mon travail, de votre participation à ce jury de thèse et de votre présence le jour de la soutenance.

Je vous remercie également pour tous vos enseignements durant ma formation initiale.

Madame La Professeure MILLET Ingrid

Je vous remercie d'avoir accepté d'évaluer mon travail et de votre participation à ce jury de thèse.

Je vous remercie également pour tous vos enseignements durant ma formation d'interne.

Monsieur le Professeur COLOMBO Pierre-Emmanuel

Professeur des universités – Chirurgie oncologique

Je vous remercie d'avoir accepté d'évaluer mon travail, de votre participation à ce jury de thèse et de votre présence le jour de soutenance.

Madame la Docteure BOURGIER Céline

Maître de conférences des universités - Radiothérapie

Je vous remercie d'avoir accepté d'évaluer mon travail, de votre participation à ce jury de thèse et de votre présence le jour de soutenance.

A tous ceux qui m'ont aidé à réaliser ce travail :

A mon frère Chouki, merci beaucoup pour ton aide immense dans la partie programmation.

A toute l'équipe de radiologie et médecine nucléaire du CHU de Montpellier pour leur soutien durant ce travail.

A l'équipe des attachés de recherche clinique de m'avoir donné accès à leurs bases de données.

A l'ensemble des équipes avec lesquelles j'ai travaillé durant mon internat :

A tous les radiologues, médecins nucléaires, et cliniciens qui m'ont accompagné tout au long de l'internat : merci de vos enseignements de grande qualité, de votre implication dans ma formation et de m'avoir transmis vos connaissances et votre passion de la médecine. C'est grâce à vous que j'en suis là aujourd'hui.

A tous les autres membres des équipes médicales et para-médicales (manipulateurs, secrétaires, infirmières, physiciens, techniciens...) : merci pour votre collaboration de chaque jour et de votre implication dans votre travail, toujours dans une bonne ambiance et au service du patient. Sans vous, rien de ce que nous faisons ne serait possible.

A tous mes collègues et amis :

A mes co-internes de Montpellier et mes camarades de la promotion de médecine de 2014 avec lesquels j'ai passé de supers bons moments en apprenant la médecine et la radiologie. Dans cette nouvelle vie professionnelle, j'espère continuer à vous voir et à travailler avec vous.

A tous mes amis et camarades de Montpellier qui m'ont accompagné durant mes premiers pas à la faculté et durant l'externat, avec lesquels j'ai passé de superbes années aussi bien dans les moments de détente qu'en stage ou durant les périodes plus difficiles de révisions. Même si nous sommes maintenant pour la plupart séparés par le fait de l'internat, je pense à vous.

A tous mes autres amis hors médecine : à mes autres amis du lycée ou d'ailleurs, d'avant ou d'aujourd'hui...

A ma famille :

A ma chère épouse Inès et à notre futur bébé, d'avoir ramené la joie et le bonheur dans ma vie.

A mes parents, à mes frères et sœurs (en particulier à mon frère Chouki) qui ont toujours été là pour moi et qui m'ont soutenu tout au long de mes études. Je vous aime et souhaite que notre famille reste toujours unie.

A tout le reste de ma grande famille en Algérie.

Table des matières

	Liste des abréviations	14
I.	Introduction	15
II.	Généralités et définitions	18
III.	État de l’art	20
IV.	Matériel et méthodes	
	III.1 Population de l’étude.....	23
	III.2 L’imagerie	23
	III.3 Les différents modèles testés	24
	III.4 Les données brutes	24
	III.5 Modèles d’apprentissage conventionnel	
	a. Prétraitement des images.....	25
	b. Extraction des données de radiomique	26
	c. Sélection des données	26
	d. Construction du modèle	27
	e. Analyse de la performance	28
	f. Comparaison des performances.....	28
	III.6 Modèle d’apprentissage profond	
	a. Généralités sur les réseaux de neurones artificiels	29
	b. Le choix du modèle	31
	c. L’architecture du réseau	31
	d. Les différentes couches de neurones	32
	e. Les fonctions d’activation	33
	f. Organisation des données	34
	g. Data augmentation	34
V.	Résultats	
	IV.1 Caractéristiques de la population	35
	IV.2 Performances des modèles d’apprentissage conventionnel.....	35
	IV.3 Comparaison des performances des modèles d’apprentissage conventionnel	40
	IV.4 Performance du modèle d’apprentissage profond	40
VI.	Discussion des résultats	42
VII.	Conclusion	45
VIII.	Bibliographie	46
IX.	Annexes	48
X.	L’Article	56

Liste des abréviations

IRM : Imagerie par Résonance Magnétique

TEP : tomographie par émission de positons

ACE : antigène carcino-embryonnaire

CNN : convolutional neural network

ROC : receiver operating curve

AUC : area under the curve

CI : confidence interval

AJCC : American joint committee on cancer

DICOM : digital imaging and communications in medicine

VGG : visual geometry group

ILSVRC : ImageNet Large Scale Visual Recognition Competition

RCT : radiochimiothérapie

Introduction

Devant les progrès de l'informatique et de l'imagerie médicale, les chercheurs ont commencé à explorer de nouvelles pistes pour exploiter au mieux les informations enfouies dans les images médicales. Ainsi, la radiomique a vu le jour comme un domaine de recherche à part entière et capté l'attention des chercheurs. Plus d'un millier d'articles a été publié jusqu'à ce jour. Leur nombre a même été multiplié par trois ces 4 dernières années, objectivant l'enthousiasme des chercheurs porté à ce sujet.

La radiomique est l'extraction d'une quantité massive de données à partir d'images médicales conventionnelles, comme les radiographies standards, l'échographie, le scanner, l'IRM ou encore le TEP-scan, en corrélation avec : 1) le diagnostic, 2) le stade de la maladie, 3) la réponse thérapeutique, 4) les données génomiques ou tout simplement 4) le pronostic [1]. Elle a émergé essentiellement à partir du champ de la cancérologie où fournir une information spécifique pour une thérapie personnalisée est primordiale. En effet, le même grade du même type histologique d'une tumeur peut se comporter différemment d'un patient à un autre, d'où l'importance d'une thérapie personnalisée [2]. Il est actuellement admis que les tumeurs solides ne consistent pas en une seule entité homogène, mais plutôt plusieurs clones ou sous-populations de cellules cancéreuses manifestant une certaine hétérogénéité spatiale, qui pourrait donner une information importante sur la tumeur elle-même [3].

Une relation entre les données radiomiques et les données génomiques, moléculaires et immunologiques a également été suggérée, faisant de la radiomique une approche intéressante non seulement en raison de son potentiel de fournir une information personnalisée mais également par son potentiel de se substituer à des procédures relativement coûteuses et invasives [4]. La caractérisation cellulaire et/ou moléculaire par des biopsies, outre son inconvénient d'être invasive et coûteuse, fournit une caractérisation limitée de la tumeur puisque l'échantillon fourni par la biopsie peut correspondre à seulement un clone d'une sous-population et non pas à la population entière. La radiomique peut pallier à ce problème en évaluant l'ensemble de la masse tumorale, dans n'importe quelle partie du corps humain et le nombre de fois nécessaire [4].

Les applications de la radiomique en cancérologie sont nombreuses. Le cancer rectal fait partie des types de cancer les plus étudiés par les chercheurs en radiomique. Il constitue le troisième cancer en termes de morbidité et de mortalité [5]. Bien que le développement des stratégies thérapeutiques et la prise en charge multidisciplinaire ont permis d'améliorer son pronostic, le taux de récurrence reste relativement élevé [6].

La chirurgie d'exérèse est le traitement de référence pour les récurrences précocement détectées. Elle a démontré un meilleur pronostic, offrant une chance de guérir les patients [7]. Mais, une bonne proportion des patients pourrait perdre cette chance de guérison si la récurrence n'est pas détectée à un stade assez précoce. Ainsi, le dépistage de patients à haut risque de récurrence pourrait offrir une opportunité de sélection des patients à risque, de les suivre de manière rapprochée, pour pouvoir détecter précocement une récurrence et la traiter convenablement. A

ce jour, il n'existe pas de marqueur biologique efficace pour nous offrir cette opportunité de dépistage de patients à haut risque de récurrence.

Des études antérieures avaient rapporté que certains facteurs clinico-biologiques et des éléments d'imagerie médicale conventionnels pourraient détenir une certaine valeur pronostique mais un consensus n'a pas été clairement établi [8]. Le rôle de la radiomique comme potentiel marqueur prédictif de récurrence a donc été évoqué. L'intelligence artificielle et l'apprentissage automatique (*machine learning*) pourraient aider à l'évaluation des données de radiomique. En particulier, l'apprentissage profond (ou *deep learning*) avec les réseaux de neurones convolutionnels (CNN pour l'acronyme anglais *convolutional neural networks*) pourrait exécuter des analyses massives de texture d'image avec un minimum d'entrée humaine.

Il existe toutefois des limitations et des challenges à surmonter avant que la radiomique puisse être implémentée en routine clinique.

Les variations des protocoles et des machines d'imagerie entre les différentes institutions et au sein même d'une seule institution pourrait réduire la robustesse des modèles prédictifs, construits à l'aide de l'apprentissage conventionnel. Il est à noter qu'à ce jour, aucun consensus n'a été établi concernant une extraction standardisée des données radiomiques [9]. La validité interne et externe des modèles d'apprentissage conventionnel pourrait se trouver compromise en particulier si les échantillons sont relativement petits. En effet, la plupart des études de radiomique, conduites à l'aide d'apprentissage conventionnel et publiées ont été conduites avec moins de 100 patients [9].

On peut voir dans le deep learning une alternative intéressante aux techniques de machine learning conventionnel exploitant le potentiel de la radiomique, dans le sens où il permet d'utiliser une faible quantité de données brutes, peu d'images ou de patients, sur lesquelles on pourrait par la suite appliquer un facteur d'augmentation afin d'accroître le nombre de données sans pour autant exposer le modèle au sur-apprentissage. On s'en passera par la même occasion, de la segmentation manuelle de la tumeur d'intérêt, de l'extraction manuelle des données radiomiques et du problème lié à la variation des protocoles de prétraitement d'images [10].

Notre étude a été conduite dans l'objectif d'évaluer et de comparer le potentiel prédictif d'algorithmes d'apprentissage conventionnel et profond, appliqués à des IRM de patients atteints de tumeurs rectales localement avancées, en corrélation avec la récurrence.

Pour tester notre hypothèse, on a été confronté à deux problèmes :

- Allions-nous utiliser des images 2D ou 3D ?
- Allions-nous utiliser que les pixels de la tumeur, ou allions nous inclure l'environnement péri-tumoral dans une bounding-box ?

Dans un premier temps, sur des modèles conventionnels nous avons testé la même base de données pour la prédiction de récurrence, sur des algorithmes 2D versus 3D, afin de montrer la non-infériorité voire la supériorité des modèles 2D. Dans un deuxième temps, nous avons testé la même base de données pour la prédiction de récurrence, sur des modèles conventionnels utilisant des masques de la segmentation manuelle de la tumeur versus des masques de bounding box pour extraire les données de radiomique.

NB : la segmentation renvoie à la notion où l'on prend uniquement la tumeur contourée comme

zone d'intérêt. Alors que l'idée de la bounding box prend en plus de la tumeur, l'environnement péri-tumoral immédiat. On suppose que cet environnement péri-tumoral immédiat détient une information avec une certaine capacité prédictive.

Et en dernier, nous avons construit le modèle de CNN en s'appuyant sur les résultats obtenus lors des tests sur les algorithmes de machine learning conventionnel.

Dans la suite de ce rapport, on va présenter quelques généralités sur la radiomique et les techniques de machine learning, suivies de l'état d'art. On détaillera la méthodologie employée pour la réalisation de cette étude. Ensuite, on exposera les résultats obtenus et on discutera les données clés.

Généralités et définitions

Le mot radiomique, comme tous les domaines de type -omique, suppose l'existence d'une large quantité de données. Faire appel à de l'intelligence artificielle, en particulier aux différentes techniques d'apprentissage automatique est une étape nécessaire pour une exploitation meilleures des données [11]. Classiquement, les chercheurs dans ce domaine de radiomique faisaient plutôt appel à des techniques de machine learning conventionnelles (le *random forest*, par exemple). Plus récemment, le deep learning, sous-domaine du machine learning a vu le jour. Ses applications sont en augmentation croissante et les résultats obtenus jusqu'à présent ont démontré leur grande efficacité [12].

Dans la suite de cette section, nous allons décrire globalement le processus du machine learning conventionnel appliqué à la radiomique ainsi que de façon schématique ce qu'est le deep learning, afin de souligner la différence entre les deux approches.

a. Radiomique et apprentissage conventionnel :

Le processus d'exploitation des données radiomiques par le machine learning conventionnel passe par plusieurs étapes, comme illustré dans le schéma de la figure 1 :

- A partir des images, on réalise une segmentation automatique, semi-automatique ou manuelle de la zone d'intérêt.
- On réalise un prétraitement des images, afin de standardiser les protocoles.
- Suivis d'une extraction des données radiomiques à l'aide de logiciels dédiés.
- Afin d'éviter le risque de sur-apprentissage du modèle, il convient d'éliminer les données radiomiques redondantes, non reproductibles et/ou sans caractère prédictif. En fin de processus de sélection, on garde moins de 30 données, constituant ce qu'on appelle la *signature radiomique*.
- On entraîne l'algorithme d'apprentissage.
- Enfin, on teste la performance du modèle, déjà entraîné, sur une cohorte de validation.

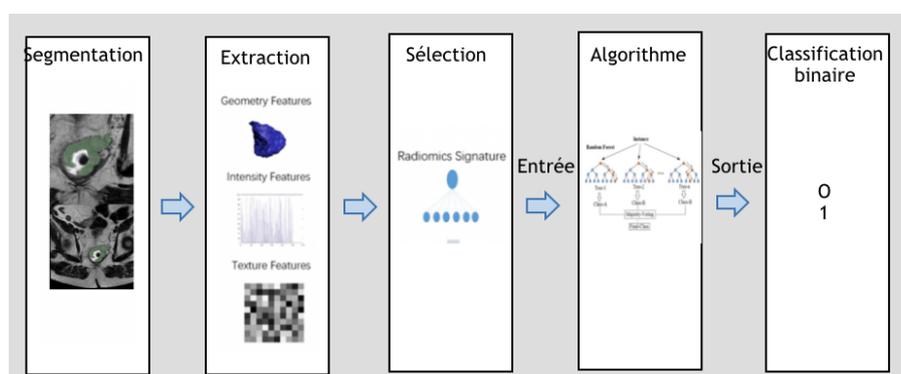


Fig 1 : la radiomique en apprentissage

b. Radiomique et apprentissage profond :

Le deep learning ou apprentissage profond implique des algorithmes inspirés de la structure et du fonctionnement du cerveau humain, appelés réseaux de neurones artificiels [12].

A l'ère actuelle, des ordinateurs très puissants sont capables d'entraîner des réseaux de neurones, avec des données de la dimension « -omique » et voir la performance des algorithmes augmenter au fur et à mesure que l'on rajoute des données supplémentaires. Comme illustré par la figure 2, ceci est différent lorsqu'on utilise les autres techniques d'apprentissage conventionnel, qui atteignent une performance en plateau au bout d'une certaine quantité de données, à cause de leur susceptibilité au risque de sur-apprentissage.

Le sur-apprentissage est défini par un modèle performant en phase d'apprentissage, mais avec une généralisation médiocre lors du test sur la cohorte de validation.

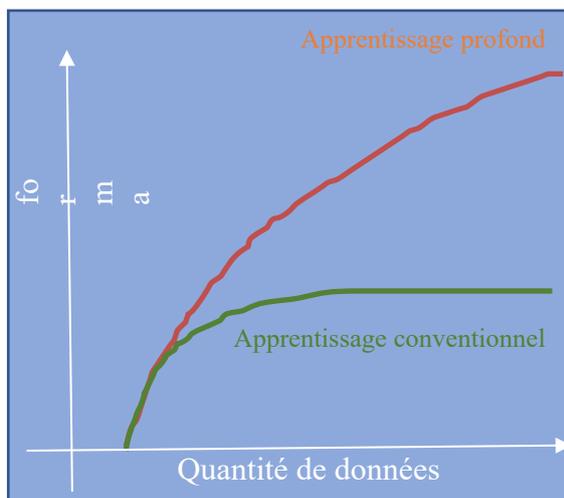


Fig 2 : Pourquoi l'apprentissage profond ?

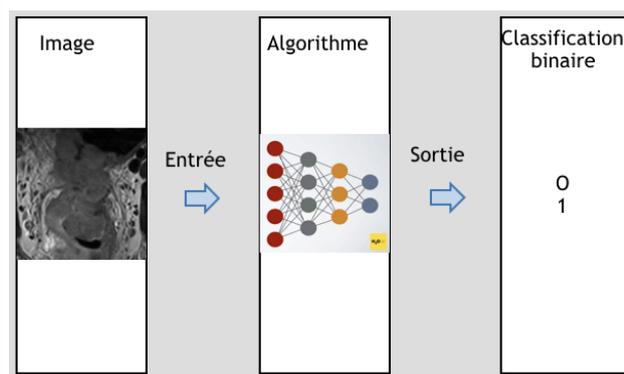


Fig 3 : la radiomique en apprentissage conventionnel

L'autre avantage du deep learning est sa capacité à extraire automatiquement, à partir des images brutes, des données de texture (aussi appelée en langage anglosaxon : feature learning). A l'inverse, les techniques d'apprentissage conventionnelles nécessitent la participation humaine à l'extraction des données radiomiques.

Pourquoi, parle-t-on de deep learning et non pas tout simplement de réseaux de neurones artificiels ? La notion de deep, renvoie à la profondeur et la complexité du réseau de neurones, en d'autres termes le nombre de couches.

Les réseaux de neurones artificiels, en fonction de leur complexité, peuvent faire l'objet d'une application d'apprentissage conventionnel, ou une application de type apprentissage profond. Par opposition, les autres techniques de machine learning conventionnel ne peuvent pas trouver d'applications en apprentissage profond.

Ainsi, l'apprentissage profond excelle dans des domaines où l'entrée n'est pas simplement une quantité modérée de données organisées sous la forme de tableaux mais plutôt des images de pixels, des documents de textes ou alors des fichiers audios [13].

La figure 3 ci-dessus, illustre la simplicité du modèle d'apprentissage profond et le peu « d'input humain » dans la construction du modèle.

État de l'art

Plusieurs études antérieures ont exploré des applications potentielles de la radiomique dans le cancer colorectal.

Ces applications potentielles peuvent être regroupées en plusieurs catégories :

- Évaluation de la reproductibilité des données de textures
- Prédiction de la réponse au traitement
- Prédiction de la survenue des métastases
- Prédiction de la survie.

Peu d'études ont toutefois exploré le potentiel de la radiomique dans la prédiction de la survie sans récurrence. En effet, dans une revue publiée par Horvat et al, on s'aperçoit que seulement 5 études, listées ici dans le tableau I, ont étudié cette question. 3 de ces études se sont servies de données IRM [14].

Author	n	Aim	IM	Segmentation	Readers (n)	FE	M	Main results	V
Van Helde et al. (2018)	99	Predict response and survival in patients with mCRC	¹⁸ F-FDG PET/CT	Semiautomatic VOI All tu lesions Pre-treatment	1	Morf FO TF	N	Some radiomics features correlates with survival	N
Liu et al. (2018)	177	Predict synchronous metastases in patients with rectal cancer	MRI (3T)	Manual VOI (T2WI) Rectal tu Pre-treatment	2	Morf FO TF SO TF	Y	ICC > 0.80 AUC clinical (T): 0.79 AUC clinical (V): 0.77 AUC clinical-radiomics (T): 0.85 AUC clinical-radiomics (V): 0.83	I
Meng et al. (2018)	108	Predict DFS	MRI (3T)	Manual VOI (DCE) Rectal tu Pre-treatment	2 Consensus	Morf FO TF SO TF SupO	Y	ICC > 0.80 C-index radiomics (T): 0.83 C-index clinical (T): 0.66 C-index both (T): 0.80 C-index radiomics (V): 0.77 C-index clinical (V): 0.64 C-index both (V): 0.79	I
Lovinfosse et al. (2017)	86	Predict survival in patients with LARC	¹⁸ F-FDG PET/CT	Semiautomatic VOI Rectal tu Pre-treatment	1	FO TF SO TF SupO	N	Coarseness was associated with DFS and DSS	N
Jalil et al. (2016)	56	Predict long term survival in patients treated with CRT	MRI (1.5T)	Manual ROI (T2WI) Rectal tu Pre- and post-CRT	1	FO TF	N	Several texture features were associated with long term survival, particularly mean of positive pixels pre-CRT and kurtosis post-CRT	N

All studies were single-center studies

Tableau I : caractéristiques des études évaluant la radiomique en corrélation à la survie et la survenue de métastases (tableau tiré de

Meng et al. Ont conduit une étude rétrospective, comparant la capacité de la signature radiomique à prédire la survie sans maladie, versus un modèle basé sur des données clinico-radiologiques, chez des patients atteints de cancer rectal localement avancé [15].

Ils ont inclus 109 patients, avec un cancer rectal localement avancé (répartis dans les groupes entraînement/validation selon un ratio de 1/1) pour lesquels il y avait une indication d'une

chimio radiothérapie néoadjuvante, suivie d'une chirurgie d'exérèse. Après segmentation manuelle des tumeurs, sur des IRM baseline, les données de radiomique ont été extraites puis filtrées pour constituer la signature radiomique. Cette signature radiomique a réussi à séparer les patients en deux groupes différents en termes de survie sans récurrence (hazard ratio = 6.83 ; $p < 0.001$).

Jalil et al. ont étudié l'analyse de texture de tumeurs rectales en IRM et sa capacité à prédire la survie à long terme, chez des patients traités par une radiochimiothérapie prolongée en adjuvant. Les données de textures, en utilisant des images IRM sur des séquences en pondération T2, avant et 6 mois après la radiochimiothérapie ont été extraites après segmentation manuelle des tumeurs. Plusieurs données de texture avaient un certain potentiel prédictif en termes de survie globale, survie sans maladie, et survie sans récurrence. Jalil et al. ont suggéré qu'un modèle d'apprentissage basé sur l'analyse de texture de tumeurs rectales pourrait prédire la survie avant la chirurgie et pourrait potentiellement sélectionner les patients en vue d'une thérapie personnalisée [16].

La segmentation manuelle sur une pile d'image (version 3D) est relativement chronophage, et le calcul sur des données 3D pourrait être relativement complexe, nécessitant des capacités computationnelles conséquentes. En littérature de traitement d'image et de radiomique, il y a eu qu'une seule étude comparant les performances des données radiomiques 3D versus données 2D. Shen et al. dans leur article « 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer » ont comparé des données radiomiques 2D et 3D en terme de performance prédictive sur des images scannographiques de cancers bronchopulmonaires non à petites cellules. Ils avaient inclus 588 patients à partir de 3 cohortes différentes. 2 cohortes avec un ensemble de 463 patients, de 2 institutions différentes, ont été utilisées pour l'apprentissage. Une cohorte de 125 patients a été utilisée comme cohorte de validation. Au total 1014 données radiomiques (207 données 2D et 507 données 3D) ont été évaluées. L'aire sous la courbe des courbes de ROC a été utilisée pour évaluer la performance des modèles. Les 2 modèles 2D et 3D avaient montré une certaine performance prédictive. Le modèle 2D présentait toutefois une performance meilleure. Shen et al. recommandent l'utilisation des données 2D [17].

Le processus de sélection des données et des attributs de classification reste débattu ce jour. La performance des différentes techniques dépend considérablement de la distribution des données et de la taille de l'échantillon.

Edward R Dougherty et al. avaient suggéré qu'un algorithme pourrait être meilleur pour une distribution de données particulière et donner paradoxalement de mauvais résultats sur un autre jeu de données [18]. Parmar et al. avaient conduit une étude afin d'évaluer la performance de différents algorithmes. Dans leur étude, 14 techniques de sélection des données et 14 autres techniques de classifications ont été examinées en terme de performance et de stabilité pour prédire la survie globale. Un total de 440 données radiomiques ont été extraites sur des scanners pré-thérapeutiques pour 464 patients atteints de cancer pulmonaire. Deux cohortes indépendantes ont été utilisées pour l'apprentissage ($n = 310$ patients) et pour la validation ($n = 154$ patients). Ils avaient démontré que le test de Wilcoxon et une méthode de classification de type random forest avaient la meilleure performance avec respectivement, une stabilité = 0.84 ± 0.05 , AUC = 0.65 ± 0.02 , une déviation standard relative = 3.52% , AUC = 0.66 ± 0.03 [19]. L'identification de la technique optimale de machine-learning pour les applications de la radiomique est une étape primordiale, pour avoir des modèles stables et cliniquement pertinents. Comme les techniques de machine learning conventionnelles sont sujettes à un manque de standardization des protocoles, le deep learning a été suggéré comme une potentielle

solution. La méthode la plus utilisée en apprentissage profond est le réseau de neurones convolutionnels (CNN) qui a commencé à devenir dominant dans les tâches de reconnaissance de visage et pleins d'autres domaines, y compris la médecine et en particulier la radiologie.

Les CNN sont conçus d'une manière à s'adapter automatiquement à l'apprentissage de la hiérarchie spatiale des données par rétropropagation de gradient et en utilisant de multiples blocs de couches de neurones. Son intérêt a été initialement démontré dans la rétinopathie diabétique, la classification des lésions cutanées et la détection des métastases ganglionnaires respectivement. Des rapports récents ont démontré la capacité des CNN à analyser l'information texturale présente dans les images radiologiques, grâce à des couches de neurones convolutionnels. Cette information texturale pourrait être visualisée en utilisant des techniques avancées pour interpréter le réseau et afin de mieux comprendre la différence entre les méthodes de radiomique (apprentissage conventionnel vs apprentissage profond). A titre d'exemple, la visualisation de la différence en données de texture pour des images avec des entropies basses et élevées pourrait nous fournir une information sur l'hétérogénéité du tissu tumoral [20]. Des algorithmes variés de « computer vision » ont été développés avec des performances satisfaisantes (Resnet, Densent et Inception) mais leurs architectures restent adaptées à des images RGB (couleurs Red, Green et blue). Leurs performances sur des images radiologiques, c'est-à-dire en échelles de gris n'a pas encore été formellement démontrée. La traduction de ces architectures dans le domaine médical pourrait ne pas offrir un résultat optimal, particulièrement pour l'imagerie multimodale. Il y a principalement deux éléments pouvant faire obstacle à l'application de l'apprentissage. La première est la qualité du pré-traitement d'image avant l'apprentissage du réseau de neurones. Plusieurs articles avaient démontré le rôle de la standardisation pour l'apprentissage et la validation. Le second élément est l'architecture du réseau de neurones lui-même.

Globalement, les applications du deep learning en radiologie pourraient concerner : des problèmes de classification [21], de segmentation [22], de détection [23] et autres. Les méthodes de deep learning ont montré des performances supérieures à celles des méthodes d'apprentissage conventionnel exploitant le potentiel de la radiomique, notamment dans la détection des lésions mammographiques et le recalage de l'imagerie multimodale [23]. Les applications de l'apprentissage profond dans la prédiction du pronostic à partir des images médicales (échelles de gris) sont mal connues et les études de ce sujet restent à ce jour peu nombreuses. Une équipe du département de radiologie de l'hôpital général de Boston a conduit une étude dans l'objectif de construire un modèle de CNN pour la prédiction de pronostic sur des patients atteints de cancers pulmonaires. Hosny A et al. avaient donc conduit cette étude rétrospective multientrique sur 1194 patients atteints de carcinome broncho-pulmonaire non à petites cellules. En utilisant des données TDM, ils avaient identifié des signatures pronostiques, en s'aidant de CNN 3D. Ils avaient trouvé que les prédictions des CNN étaient significativement associées à la survie globale à 2 ans, à partir du début de traitement de radiothérapie (AUC = 0,70 [95% CI 0.63-0.78], $p < 0.001$) et de la chirurgie (AUC = 0,71 [95% CI 0.60-0.82], $p < 0.001$). Le CNN était également capable de stratifier les patients en deux groupes différents, faible et haut risque de mortalité pour les deux catégories de patients, traités par radiothérapie ($p < 0.001$) et par chirurgie ($p = 0.03$). En outre, le CNN avait une performance meilleure qu'un random forest construit sur des paramètres cliniques, incluant l'âge, le sexe et l'invasion ganglionnaire. Afin d'avoir une meilleure compréhension des caractéristiques de texture de l'image capturées par le CNN, ils avaient identifié les régions contribuant à la prédiction et objectivant l'importance de l'environnement péri-tumoral dans la stratification des patients [24].

Matériels et méthodes

1. *Population de l'étude :*

Les patients ont été sélectionnés à partir d'une base de données regroupant 330 patients, ayant consulté entre le 01 janvier 2011 et le 30 décembre 2014 dans différents CHU (centre hospitalo-universitaire) et institut de cancérologie sur l'ensemble du territoire français et pour lesquels un diagnostic de cancer rectal avait été posé. Après revue de la base de données 130 patients étaient éligibles, dont seulement 98 remplissaient l'ensemble des critères d'inclusion.

Les critères d'inclusion dans l'étude étaient :

- Un adénocarcinome rectal prouvé histologiquement.
- Un stade localement avancé, pour lequel il y avait une indication d'un traitement néo-adjuvant par radio-chimiothérapie.
- Le stade localement avancé était défini par un statut T3 ou T4, selon la classification de l'AJCC actuellement en vigueur (TNM 8^{ème} édition 2017).
- La présence d'une première IRM de référence : IRM-baseline.
- Cette première IRM avait servi comme référence mais également pour la stadification de la tumeur.
- La présence d'une deuxième IRM, après traitement néo-adjuvant : IRM-post.
- Un traitement par chirurgie complémentaire, à la fin de la radio-chimiothérapie. Ce traitement chirurgical consistait en l'exérèse monobloc de la tumeur, à marges carcinologiques avec exérèse du méso-rectum adjacent.
- Un suivi ultérieur sur au moins 4 ans pour le recueil des données cliniques.

Les critères d'exclusion étaient :

- La présence d'un autre type histologique de cancer synchrone (2 patients).
- La présence de métastases synchrones à distance (5 patients).
- L'absence de chirurgie complémentaire, à la fin de la radio-chimiothérapie (25 patients).
- Des images IRM de mauvaise qualité, en raison d'artéfacts de mouvements principalement (10 patients).
- Ou lorsqu'une des deux imageries (IRM-baseline ou IRM-post) était manquante (3 patients).

L'étude a reçu une dérogation de la part du comité d'éthique de l'établissement (Institut de cancérologie de Montpellier) concernant le consentement des patients.

2. *L'imagerie :*

Au total 196 IRM ont été récupérées. 2 IRM par patient, avec une IRM-baseline et une IRM-post. Toutes les IRM ont été réalisées sur des machines 1.5 Tesla.

Les protocoles variaient en fonction des machines utilisées pour l'acquisition des images et en fonction des institutions. Cette variation a été prise en compte dans l'analyse des données. Parmi les différentes séquences IRM disponibles, la séquence T2 a été choisie pour l'analyse. Ceci, pour diverses raisons :

- Son caractère informatif. En effet, les radiologues dans leur travail quotidien se basent sur ce type de séquences pour la majeure partie de l'interprétation.
- Séquence ubiquitaire. Tous les protocoles d'IRM incluaient des séquences T2.
- Son intérêt particulier en ce qui concerne la radiomique en cancer rectal comme il a déjà été démontré par de nombreuses études antérieures.

Pour des raisons de simplicité et de difficultés de calcul, il a été décidé d'ignorer les autres types de séquences.

3. Les différents modèles testés :

- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 2D d'une image d'intérêt de l'IRM-baseline, après segmentation manuelle (model 1).
- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 2D d'une bounding box de l'IRM-baseline (model 2).
- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 3D d'une image d'intérêt de l'IRM-baseline, après segmentation manuelle (model 3).
- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 2D d'une image d'intérêt de l'IRM-post, après segmentation manuelle (model 4).
- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 2D d'une bounding box de l'IRM-post (model 5).
- Modèle d'apprentissage conventionnel, basé sur une analyse radiomique 3D d'une image d'intérêt de l'IRM-post, après segmentation manuelle (model 6).
- Modèle d'apprentissage profond, basé sur une analyse 2D des deux IRM (baseline + post-thérapeutique).

4. Les données brutes :

Pour chaque patient, un répertoire a été créé, pour organiser les données brutes. Ce répertoire contenait, 6 fichiers :

- Un fichier DICOM pour l'acquisition T2 (3D) dans le plan axial de l'IRM-baseline.
- Un fichier NRRD pour la segmentation 3D du volume tumoral sur l'IRM-baseline.
- Un fichier NRRD pour la segmentation 2D de la surface tumorale sur l'IRM-baseline. La segmentation a été réalisée en choisissant une image d'intérêt parmi la pile d'images du volume d'acquisition, idéalement celle passant par l'épicentre de la tumeur.
- Un fichier DICOM pour l'acquisition T2 (3D) dans le plan axial de l'IRM-post.
- Un fichier NRRD pour la segmentation 3D sur l'IRM-post.
- Un fichier NRRD pour la segmentation 2D sur l'IRM-post. La segmentation a été réalisée en choisissant une image d'intérêt parmi la pile d'images du volume d'acquisition, idéalement celle passant par l'épicentre de la tumeur.

La lecture des images DICOM et la réalisation de la segmentation ont été exécutées à l'aide d'un logiciel gratuit, d'usage commun : 3D Slicer, version 4.10.2

Sur l'imagerie Baseline, le volume ou la surface d'intérêt étaient définis par tout épaississement pariétal ou de syndrome de masse appendu à la paroi rectale, apparaissant de signal T2 intermédiaire, en restriction de diffusion et rehaussé après injection de gadolinium.

Sur l'imagerie post-thérapeutique, le volume ou la surface d'intérêt étaient définis par toute anomalie morphologique et/ou de signal en lieu et place de la tumeur traitée.

Le contourage des lésions était réalisé par un radiologue de façon manuelle, image par image pour la segmentation 3D et de façon manuelle sur une seule image pour la segmentation 2D. Là où existait un doute sur le caractère pathologique des pixels, ceux-ci n'étaient pas pris en compte dans la segmentation.

Un fichier Excel a été créé en plus et sur lequel étaient notés les épicesentres des tumeurs (pour l'IRM-baseline) et les épicesentres de remaniements post-thérapeutiques (pour l'IRM-post). Ainsi, après lecture de l'ensemble des IRM, les coordonnées X, Y et Z étaient recueillies sur ce fichier. Ces coordonnées vont servir pour la création des bounding box. Cette dernière est créée de manière automatisée pour l'ensemble des patients à la fois sur l'IRM-baseline et l'IRM-post.

La figure 4 illustre les 3 types d'images utilisées à partir d'une séquence T2 axiale de l'IRM-baseline du premier patient de la cohorte d'entraînement



Fig 4 : exemple illustrant la segmentation 2D et la bounding-box avec l'image native correspondante sur une IRM-baseline image native à gauche, segmentation 2D au milieu, bounding-box à droite

5. Modèles d'apprentissage conventionnel :

Pour chacun de ces modèles, un code a été implémenté sur python (version 2.7), dont le détail retrouvable en Annexe II et dont les objectifs sont les suivants :

- Automatiser la lecture des images brutes.
- Automatiser la création des bounding box.
- Automatiser le prétraitement des images.
- Automatiser l'extraction des données radiomiques.
- Réaliser la sélection des données informatives.
- Construire un modèle de random forest pour la classification binaire
- La sortie du modèle est exprimée sous la forme d'une variable binaire (0 pour l'absence de récurrence et 1 pour la récurrence)
- Entraîner le modèle.
- Tester la performance du modèle.

A. Prétraitement des images :

Une étape de *ré-échantillonnage*. En raison de la variation des protocoles et de l'inhomogénéité en taille des pixels entre les différents patients et les différentes imageries, cette étape était nécessaire.

Avant ré-échantillonnage, la taille des pixels était comprise entre 0,5 et 0,9 dans les dimensions X et Y, et entre 2,5 et 4 dans la dimension Z.

Le ré-échantillonnage a fait appel à une fonction disponible sur la librairie pyradiomics de python, avec comme sortie des images de pixels avec des dimensions de 1x1mm dans le plan xy et 4mm en profondeur z.

Une étape de *normalisation*, en faisant appel à la fonction 'normalize' de pyradiomics.

Pour rappel, la normalisation est un processus qui permet de changer la dynamique d'intensité des pixels, afin que les échantillons soient comparables. Dans notre cas, la dynamique des intensités des pixels a été fixée sur un intervalle de **0 à 255**.

En plus de l'image originale, plusieurs *filtres* ont été appliqués pour augmenter le nombre de données à extraire et exploiter au mieux l'information de l'image.

Au total 8 filtres ont été appliqués : Wavelet, LoG (Laplacian of Gaussian), Square, SquareRoot, Logarithm, Exponential, Gradient et LBP2D ou LBP3D.

- Le filtre Wavelet retourne plusieurs décompositions (toutes les combinaisons sont possibles en appliquant un filter pass-haut ou pass-bas, l'intérêt étant d'enlever du bruit).
- Le Laplacian Of Gaussian filter servira pour le rehaussement des contours, en repérant des zones de changement d'intensité.
- Le Square filter met au carré les intensités des pixels.
- Le SquareRoot prend la racine carrée des intensités des pixels.
- Le filtre Logarithm prend le logarithme des intensités des pixels.
- Le filtre Exponential comme son nom l'indique là encore, met en exponentiel les intensités des pixels.
- Le filtre gradient retourne la magnitude du gradient.
- Le filtre LBP2D retourne un pattern local binaire en 2D.
- Le filtre LBP3D retourne un pattern local binaire en 3D en utilisant des harmoniques sphériques. La dernière image retournée correspond à la map kurtosis.

B. Extraction des données de radiomique :

Les données ont été extraites, de façon automatisée à l'aide de l'algorithme implémenté sur python, à partir de l'image originale et les images construites en appliquant les 8 filtres précédemment mentionnés.

Le processus d'extraction a été réalisé pour chaque modèle, avec plus de 1000 données radiomiques récupérées/ patient.

L'ensemble a été organisé dans un dataframe sur python.

C. La sélection des données :

La question de sélection de données ou d'attributs pour la classification est un axe de recherche très actif dans le domaine du data mining.

Cette sélection permet d'identifier et d'éliminer les variables qui pénalisent les performances d'un modèle complexe dans la mesure où elles peuvent être bruitées, non informatives, redondantes ou non (voire peu) reproductibles. En outre la mise en évidence de variables pertinentes facilite considérablement l'interprétation et la compréhension des aspects radiologiques des tumeurs. Elle permet également d'améliorer la performance de prédiction de

l'algorithme de classification et de passer outre le fléau de la haute dimensionnalité des données (the curse of dimensionality).

Dans notre étude, le nombre de variables était largement supérieur vis-à-vis du nombre de patients ou d'observations (un facteur de 10-15), rendant nécessaire la « sélection ». Dans la littérature du machine learning, trois approches ont été décrites : l'approche filter, wrapper et embedded. Ces deux dernières, comme l'illustre la figure ci-dessous, et à l'opposé de la première, sélectionnent de façon implicite les variables lors du processus d'apprentissage. La première consiste à parcourir l'ensemble des données avant le processus d'apprentissage.

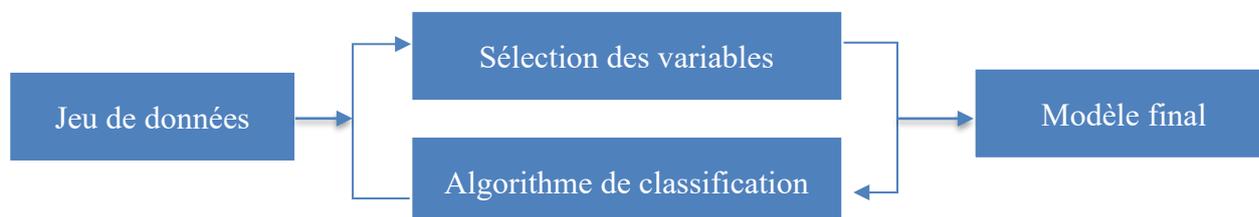


Fig 5 : Sélection des données par approche wrapper et embedded

Dans ce contexte, nous avons opté pour une technique combinatoire, utilisant à la fois algorithme de sélection (Recursive feature elimination ou RFE) avec algorithme de classification (random forest ou RF). Cette approche est d'implémentation relativement facile sur python et d'une efficacité déjà démontrée.

La RFE est une technique qui sélectionne les données prédictives par voie rétrograde. Elle commence par construire le modèle RF, en utilisant l'ensemble des données radiomiques disponibles sur le jeu dédié à l'entraînement. Elle calcule un facteur d'importance pour chaque donnée. Les données avec les facteurs d'importance les plus bas sont éliminées à chaque itération. Un paramètre permet de régler le nombre de variables éliminées sur chaque itération. Dans notre étude il a été fixé à 50. Un re-calcule des facteurs d'importance pour les données restantes est réalisé lors de l'itération suivante et ainsi de suite, jusqu'à l'obtention des données les plus prédictives.

Le RF est souvent utilisé avec la RFE, premièrement parce qu'il n'exclut pas les variables de l'équation de prédiction et deuxièmement parce que le RF a une méthode interne bien connue pour le calcul de l'importance des données. L'autre avantage avec cette technique est que le nombre optimal des données à sélectionner pour la construction du modèle prédictif est donné automatiquement en fin d'analyse.

D. La construction du modèle :

La construction du modèle prédictif a fait appel à l'algorithme des « forêts aléatoires » (ou RF pour Random Forest).

Le RF a été premièrement proposé par HO en 1995, puis formellement proposé en 2001 par Leo Breiman et Adèle Culter. C'est un algorithme combinant de nombreux arbres de décisions dans une approche de type bagging. Le bagging ou bootstrap est un groupe de méthodes d'inférence statistique basées sur la réplication multiple des données à partir du jeu de données étudié, selon les techniques de ré échantillonnage. Ainsi chaque arbre décisionnel reçoit une partie du jeu de données initiales. Un arbre de décision est une visualisation graphique sous forme d'arbre, d'une suite de décisions/possibilités. Chaque point est un nœud et chaque lien entre les nœuds est une branche.

Le point de départ se situe au sommet de l'arbre, et la décision/l'état final se trouve à l'autre extrémité : ceci étant atteint en suivant un chemin défini par les étapes intermédiaires qui sont à chaque nœud séparé en deux sous-groupes.

Le RF attribue une probabilité à chaque combinaison chemin/point de sortie. Le critère le plus connu de segmentation pour le problème de classification est l'indice d'impureté de Gini. Le concept de pureté fait référence au caractère discriminant de la séparation effectuée par un nœud. En clair, une séparation est dite pure quand chaque partie post-split contient des éléments d'une même classe. A l'inverse, le maximum d'impureté est atteint lorsque chaque séparation contient la même probabilité d'éléments de chaque classe.

La figure suivante illustre de façon très schématique un ensemble de quelques arbres décisionnels dans l'objectif d'une classification binaire. Dans notre étude le nombre d'arbre a été fixé à 1000.

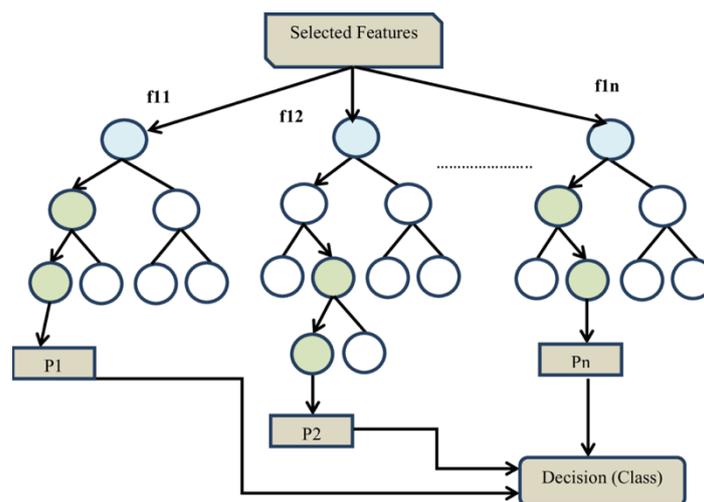


Fig 6 : Schéma tiré de article-Random forest for big data classification in the internet of things using optimal features by Lakshmanrabu et al

E. L'analyse de la performance :

Pour chaque modèle entraîné, on a réalisé une itération afin d'acquérir la matrice de confusion et calculer un facteur de précision sur une cohorte de test représentant 50% du jeu de données initial. Ce facteur de précision est égal à la somme des vrais positifs et vrais négatifs divisée par le nombre total des observations. Dans un second temps, on réalise une k-fold cross-validation sur l'ensemble du jeu de données. La k-fold cross-validation procède en divisant le jeu de données en k échantillons. Elle sélectionne l'un des k échantillons comme ensemble de validation avec les k-1 échantillons comme ensemble d'entraînement (apprentissage). On schématise la courbe de ROC (Receiver Operating Characteristic) résultante et on calcule l'aire sous la courbe (AUC) correspondante. Cette opération est répétée k fois. Dans notre étude, le facteur k est fixé à 5, comme dans les études de la littérature. Ainsi pour chaque modèle il existe 5 courbes de ROC avec les AUC correspondantes.

F. La comparaison des performances :

Elle a été réalisée en comparant les 5 valeurs d'AUC résultantes pour chacun des 6 modèles. Compte tenu de la distribution non normale et la comparaison de différents groupes de valeurs (>2), un test non paramétrique type Kruskal Wallis était nécessaire. Le seuil du risque alpha pour conclure à une différence était fixé à 0.05.

6. *Modèle d'apprentissage profond :*

Voir en Annexe III le code de programmation du modèle.

A. Quelques généralités sur les réseaux de neurones artificiels :

Les réseaux de neurones artificiels (ou ANN pour l'acronyme anglais, artificial neural networks) ont été comme leur nom indique, inspirés par le fonctionnement du cerveau humain. Ils consistent en un nombre défini de neurones. La version la plus simplifiée du modèle avec un neurone est appelé « Perceptron », proposé par Frank Rosenblatt en 1957. Le neurone comme on le verra dans le chapitre des fonctions d'activation est une fonction mathématique qui prend une ou plusieurs entrées. Pour chaque entrée le neurone assigne un nombre : un facteur de pondération. La sortie du neurone sera la somme des entrées pondérées. Ce sont ces facteurs de pondération qui rendent chaque neurone unique. Ils sont fixes lors de la phase de validation mais durant la phase d'apprentissage ils vont changer au fur et à mesure qu'on essaie d'améliorer la performance du modèle.

Le réseau de neurones est également une fonction mathématique, définie par les neurones le constituant. Ainsi chaque sortie de neurone va constituer une entrée pour le neurone de la couche suivante.

La fonction de perte : c'est une fonction qui nous renseigne sur la performance du modèle pour une tâche donnée. Le moyen le plus intuitif pour l'avoir, est de prendre chaque exemple d'apprentissage, de le basculer sur le modèle de réseau de neurone et récupérer la sortie à la fin du réseau. On soustrait cette valeur de sortie de la valeur réelle pour l'exemple qu'on a basculé sur le réseau et on la met au carré, comme illustré sur la formule suivante :

$$L(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

Sur cette formule : y désigne le label où la valeur réelle correspondant à l'exemple.

\hat{y} avec le chapeau, correspond au label où la valeur donnée par le réseau. i représente l'index de l'exemple.

L'apprentissage : lorsqu'on initialise l'entraînement du réseau, les facteurs de pondération sont des nombres randoms, la performance étant mauvaise au début. Au fur et à mesure, les facteurs de pondérations sont modifiés et la performance s'améliore. La fonction de perte suit la direction inverse, c-à-d commence avec des valeurs hautes, et se termine avec des valeurs basses en fin d'apprentissage. L'intérêt de l'entraînement est de minimiser cette fonction de perte. Il existe plusieurs algorithmes pour optimiser la fonction de perte. Ces algorithmes peuvent être basés sur le gradient ou pas, dans le sens où ils ne servent pas uniquement de l'information fournie par la fonction mais également par le gradient.

L'algorithme se servant du gradient le plus simple et le plus utilisé, est le « stochastic gradient descent » ou SGD, celui qu'on utilise pour la construction de notre modèle.

Comment fonctionne le SGD ?

La dérivative de notre fonction de perte nous renseigne sur la vitesse de changement de la fonction lorsqu'on prend des petites étapes dans la direction positive. Mathématiquement, elle suit la formule suivante :

$$\nabla L \approx \frac{\partial L}{\partial x_i} \nabla x_i$$

Ce qui veut dire que : le changement de notre fonction équivaut approximativement à la dérivée de la fonction pour une variable donnée x , multipliée par le changement de la variable x . Cette approximation sera exacte lorsqu'on prend des étapes tellement petites.

Le gradient est un vecteur de dérivées partielles, dont les éléments contiennent des dérivées de certaines variables dont la fonction dépend.

Voici un exemple simple, afin de mieux comprendre le processus d'optimisation de la fonction de perte. Considérant une fonction $f(x) = x^2$. Le minimum de cette fonction serait au point $x = 0$; mais comment l'algorithme va reconnaître ce point ? Supposons qu'on commence par une valeur donnée de $x = 2$. La dérivée de la fonction serait égale à 4. Donc on fait un pas vers une direction positive et le changement de notre fonction sera proportionnelle à 4. Donc elle va augmenter, alors qu'on cherche à la minimiser. L'objectif est de faire un pas vers la direction négative. L'importance du pas est difficile à déterminer, voir impossible. Elle est contrôlée par une sorte d'hyper-paramètre appelé : learning rate. Maintenant essayons de partir de $x = -2$. La dérivée sera égale à -4, donc elle décroît et c'est ce qu'on cherche à faire. Quand $x > 0$, la dérivée est supérieure à 0, on doit aller en direction négative, et inversement. On doit toujours aller dans la direction inverse de la dérivée. Le même principe s'applique au gradient qui est un vecteur pointant vers une direction donnée de l'espace. On prendra ainsi un pas dans la direction opposée au gradient.

En réseaux de neurones on va supposer que les entrées (x) et sorties (y) sont fixes. Les dérivées vont ainsi dépendre des facteurs de pondération (w) puisque ce sont les valeurs qu'on cherche à modifier pour améliorer le modèle. Si on calcule le gradient de la fonction de perte et qu'on fait des pas dans la direction opposée du gradient, la fonction de perte diminue jusqu'à convergence vers un point minimum.

Pour chaque itération, on met à jour les facteurs de pondération selon la formule suivante :

$$w_j = w_j - lr \frac{\partial L}{\partial w_j}$$

Le learning rate (lr) contrôle l'importance du pas à chaque itération. C'est l'hyper-paramètre le plus important. Si sa valeur est importante, il va sauter à travers le minimum. S'il est très faible, l'algorithme risque de ne pas converger vers le minimum ou alors prendra un certain temps pour y parvenir.

Dans notre cas, le **learning rate est fixé à 0.01**

L'autre hyperparamètre important est le batch size. C'est-à-dire le nombre d'exemple/observations incluses dans le calcul du gradient. Dans notre cas il a été fixé au nombre total des observations d'apprentissage/2.

Le dernier élément à savoir sur le fonctionnement des réseaux de neurones type CNN est la backpropagation (ou rétro-propagation en langage francophone). C'est le re calcul des facteurs de pondération en se basant sur le taux d'erreur (fonction de perte) de l'itération précédente, dans le but de minimiser ce taux d'erreur sur l'itération suivante.

B. Le choix du modèle :

En réseaux de neurones, les réseaux de neurones convolutionnels (ou CNN pour l'acronyme anglais Convolutional Neural Network) est l'une des catégories de réseaux de neurones artificielles les plus utilisées pour la reconnaissance des images et leurs classifications. Typiquement, la classification de l'image par CNN prend une image d'une dimension donnée et la voit plutôt comme une matrice de pixels. Chaque image d'entrée passe à travers des séries de couches de convolutions avec des filtres (ou kernels), des couches de pooling et des couches denses (fully connected layers) et applique à la fin une fonction (Softmax ou Sigmoid) pour classer l'image avec une probabilité entre 0 et 1.

En ligne, il existe des réseaux neurones déjà pré-entraînés et performants, avec un risque d'erreur de classification de moins de 7%. Parmi ces réseaux disponibles sur le net, on a opté pour le modèle nommé : VGG.

Le VGG a été inventé par Visual Geometry Group, de l'université d'Oxford, pour le présenter en compétition de l'ILSVRC (ImageNet Large Scale Visual Recognition Competition) en 2014. Le VGGnet a battu celui de google (GoogLeNet) et a gagné la compétition ILSVRC pour la tâche de localisation. Cette année était celle où les modèles de deep learning commençaient à avoir un taux d'erreur de moins de 10%. Plusieurs types de modèles VGG étaient disponibles. On a choisi le modèle VGG-16, c'est-à-dire un modèle à 16 couches, premièrement pour ses performances déjà démontrées sur d'autres jeux de données, deuxièmement parce qu'il était le plus adapté au jeu de données qu'on avait, pas trop compliqué avec un risque d'overfitting (comme pour le VGG-19) et pas trop obsolète (comme pour le VGG-11).

Similaire à d'autres réseaux (comme AlexNet), il utilise des convolutions de 3x3, mais avec beaucoup de filtres. Il est le modèle préféré en communauté pour extraire des données à partir d'une image. Bien évidemment la transposition de ce modèle déjà pré-entraîné en ligne directement vers les images médicales n'est pas réalisable. Ses facteurs de pondération ont été mis à jour avant de lancer l'apprentissage sur notre jeu de données.

C. L'architecture du réseau :

La première couche devait recevoir une image d'entrée avec des dimensions de 224x224 en format RGB.

Comme on travaillait sur des échelles de gris et non pas des échelles de couleurs, une fonction sur python nous a permis de répéter l'image 3 fois sur chaque input pour pallier à ce problème du format RGB.

On a ré-échantillonné l'ensemble des images, à partir de matrices de 512x512 et de 256 x 256 vers des matrices de 224 x 224.

L'image passe ensuite à travers d'une série de couches convolutionnelles, où des filtres sont utilisés avec des champs relativement étroits : 3x3 (c'est-à-dire le champ le plus petit pour capturer la notion du gauche/droit, haut/bas, centre).

Le pas de la convolution est fixé à un pixel, ainsi le remplissage spatial par convolution est tel que la résolution spatiale est conservée après convolution.

Le spatial pooling est réalisé par 5 couches de max-pooling, qui font suite à quelques couches convolutionnelles. Le max-pooling est réalisé sur des fenêtres de 2x2 avec des pas de 2. Les 3 couches fully-connected font suite à la série des couches convolutionnelles. Les deux premières ont 4096 canaux chacune. La dernière contient deux neurones avec une fonction sigmoïde, pour une classification binaire, soit 0 (absence de récurrence) soit 1 (récurrence). Toutes les couches cachées sont des équipées de la fonction ReLU.

La figure ci-dessous schématise l'architecture globale du réseau.

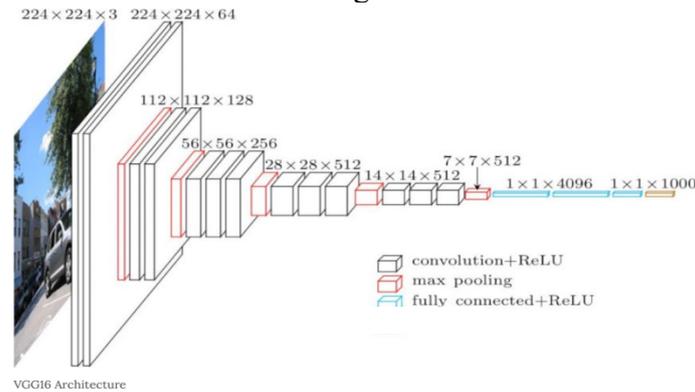


Fig 7 : Architecture du réseau, image tirée de neurohive.com

D. Les différentes couches de neurones :

La couche de convolution : elle constitue la première couche pour extraire les données à partir de l'image. C'est une opération mathématique qui prend l'image d'entrée et un filtre (ou kernel).

La couche de padding : rajoute des pixels à l'image lorsque le filtre couvre en excès ou ne couvre pas une partie de l'image.

La couche pooling : cette couche sert à réduire le nombre de paramètres quand l'image est trop large. Le spatial pooling, aussi appelé sous-échantillonnage réduit la dimensionnalité de chaque image mais retient les informations les plus importantes. Il existe plusieurs types de couches pooling : Max Pooling, Average Pooling et Sum Pooling. Max Pooling est la plus utilisée et a été choisie pour la construction de notre modèle. Elle prend les pixels d'intensité la plus élevée. Sa taille a été fixée à 2×2 dans notre étude, avec un pas de 2.

La couche Fully Connected :

La dernière couche Max Pooling retourne une matrice qui sera transformée par la suite en un vecteur (x_1, x_2, x_3, \dots), puis balancé à son tour à l'ensemble des couches denses (Fully connected).

- An image matrix (volume) of dimension $(h \times w \times d)$
- A filter $(f_h \times f_w \times d)$
- Outputs a volume dimension $(h - f_h + 1) \times (w - f_w + 1) \times 1$

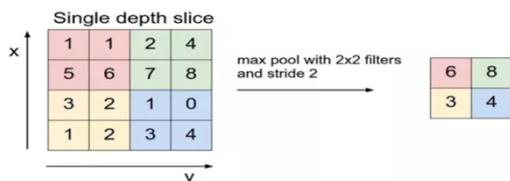
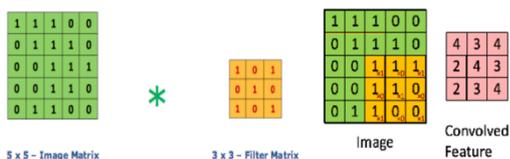
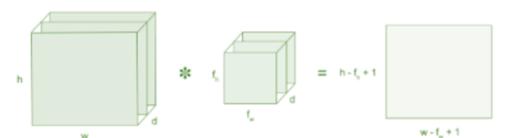
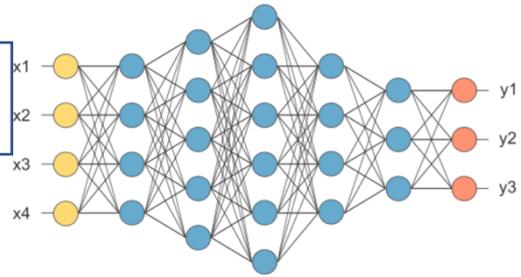


Fig 8 : Schéma expliquant le processus de convolution (tiré de medium.com)

Fig 9 : Schéma expliquant le processus de Max Pooling (tiré de medium.com)

Fig 10 : Schéma illustrant la configuration des couches fully connected (tiré de medium.com)



E. Les fonctions d'activation :

La fonction d'activation a été inspirée du potentiel d'action des neurones biologiques. En milieu biologique, le neurone reçoit des signaux provenant d'autres neurones à travers les dendrites. Le poids associé à une dendrite donnée, appelé poids synaptique, est multiplié par le signal entrant. Ainsi les signaux sont accumulés dans le corps cellulaire et si la force du signal résultant dépasse un certain seuil, le neurone transmet le message à l'axone.

La fonction d'activation prend la décision de transmettre ou non le signal. Dans ce cas, il s'agit d'une fonction simple avec un seul paramètre (c'est-à-dire le seuil). Il est à noter que lorsque l'être humain, apprend quelque chose de nouveau, le seuil et la probabilité de connexion (appelée poids synaptique) de certains neurones changent. Cela crée de nouvelles connexions entre les neurones, ce qui permet au cerveau d'apprendre de nouvelles choses.

Pour comprendre ce qu'il se passe avec les réseaux de neurones artificiels, il faut imaginer les valeurs entrantes dans un neurones ($x_1, x_2, x_3, \dots, x_n$) sont multipliées avec leur poids ou facteurs de pondération (référence au poids synaptique) qui leur sont associés ($w_1, w_2, w_3, \dots, w_n$). On fait ensuite la somme de ces multiplications et on ajoute enfin le biais (référence au seuil), selon la formule suivante :

$$Z(x) = \sum_i^N (w_i x_i) + b$$

La fonction $Z(x)$ correspond à la pré-activation, c'est à dire l'étape qui précède l'activation. Ensuite, le résultat z de la fonction de pré-activation $Z(x)$ est interprété par une fonction d'activation produisant en sortie un résultat y . Pour le choix des fonctions d'activations des neurones constituant notre réseau, on a fait appel à deux fonctions :

- La fonction ReLU : elle constitue la fonction la plus utilisée en réseaux de neurones artificiels, interprétée par la formule suivante : $f(x) = \max(0, x)$. Si l'entrée est négative la sortie est de 0 et si elle est positive alors la sortie sera de x .
- La fonction sigmoïde : Son objectif premier est de réduire la valeur d'entrée pour la situer entre 0 et 1. Elle exprime la valeur sous forme de probabilité, si la valeur en entrée est un très grand nombre positif, la fonction convertira cette valeur en une probabilité égale à 1, et inversement.

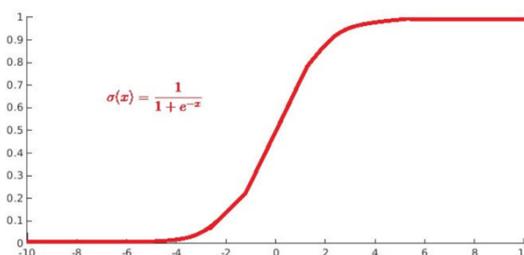
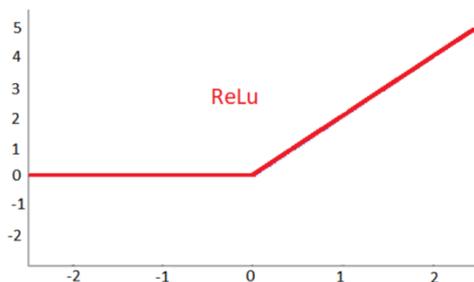


Fig 11 : Schéma expliquant la fonction d'activation ReLU (tiré du site de Supinfo, International University)

Fig 12 : Schéma expliquant la fonction d'activation Sigmoid (tiré du site de Supinfo, International University)

F. L'organisation des données :

Pour construire un réseau de neurones performant et efficace, une large base de données était nécessaire. Pour cela, on a regroupé l'ensemble des images en une seule base de données, à savoir les 98 images de l'IRM-baseline x 2 (brutes et bounding box) et les 98 images de l'IRM-post x 2 (brutes et bounding box). Au total, on avait 392 images. Les patients étaient randomisés en cohortes d'entraînement et de validation, selon un ratio de 0.8.

Les images ont été ré-échantillonnées pour avoir des dimensions de 224 x 224 pour éviter de changer l'architecture de base du réseau de neurones.

G. La data augmentation :

Pour profiter pleinement du potentiel du réseau de neurones, on a rajouté un facteur d'augmentation du data. La librairie keras de Python contenait des fonctions capables de prendre les images initiales et exécuter quelques modifications afin de construire des images avec un autre type d'information.

Cette augmentation des données, a fait appel à :

- une randomisation pour chacune des données (celle-ci étant obtenue en divisant la valeur par la moyenne de l'ensemble des valeurs)
- des rotations random dans un intervalle de 20°
- des déplacements dans les axes x et y sur des longueurs correspondant à 0.2 de la dimension matricielle
- et en dernier à des flips horizontaux.

Résultats

1. Description de la population de l'étude

Le tableau II en Annexe I résume les caractéristiques de la population étudiée, comportant au total 98 patients. Compte tenu du déséquilibre entre le nombre de patients ayant récidivé (24/98) et le nombre de patients sans récidive (64/98), une technique d'échantillonnage augmentant la dimension de l'échantillon minoritaire était nécessaire afin d'avoir le même nombre d'observation par échantillon. Pour cela, nous avons appliqué SMOTE (pour l'acronyme anglais Synthetic minority oversampling technique).

2. Performance des modèles d'apprentissage conventionnel

Le récapitulatif des résultats de performance pour chaque modèle est représenté en tableau III.

Modèles	Nombre de données radiomiques extraites	Nombre de données sélectionnées et incluses dans la signature radiomique	Précision du modèle sur la matrice de confusion	AUC moyenne sur les 5 itérations	p-value
1	1045	28	0.91	0.83 +/- 0.08	0.001
2	1046	22	0.68	0.8 +/- 0.12	0.001
3	1632	8	0.65	0.76 +/- 0.09	0.001
4	1045	21	0.73	0.8 +/- 0.08	0.001
5	1046	22	0.81	0.83 +/- 0.04	0.001
6	1632	8	0.74	0.83 +/- 0.08	0.001

Tableau III : récapitulatif résultats des performances modèles d'apprentissage conventionnel

Modèle 1

1045 données radiomiques ont été extraites pour chaque patient.

Après application de l'algorithme de sélection, 28 données ont été gardées comme signature radiomique.

```
('Optimal number of features :', 28)
('Best features :', Index([ 33, 40, 45, 52, 67, 71, 76, 86, 98, 106, 114, 120,
125, 149, 151, 186, 205, 254, 279, 355, 935, 959, 963, 964,
979, 991, 1026, 1041],
dtype='object'))
```

Fig13 (Capture d'écran affichage python) : Résultats de la sélection des données radiomiques pour le modèle 1

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion.

On notera que sur cette matrice de confusion la performance était supérieur à 0.9

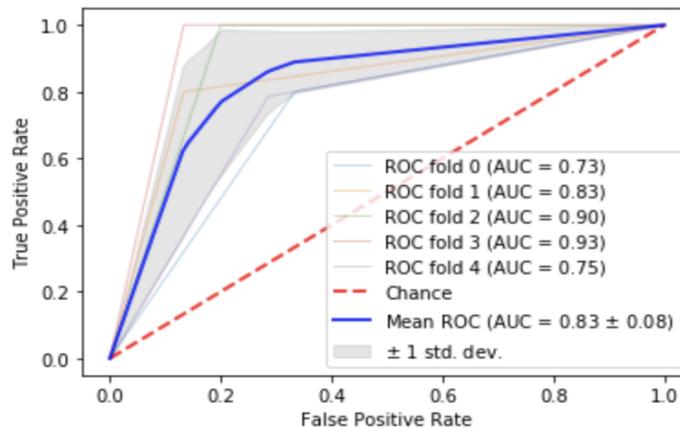
- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC (receiver operating characteristic).

On notera en particulier une aire sous la courbe moyenne estimée à 0.83 +/- 0.08

```
=== Confusion Matrix ===
[[ 36  6]
 [  0 32]]
```

Fig14 (Capture d'écran affichage python) : matrice de confusion pour le modèle 1

Sur cette matrice de confusion, sur la première colonne on retrouve le nombre de patients sans récurrence, sur la deuxième colonne le nombre de patients avec récurrence, sur la première ligne le nombre de patients étiquetés par le modèle comme étant à faible risque de récurrence, et sur la deuxième ligne le nombre de patients étiquetés par le modèle comme étant à haut risque de récurrence.



```
1 from scipy.stats import mannwhitneyu
2 Model = 0.73, 0.83, 0.90, 0.93, 0.75]
3 chance = 0.5, 0.5, 0.5, 0.5, 0.5, 0.5
4 stat, p = mannwhitneyu(model, chance)
5 print('p=%0.3f' % (p))
```

p=0.001

Fig 15 (Capture d'écran affichage python) : courbes de ROC pour le modèle 1

Modèle 2

1046 données radiomiques ont été extraites pour chaque patient.

Après application de l'algorithme de sélection, 22 données ont été gardées comme signature radiomique.

```
('Optimal number of features :', 22)
('Best features :', Index([ 31, 35, 36, 41, 45, 47, 49, 56, 61, 76, 80, 83, 102, 110,
112, 114, 135, 212, 253, 776, 984, 991],
dtype='object'))
```

Fig 16 : Résultats de la sélection des données radiomiques pour le modèle 2

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion.

On notera que sur cette matrice de confusion la performance était supérieur à 0.65

- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC. On notera en particulier une aire sous la courbe moyenne estimée à 0.80 ± 0.12

```

=== Confusion Matrix ===
[[27 12]
 [14 21]]

```

Fig 17 :
matrice de confusion pour le modèle 2

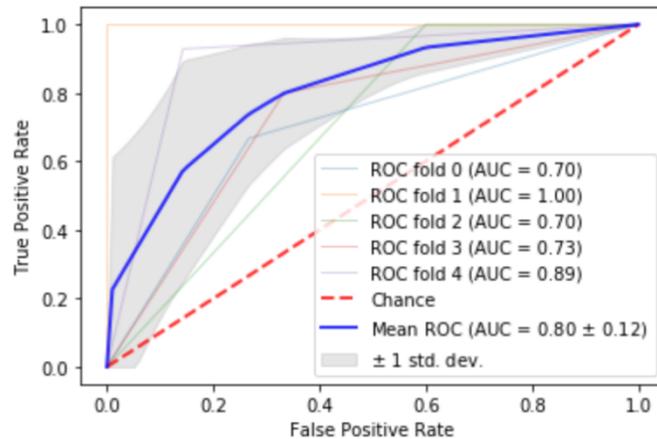


Fig 18 :
courbes de ROC pour le modèle 2

Modèle 3

1632 données radiomiques ont été extraites pour chaque patient.

Après application de l'algorithme de sélection, 8 données ont été gardées comme signature radiomique.

```

Optimal number of features : 8
Best features : Int64Index([42, 53, 112, 174, 195, 1490, 1541, 1596], dtype='int64')

```

Fig 19 : Résultats de la sélection des données radiomiques pour le modèle 3

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion. On notera que sur cette matrice de confusion la performance était supérieur à 0.65
- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC. On notera une aire sous la courbe moyenne estimée à 0.76 ± 0.09

```

=== Confusion Matrix ===
[[24  9]
 [17 24]]

```

Fig 20 :
matrice de confusion pour le modèle 3

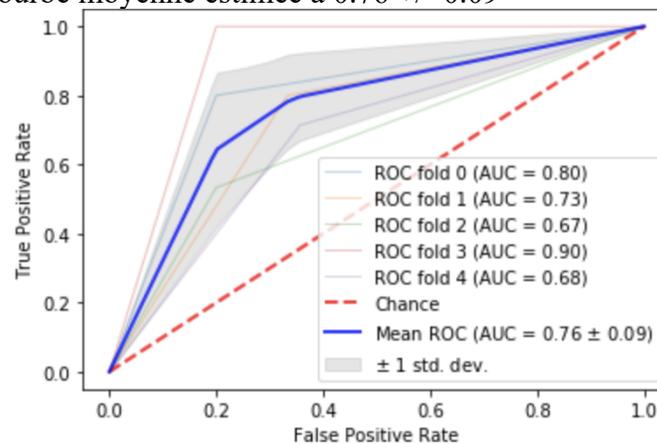


Fig 21 :
courbes de ROC pour le modèle 3

Modèle 4

1045 données radiomiques ont été extraites pour chaque patient.
Après application de l'algorithme de sélection, 21 données ont été gardées comme signature radiomique.

```
('Optimal number of features :', 21)
('Best features :', Int64Index([ 41, 44, 70, 71, 75, 80, 84, 85, 90, 104, 109,
113, 194, 873, 912, 926, 941, 947, 966, 967, 1024],
dtype='int64'))
```

Fig 22 : Résultats de la sélection des données radiomiques pour le modèle 4

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion.

On notera que sur cette matrice de confusion la performance était supérieur à 0.73

- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC.

On notera en particulier une aire sous la courbe moyenne estimée à 0.80 +/- 0.08

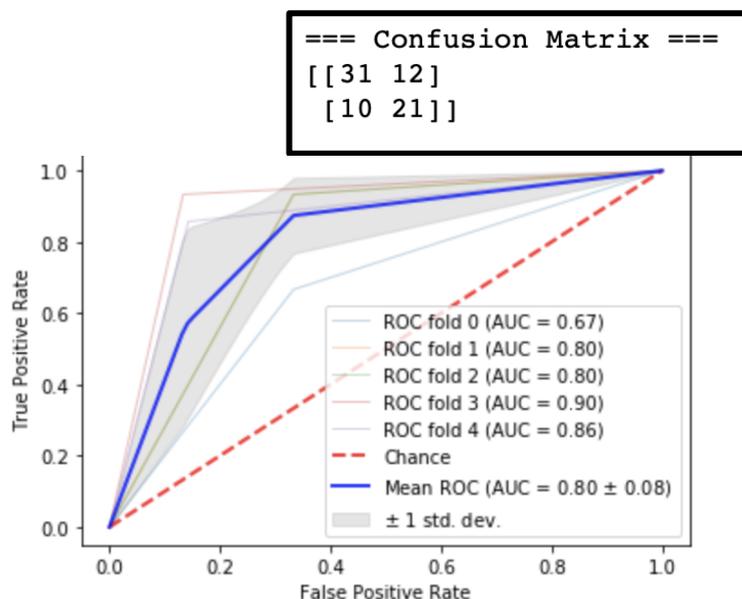


Fig 23 :
matrice de confusion pour le modèle 4

Fig 24 :
courbes de ROC pour le modèle 4

Modèle 5

1046 données radiomiques ont été extraites pour chaque patient.
Après application de l'algorithme de sélection, 22 données ont été gardées comme signature radiomique.

```
('Optimal number of features :', 22)
('Best features :', Index([ 31, 36, 41, 60, 61, 65, 74, 76, 96, 128, 133, 140,
200, 213, 230, 252, 328, 950, 964, 1012, 1020, 1032],
dtype='object'))
```

Fig 25 : Résultats de la sélection des données radiomiques pour le modèle 5

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion.

On notera que sur cette matrice de confusion la performance était supérieur à 0.81

- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC (receiver operating characteristics).

On notera en particulier une aire sous la courbe moyenne estimée à 0.83 +/- 0.04

```
=== Confusion Matrix ===
[[27 11]
 [ 3 33]]
```

Fig 26 :
matrice de confusion pour le modèle 5

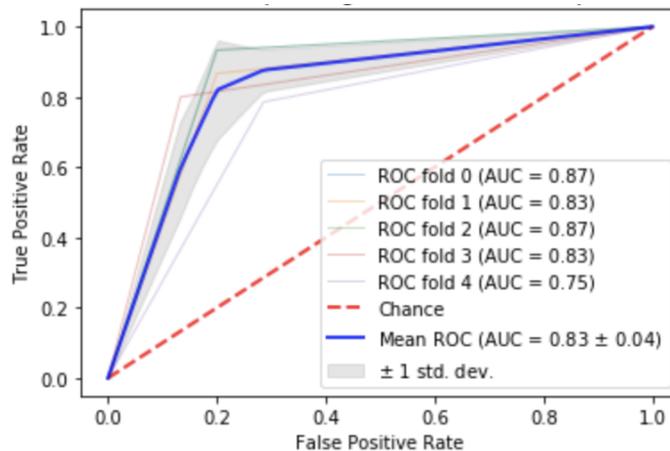


Fig 27 :
courbes de ROC pour le modèle 5

Modèle 6

1632 données radiomiques ont été extraites pour chaque patient.

Après application de l'algorithme de sélection, 8 données ont été gardées comme signature radiomique.

```
Optimal number of features : 8
Best features : Int64Index([365, 405, 462, 587, 658, 689, 720, 728], dtype='int64')
```

Fig 28 : Résultats de la sélection des données radiomiques pour le modèle 6

Le résultat du test de la performance du modèle sur la cohorte de validation est illustré :

- Sur une itération sous la forme d'une matrice de confusion.

On notera que sur cette matrice de confusion la performance était supérieur à 0.74

- Sur l'ensemble des itérations de la cross-validation sous la forme de courbes de ROC (receiver operating characteristics).

On notera en particulier une aire sous la courbe moyenne estimée à 0.83 +/- 0.08

```

=== Confusion Matrix ===
[[24 12]
 [ 7 31]]

```

Fig 29 :
matrice de confusion pour le modèle 6

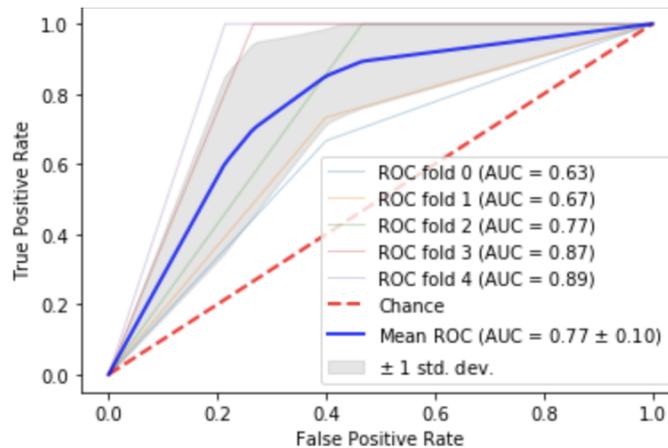


Fig 30 :
courbes de ROC pour le modèle 6

3. Comparaison des performances des modèles d'apprentissage conventionnel :

La performance en terme d'AUC a été comparée entre les 6 modèles, en utilisant le Kruskal-Wallis test. Le résultat est illustré en capture d'écran ci-dessous.

La figure en effet montre qu'il n'y a aucune différence entre les performances des 6 modèles.

```

1 from scipy.stats import kruskal
2 X6 = [0.63, 0.67, 0.77, 0.87, 0.89]
3 X5 = [0.75, 0.83, 0.87, 0.83, 0.87]
4 X4 = [0.86, 0.9, 0.8, 0.8, 0.67]
5 X3 = [0.68, 0.9, 0.67, 0.73, 0.8]
6 X2 = [0.89, 0.73, 0.7, 1, 0.7]
7 X1 = [0.75, 0.93, 0.9, 0.83, 0.73]
8 stat, p = kruskal(X1, X2, X3, X4, X5, X6)
9 print('Statistics=%.3f, p=%.3f' % (stat, p))
10 # interpret
11 alpha = 0.05
12 if p > alpha:
13     print('Same distributions (fail to reject H0)')
14 else:
15     print('Different distributions (reject H0)')

```

Statistics=2.675, p=0.750
Same distributions (fail to reject H0)

Fig 31 : comparaison de performances des 6 modèles d'apprentissage conventionnel

4. Performance du modèle d'apprentissage profond

L'ensemble des données entraînables a été calculé à 134268738 (voir fig 32).

Discussion

1. Apprentissage conventionnel

Performance de chaque modèle :

Les résultats des modèles d'apprentissage conventionnel sont en accord avec ce qui a été publié auparavant. En effet, les 6 modèles ont montré une certaine capacité prédictive, avec des p-values restant inférieures au seuil alpha de 0.05, soulignant encore une fois le potentiel de la radiomique comme facteur prédictif, en particulier du risque de récurrence de néoplasies rectales localement avancées et de son intérêt dans la sélection des patients à haut risque.

Comparaison des performances de modèles :

Dans notre étude, le test de Kurskal Wallis souligne deux résultats principaux :

- 1- L'absence de différence significative entre les modèles utilisant des données 2D vs les modèles utilisant les données 3D.
- 2- La non-infériorité des modèles utilisant la bounding-box par rapport aux modèles utilisant le contourage de la tumeur.

Qu'est ce qu'on apporte de plus par rapport à la littérature ? Shen et Al. dans leur étude évaluant les données 2D vs 3D, ont utilisé des images TDM sur des coupes pulmonaires [17]. Or, on sait que ce qu'il s'applique sur une modalité d'imagerie données (TDM, IRM, échographie, ...etc) ne s'applique pas forcément pour une autre modalité. De plus, l'anatomie est un facteur à prendre en compte, puisque là où le scanner est performant pour l'analyse du parenchyme pulmonaire, il est très mauvais pour l'évaluation locale du cancer rectal. A l'inverse l'IRM est très performante pour l'évaluation du cancer rectal, avec des indications très limitées dans l'évaluation du cancer broncho-pulmonaire. D'après nos connaissances, aucune étude antérieure n'a évalué la performance des données de texture 2D vs 3D sur des images IRM. L'avantage des données 2D est indéniable en termes de temps de calcul et de simplicité des modèles. Les résultats de notre étude sont donc en **faveur de l'utilisation des données 2D**.

La performance des modèles utilisant la bounding-box reste comparable aux autres modèles d'apprentissage conventionnel. Bien qu'elle ne soit pas inférieure, elle ne s'est pas montrée supérieure en terme de prédiction, dans les limites de la taille des échantillons. Ce rôle de bounding-box a déjà été évoqué par Hosny A et Al. qui ont suggéré la présence d'information au niveau de l'environnement péri-tumoral immédiat [20]. Bien que dans notre étude, on n'a pas réussi à démontrer l'existence d'une information prédictive au sein de cet environnement immédiat, cette idée de bounding-box n'est pas obsolète dans le sens où elle limite l'input humain. En d'autres termes, le radiologue n'a pas à segmenter la tumeur, mais il n'a qu'à cliquer sur l'épicentre de la tumeur et la récupération des coordonnées x, y et z de cet épicentre se fait automatiquement. On rajoutera des pixels de part et d'autre de ces coordonnées selon un nombre qu'on définit.

Dans le cancer rectal, la segmentation automatique n'est pas réalisable, en raison de l'anatomie de l'ampoule rectale et du tissu environnant. A ce jour, il n'existe aucun algorithme de segmentation automatique efficace applicable aux tumeurs rectales. Là où la segmentation manuelle des tumeurs rectales prendrait en moyenne 6-7 minutes, désigner l'épicentre de la tumeur pour construire une bounding-box autour prendrait moins de 5 secondes. Les résultats de notre étude donnent ainsi **l'avantage à l'utilisation des bounding boxes.**

2. Problème de l'apprentissage profond

Notre modèle de CNN n'a pas montré de potentiel prédictif.

Pour comprendre au mieux ce résultat, dans un premier temps, on a vérifié les sorties du modèle. Celles-ci étaient toutes de la même classe : soit c'est du 0 soit c'est du 1.

Dans un deuxième temps, on a cherché à comprendre les raisons derrière ce réseau de neurones donnant comme sorties une classe unique. Un petit tour de la littérature nous permet de recueillir quelques hypothèses concernant ce problème :

- Erreur liée au type de valeurs de pixels
- L'architecture du réseau de neurone lui-même.
- Le learning rate trop haut ou trop bas
- Le pré-traitement des images
- « a dying ReLU »
- Un réseau de neurone trop profond
- Et enfin, l'absence de corrélation entre l'entrée (images IRM) et le pronostic

A/ On affirme qu'il n'existe pas d'erreur liée au type de valeurs de pixels. Celles-ci ont été vérifiées à plusieurs reprises. Les valeurs ont été exprimées en float64.

B/ On ne pense pas que l'erreur soit liée à l'architecture du réseau. On a essayé plusieurs expérimentations, avec rajout et retrait de couches différentes.

C/ Concernant le learning rate, là encore plusieurs expérimentations ont été testées avec des valeurs entre 0,001 et 0,5. Le problème persiste avec l'ensemble des valeurs testées.

D/ Le prétraitement des images a correctement été réalisé, selon les données de l'état d'art.

E/ Le concept de dying ReLU fait référence à la fragilité de la fonction d'activation ReLU. Quand un large gradient passe à travers le neurone ReLU, il risque de changer les poids d'une certaine façon que ce neurone ne s'activera plus durant les itérations suivantes. Le résultat est que le neurone ReLU mort donnera toujours la même sortie. Pour pallier à ce problème, on a testé en lieu et place des fonctions ReLU la fonction « leakyReLU ». Celle-ci était censée donner un petit gradient positif quand l'entrée était négative selon la formule suivante ($y = 0.03x$ quand $x < 0$, avec x comme entrée et y comme sortie). Cette fonction leakyReLU était supposée régler le problème de la mort neuronale, Or le modèle reste toujours en prédiction mono-classe, même après changement des fonctions d'activation.

F/ La profondeur du réseau ne semble pas être un problème puisque différentes profondeurs ont été testées, du modèle à 11 couches jusqu'au modèle à 19 couches.

G/ Les modèles de machine learning conventionnel ont été capables de capter les informations prédictives enfouies au sein des images IRM de notre base de données en corrélation avec le

risque de récurrence. De ce fait, on rejette l'hypothèse selon laquelle l'échec du modèle de deep learning puisse être justifié par le manque de corrélation entre les données et le pronostic.

3. Perspectives

Par conséquent, il s'avère que les réseaux de neurones convolutionnels traitent les informations provenant des images IRM d'une manière complètement différente, par opposition aux techniques d'apprentissage conventionnel. Là où les techniques conventionnelles reçoivent comme entrée des données de texture résultat d'une forme d'ingénierie claire et relativement facile à comprendre, l'analyse de texture réalisée dans le noir par les CNN semble être différente et difficilement compréhensible. Ainsi les CNN ne constituent pas à ce jour d'équivalent automatique aux techniques d'apprentissage conventionnel, contrairement à ce qu'on avait supposé au début de l'étude. Bien qu'ils présentent l'avantage d'une certaine automaticité et de simplicité d'exécution, ils méritent leur qualificatif de « boîte noire ».

Dans un avenir proche, nous essaierons sur la même base de données de comprendre et déchiffrer les calculs qui se font à l'intérieur du réseau, dans l'objectif d'ouvrir cette boîte noire. Nous emploierons les cartographies d'activation, après chaque couche de neurones, afin de comprendre à quels pixels la couche de neurone attribue l'information. Cette étape nous semble en effet indispensable avant de conclure que l'apprentissage profond est incapable de déchiffrer l'information complexe des images IRM de tumeurs rectales.

Conclusion

Dans cette étude nous avons évalué et comparé 6 modèles d'apprentissage conventionnel et un modèle d'apprentissage profond, basés sur l'analyse texturale d'IRM de patients atteints de tumeurs rectales localement avancées, en corrélation avec le risque de récurrence.

En apprentissage conventionnel, nous avons comparé des modèles d'analyse d'images 2D vs des modèles d'analyse d'images 3D, des modèles basés sur une analyse texturale de la tumeur versus des modèles prenant en compte l'environnement péri-tumoral en plus de la tumeur elle-même.

En apprentissage profond, nous avons construit un modèle de réseau de neurones convolutionnels à 16 couches, entraîné par une base de données d'images IRM 2D comportant à la fois les images natives plus la bounding-box correspondant à chacun des images.

L'apprentissage conventionnel s'est montré nettement performant, avec pour chacun des modèles des signatures radiomiques capables de prédire avec précision le risque de récurrence.

A l'inverse l'apprentissage profond était incapable d'apprendre des patterns corrélés au pronostic. Il ne constitue pas de ce fait un substitut automatique des techniques plus conventionnelles contrairement à ce qui a été suggéré.

La comparaison des performances des modèles d'apprentissage conventionnel entre eux souligne deux faits principaux. Premièrement, là où les données de texture 3D présentent l'inconvénient d'être complexes et nécessitant du temps et des capacités de calcul conséquentes, les données de texture 2D ont montré des performances équivalentes avec l'avantage de la simplicité et le moindre coût en capacités de calcul.

Deuxièmement, au risque d'être chronophage, la segmentation manuelle préalable à l'extraction des données de texture en apprentissage conventionnel peut être remplacée par la création quasi-automatique des bounding-box, moins coûteuse en temps et en énergie, et incluant un environnement péri-tumoral potentiellement précieux pour la performance du modèle.

Bibliographie

- [1] Larue RT et al. Quantitative radiomics studies for tissue characterization : a review of technology and methodological procedures. *Br J Radiol.* 2017 Feb;90. doi: 10.1259/bjr.20160665.
- [2] Lambin et al. .Radiomics : the bridge between medical imaging and personalized medicine. *Nat Rev Oncol.* 2017 Dec;14(12):749-762. Doi:10.1038/nrclinonc.2017.141
- [3] Cook GJR et al. .What can artificial intelligence teach us about the molecular mechanisms underlying disease? *Eur J Nucl Med Mol Imaging.* 2019 Jun 12. Doi: 10.1007/s00259-019-04370-z
- [4] Larue RT et al. .Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *Br J Radiol.* 2017 Feb;90(1070):20160665. Doi: 10.1259/bjr.20160665
- [5] Bailey CE et al. .Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975-2010. *JAMA Surg.* (2015) 150:17–22. 10.1001/jamasurg.2014.1756
- [6] Titu LV et al. .Routine follow-up by magnetic resonance imaging does not improve detection of resectable local recurrences from colorectal cancer. *Ann Surg.*2006 ;243 :348-352.
- [7] Avanish P Saklani et al. .Magnetic resonance imaging in rectal cancer : A surgeon’s perspective. *World J Gastroenterol.* 2014 Feb 28; 20(8): 2030-2041. Doi: 10.3748/wjg.v20.i8.2030
- [8] Beckers R et al. Whole liver CT texture analysis to predict the development of colorectal liver metastases-A multicenter study. *Eur J radiol* 2017;92:64-71
- [9] Michele Avanzo et al. .Beyond imaging : The promise of radiomics. *European Journal of Medical Physics.* June 2017. Volume 38, Pages 122-139
Doi : <https://doi.org/10.1016/j.ejmp.2017.05.071>
- [10] Giger ML et al. Machine learning in medical imaging. *J Am Coll Radiol.* 2018 Mar;15(3 Pt B):512-520. Epub 2018 Feb 2. Doi: 10.1016/j.acr.2017.12.028.
- [11] Reza Forghani et al. . Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology. *Computational and Structural Biotechnology Journal.* Vol17, 2009, Pages 995-1008

- [12] June-Goo Lee et al. . Deep learning in Medical Imaging : General Overview. Korean J radiol. 2017 Jul-Aug; 18(4): 570-584. Doi : 10.3348/kjr.2017.18.4.570
- [13] Vishwa et al. . Deep learning and radiomics in precision medicine. Expert Rev Precis Med Drug Dev. 2019; 4(2): 59-72. Doi: 10.1080/23808993.2019.1585805
- [14] Horvat N et al. .Novel imaging techniques of rectal cancer : what do radiomics and radiogenomics have to offer? A literature review. Abdominal Radiology. <https://doi.org/10.1007/s00261-019-002042-y>
- [15] Liang M et al. .Machine Learning-based Analysis of Rectal Cancer MRI Radiomics for Prediction of Metachronous Liver Metastasis. Acad Radiol. 2019 Jan 30. Doi : 10.1016/j.acra.2018.12.019.
- [16] Jalil et al. Magnetic resonance based texture parameters as potential imaging biomarkers for predicting long-term survival in locally advanced rectal cancer treated by chemoradiotherapy. Colorectal Dis. 2017 apr;19(4):349-362. Doi : 10.1111/codi.13496
- [17] Shen et al. .2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. Transl Oncol. 2017 Dec;10(6):886-894. Doi : 10.1016/j.tranon.2017.08.007.
- [18] Edward R Dougherty et al. .Performance of Feature Selection Methods. Curr Genomics. 2009 Sep; 10(6): 365-374. Doi : 10.2174/138920209789177629
- [19] Chintan Parmar et al. .Machine Learning methods for Quantitative Radiomics Biomarkers. Sci Rep. 2015; 5:13087. Doi:10.1038/srep13087
- [20] Hosny et al. .Artificial intelligence in radiology. Nat Rev Cancer. 2018 Aug; 18(8): 500-510. Doi: 10.1038/s41568-018-0016-5
- [21] Forsberg D et al. Detection and labeling of vertebrae in MR images using deep learning with clinical annotations as training data. J Digit imaging. 2017;30:406-12. Doi: 10.1007/s10278-017-9945-x
- [22] Ghafoorian M et al. .Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. Sci Rep. 2017;7:5110. Doi : 10.1038/s41598-017-05300-5
- [23] Miao S et al. .A CNN regression approach for real-time 2D/3D registration. IEEE Trans Med Imaging. 2016;35:1352-63. Doi : 10.1109/TMI.2016.2521800
- [24] Ahmed Hosny et al. .Deep learning for lung cancer prognostication : A retrospective multi-cohort radiomics study. PLoS Med. 2018 Nov;15(11): e1002711. Doi : 10.1371/journal.pmed.1002711

Annexes

Annexe 1 :

Facteurs pronostiques	Résultats descriptifs
Age	Moyenne = 60 ans. Intervalle = [21-88] ans
Sexe	sexe ratio H/F = 2.0625
Délai diagnostic-chirurgie Délai fin RCT-chirurgie	Médiane = 13,5 mois. Intervalle = [2-18] mois Médiane = 6,8 mois. Intervalle = [3-10] mois
Stade TNM : - T - N	- T4 : 4 patients - T3 : 9 patients - T2 : 3 patients - T1 : 23 patients - T0 : 23 patients - Tx : 17 patients - N0 : 66 patients - N1 : 32 patients
Tableau II : caractéristiques de la population étudiée	

Annexe II : Programmation du modèle 1

```

import numpy as np
import pydicom
import os
from glob import glob
#from mpl_toolkits.mplot3d.art3d import Poly3D Collection
import scipy.ndimage
from skimage import morphology
from skimage import measure
from skimage.transform import resize
from sklearn.cluster import KMeans
from skimage.io import imread,imsave
import numpy.ma as ma
import SimpleITK as sitk
import six
import pandas as pd
from __future__ import absolute_import
import matplotlib.pyplot as plt
from radiomics import featureextractor
import keras
from xlr import open_workbook
from sklearn.model_selection import StratifiedKFold
from sklearn.metrics import roc_curve, auc
from scipy import interp
from sklearn.ensemble import RandomForestRegressor
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFECV
from sklearn.ensemble import RandomForestClassifier

```

Préparation et pré-traitement d'images

```
xls_table = open_workbook('/Users/hichem/Desktop/Master project/Greccar/Tableau clin.xlsx')
df = pd.read_excel(xls_table, sheetname=0, header=0, index_col=False, keep_default_na=True)
main_path="/Users/hichem/Desktop/Master project/Greccar_Martina/Greccar/"
RAW_DATA=main_path+"RAW_2D_Baseline/"
SEG_DATA=main_path+"SEG_2D_Baseline/"
OUTPUT_SIZE=256
#def readRAWSlice(ipatient,s):
#PathDicomP = main_path+str(ipatient)+"/Ax T2/"
#print("PathDicomP " +PathDicomP)
#for dirName, subdirList, fileList in os.walk(PathDicomP): #JUST TO GET FOLDER NAME
#print("dirName " +dirName)
#for filename in fileList: #EACH DICOM FILE IS ONLY ONE SLICE
#print("TEST RAW " +os.path.join(dirName,filename));
#if "IMG"+str(s).zfill(4)+".dcm" in filename: #SLICE OF INTERSET
#print("READ RAW " +os.path.join(dirName,filename));
#ds = pydicom.read_file(os.path.join(dirName,filename))
#return ds
#return None

#def readRAWSeg2D(ipatient,s):
#PathSeg2DP = main_path+str(ipatient)+"/Seg 3D/"
#for dirName, subdirList, fileList in os.walk(PathSeg2DP): #JUST TO GET FOLDER NAME
#for filename in fileList: #EACH DICOM FILE IS ONLY ONE SLICE
#if "IMG"+str(s).zfill(4)+".dcm" in filename: #SLICE OF INTERSET
#print("READ SEG " +os.path.join(dirName,filename));
#ds = pydicom.read_file(os.path.join(dirName,filename))
#return ds
#return None

#for i in range(1,99):
#s = df.iloc [i-1,5] #SLICE OF INTEREST
#o = df.iloc [i-1,17] #RECURRENCE
#s = int(s)
#o = int(o)
#print('Process Patient '+str(i) + " with slice="+str(s)+ " -> "+str(o))
#ds=readRAWSlice(i,s) #READ 2D SLICE
#dss=readRAWSeg2D(i,s) #READ 2D SEGEMENTATION
#if ds is not None and dss is not None:
#ConstPixelDims = (int(ds.Rows), int(ds.Columns)) #MATRIX N x M (not ++)
#ConstPixelSpacing = (float(ds.PixelSpacing[0]), float(ds.PixelSpacing[1]))
#print("-> " +str(ConstPixelDims)+ " , " +str(ConstPixelSpacing));

#if ds.Rows==ds.Columns: #SQAURE IMAGE
#im=resize(ds.pixel_array,[OUTPUT_SIZE,OUTPUT_SIZE],preserve_range=True).astype(dtype=ds.pixel_array.dtype)
#seg=resize(dss.pixel_array,[OUTPUT_SIZE,OUTPUT_SIZE],preserve_range=True).astype(dtype=dss.pixel_array.dtype)
#ratio=float(OUTPUT_SIZE)/float(ds.Columns)
#else:
#maxHW=max(ds.Rows,ds.Columns)
#ratio=float(OUTPUT_SIZE)/float(maxHW)
#im=np.zeros([OUTPUT_SIZE,OUTPUT_SIZE],dtype=ds.pixel_array.dtype)
#seg=np.zeros([OUTPUT_SIZE,OUTPUT_SIZE],dtype=dss.pixel_array.dtype)
#if ds.Rows<ds.Columns:
#newW=int(ds.Rows*ratio)
#imsmall=resize(ds.pixel_array,[newW,OUTPUT_SIZE],preserve_range=True).astype(dtype=ds.pixel_array.dtype)
#imsmalls=resize(dss.pixel_array,[newW,OUTPUT_SIZE],preserve_range=True).astype(dtype=dss.pixel_array.dtype)
#w=int( ( OUTPUT_SIZE-newW ) / 2.0 )
#im[w:w+newW,:]=imsmall
#seg[w:w+newW,:]=imsmalls
#print("ds.Rows -> "+str(im.shape))
#else:
#newW=int(ds.Columns*ratio)
#imsmall=resize(ds.pixel_array,[OUTPUT_SIZE,newW],preserve_range=True).astype(dtype=ds.pixel_array.dtype)
#imsmalls=resize(dss.pixel_array,[OUTPUT_SIZE,newW],preserve_range=True).astype(dtype=dss.pixel_array.dtype)
#w=int( ( OUTPUT_SIZE-newW ) / 2.0 )
#im[:,w:w+newW]=imsmall
#seg[:,w:w+newW]=imsmalls
#print("ds.Columns -> "+str(im.shape))

#imsave(RAW_DATA+"RAW_"+str(i)+"_"+str(o)+".tiff",im)
#imsave(SEG_DATA+"SEG_"+str(i)+"_"+str(o)+".tiff",seg)
#ArrayDicom[:, :, lstFilesDCM.index(filenameDCM)] = ds.pixel_array
```

```

    #print(" converted -> " +str(np.array(ConstPixelSpacing)/ratio));
#else:
    #print('----->>>>> SKIP '+str(i))
    #if ds is None:
        #print('Miss RAW')
    #if dss is None:
        #print('Miss SEG')

# Extraction des données de radiomique
iteration = 0
for i in range(1,99):
    iteration+=1
    s = df.iloc [i-1,5] #SLICE OF INTEREST
    o=df.iloc [i-1,17] #RECURRENCE
    s = int(s)
    o = int(o)
    patient = i
    #testPath = imagePath
    #testPath += paths(imagePath)

    imagePath = sitk.ReadImage(RAW_DATA+"RAW_"+str(i)+"_"+str(o)+".tiff")

    #testPath = labelPath
    #testPath += paths(testPath)

    labelPath = sitk.ReadImage(SEG_DATA+"SEG_"+str(i)+"_"+str(o)+".tiff")

    params = {'minimumROISize': None, # Skip testing the ROI size by default
              'normalize': True,
              'normalizeScale': 255,
              'removeOutliers': None,
              'interpolator': 'sitkBSpline', # Alternative: sitk.sitkBSpline
              'padDistance': 5,
              'distances': [1],
              'force2D': False,
              'force2Ddimension': 0,
              'resegmentRange': None, # No resegmentation by default
              'label': 1,
              'enableCExtensions': True,
              'additionalInfo': True}
    extractor = featureextractor.RadiomicsFeatureExtractor()
    extractor.enableImageTypeByName("Square")
    extractor.enableImageTypeByName("Wavelet")
    extractor.enableImageTypeByName("SquareRoot")
    extractor.enableImageTypeByName('LoG')
    extractor.enableImageTypeByName('Logarithm')
    extractor.enableImageTypeByName('Exponential')
    extractor.enableImageTypeByName('Gradient')
    extractor.enableImageTypeByName('LBP2D')
    result = extractor.execute(imagePath, labelPath)
    #k=r.keys()
    #v=r.values()
    #table[i,:]=v
    if iteration == 1:
        table_result = pd.DataFrame(six.iteritems(result))
        table_result.rename(columns={0:'parameters',1:'patient_'+str(i)}, inplace=True)
    else :
        table = pd.DataFrame(six.iteritems(result))
        title = 'patient_'+str(i)
        table_result[title]=table.iloc[:,1]
    #pd.concat(objs, axis=0, join='outer', join_axes=None, ignore_index=False, keys=None, levels=None, names=None,
    verify_integrity=False, sort=None, copy=True)[source]
    #for key, val in six.iteritems(result):
        #print("\t%s: %s" %(key, val))
    table_fin = table_result.T
    outcome=[]
    outcome.append('outcome')

for i in range (1,99):
    o = df.iloc[i-1,17]
    o = int(o)
    outcome.append(o)

```

```

table_fin.insert(loc=0, column='outcome', value=outcome)
table_fin.drop('outcome', axis=1,inplace=True)
X = table_fin.iloc[1:]
X.drop([0,1,2,3,4,5,6,7,8,9,10,14,15,16,17,20,21], axis=1, inplace=True)
from sklearn.ensemble import ExtraTreesClassifier
Y=[]
for i in range (1,99):
    o = df.iloc[i-1,17]
    o = int(o)
    Y.append(o)

```

Sélection des données et classification

```

# The "accuracy" scoring is proportional to the number of correct classifications
clf_rf_4 = RandomForestClassifier()
rfecv = RFECV(estimator=clf_rf_4, step=50, cv=5,scoring='accuracy') #5-fold cross-validation
rfecv = rfecv.fit(X, Y)
print('Optimal number of features :', rfecv.n_features_)
print('Best features :', X.columns[rfecv.support_])
columns = X.columns[rfecv.support_]
X = X.loc[:, X.columns.isin(list(columns))]
sm = SMOTE(random_state=42)
X, Y = sm.fit_sample(X, Y)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.5)
rfc = RandomForestRegressor(n_estimators = 1000, random_state = 42)
rfc.fit(X_train, Y_train)
rfc_predict = rfc.predict(X_test)
rfc_predict = rfc_predict.round()
print (rfc_predict)
print (Y_test)

```

Test de la performance et obtention des courbes de ROC

```

cv = StratifiedKFold(n_splits=5)
classifier = RandomForestRegressor(n_estimators = 1000, random_state = 42)
tprs = []
aucs = []
mean_fpr = np.linspace(0, 1, 100)

i = 0
for train, test in cv.split(X, Y):
    probas_ = classifier.fit(X[train], Y[train]).predict(X[test])
    probas_ = probas_.round()
    # Compute ROC curve and area the curve
    fpr, tpr, thresholds = roc_curve(Y[test], probas_)
    tprs.append(interp(mean_fpr, fpr, tpr))
    tprs[-1][0] = 0.0
    roc_auc = auc(fpr, tpr)
    aucs.append(roc_auc)
    plt.plot(fpr, tpr, lw=1, alpha=0.3,
             label='ROC fold %d (AUC = %0.2f)' % (i, roc_auc))

    i += 1
plt.plot([0, 1], [0, 1], linestyle='--', lw=2, color='r',
         label='Chance', alpha=.8)

mean_tpr = np.mean(tprs, axis=0)
mean_tpr[-1] = 1.0
mean_auc = auc(mean_fpr, mean_tpr)
std_auc = np.std(aucs)
plt.plot(mean_fpr, mean_tpr, color='b',
         label='r'Mean ROC (AUC = %0.2f %pm$ %0.2f)' % (mean_auc, std_auc),
         lw=2, alpha=.8)

std_tpr = np.std(tprs, axis=0)
tprs_upper = np.minimum(mean_tpr + std_tpr, 1)
tprs_lower = np.maximum(mean_tpr - std_tpr, 0)
plt.fill_between(mean_fpr, tprs_lower, tprs_upper, color='grey', alpha=.2,
                 label='r'$ %pm$ 1 std. dev.')

plt.xlim([-0.05, 1.05])
plt.ylim([-0.05, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic example')

```

```
plt.legend(loc="lower right")
plt.show()
```

Annexe III : Programmation du modèle d'apprentissage profond

```
import numpy as np
import pydicom
import os
from glob import glob
import scipy.ndimage
from skimage import morphology
from skimage import measure
from skimage.transform import resize
from sklearn.cluster import KMeans
from skimage.io import imread,imsave
import numpy.ma as ma
import SimpleITK as sitk
import six
import pandas as pd
import keras
import numpy as np
from future import absolute_import
from sklearn import preprocessing
from keras.utils import np_utils
```

Préparation et prétraitement d'images

```
cwd = os.getcwd()
os.chdir("/Users/hichem/Desktop/Master project/Greccar")
from xlr import open_workbook
xls_table = open_workbook("/Users/hichem/Desktop/Master project/Greccar/Tableau clin.xlsx")
df = pd.read_excel(xls_table, sheetname='Sheet1', header=0, index_col=False, keep_default_na=True)
main_path="/Users/hichem/Desktop/Master project/Greccar/Post/"
PATH_DATA=main_path+"Path_response/"
OUTPUT_SIZE=224
#LIST ALL DATAS
datas=[]
for filename in os.listdir(PATH_DATA):
    if filename.find("BB_")>=0:
        datas.append(filename)
for filename in os.listdir(PATH_DATA):
    if filename.find("BB2_")>=0:
        datas.append(filename)
for filename in os.listdir(PATH_DATA):
    if filename.find("RAW2")>=0:
        datas.append(filename)
for filename in os.listdir(PATH_DATA):
    if filename.find("RAW_")>=0:
        datas.append(filename)

import random
random.shuffle(datas)
NbImagesTotal=len(datas)

#PREPARE TRAIN DATASET
traintestsplit=0.8 #80% FOR TRAIN
NbImagesTrain=int(NbImagesTotal*traintestsplit)
NbImagesTest=NbImagesTotal-NbImagesTrain

#PREPARE TRAIN DATASET
TRAIN_DATAS = np.zeros([ NbImagesTrain,OUTPUT_SIZE,OUTPUT_SIZE]).astype('float32') #X TRAIN
OUTPUT_TRAIN_DATAS = np.zeros([ NbImagesTrain,1]).astype('float32') # Y TRAIN

for i in range(NbImagesTrain):
    rawname=datas[i]
    #print(PATH_DATA+rawname)
    raw=imread(PATH_DATA+rawname)
    #segname=rawname.replace('RAW','SEG')
    #seg=imread(segname)
    raw = resize(raw, [224, 224])
    ratio = raw.max()/255.0
    TRAIN_DATAS[i,:,:]=raw/ratio
    #[indx, indy]=np.where(TRAIN_DATAS[i,:,:]<0)
```

```

#TRAIN_DATAS[indx, indy]=0.
#TRAIN_DATAS[i,:]=raw
#DATAS[i,:]=seg
o=int(rawname.replace('.tiff','').replace("RAW2_','").replace("BB2_','").replace("RAW_','").replace("BB_','").split("_")[1])
OUTPUT_TRAIN_DATAS[i]=int(o)
from imblearn.over_sampling import SMOTE
nsamples, nx, ny = TRAIN_DATAS.shape
TRAIN_DATAS = TRAIN_DATAS.reshape((nsamples,nx*ny))
sm = SMOTE(random_state=42)
TRAIN_DATAS, OUTPUT_TRAIN_DATAS = sm.fit_sample(TRAIN_DATAS, OUTPUT_TRAIN_DATAS)
print (TRAIN_DATAS.shape)
ns, nz = TRAIN_DATAS.shape
TRAIN_DATAS = TRAIN_DATAS.reshape((ns,nx,ny))
keras.backend.image_data_format()
TRAIN_DATAS = np.stack((TRAIN_DATAS,)*3, axis=-1)
from keras.utils import np_utils
OUTPUT_TRAIN_DATAS = np_utils.to_categorical(OUTPUT_TRAIN_DATAS, 2)
print(OUTPUT_TRAIN_DATAS)

```

Construction et entraînement du modèle

```

# NETWORK
from keras.models import Sequential
from keras.layers.core import Flatten, Dense, Dropout
from keras.layers import LeakyReLU
from keras.layers.convolutional import Convolution2D, MaxPooling2D, ZeroPadding2D
from keras.optimizers import SGD
import cv2, numpy as np
from keras.layers import LeakyReLU
from keras.regularizers import l2
#import tensorflow as tf
#run_opts = tf.RunOptions(report_tensor_allocations_upon_oom = True)

def VGG(weights_path=None):
    model = Sequential()
    model.add(ZeroPadding2D((1,1),input_shape=(OUTPUT_SIZE,OUTPUT_SIZE, 3))) # with OUTPUT_SIZE = 224
    model.add(Convolution2D(64, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(64, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(128, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(128, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(256, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(Dropout(rate = 0.25))
    model.add(MaxPooling2D((2,2), strides=(2,2)))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(ZeroPadding2D((1,1)))
    model.add(Convolution2D(512, 3, 3, activation='relu'))
    model.add(MaxPooling2D((2,2), strides=(2,2)))
    model.add(Flatten())
    model.add(Dense(4096, activation='relu'))
    model.add(Dropout(rate = 0.5))
    model.add(Dense(4096, activation='relu'))
    model.add(Dropout(rate = 0.5))
    model.add(Dense(2, activation='sigmoid'))

```

```

if weights_path:
    model.load_weights(weights_path)
    optimizer = SGD(lr=0.09, momentum=0.9, nesterov=True)
    model.compile(optimizer=optimizer, loss='binary_crossentropy', metrics=['accuracy'])#, #options = run_opts) # SCC instead of MSE
    model.summary()
    model.save_weights('model.h5')
    return model
model = VGG()

TRAIN_DATAS = TRAIN_DATAS/255.0

from keras.preprocessing.image import ImageDataGenerator
datagen = ImageDataGenerator(featurewise_center=True, featurewise_std_normalization=True, rotation_range=20,
                             width_shift_range=0.2, height_shift_range=0.2, horizontal_flip=True)

# compute quantities required for featurewise normalization
# (std, mean, and principal components if ZCA whitening is applied)
datagen.fit(TRAIN_DATAS)

# fits the model on batches with real-time data augmentation:
model.fit_generator(datagen.flow(TRAIN_DATAS, OUTPUT_TRAIN_DATAS, batch_size=batch_size),
                    steps_per_epoch=len(TRAIN_DATAS) / batch_size, epochs=epochs)

```

Test de la performance sur la cohorte de validation

```

TEST_DATAS = np.zeros([ NbImagesTest,OUTPUT_SIZE,OUTPUT_SIZE]) #X TEST
OUTPUT_TEST_DATAS = np.zeros([ NbImagesTest,1]) # Y TEST

for i in range(NbImagesTest):
    rawname=datas[i]
    #print(PATH_DATA+rawname)
    raw=imread(PATH_DATA+rawname)
    #segname=rawname.replace('RAW','SEG')
    #seg=imread(segname)
    ratio = raw.max()/255.0
    TEST_DATAS[i,:,:]=raw/ratio
    #DATAS[i,:,:,1]=seg
    o=int(rawname.replace('.tiff','').replace("RAW2_',"").replace("BB2_',"").replace("RAW_',"").replace("BB_',"").split("_")[1])
    OUTPUT_TEST_DATAS[i]=o
from imblearn.over_sampling import SMOTE
nsamples, nx, ny = TEST_DATAS.shape
TEST_DATAS = TEST_DATAS.reshape((nsamples,nx*ny))
ns, nz = TEST_DATAS.shape
sm = SMOTE(random_state=42)
TEST_DATAS, OUTPUT_TEST_DATAS = sm.fit_sample(TEST_DATAS, OUTPUT_TEST_DATAS)
ns, nz = TEST_DATAS.shape
TEST_DATAS = TEST_DATAS.reshape((ns,nx,ny))
TEST_DATAS = np.repeat(TEST_DATAS[... , np.newaxis], 3, -1) # to solve the 3 channels issue
OUTPUT_TEST_DATAS = np_utils.to_categorical(OUTPUT_TEST_DATAS, 2)
print(OUTPUT_TEST_DATAS.shape)
from sklearn.metrics import confusion_matrix
from __future__ import print_function
from sklearn.metrics import accuracy_score
predictions = model.predict(TRAIN_DATAS)
print (predictions.round())

```

L'article

Evaluation of different radiomics approaches to predict patient outcome in locally advanced rectal cancerH Tibermacine^{1,2}, P Rouanet³, M Sbarra^{4,5}, C Reinhold⁶, S Nougaret^{1,2}

¹ Radiology Department, Institut du Cancer de Montpellier, Univ. Montpellier, Montpellier, France

² Institut de Recherche en Cancérologie de Montpellier (IRCM), INSERM, U1194, Montpellier, France.

³ Surgical oncology Department, Institut du Cancer de Montpellier, Univ. Montpellier, Montpellier, France

⁴ Radiologia Diagnostica e Interventistica Generale, Area Diagnostica per Immagini, Dipartimento Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Fondazione Policlinico Universitario A. Gemelli IRCCS, Largo A. Gemelli, 8, Rome, 00168, Italy

⁵ Istituto di Radiologia, Università Cattolica del Sacro Cuore, Largo A. Gemelli, 8, Rome, 00168,

⁶ Department of Radiology, McGill University, Montreal, QC, Canada.

Correspondance to:

Dr Stephanie Nougaret
Institut du Cancer de Montpellier (ICM)

208 rue des Apothicaires
34 298 Montpellier, France
Tel: +33 4 67 61 47 25

Category: Submitted as an original article

Support/Funding: French National Cancer Institute (INCa).
Grant number INCa-DGOS_5506: PHRC-K 2012-112.

List of where the study was presented, if applicable: Not applicable

Disclaimers: The authors claim no conflict of interest regarding this manuscript

Abstract:

Purpose: To compare different radiomics approaches to predict patients outcome in locally advanced rectal cancer (LARC) using MRI radiomics engineered features or deep learning models at baseline and after neoadjuvant chemoradiotherapy (CRT).

Methods: 98 patients from a phase II, prospective, multicenter, randomized study (GRECCAR4-NCT01333709) were included in this study. T2-weighted sequences in conjunction with diffusion weighted images (DWI) from baseline and post CRT images were used for the analysis of 3 radiomics models at baseline and after CRT using conventional machine learning (CML) techniques. For both time point, features were extracted from 2D manual segmentation (MS), 3D MS and from bounding boxes (BBs). Recursive feature elimination and a random forest classifier were used to select features and build each model. Additionally, a deep learning model (DL), using a 16-layer convolutional neural network (CNN) was evaluated as well.

Results: All 6 models of CML were able to predict patient outcome with AUCs ranging from 0.69 to 0.92. Comparison of performances among the 3 baseline models and after CRT did not show significant differences. In contrast, deep learning model demonstrated hazardous performance with an average accuracy of 0.51 +/- 0.3. BB Model had the highest stability across the 5 iterations of the cross-validation, with an AUC of 0.8 +/- 0.12 at baseline and 0,83 +/- 0.04 after CRT to predict patient outcome.

Conclusion: Among the 3 radiomics approaches to predict patient outcome using CML techniques, BBs were the easiest to obtain and be translated to a busy routine practice.

Keywords: *RECTAL CANCER; RADIOMICS; OUTCOME; MACHINE LEARNING;*

INTRODUCTION

The era of personalized cancer therapy has resulted in substantial changes in the management of patients with locally-advanced rectal cancer (LARC) [1, 2]. In the past, standard chemoradiotherapy was followed by surgical resection in nearly all patients, irrespective of tumor response. Nowadays, patient's stratification according to the likelihood of their early treatment response is part of the new therapeutic strategies offering minimally-invasive treatments for potentially good responders and following a more "aggressive" approach for patients likely to show a poor response [3-5]. In this particular context, finding reliable prognostic indicators for monitoring treatment response in order to switch to alternative strategies is critical and could enable treatment personalization [12].

Multiple MRI approaches have been proposed to assess tumor response, including morphological changes with tumor regression assessed on MRI (mrTRG) [13] and volumetry [12, 14-19], or functional changes using diffusion-weighted imaging (DWI) [14, 15, 20-24] and dynamic contrast-enhanced MRI (DCE-MRI) [25-29]. Recently, radiomics has been introduced as an emergent tool for postprocessing CT or MR images and developing new quantification metrics linking qualitative and/or quantitative imaging data to clinical endpoints [30-36]. This may allow development of new biomarkers for diagnosis, prognosis and response

evaluation [37-38]. Radiomics has already demonstrated considerable potential for lesion characterization or tumor aggressiveness [31, 52-56]. In rectal cancer, T2-derived texture metrics extracted from the whole tumor volume have been shown to outperform the combination of T2 and diffusion-weighted images to assess complete response using either conventional machine learning (CML) [57, 58] or deep learning (DL) techniques [59]. However, conventional machine learning using 3D features extracted from manual segmentation might be time consuming and difficult to transfer to a busy routine practice, in contrast 2D features could serve as a quicker alternative. Similarly, 2D bounding box (BB) could be as well a substitute to manual segmentation. BB is a tool used for object recognition and image classification [60]. After the lesion epicenter is chosen in one click, an automated rectangular box with a chosen surface size is drawn. Finally, to further limit human input, and circumvent the lack of standard protocols in terms of feature selection and classification algorithms, convolutional neural network (CNN) could be an alternative.

Thus, our study's aim was to test different radiomics approaches in rectal cancer, to predict patient outcome with the potential to make it a more automated and standardized process.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the Ethics Committee of our institution, and informed consent was obtained for all patients. All patients initially enrolled in the original prospective II multicenter randomized study (n=206) (GRECCAR4-NCT01333709) were included. The inclusion and exclusion criteria are detailed in appendix 1.

The final cohort included all the 98 patients (66 male and 32 female; mean age: 60 ys; range [21-88 years]) who had both baseline and post CRT MRI. Among them, 74 had no evidence of local and/or distant recurrence during the follow-up period (median of 55 months; range [6-77 months]), and 24 developed a local and/or distant recurrence. Recurrence was diagnosed by histopathology or follow-up imaging with new lesions.

MR Acquisitions:

Imaging was performed using a phased array body coil at 1.5 or 3T (Supplementary Table 1 in appendix) include the MRIs parameters reported in the 16 participating centers; these parameters were in accordance with the European Society of Gastrointestinal and Abdominal Radiology guidelines ^{40,41}).

MR post processing method :

Conventional machine learning using engineered radiomics features:

Image Segmentation and Bounding Box

Regions of interest (ROI) were manually segmented by a radiologist with 2 years of experience in rectal imaging. MS was performed using free open source segmentation software (3D slicer (<https://www.slicer.org/>)). 2D and 3D MS approaches were tested at baseline and after CRT.

T2WI sequence was used for MS in conjunction with DWI. ROIs were drawn to cover the whole rectal tumor on each consecutive slice for 3D models, on a chosen slice covering the center of the lesion on 2D models, as illustrated on figure 1.

A third segmentation approach (semi-automated) was assessed using a 2D bounding-box (BB). BB is a technique used by deep learning engineers, particularly for object detection, recognition and image classification tasks. In this approach, the process of delineation starts with the designation, by the radiologist, of one seeding point, corresponding to the epicenter of the lesion of interest. From there, 1 x 1 mm dimensions pixels were added to cover an overall surface of 100 x 100 pixels, as illustrated on figure 1.

In order to standardize image intensity and size as suggested by Rizzo et al. [61], image intensity was binned by 20 and patches were normalized to a 0-255 range. Image resampling was performed using linear interpolation to get pixels of 1 x 1 mm dimensions, with 1mm depth for the 3D model. 224 x 224 pixels images were used for 2D and 3D MS models and 100 x 100 pixels images were used for the BB models.

Figure 2 illustrates the analytical process and gives an overview of the approaches/models being tested.

Features extraction

More than 1000 radiomic features were automatically extracted from MS ROIs and BBs regarding tumor shape and size, intensity statistics, and texture information (including the gray level co-occurrence matrix, gray level run-length matrix, and gray-level size zone matrix). In addition, the whole radiomic-feature set also contained higher order statistical features, including the intensity and texture features derived from the images processed with eight types of filters (Wavelet, Laplacian of Gaussian, Square, Square Root, Logarithm, Exponential, Gradient, LBP2D for 2D models or LBP3D for 3D models). Features were computed using PyRadiomics [62], an open-source radiomics package.

Softwares, Statistical Analysis and Radiomic-Model Building

Statistical analyses, feature selection and model development were performed using python (<https://www.python.org/>) in Anaconda3 platform (<https://www.anaconda.com>) with Scikit-learn (<https://scikit-learn.org/>) and Matplotlib packages (<https://matplotlib.org/>).

Due to unbalanced sampling between the non -relapse and relapse group, we applied SMOTE (sample minority oversampling technique) method to resolve the balance issue by oversampling the recurrent group [63]. Each sample was assigned one predictive value, either a 0 for absence of recurrence or a 1 for recurrence.

The recursive feature elimination as a feature selection technique was integrated within a random forest (RF) classifier. During each iteration the number of removed features was set to 50. Five-fold cross-validation and 1000 trees were used to train and test the different models.

The diagnostic performance of the models was evaluated by receiver operating characteristic (ROC) curves and confusion matrix with indicators of accuracy, sensitivity, specificity, and

area under the curve (AUC). Finally, a comparison was made between the models performances using the Kruskal-Wallis test.

Deep learning

We employed a 2D CNN (convolutional neural network) architecture inspired by the VGG-16 proposed by K. Simonyan and A. Zisserman [64]. The network comprised a total of five 2D convolutional layers of 64, 128, 256, 512 and 512 filters with kernel sizes of 3 x 3 x 3. The 5 max pooling layers of kernel size 2 x 2 were applied after each convolutional layer. A series of 2 fully connected layers with 4096 units each provided high-level reasoning before the prediction probabilities were calculated in the final sigmoid classifier layer.

Deep learning training details:

A gradient-based stochastic optimizer Adam with a learning rate of 0.001 without decay [65], a batch size of 49, and dropout of 50% on the fully connected layers was used. Rectified linear units (ReLUs) were the activation function across the entire network prior to the final sigmoid activation. In training, BB dataset was used to train and test our model. Data split ratio was set to a 4:1 with 80% of patients used for training and a 20% independent set for validation.

Data preprocessing for deep learning

Resampled BBs were normalized to a 0-1 range, then were subject to an augmentation function, incorporated within the neural network. This augmentation was done in real time during training and included random rotations in an interval of 20°, random translations ± 10 pixels in the horizontal axis, and horizontal flip.

Data analysis:

Model performance was assessed across 5 iterations, with results returning an average AUC, an accuracy representing the ratio of correct predictions over all predictions, a sensitivity corresponding to the ratio of recurrence predictions over the number of recurrent patients, a specificity consisting on the number of non-recurrence predictions over the number of non-recurrent patients.

RESULTS

Demographic and baseline patients' characteristics are summarized in supplementary *Table 2 (appendix)*

Conventional machine learning

At baseline

1045 features were automatically extracted from each 2D and 3D MS ROIs and 1632 from the BBs. After applying the recursive feature elimination process 28 features for 2D MS, 22 for 3D MS and 8 for BBs were selected. Details of selected features are summarized in table 3. Models from 2D MS, 3D MS and BBs were then created using those selected features and RF classifier as detailed in M&M.

All three models showed good performance with AUCs ranging from 0.69 to 0.92 (figure 3). Performance details for conventional machine learning models are summarized in table 4.

No statistical difference was found between the 3 models.

At presurgical (post-CRT) assessment

1045 features were automatically extracted from each 2D, 3D MS ROIs and BBs. After applying the recursive feature elimination process 21 features for 2D MS, 22 for 3D MS and 8 for BBs were selected. Details of selected features are summarized in table 5.

All 3 models showed good performance, in particular models using 2D features and BB demonstrating high performance (figure 4). Performance details for conventional machine learning models are summarized in table 4. BB Model had the highest stability across the 5 iterations of the cross-validation, with an AUC of 0,83 (95% CI: 0.79-0.87), an accuracy, a sensitivity and a specificity of 0.81, 90, 75 respectively. However, no statistical difference was found between the 3 models.

Deep Learning

CNN architecture resulted in a number of 134.268.738 trainable parameters.

During the training phase, the model failed to make the function loss converge, and the accuracy was random. During the validation phase, CNN showed poor performance, with mono-class predictions, yielding to hazardous results, with an average accuracy of 0.51 (+/- 0.3), sensitivity and specificity of 0.6, 0.39 respectively.

DISCUSSION

This study evaluated different radiomics approaches to predict patient outcome in rectal cancer. First, conventional machine learning models were able to accurately predict the risk of recurrence in rectal cancer at baseline and after CRT. Second, there were no significant differences among the models used. Finally, deep learning model failed to predict patient outcome.

Radiomics has been introduced as an emergent tool for post processing CT or MR images and developing new quantification metrics [4-10] linking qualitative and/or quantitative imaging data to clinical endpoints. Several studies have now been published showing a potential interest of radiomics in predicting treatment response and outcome in rectal cancer. For example, at baseline, studies have suggested the role of pretherapeutic radiomic features in the prediction of non-response to neoadjuvant therapy in patients with LARC [66-71]. Lovinfosse et al investigated the prognostic value of baseline ¹⁸F-FDG PET/CT textural features and demonstrated that such features might provide strong independent predictors of survival in patients with LARC [66]. Texture parameters homogeneity and coarseness were significantly associated with disease free survival in a multivariate analysis. Similarly, dissimilarity and contrast from the neighborhood intensity-difference matrix (contrastNGTDM) were significantly associated with overall survival. Additionally, in a study conducted by Lucia et al. Radiomics features from both baseline MR and ¹⁸F-FDG PET/CT were found potential indicators for local recurrence and loco-regional control [71] with features such as Grey Level Non Uniformity_{GLRLM} in PET and Entropy_{GLCM} in ADC maps from DWI MRI

identified as independent prognostic factors. These features had significantly higher predictive value than clinical parameters in the testing cohort with accuracy of 94% for predicting recurrence and 100% for predicting lack of loco-regional control (versus ~50-60% for clinical parameters). These studies hint to the importance of baseline radiomics as a potential indicator of patient outcome. In our study, we found that conventional machine learning was able to predict patient outcome with a range accuracy of 0.69-0.9 at baseline. Other studies, have highlighted the emergent role of post-CRT MR textural features as a successful prognostic indicators [57, 72]. In these settings, our results are in agreement with Horvat et al who found that T2-weighted-based radiomics from post CRT MRI studies were associated with better classification performance compared with qualitative assessment at T2-weighted and DW imaging for diagnosing pathologic complete response in patients with LARC.

In our study, no difference was found regarding the segmentation approach used. To date, nearly all studies published on rectal cancer radiomics have used MS as the technical approach of tumor segmentation. Although, manual delineation is a relatively simple process when carried out by experts, it's a highly subjective approach, which requires maximal human input and time to process. 2D MS has been evaluated with encouraging results. Shen et al. for example [73] compared models performances using 2D vs models using 3D features in non-small cell lung cancer using CT textural features. Although both type of features performed well, they recommend using 2D features. At the opposite, both studies conducted by Yang et al. And Xu et al. were in favor of 3D features [74, 75]. In our study, we did not find any difference between 3D and 2D models both at baseline and after CRT. As such, 2D models could be a quicker alternative.

In the aim to make the technique easier, we evaluated BB as well. Using the same slice of interest, the radiologist could designate the center of the lesion and an algorithm adds a variable number of pixels in order to obtain a surface size of our choice. The advantages of such approach are mainly represented by the gain of time and the incorporation of a peritumoral environment potentially valuable for model's performance with minimal human input [76]. While 2D MS of a rectal tumor could take between 30s and 60s, the 2D BB creation is usually done in less than 10s. As such BBs could be recommended since they are less time consuming, might contain a valuable peri-tumoral environment, and exclude peripheral foggy areas jeopardizing model's performance. Additionally, applying a BB is a valuable option to get through the lack of a standard protocol in terms of FOV size. Indeed, in our sample we had a significant variability in terms of FOV, pixels and voxels sizes. While comparing images obtained with MS versus BBs, we realized that the latter option offers the chance to get rid of the zoom effect related to the FOV variability. To our knowledge, there was no prior study comparing MS to 2D BB models. However, Hosny et al. [76] in their attempt in making deep learning interpretability more explicit applied activation heatmaps at the final convolutional layer of a CNN model trained in a group of lung cancer patients with CT scan images. Heatmaps indicated regions in the input image having the most impact on prediction. They demonstrated that tissue within and surrounding the tumor were both crucial for prediction, highlighting the importance of the peritumoral environment included within the BBs and rising questions regarding the tradition to limit segmentation to the tumoral area.

Lastly, deep learning model failed to predict patient outcome in our study. Deep learning with its apparent simplicity could offer an alternative to conventional machine learning techniques and their lack of standardization. Both deep learning and conventional techniques perform massive textural analyses on images, but without segmentation and feature selection for the former [77]. Deep learning is a new area of research, dealing with development

of deep artificial neural networks, inspired by biological networks in human brain [78]. It has become a very active field of research in computer-assisted radiological support in the last decade, with excellent preliminary results [79-82]. In our study, the CNN model demonstrated hazardous performance. Although deep learning offers the advantage of an apparent simplicity, without the segmentation, features extraction and features selection processes, its analysis process deserves its qualifier of “a black box” and as demonstrated in our study, what might work with a conventional machine learning technique might not work with a deep learning technique. One of the major obstacles for successful translation of DL algorithms from research to practice in precision medicine is their interpretability, as a large number of studies have shown that deep learning can be easily tricked [83-84]. As such, although, a number of studies have attempted to open the deep learning black box [85-86] comprising the major techniques developed to improve network interpretability, their interpretability has been so far and remains a question mark.

Our study has several limitations. First, we included only 98 patients for whom we had all needed radio-clinical data. Second, considering that only one reader segmented the images, it is impossible to assess the reproducibility of segmentation. Third, radiomics was performed only on T2-weighted images to minimize variability in acquisition parameters. However, this segmentation was performed in conjunction with DWI. We anticipate that the performance of radiomics analyses may improve with the inclusion of other MR imaging sequences. Forth, we did not perform external validation. Therefore, further studies are needed to overcome these limitations and to validate the reported data to provide a better generalization of our results.

CONCLUSION

In conclusion, our preliminary study shows that the radiomics measures by using Conventional machine learning were able to predict outcome in patients with locally advanced rectal cancer. The use of BBs is a promising alternative approach to MS which offer the advantages of micro-environment evaluation and minimal human input. Although promising, these results are preliminary and require validation on a larger and independent data set to assess the potential for clinical translation.

References

- 1.Habr-Gama A, Gama-Rodrigues J, Sao Juliao GP, et al: Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control. *Int J Radiat Oncol Biol Phys* 88:822-8, 2014
- 2.Habr-Gama A, Sabbaga J, Gama-Rodrigues J, et al: Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management? *Dis Colon Rectum* 56:1109-17, 2013
- 3.Pacelli F, Sanchez AM, Covino M, et al: Improved outcomes for rectal cancer in the era of preoperative chemoradiation and tailored mesorectal excision: a series of 338 consecutive cases. *Am Surg* 79:151-61, 2013
- 4.Valentini V, Lambin P, Myerson RJ: Is it time for tailored treatment of rectal cancer? From prescribing by consensus to prescribing by numbers. *Radiother Oncol* 102:1-3, 2012

5. Wibe A, Law WL, Fazio V, et al: Tailored rectal cancer treatment--a time for implementing contemporary prognostic factors? *Colorectal Dis* 15:1333-42, 2013
6. Cho H, Kim JE, Kim KP, et al: Phase II Study of Preoperative Capecitabine and Oxaliplatin-based Intensified Chemoradiotherapy With or Without Induction Chemotherapy in Patients With Locally Advanced Rectal Cancer and Synchronous Liver-limited Resectable Metastases. *Am J Clin Oncol* 39:623-629, 2016
7. Rouanet P, Rullier E, Lelong B, et al: Tailored Treatment Strategy for Locally Advanced Rectal Carcinoma Based on the Tumor Response to Induction Chemotherapy: Preliminary Results of the French Phase II Multicenter GRECCAR4 Trial. *Dis Colon Rectum* 60:653-663, 2017
8. van Zoggel D, Bosman SJ, Kusters M, et al: Preliminary results of a cohort study of induction chemotherapy-based treatment for locally recurrent rectal cancer. *Br J Surg*, 2017
9. Bujko K, Wyrwicz L, Rutkowski A, et al: Long-course oxaliplatin-based preoperative chemoradiation versus 5 x 5 Gy and consolidation chemotherapy for cT4 or fixed cT3 rectal cancer: results of a randomized phase III study. *Ann Oncol* 27:834-42, 2016
10. De Felice F, Benevento I, Magnante AL, et al: Clinical benefit of adding oxaliplatin to standard neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a meta-analysis : Oxaliplatin in neoadjuvant treatment for rectal cancer. *BMC Cancer* 17:325, 2017
11. Haddad P, Miraie M, Farhan F, et al: Addition of oxaliplatin to neoadjuvant radiochemotherapy in MRI-defined T3, T4 or N+ rectal cancer: a randomized clinical trial. *Asia Pac J Clin Oncol*, 2017
12. Nougaret S, Fujii S, Addley HC, et al: Neoadjuvant chemotherapy evaluation by MRI volumetry in rectal cancer followed by chemoradiation and total mesorectal excision: Initial experience. *J Magn Reson Imaging* 38:726-32, 2013
13. Patel UB, Taylor F, Blomqvist L, et al: Magnetic resonance imaging-detected tumor response for locally advanced rectal cancer predicts survival outcomes: MERCURY experience. *J Clin Oncol* 29:3753-60, 2011
14. Carbone SF, Pirtoli L, Ricci V, et al: Diffusion-weighted MR volumetry for assessing the response of rectal cancer to combined radiation therapy with chemotherapy. *Radiology* 263:311, 2012
15. Carbone SF, Pirtoli L, Ricci V, et al: Assessment of response to chemoradiation therapy in rectal cancer using MR volumetry based on diffusion-weighted data sets: a preliminary report. *Radiol Med* 117:1112-24, 2012
16. Curvo-Semedo L, Lambregts DM, Maas M, et al: Rectal cancer: assessment of complete response to preoperative combined radiation therapy with chemotherapy--conventional MR volumetry versus diffusion-weighted MR imaging. *Radiology* 260:734-43, 2011

17. Ha HI, Kim AY, Yu CS, et al: Locally advanced rectal cancer: diffusion-weighted MR tumour volumetry and the apparent diffusion coefficient for evaluating complete remission after preoperative chemoradiation therapy. *Eur Radiol* 23:3345-53, 2013
18. Kim YC, Lim JS, Keum KC, et al: Comparison of diffusion-weighted MRI and MR volumetry in the evaluation of early treatment outcomes after preoperative chemoradiotherapy for locally advanced rectal cancer. *J Magn Reson Imaging* 34:570-6, 2011
19. Lambregts DM, Rao SX, Sassen S, et al: MRI and Diffusion-Weighted MRI Volumetry for Identification of Complete Tumor Responders After Preoperative Chemoradiotherapy in Patients With Rectal Cancer: A Bi-institutional Validation Study. *Ann Surg*, 2014
20. Barbaro B, Vitale R, Valentini V, et al: Diffusion-weighted magnetic resonance imaging in monitoring rectal cancer response to neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 83:594-9, 2012
21. Bauerle T, Seyler L, Munter M, et al: Diffusion-weighted imaging in rectal carcinoma patients without and after chemoradiotherapy: a comparative study with histology. *Eur J Radiol* 82:444-52, 2013
22. Boone D, Taylor SA, Halligan S: Diffusion weighted MRI: overview and implications for rectal cancer management. *Colorectal Dis* 15:655-61, 2013
23. Cai G, Xu Y, Zhu J, et al: Diffusion-weighted magnetic resonance imaging for predicting the response of rectal cancer to neoadjuvant concurrent chemoradiation. *World J Gastroenterol* 19:5520-7, 2013
24. Dzik-Jurasz A, Domenig C, George M, et al: Diffusion MRI for prediction of response of rectal cancer to chemoradiation. *Lancet* 360:307-8, 2002
25. Hotker AM, Schmidtman I, Oberholzer K, et al: Dynamic contrast enhanced-MRI in rectal cancer: Inter- and intraobserver reproducibility and the effect of slice selection on pharmacokinetic analysis. *J Magn Reson Imaging* 40:715-22, 2014
26. Intven M, Reerink O, Philippens ME: Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. *J Magn Reson Imaging* 41:1646-53, 2015
27. Shen FU, Lu J, Chen L, et al: Diagnostic value of dynamic contrast-enhanced magnetic resonance imaging in rectal cancer and its correlation with tumor differentiation. *Mol Clin Oncol* 4:500-506, 2016
28. Tong T, Sun Y, Gollub MJ, et al: Dynamic contrast-enhanced MRI: Use in predicting pathological complete response to neoadjuvant chemoradiation in locally advanced rectal cancer. *J Magn Reson Imaging* 42:673-80, 2015
29. Zhao Q, Liu L, Wang Q, et al: Preoperative diagnosis and staging of rectal cancer using diffusion-weighted and water imaging combined with dynamic contrast-enhanced scanning. *Oncol Lett* 8:2734-2740, 2014

30. Hu T, Wang S, Huang L, et al: A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules. *Eur Radiol*, 2018
31. Ortiz-Ramon R, Larroza A, Ruiz-Espana S, et al: Classifying brain metastases by their primary site of origin using a radiomics approach based on texture analysis: a feasibility study. *Eur Radiol*, 2018
32. She Y, Zhang L, Zhu H, et al: The predictive value of CT-based radiomics in differentiating indolent from invasive lung adenocarcinoma in patients with pulmonary nodules. *Eur Radiol*, 2018
33. Tan X, Ma Z, Yan L, et al: Radiomics nomogram outperforms size criteria in discriminating lymph node metastasis in resectable esophageal squamous cell carcinoma. *Eur Radiol*, 2018
34. Wang J, Wu CJ, Bao ML, et al: Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol* 27:4082-4090, 2017
35. Matzner-Lober E, Suehs CM, Dohan A, et al: Thoughts on entering correlated imaging variables into a multivariable model: Application to radiomics and texture analysis. *Diagn Interv Imaging* 99:269-270, 2018
36. Soyer P: Agreement and observer variability. *Diagn Interv Imaging* 99:53-54, 2018
37. Bonekamp D, Kohl S, Wiesenfarth M, et al: Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. *Radiology*:173064, 2018
38. Kickingeder P, Burth S, Wick A, et al: Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. *Radiology* 280:880-9, 2016
39. Feng Z, Rong P, Cao P, et al: Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. *Eur Radiol* 28:1625-1633, 2018
40. Giganti F, Antunes S, Salerno A, et al: Gastric cancer: texture analysis from multidetector computed tomography as a potential preoperative prognostic biomarker. *Eur Radiol* 27:1831-1839, 2017
41. Kim BR, Kim JH, Ahn SJ, et al: CT prediction of resectability and prognosis in patients with pancreatic ductal adenocarcinoma after neoadjuvant treatment using image findings and texture analysis. *Eur Radiol*, 2018

- 42.Lakhman Y, Veeraraghavan H, Chaim J, et al: Differentiation of Uterine Leiomyosarcoma from Atypical Leiomyoma: Diagnostic Accuracy of Qualitative MR Imaging Features and Feasibility of Texture Analysis. *Eur Radiol* 27:2903-2915, 2017
- 43.Lisson CS, Lisson CG, Flosdorf K, et al: Diagnostic value of MRI-based 3D texture analysis for tissue characterisation and discrimination of low-grade chondrosarcoma from enchondroma: a pilot study. *Eur Radiol* 28:468-477, 2018
- 44.Liu S, Liu S, Ji C, et al: Application of CT texture analysis in predicting histopathological characteristics of gastric cancers. *Eur Radiol* 27:4951-4959, 2017
- 45.Shen Q, Shan Y, Hu Z, et al: Quantitative parameters of CT texture analysis as potential markers for early prediction of spontaneous intracranial hemorrhage enlargement. *Eur Radiol*, 2018
- 46.Wibmer A, Hricak H, Gondo T, et al: Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol* 25:2840-50, 2015
- 47.Hodgdon T, McInnes MD, Schieda N, et al: Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images? *Radiology* 276:787-96, 2015
- 48.Imbriaco M, Cuocolo R: Does Texture Analysis of MR Images of Breast Tumors Help Predict Response to Treatment? *Radiology* 286:421-423, 2018
- 49.Miles KA, Ganeshan B, Griffiths MR, et al: Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology* 250:444-52, 2009
- 50.Ng F, Ganeshan B, Kozarski R, et al: Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology* 266:177-84, 2013
- 51.Ueno Y, Forghani B, Forghani R, et al: Endometrial Carcinoma: MR Imaging-based Texture Model for Preoperative Risk Stratification-A Preliminary Analysis. *Radiology* 284:748-757, 2017
- 52.Kjaer L, Ring P, Thomsen C, et al: Texture analysis in quantitative MR imaging. Tissue characterisation of normal brain and intracranial tumours at 1.5 T. *Acta Radiol* 36:127-35, 1995
- 53.Skogen K, Schulz A, Helseth E, et al: Texture analysis on diffusion tensor imaging: discriminating glioblastoma from single brain metastasis. *Acta Radiol*:284185118780889, 2018
- 54.Li Z, Mao Y, Li H, et al: Differentiating brain metastases from different pathological types of lung cancers using texture analysis of T1 postcontrast MR. *Magn Reson Med* 76:1410-1419, 2016

55. Haider MA, Vosough A, Khalvati F, et al: CT texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging* 17:4, 2017
56. Scrima AT, Lubner MG, Abel EJ, et al: Texture analysis of small renal cell carcinomas at MDCT for predicting relevant histologic and protein biomarkers. *Abdom Radiol (NY)*, 2018
57. Horvat N, Veeraraghavan H, Khan M, et al: MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response after Neoadjuvant Therapy. *Radiology* 287:833-843, 2018
58. Machine Learning methods for Quantitative Radiomic Biomarkers. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL *Sci Rep.* 2015 Aug 17; 5():13087.
59. Bibault J, Housset M, Durdux C, Taieb J, Berger A, Coriat R, Chaussade S, Dousset B, Nordlinger B, Burgun A: Deep learning and Radiomics predict complete response after neoadjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep.* 2018 Nov 12
60. Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue
Jiejie Zhou, Yang Zhang, Kai-Ting Chang, Kyoung Eun Lee, Ouchen Wang, Jiance Li, Yezhi Lin, Zhifang Pan, Peter Chang, Daniel Chow, Meihao Wang, Min-Ying Su
J Magn Reson Imaging. 2019 Nov 1. doi: 10.1002/jmri.26981.
61. Radiomics: the facts and the challenges of image analysis.
Stefania Rizzo, Francesca Botta, Sara Raimondi, Daniela Origgi, Cristiana Fanciullo, Alessio Giuseppe Morganti, and Massimo Bellomi.
Eur Radiol Exp. 2018 Dec, 2:36. Doi: 11.1186/s41747-018-0068-z
62. Computational Radiomics System to Decode the Radiographic Phenotype.
van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL
Cancer Res. 2017 Nov 1; 77(21):e104-e107.
63. SMOTE for high-dimensional class-imbalanced data.
Rok Blagus and Lara Lusa.
BMC Bioinformatics. 2012; 14: 106. Doi: 10.1186/1471-2105-14-106
64. K. Simonyan, A. Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition. Published in ICLR 2014
65. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980. arXiv; 2014.
66. Lovinfosse et al. Delta-radiomics signature predicts treatment outcomes after preoperative chemoradiotherapy and surgery in rectal cancer.
Eur J Nucl Med Mol Imaging. 2018 Mar;45(3):365-375. doi: 10.1007/s00259-017-3855-5.

67. Jeon et al. FDG PET/CT radiomics for predicting the outcome of locally advanced rectal cancer. *Radiat Oncol.* 2019 Mar 12;14(1):43. doi: 10.1186/s13014-019-1246-8.
68. Shayesteh et al. Prediction of Response to Neoadjuvant Chemoradiotherapy by MRI-Based Machine Learning Texture Analysis in Rectal Cancer Patients. *J Gastrointest Cancer.* 2019 Aug 27. doi: 10.1007/s12029-019-00291-0.
69. Zhou et al. Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer. *Ann Surg Oncol.* 2019 Jun;26(6):1676-1684. doi: 10.1245/s10434-019-07300-3. Epub 2019 Mar 18.
70. Cui et al. Radiomics analysis of multiparametric MRI for prediction of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Eur Radiol.* 2019 Mar;29(3):1211-1220. doi: 10.1007/s00330-018-5683-9. Epub 2018 Aug 20.
71. Lucia et al. Prediction of outcome using pretreatment ¹⁸F-FDG PET/CT and MRI radiomics in locally advanced cervical cancer treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging.* 2018 May;45(5):768-786. doi: 10.1007/s00259-017-3898-7. Epub 2017 Dec 9.
72. Liu et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res.* 2017 Dec 1;23(23):7253-7262. doi: 10.1158/1078-0432.CCR-17-1038. Epub 2017 Sep 22.
73. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer : Shen et al. *Transl Oncol.* 2017 Dec; 10(6): 886–894. Published online 2017 Sep 18. doi: 10.1016/j.tranon.2017.08.007
74. Yang et al. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol.* 2019 May;29(5):2196-2206. doi: 10.1007/s00330-018-5770-y. Epub 2018 Dec 6.
75. Xu et al. A multi-organ cancer study of the classification performance using 2D and 3D image features in radiomics analysis. *Phys Med Biol.* 2019 Nov 4;64(21):215009. doi: 10.1088/1361-6560/ab489f.
76. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study : Hosny A et al. *PLoS Med.* 2018 Nov 30;15(11):e1002711. doi: 10.1371/journal.pmed.1002711. eCollection 2018 Nov.
77. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *arXiv:1511.00561.* 2015:1–14.
78. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60–88.

79. Cao C, Liu F, Tan H, et al. Deep learning and its applications in biomedicine. *Genomics Proteomics Bioinformatics*. 2018.
80. Schmidhuber J Deep learning in neural networks: an overview. *Neural Networks*. 2015. [cited 2015 January 01];61:85–117.
81. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. *Comput Vision Pattern Recognition*. 2015;427–436.
82. Brown TB, Mané D, Roy A, et al. Adversarial patch. *arXiv:171209665*. 2017:1–6.
83. Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. *Comput Vision Pattern Recognit*. 2015;427–436.
84. Moosavi-Dezfooli S-M, Fawzi A, Frossard P. Deepfool: a simple and accurate method to fool deep neural networks. *Comput Vision Pattern Recognit*. 2016;2574–2582.
85. Mahendran A, Vedaldi A. Understanding deep image representations by inverting them. *Comput Vision Pattern Recognit*. 2015;5188–5196.
86. Bau D, Zhou B, Khosla A, et al. Network dissection: quantifying interpretability of deep visual representations. *arXiv:170405796*. 2017:1–9.

Supplementary figures and tables

Table 1: Acquisition parameters summary in line with the European Society of Gastrointestinal and Abdominal Radiology Guidelines.

	Slice Thickness	T2 HR Non FS	Perpendicular Axis	Dwi
IRM 1 (n=133)	3 [1.4-4.5]	133	133	107
IRM2 (n=94)	3 [1.4-4.5]	94	94	87

Table 2: Studied population characteristics

Clinical characteristics	
Age (median,range)	60 yrs, [21-88]
Sex	Male : 66 Female : 32
Median from diagnosis to surgery	13.5 months [2 -18]
Median from the end of n-RCT to surgery	6.8 months [3 -10]
Median follow-up period	55 months
pTNM classification :	<ul style="list-style-type: none"> • T4 : 4 • T3 : 9 • T2 : 3 • T1 : 23 • T0 : 23 • Tx : 17 • N0 : 66 • N1 : 32

Figure 1. An example of manual segmentation in a 68-year-old patient with rectal cancer. Image A represents native image. Contouring was drawn to cover the whole rectal tumor on each consecutive slice (3D models), on one slice of interest (2D models) as shown in image B. The segmented tumor is within the green area. Images C and D show the bounding box created from the native image.

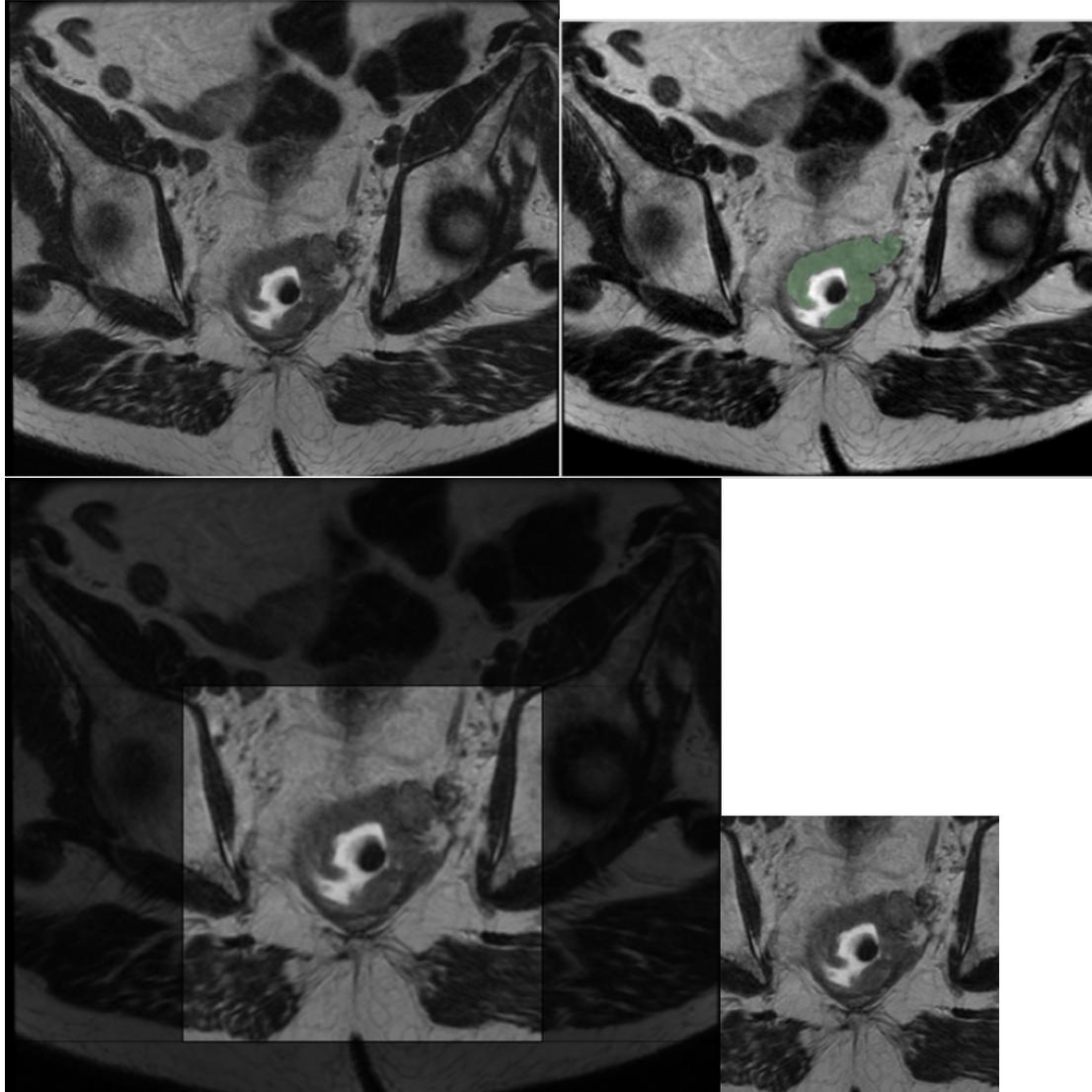


Figure 2. Summary of the analytical setup with an overview of the models being tested in this study.

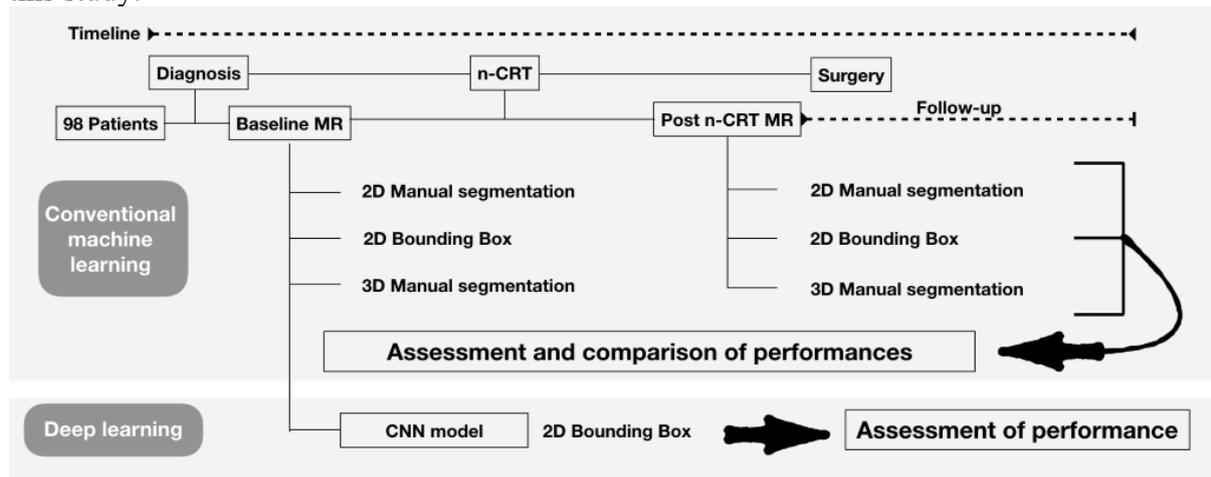


Figure 3. Prognostic power (AUC) of CML using selected MR radiomic features from baseline dataset

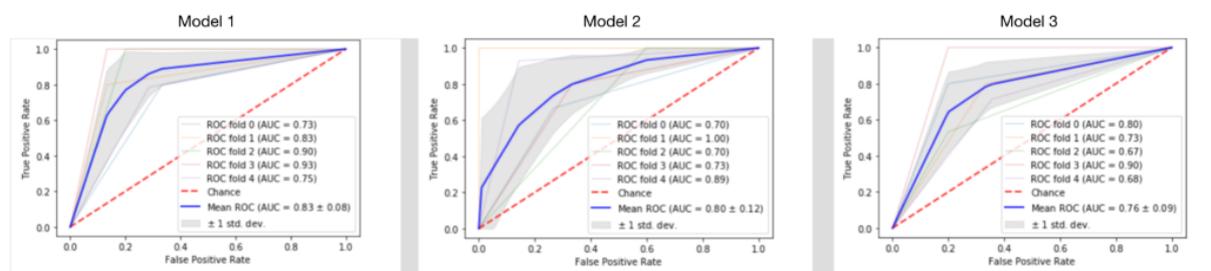


Figure 4. Prognostic power (AUC) of CML using selected MR radiomic features from Post CRT dataset

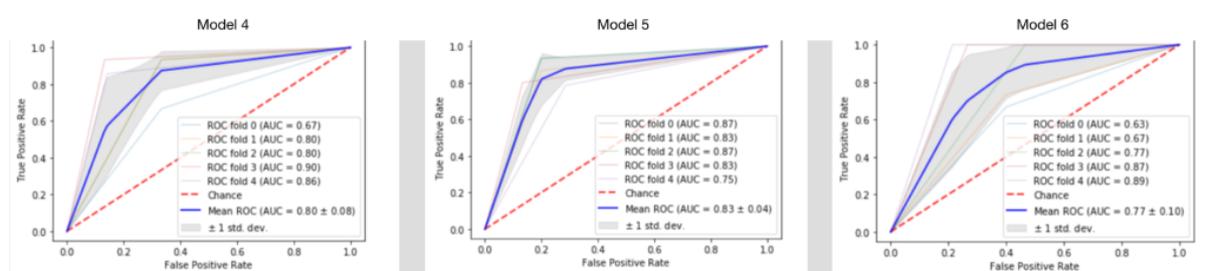


Table 3. Details of selected features for conventional machine learning models at baseline

Models	1	2	3
Number of selected features	28	21	8
Selected features	<ul style="list-style-type: none"> - Logarithm gldm Large Dependence High Gray Level Emphasis - Logarithm gldm Small Dependence Low Gray Level Emphasis - Logarithm glcm Joint Entropy - Logarithm glcm Maximum Probability - Logarithm glcm Joint Energy - Logarithm first order Skewness - Logarithm first order Median - Logarithm first order Total Energy - Logarithm first order Root Mean Squared - Logarithm glrlm Run Length Non Uniformity - Logarithm glrlm Long Run High Gray Level Emphasis - Square gldm Small Dependence Emphasis - Square glcm Difference Entropy Square glcm Imc2 - Square firstorder Skewness - Square firstorder Range - Original gldm Gray Level Non Uniformity - Original gldm Small Dependence Low Gray Level Emphasis - Original glcm Idmn - Original glcm Inverse Variance - Original glcm Correlation - Original firstorder Skewness - Original firstorder Total Energy - Original glrlm Short Run Low Gray Level Emphasis - Original glrlm Gray Level Non Uniformity Original glrlm Long Run High Gray Level Emphasis - Original glrlm Run Length Non Uniformity Normalized - Original glszm Size Zone Non Uniformity 	<ul style="list-style-type: none"> - Logarithm gldm Large Dependence Low Gray Level Emphasis - Logarithm glcm Cluster Shade - Logarithm glcm Idmn - Logarithm glcm Id - Logarithm firstorder Skewness - Logarithm firstorder Total Energy Logarithm firstorder Minimum - Logarithm firstorder Kurtosis - Logarithm glrlm Low Gray Level Run Emphasis - Logarithm glrlm Run Entropy - Logarithm glszm Gray Level Non Uniformity - Logarithm ngtdm Contrast - Logarithm ngtdm Busyness - Square glcm Joint Average - Square glcm MCC - Exponential gldm Dependence Entropy - Exponential gldm Large Dependence Emphasis - Exponential glcm Correlation - lbp-2D first order Variance - Original glcm Correlation - Original glcm MCC - Original firstorder Kurtosis 	<ul style="list-style-type: none"> - Original shape Least Axis Length - Logarithm gldm Dependence Entropy - Logarithm glrlm Gray Level Non Uniformity - Square glcm MCC - Square firstorder Range - lbp-3D-m2 firstorder Mean Absolute Deviation - lbp-3D-k gldm Dependence Entropy - lbp-3D-k glrlm Gray Level Variance

Table 4. Performance details for conventional machine learning models

Models	Number of extracted features	Number of selected features	Accuracy	Sensitivity (%)	Specificity (%)	AUC (95%)	p-value
Baseline							
1	1045	28	0.91	97	89	0.83 +/- 0.08	< 0.001
2	1046	22	0.68	66	64	0.8 +/- 0.12	< 0.001
3	1632	8	0.65	59	73	0.76 +/- 0.09	< 0.001
Presurgery							
4	1045	21	0.73	76	64	0.8 +/- 0.08	< 0.001
5	1046	22	0.81	90	75	0.83 +/- 0.04	< 0.001
6	1632	8	0.74	77	72	0.83 +/- 0.08	< 0.001

Table 5. Details of selected features for conventional machine learning models at pre-surgical assessment

Models	4	5	6
Number of selected features	21	22	8
Selected features	<ul style="list-style-type: none"> - Logarithm glcm Idmn - Logarithm glcm Difference Entropy - Logarithm firstorder 90 Percentile - Logarithm firstorder Minimum - Logarithm firstorder 10 Percentile - Logarithm glrlm Low Gray Level Run Emphasis - Logarithm glrlm Long Run Emphasis - Logarithm glrlm Short Run High Gray Level Emphasis - Logarithm glrlm Long Run LowGray Level Emphasis - Logarithm glszm Large Area High Gray Level Emphasis - Logarithm glszm Small Area Low Gray Level Emphasis - Logarithm ngtdm Contrast - Square glszm Small Area HighGray Level Emphasis - wavelet-LL glcm Joint Average 	<ul style="list-style-type: none"> - Logarithm gldm Large Dependence Low GrayLevel Emphasis - Logarithm glcm joint Average - Logarithm glcm Idmn - Logarithm firstorder Interquartile Range - Logarithm firstorder Skewness - Logarithm firstorder Robust Mean Absolute Deviation - Logarithm firstorder Variance - Logarithm firstorder Kurtosis - Logarithm glszm Gray Level Non Uniformity Normalized - Square gldm Low GrayLevel Emphasis - Square glcm Maximum Probability - Square glcm Idn - Square glszm Low GrayLevel Zone Emphasis - Exponential gldm SmallDependence Emphasis - Exponential glcm Difference Entropy - Exponential firstorder Mean Absolute Deviation - Gradient glcm Correlation - Wavelet-LL ngtdm Contrast 	<ul style="list-style-type: none"> - Gradient glcm Difference Average - Gradient glszm Size Zone Non Uniformity Normalized - Square root firstorder Skewness - Wavelet-HLL glrlm Run Length Non Uniformity Normalized - Wavelet-LHL firstorder Minimum - Wavelet-LHL glszm Zone Percentage - Wavelet-LHH glcm Maximum Probability - Wavelet-LHH glcm Idm

	<ul style="list-style-type: none"> - wavelet-LL firstorder 10Percentile - wavelet-LL glrlm Run Percentage - wavelet-LL glszm Large Area High Gray Level Emphasis - wavelet-LL ngtdm Coarseness - Original glcm Joint Average - Original glcm Sum Average - Original glszm Gray Level Variance 	<ul style="list-style-type: none"> - Original gldm Small Dependence Low Gray Level Emphasis - Original glrlm Long Run Low Gray Level Emphasis - Original glrlm Run Variance - Original glszm Zone Percentage 	
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

SERMENT

➤ *En présence des Maîtres de cette école, de mes chers condisciples et devant l'effigie d'Hippocrate, je promets et je jure, au nom de l'Être suprême, d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la médecine.*

➤ *Je donnerai mes soins gratuits à l'indigent et n'exigerai jamais un salaire au-dessus de mon travail.*

➤ *Admis (e) dans l'intérieur des maisons, mes yeux ne verront pas ce qui s'y passe, ma langue taira les secrets qui me seront confiés, et mon état ne servira pas à corrompre les mœurs, ni à favoriser le crime.*

➤ *Respectueux (se) et reconnaissant (e) envers mes Maîtres, je rendrai à leurs enfants l'instruction que j'ai reçue de leurs pères.*

➤ *Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert (e) d'opprobre et méprisé (e) de mes confrères si j'y manque.*

Résumé de l'article

Objectif : Évaluer et comparer différentes approches d'analyse radiomique dans la prédiction du risque de récurrence chez des patients atteints de tumeurs rectales localement avancées, en utilisant des modèles d'apprentissage conventionnel et d'apprentissage profond basés sur des données IRM en baseline et après radio chimiothérapie néo adjuvante (nRCT).

Méthodes : 98 patients d'une étude prospective, multicentrique, randomisée, en phase II (GRECCAR4-NCT01333709) étaient inclus dans l'étude. Des séquences IRM en pondération T2 étaient utilisées pour l'analyse de 3 modèles radiomiques en baseline et de 3 modèles radiomiques après nRCT, en s'appuyant sur des techniques d'apprentissage conventionnel. Pour chaque groupe de modèles, les données étaient extraites après une segmentation manuelle 2D, une segmentation manuelle 3D et à partir de bounding boxes 2D. Un classifieur de type « random forest », intégrant une technique de sélection de données type « recursive feature elimination » étaient utilisés pour sélectionner les données et construire chacun des modèles. De plus, un modèle d'apprentissage profond, basé sur un réseau de neurones convolutionnels à 16 couches a également été évalué.

Résultats : Les 6 modèles d'apprentissage conventionnel étaient capables de prédire le risque de récurrence avec des aires sous la courbe (AUC) allant de 0.69 à 0.92. La comparaison des performances entre les 3 modèles de baseline et après radio chimiothérapie néo adjuvante n'a pas montré de différence significative. Le modèle d'apprentissage profond a toutefois montré des performances médiocres, avec une précision de 0.51 +/- 0.3. Le Modèle utilisant une bounding box avait une plus grande stabilité à travers les 5 itérations de la validation croisée, avec une AUC de 0.83 +/- 0.04 après nRCT.

Conclusion : Parmi les 3 approches testées en apprentissage conventionnel, les bounding boxes étaient plus simples à obtenir, plus stables en termes de performance et sont plus recommandées à utiliser en pratique clinique.

Mots clés : *CANCER RECTAL; RADIOMIQUES; RECIDIVE; APPRENTISSAGE CONVENTIONNEL; APPRENTISSAGE PROFOND;*