



HAL
open science

Modèles radiomiques hybrides d'apprentissage automatique pour le diagnostic des glioblastomes et métastases cérébrales en IRM

Philippe Naggara

► **To cite this version:**

Philippe Naggara. Modèles radiomiques hybrides d'apprentissage automatique pour le diagnostic des glioblastomes et métastases cérébrales en IRM. Médecine humaine et pathologie. 2021. dumas-03514039

HAL Id: dumas-03514039

<https://dumas.ccsd.cnrs.fr/dumas-03514039v1>

Submitted on 6 Jan 2022

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

UNIVERSITE DE MONTPELLIER
FACULTE DE MEDECINE MONTPELLIER-NIMES

THESE

Pour obtenir le titre de
DOCTEUR EN MEDECINE

Présentée et soutenue publiquement

Par

Philippe NAGGARA

Le 06/12/2021

Modèles radiomiques hybrides d'apprentissage
automatique pour le diagnostic des glioblastomes et
métastases cérébrales en IRM

Directeur de thèse : Pr. Nicolas Menjot de Champfleur

JURY

Présidente : Pr. Ingrid Millet

Assesseur : Pr. Alain Bonafé

Assesseur : Pr. Nicolas Menjot de Champfleur

Assesseur : Pr. Alexandre Krainik

Invitée : Dr. Emmanuelle Le Bars

Invité : Dr. Jérémie Deverdu

UNIVERSITE DE MONTPELLIER
FACULTE DE MEDECINE MONTPELLIER-NIMES

THESE

Pour obtenir le titre de
DOCTEUR EN MEDECINE

Présentée et soutenue publiquement

Par

Philippe NAGGARA

Le 06/12/2021

Modèles radiomiques hybrides d'apprentissage
automatique pour le diagnostic des glioblastomes et
métastases cérébrales en IRM

Directeur de thèse : Pr. Nicolas Menjot de Champfleur

JURY

Président : Pr. Ingrid Millet

Assesseur : Pr. Alain Bonafé

Assesseur : Pr. Nicolas Menjot de Champfleur

Assesseur : Pr. Alexandre Krainik

Invité : Dr. Emmanuelle Le Bars

Invité : Dr. Jérémie Deverduin

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

Professeurs Honoraires

ALBAT Bernard	BRUNEL Michel	HUMEAU Claude	MIRO Luis
ALLIEU Yves	CANAUD Bernard	JAFFIOL Claude	NAVARRO Maurice
ALRIC Robert	CHAPTAL Paul-André	JANBON Charles	NAVRATIL Henri
ARNAUD Bernard	CIURANA Albert-Jean	JANBON François	OTHONIEL Jacques
ASENCIO Gérard	CLOT Jacques	JARRY Daniel	PAGES Michel
ASTRUC Jacques	COSTA Pierre	JOURDAN Jacques	PEGURET Claude
AUSSILLOUX Charles	D'ATHIS Françoise	KLEIN Bernard	PELISSIER Jacques
AVEROUS Michel	DEMAILLE Jacques	LAFFARGUE François	PETIT Pierre
AYRAL Guy	DESCOMPS Bernard	LALLEMANT Jean Gabriel	POUGET Régis
BAILLAT Xavier	DIMEGLIO Alain	LAMARQUE Jean-Louis	PUJOL Henri
BALDET Pierre	DUBOIS Jean Bernard	LAPEYRIE Henri	RABISCHONG Pierre
BALDY-MOULINIER Michel	DUJOLS Pierre	LEROUX Jean-Louis	RAMUZ Michel
BALMES Jean-Louis	DUMAS Robert	LESBROS Daniel	REBOUL Jean
BANSARD Nicole	DUMAZER Romain	LOPEZ François Michel	RIEU Daniel
BAYLET René	ECHENNE Bernard	LORIOT Jean	ROCHEFORT Henri
BILLIARD Michel	FABRE Serge	LOUBATIERES Marie Madeleine	ROUANET DE VIGNE LAVIT Jean Pierre
BLARD Jean-Marie	FREREBEAU Philippe	MAGNAN DE BORNIER Bernard	SAINT AUBERT Bernard
BLAYAC Jean Pierre	GALIFER René Benoît	MARTY ANE Charles	SANCHO-GARNIER Hélène
BLOTMAN Francis	GODLEWSKI Guilhem	MARY Henri	SANY Jacques
BONNEL François	GRASSET Daniel	MATHIEU-DAUDE Pierre	SEGNARBIEUX François
BOURGEOIS Jean-Marie	GUILHOU Jean-Jacques	MEYNADIER Jean	SENAC Jean-Paul
BOUSQUET Jean	GUITER Pierre	MICHEL François-Bernard	SERRE Arlette
BRUEL Jean Michel	HEDON berbard	MION Charles	SOLASSOL Claude
BUREAU Jean-Paul	HERTAULT Jean	MION Henri	VIDAL Jacques
			VISIER Jean Pierre



Professeurs Emérites

ARTUS Jean-Claude	LE QUELLEC Alain
BLANC François	MARES Pierre
BONAFE Alain	MAUDELONDE Thierry
BOULENGER Jean-Philippe	MAURY Michèle
BOURREL Gérard	MESSNER Patrick
BRINGER Jacques	MILLAT Bertrand
CLAUSTRES Mireille	MONNIER Louis
DAURES Jean-Pierre	MOURAD Georges
DAUZAT Michel	PREFAUT Christian
DAVY Jean-Marc	PUJOL Rémy
DEDET Jean-Pierre	RIBSTEIN Jean
ELEDJAM Jean-Jacques	SCHVED Jean-François
GROLLEAU RAOUX Robert	SULTAN Charles
GUERRIER Bernard	TOUCHON Jacques
GUILLOT Bernard	UZIEL Alain
JONQUET Olivier	VOISIN Michel
LANDAIS Paul	ZANCA Michel
LARREY Dominique	

Docteurs Emérites

PRAT Dominique
PUJOL Joseph

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

Professeurs des Universités - Praticiens Hospitaliers

PU-PH de classe exceptionnelle

ALRIC Pierre	Chirurgie vasculaire ; médecine vasculaire (option chirurgie vasculaire)
BACCINO Eric	Médecine légale et droit de la santé
BASTIEN Patrick	Parasitologie et mycologie
BEREGI Jean-Paul	Radiologie et imagerie médicale
BLAIN Hubert	Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
BOULOT Pierre	Gynécologie-obstétrique ; gynécologie médicale
CAPDEVILA Xavier	Anesthésiologie-réanimation et médecine péri-opératoire
CHAMMAS Michel	Chirurgie orthopédique et traumatologique
COLSON Pascal	Anesthésiologie-réanimation et médecine péri-opératoire
COMBE Bernard	Rhumatologie
COSTES Valérie	Anatomie et cytologie pathologiques
COTTALORDA Jérôme	Chirurgie infantile
COUBES Philippe	Neurochirurgie
COURTET Philippe	Psychiatrie d'adultes ; addictologie
CRAMPETTE Louis	Oto-rhino-laryngologie
CRISTOL Jean Paul	Biochimie et biologie moléculaire
CYTEVAL Catherine	Radiologie et imagerie médicale
DE LA COUSSAYE Jean Emmanuel	Médecine d'urgence
DE WAZIERES Benoît	Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
DELAPORTE Eric	Maladies infectieuses ; maladies tropicales
DEMOLY Pascal	Pneumologie ; addictologie
DOMERGUE Jacques	Chirurgie viscérale et digestive

DUFFAU Hugues	Neurochirurgie
ELIAOU Jean François	Immunologie
FABRE Jean Michel	Chirurgie viscérale et digestive
FRAPIER Jean-Marc	Chirurgie thoracique et cardiovasculaire
HAMAMAH Samir	Biologie et Médecine du développement et de la reproduction ; gynécologie médicale
HERISSON Christian	Médecine physique et de réadaptation
JABER Samir	Anesthésiologie-réanimation et médecine péri-opératoire
JEANDEL Claude	Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
JORGENSEN Christian	Thérapeutique ; médecine d'urgence ; addictologie
KOTZKI Pierre Olivier	Biophysique et médecine nucléaire
LABAUGE Pierre	Neurologie
LEFRANT Jean-Yves	Anesthésiologie-réanimation et médecine péri-opératoire
LEHMANN Sylvain	Biochimie et biologie moléculaire
LUMBROSO Serge	Biochimie et Biologie moléculaire
MERCIER Jacques	Physiologie
MEUNIER Laurent	Dermato-vénéréologie
MONDAIN Michel	Oto-rhino-laryngologie
MORIN Denis	Pédiatrie
PAGEAUX Georges-Philippe	Gastroentérologie ; hépatologie ; addictologie
PUJOL Pascal	Biologie cellulaire
QUERE Isabelle	Chirurgie vasculaire ; médecine vasculaire (option médecine vasculaire)
RENARD Eric	Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
REYNES Jacques	Maladies infectieuses, maladies tropicales
RIPART Jacques	Anesthésiologie-réanimation et médecine péri-opératoire
ROUANET Philippe	Cancérologie ; radiothérapie
SOTTO Albert	Maladies infectieuses ; maladies tropicales
TAOUREL Patrice	Radiologie et imagerie médicale

TOUITOU Isabelle

Génétique

VANDE PERRE Philippe

Bactériologie-virologie ; hygiène hospitalière

YCHOU Marc

Cancérologie ; radiothérapie

PU-PH de 1^{er} classe

AGUILAR MARTINEZ Patricia

Hématologie ; transfusion

ASSENAT Éric

Gastroentérologie ; hépatologie ; addictologie

AVIGNON Antoine

Nutrition

AZRIA David

Cancérologie ; radiothérapie

BAGHDADLI Amaria

Pédopsychiatrie ; addictologie

BLANC Pierre

Gastroentérologie ; hépatologie ; addictologie

BORIE Frédéric

Chirurgie viscérale et digestive

BOURDIN Arnaud

Pneumologie ; addictologie

CAMBONIE Gilles

Pédiatrie

CAMU William

Neurologie

CANOVAS François

Anatomie

CAPTIER Guillaume

Anatomie

CARTRON Guillaume

Hématologie ; transfusion

CAYLA Guillaume

Cardiologie

CHANQUES Gérald

Anesthésiologie-réanimation et médecine péri-opératoire

CORBEAU Pierre

Immunologie

COULET Bertrand

Chirurgie orthopédique et traumatologique

CUVILLON Philippe

Anesthésiologie-réanimation et médecine péri-opératoire

DADURE Christophe

Anesthésiologie-réanimation et médecine péri-opératoire

DAUVILLIERS Yves

Physiologie

DE TAYRAC Renaud

Gynécologie-obstétrique, gynécologie médicale

DE VOS John

Histologie, embryologie et cytogénétique

DEMARIA Roland	Chirurgie thoracique et cardio-vasculaire
DEREURE Olivier	Dermatologie - vénéréologie
DROUPY Stéphane	Urologie
DUCROS Anne	Neurologie
DUPEYRON Arnaud	Médecine physique et de réadaptation
FESLER Pierre	Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie
GARREL Renaud	Oto-rhino-laryngologie
GENEVIEVE David	Génétique
GUILLAUME Sébastien	Psychiatrie d'adultes ; addictologie
GUIU Boris	Radiologie et imagerie médicale
HAYOT Maurice	Physiologie
HOUEDE Nadine	Cancérologie ; radiothérapie
KLOUCHE Kada	Médecine intensive-réanimation
KOENIG Michel	Génétique
KOUYOUMDJIAN Pascal	Chirurgie orthopédique et traumatologique
LAFFONT Isabelle	Médecine physique et de réadaptation
LAVABRE-BERTRAND Thierry	Histologie, embryologie et cytogénétique
LAVIGNE Jean-Philippe	Bactériologie-virologie ; hygiène hospitalière
LE MOING Vincent	Maladies infectieuses ; maladies tropicales
LECLERCQ Florence	Cardiologie
MARIANO-GOULART Denis	Biophysique et médecine nucléaire
MATECKI Stéfan	Physiologie
MORANNE Olivier	Néphrologie
MOREL Jacques	Rhumatologie
NAVARRO Francis	Chirurgie viscérale et digestive
NOCCA David	Chirurgie viscérale et digestive
PASQUIE Jean-Luc	Cardiologie



PERNEY Pascal

Médecine interne ; gériatrie et biologie du vieillissement, médecine générale, addictologie

PRUDHOMME Michel

Anatomie

PUJOL Jean Louis

Pneumologie ; addictologie

PURPER-OUAKIL Diane

Pédopsychiatrie ; addictologie

ROGER Pascal

Anatomie et cytologie pathologiques

TRAN Tu-Anh

Pédiatrie

VERNHET Héléène

Radiologie et imagerie médicale

PU-PH de 2ème classe

BOURGIER Céline

Cancérologie; radiothérapie

CANAUD Ludovic

Chirurgie vasculaire ; médecine vasculaire (option chirurgie vasculaire)

CAPDEVIELLE Delphine

Psychiatrie d'Adultes ; addictologie

CLARET Pierre-Géraud

Médecine d'urgence

COLOMBO Pierre-Emmanuel

Cancérologie ; radiothérapie

COSTALAT Vincent

Radiologie et imagerie médicale

DAIEN Vincent

Ophthalmologie

DORANDEU Anne

Médecine légale et droit de la santé

FAILLIE Jean-Luc

Pharmacologie fondamentale ; pharmacologie clinique ; addictologie

FUCHS Florent

Gynécologie-obstétrique ; gynécologie médicale

GABELLE DELOUSTAL Audrey

Neurologie

GAUJOUX Viala Cécile

Rhumatologie

GODREUIL Sylvain

Bactériologie-virologie ; hygiène hospitalière

GUILPAIN Philippe

Médecine Interne, gériatrie et biologie du vieillissement; addictologie

HERLIN Christian

Chirurgie plastique, reconstructrice et esthétique, brûlologie

IMMEDIATO DAIEN Claire

Rhumatologie

JACOT William

Cancérologie ; Radiothérapie

JEZIORSKI Eric

Pédiatrie



JUNG Boris	Médecine Intensive-réanimation
KALFA Nicolas	Chirurgie Infantile
LACHAUD Laurence	Parasitologie et mycologie
LALLEMANT Benjamin	Oto-rhino-laryngologie
LE QUINTREC DONNETTE Mogle	Néphrologie
LETOUZEY Vincent	Gynécologie-obstétrique ; gynécologie médicale
LONJON Nicolas	Neurochirurgie
LOPEZ CASTROMAN Jorge	Psychiatrie d'Adultes ; addictologie
LUKAS Cédric	Rhumatologie
MENJOT de CHAMPFLEUR Nicolas	Radiologie et Imagerie médicale
MILLET Ingrid	Radiologie et Imagerie médicale
MURA Thibault	Biostatistiques, Informatique médicale et technologies de la communication
NAGOT Nicolas	Biostatistiques, Informatique médicale et technologies de la communication
OLIE Emille	Psychiatrie d'adultes; addictologie
PANARO Fabrizio	Chirurgie viscérale et digestive
PARIS Françoise	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
PELLESTOR Franck	Histologie, embryologie et cytogénétique
PEREZ MARTIN Antonia	Chirurgie vasculaire ; médecine vasculaire (option médecine vasculaire)
POUDEROUX Philippe	Gastroentérologie ; hépatologie ; addictologie
RIGAU Valérie	Anatomie et cytologie pathologiques
RIVIER François	Pédiatrie
ROSSI Jean François	Hématologie ; transfusion
ROUBILLE François	Cardiologie
SEBBANE Mustapha	Médecine d'urgence
SIRVENT Nicolas	Pédiatrie
SOLASSOL Jérôme	Biologie cellulaire
STOEBNER Pierre	Dermato-vénérologie



SULTAN Ariane	Nutrition
THOUVENOT Éric	Neurologie
THURET Rodolphe	Urologie
TUAILLON Edouard	Bactériologie-virologie; hygiène hospitalière
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)

LAOUDJ CHENIVESSE Dalila (Biochimie et biologie moléculaire)

VISIER Laurent (Sociologie, démographie)

PROFESSEURS DES UNIVERSITES - Médecine générale

1^{re} classe :

LAMBERT Philippe

AMOUYAL Michel

PROFESSEURS ASSOCIES - Médecine Générale

CLARY Bernard

DAVID Michel

GARCIA Marc

PROFESSEURS ASSOCIES - Médecine



BESSIS Didier (Dermato-vénéréologie)

MEUNIER Isabelle (Ophtalmologie)

MULLER Laurent (Anesthésiologie-réanimation et médecine péri-opératoire)

PERRIGAULT Pierre-François (Anesthésiologie-réanimation et médecine péri-opératoire)

QUANTIN Xavier (Pneumologie)

ROUBERTIE Agathe (Pédiatrie)

VIEL Eric (Soins palliatifs et traitement de la douleur)

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

Maitres de Conférences des Universités - Praticiens Hospitaliers

MCU-PH Hors classe - Echelon Exceptionnel

RICHARD Bruno	Médecine palliative
SEGONDY Michel	Bactériologie-virologie ; hygiène hospitalière

MCU-PH Hors classe

BADIOU Stéphanie	Biochimie et biologie moléculaire
BOULLE Nathalie	Biologie cellulaire
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
GIANSILY-BLAIZOT Muriel	Hématologie ; transfusion

MCU-PH de 1^{re} classe

BERTRAND Martin	Anatomie
BOUDOUSQ Vincent	Biophysique et médecine nucléaire
BRET Caroline	Hématologie biologique
BROUILLET Sophie	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
COSSEE Mireille	Génétique
GIRARDET-BESSIS Anne	Biochimie et biologie moléculaire
LAVIGNE Géraldine	Hématologie ; transfusion
LESAGE François-Xavier	Médecine et Santé au Travail
MARTRILLE Laurent	Médecine légale et droit de la santé
MATHIEU Olivier	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie



MOUZAT Kévin

Biochimie et biologie moléculaire

PANABIÈRES Catherine

Biologie cellulaire

RAVEL Christophe

Parasitologie et mycologie

SCHUSTER-BECK Iris

Physiologie

STERKERS Yvon

Parasitologie et mycologie

THEVENIN-RENE Céline

Immunologie

MCU-PH de 2^{ème} classe

BERGOUX Anne

Génétique

CHIRIAC Anca

Immunologie

DE JONG Audrey

Anesthésiologie-réanimation et médecine péri-opératoire

DU THANH Aurélie

Dermato-vénérologie

FITENI Frédéric

Cancérologie ; radiothérapie

GOUZI Farès

Physiologie

HERRERO Astrid

Chirurgie viscérale et digestive

HUBERLANT Stéphanie

Gynécologie-obstétrique ; Gynécologie médicale

KUSTER Nils

Biochimie et biologie moléculaire

MAKINSON Alain

Maladies infectieuses, Maladies tropicales

PANTEL Alix

Bactériologie-virologie ; hygiène hospitalière

PERS Yves-Marie

Thérapeutique; addictologie

ROUBILLE Camille

Médecine interne ; gériatrie et biologie du vieillissement ; addictologie

SZABLEWSKY

Anatomie et cytologie pathologiques



Maîtres de Conférences des Universités - Médecine Générale

MCU-MG de 1^{re} classe

COSTA David

OUDE ENGBERINK Agnès

MCU-MG de 2^{ème} classe

FOLCO-LOGNOS Béatrice

CARBONNEL François

Maîtres de Conférences associés - Médecine Générale

CAMPAGNAC Jérôme

LOPEZ Antonio

MILLION Elodie

PAVAGEAU Sylvain

REBOUL Marie-Catherine

SERAYET Philippe

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

Maitres de Conférences des Universités

Maitres de Conférences hors classe

BADIA Eric	Sciences biologiques fondamentales et cliniques
CHAZAL Nathalie	Biologie cellulaire

Maitres de Conférences de classe normale

BECAMEL Carine	Neurosciences
BERNEX Florence	Physiologie
CHAUMONT-DUBEL Séverine	Sciences du médicament et des autres produits de santé
DELABY Constance	Biochimie et biologie moléculaire
GUGLIELMI Laurence	Sciences biologiques fondamentales et cliniques
HENRY Laurent	Sciences biologiques fondamentales et cliniques
HERBET Guillaume	Neurosciences
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
RAYNAUD Fabrice	Sciences du Médicament et autres produits de santé
TAULAN Magali	Biologie Cellulaire

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

Praticiens Hospitaliers Universitaires

BARATEAU Lucie	Physiologie
BASTIDE Sophie	Epidémiologie, économie de la santé et prévention
CAZAUBON Yoann	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
DAGNEAUX Louis	Chirurgie orthopédique et traumatologique
DUFLOS Claire	Biostatistiques, informatique médicale et technologies de la communication
GOULABCHAND Radjiv	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
LATTUCA Benoit	Cardiologie
MARIA Alexandre	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
MIOT Stéphanie	Médecine interne ; gériatrie et biologie du vieillissement ; addictologie
SARRABAY Guillaume	Génétique
SOUICHE François-Régis	Chirurgie viscérale et digestive

ANNEE UNIVERSITAIRE 2020 - 2021

PERSONNEL ENSEIGNANT

PH chargés d'enseignements

ABOUKRAT Patrick	BLANCHET Catherine	COROIAN Flavia-Oana	GINIES Patrick
AKKARI Mohamed	BLATIERE Véronique	COUDRAY Sarah	GRECO Frédéric
ALRIC Jérôme	BOBBIA Xavier	CRANSAC Frédéric	GUEDJ Anne Marie
AMEDRO Pascal	BOGE Gudrun	CUNTZ Danielle	GUYON Gaël
AMOUREUX Cyril	BOURRAIN Jean Luc	DARDALHON Brigitte	HENRY Vincent
ANTOINE Valéry	BOUYABRINE Hassan	DAVID Aurore	JAMMET Patrick
ARQUIZAN Caroline	BRINGER-DEUTSCH Sophie	DE BOUTRAY Marie	JEDRYKA François
ATTALIN Vincent	BRINGUIER BRANCHEREAU Sophie	DE LA TRIBONNIÈRE Xavier	JREIGE Riad
AYRIGNAC Xavier	BRISOT Dominique	DEBIEN Blaise	KINNE Mélanie
BADR Malîha	BRONER Jonathan	DELPONT Marion	LABARIAS Coralie
BAIS Céline	CADE Stéphane	DENIS Hélène	LACAMBRE Mathieu
BARBAR Saber Davide	CAIMMI Davide Paolo	DEVILLE de PERIERE Gilles	LANG Philippe
BASSET Didier	CARR Julie	DJANIKIAN Flora	LAZERGES Cyril
BATIFOL Dominique	CARTIER César	DONNADIEU-RIGOLE Hélène	LE GUILLOU Cédric
BATTISTELLA Pascal	CASPER Thierry	FAIDHERBE Jacques	LEGLISE Marie Suzanne
BAUCHET Luc	CASSINOTTO Christophe	FATTON Brigitte	LOPEZ Régis
BENEZECH Jean-Pierre	CATHALA Philippe	FAUCHERRE Vincent	LUQUIENS Amandine
BENNYS Karim	CAZABAN Michel	FILLERON Anne	MANZANERA Cyril
BERNARD Nathalie	CHARBIT Jonathan	FITENI Frédéric	MARGUERITTE Emmanuel
BERTCHANSKY Ivan	CHEVALLIER Thierry	FOURNIER Philippe	MARTIN Lucille
BIBOULET Philippe	CHEVALLIER-MICHAUD Josyane	GAILLARD Nicolas	MATTATIA Laurent
BIRON-ANDREANI Christine	COLIN Olivier	GALMICHE Sophie	MEROUEH Fadi
BLANC Brigitte	CONSEIL Mathieu	GENY Christian	MEYER Pierre
BLANCHARD Sylvie	CORBEAU Catherine	GERONIMI Laetitia	MILESI Christophe

MORAU Estelle	SEGURET Fabienne
MOSER Camille	SENESSE Pierre
MOUSTY Eve	SKALLI EI Medhi
MOUTERDE Gaël	SOLA Christelle
PANSARD Nicole	SOULLIER Camille
PERNIN Vincent	STOEBNER DELBARRE Anne
PERRIGAULT Pierre François	TEOT Luc
PEYRON Pierre-Antoine	THIRION Marina
PICARD Eric	VACHIERY-LAHAYE Florence
PICOT Marie Christine	VERNES Eric
PIERONI Laurence	VINCENT Laure
POQUET Hélène	WAGNER Laurent
PUJOL Sarah-Lise	ZERKOWSKI Laetitia
PUPIER Florence	
QUANTIN Xavier	
RAFFARD Laurence	
RAPIDO Francesca	
RIBRAULT Alice	
RICHAUD-MOREL Brigitte	
RIDOLFO Jérôme	
RIPART Sylvie	
RONGIERES Michel	
ROULET Agnès	
RUBENOVITCH Josh	
SANTONI Fannie	
SASSO Milène	
SCHULDINER Sophie	



MORAU Estelle	SEGURET Fabienne
MOSER Camille	SENESSE Pierre
MOUSTY Eve	SKALLI EI Medhi
MOUTERDE Gaël	SOLA Christelle
PANSARD Nicole	SOULLIER Camille
PERNIN Vincent	STOEBNER DELBARRE Anne
PERRIGAULT Pierre François	TEOT Luc
PEYRON Pierre-Antoine	THIRION Marina
PICARD Eric	VACHIERY-LAHAYE Florence
PICOT Marie Christine	VERNES Eric
PIERONI Laurence	VINCENT Laure
POQUET Hélène	WAGNER Laurent
PUJOL Sarah-Lise	ZERKOWSKI Laetitia
PUPIER Florence	
QUANTIN Xavier	
RAFFARD Laurence	
RAPIDO Francesca	
RIBRAULT Alice	
RICHAUD-MOREL Brigitte	
RIDOLFO Jérôme	
RIPART Sylvie	
RONGIERES Michel	
ROULET Agnès	
RUBENOVITCH Josh	
SANTONI Fannie	
SASSO Milène	
SCHULDINER Sophie	

REMERCIEMENTS

Chers maîtres,

Mme la Professeure Ingrid Millet,

Je vous remercie d'avoir accepté la présidence du jury et vous suis reconnaissant pour vos enseignements et encouragements.

M. le Professeur émérite Alain Bonafé,

C'est un honneur d'avoir pu vous compter parmi les jurés, vous qui avez façonné la neuroradiologie montpelliéraine pour qu'elle rayonne mondialement.

M. le Professeur Alexandre Krainik,

Je vous remercie pour avoir joué un rôle clef pour la concrétisation de ce travail, et d'avoir confirmé mon intérêt pour l'apprentissage automatique via les travaux votre service et de la « french tech » grenobloise.

M. le Professeur Nicolas Menjot de Champfleur,

Je te remercie d'avoir accepté la direction de ce travail et pour les encouragements à persévérer en ce sens. Merci pour les précieux enseignements au fil des années et ton dévouement à la neuroradiologie diagnostique.

Mme la Docteure Emmanuelle Le Bars,

Je vous remercie pour votre participation à ce travail, vos conseils sont une aide bienvenue et motrice pour cette production scientifique.

M. le Docteur Jeremy Deverdun,

Je te remercie de ta participation pour ce travail : les remarques et interrogations techniques ont permis d'affiner et enrichir ce projet.

TABLE DES MATIERES

Table des matières.....	23
Abréviations.....	24
I. Introduction.....	25
1. A propos de la radiomique.....	26
2. Généralités sur les explorations radiomiques en imagerie oncologique.....	28
3. Diagnostic différentiel entre glioblastome et métastase cérébrale en routine clinique.....	30
4. Apport de la radiomique pour le diagnostic différentiel entre glioblastome et métastase cérébrale.....	35
5. Analyse des données radiomiques par apprentissage automatique.....	38
II. Matériel et méthodes.....	78
1. Design, population et critères d'inclusion.....	79
2. Objectifs.....	80
3. Matériel.....	81
4. Prétraitement.....	81
5. Segmentation.....	82
6. Extraction des index radiomiques.....	83
7. Réduction de dimensionnalité.....	83
8. Entraînement et validation du modèle de classification.....	85
9. Entraînement du CNN en parallèle.....	87
10. Performances prédictives des modèles.....	90
11. Autres tests statistiques.....	92
III. Résultats.....	93
1. Population et échantillons.....	94
2. Performances des modèles pour la tâche de classification.....	100
3. Objectifs secondaires.....	151
IV. Discussion.....	168
1. Résultats.....	169
2. Index radiomiques pour le modèle RFE11_SVM.....	173
3. Limites en design.....	182
4. Perspectives.....	183
V. Conclusion.....	186
VI. Bibliographie.....	188
Serment.....	199
Résumé.....	200

ABREVIATIONS

AUC	Area Under the receiver operating Curve
CNN	Convolutional Neural Network
CCC	Coefficient de corrélation de concordance
DISR	Double Input Symmetrical Relevance
DSC	Dice Similarity Coefficient
EGFR	Epidermal Growth Factor Receptor
FBN	Fixed Bin Number
FBW	Fixed Bin Width
FLAIR	Fluid-Attenuated Inversion Recovery
GBM	Glioblastome (multiforme)
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run-Length Matrix
GLSZM	Gray Level Size Zone Matrix
IBSI	Image Biomarker Standardization Initiative
IDH	Isocitrate Déshydrogénase
IRM	Imagerie par résonance magnétique nucléaire
kNN	k-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LR	Logistic Regression
MGMT	6-O-Méthylguanine-ADN méthyltransférase
mRMR	minimum Redundancy Maximum Relevance
NB	Naive Bayes
NGTDM	Neighbouring Gray Tone Difference Matrix
PCA	Principal Component Analysis
ReLU	Rectifier Linear Unit
RF	Random Forest
RFE	Recursive Feature Elimination
ROC	Receiver Operating Curve
SBF	Selection By Filter
SPC	Supervised Principal Component
SVM	Support Vector Machine, Séparateur à vastes marges.
TDM	Tomodensitométrie
TEP	Tomographie par émission de positons

I. INTRODUCTION

1. A propos de la radiomique

Si l'interprétation en radiodiagnostic s'est historiquement fondée sur l'analyse qualitative d'une représentation en images de son objet d'étude, la dernière décennie fut marquée par l'émergence d'un champ de recherche tentant d'identifier de nouveaux biomarqueurs quantitatifs, non immédiatement apparents sur l'imagerie de routine. L'ambition de ces biomarqueurs radiologiques est de décrire avec précision la grande variabilité phénotypique et temporelle des tissus biologiques, normaux ou pathologiques, et ce de manière non invasive. Cette caractérisation permettant hypothétiquement de faire état de la singularité d'un individu ou d'une lésion, elle améliorerait l'adaptation de la prise en charge clinique et thérapeutique, concourant à l'évolution du rationnel médical vers des stratégies personnalisées, notamment en oncologie (1,2).

Dans cette démarche s'inscrit l'essor récent de la radiomique (figure 1), ou imagerie médicale computationnelle, champ d'étude défini par l'exploitation de données quantitatives complexes, extraites de modalités conventionnelles telles que l'imagerie par résonance magnétique nucléaire (IRM), la tomodensitométrie (TDM), l'échographie, la radiographie, ou encore la tomographie par émissions de positons (TEP). Ces données ou index radiomiques décrivent la morphologie tridimensionnelle, la distribution de l'intensité du signal par histogrammes, et les relations spatiales entre les niveaux de gris (index de texture), en complément des métriques traditionnelles (1,2). Ces index sont de haute dimension, se comptant fréquemment par centaines ou milliers.

Cette approche méthodologique s'appuie sur l'hypothèse qu'un sous-ensemble de ces données est susceptible de correspondre statistiquement à un profil biochimique, génomique, histologique etc. constituant alors une signature ou un biomarqueur radiomique. Ce sous-ensemble pourrait également correspondre à une information indépendante supplémentaire, associée par exemple à une survie globale. Cette information permettrait alors d'augmenter les performances d'un score clinique, biologique ou une analyse génomique. La computation de ces index est possible grâce aux outils et matériels informatiques courants, les processeurs ayant bénéficié ces dernières décennies d'une progression exponentielle de leurs capacités opératoires, de concert avec l'essor des processeurs graphiques, permettant un haut débit de calcul à la volée de ces index (1,2).

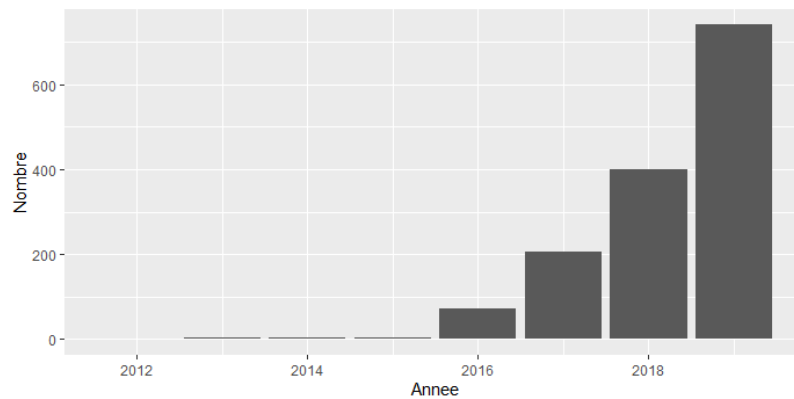


Figure 1. Nombre de publications par année sur le thème de la radiomique. Augmentation exponentielle du nombre de publications avec plus de 700 articles en 2019 contre moins de 100 en 2016. Recherche du terme « *radiomics* » sur la plateforme NIH PubMed en Janvier 2021 (3).

La sélection d'un sous-ensemble d'index est tout d'abord réalisée par différentes méthodes ayant recours à des tests statistiques ou des algorithmes : elle a pour but de réduire la dimension des données, en identifiant les variables susceptibles d'être les plus explicatives. Puis, des algorithmes d'apprentissage automatique (*machine learning*), supervisés ou non supervisés, ou encore d'apprentissage profond (*deep learning*) permettent de générer et d'entraîner des modèles capables de prédire soit la classe des sujets, soit une variable quantitative par régression (1,2). Ces modèles peuvent être purement radiomiques, c'est-à-dire ne prenant en compte que ces index pour la création du modèle dans le cadre d'un processus de découverte : on parle alors de modèle uniquement conduit par les données. Une approche associant d'autres données clinico-biologiques à celles radiomiques est souvent employée dans le cadre d'une considération plus globale du sujet, dans l'hypothèse d'une translation en routine clinique : il s'agit alors d'un modèle mixte ou hybride, construit sur les données et les hypothèses issues de l'observation traditionnelle (1,2). Les performances diagnostiques, prédictives, ou pronostiques des modèles ainsi créés restent alors à valider dans une cohorte indépendante.

Lorsque l'on cherche à corréler une entité biologique à un profil radiomique, l'exploration d'un grand nombre de paramètres requiert également une multiplicité d'observations et de sujets, pour éviter l'écueil d'une association statistique issue du hasard, aussi appelé fléau de la dimension (4). En oncologie, toutes ces observations sont accessibles, le concours de l'imagerie étant systématique ou presque, précédant, accompagnant et suivant le

temps thérapeutique. De plus, ces données sont incluses par défaut dans les protocoles d'imagerie habituels et ne nécessitent donc pas d'acquisitions supplémentaires, encourageant les études rétrospectives (2).

Chronologiquement, le terme *radiomics* est employé par Gillies et al. en 2010 ainsi que par Lambin et al. en 2012, concomitamment au terme *radiogenomics*, qui s'intéresse plus précisément aux corrélations entre les données radiologiques, conventionnelles ou radiomiques, et le génotype exprimé par la tumeur (1,5). *Radiogenomics* pouvant toutefois porter à confusion, puisqu'il était préalablement utilisé pour décrire les analyses génomiques étudiant les variations de radiosensibilité tumorale en radiothérapie (2). Le choix du suffixe *-omics* s'est manifestement fait par analogie avec d'autres disciplines de biologie moléculaire (génomique, protéomique, métabolomique etc.) ayant recours aux données quantitatives complexes (2). Gillies et al. ont également recours au terme d'*anatomie* pour décrire la présentation radiomique singulière d'un tissu ou d'une lésion (5).

2. Généralités sur les explorations radiomiques en imagerie oncologique

En imagerie oncologique, de nombreuses spécialités d'organes explorent les potentiels bénéfiques d'un transfert clinique de cette méthodologie : l'analyse radiomique en IRM prostatique a montré une capacité à la localisation automatique de la lésion de haut grade, mais permet également d'évaluer l'agressivité tumorale et l'hypoxie tissulaire (6).

Concernant la prise en charge des nodules pulmonaires, plusieurs modèles sont prometteurs pour la détection automatique, la stratification du risque, et l'identification d'une histologie maligne (7–9). Plusieurs études ont évalué la capacité à prédire le statut de mutation de l'epidermal growth factor receptor (EGFR) des adénocarcinomes pulmonaires, avec des AUC variant de 0,65 à 0,80 : cette approche pourrait servir à indiquer une thérapeutique ciblée (inhibiteurs de tyrosine kinase), tout en aidant au repérage du site de prélèvement biopsique, réduisant ainsi le nombre de gestes invasifs (10,11).

A propos des néoplasies mammaires, Choi et al. relèvent des valeurs significativement plus élevées d'index de premier ordre (Kurtosis) sur les cartes d'ADC des lésions triples négatives comparativement à celles exprimant un récepteur hormonal, en analyse multivariée

(12). Leithner et al. utilisent un modèle s'appuyant sur des index morphologiques et de texture pour différencier avec une exactitude supérieure à 90% les différents sous types de lésions mammaires en fonction des récepteurs exprimés (13). Li et al. ainsi que Waugh et al. montrent également le potentiel d'un modèle radiomique à différencier les lésions triples négatives des autres sous types moléculaires, en exploitant différents index radiomiques en séquence T1 injectée avec des AUC à 0,67 et 0,75 respectivement (14,15). Zhu et al. observent une corrélation entre des index morphologiques en séquence T1 injectée et certaines activités transcriptionnelles, suggérant que les lésions mammaires de contours irréguliers et mal définis étaient plus agressives (16).

Concernant les tumeurs ovariennes, de nombreuses études ont exploré les liens entre les index radiomiques et la survie, l'hétérogénéité tumorale, le caractère bénin ou malin, le grade tumoral ou encore les corrélations radiogénomiques et transcriptomiques, en scanner et IRM (17). Meier et al. ainsi que Vargas et al. observent qu'une plus grande variabilité inter-sites tumoraux de certains index de texture est associée à une baisse de la survie globale et sans progression, et à une résection chirurgicale incomplète (18,19). Lu et al. identifient 42 index radiomiques permettant de produire un score capable de dépister au diagnostic initial les 5% des sujets qui présenteront la survie globale la plus basse parmi 364 patients (20). De multiples études ont exploré avec succès la capacité de modèles radiomiques conduits par des algorithmes d'apprentissage conventionnels à discriminer les lésions bénignes des lésions borderline ou malignes, ainsi que le grade tumoral d'envahissement en IRM (17). Celles-ci présentent toutes des performances d'exactitude et d'AUC supérieures à 80% (17,21–25).

Au sujet des néoplasies colorectales, plusieurs thèmes d'études ont été explorés avec des données radiomiques scanographiques et IRM, parmi lesquels le statut de mutation KRAS ou la présence d'un envahissement ganglionnaire (26). Huang et al. rapportent chez 326 patients atteints de cancers colorectaux la capacité d'un modèle de régression logistique basé sur 24 index radiomiques en scanner à prédire la présence de métastases ganglionnaires, avec un C-Index à 0,78 (27). Yang et al. proposent un modèle de séparateurs à vastes marges entraîné sur 346 index radiomiques de 117 sujets, se montrant capable de différencier en scanner la présence d'une mutation KRAS, NRAS ou BRAF avec un AUC à 0,83 (28). Golia Pernicka et al. observent la performance d'un modèle Random Forest entraîné sur 254 index radiomiques et 198 sujets à identifier les néoplasies coliques qui présentent un instabilité génomique des microsatellites avec un AUC à 0,79 (29). Avec une cohorte de 494 patients, Liang et al. rapportent un modèle basé sur une régression logistique associant 16 index radiomiques de

premier ordre, permettant de départager les néoplasies colorectales en scanner selon leurs stades localisé I-II ou avancé III-IV avec un AUC à 0,79 (30).

Plusieurs études ont montré un apport de la radiomique concernant les néoplasies tête et cou concernant la capacité à prédire le stade tumoral et l'atteinte ganglionnaire (31). Wang et al. proposent un modèle de séparateurs à vastes marges entraîné sur les données radiomiques scanographiques de 211 patients atteints de carcinomes laryngés, capable de discerner un stade T3 d'un stade T4 avec un AUC de 0,86 (32). Ren et al. font état d'un modèle de régression logistique basé sur 22 index radiomiques IRM de 127 patients, pouvant différencier les stades I-II des stades III-IV des carcinomes épidermoïdes tête et cou, avec un AUC de 0,85 (33). Wu et al. proposent un modèle Random Forest avec sélection des index par une analyse en composantes principales, capable de distinguer les carcinomes épidermoïdes oro-pharyngés bien différenciés de ceux peu différenciés avec une cohorte scanographique de 206 patients (34). Mukherjee et al. observent des performances modérées d'un modèle radiomique par régression logistique régularisée pour identifier le grade tumoral, l'envahissement péri-neural et vasculaire, le statut viral HPV, ainsi que l'extension extra-capsulaire dans les carcinomes épidermoïdes tête et cou parmi 184 sujets, avec des AUC variables entre 0,65 et 0,80 (35).

3. Diagnostic différentiel entre glioblastome et métastase cérébrale en routine clinique

En neuro-oncologie, une problématique récurrente faisant l'objet de ce travail, est celle de la différenciation entre glioblastome (GBM) et métastase cérébrale (MC), dans le but d'adapter la prise en charge du patient au plus tôt. Le glioblastome est la plus fréquente des tumeurs cérébrales nouvellement diagnostiquées, comptant pour 16% d'entre elles (36). Il s'agit de lésions astrocytaires de haut grade (grade IV, classification OMS 2021) qui représentent 50% à 60% de tous les astrocytomes (37). Son pronostic est sombre, avec une médiane de survie globale évaluée à 15 mois, et une survie à 2 ans de l'ordre de 26-33% (36).

Le traitement d'une lésion unique est standardisé, ayant recours à une résection chirurgicale maximale cherchant à épargner les régions fonctionnellement éloquentes. Il est suivi d'un protocole Stupp (38) consistant en une radiothérapie externe et une chimiothérapie concomitante puis adjuvante par alkylants (temozolomide, TMZ) (36). L'adjonction d'une thérapie ciblée par bevacizumab en première ligne a montré une tendance à l'amélioration de la

médiane de survie sans progression, toutefois sans prouver l'amélioration de la survie globale, lors de deux essais de phase III (36,39,40). L'âge, La dégradation du statut clinique et neurologique (Karnofsky Performance Score, Mini Mental State Examination), les faibles marges de résection chirurgicales, l'absence de mutation IDH (isocitrate déshydrogénase), et l'absence de méthylation du promoteur de l'enzyme de réparation de l'ADN 6-O-Méthylguanine-ADN méthyltransférase (MGMT) sont autant de facteurs de mauvais pronostic identifiés (36).

Les métastases cérébrales font également partie des tumeurs cérébrales les plus fréquentes, puisque 6 à 40% de toutes les néoplasies diagnostiquées dissémineront jusqu'à l'étage encéphalique (41,42). Suivant les sources, les métastases encéphaliques représentent 15 à 40% de l'ensemble des tumeurs cérébrales (43,44). Les atteintes primitives les plus pourvoyeuses de métastases cérébrales sont les néoplasies pulmonaires, qui représentent 45% d'entre elles, tandis que les lésions mammaires, mélaniques et colorectales comptent pour 15%, 10% et 5% respectivement (42). Le pronostic est sombre, avec une médiane de survie globale inférieure à un an (42). Une exérèse chirurgicale en bloc sera envisagée lorsqu'une seule lésion accessible est retrouvée, plus ou moins associée à une radiothérapie encéphalique externe. Lorsque les lésions sont multiples, une chimiothérapie systémique reste d'efficacité variable, notamment du fait de la faible diffusion à travers la barrière hémato-encéphalique (45). Le recours à une thérapie ciblée ou immunothérapie dépend des cibles moléculaires exprimées par le primitif, et de l'hétérogénéité moléculaire de la métastase par rapport au primitif (45). Une radiochirurgie stéréotaxique est envisageable lorsque les lésions sont de petite taille, inférieures à 3 - 4 cm, et en faible nombre, jusqu'à 3 lésions, d'autant plus lorsque le primitif est considéré radio-résistant (45). Certaines thérapies ciblant l'expression de mutations sembleraient être plus efficaces, par exemple les inhibiteurs ALK et EGFR pour les lésions pulmonaires concernées, les inhibiteurs BRAF pour les lésions mélaniques ou encore les inhibiteurs HER2 pour les lésions mammaires (42).

En l'absence de caractéristiques pathognomoniques en IRM cérébrale, la démarche diagnostique s'appuie sur un faisceau d'arguments inconstants, faisant état du nombre de lésions, leur localisation, la morphologie du signal T2 FLAIR (fluid-attenuated inversion recovery), la perfusion ou encore la spectroscopie (tableau 1). La multiplicité des lésions cérébrales plaide en faveur d'un caractère métastatique, puisqu'elle est observée chez 70% à 80% des sujets porteurs de cette atteinte secondaire (42,46). A contrario, le caractère multiple des lésions ne représente que 2% à 20% des cohortes de glioblastomes, pour lesquels la taxonomie précisera la forme

multifocale et/ou multicentrique en fonction du nombre d'atteinte par lobe et du nombre de lobes atteints, respectivement (47,48). Les rehaussements T1 inclus dans l'hypersignal FLAIR de la lésion principale sont considérés comme lésions satellites (49).

Tableau 1. Faisceau d'arguments orientant vers un glioblastome ou une métastase cérébrale en IRM multimodale de routine (11-20).

	Glioblastome	Métastase cérébrale
Nombre de lésion	Lésion unique (80-98%).	Lésions multiples (70-80%).
Morphologie FLAIR	Hypersignal FLAIR infiltrant (tumoral)	Hypersignal FLAIR non infiltrant (œdème vasogénique).
Localisations	Cortex (50%), Corps calleux (3%), Noyaux gris centraux (30%).	Jonction cortico-sous-corticale.
Perfusion	↑ rCBV (hypersignal FLAIR péri-rehaussement).	Rupture de barrière hémato-encéphalique.
Spectroscopie	↑ Cho/Cr > 1,24 (hypersignal FLAIR péri-rehaussement).	↑ Lip/Cr (lésion rehaussée).

La morphologie de l'hypersignal T2 FLAIR aux pourtours du rehaussement oriente vers une lésion gliale infiltrante lorsqu'une atteinte corticale (50%), des noyaux gris centraux (30%) ou du corps calleux (3%) est constatée (figure 2) (50–52). Concernant les métastases, l'hypersignal FLAIR est de topographie sous corticale, correspondant à un œdème vasogénique, donnant l'aspect consacré en « doigts de gants » (figure 5).

Concernant la localisation lésionnelle, le rehaussement tumoral est plus fréquemment observé à la jonction cortico-sous-corticale pour les métastases, tandis que le tropisme du glioblastome intéresse la substance blanche sus-tentorielle sous corticale, avec extension périventriculaire, le long des faisceaux tels que les tractus cortico-spinaux, avec un risque de bilatéralisation via une atteinte profonde passant par le corps calleux (53).

L'analyse de perfusion retrouve un volume sanguin cérébral relatif (rCBV) fréquemment plus élevé au sein de l'hypersignal FLAIR non rehaussé pour les glioblastomes, faisant état de l'infiltrat tumoral avec vaisseaux anormaux de la lésion gliale (figure 4) (54). Il est parfois observé un défaut de retour à la ligne de base pour les métastases, traduisant une rupture ou une simple absence de barrière hémato-encéphalique (54).

L'analyse spectroscopique observe une augmentation du pic de choline au sein de l'hypersignal FLAIR péri-rehaussement pour les glioblastomes, avec une valeur seuil de Choline/Créatinine > 1.24 ayant une spécificité de 88,9% (figure 3) (55).

Lorsque la probabilité d'une métastase prévaut, la démarche diagnostique peut s'appuyer sur une ou plusieurs imageries systémiques, à la recherche d'une lésion primitive, parmi lesquels : un scanner cervical et thoraco-abdomino-pelvien, un TEP-scanner au ^{18}F FDG, une mammographie ou encore une échographie thyroïdienne. Ces explorations peuvent être complétées par un examen cutané ou encore une coloscopie. Cette batterie d'examens est susceptible de retarder la prise en charge thérapeutique du patient, avec un risque théorique de perte de chances (56).

Figure 2. Exemple de glioblastome en IRM avec atteinte du corps calleux, sans atteinte corticale. A gauche, séquence T1 injectée, à droite, T2 FLAIR.

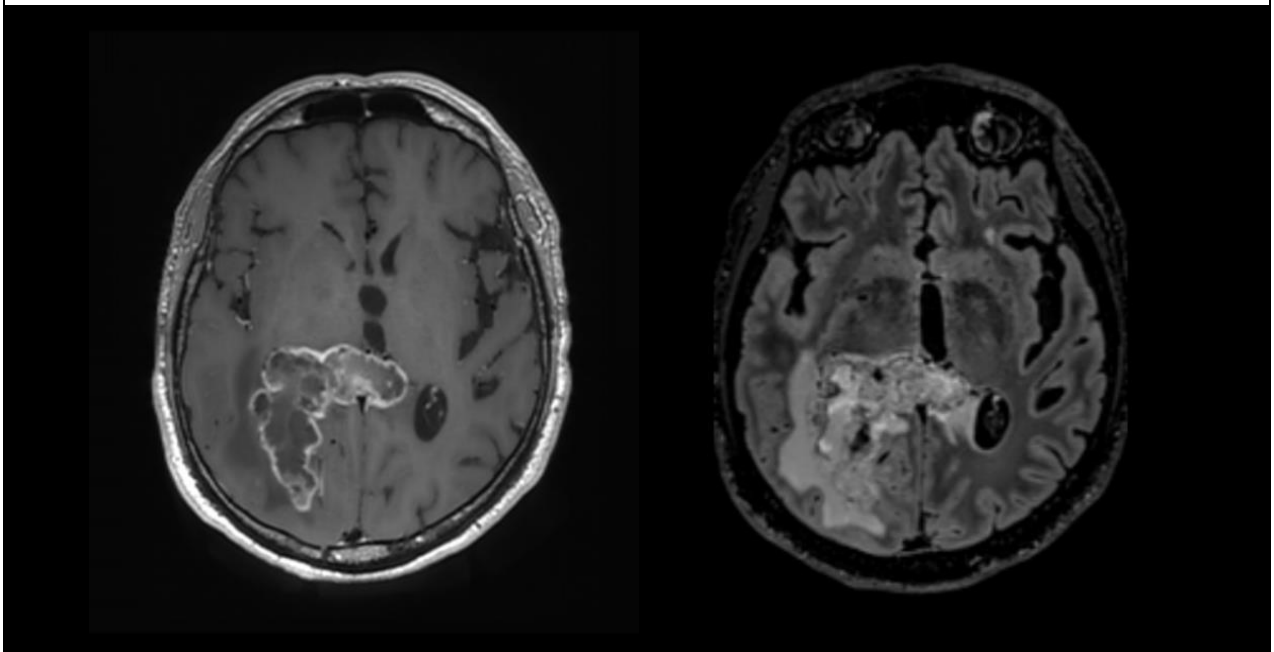


Figure 3. Illustration d'une augmentation du rapport Choline/Créatine dans l'hypersignal FLAIR en spectroscopie, dans le cadre d'une lésion gliale de haut grade.

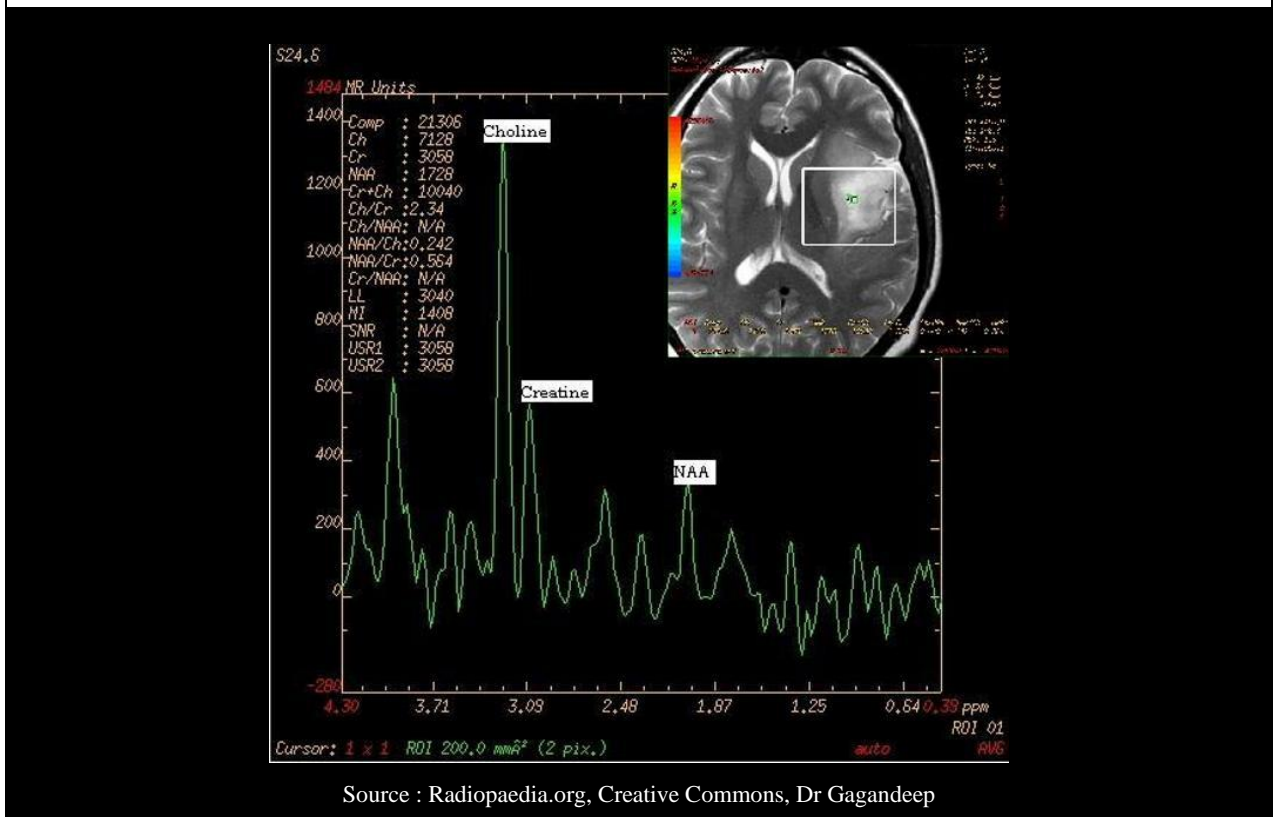


Figure 4. Illustration d'une augmentation du volume cérébral sanguin intra et péri rehaussement, ici dans le cadre d'une lésion gliale de haut grade.

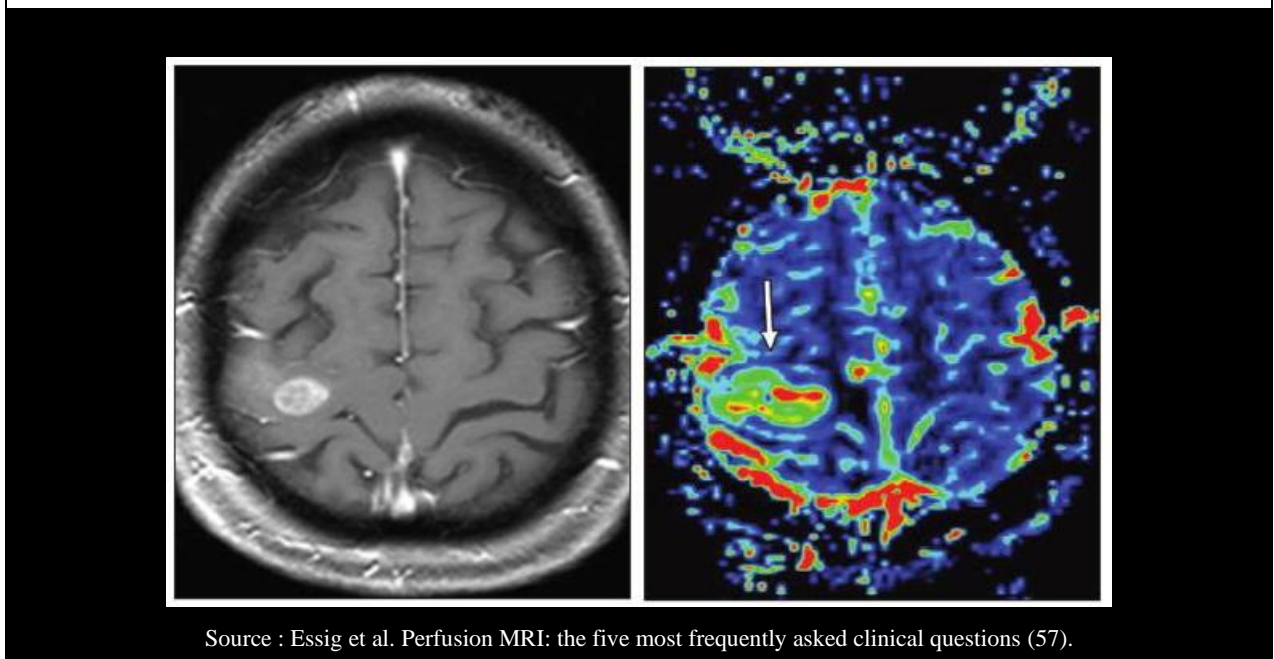
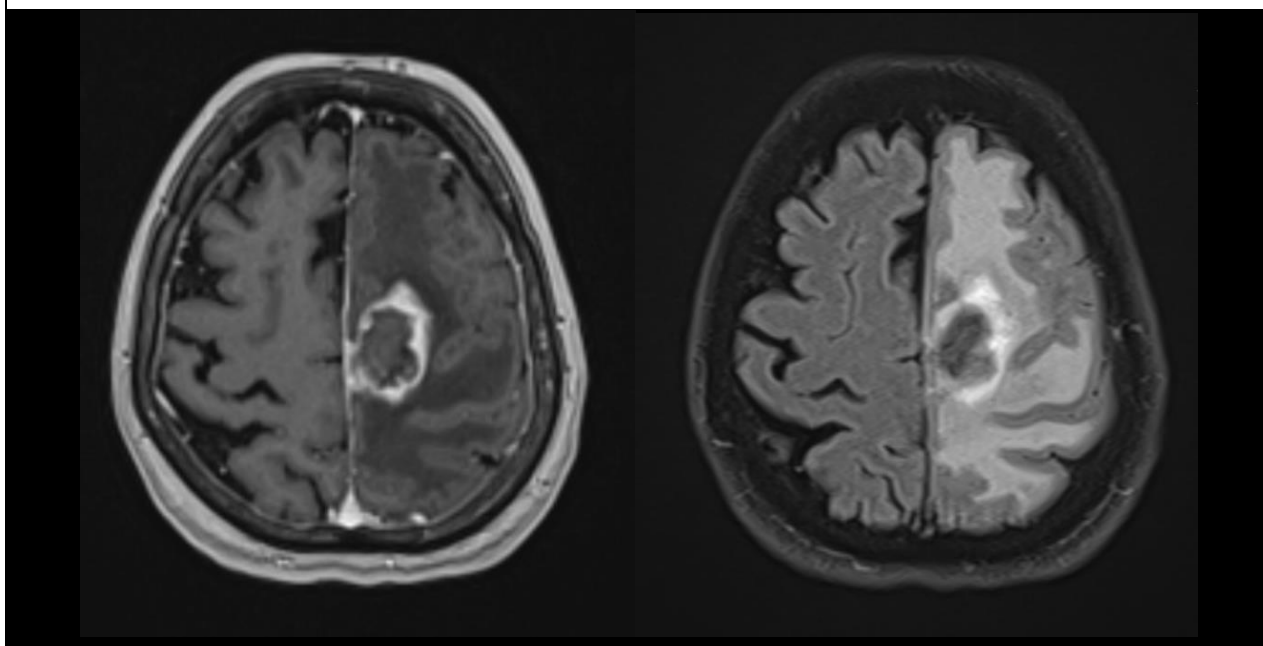


Figure 5. Illustration d'une métastase cérébrale. A gauche, séquence T1 injectée, à droite, séquence T2 FLAIR. Hypersignal FLAIR d'œdème gyriforme en « doigts de gant ».



4. Apport de la radiomique pour le diagnostic différentiel entre glioblastome et métastase cérébrale

Quatre études ont récemment abordé la contribution des données radiomiques et de l'apprentissage automatique à la problématique de distinction entre GBM et MC, résumés dans le tableau 2. Ainsi, Artzi et al. ont montré la capacité d'un modèle d'apprentissage supervisé généré par Support Vector Machine (SVM) à classer glioblastomes et métastases cérébrales sur une séquence T1 injectée conventionnelle, avec un AUC à 0,96 dans une cohorte rétrospective de 439 patients, en utilisant une analyse en composantes principales (PCA) pour la réduction de dimension (58). Qian et al. ont également eu recours à un modèle SVM avec réduction de dimension par régression logistique pénalisée (LASSO), pour différencier ces lésions avec un AUC à 0,90 dans une cohorte rétrospective de 412 patients (59). Han et al. se sont appuyés sur une cohorte de 471 patients, en utilisant un classifieur Random Forest (RF) avec une réduction de dimension par DISR (double input symmetrical relevance), avec un AUC à 0,94 pour cette même question (60). Enfin, Bae et al. proposent plusieurs modèles radiomiques pour cette différenciation entre GBM et MC, notamment avec un réseau de neurones performant avec un

AUC à 0,96 et un algorithme AdaBoost après sélection par arbres décisionnels avec un AUC à 0,89 .

Différents modèles radiomiques se sont montrés capables de prédire une caractéristique génomique, protéomique ou clinique des glioblastomes en IRM. Ainsi, Kong et al. relèvent des profils radiomiques correspondant à des profils transcriptomiques, permettant d'identifier des sous types moléculaires de GBM (classique, proneural, neural, et mésenchymateux), eux-mêmes associés à des pronostics variables (61). Concernant la stratification du risque lors de l'IRM initial, Kickingeder et al. identifient une amélioration de l'évaluation de la survie globale et sans progression avec un modèle sélectionnant 11 index radiomiques basé sur une analyse en composantes principales supervisée (SPC) (62).

Beig et al. montrent le potentiel des index radiomiques à représenter l'hypoxie tissulaire au sein des GBM en fonction de sous régions tumorales, suggérant leur emploi possible pour le monitoring de l'apparition d'une résistance au traitement anti-angiogénique, et par ce biais, d'évaluer la survie (63). Lao et al. proposent un modèle d'apprentissage profond basé sur 6 index radiomiques sélectionnés par LASSO permettant de stratifier le risque concernant la survie globale, avec un C-Index à 0,71 (64). Lee et al. font état du potentiel des index de premier ordre étudiés selon leur distribution spatiale à prédire la survie à 12 mois des glioblastomes avec un AUC à 0,74, ainsi qu'à identifier le statut d'EGFR avec un AUC à 0,85, en se basant sur des séquences T1 injectées et T2 Flair parmi 68 sujets (65).

Hu et al., en s'appuyant sur des index de texture et un modèle d'arbres décisionnels, montrent la possibilité de détecter le statut de proto-oncogènes parmi lesquels EGFR et TP53, en séquences de diffusion et de perfusion, avec une exactitude très variable de 37,5% à 87,5% en analyse univariée, pour 48 biopsies de glioblastomes (66).

La morphologie tumorale diffère entre GBM et MC telle que rapportée par une étude de Yang et al. montrant que deux index morphométriques permettent de classer ces lésions avec une exactitude de 95,8% parmi 48 sujets, en se basant sur un modèle SVM (67). Les auteurs suggèrent que les GBM présentent des contours plus irréguliers en rapport avec une vitesse de prolifération plus élevée et la capacité à s'infiltrer au sein notamment de la substance blanche, tandis que les MC ont généralement une prolifération plus homogène, sphérique (67).

Tableau 2. Etudes abordant la différenciation entre GBM et MC. Ces études ont recours à des modèles radiomiques, des algorithmes de sélection et d'apprentissage automatique en IRM. Selon leurs auteurs la meilleure performance sur échantillon test est rapportée.

Etude	N	Séquences	Sélection	Classifieurs	Performance
Artzi et al. 2019	N = 439 Training 80% 5-fold cross validation	3.0T / 1.5T 3D T1 EG injectée	18 index par PCA	SVM k-NN Arbres Ensemble	SVM : AUC = 0,96 Acc = 0,85
Qian et al. 2019	N = 412 Training 55% 5-fold cross validation	3.0T T1 T2 T1 injectée	Nombre d'index optimal selon méthode. 12 méthodes (LASSO, SBF, Relief, IFGN, GNRO, Euclidian D, F-Anova, WLCX, LR, MIM, SVM, RF)	SVM Adaboost kNN MLP Arbres Random Forest Naive Bayes	LASSO-SVM : AUC = 0,90 Acc = 0,83
Han et al. 2019	N = 471 Training 67% 5-fold cross validation	3.0T T1 injectée	20 premiers index 7 méthodes (DISR, relief, RFS, LCFS, Fisher, cMIM, mRMR)	Random Forest LR SVM Arbres	DISR-RF : AUC = 0,94 Acc = 0,82
Bae et al. 2020	N = 248 Training 67% 10-fold cross validation	3.0T T2 T1 injectée	Nombre d'index optimal selon méthode. 5 méthodes (Fscore, MIM, RFE, LASSO, Arbres)	kNN NB RF AdaBoost Linear SVM Radial SVM LDA ANN	ANN : AUC = 0,96 Acc = 0,89 Arbres-AdaBoost : AUC = 0,89 Acc = 0,83
Abréviations des classifieurs : SVM : Support Vector Machine ; k-NN : k-Nearest Neighbors ; MLP : Multilayer perceptron ; LR : Logistic Regression ; DISR : double input symmetrical relevance ; ANN : Artificial neural network ; LDA : Linear Discriminant Analysis.					

5. Analyse des données radiomiques par apprentissage automatique

L'analyse radiomique se décompose en une série d'étapes qu'il est nécessaire de décrire afin d'évaluer les sources potentielles de biais et de variance, et pour permettre une reproductibilité méthodologique (1,2). Les phases successives sont résumées dans la figure 6 et se déclinent en les points suivants (1,2) :

- (i)** L'acquisition des images et leur prétraitement ;
- (ii)** La segmentation et/ou l'étiquetage des volumes d'intérêt ;
- (iii)** Le prétraitement particulier des volumes d'intérêt ;
- (iv)** L'extraction des index radiomiques ;
- (v)** La sélection ou réduction statistique du nombre de caractéristiques extraites ;
- (vi)** L'apprentissage à la tâche de classification ou régression sur un échantillon d'entraînement ;
- (vii)** L'évaluation du modèle sur un échantillon test non utilisé pour l'entraînement ou la sélection des variables, puis analyse statistique des performances.

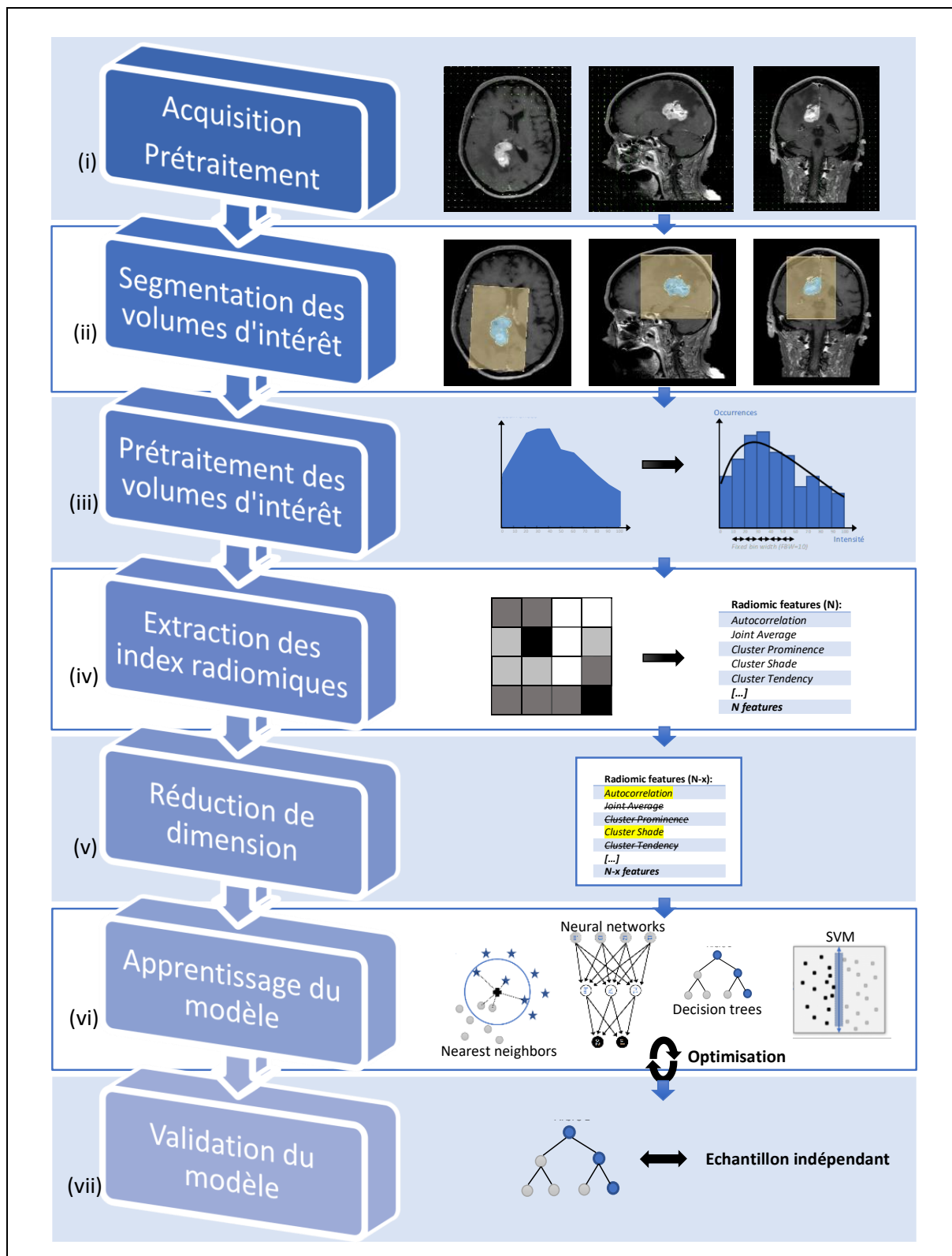


Figure 6. Diagramme de flux des étapes de l'analyse radiomique avec apprentissage automatique.

5.1. Acquisition des images et prétraitement

Pour une présentation radiologique d'une lésion, il existe une variabilité quantitative des données radiomiques qui lui sont associées. De ce fait, il est nécessaire de connaître la variance et l'intervalle des valeurs prises par chacune des mesures et calculs radiomiques. Il faut alors établir des recommandations méthodologiques, concernant le matériel et le protocole d'acquisition ainsi que le traitement des images (68–72). Plusieurs groupes de travail ont été constitués dans une volonté d'harmonisation, parmi lesquels le *Quantitative Imaging Network* (QIN, *National Cancer Institute*, Etats-Unis) (73), le *Quantitative Imaging Biomarkers Alliance* (QIBA, *Radiological Society of North America*) (74), les groupes européens de l'*European Society of Radiology* (ESR) (75), l'*European Organisation for Research and Treatment of Cancer* (EORTC) (76) ou encore l'*Image Biomarker Standardization Initiative* (IBSI) (77).

Toutes les modalités d'acquisition radiologiques de routine sont théoriquement exploitables, le protocole d'acquisition devant être reproductible au sein d'un même centre pour une analyse inter ou intra sujet (1,2). Idéalement, les données brutes, non traitées, non compressées, et sans reconstruction spatiales peuvent être utilisées comme base de travail. Certains prétraitements peuvent être effectués en fonction de la modalité employée. Une optimisation du rapport signal sur bruit est réalisable par différents algorithmes de réduction du bruit, parmi lesquels les filtres médians, les transformées en ondelettes ou le partitionnement par K-moyennes sont les plus employés (78).

La correction des inhomogénéités de champ magnétique B_0 est couramment faite par l'application d'algorithmes tels que N3 ou N4ITK, qui opèrent une normalisation d'intensités par ajustement d'une courbe b-spline par la méthode des moindres carrés (79). En neuro-imagerie, un dépouillement du crâne osseux et du support de tête peut être effectué, en particulier lorsque l'on souhaite réaliser une segmentation automatique par la suite (80). D'autres algorithmes de correction sont ponctuellement appliqués lorsqu'il y a par exemple des artéfacts de susceptibilité magnétique par distorsion du champ en IRM (81), ou des artéfacts de durcissement du faisceau en TDM chez les sujets porteurs de matériel prothétique (82) : il faut cependant se questionner sur la pertinence de l'inclusion de ces examens selon le design choisi de l'étude : en situation réelle ou en processus de découverte.

Lorsqu'il y a plusieurs acquisitions au cours d'un ou plusieurs examens chez le même sujet, on peut être amené à réaliser un recalage des images, fréquemment par une transformation

non linéaire employant une fonction de type affine et/ou b-spline à nombre élevé de degrés de libertés. Le type de transformation s'adaptera aux différents types de séquence IRM employés (spin echo, echo de gradient etc.) et aux risques liés au déroulement de l'acquisition (mouvements du patient, matériel étranger, sortie du tunnel etc.).

Plusieurs algorithmes sont employés, tels que *BRAINSFit ITK* via *3D Slicer* ou encore via le logiciel *SPM12* (83,84). En TDM, les données quantitatives sont normalisées par l'emploi d'une transformation linéaire prenant en compte le coefficient d'atténuation de l'eau et de l'air, ne nécessitant théoriquement pas de normalisation supplémentaire, en dehors d'autres sources d'artéfacts (85). En revanche, les séquences de routine en IRM n'ont pas de valeurs de signal normalisées : selon la méthodologie, en particulier lorsque l'étude est multicentrique avec matériels et protocoles différents, on peut appliquer une loi centrée réduite (Z-score) de la distribution d'intensités du signal, avec cependant un risque de perte d'information.

5.2. Segmentation des volumes d'intérêt

La segmentation manuelle coupe à coupe par un ou plusieurs radiologues experts reste la méthode de référence pour la délimitation du volume d'intérêt. La nécessité d'un grand nombre d'observations en radiomique et le caractère chronophage de la tâche motivent l'emploi de méthodes semi-automatiques ou automatiques. La segmentation semi-automatique permet d'accélérer cette étape, et semble supérieure à la segmentation manuelle en termes de reproductibilité inter opérateurs des paramètres radiomiques (86,87). Elle apparaît également préférable pour la reproductibilité intra opérateur avec une variabilité inférieure du volume et des limites de la lésion pour deux études explorant des cancers pulmonaires (88,89).

Le logiciel *3D Slicer* en accès libre est couramment utilisé, entre autres traitements, pour la segmentation semi-automatique avec l'algorithme *GrowCut* (90). Celui-ci consiste en un étiquetage de la lésion par l'utilisateur sur quelques coupes, associé à un cadre de délimitation (*bounding box*) approximatif au niveau du tissu sain avoisinant les bords lésionnels. L'algorithme associe alors chaque voxel à une catégorie par propagation : lésion, tissu non lésionnel, et tissu indéterminé. Cet algorithme est un automate cellulaire, modèle qui détermine la valeur d'un voxel au temps t en fonction de la valeur du voisinage au temps $t-1$ (figure 7) (90). Après multiples itérations et en faisant appel à une fonction de croissance ou propagation des

régions étiquetées qui sont en compétition, le système atteint un équilibre où les voxels sont labélisés de manière stable, en fonction des voxels initiaux marqués par l'utilisateur (figure 8) (89). Dans une démarche de découverte, il peut être intéressant de segmenter les sous régions lésionnelles en plus d'une segmentation globale, en faisant l'assomption qu'elles peuvent être individuellement plus informatives.

Figure 7. Représentation simplifiée bidimensionnelle d'un automate cellulaire pour une tâche de segmentation telle que réalisée par l'algorithme *GrowCut*.

Initialement, l'utilisateur indique deux régions : le centre approximatif de la lésion et le tissu non lésionnel adjacent. Puis, au temps $t=0$ de l'automate, celui-ci attribue une catégorie à chaque voxel (lésion, non lésion, indéterminé). A $t=1$, la valeur des voxels est confrontée à celle des voxels voisins au temps $t-1$, avec un changement de catégorie lorsqu'une classe voisine devient majoritaire. Ainsi, certains voxels indéterminé (?) à $t=0$ sont classés en lésion ou non lésion au temps $t=1$, tandis que d'autres nécessitent une itération supplémentaire pour lever l'indétermination. Le système atteint un point d'équilibre où les changements de classe deviennent non significatifs ou cycliques : par exemple le voxel marqué d'une étoile (☆) qui est classé bleu à $t=2$ mais restera indéterminé sur les itérations suivantes du fait de l'égalité en nombre des classes des voisins (89,90).

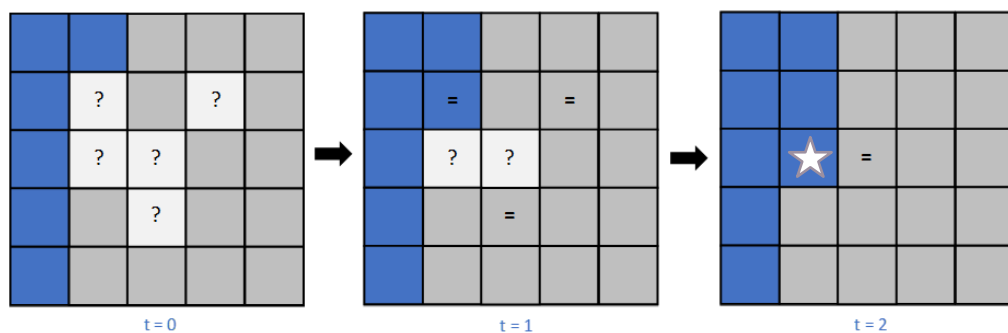
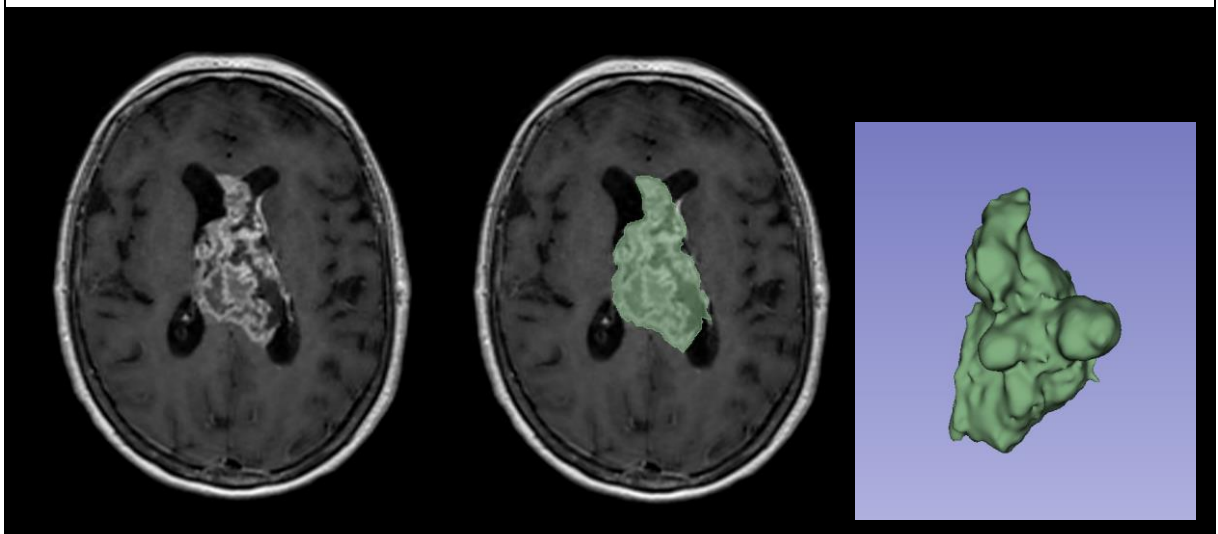
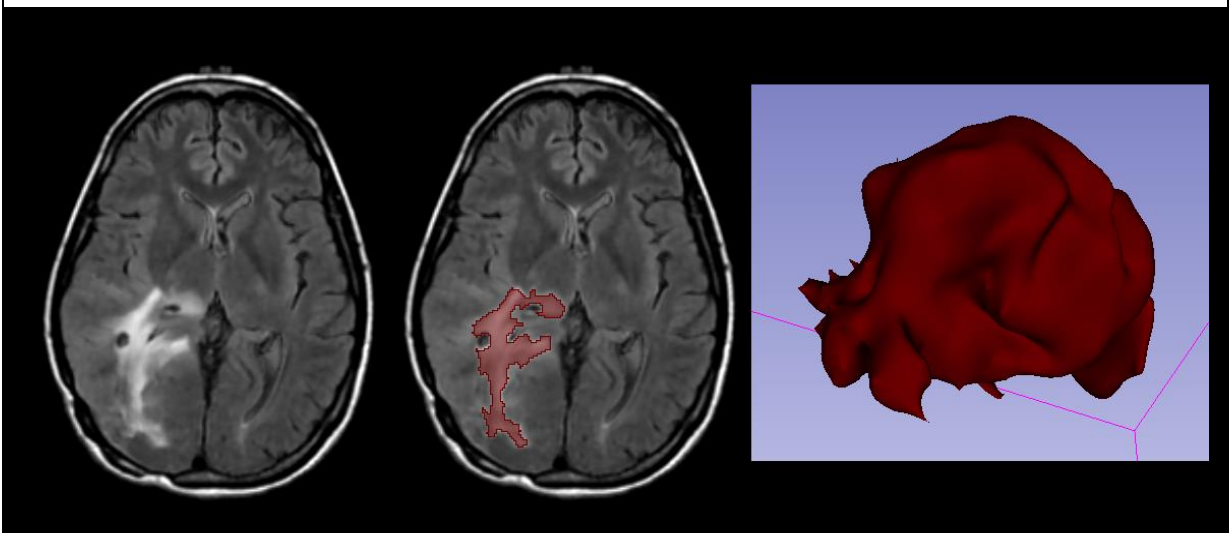


Figure 8. Exemple de segmentation avec algorithme *GrowCut* via *3D Slicer*.



Un autre algorithme de segmentation semi-automatique est le *LevelTracing* qui permet d'ajouter au masque tous les voxels connexes de valeurs proches ou égales. Empiriquement, il apparaît efficace pour la segmentation d'anomalies d'intensité homogène, tel qu'un hypersignal Flair péri-rehaussement d'un GBM ou encore une structure anatomique homogène (figure 9) (91).

Figure 9. Exemple de segmentation semi-automatique par *LevelTracing* sur *3D Slicer*. A gauche, Hypersignal Flair péri-rehaussement d'un GBM. Au milieu, l'algorithme sélectionne les voxels d'intensité proche ou égale du voisinage et les ajoute au masque de segmentation. A droite, rendu 3D à la volée en vue sagittale gauche (91).



5.3. Prétraitement des volumes d'intérêt

5.3.1. Discrétisation de l'intensité

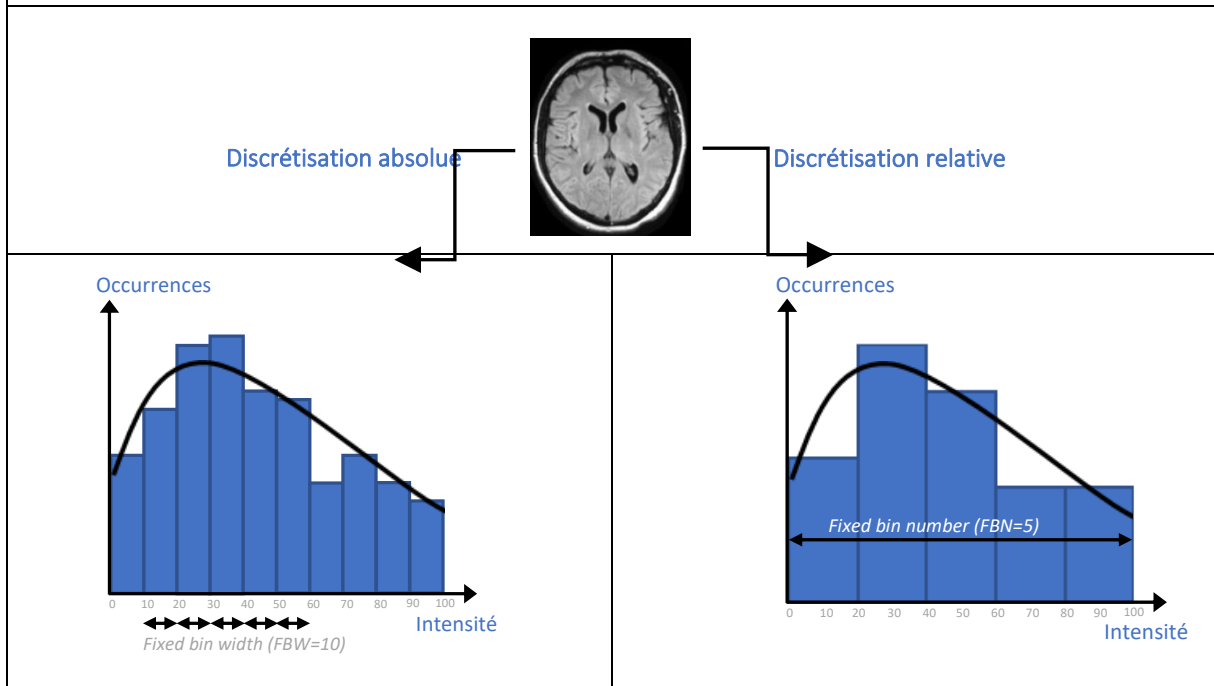
L'analyse des données de deuxième ordre (texture) requiert une étude des différentes matrices de co-occurrence des niveaux de gris (*gray level co occurrence matrix*, GLCM) ainsi que leurs transformées. Pour ce faire, il est nécessaire de discrétiser les valeurs d'intensité des voxels, en les regroupant par intervalles de valeur ou classes (*bins*) d'histogramme. Deux méthodes sont habituellement employées : une relative (*fixed bin number*, FBN), et une absolue (*fixed bin width*, FBW, ou *fixed bin size*, FBS) (68,69). L'approche relative (FBN) correspond à une normalisation de l'intensité du voxel par rapport à l'amplitude total de l'histogramme, multipliée par le rang de l'intervalle de valeur (figure 10) (69). Ainsi, le nombre d'intervalles étant fixe, l'amplitude de chaque intervalle est amené à varier d'un volume d'intérêt à un autre. L'intensité discrétisée s'exprime alors par l'équation (69) :

$$I_{bn}(x) = bn * \frac{I(x) - \min(I(x))}{\max(I(x)) - \min(I(x))}$$

Où :

- $I_{bn}(x)$ représente l'intensité au rang bn pour tout $I(x) > \min(I(x))$;
- $I_{bn}(x)$ vaut 1 lorsque $I(x) = \min(I(x))$;
- bn représente le rang de l'intervalle ;
- $I(x)$ représente l'intensité initiale à discrétiser ;
- $\max(I(x))$ et $\min(I(x))$ représente les bornes supérieures et inférieures des intensités à discrétiser.

Figure 10. Discrétisation de l'intensité en niveaux de gris, par méthode absolue (*fixed bin width, FBW*, schéma de gauche) et méthode relative (*fixed bin number, FBN*, schéma de droite). Pour une même image, avec un profil d'intensités approximé par la courbe, la discrétisation en niveaux de gris par méthode absolue (ici, $FBW=10$ unités d'intensités) produit 10 intervalles de classe, cela pouvant varier suivant l'examen et le type de séquence IRM. Concernant la méthode relative (ici, $FBN=5$ intervalles), on obtient des intervalles ayant une largeur de 20 unités d'intensités, cette valeur pouvant varier suivant l'amplitude totale de l'intensité (69).



Pour l'approche absolue (FBW ou FBS), l'amplitude de chaque intervalle est fixe, la normalisation est réalisée en estimant l'écart entre l'intensité du voxel et l'intensité minimale, rapportées à l'amplitude de l'intervalle (69). La quantité d'intervalles employés peut donc varier d'un volume à l'autre. L'intensité discrétisée s'exprime alors (69) :

$$I_{bs}(x) = \frac{I(x)}{bs} - \min\left(\frac{I(x)}{bs}\right) + 1$$

Où :

- $I_{bs}(x)$ représente l'intensité liée à l'amplitude d'intervalle fixée ;
- $I(x)$ représente l'intensité initiale à discrétiser ;
- bs représente l'amplitude de l'intervalle ;
- $\min(I(x) / bs)$ représente le minimum de l'intensité normalisée par l'amplitude de l'intervalle ;

- + 1 permet d'éviter les intensités négatives.

Il n'existe pas de recommandations pour l'emploi préférentiel d'une méthode par rapport à l'autre. *L'Image Biomarker Standardisation Initiative* (IBSI) rappelle toutefois qu'il est intéressant de multiplier les analyses avec des valeurs variables de FBN ou FBW dans un processus de découverte et d'optimisation (77). Ainsi, il est fréquemment d'usage d'employer des FBN variant entre 1 et 100, couramment 25, et des FBW entre 8 et 1024 (77).

5.3.2. Rééchantillonnage des acquisitions par interpolation (*Resampling*)

En IRM, la résolution spatiale définie par le volume de chaque voxel diffère selon les méthodes d'acquisition choisies. Il faut ainsi renseigner la définition spatiale dans le plan, c'est-à-dire le nombre de pixels par plan via la taille de la matrice d'acquisition, mais aussi l'épaisseur de coupe, et le caractère jointif ou non des coupes. Puisque le nombre et la taille d'un voxel au sein d'un même volume d'intérêt changent selon la séquence, la quantité d'information analysable est appelée à varier. Lorsque l'on souhaite homogénéiser la quantité d'information tout en ayant des voxels reproductibles entre différentes méthodes d'acquisition, on effectue un rééchantillonnage de l'image par interpolation, en transformant la matrice contenant l'information. Ce rééchantillonnage est isotropique, permettant d'obtenir la même définition en rotation et translation dans les trois plans de l'espace. Lorsque l'image est faiblement définie, un sur-échantillonnage (*oversampling*) augmente la quantité de voxels et d'information, par interpolation. A l'inverse, si l'image est fortement définie, un sous échantillonnage (*undersampling*) diminue cette quantité de voxels et d'information (92).

5.4. Calcul et extraction des index radiomiques

Le travail de standardisation réalisé par l'IBSI a permis d'établir une liste de caractéristiques radiomiques reproductibles et faisant consensus, quantifiant différents types d'information, regroupés en familles (77). Cette liste est issue entre autres des travaux d'Aerts et al. (93), Hatt et al (94). Pour ce travail, nous nous appuyons sur les index radiomiques fournis

par *Pyradiomics*, prenant en compte cet effort de standardisation ou proposant l'équivalent IBSI lorsque ce n'est pas le cas (95).

5.4.1. Index morphologiques (*Shape*)

Les index morphologiques font état de la géométrie 2D et 3D du volume d'intérêt. La morphologie est déterminée par des mesures métriques simples, telles que le volume, la surface, les longueurs maximales et minimales (77,95). Elles permettent de quantifier la sphéricité, l'allongement, la platitude de la région d'intérêt (tableau 3). Ces caractéristiques sont fréquemment rapportées en oncologie pulmonaire comme étant prédictives d'une histologie maligne, notamment l'absence de sphéricité témoignant des contours spiculés et irréguliers de lésions primitives (8,9).

Tableau 3. Liste des index morphologiques (*Pyradiomics*)

Index morphologiques :
Mesh Volume
Voxel Volume
Sphericity
Compactness
Spherical Disproportion
Maximum 3D Diameter
Maximum 2D Diameter (axial)
Maximum 2D Diameter (coronal)
Maximum 2D Diameter (sagittal)
Major Axis length
Minor Axis Length
Least Axis Length
Elongation
Flatness

Définitions des index sur <https://pyradiomics.readthedocs.io/>

5.4.2. Index de premier ordre

Les index statistiques de premier ordre décrivent la distribution des valeurs d'intensité au sein du volume d'intérêt, en ayant recours à des données métriques conventionnelles (95). On retrouve des index classiques tels que les valeurs d'intensité minimales, maximales, moyennes et médianes, ainsi que l'écart-type et la variance associés (tableau 4). Ces index ne prennent pas en

considération les relations spatiales entre les différents voxels, contrairement aux index d'ordres supérieurs (95).

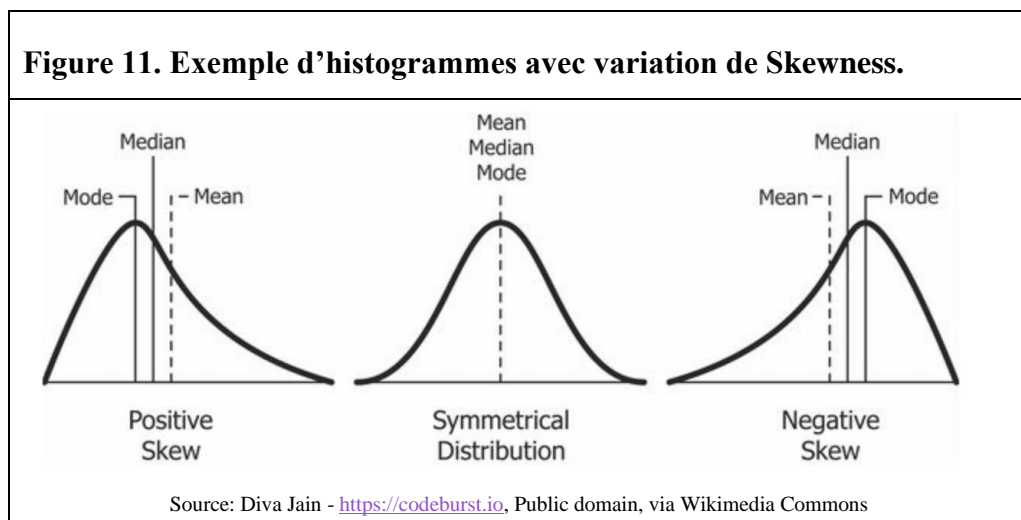
Tableau 4. Liste des index de premier ordre (*Pyradiomics*)

Energy
Total Energy
Entropy
Minimum
10th percentile
90th percentile
Maximum
Mean
Median
Interquartile Range
Range
Mean Absolute Deviation
Robust Mean Absolute Deviation
Root Mean Squared
Standard Deviation
Skewness
Kurtosis
Variance
Uniformity

Définitions des index sur <https://pyradiomics.readthedocs.io/>

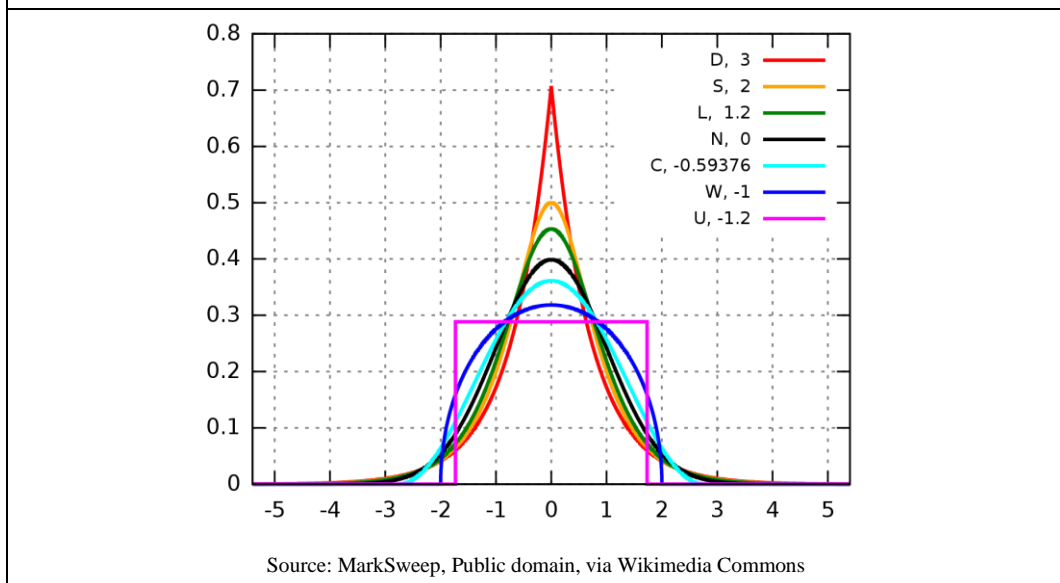
L'entropie de la région d'intérêt permet de représenter l'incertitude, le désordre, ou encore le caractère aléatoire de la distribution des valeurs dans l'image, et s'estime en fonction de la quantité d'information nécessaire pour encoder toutes les valeurs d'une région. Ainsi, en théorie de l'information, l'entropie correspond à la limite de compression possible sans perte des données sources. L'entropie est considérée comme un biomarqueur prometteur représentatif de l'hétérogénéité tumorale, intra et inter-patient (71). L'hétérogénéité tumorale est reconnue comme facteur pronostique et de réponse au traitement, sachant qu'elle est variable au cours du temps, selon les traitements employés ou encore selon son instabilité génomique (96–98). Elle est extensivement explorée en anatomopathologie, biologie cellulaire et moléculaire, ainsi qu'en génétique en exploitant des prélèvements biopsiques. L'avantage de son exploration par imagerie serait d'en rendre compte sans nécessiter des gestes invasifs répétés chez le même patient. L'entropie a été associée à la réponse thérapeutique, la survie, le métabolisme ou encore le stade tumoral notamment pour les lésions pulmonaires, colo-rectales, œsophagiennes ou ORL (71,99–102).

Skewness représente l'asymétrie de la distribution des valeurs d'intensité avec une médiane significativement différente de la moyenne, se traduisant par un déplacement à gauche ou à droite de la crête (ou mode) de la distribution par rapport à une distribution normale (figure 11). Le plus grand nombre d'occurrences se retrouve du côté de la valeur médiane, tandis qu'à l'opposé, il est observé un allongement et une augmentation des occurrences de la queue de distribution (95). Une Skewness positive, avec déplacement vers les intensités plus basses, pourraient par exemple correspondre à une tumeur avec important contingent nécrotique, avec en T1 injectée une forte portion en hyposignal T1 qui se situerait à hauteur du mode de distribution, tandis qu'un faible contingent rehaussé et hétérogène serait situé plus proche de la queue de distribution droite, selon un mode de distribution secondaire. Pour une lésion très homogène, la distribution attendue serait normale, avec une faible Skewness.



Kurtosis fait état du regroupement des valeurs d'intensité au niveau de la crête moyenne, ou à l'inverse autour des queues de distribution, et traduit donc l'aspect en pic, ou à l'inverse l'aplatissement de la courbe (figure 12) (95). Lorsqu'il y a un aspect en pic avec des queues de distribution marquées, le Kurtosis est majoré et la distribution est dite leptokurtique. Elle est désignée comme platikurtique lorsque le Kurtosis est faible, et la courbe présente un aspect arrondi avec faible queues de distribution et faible pic (95).

Figure 12. Exemple d’histogrammes de différentes lois avec variation de Kurtosis. Kurtosis maximal pour une loi de Laplace (D, 3), minimale pour une loi uniforme continue (U, -1,2). Loi Normale (N, 0).



Kurtosis et Skewness, en illustrant la dispersion par rapport à la valeur moyenne, peuvent être considérés comme complémentaires de l’entropie pour constater une hétérogénéité tumorale. Au total, une lésion hétérogène présentera une majoration de l’entropie (majoration de la quantité d’information), une diminution relative du Kurtosis (aspect aplati de la distribution), une valeur négative ou positive de Skewness (déplacement à gauche ou à droite de la crête de la distribution). A l’inverse, une lésion homogène présentera une faible entropie, un Kurtosis relativement majoré, et un Skewness nul ou presque.

5.4.3. Index de deuxième ordre, troisième ordre, et d’ordres supérieurs (index de texture)

Les index de deuxième ordre, troisième ordre et d’ordres supérieurs, également désignés par le terme d’index de texture, font état de la distribution spatiale des intensités des voxels étudiés au sein d’une région. Les index de deuxième et troisième ordre sont calculés par l’intermédiaire de matrices qui quantifient chacune un type de relation spatiale, telle qu’une série de valeurs identiques dans une direction de l’espace. Les index d’ordres supérieurs requièrent

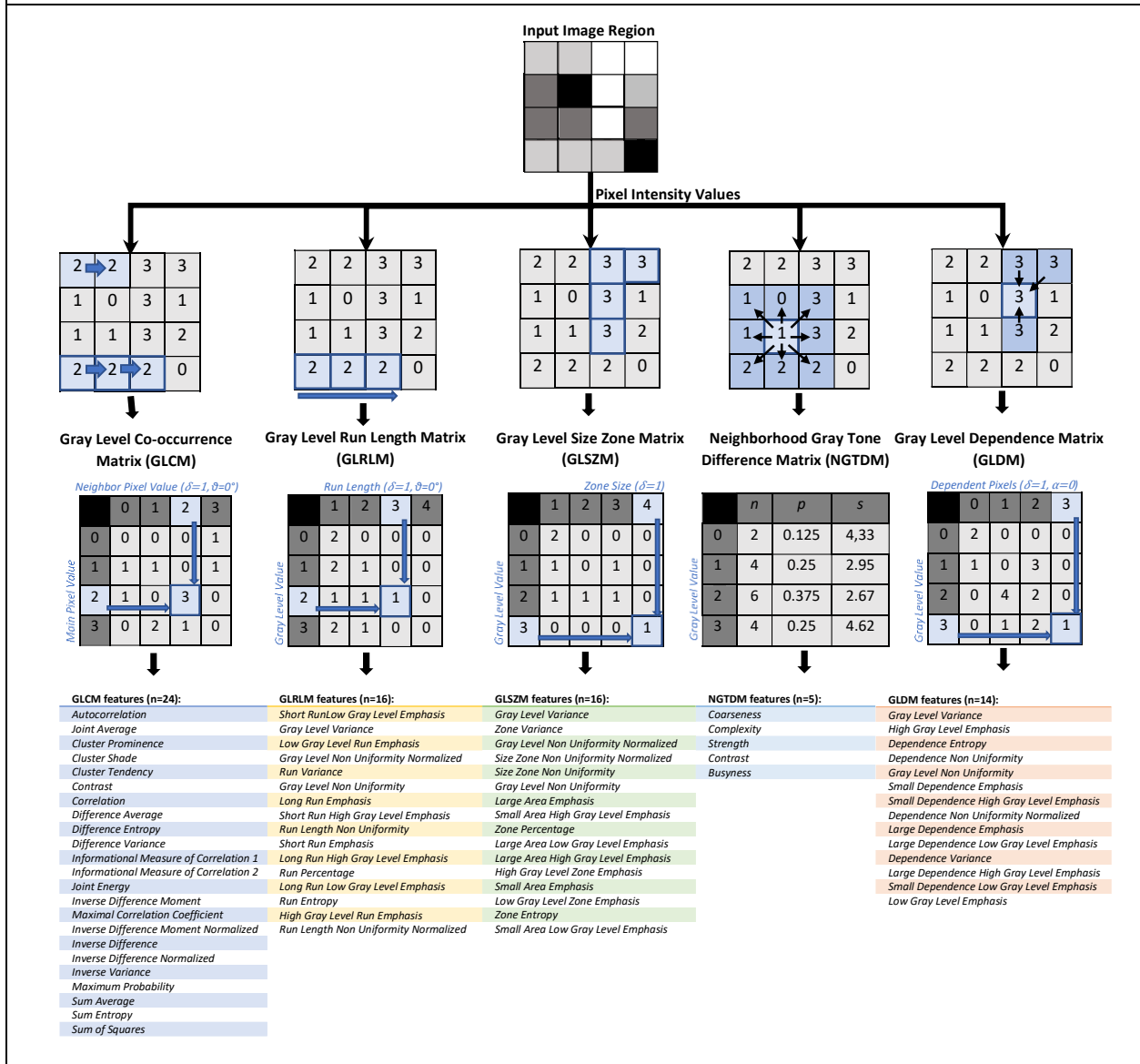
une transformation, les plus employées étant la transformée en ondelettes, ou la transformée de Laplace d'une gaussienne (77,95).

En ayant recours à *Pyradiomics*, cinq matrices sont générées pour calculer les index de deuxième et troisième ordre, illustrées dans la figure 13 (95) :

- i. **La matrice de co-occurrence des niveaux de gris** (*Gray Level Co-occurrence Matrix*, GLCM) exprime la distribution tridimensionnelle de combinaisons d'intensités de voxels voisins. Cette distribution est mesurée dans chacune des directions de l'espace, correspondant à 13 angles en 3D (connectivité 26), ou 4 angles en 2D (connectivité 8). Ainsi, l'élément (i,j) de la GLCM renseigne le nombre d'occurrence de l'association d'intensités i et j de voxels éloignés d'une distance $\delta=1$, dans l'une des 13 directions définie par l'angle θ (95).
- ii. **La matrice de longueur de série de niveaux de gris homogènes** (*Gray Level Run-Length Matrix*, GLRLM) quantifie les séries de voxels consécutifs (distance $\delta=1$) qui présentent le même niveau de gris dans une direction. L'élément (i,j) représente le nombre de séries de niveau de gris i et de longueur j, dans l'une des 13 directions définie par l'angle θ du volume (95).
- iii. **La matrice de taille de zone de niveaux de gris homogènes** (*Gray Level Size Zone Matrix*, GLSZM) quantifie également les séries de voxels consécutifs de même niveau de gris, mais indépendamment de la direction d'angle θ : une seule matrice est calculée pour les 13 directions de l'espace. Ainsi, l'élément (i,j) représente le nombre de zones de niveau de gris i et de taille j (95).
- iv. **La matrice de différence de tons de gris des voxels du voisinage** (*Neighbouring Gray Tone Difference Matrix*, NGTDM) quantifie la moyenne des différences absolues de niveau de gris de chaque voxel par rapport à ses voxels voisins d'une distance δ (95). Puis ces moyennes sont sommées (s) en fonction de chaque valeur d'intensité du voxel de référence. Cette somme pourra être pondérée par la probabilité p d'occurrence de chaque valeur d'intensité.
- v. **La matrice de dépendance de niveaux de gris** (*Gray Level Dependence Matrix*, GLDM) quantifie le nombre de voxels à une distance δ qui présentent un niveau

de gris de $\pm \alpha$ par rapport au voxel interrogé. L'élément (i,j) représente le nombre de voxels de niveau de gris i qui ont j voxels de niveaux $\pm \alpha$ au sein de l'image (95).

Figure 13. Diagramme de flux du calcul des matrices et des index de deuxième et troisième ordre.



L'ensemble des index radiomiques de différents ordres est ensuite colligé sous la forme d'un vecteur. En utilisant *Pyradiomics*, ce vecteur sera composé de 18 index de premier ordre, 24 index de deuxième ordre, 51 index de troisième ordre et 14 index de forme 2D ou 3D pour une région d'intérêt et une modalité d'imagerie, soit un total de 107 index (95). A ceux-ci, il faut

ajouter les valeurs transformées de tous ces index, à l'exception des index morphologiques. La transformée en ondelettes applique un filtre de fréquences soit passe haut, soit passe bas, dans chacune des trois directions de l'espace, donnant 8 ensembles de valeurs supplémentaires, soit 744 valeurs (93,95). Le filtre Laplacian of Gaussian (LoG) met l'accent sur les bords et contours, c'est-à-dire les régions avec un changement net de niveaux de gris. Ce filtre peut être paramétré selon que l'on choisisse de mettre l'emphase sur des textures bien définies et limitées (taille de kernel bas) ou sur des textures plus lissées (taille de kernel plus élevé) (95,103). Plusieurs tailles de kernel sont employées, couramment entre 0.5 et 5mm, les valeurs supérieures semblent permettre une meilleure reproductibilité (95,103). Chaque taille de kernel ajoute un ensemble de 93 valeurs. Toutes ces valeurs d'index, originales et transformées, étant explorées pour chaque région et sous-région d'intérêt, ainsi que pour chaque séquence ou type d'acquisition, le vecteur final peut contenir plusieurs milliers d'index.

5.5. Sélection des index radiomiques et réduction de dimensionnalité

Les valeurs quantitatives sont standardisées par Z-score qui consiste à centrer la moyenne arithmétique à 0 et réduire l'écart type à 1, pour permettre une comparaison plus robuste notamment entre patients, tout en imposant un même ordre de grandeur pour toutes les variables :

$$Z - score = \frac{x - \mu}{\sigma}$$

Où :

- x est la valeur mesurée ;
- μ est la moyenne arithmétique observée de l'échantillon ;
- σ est l'écart-type de cet échantillon.

En supposant que la distribution des observations est normale, un Z-score inférieur à -1,96 ou supérieur à 1,96 correspond alors à des valeurs respectivement situées dans les 2,5 premiers ou derniers percentiles de l'échantillon observé.

Ainsi, cette méthode de standardisation, en mesurant la distance d'un point par rapport à la moyenne et en la pondérant par l'écart-type, impose une même dimension pour les différentes

variables. Cela permet d'éviter qu'une grande variance de l'une de ces variables, d'autant plus si elle est d'une plus grande dimension par rapport aux autres, ne prenne un poids disproportionné dans le modèle prédictif (104).

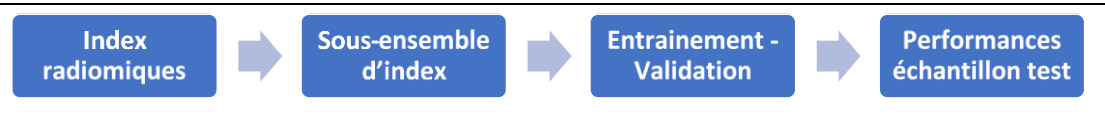
Il est alors possible d'injecter des données complémentaires au sein du vecteur, notamment un critère de jugement, des renseignements cliniques ou encore des informations venant d'autres types d'examen (2). Cette injection peut être avant la réduction statistique du nombre d'index lorsque l'on souhaite découvrir un modèle conduit par les données (*data-driven*), ou après réduction, si l'on souhaite évaluer un modèle hybride.

Etant donné la complexité du vecteur d'index, il existe un risque de sur-apprentissage ou surajustement (*overfitting*) du modèle, c'est-à-dire de sélectionner un sous-ensemble trop complexe d'index corrélés au critère de jugement, uniquement représentatif de l'échantillon d'entraînement, mais ne pouvant être généralisable à la population ou à un autre échantillon (2,105). Le sur-apprentissage mène donc à l'identification erronée d'index faussement informatifs ou faux positifs (105). Pour éviter cet écueil, il est indispensable de réduire la quantité de variables en identifiant les index porteurs d'informations redondantes (2,105). Pour ne pas biaiser les performances du modèle, la recherche des variables les plus informatives ne doit pas comporter l'échantillon de test pour éviter une fuite de données. A posteriori, le surajustement peut être estimé en observant les valeurs d'AUC d'entraînement et de test : il sera recherché de trop bonnes performances en entraînement associées à un résultat médiocre sur l'échantillon test.

Plusieurs méthodes statistiques permettent de réduire la dimensionnalité du vecteur radiomique (105–112) :

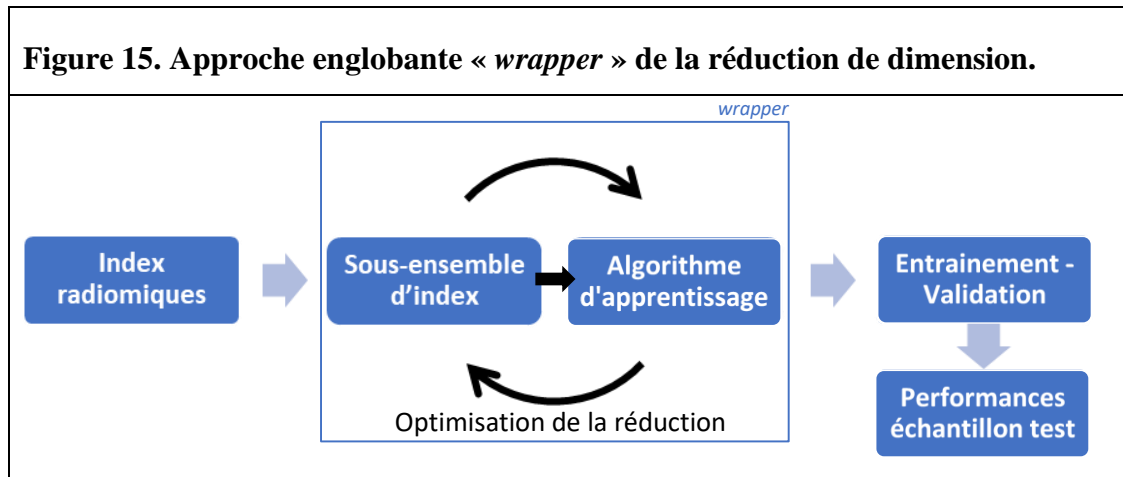
- i. **L'approche « par filtre »** (figure 14) permet de sélectionner les variables en amont et indépendamment du modèle d'apprentissage, en se basant sur les corrélations des index entre eux et avec la variable à prédire. Cette approche a donc recours aux tests statistiques habituels de corrélation (coefficient de Pearson, rangs de Spearman, analyse de variance etc.). Cette méthode est robuste face au risque d'*overfitting* du fait de son indépendance vis-à-vis du modèle d'apprentissage. Une approche fréquemment employée est la méthode *minimum redundancy maximum relevance* (mRMR) qui a recours aux calculs d'une matrice d'information mutuelle, de corrélations linéaires et de distances entre les index, dans le but d'identifier et de pénaliser les index redondants les moins informatifs (109).

Figure 14. Approche par filtre de la réduction de dimension.



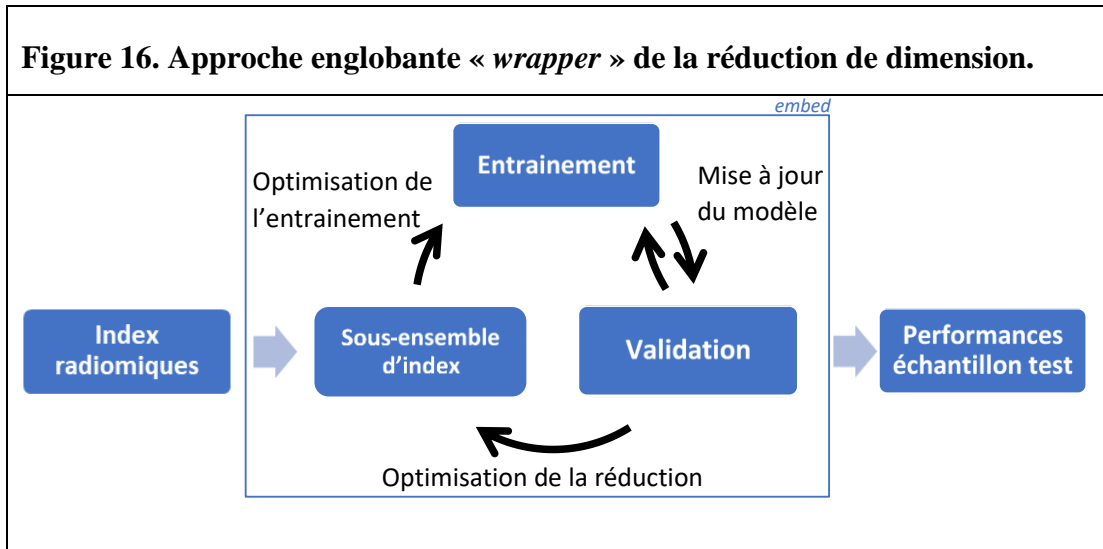
- ii. **L'approche « englobante »** (*wrapper*) (figure 15) sélectionne un ou plusieurs sous-ensembles d'index dont les performances sont itérativement testées par un algorithme d'apprentissage sur les données d'entraînement, par exemple une forêt aléatoire. Cet algorithme d'apprentissage est propre à cette étape de réduction, il est « englobé », et diffère de celui employé en aval pour la tâche de classification même s'il est conseillé qu'il soit de la même famille. Il s'agit donc d'anticiper le comportement du modèle d'apprentissage situé en aval en le confrontant à de multiples sous-ensembles de variables. Cela permet d'identifier un sous-ensemble optimal en fonction des erreurs de classification entre les différents sous-ensembles. Ces jeux de variables sont donc dépendants de la méthode d'apprentissage employée et peuvent de ce fait contribuer à un *overfitting*. Lorsque les données sont de haute dimension, les sous-ensembles à tester deviennent très nombreux et les temps de calcul augmentent significativement : on parle d'algorithme « glouton ». Cette approche ne permet pas une optimisation de la sélection lors de l'entraînement du modèle final (110). Trois approches « englobantes » sont notables : la sélection ascendante (*forward stepwise selection*), l'élimination descendante (*backward stepwise elimination*), et l'une de ses variations, l'élimination descendante récursive de variables (*recursive feature elimination*, RFE). La sélection ascendante consiste à débiter avec un sous-ensemble constitué de la variable la plus explicative. Puis, à chaque cycle, on teste l'association d'une variable à ce sous ensemble : est retenue l'association qui améliore le plus le modèle. Cette sélection est réitérée jusqu'à l'épuisement des variables, et seulement si le modèle continue d'être optimisé. À l'inverse, l'élimination descendante consiste à débiter avec toutes les variables, puis on teste le retrait de chacune d'entre elle pour éliminer la moins informative. On procède jusqu'à la performance maximale de l'algorithme d'apprentissage englobé. Une approche mixte consiste à considérer à chaque étape le retrait ou l'ajout d'une variable, on parle alors de sélection bidirectionnelle. Enfin, l'élimination descendante récursive de variables est fréquemment employée pour ses bonnes performances notamment avec les SVM

(110). Son principe est de partir du modèle maximal avec l'ensemble des variables, d'ajuster le modèle, puis d'éliminer de manière récursive la variable contribuant le moins au modèle. A l'épuisement des variables, celles-ci peuvent alors être classées par rang d'élimination. Il est alors possible de choisir le jeu de variables présentant la meilleure performance, ou encore un compromis entre le plus faible nombre de variables et une performance acceptable (110).



- iii. **L'approche « intégrative »** (*embedded*) (figure 16) prend en compte les performances de classification sur les échantillons d'entraînement du modèle d'apprentissage pour ajuster la sélection d'index et leurs poids. Un exemple est l'approche FRMT pour la prédiction de stades oncologiques sur des données protéomiques (111). L'approche par LASSO (*Least Absolute Shrinkage and Selection Operator*) consiste à réaliser une régression logistique des variables en fonction du critère de jugement, dont les paramètres ont une somme maximale finie. En conséquence, les paramètres des variables les moins informatives sont réduits à zéro, éliminant les variables correspondantes de l'équation. Les paramètres non nuls permettent alors de sélectionner les variables potentiellement explicatives. Cela permet de réduire le surajustement mais ajoute le risque d'éliminer une variable d'intérêt par excès (112). Une autre approche est la régularisation Ridge, réalisant une régression linéaire dont les paramètres les plus grands sont pénalisés, mais aucun paramètre n'est réduit à zéro : il n'y a par conséquent aucune élimination de variable. Fréquemment employée, l'approche Elastic-net est constituée par l'association de ces deux méthodes de régression Lasso et Ridge, en cherchant un

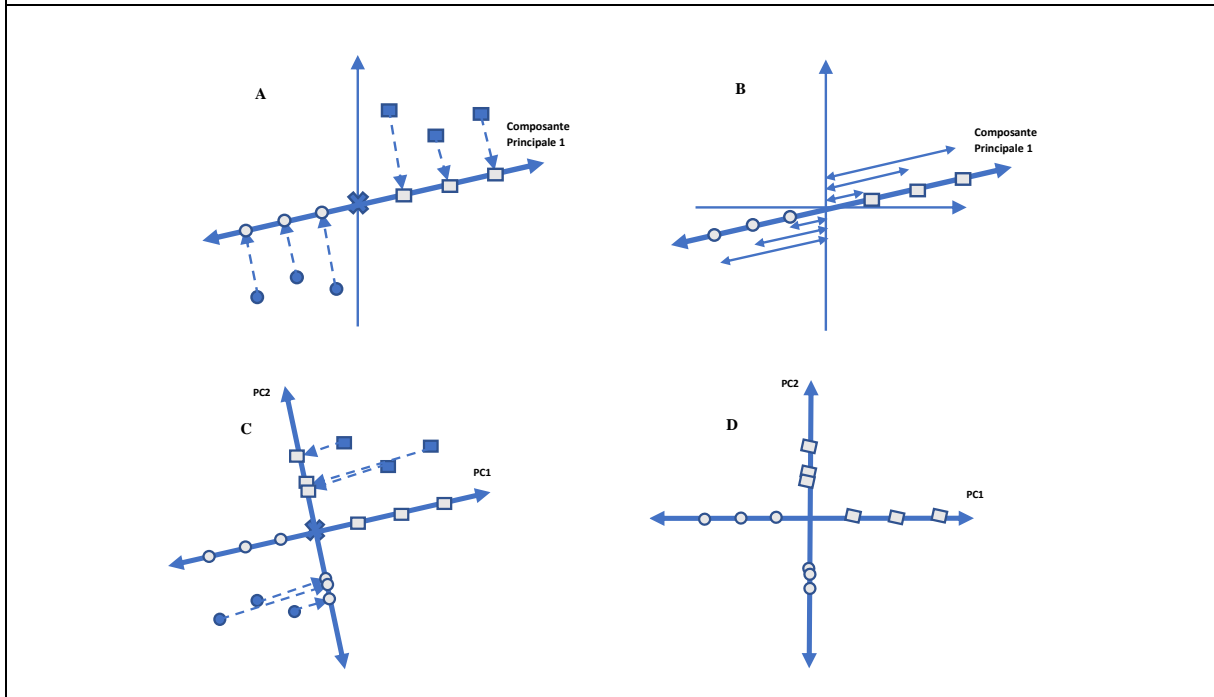
équilibre optimal entre leurs systèmes de pénalisation, évitant une élimination par excès de variables informatives et le surajustement aux index les plus importants (112).



D'autres méthodes telles que l'analyse en composantes principales (*Principal Component Analysis*, PCA) ou l'analyse linéaire discriminante (*Linear Discriminant Analysis*, LDA) permettent de réduire la dimension des variables d'intérêt en les transformant.

La PCA consiste à combiner linéairement les variables deux à deux, et débute par le centrage des variables sur leurs valeurs moyennes. Puis, elle retient l'équation linéaire, ou composante principale, permettant de maximiser la distance entre les projections des observations et la moyenne des données (figure 17). Une deuxième composante principale correspondant à l'orthogonale de la première équation, permet de tracer un nouveau repère bidimensionnel de projection des données. Enfin, la PCA analyse la part explicative de chacune des composantes principales en comparant la moyenne des distances à l'origine sur chaque axe, dans le nouveau repère (113). En pratique, la PCA continue d'ajouter n composantes principales, chacune étant dans un plan orthogonal au plan précédent. La PCA peut également avoir recours à des fonctions non linéaires (kernel) pour combiner les variables entre elles, la méthode étant alors désignée kernel-PCA.

Figure 17. Représentation schématique de la transformation des variables selon les composantes principales, avec linéarité des variables.



L'objectif de l'analyse discriminante linéaire est de combiner linéairement les variables de manière à maximiser la différence entre les moyennes des groupes, en projetant les données sur la droite de régression optimale (114). Dans le même temps, la LDA tente de minimiser la variance entre les groupes.

Les méthodes de PCA et de LDA peuvent également être utilisées pour une tâche de classification (113).

5.6. Apprentissage et optimisation du modèle

L'analyse s'appuie sur l'apprentissage automatique (*machine learning*) se déclinant en apprentissage supervisé et non supervisé, ainsi que l'apprentissage profond (*deep learning*). L'apprentissage est dit supervisé lorsque le critère de jugement principal, c'est-à-dire une classe ou une valeur de régression, est connu pour chaque sujet lors de l'entraînement. A l'inverse, il est dit non supervisé lorsque l'entraînement est effectué en aveugle de ce critère de jugement. Ces méthodes ont recours à des algorithmes pour entraîner un modèle capable d'effectuer une

tâche de classification ou de prédire une valeur quantitative par régression, en fonction des données qualitatives et quantitatives qui lui sont fournies. Idéalement, la cohorte est divisée en un échantillon d'entraînement, et un échantillon de validation, tous deux indépendants, tirés au sort (115–117). Pour s'affranchir de facteurs de confusion, il est préférable de vérifier l'absence de différences statistiques entre échantillons entraînement et test.

Lorsque l'apprentissage est supervisé pour une tâche de classification, La source d'informations est annotée au préalable avec le critère de jugement, par exemple un grade histologique, le statut d'une mutation, l'existence de métastases, un score clinique, une durée de survie globale etc. Les algorithmes vont ensuite ajuster les poids des différentes variables potentiellement explicatives jusqu'à atteindre un état d'équilibre où il est capable de segmenter au mieux l'échantillon d'entraînement selon les classes imposées (115–117). Les valeurs de ces poids correspondent aux paramètres du modèle, obtenus par apprentissage, par exemple les poids de liaison des couches d'un réseau de neurones : ils ne sont pas directement manipulables par l'utilisateur. A l'inverse, les hyperparamètres d'un modèle peuvent être configurés par l'opérateur avant la tâche d'apprentissage. Ces hyperparamètres sont propres à chaque type d'algorithme et peuvent eux même être optimisés lors des répétitions de l'entraînement. Ils doivent être renseignés pour assurer une reproductibilité (118).

Les algorithmes d'apprentissage supervisé se déclinent en prédicteurs linéaires et non linéaires. Les prédicteurs linéaires incluent l'analyse linéaire, les classifieurs bayésiens naïfs, la régression logistique ou encore les machines à vecteurs de support (115–117). Les prédicteurs non linéaires sont représentés par les arbres décisionnels, la méthode des k plus proches voisins, et les réseaux de neurones (115–117).

5.6.1. Prédicteurs linéaires

Ces prédicteurs ou classifieurs réalisent la tâche de classification en s'appuyant sur le résultat d'une combinaison linéaire, en attribuant un poids variable à chaque paramètre de manière à séparer deux ensembles de points correspondants aux deux classes par un plan séparateur discriminant. Ces prédicteurs sont idéaux pour répondre à un problème de faible complexité. Ces classifieurs sont de deux types : les modèles génératifs et les modèles discriminants (115,119). Les modèles génératifs ont pour but de modéliser la densité de

probabilité conditionnelle d'une classe en fonction d'un élément : les chefs de file sont le classifieur bayésien naïf en distribution binomiale et l'analyse linéaire en distribution gaussienne (115,119). Les modèles discriminants cherchent à maximiser la performance de classification lors de l'entraînement, puis à minimiser les erreurs par une fonction de coût. Les plus employés sont les modèles de régression logistique, les machines à vecteurs de support et le perceptron (115,119).

5.6.1.1. Classifieur bayésien naïf

Le classifieur bayésien naïf est issu du théorème de Bayes formulant qu'une probabilité conditionnelle $P(c|x)$ de l'appartenance à une classe c en connaissant x la variable prédictive de l'élément à classer s'exprime telle que :

$$P(c|x) = \frac{[P(c) \times P(x|c)]}{P(x)}$$

$P(c)$ est la probabilité *a priori* d'appartenir à la classe c , $P(x|c)$ la fonction de vraisemblance de x et $P(x)$ la probabilité *a priori* de x (115,120). Les variables étant multiples, la probabilité finale *a posteriori* $P(c|x)$ résultante correspond à la multiplication de l'ensemble des probabilités conditionnelles. L'avantage du modèle bayésien est donc d'être facilement compréhensible et utilisable, d'autant plus qu'il est adapté aux petits échantillons et aux données manquantes. Cette approche est dite naïve : elle fait l'hypothèse que toutes les variables sont indépendantes et d'importances égales, ce qui est exceptionnellement le cas (115,120). Pour cette méthode, il faut souligner l'importance des probabilités *a priori* issues de l'échantillon : si la classe c a une prévalence de 95%, le modèle a de fortes chances de classer comme c tous les éléments avec une précision faussement excellente, nécessitant d'analyser alors la sensibilité et la spécificité pour éviter cet écueil (115). Il est également important d'adapter le design d'étude en fonction de ces probabilités *a priori* : faut-il opter répliquer la prévalence de chaque pathologie pour un modèle en situation réelle, ou plutôt représenter chaque maladie à parts égales pour analyser la capacité du modèle à les différencier ?

5.6.1.2. Régression logistique

La régression logistique est un modèle probabiliste employant le résultat d'une fonction de maximum de vraisemblance pour classer un élément, selon une distribution binomiale. Le modèle produit une valeur de probabilité d'appartenance d'un élément à une classe en fonction de la combinaison linéaire des variables explicatives par une fonction logistique (figure 18) avec un seuil discriminant à $p = 0,5$ (120). Le principal avantage de ce modèle réside en la facilité d'interprétation d'une probabilité – et son intervalle de confiance – permettant d'établir le tableau de contingence, ainsi qu'une matrice de confusion estimant la performance de classification (120). Il est également possible de mettre à jour le modèle facilement avec de nouvelles données. Cette approche est sensible aux variables fortement corrélées, il est ainsi nécessaire d'identifier ces corrélations pour ne conserver que les variables mutuellement indépendantes (120).

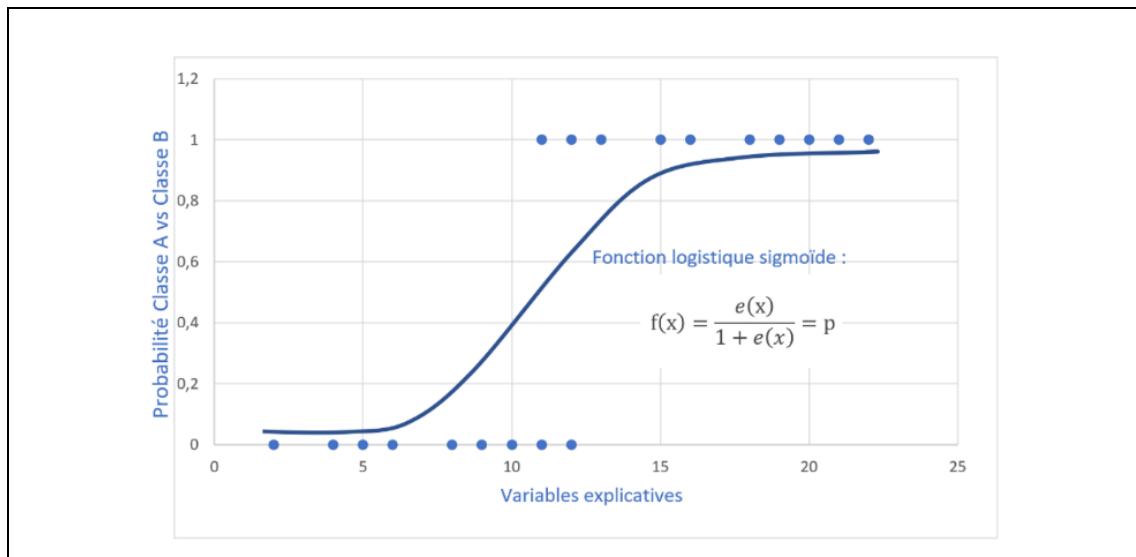
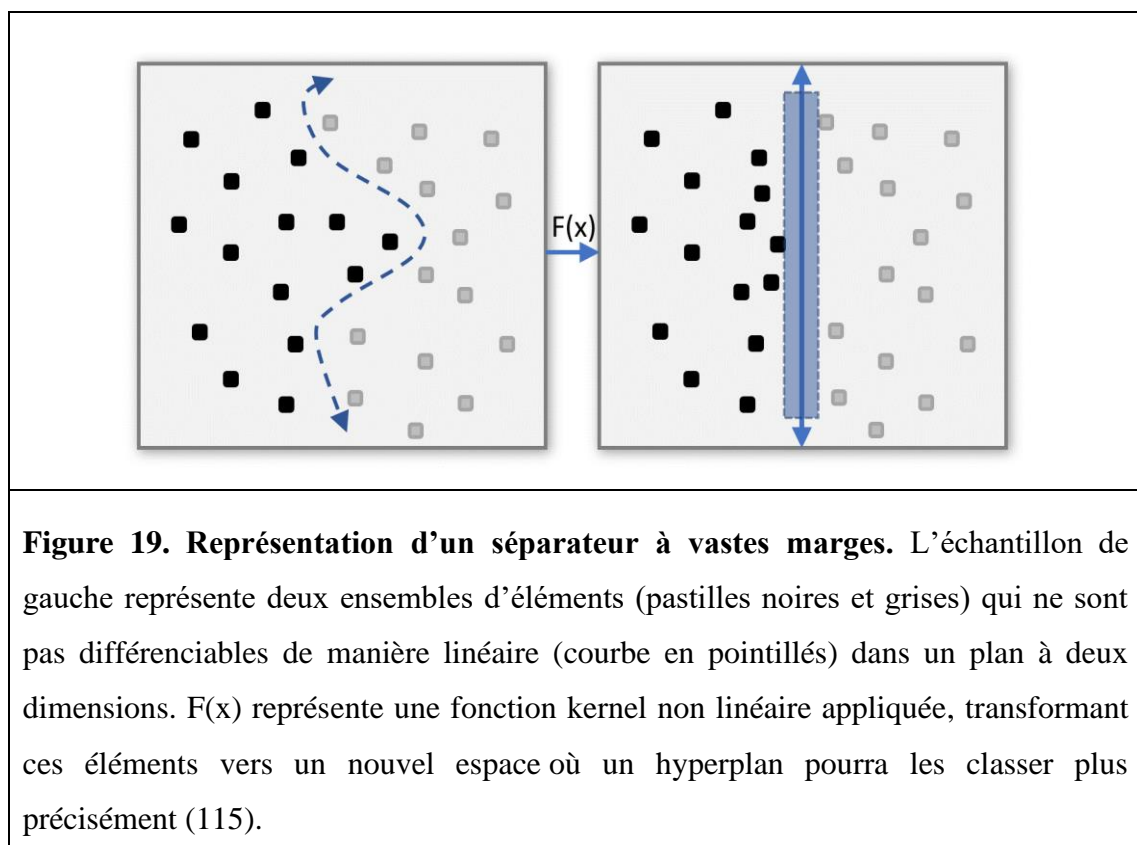


Figure 18. Régression logistique binomiale. Graphique représentant la probabilité de classe A (lorsque $p = 1$) ou de classe B (lorsque $p = 0$). Les pastilles correspondent au jeu d'entraînement étiqueté avec $p = 0$ ou 1 . Les nouveaux éléments sont situés selon les variables explicatives par une fonction logistique sigmoïde de probabilité avec un seuil à $p = 0,5$.

5.6.1.3. Machines à vecteurs de support

Une machine à vecteurs de support, ou séparateur à vastes marges (SVM), maximise les marges d'un plan séparateur qui discrimine linéairement les points des deux classes. Il est

fréquent que l'on ne puisse résoudre un problème avec une solution linéaire. Pour remédier à cela, les SVM ont recours à l'astuce du noyau (*kernel trick*) consistant à transformer de manière non linéaire les valeurs vers un espace de plus grande dimension, cette transformée étant réalisée par une fonction dite kernel, par exemple une fonction polynomiale, gaussienne ou encore une fonction à base radiale. Dans l'espace de dimension supérieure, le SVM procédera alors à identifier l'hyperplan séparateur pour réaliser la discrimination des éléments (figure 19) (115). Outre le choix de la fonction kernel, deux hyperparamètres sont modulables par l'utilisateur : la variable C de régularisation et la variable γ (ou σ). La variable C influence la largeur souhaitée des marges séparatrices : une valeur plus basse est en faveur de marges plus grandes, se traduisant par une précision de classification accrue au prix d'une exactitude inférieure. La variable γ définit la sphère d'influence de la classe d'un point sur un autre : une valeur basse est en faveur d'une plus forte influence à distance.



5.6.1.4. Perceptron

Le principe du perceptron est développé au chapitre 5.6.2.3 traitant des réseaux de neurones.

5.6.2. Prédicteurs non linéaires

5.6.2.1. Arbres décisionnels et forêts aléatoires

Les arbres décisionnels, de par leur utilisation courante en routine clinique en dehors des modèles d'apprentissage automatique, ont l'avantage de présenter de manière très graphique le processus de classification. Ils sont constitués d'une succession de nœuds représentant chacun une opération logique confrontant la valeur d'un index à celle d'un seuil : par exemple la supériorité ou l'infériorité pour un index quantitatif, ou encore le statut d'un index qualitatif (115,120). La réponse binaire oui/non détermine le chemin pris à chaque division de l'arbre. Ces embranchements relient la racine de l'arbre qui est la variable la plus pertinente pour la classification – dite variable la moins impure - aux feuilles périphériques, qui permettront la décision finale (figure 20). Certaines variables et les opérations logiques associées peuvent apparaître plusieurs fois au sein d'un même arbre à différents embranchements. L'algorithme testera alors plusieurs combinaisons pour déterminer le modèle le plus optimal (115,120).

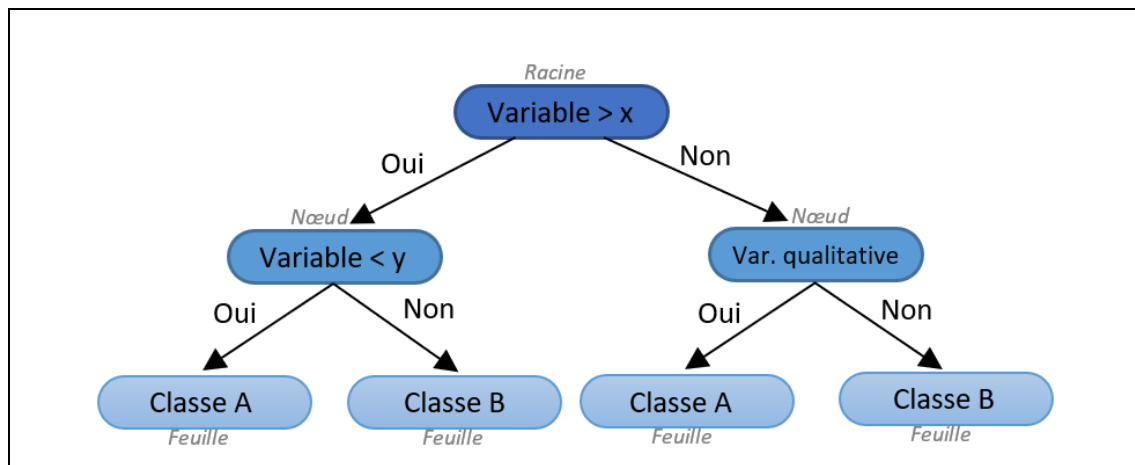


Figure 20. Exemple d'arbre décisionnel (115). Le nœud racine est le point d'entrée de l'arbre, il correspond à la variable initiale permettant la première dichotomie la plus pertinente. Les nœuds décisionnels internes orientent ensuite vers les nœuds feuilles finaux qui permettent la classification. Les nœuds peuvent être des variables quantitatives avec un seuil ou qualitatives. La décision est binaire.

Toutefois, l'arbre décisionnel est limité puisqu'il ne propose en sortie qu'une seule classification pour chaque entrée. De plus, une fois constitué, l'arbre décisionnel est difficilement actualisable avec de nouvelles données. Pour pallier à ces restrictions, des méthodes dites d'ensembles telles que le bagging, le boosting ou les forêts aléatoires (random forest, RF) proposent de multiplier les arbres de classification pour une même tâche : on obtient alors plusieurs résultats pour une même entrée, qui permettront la classification finale selon le vote majoritaire, ou la tâche de régression par la moyenne numérique des sorties (120) :

- Le **bagging** est une méthode se basant sur un rééchantillonnage des données d'entraînement pour la génération de chacun des arbres décisionnels qui participeront à la votation finale. La particularité étant que ces rééchantillonnages itératifs sont dits *bootstrapped*, c'est-à-dire que les jeux d'entraînement sont constitués aléatoirement, cependant chaque observation peut être tiré au sort plusieurs fois : cela permet d'utiliser les mêmes données dans plusieurs échantillons. On considère alors que ces échantillons d'entraînement sont indépendants des arbres décisionnels constitués. Pour chaque nœud décisionnel, toutes les variables de l'échantillon d'entraînement sont considérées.
- Le **boosting** diffère du bagging du fait que chacun des arbres décisionnels créés a pour but de compenser les erreurs de classification de l'arbre précédent, en attribuant un poids plus important aux fautes, tout en diminuant celui des sujets correctement classés. Les arbres sont donc interdépendants du fait des itérations successives, à l'inverse du bagging où les arbres ne s'influencent pas mutuellement.
- Les **forêts aléatoires** sont constituées d'arbres décisionnels formés par un sous ensemble d'index tirés au sort (figure 21). Pour constituer chaque arbre, on peut choisir de rééchantillonner les données en sous-échantillons indépendants par *bootstrap*. La création d'une forêt aléatoire dépend essentiellement de deux hyperparamètres : le nombre prédéfini d'arbres et le nombre de variables considérées à chaque nœud, donc de leurs proportions respectives. Un trop grand nombre de données explorées par un faible nombre d'arbres fait prendre le risque de méconnaître des variables hautement explicatives mais sous utilisées. De la même manière, si le sous échantillonnage des données est trop restrictif ajoutés à un nombre trop faible d'arbres, les performances du modèle seraient inférieures son potentiel. Les forêts

aléatoires ont l'avantage d'être facile à calculer, tout en étant peu sensibles aux données manquantes et à l'*overfitting*. En pratique, il convient de rechercher l'optimum des hyperparamètres par un algorithme d'optimisation tel que *grid search* et une méthode de validation d'entraînement comme la validation croisée.

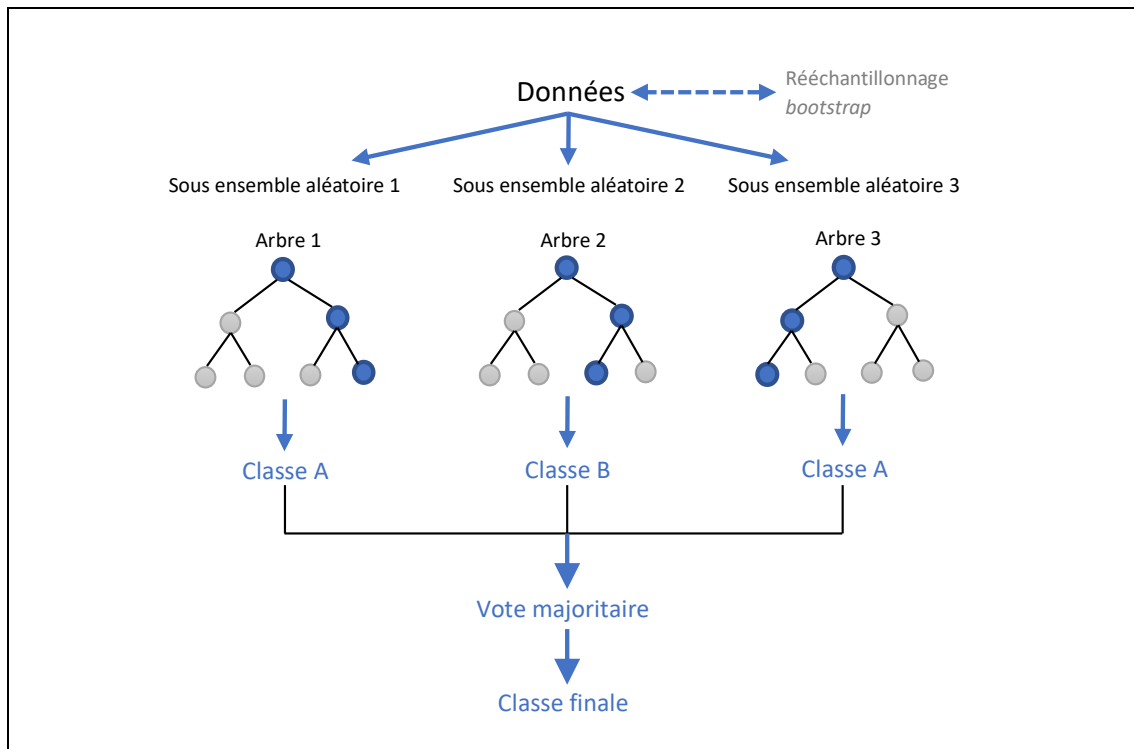


Figure 21. Exemple de forêt aléatoire pour une tâche de classification (115). Un sous ensemble de variables est sélectionné aléatoirement pour chaque arbre décisionnel puis, pour chaque nœud, la variable la plus discriminante est identifiée. Chaque arbre propose une classification finale, la votation majoritaire déterminant la classe retenue. Pour une tâche de régression, la moyenne pondérée de tous les arbres sera la valeur de sortie.

5.6.2.2. Méthode des k plus proches voisins

La méthode des k plus proches voisins (*k-nearest neighbors*, kNN) a pour principe de classer un nouveau sujet en fonction de la distance de celui-ci avec les deux classes préalablement étiquetées. K désigne le nombre de voisins proches à prendre en considération, et le nombre majoritaire de voisins proches de l'une des classes déterminent la classe finale (figure

22) (115,120). La distance peut être évaluée par une fonction dite de similarité, mesurant l'éloignement (distance euclidienne ou Minkowski) entre le vecteur évalué et les autres vecteurs préalablement classés. Ainsi, lorsque $k = 1$, le sujet est classé selon l'étiquette du seul voisin le plus proche, avec un risque d'*overfitting*. A l'inverse, choisir un k plus grand expose à l'écueil d'un modèle non ajusté (*underfitting*) avec impossibilité de classer significativement. Une validation croisée du modèle permet d'ajuster une valeur de k pertinente (120). Puisqu'elle a recours aux distances dans un espace vectoriel, cette méthode est très sensible aux données non normalisées.

Cette approche revêt plusieurs avantages, dont le fait de ne pas avoir de « coût » d'apprentissage du modèle, puisque les éléments non classés sont immédiatement évalués en fonction d'une distance calculée. De plus, s'agissant d'une approche non paramétrique, il n'est pas nécessaire d'évaluer le type de distribution des données (120). Enfin, n'importe quel type de données est utilisable, dès lors qu'une mesure de cette distance vectorielle est réalisable. Toutefois, l'interprétation et la lisibilité d'un résultat de classification sont difficiles, car il n'est obtenu en sortie que les proportions de classification dans chaque classe (120). De même, une mesure de distance reste peu tangible (120). La méthode peut être limitée lorsqu'il s'agit d'un problème de dimension supérieure avec un jeu de données de grande taille : il est alors nécessaire de réduire la dimension pour éviter un temps de calcul rédhibitoire (120).

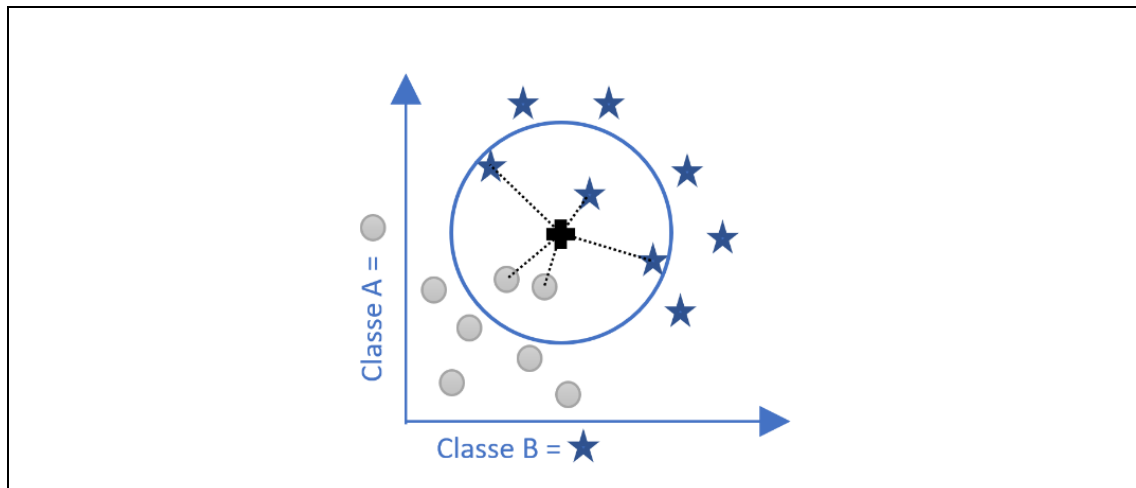
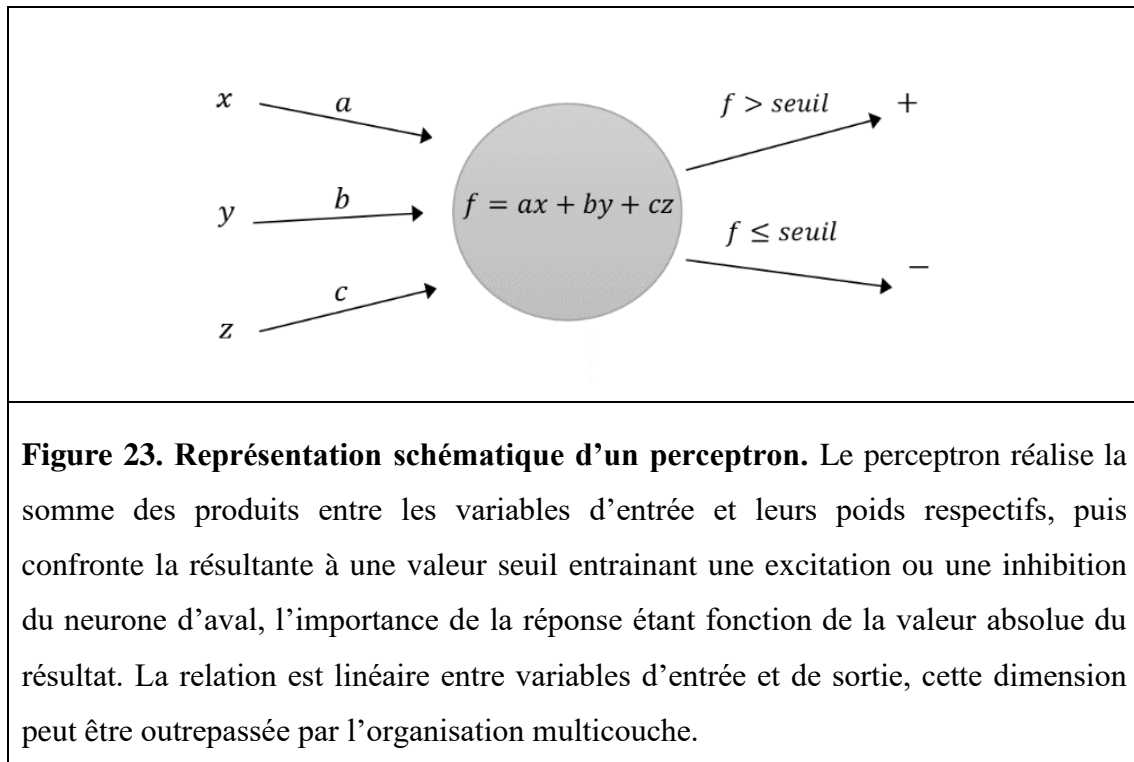


Figure 22. Exemple de classification par la méthode des k plus proches voisins (115). La croix (+) représente le nouvel élément à classer, le cercle illustre ses plus proches voisins dans l'espace vectoriel. Les pointillés montrent la distance euclidienne aux plus proches voisins. En choisissant $k = 3$, l'élément sera classé A (rond, ○), car les 3 plus proches voisins sont deux classes A (ronds, ○) contre une classe B (étoile, ★). Cependant, en choisissant $k = 5$, cet élément serait classé B.

5.6.2.3. Apprentissage par réseaux de neurones

L'apprentissage par réseaux de neurones s'inspire de l'architecture neuronale humaine et animale, notamment le cortex visuel, en reproduisant un agencement en couches de neurones artificiels. Ces derniers, également nommés perceptrons, sont une figuration mathématique d'un neurone biologique, avec plusieurs entrées, sorties, et fonctions de transformation des données, correspondant respectivement aux dendrites, aux cônes d'émergence et aux corps neuronaux (figure 23) (115,118). Dans le modèle classique, la couche d'entrée initie le traitement des données, les couches intermédiaires appliquent différentes opérations logiques et la couche de sortie effectue la tâche de classification ou de régression (120).



L'importance de la connexion entre ces perceptrons est représentée par des poids modulant les variables, ces poids étant d'intensités ajustables lors de l'entraînement, en fonction des performances (120). Dans le schéma classique du perceptron, le neurone artificiel effectue la somme des produits des variables entrantes avec les poids associés, puis compare la résultante à une valeur seuil, déclenchant une excitation ou une inhibition du neurone d'aval (120) (figure 23). Les poids associés aux variables d'entrée sont ainsi une fonction de chaque réponse de sortie : ils représentent donc une approximation des corrélations entre données initiales et classes ou régressions finales (121). Du fait de leur organisation multicouche, les réseaux de neurones permettent de dépasser la dimension linéaire d'un seul perceptron, et peuvent ainsi faire état de relations non linéaires entre variables et classes, dépendantes ou indépendantes (120,121). Ce sont des modèles non paramétriques qui permettent de s'affranchir en partie d'erreurs et de variations des mesures (120).

L'entraînement d'un réseau de neurones classique consistera donc à optimiser les valeurs de ces poids, en fonction du nombre de couches, des variables d'entrée plus ou moins explicatives et du nombre de classes définies (121). Pour ce faire, la méthode par propagation rétrograde (*backpropagation*) permet de quantifier la différence entre les prédictions de l'algorithme et les classes réelles au fur et à mesure de l'entraînement, concrétisée par une fonction de perte (ou de « coût ») dépendant de la valeur des poids pour chaque couple variable –

classe (figure 24). Cela permet d'optimiser le réseau via une fonction de descente de gradient (*gradient descent*) qui recherche la situation optimale où la fonction de perte atteint un minimum (figure 25) (122).

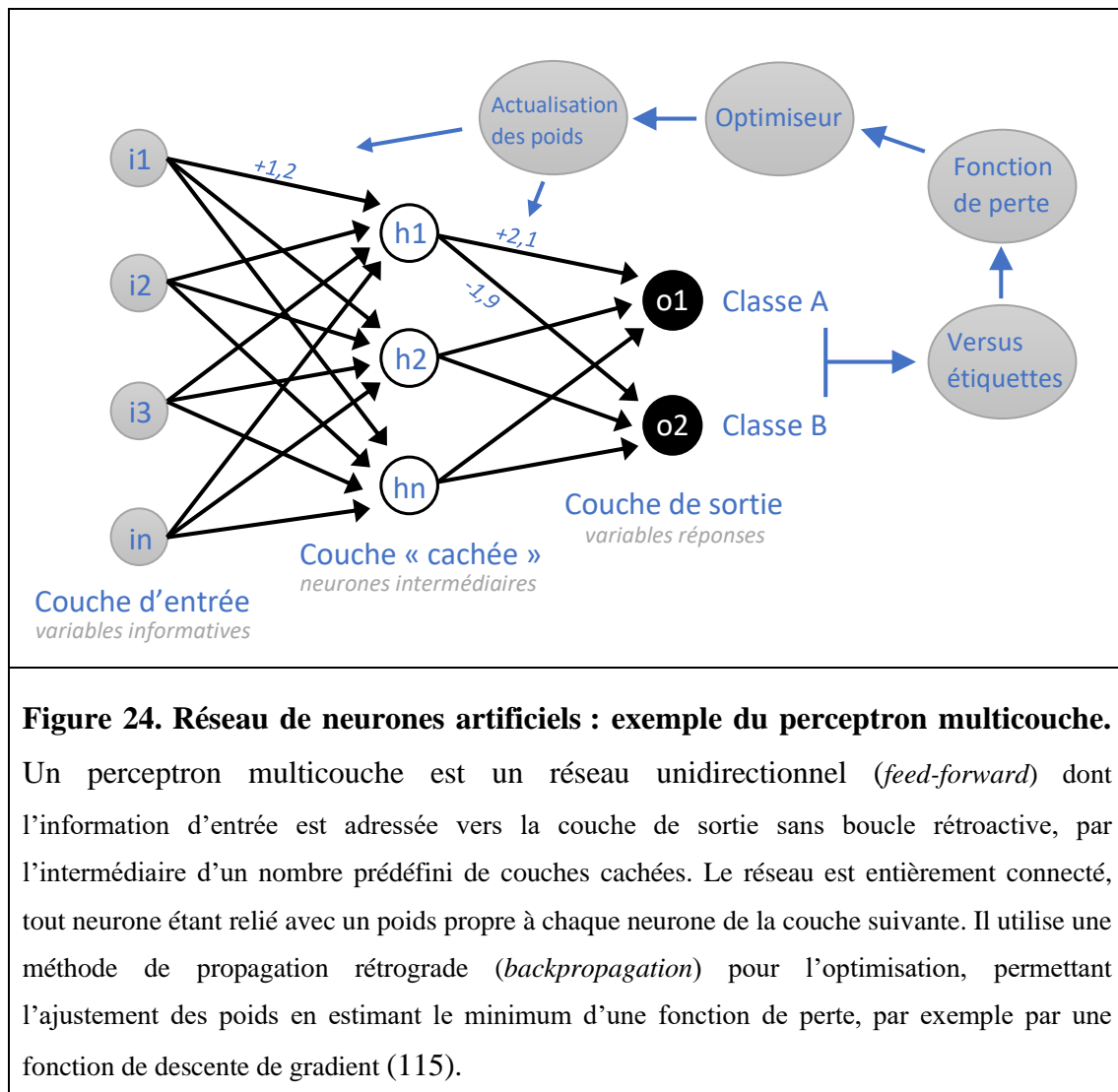
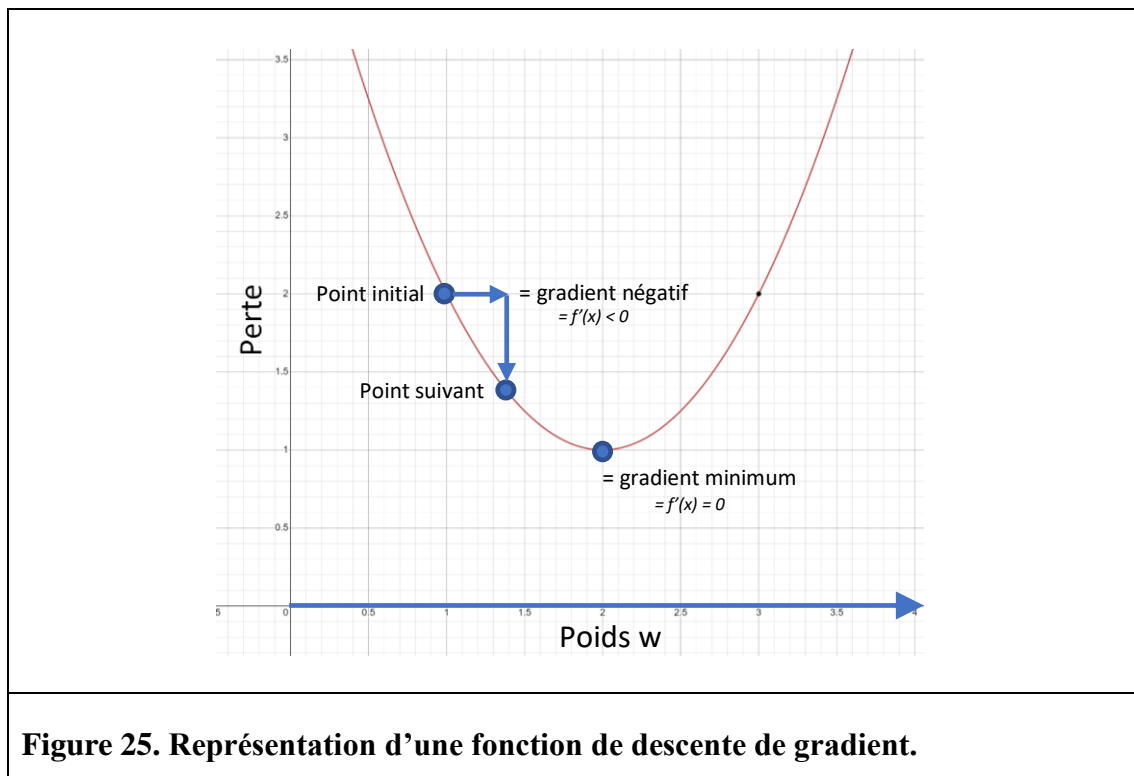


Figure 24. Réseau de neurones artificiels : exemple du perceptron multicouche.

Un perceptron multicouche est un réseau unidirectionnel (*feed-forward*) dont l'information d'entrée est adressée vers la couche de sortie sans boucle rétroactive, par l'intermédiaire d'un nombre prédéfini de couches cachées. Le réseau est entièrement connecté, tout neurone étant relié avec un poids propre à chaque neurone de la couche suivante. Il utilise une méthode de propagation rétrograde (*backpropagation*) pour l'optimisation, permettant l'ajustement des poids en estimant le minimum d'une fonction de perte, par exemple par une fonction de descente de gradient (115).



Initialement, les réseaux de neurones comportaient un faible nombre de couches, limités par la fréquence de calcul des processeurs (115). Toutefois, les récentes évolutions de ces derniers et le recours aux processeurs graphiques permettent d'employer plusieurs dizaines de strates, l'apprentissage est alors arbitrairement considéré comme profond lorsqu'il emploie des réseaux à plus de 20 couches (115). A titre de comparaison, les réseaux employés pour la reconnaissance d'images par les moteurs de recherche sont constitués de 50 à plus de 100 couches. Grâce à ces réseaux, il est ainsi possible de traiter efficacement les données de masse de grandes cohortes, en exploitant notamment l'architecture parallèle des processeurs qui permettent d'effectuer de multiples opérations simultanément. Du fait de l'importante quantité de calculs et de données générés par ces modèles, il est nécessaire d'avoir recours à une cohorte étendue, même si certaines techniques permettent d'augmenter artificiellement les données en appliquant notamment des rotations ou modifications de contraste aux images (115).

Les réseaux de neurones sont employés pour diverses tâches, dont le dépistage de lésion, la classification diagnostique, la stratification du risque, ou encore la segmentation (115). Il existe plusieurs types d'algorithmes d'apprentissage profonds caractérisés par leurs architectures : perceptron multicouche, réseaux de neurones récurrents, auto-encodeurs,

machines de Boltzmann ou encore réseaux de neurones convolutifs (CNN, *convolutional neural networks*). Ces derniers sont les plus employés en recherche pour l'analyse directe des images ou vision par ordinateur (*computer vision*), en particulier depuis 2012 et les excellents résultats de reconnaissance d'objets dans le cadre du défi *ImageNet* avec des taux d'erreur inférieurs à 5% (123,124).

Le principe d'un réseau de neurones convolutifs de type CNN réside en la lecture itérative des pixels par plusieurs types de filtres se déplaçant sur l'image. Ces filtres, également appelés kernels ou *features detectors*, effectuent une opération de convolution et génèrent une carte d'activation (*feature map*) pour chaque type de kernel (figure 26). En pratique, ces cartes d'activation dénombrent le nombre de co-occurrences de pixels de même valeur entre un filtre et l'image d'entrée, effectuant un sous échantillonnage des données. Les filtres sont généralement de taille 3x3 et 5x5, et leurs valeurs peuvent correspondre à la recherche d'une particularité telle qu'un contour. Toutefois, la plupart des filtres générés par le modèle ne peuvent être explicitement traduits. Les images étant toujours filtrées dans le sens de lecture matriciel, l'information spatiale des données persiste au sein des cartes d'activation (118).

Puis, les données convoluées sont corrigées par une unité de rectification linéaire ou redresseur (*Rectifier Linear Unit, ReLU*) qui permet de s'affranchir d'une analyse uniquement linéaire, la plupart des caractéristiques d'une image telles que les contours, les variations d'intensités ou de couleurs, étant généralement non linéaires (118). Cette fonction permet de transformer des données continues en une variable binaire, donnant un signal d'activation ou de non-activation en sortie (figure 26).

Ensuite, les données passent par une couche de sous échantillonnage, permettant la réduction des données et ainsi une optimisation de la puissance de calcul. Le plus fréquent le filtre 2x2 de *max pooling*, qui récupère la valeur maximale dans sa zone de lecture, résultant en une compression des données par un facteur 4 et permet de diminuer le risque d'*overfitting* (figure 27) (118). L'ensemble des couches de convolution avec activation, et de *pooling* constitue un bloc unitaire qui peut être répété en série, ajoutant de la profondeur au réseau.

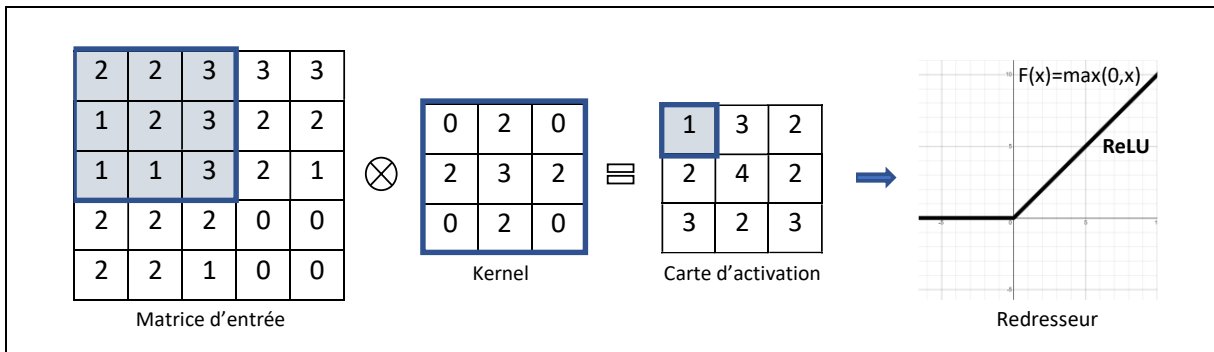


Figure 26. Exemple de couche de convolution dans un CNN. La lecture par le kernel se fait dans le sens matriciel. Dans cet exemple, on ne retrouve qu'une seule correspondance entre la matrice d'entrée et le kernel, résultant en la valeur 1 sur la carte d'activation. Selon les modèles, la distance entre deux positions successives du kernel est dénommé *stride*, ou « pas » : il est habituellement de 1, correspondant alors à une lecture matricielle habituelle. Il est parfois de 2 ou plus lorsque l'on souhaite sous échantillonner les données. La lecture du filtre est limitée en périphérie par l'impossibilité du kernel d'être en dehors de la matrice d'entrée : autrement dit, le centre d'un kernel 3x3 ne peut lire les premières et dernières colonnes ou rangées. Pour pallier à cela, la méthode du *zero padding* ajoute des deux colonnes et deux rangées de valeurs nulles aux extrémités de la matrice, permettant la lecture en périphérie par le kernel. Cet ajustement est important lorsque l'architecture est constituée de plusieurs convolutions successives qui auraient, sans padding, abouti à une importante perte de données (118).

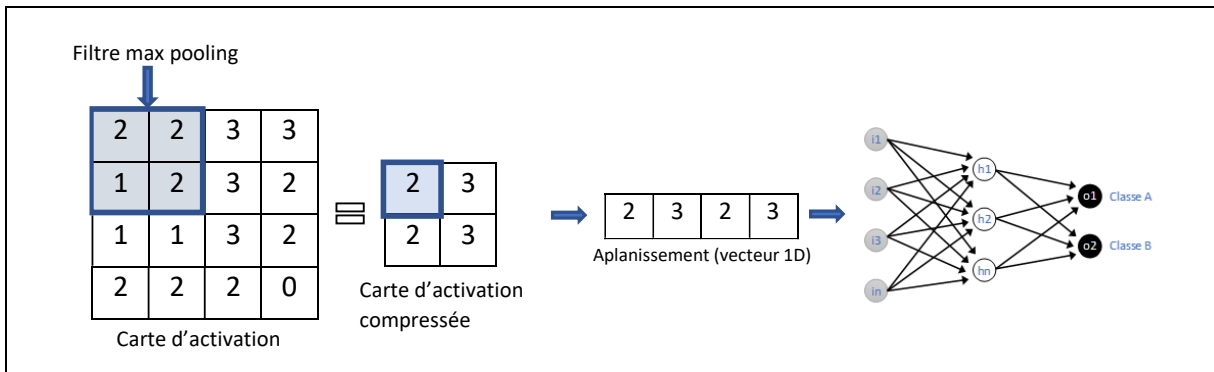


Figure 27. Exemple de couche de compression (*pooling*) de la carte d'activation dans un CNN. Réduction des données de la carte d'activation par un filtre *max pooling* 2x2 : seule la valeur maximale est retenue. D'autres filtres sont possibles comme l'*average pooling*. Puis la carte d'activation compressée est aplanie en un vecteur unidimensionnel fourni en entrée à un perceptron multicouche. (118)

Les matrices sont alors aplanies en plusieurs vecteurs unidimensionnels (étape de *flattening*), ceux-ci étant adressés à l'entrée d'un réseau de neurones à couches entièrement connectées, par exemple de type perceptron multicouche (118).

Enfin, en sortie de réseau, la fonction *Softmax* normalise les probabilités des deux classes, spontanément indépendantes, en deux probabilités complémentaires de somme 1. Cela permet d'appliquer la fonction *Cross-Entropy*, qui quantifie l'erreur en fonction du logarithme des probabilités. Cette fonction est préférable pour les tâches de régression, tandis qu'on emploiera l'erreur quadratique moyenne pour les régressions (118).

L'optimisation du modèle viendra alors mettre à jour les poids des liaisons du réseau entièrement connecté mais également les kernels du CNN, donc les cartes d'activation correspondantes. A l'instar des perceptrons multicouches, cette optimisation a également recours à un optimiseur type gradient de descente cherchant le minimum de la fonction de perte, faisant suite à la confrontation entre prédiction et étiquette réelle (118). La figure 28 propose un aperçu récapitulatif des différentes étapes d'un CNN standard.

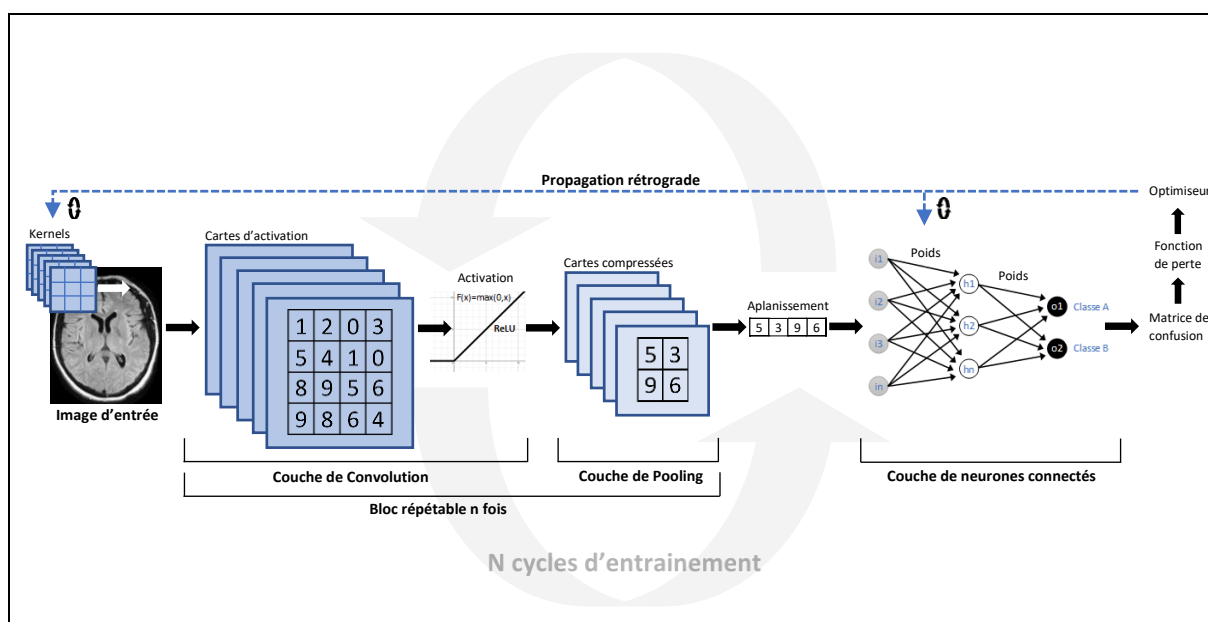


Figure 28. Aperçu récapitulatif des couches successives d'un CNN standard. Les kernels effectuant les convolutions et les poids des liaisons entre les neurones de la couche dense sont mis à jour après optimisation à chaque cycle d'entraînement (*epoch*). Les blocs *convolution-pooling* peuvent être répétés en série. Certains modèles proposent des blocs *convolution-pooling* agencés séparément, en branches parallèles (118).

Par convention, les paramètres d'un CNN correspondent aux variables optimisées par le modèle au fil des entraînements, tandis qu'on nomme hyperparamètres les variables préalablement fixées par l'utilisateur, avant tout entraînement. Ainsi, les valeurs des kernels et

des poids des couches de neurones sont des paramètres, tandis que, par exemple, la taille d'un kernel constitue un hyperparamètre parmi d'autres (tableau 5).

De nombreux paramètres et hyperparamètres sont ajustables au sein d'un CNN (tableau 5), rendant possible un grand nombre d'architecture : ainsi, depuis 2012, plus de 250 modèles de CNN ont été proposés pour le seul défi de classification *ImageNet* (125).

Tableau 5. Paramètres ajustables pour tout CNN avant entraînement (118).

Couche	Paramètres	Hyperparamètres
Convolution	Kernels	Nombre, taille, valeurs et <i>stride</i> des kernels. <i>Zero Padding</i> de la matrice d'entrée. Fonction d'activation (ReLU, leaky ReLU, tanh, sigmoid)
Pooling	-	Taille et opération du filtre (max, average), <i>stride</i> , <i>padding</i> .
Réseau de neurones	Poids	Nombre de poids, de couches et fonction d'activation.
Architecture		Nombre et agencement des blocs de <i>convolution-pooling</i> , type d'optimiseur, fonction de perte, nombre d'itération d'entraînement (<i>epochs</i>), taille des sous-échantillons par cycle (<i>batch size</i>), division du jeu de données.

5.7. Evaluation et validation du modèle entraîné

L'évaluation de l'entraînement du modèle consiste à quantifier sa capacité à classer justement de nouveaux éléments issus d'un échantillon indépendant. Cette évaluation permettra d'éviter un modèle trop ou trop peu ajusté au jeu d'entraînement (*overfitting* et *underfitting*) et de déterminer le modèle plus optimal. Pour cela, la technique de validation croisée à k sous-ensembles ou blocs (k-fold cross validation, kfCV) est fréquemment utilisée (figure 29) (120). Elle consiste à diviser l'échantillon d'entraînement en k sous-ensembles d'éléments égaux, parmi lesquels k-1 constituent l'échantillon d'entraînement, tandis que le dernier sous-ensemble sert de jeu de validation. Une fois les k sous-ensembles formés, une boucle répétée k fois permet d'effectuer les k entraînements et validations, chaque sous-ensemble étant utilisé une fois pour validation. En découlent k évaluations de performance, par exemple sous la forme de k valeurs d'exactitude dont on peut estimer la moyenne et la variance. Habituellement, la valeur de k employée est de 10, mais dépend essentiellement du nombre de sujets dans la cohorte (120).

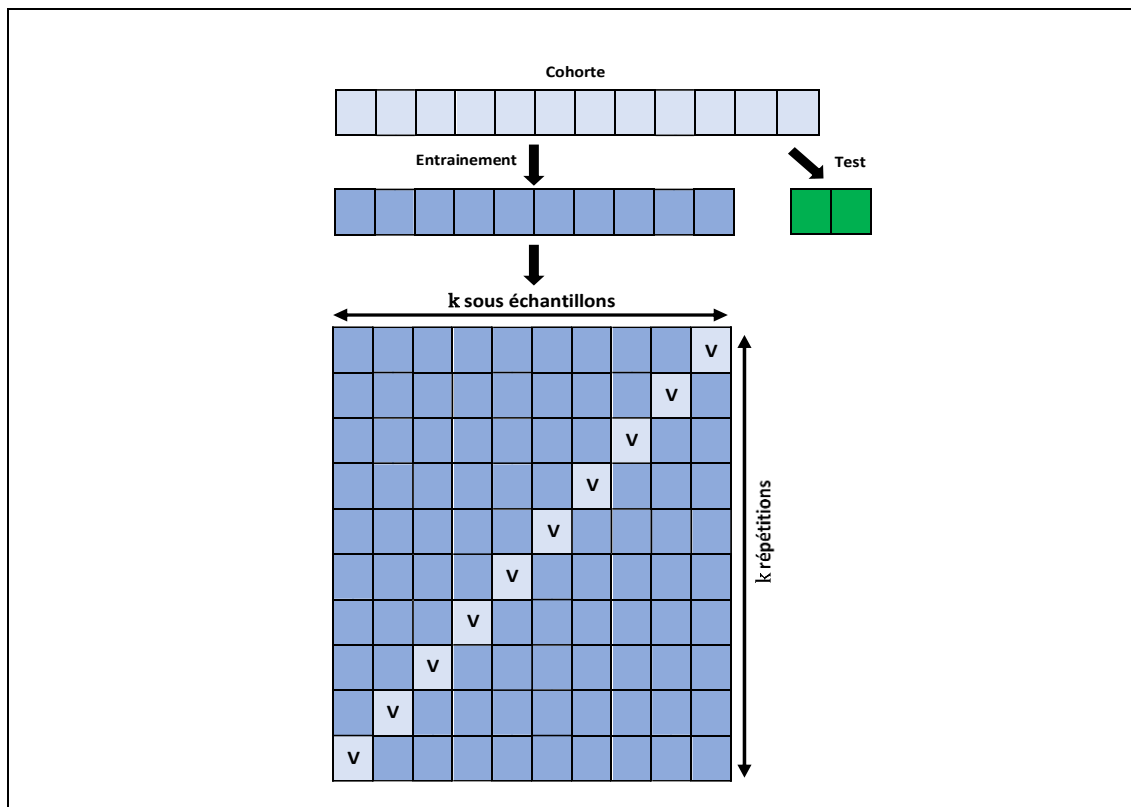


Figure 29. Principe de la validation croisée à k blocs (*k fold cross validation*). La cohorte de données est divisée en un échantillon d’entraînement et un échantillon de test, en proportion variable. L’échantillon d’entraînement est lui-même divisé en $k = 10$ sous échantillons, chacun des sous échantillons servant de jeu de validation pour un entraînement répétés k fois.

Pour explorer les performances du modèle, un tableau de contingence, ou matrice de confusion, collige les sujets en fonction des classes prédites et réelles, et permet de calculer entre autres : sensibilité, spécificité, valeurs prédictives positives et négatives, rapports de vraisemblances, exactitude (120).

L’évaluation passe également par l’estimation de l’importance de chacune des variables du modèle entraîné, en s’appuyant plus particulièrement sur leurs corrélations linéaires et non linéaires respectives avec chacune des classes finales. Il est aussi possible de réévaluer les corrélations entre les différents index, via une matrice de covariance, afin de ne conserver que celui potentiellement explicatif et éliminer ceux redondants (120).

La courbe *Receiver Operating Characteristic* (ROC) fait état de la capacité à classer de manière binaire en fonction des variations de sensibilité et de spécificité du modèle (ou de ses

valeurs prédictives positives et négatives). L'intégrale de la fonction ROC permet d'obtenir l'aire sous la courbe (Area Under the Curve, AUC) permettant de quantifier et comparer les performances de différents modèles (120).

5.8. Stabilité des index radiomiques

La stabilité ou robustesse d'un index désigne la faible modification de la variance, et donc de l'intervalle de valeurs qu'il représente, lors de la répétition de la mesure. La répétabilité désigne alors la stabilité d'un index mesuré à nouveau, dans des conditions méthodologiques inchangées. La reproductibilité se définit par la stabilité de cet index lorsque l'on change la méthode, par exemple l'emploi d'un différent imageur, un nouveau logiciel, un autre opérateur etc. en faisant l'hypothèse que la modification méthodologique ne modifie pas elle-même l'index (126).

La reproductibilité et la répétabilité des index radiomiques est rendue difficile par de multiples biais, parmi lesquels l'absence de standardisation des outils de mesures et des méthodes de validation statistiques, la démarche méthodologique imprécisément décrite, ou encore l'absence de détection de facteurs confondants au sein de la base de données, d'autant plus si l'étude est rétrospective (2).

Pour éviter ces écueils, un modèle généré sur des données rétrospectives doit être validé avec un nouvel échantillon indépendant, dont l'acquisition aura été faite sur une différente machine (2). Idéalement, la validation d'un modèle ou d'un profil radiomique comme biomarqueur devrait bénéficier d'une étude multicentrique prospective employant des systèmes d'acquisition de différents constructeurs et différents investigateurs (2). Cette approche permet d'évaluer les performances du modèle en validation externe tout en évaluant la reproductibilité des index (2).

La répétabilité des index radiomiques est évaluable par la méthode du test-retest, qui permet de quantifier la variance de la mesure faite lors de plusieurs acquisitions successives, réalisées à quelques minutes d'intervalles et avec des paramètres inchangés (127). Cette analyse fait l'hypothèse que le tissu lésionnel et les index radiomiques en rapport ne varient pas lors de ce faible laps de temps. Toutefois, la stabilité d'un index radiomique pourrait dépendre de l'histologie de la lésion, et de ses différents contingents, par exemple nécrotiques ou

hémorragiques (127). Idéalement, il faudrait réaliser un test-retest pour tout type et sous type de lésion, à chaque étape du suivi en imagerie (127).

L'évaluation de la répétabilité et de la reproductibilité s'appuie le plus souvent sur le coefficient de corrélation de concordance (CCC) de Lin faisant l'hypothèse de l'indépendance des observations (128). Dans une situation idéale, deux mesures successives d'une même variable seront égales et le CCC vaudra 1. Le seuil inférieur de stabilité d'un index radiomique est variable suivant les études, avec un CCC minimum allant de 0,75 à 0,90 (126). Une variante de cet indice est le coefficient de corrélation intra-classe (CCI), utilisé pour l'analyse de la reproductibilité lorsque l'on s'attend à une forte corrélation de la variable au sein d'une classe mais une faible corrélation de la même variable entre différentes classes (126).

Une méta-analyse de la stabilité des index radiomiques par Traverso et al. portant sur 41 études, incluant essentiellement des cancers pulmonaires et oropharyngés, rapporte une meilleure reproductibilité et répétabilité des index de premier ordre, parmi lesquels l'index Entropy apparaît le plus stable, suivi des index Skewness et Kurtosis (126). Les index morphologiques et d'ordres supérieurs semblent moins robustes, en particulier Coarseness et Contrast avec cependant une seule étude multicentrique dans l'analyse (126).

II. MATERIEL ET METHODES

1. Design, population et critères d'inclusion

Il s'agit d'une étude rétrospective cherchant à évaluer les performances diagnostiques de modèles radiomiques hybrides d'apprentissage automatique répondant à la question de la classification binaire entre glioblastome et métastase cérébrale en IRM pré-thérapeutique.

Les critères d'inclusion sont les suivants :

- (1) Sujets souffrant d'un glioblastome (GBM) unifocal, multifocal ou multicentrique, primitif (*IDH wildtype*) ou secondaire (*IDH mutant*) ou d'une ou plusieurs métastases cérébrales (MC) ;
- (2) Diagnostic de certitude histologique disponible ;
- (3) IRM cérébrale pré-thérapeutique et pré-biopsie avec séquences 3D ou 2D T1 injectée et 3D ou 2D T2 FLAIR ;
- (4) Pas de chirurgie ou de lésions préexistantes traitées du site de segmentation susceptible d'altérer les données.

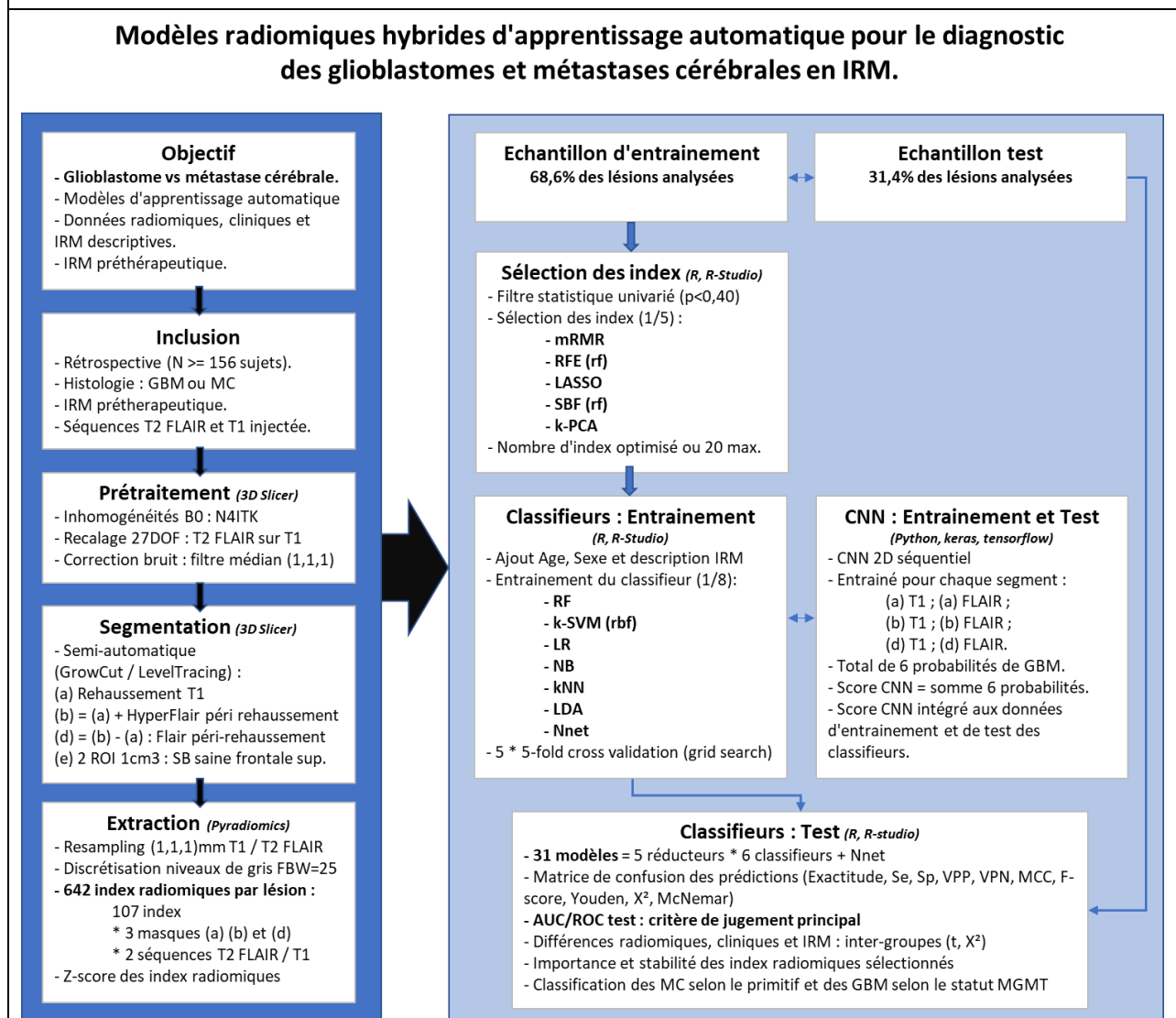
Le schéma de flux récapitulatif de la méthodologie est représenté par la figure 1.

Les sujets sont exclus de l'analyse lorsque :

- L'atteinte métastatique est exclusivement extra-axiale ;
- L'acquisition IRM est de mauvaise qualité avec entre autres artéfacts (mouvement, repliement, distorsion de champ etc.) ;
- Un traitement et/ou une biopsie a été réalisé.
- L'atteinte est un gliome diffus du tronc cérébral pédiatrique, cette entité pouvant être considérée comme borderline et non représentative des GBM en général.

L'âge, le sexe, le statut de méthylation MGMT pour les GBM, le statut IDH, et l'origine des métastases étaient recueillis.

Figure 1. Diagramme de flux du design de l'étude.



2. Objectifs

L'objectif principal est la génération, l'entraînement, et la validation de plusieurs modèles d'apprentissage automatique, entraînés avec des données radiomiques, l'âge, le sexe, la description qualitative IRM, et augmentés par un CNN, ayant pour but de différencier un glioblastome d'une métastase cérébrale sur une IRM pré-thérapeutique.

Le critère de jugement principal est l'analyse des courbes ROC et leurs AUC sur l'échantillon test.

Plusieurs objectifs secondaires sont liés à ce travail :

- (1) Établir un profil radiomique compréhensible des métastases cérébrales et du glioblastome en évaluant l'importance des index radiomiques sélectionnés par les modèles puis en proposant une correspondance avec des caractéristiques morphologiques ou histologiques ;
- (2) Évaluer la stabilité des index radiomiques ;
- (3) Evaluer la capacité du modèle à classer les métastases selon leurs primitifs sous la forme binaire, origine pulmonaire versus non pulmonaire ;
- (4) Evaluer la capacité du modèle à classer les glioblastomes selon le statut de méthylation MGMT.

Un nombre de sujets nécessaires est estimé à 156 avec 52 patients pour l'échantillon de test et 104 sujets pour l'échantillon d'entraînement, pour une puissance de 80%, un risque de premier ordre de 5%, une AUC minimale de 80%, et une exactitude minimale de classement sur l'échantillon test de 70%.

3. Matériel

Pour chaque sujet inclus, le centre d'origine, le modèle d'IRM, et les paramètres de l'acquisition de chaque séquence T1 injectée et T2 FLAIR sont relevés.

4. Prétraitement

Le prétraitement se base sur les images acquises en format *dicom*, en utilisant le logiciel *3D Slicer* v.4.10.2. Les corrections d'inhomogénéités du champ magnétique B_0 sont effectuées par l'extension N4ITK. Le recalage des images est réalisé par une transformation affine et b-spline avec 27 degrés de liberté, de la séquence T2 FLAIR en considérant la séquence T1 injectée comme référence. La réduction du bruit est réalisée à l'aide d'un filtre médian de 1mm.

5. Segmentation

La segmentation semi-automatique est réalisée à l'aide du logiciel *3D Slicer* par l'algorithme *GrowCut* pour le rehaussement T1 et par l'algorithme *GrowCut* ou *LevelTracing* pour l'anomalie de signal FLAIR péri-rehaussement.

Les régions d'intérêt segmentées sont :

(a) La lésion rehaussée en séquence T1 injectée, comprenant une éventuelle région nécrotico-hémorragique non rehaussée. Lorsque la lésion n'était pas rehaussée, la segmentation s'effectuait sur l'hyposignal T1 tumoral ;

(b) La lésion rehaussée (a) et l'hypersignal FLAIR péri-rehaussement contigu ;

(d) L'hypersignal FLAIR péri-rehaussement correspondant à la soustraction de la région (a) à la région (b) ;

(e) deux régions de référence symétriques d'environ 1 cm³ au sein de la substance blanche profonde saine frontale supérieure en séquence FLAIR, à distance de toute lésion, homolatérale (e1) et controlatérale (e2) à la tumeur principale pour évaluer la stabilité des index non morphologiques.

Les bordures des ROI segmentées étaient agrandies de 3x3x3mm afin d'inclure avec certitude l'ensemble du rehaussement T1 et minimiser un éventuel effet de volume partiel. Un filtre médian de smoothing 3mm était finalement appliqué aux ROI segmentées, permettant notamment d'éliminer les voxels « *outliers* » en marge.

Les régions d'intérêt sont ensuite reportées sur les deux séquences pour l'extraction des données. Lorsque les lésions sont multiples, en particulier pour les métastases cérébrales, au maximum 3 lésions sont mesurées pour éviter le risque de *clustering* des données et d'*overfitting* à l'entraînement tout en permettant d'augmenter les données. Toutes les lésions devaient être supra centimétriques.

Les lésions satellites des GBM cerclées par l'hypersignal FLAIR péri-rehaussement sont incluses dans la segmentation de la lésion principale dans la région (b).

La segmentation est répétée à l'aveugle pour 30 lésions tirées au sort par le même opérateur après une période d'oubli de 2 mois.

6. Extraction des index radiomiques

Les index radiomiques sont calculés pour chacune des régions segmentées sur les séquences T1 injectée et T2 FLAIR. Cette extraction est effectuée aux moyens de l'intégration *Pyradiomics v.2.2.0 (Python v.2.7.13)* au sein du logiciel *3D Slicer v.4.10.2*, suivant les recommandations établies par l'Image Biomarker Standardization Initiative (IBSI) (77).

Les séquences sont rééchantillonnées de manière isotropique à 1x1x1mm pour les séquences T1 injectée et T2 FLAIR. Les niveaux de gris sont discrétisés de manière absolue à 25 (FBW = 25).

Pour chaque région d'intérêt et chaque séquence, 107 index sont calculés : 14 morphologiques, 93 de premier ordre, deuxième ordre et d'ordres supérieurs. Il n'était pas effectué d'exploration par transformée des index radiomiques compte-tenu du petit effectif. Avant la réduction de dimension, le jeu d'index radiomiques des régions (a), (b) et (d), pour les deux séquences d'acquisition, correspond donc à 6 fois 107, soit 642 index radiomiques.

7. Réduction de dimensionnalité

Les index sont standardisés par un loi centrée réduite (Z-Score). Les sujets sont alors randomisés soit dans l'échantillon d'entraînement (2/3) soit dans l'échantillon test (1/3). Toutes les mesures d'un sujet sont obligatoirement dans le même échantillon pour éviter une fuite de données. Un premier filtre statistique appliqué sur l'échantillon d'entraînement conservait les index avec $p < 0,40$ par test de Student.

Puis, une sélection des index radiomiques informatifs est effectuée par un algorithme parmi les cinq suivants :

- l'algorithme *Minimum Redundancy Maximum Relevance* (mRMR), se basant sur la matrice d'information mutuelle, en limitant manuellement aux 15 premiers index.

- l'algorithme *Recursive Feature Elimination* (RFE), englobant un algorithme de type Random Forest, en se limitant aux index minimisant l'écart-type des résidus du modèle (Residual Mean Square Error, RMSE), parmi les 20 premiers index par rang.
- l'algorithme *Least Absolute Shrinkage and Selection Operator* (LASSO), effectuant une régression logistique pénalisant les variables peu informatives, et en conservant le nombre optimal inférieur ou égal à 20.
- l'algorithme *Kernel Principal Component Analysis* (K-PCA), transformant les données d'entraînement et de test en composantes principales par une fonction de base radiale (Gaussian Radial Basis) explicatives de la variance de l'échantillon. Le nombre de composantes principales est de 15.
- Ou l'algorithme *Selection By Filter* (SBF), réalisant une réduction statistique en englobant un algorithme Random Forest, et en effectuant une comparaison univariée des moyennes de index retenus (t-test, Wilcoxon). Le nombre d'index retenus est le nombre optimal de l'algorithme inférieur ou égal à 20.

Cette sélection est effectuée avec le langage R v.3.6.2 et le logiciel R-Studio v.1.3.1056.

La distribution normale des variables sélectionnées est évaluée avec le test de Shapiro-Wilk pour le meilleur modèle.

L'âge et le sexe sont ajoutés aux données analysées ainsi qu'un score CNN basé sur une prédiction par analyse directe des images en 2D, détaillé dans le chapitre dédié ci-après.

La description IRM des lésions y est également associée, en décrivant qualitativement plusieurs caractéristiques d'intérêt :

- (1) La présence d'une atteinte profonde se caractérisant par une lésion T1 ou FLAIR infiltrant les noyaux gris centraux, la région périventriculaire, le corps calleux, le tronc cérébral ou le vermis.
- (2) La présence d'une atteinte corticale ou juxta-corticale, définie par une atteinte en T1 du cortex et/ou de la jonction cortico-sous-corticale.
- (3) L'atteinte sous tentorielle, caractérisée par toute extension sous tentorielle d'une lésion sus tentorielle ou simplement une lésion dont l'épicentre est en fosse postérieure.
- (4) La présence d'un rehaussement T1 perceptible, en considérant tout hypersignal T1 lésionnel comme rehaussement en l'absence de séquence non injectée.

- (5) Le nombre de lésions, caractérisé par tout rehaussement T1 tumoral perceptible, en comptant les lésions satellites des GBM comme lésion individualisée.
- (6) La présence d'un franchissement de l'hypersignal FLAIR au niveau de la ligne médiane, cherchant en particulier à souligner l'atteinte controlatérale via le splenium ou le genou du corps calleux.

Les informations confondantes sont masquées (statut MGMT, primitif de la métastase, numéro d'inclusion etc.) pour l'entraînement ou le test.

8. Entraînement et validation du modèle de classification

Plusieurs types d'algorithmes de classification sont entraînés sur les index radiomiques, la description IRM et le score CNN :

- (1) Une forêt aléatoire ou Random Forest (RF),
- (2) Un séparateur à vastes marges (SVM),
- (3) Un modèle de régression logistique (LR),
- (4) Un modèle bayésien naïf (NB),
- (5) Un classifieur selon les k plus proches voisins (kNN),
- (6) Une analyse discriminante linéaire (LDA)
- (7) Un réseau de neurone monocouche – multi-neurones : NNet.

Ces modèles sont réalisés en langage R v.3.6.2 en s'appuyant notamment sur l'environnement du package caret v.6.0-84 et des packages correspondant à chaque algorithme. La structure du code de ces modèles est résumée dans le tableau 1.

L'apprentissage est effectué sur l'échantillon d'entraînement avec une validation croisée à 5 blocs répétée 5 fois, avec calcul de l'exactitude, de la *Receiver Operator Curve* (ROC), et l'aire sous la courbe correspondante (AUC). Cette validation croisée est répétée 5 fois avec un nouveau tirage au sort des blocs pour minimiser le risque lié aux données proches (*data clustering*) notamment lorsqu'il y a plusieurs lésions mesurées chez le même patient, pouvant exposer à un surajustement du modèle.

Les hyperparamètres des différents classifieurs sont optimisés par l'algorithme de recherche aléatoire par quadrillage (*grid search*) lors de la validation croisée et sont répertoriés dans le tableau 2.

Tableau 1. Exemple de la structure résumée du code d'un modèle de classification conventionnel.

```
Load libraries: caret, CaTools, pROC, e1071, readxl, classifair
Tune hyperparameters & stats
Import & factor dataset
Split dataset: training & test
Define training control settings
Define classifier settings
Train classifier
Summarize classifier's training performance & tuning
Make predictions: classes & probabilities
Set up the confusion matrix
Compute statistics:
  Accuracy
  Recall = Sensitivity
  Specificity
  Precision = Positive Predictive Value
  Negative Predictive Value
  Chi2
  McNemar
  Flscore
  Matthews Correlation Coefficient
  ROC, AUC
  Youden
Plot and compare training ROC and test ROC.
Compute features importance if avail.
End
```

Tableau 2. Valeurs des hyperparamètres des différents classifieurs, optimisés par *grid search*.

Classifieur	Hyperparamètres	Package (<i>Method</i>)
RF	<ul style="list-style-type: none"> - Nombre d'arbres - Critère de décision à chaque nœud - Nombre de branches de chaque arbre (profondeur) - Nombre maximum de feuilles - Nombre de variables à chaque nœud (<i>mtry</i>) - Taille du sous-ensemble pour chaque arbre - Bootstrap 	Caret, RandomForest (<i>rf</i>)
SVM	<ul style="list-style-type: none"> - Régularisation C - γ ou ν - Kernel 	Caret (<i>svmRadial</i>)
LR	<ul style="list-style-type: none"> - Type de régularisation (pénalité) - Itérations - Seuil de régression 	Caret (<i>LogitBoost</i>)
NB	<ul style="list-style-type: none"> - Distribution des probabilités (<i>usekernel</i>) - Lissage laplacien (<i>laplace</i>) 	Caret, naivebayes (<i>naive_bayes</i>)
kNN	<ul style="list-style-type: none"> - nombre maximal de points voisins considérés (<i>kmax</i>) - distance de Minkowski entre deux points (<i>distance</i>) - kernel (<i>kernel</i>) 	Caret, kknn (<i>kknn</i>)
LDA	<ul style="list-style-type: none"> - N/A 	Caret, MASS (<i>lda</i>)
MLP NNet	<ul style="list-style-type: none"> - Nombre de neurones cachés (<i>size</i>) - Régularisation par dégradation des poids (<i>decay</i>) 	Caret, nnet (<i>nnet</i>)

9. Entraînement du CNN en parallèle

En parallèle, un modèle basé sur un réseau de neurones convolutifs 2D séquentiel calcule une probabilité d'appartenance à chacune des classes par l'analyse directe d'images. Pour cela, une image par millimètre environ est extraite dans le plan axial pour la séquence T1, pour les trois régions segmentées. En FLAIR, le même processus est réalisé mais avec une image tous les 4mm environ, pour s'adapter à la résolution axiale. Pour les données test, une image par région et par séquence est proposée au modèle, située à mi-hauteur de la lésion segmentée en (a).

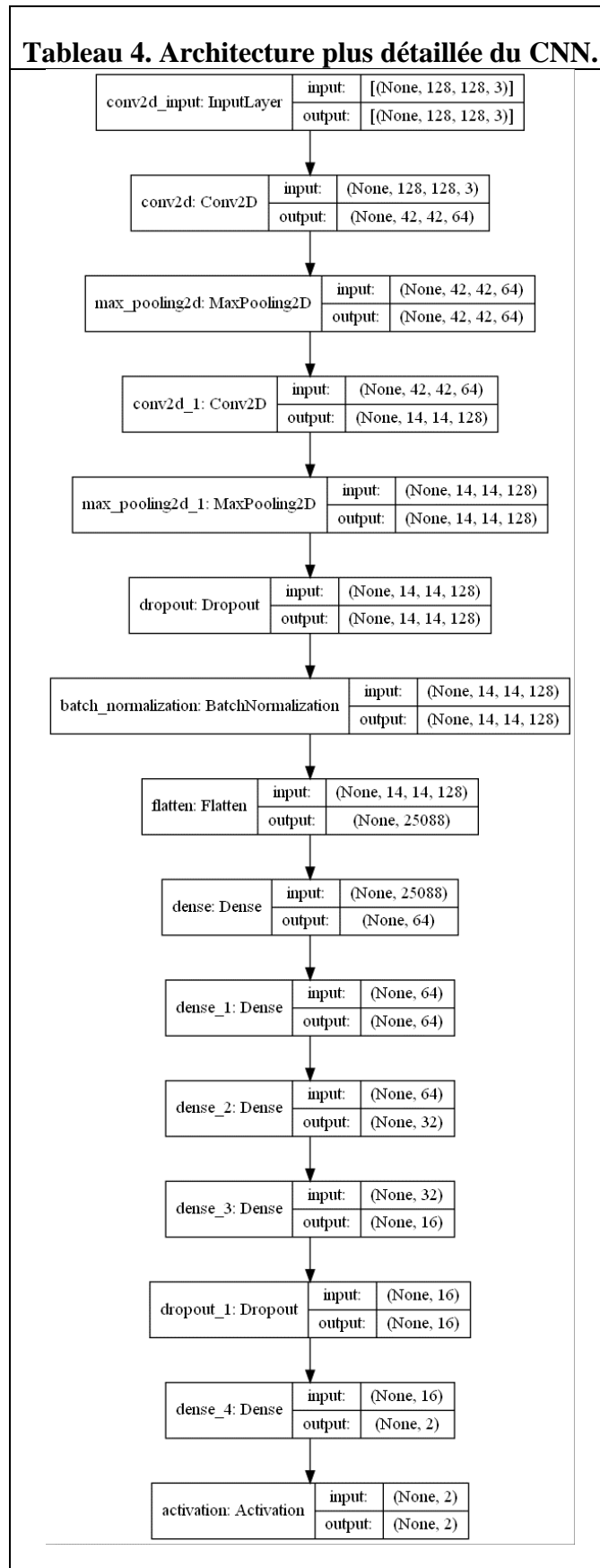
Le modèle est développé en Python v.3.8 avec l'intégration Keras v2.4.3 Tensorflow v.2.4.1 et s'inspire du travail d'optimisation de D'Souza et al. dédié aux petits échantillons (129). L'apprentissage utilise l'échantillon d'entraînement (2/3) avec un échantillon interne de validation d'environ 25% des mesures totales. La probabilité de classe GBM (versus MC en probabilités mutuellement exclusive) des trois régions d'intérêt (a), (b) et (d) pour 2 séquences, c'est à dire 6 probabilités, sont sommées, puis centrées et réduites. Ceci permet de garder la même dimension que les index radiomiques après Z-score. Ce score est ensuite associé au reste des données après la sélection des index radiomiques. Le score est propagé à l'échantillon d'entraînement en évitant le surapprentissage par arrêt et restauration des poids optimaux lorsque la fonction de perte de validation n'est plus optimisée après 10 cycles.

L'architecture est schématisée dans les tableaux 3 et 4.

Tableau 3. Architecture simplifiée du CNN. Type de couche associé à au nombre de filtres et leur taille ou au nombre de neurones. Conv2D : couche de convolution 2D ; MP : couche de Max pooling ; FC : couche de neurones connectés. Input avec taille de la matrice d'image.

Input, 128, 3 → Conv2D, 64, 3 → MP2 → Conv2D, 128, 3 → MP2 → FC, 64 → FC, 32 → FC, 32 → FC, 10 → Activation Softmax.

Tableau 4. Architecture plus détaillée du CNN.



Les hyperparamètres du CNN sont résumés dans le tableau 5.

Tableau 5. Hyperparamètres du CNN.	
Learning rate	0.01
Regularization L2	0.01
Decay	1.10^{-6}
Optimizer	Stochastic Gradient Descent
Batch size	32
Metric	Accuracy (binary)
Epochs	Maximum 50 Stop après 10 epochs sans optimisation de fonction de perte de validation et restauration des poids optimaux

10. Performances prédictives des modèles

Les modèles entraînés sont confrontés aux données de l'échantillon test. Les performances prédictives sont évaluées par une courbe ROC avec calcul de l'AUC. Une comparaison des courbes ROC à l'entraînement et en test est effectuée pour dépister un éventuel sur ou sous apprentissage.

Une matrice de confusion permet de calculer sensibilité, spécificité, valeurs prédictives positives et négatives, ainsi que l'exactitude (tableau 6).

Le F-score établit la moyenne de la sensibilité et de la valeur prédictive positive de la classification : une valeur maximale de 1 correspondra à une parfaite similarité entre les classes réelles et celles prédites. Cette évaluation est complétée par l'index de Youden, calculé par la somme de la sensibilité et de la spécificité moins un, pour tous les points de la courbe ROC, et sa valeur maximale définit l'index J de Youden optimal retenu, avec une valeur maximale de 1 correspondant à une parfaite sensibilité et spécificité.

Le coefficient de corrélation de Matthews (MCC) permet également de mesurer la performance lors d'une classification binaire, et s'appuie sur le calcul de la variable test du X^2 . Il a pour intérêt de prendre en compte les vrais négatifs qui sont omis dans le calcul du F-score.

Le test du Chi2 et de McNemar avec correction de Yates pour les matrices 2x2 évaluent l'hypothèse nulle H_0 d'une distribution indépendante des classes réelles et prédites, en suivant une loi du X^2 à un degré de liberté.

Pour le meilleur modèle, les performances sont testées à nouveau en ne sélectionnant que les données cliniques (âge et sexe), puis uniquement les données descriptives IRM, le score CNN ou encore les index radiomiques pour estimer le poids de chaque type de données.

L'importance des index radiomiques est calculé en les retirant un par un du meilleur modèle. Puis, l'impact de ces retraits est évalué en calculant la variation relative d'AUC d'entraînement par rapport à l'AUC de référence.

Tableau 6. Description d'une matrice de confusion associée aux mesures statistiques d'intérêt pour une tâche de prédiction.

	Classe réelle			
	Classe réelle positive	Classe réelle négative		
Classe prédite positive	Vrais positifs	Faux positifs	Valeur prédictive positive (VPP, Précision) : $VPP = \frac{\Sigma \text{Vrais positifs}}{\Sigma \text{Classe prédite positive}}$	Coefficient de corrélation de Matthews (MCC) : $MCC = \sqrt{\left(\frac{\chi^2}{\text{Echantillon}}\right)}$
Classe prédite négative	Faux négatifs	Vrais négatifs	Valeur prédictive négative (VPN) : $VPN = \frac{\Sigma \text{Vrais négatifs}}{\Sigma \text{Classe prédite négative}}$	F-score : $F - \text{Score} = 2 * \frac{\text{Précision} * \text{Sensibilité}}{\text{Précision} + \text{Sensibilité}}$
	Sensibilité : $Se = \frac{\Sigma \text{Vrais positifs}}{\Sigma \text{Classe réelle positive}}$	Spécificité : $Sp = \frac{\Sigma \text{Vrais négatifs}}{\Sigma \text{Classe réelle négative}}$	Exactitude : $\text{Exactitude} = \frac{\Sigma \text{Vrais positifs} + \Sigma \text{Vrais négatifs}}{\text{Echantillon}}$	Index J de Youden: $J = \text{Sensibilité} + \text{Spécificité} - 1$

11. Autres tests statistiques

Le test t paramétrique de Student estime la différence des moyennes des deux groupes GBM et MC pour les index quantitatifs. Le filtre statistique utilisant le test t n'est réalisé que sur les mesures de l'échantillon d'entraînement. Le test de Shapiro-Wilk permet d'évaluer le caractère normal de la distribution de ces index quantitatifs, il est réalisé pour les index sélectionnés par le meilleur modèle. Le test du X^2 recherche une différence de distribution entre les deux groupes pour les variables catégorielles.

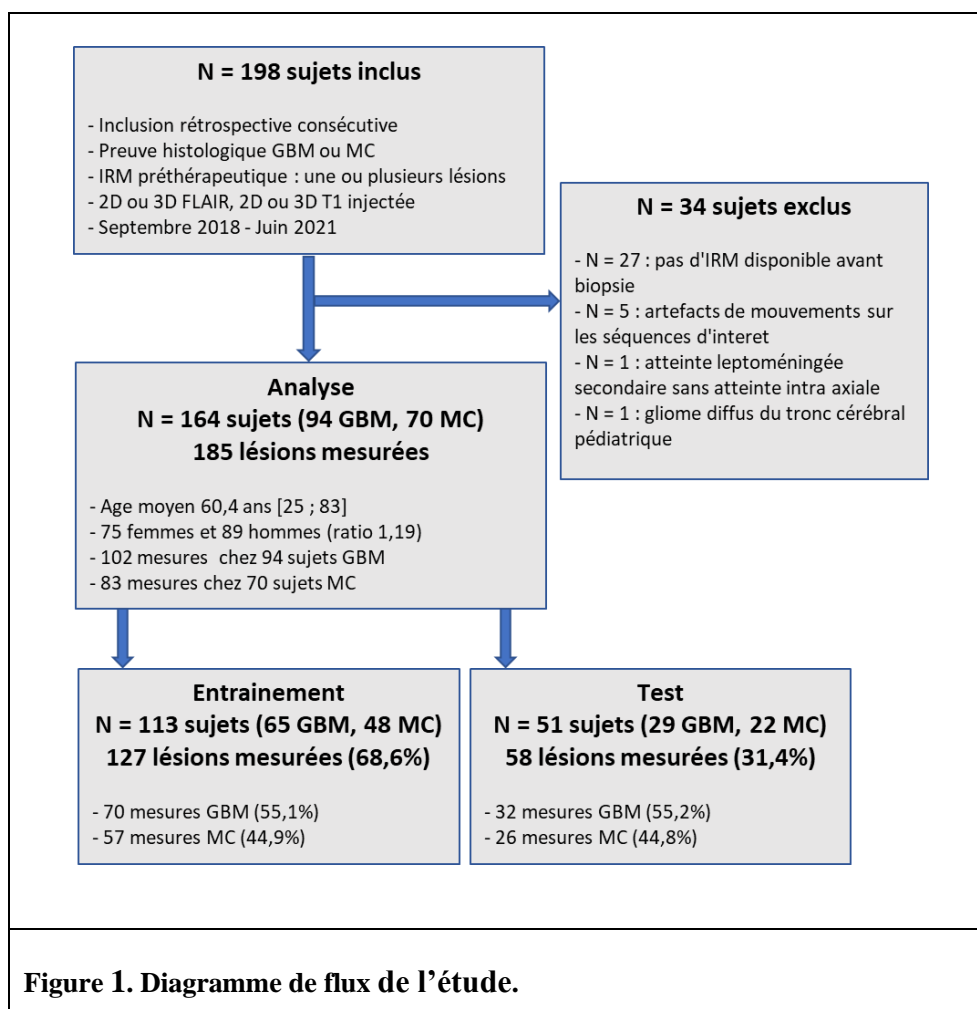
Le coefficient de corrélation de concordance (CCC) permet d'évaluer la stabilité des index dans les segments de référence en substance blanche saine homo et controlatérale. Le CCC est également calculé après la répétition de la segmentation dans les régions lésionnelles.

III. RESULTATS

1. Population et échantillons

198 sujets ont été inclus entre septembre 2018 et Juin 2021 : les données descriptives démographiques et IRM sont résumées dans le diagramme de flux (figure 1) et les tableaux 1 et 2. 34 sujets furent exclus de l'analyse : 27 n'avaient pas d'IRM exploitable pré-biopsie ; 5 sujets avaient une IRM avec artéfacts de mouvement ; un sujet présentait une atteinte unique de la leptoméninge et un patient pédiatrique souffrait d'un gliome diffus avec atteinte du tronc cérébral.

164 sujets étaient éligibles à l'analyse : 94 (57,3%) étaient porteurs d'un glioblastome et 70 (42,7%) souffraient d'une ou plusieurs lésions cérébrales secondaires. L'âge moyen de la cohorte était de 60,4 ans \pm 12,2 (intervalle de 25 - 83 ans). Le sexe ratio était de 1,19 avec 89 hommes pour 75 femmes.



Concernant les sujets atteints de GBM, l'âge moyen était de 61,6 ans \pm 12,6 (25 - 83 ans), contre 58,8 ans \pm 11,5 pour ceux pâtissant d'une métastase cérébrale (30 - 80 ans) (tableau 1). Le sexe ratio était de 1,35 pour les GBM (54 hommes, 40 femmes) et de 1 pour les MC (35 hommes, 35 femmes). 64/94 (68,1%) sujets GBM présentaient une lésion unique et 30/94 (31,9%) deux lésions ou plus, en considérant les rehaussements satellites comme lésion. Le nombre moyen de lésions GBM était de 1,56 \pm 1,11. 47/70 (67,1%) des sujets MC présentaient une lésion unique et 23/70 (32,9%) une atteinte multiple, correspondant à une moyenne de 1,5 \pm 0,86.

Tableau 1. Caractéristiques des populations GBM et MC selon les échantillons d'entraînement et de test. Les valeurs sont données selon la moyenne (écart-type). La comparaison des moyennes est réalisée par le test t de Student et la comparaison des catégories par le test du X ² avec correction de Yates.			
	GBM	MC	p-value (X² ou t)
Cohorte entière (N=164)			
Nombre de sujets	94/164 (57,3%)	70/164 (42,7%)	
Age	61,6 (12,6)	58,8 (11,5)	0,470
Médiane et intervalle d'âge	64 [25 ; 83]	59 [30 ; 80]	
Sexe h / f	54 / 40	35 / 35	0,430
Ratio sexe h / f	1,35	1	
Entraînement (N=113)			
Nombre de sujets	65/94 (69,1%)	48/70 (68,6%)	1
Age	59,3 (13,4)	57,8 (11,5)	0,51
Sexe h / f	36/29	22/26	1
Ratio sexe h / f	1,24	0,85	
Nombre de mesures (N=127)	70/127 (55,1%)	57/127 (44,9%)	
Test (N=51)			
Nombre de sujets	29/94 (30,9%)	22/70 (31,4%)	
Age	66,7 (8,9)	61 (11,2)	0,055
Sexe h / f	18/11	13/9	1
Ratio sexe h / f	1,64	1,44	
Nombre de mesures (N=58)	32/58 (55,2%)	26/58 (44,8%)	

Au total, 185 lésions étaient mesurées : 102 (55,1%) lésions GBM et 83 (44,9%) lésions MC. Ainsi, pour 8/94 (8,5%) patients GBM et 12/70 (17,1%) patients MC, deux mesures étaient effectuées avec un maximum de 3 mesures pour un seul sujet MC.

L'échantillon d'entraînement comprenait 113/164 sujets et 127/185 (68,6%) mesures, avec 70/127 (55,1%) mesures GBM et 57/127 (44,9%) mesures MC. Lorsque plusieurs mesures

étaient effectuées, celles-ci devaient être dans le même échantillon pour éviter une fuite de données entre échantillons. L'échantillon test incluait 51/164 patients et 58/185 (31,4%) mesures avec 32/58 (55,2%) cibles GBM et 26/58 (44,8%) cibles MC.

Tableau 2. Caractéristiques descriptives des lésions GBM et MC mesurées. Les valeurs sont données selon la moyenne (écart-type). La comparaison des moyennes est réalisée par le test t de Student et la comparaison des catégories par le test du X ² avec correction de Yates.			
	GBM	MC	p-value (X² ou t)
Cohorte entière			
Nombre de mesures	102	83	
IRM qualitatif :			
Atteinte profonde	54/102 (52,9%)	28/83 (33,7%)	0,014
Atteinte corticale ou juxta corticale	80/102 (78,4%)	80/83 (96,4%)	0,0008
Atteinte sous-tentorielle	3/102* (2,9%)	9/83 (10,8%)	0,061
Rehaussement perceptible	96/102 (94,1%)	83/83 (100%)	0,034**
Croisement FLAIR	27/102 (26,5%)	20/83 (24,1%)	0,84
Acquisition IRM :			
3T	39/102	25/83	0,318
1.5T	63/102	58/83	
3D T1	102/102	83/83	1
3D FLAIR	37/102	19/83	0,07
Mesures spatiales :			
Grand axe 3D T1 (mm)	52,8 ± 17,4	40,2 ± 17,1	1.10 ⁻⁶
Volume T1 (cm ³)	33,2 ± 29,4	21,7 ± 23,1	0.003
Grand axe 3D FLAIR	77,9 ± 22,8	78,4 ± 25,8	0,89
Volume FLAIR	77,0 ± 54,9	89,0 ± 63,9	0,18
Ratio T1/FLAIR	0.45 ± 0.22	0.28 ± 0.22	5.10 ⁻⁷
Quantité de lésions :			
Nombre total de lésions en IRM / nombre de patients	147/94	105/70	
Nombre total de lésions en IRM par patient	1,56 (1,11)	1,5 (0,86)	0,680
Lésion unique	64/94 (68,1%)	47/70 (67,1%)	1
Deux lésions ou plus	30/94 (31,9%)	23/70 (32,9%)	
Quantité de lésions mesurées :			
Nombre de lésions mesurées	102/185 (55,1%)	83/185 (44,9%)	0,569
Nombre de lésions mesurées par patient	102/94	83/70	0,763
Mesure unique	86/94 (91,5%)	58/70 (82,9%)	0,299
Deux mesures ou plus	8/94 (8,5%)	12/70 (17,1%)	

* Trois atteintes sus-tentorielles avec extension au tronc cérébral. **Test exact de Fisher

Les lésions GBM montraient une atteinte profonde dans 54/102 (52,9%) cas contre 28/83 (33,7%) pour les mesures MC (p=0,014) (tableau 2). L'atteinte corticale ou juxta-corticale intéressait 80/102 (78,4%) des mesures GBM contre 80/83 (96,4%) des cibles MC (p=0.0008). L'atteinte sous-tentorielle des GBM concernait 3/102 (2,9%) lésions, se caractérisant par une extension sous-tentorielle, contre 9/83 (10,8%) des atteintes MC (p=0,061). Un rehaussement était perceptible pour 96/102 (94,1%) mesures GBM contre 83/83 (100%) des cibles MC (p=0,034). Le croisement de la ligne médiane par l'hypersignal FLAIR concernait 27/102 (26,5%) GBM contre 20/83 (24,1%) MC sans différence significative (p=0.84).

La lésion primitive des sujets MC, le statut IDH et MGMT des sujets GBM sont répertoriés dans les tableaux 3 et 4.

46/70 (65,7%) des sujets MC présentaient une atteinte primitive pulmonaire, tandis que pour 10/70 (14,3%) l'atteinte initiale était mammaire. D'autres origines néoplasiques étaient à la marge : primitif digestif (6/70), ovarien (4/70), prostatique (2/70), endométrial (1/70) et vésical (1/70). Les lésions MC étaient de type adénocarcinome peu à bien différenciés.

Pour 9/92 (9,7%) sujets GBM, un statut IDH mutant était constaté avec données manquantes pour 2 patients. 43/80 (51,8%) des sujets GBM présentaient un état méthylé MGMT avec données manquantes pour 14 patients.

Tableau 3. Statut de la mutation IDH et méthylation MGMT des sujets GBM.

Proportions des sujets dans la cohorte entière, à l'entraînement et lors du test. Données manquantes pour 14 sujets concernant le statut MGMT et pour 2 sujets pour le statut IDH.

	GBM	Entrainement	Test	p-value
IDH mutant / total (%)	9/92 (9,7%)	7/65	2/25	1*
MGMT méthylé / total (%)	43/80 (51,8%)	32/59	11/21	1

*Test exact de Fisher

Tableau 4. Lésions primitives des sujets MC. Proportions des sujets dans la cohorte entière, à l'entraînement et lors du test. Les lésions histologiques étaient de type adénocarcinome peu à bien différenciés.

	MC	Entraînement	Test	p-value
Origine pulmonaire	46/70 (65,7%)	35/48 (72,9%)	11/22 (50%)	0,110
Autre origine	24/70 (34,3%)	13/48 (27,1%)	11/22 (50%)	
Sein	10 (14,3%)	7	3	
Colon	6 (8,5%)	2	4	
Ovaire	4 (5,6%)	2	2	
Prostate	2 (2,8%)	1	1	
Endomètre	1 (1,4%)	1	0	
Vessie	1 (1,4%)	0	1	
Nombre de lésions pulmonaires mesurées	54/83 (65,1%)	41/57 (71,9%)	13/26 (50%)	0,09
Nombre de lésions autres mesurées	29/83 (34,9%)	16/57 (28,1%)	13/26 (50%)	

L'imagerie provenait de 20 centres avec un total de 27 IRM, 8 modèles d'IRM et 3 constructeurs différents : ces données sont résumées dans le tableau 5. 112/185 (55,6%) mesures étaient issues du centre universitaire de l'étude sur deux IRM 1.5 et 3T. Les autres origines étaient très diverses représentant chacune au maximum 6,5% de la cohorte. Toute l'imagerie T1 injectée était acquise en 3D millimétrique. 63/185 mesures concernaient une acquisition T2 FLAIR en 3D millimétrique, contre 122/185 (65,9%) en 2D 4mm en moyenne. 65/185 (35,1%) des mesures étaient issues d'une imagerie à 3T contre 120/185 (64,9%) à 1.5T.

La première source d'imagerie était une IRM Siemens Magnetom Prisma 3T correspondant à 29,7% des mesures (37 GBM, 18MC). L'acquisition 3D T1 injectée était réalisée en séquence Turbo Spin Echo SPACE (Sampling Perfection with Application optimized Contrasts by using different flip angle Evolutions, Siemens Healthcare), avec Temps de Répétition (TR) = 580ms, Temps d'Echo (TE) = 15ms, matrice = 256x256, épaisseur de coupe=1mm, field of view (FOV) = 256-256). L'acquisition 3D FLAIR était réalisée en séquence Turbo Spin Echo SPACE, avec Temps d'Inversion (TI) = 1600ms, TR = 5000ms, TE = 415ms, matrice = 256x228, épaisseur de coupe=1mm, FOV = 240-256). L'agent de contraste était le DOTAREM® (Guerbet) 0,5 mmol/mL (acide gadotérique) en solution injectable, 0,2 mL/kg.

La seconde source d'imagerie était une IRM Siemens Magnetom Aera 1.5T représentant 25,9% des mesures (20 GBM, 28 MC). L'acquisition 3D T1 injectée était réalisée en séquence

Echo de gradient rapide MP-RAGE (Magnetization Prepared - RAPid Gradient Echo), avec TR = 2200ms, TE = 2.67ms, TI = 900ms, matrice = 256x192, épaisseur de coupe=1mm, field of view (FOV) = 187-250). L'acquisition T2 FLAIR était réalisée en séquence Spin Echo axiale, avec TI = 2500ms, TR = 9000ms, TE = 78ms, matrice = 320x168, épaisseur de coupe=1mm, FOV = 180-240). L'agent de contraste était le DOTAREM® (Guerbet) 0,5 mmol/mL (acide gadotérique) en solution injectable, 0,2 mL/kg.

Tableau 5. Description des centres et machines à l'origine des données analysées. Au total, 20 centres, 27 machines, 8 modèles et 3 constructeurs différents.

Centre	Sujets (mesures)	%mesures	Machine	T1 injectée	T2 FLAIR
CHRU Gui de Chauliac - Montpellier	41 (48) 49 (55) 8 (9) Tot. 98 (112)	25,9% 29,7% 4,9%	Siemens Aera 1.5T Siemens Prisma 3T Siemens Skyra 3T	3D 1mm 3D 1mm 3D 1mm	2D 4mm 3D 1mm 2D 3-4mm
CHRU Arnaud de Villeneuve - Montpellier	1 (1) 1 (1) Tot. 2 (2)	0,5% 0,5%	Siemens Aera 1.5T Siemens Sola 1.5T	3D 1mm 3D 1mm	2D 5mm 2D 4mm
CH Béziers	2 (2) 11 (12) 1 (1) Tot. 14 (14)	1,1% 6,5% 0,5%	Siemens Aera 1.5T Siemens Sola 1.5T Siemens Sempra 1.5T	3D 1mm	2D 4mm
CCIM Perpignan	2 (2)	1,1%	Siemens Aera 1.5T	3D 1mm	2D 4mm
CH Aurillac	1 (1)	0,5%	Siemens Aera 1.5T	3D 1mm	2D 4mm
CH Carcassonne	1 (1)	0,5%	Siemens Sola 1.5T	3D 0.6mm	2D 4mm
CH Mende	3 (3)	1,6%	Siemens Aera 1.5T	3D 0.5mm	2D 4mm
CH Narbonne	4 (5)	2,7%	GE Optima 1.5T	3D 1.2mm	2D 5mm
GHU Créteil	1 (1)	0,5%	Siemens Skyra 3T	3D 0.9mm	2D 3mm
Clinique du Parc - Montpellier	2 (2)	1,1%	Siemens Aera 1.5T	3D 0.9-1mm	2D 4-4.5mm
Clinique Kennedy - Nimes	2 (3)	1,6%	Philips Ingenia 1.5T	3D 1mm	2D 5mm
Clinique Millénaire - Montpellier	2 (2)	1,1%	Siemens Sola 1.5T	3D 1mm	2D 5mm
Clinique Mutualiste Beausoleil	2 (2)	1,1%	Philips Ingenia 1.5T	3D 1.4mm	2D 5mm
CTR Millau	2 (2)	1,1%	Siemens Aera 1.5T	3D 0.5mm	2D 0.5-4mm
GIE IRM St Eloi	1 (2) 6 (6) Tot. 7 (8)	1,1% 3,2%	Siemens Aera 1.5T GE Optima 1.5T	3D 1mm 3D 1.4mm	2D 4mm 2D 5mm
GIE Etang de Thau	3 (4)	2,2%	Siemens Aera 1.5T	3D 1mm	2D 5mm
ICM Val d'Aurelle - Montpellier	7 (8) 8 (8) Tot. 15 (16)	4,3% 4,3%	Siemens Sola 1.5T GE Optima 1.5T	3D 0.5-1.5mm 3D 1.2mm	3D 1.1mm 2D 5mm
Les Franciscaines - Nimes	1 (2)	1,1%	Philips Ingenia 1.5T	3D 1mm	2D 5mm
Médipôle - Perpignan	1 (1)	0,5%	Siemens Altea 1.5T	3D 1mm	2D 4mm
Pôle Santé - Lunel	1 (1)	0,5%	Siemens Aera 1.5T	3D 0.6mm	2D 4mm

2. Performances des modèles pour la tâche de classification

2.1. Résumé des meilleures performances des différents modèles

642 index radiomiques furent extraits des régions d'intérêt sur 185 mesures. Une analyse univariée sur l'échantillon d'entraînement permettait de réduire à 425 les index radiomiques présentant une association avec $p < 0.40$.

31 modèles d'apprentissage automatique furent entraînés sur une sélection d'index radiomiques, auxquels furent adjoints le score de probabilité du CNN, les variables descriptives IRM, ainsi que l'âge et le sexe. Les principaux résultats de test et d'entraînement sont colligés dans les tableaux 6 et 7 puis plus en détails dans les chapitres correspondants. Les modèles sont nommés avec le format « réducteur-nombre d'index_classifieur ».

Tableau 6. Valeurs d’AUC sur les échantillons de test et d’entraînement des classifieurs en fonction des réducteurs employés, avec intervalle de confiance à 95%. NNet n’a pas été soumis aux algorithmes de sélection d’index.

AUC (Test)					
Classifieur	Réducteur				
	mRMR15	RFE11	LASSO15	SBF10	PCA15
RF	0.897	0.912	0.877	0.844	0.852
	[0.805 ; 0.988]	[0.829 ; 0.996]	[0.780 ; 0.973]	[0.733 ; 0.956]	[0.753 ; 0.950]
SVM	0.905	0.928	0.891	0.865	0.873
	[0.818 ; 0.992]	[0.853 ; 1]	[0.804 ; 0.977]	[0.768 ; 0.963]	[0.784 ; 0.961]
LR	0.880	0.850	0.849	0.791	0.755
	[0.783 ; 0.977]	[0.738 ; 0.961]	[0.741 ; 0.957]	[0.673 ; 0.910]	[0.635 ; 0.876]
NB	0.904	0.847	0.816	0.786	0.793
	[0.817 ; 0.992]	[0.736 ; 0.959]	[0.700 ; 0.932]	[0.644 ; 0.928]	[0.672 ; 0.914]
kNN	0.903	0.887	0.859	0.844	0.815
	[0.826 ; 0.979]	[0.806 ; 0.968]	[0.770 ; 0.949]	[0.750 ; 0.937]	[0.715 ; 0.915]
LDA	0.881	0.924	0.893	0.880	0.862
	[0.798 ; 0.964]	[0.856 ; 0.992]	[0.813 ; 0.973]	[0.796 ; 0.963]	[0.753 ; 0.939]
Nnet425	0.843 [0.736 ; 0.949]				

AUC (Entraînement)					
Classifieur	Réducteur				
	mRMR15	RFE11	LASSO15	SBF10	PCA15
RF	0.935	0.929	0.924	0.885	0.808
	[0.918 ; 0.953]	[0.909 ; 0.950]	[0.903 ; 0.945]	[0.857 ; 0.914]	[0.739 ; 0.876]
SVM	0.943	0.942	0.939	0.897	0.827
	[0.926 ; 0.960]	[0.922 ; 0.962]	[0.918 ; 0.960]	[0.871 ; 0.923]	[0.794 ; 0.860]
LR	0.902	0.883	0.907	0.844	0.811
	[0.879 ; 0.924]	[0.857 ; 0.910]	[0.884 ; 0.930]	[0.813 ; 0.874]	[0.777 ; 0.845]
NB	0.906	0.896	0.808	0.837	0.736
	[0.881 ; 0.932]	[0.870 ; 0.921]	[0.771 ; 0.845]	[0.805 ; 0.869]	[0.696 ; 0.776]
kNN	0.892	0.880	0.861	0.875	0.804
	[0.865 ; 0.919]	[0.851 ; 0.909]	[0.829 ; 0.893]	[0.847 ; 0.904]	[0.769 ; 0.839]
LDA	0.902	0.901	0.912	0.828	0.768
	[0.877 ; 0.927]	[0.877 ; 0.926]	[0.889 ; 0.935]	[0.794 ; 0.861]	[0.729 ; 0.807]
Nnet425	0.887 [0.860 ; 0.914]				

Tableau 7. Valeurs d'exactitude sur les échantillons de test et d'entraînement des classifieurs en fonction des réducteurs employés, avec intervalle de confiance à 95%. NNet n'a pas été soumis à la réduction d'index.

Exactitude (Test)					
Classifieur	Réducteur				
	mRMR15	RFE11	LASSO15	SBF10	PCA15
RF	0.862	0.897	0.810	0.793	0.448
	[0.773 ; 0.951]	[0.818 ; 0.975]	[0.709 ; 0.911]	[0.689 ; 0.897]	[0.320 ; 0.576]
SVM	0.828	0.914	0.828	0.828	0.741
	[0.730 ; 0.925]	[0.842 ; 0.986]	[0.730 ; 0.925]	[0.730 ; 0.925]	[0.629 ; 0.854]
LR	0.776	0.793	0.741	0.690	0.672
	[0.669 ; 0.883]	[0.689 ; 0.897]	[0.629 ; 0.854]	[0.571 ; 0.809]	[0.552 ; 0.793]
NB	0.759	0.793	0.759	0.741	0.448
	[0.648 ; 0.869]	[0.689 ; 0.897]	[0.648 ; 0.869]	[0.629 ; 0.854]	[0.320 ; 0.576]
kNN	0.862	0.845	0.828	0.828	0.759
	[0.773 ; 0.951]	[0.752 ; 0.938]	[0.730 ; 0.925]	[0.730 ; 0.925]	[0.648 ; 0.869]
LDA	0.862	0.828	0.793	0.828	0.759
	[0.773 ; 0.951]	[0.730 ; 0.925]	[0.689 ; 0.897]	[0.730 ; 0.925]	[0.648 ; 0.869]
Nnet425	0.793 [0.689 ; 0.897]				

Exactitude (Entraînement)					
Classifieur	Réducteur				
	mRMR15	RFE11	LASSO15	SBF10	PCA15
RF	0.859	0.914	0.868	0.783	0.783
	[0.798 ; 0.919]	[0.790 ; 0.914]	[0.809 ; 0.927]	[0.711 ; 0.855]	[0.711 ; 0.855]
SVM	0.881	0.890	0.899	0.838	0.792
	[0.824 ; 0.937]	[0.835 ; 0.944]	[0.847 ; 0.952]	[0.774 ; 0.902]	[0.722 ; 0.863]
LR	0.824	0.814	0.841	0.786	0.761
	[0.757 ; 0.890]	[0.747 ; 0.882]	[0.777 ; 0.904]	[0.715 ; 0.857]	[0.687 ; 0.835]
NB	0.831	0.817	0.775	0.788	0.681
	[0.766 ; 0.897]	[0.750 ; 0.884]	[0.703 ; 0.848]	[0.717 ; 0.859]	[0.599 ; 0.762]
kNN	0.833	0.827	0.835	0.828	0.757
	[0.768 ; 0.898]	[0.761 ; 0.893]	[0.770 ; 0.899]	[0.763 ; 0.894]	[0.682 ; 0.831]
LDA	0.837	0.816	0.846	0.751	0.742
	[0.773 ; 0.902]	[0.748 ; 0.883]	[0.783 ; 0.908]	[0.676 ; 0.826]	[0.666 ; 0.818]
Nnet425	0.750 [0.639 ; 0.861]				

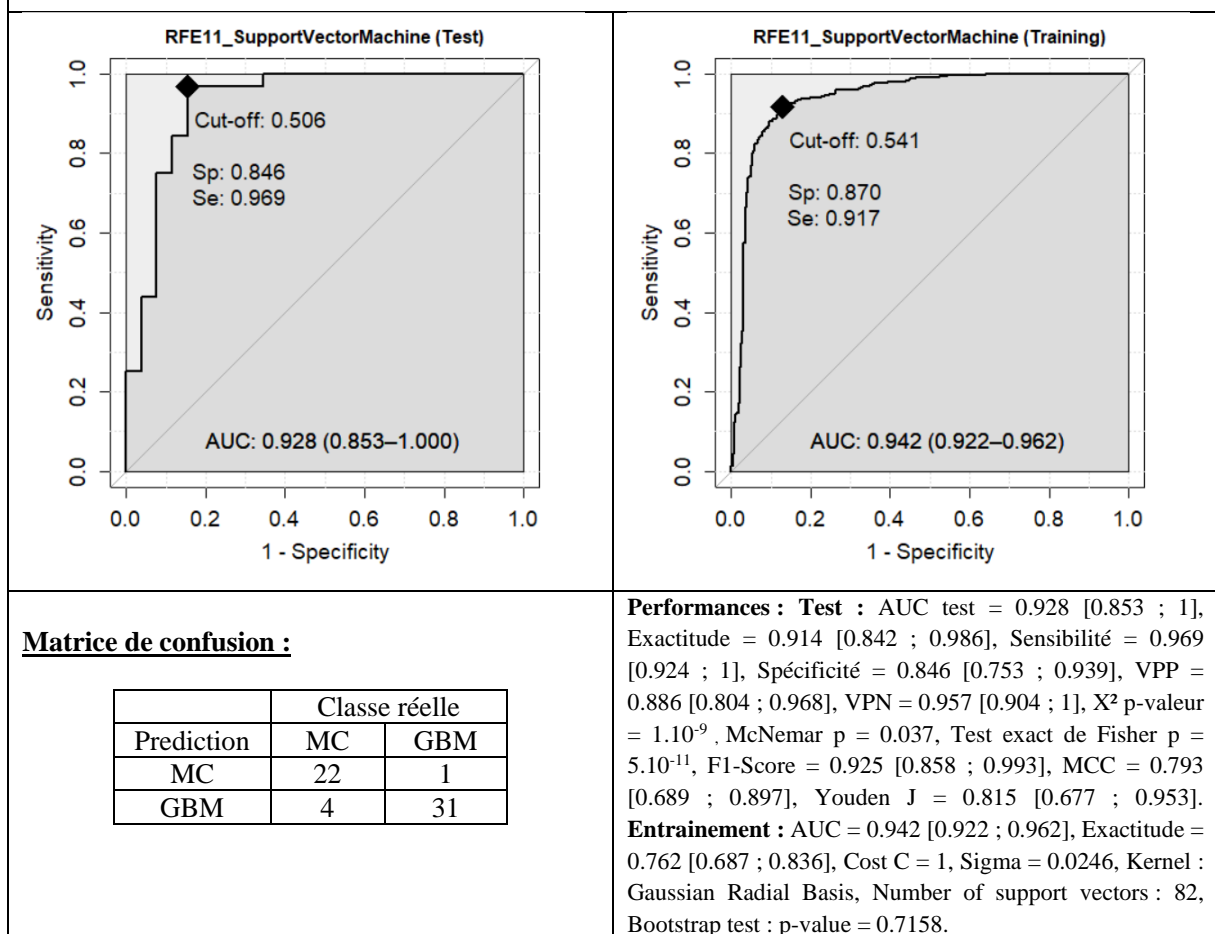
Après confrontation à l'échantillon test, le modèle le plus performant associait le réducteur Recursive Feature Elimination avec 11 index radiomiques au classifieur Support

Vector Machine (RFE11_SVM) (tableau 8). Ce modèle présentait la meilleure valeur d'AUC_{test} à 0.928 IC_{95%}[0.853 ; 1] avec une exactitude à 0.914 [0.842 ; 0.986], une sensibilité de 0.969 [0.924 ; 1], une spécificité de 0.846 [0.753 ; 0.939], une valeur prédictive positive de 0.886 [0.804 ; 0.968] et une valeur prédictive négative de 0.957 [0.904 ; 1].

La matrice de confusion de la classification montre une distribution significative (McNemar $p=0.037$, $X^2 p=1.10^{-9}$, test exact de Fisher $p=5.10^{-11}$). L'AUC à l'entraînement était de 0.942 [0.922 ; 0.962] avec une exactitude de 0.762 [0.687 ; 0.836]. Il n'y avait pas de différence entre l'AUC d'entraînement et de test (Bootstrap test : $p\text{-value} = 0.716$) (tableau 8).

Le F1-Score était de 0.925 [0.858 ; 0.993], Le coefficient de corrélation de Matthews de 0.793 [0.689 ; 0.897], et l'index de Youden J de 0.815 [0.677 ; 0.953].

Tableau 8. Courbes ROC avec AUC à l'entraînement et lors du test pour le modèle RFE11_SVM.



Lorsque le modèle RFE11_SVM n'était entraîné que sur les 11 index radiomiques, l'AUC test baissait de 11,4% à 0.822 [0.715 ; 0.929] (tableau 9). De même, lorsque le modèle SVM n'était basé que sur le score CNN, l'AUC test était amoindrie de 15,73% à 0.782 [0.655 ; 0.910].

Le modèle SVM avec uniquement l'âge et le sexe présentait un AUC test de 0.608 [0.458 ; 0.757] . Prises isolées, les données descriptives IRM par SVM correspondaient à un AUC test de 0.612 [0.468 ; 0.757].

Tableau 9. AUC test et entraînement des index radiomiques RFE11, du score CNN et des données cliniques et IRM, chacun pris isolément du reste des données, avec le classifieur SVM.				
Référence : RFE11_SVM modèle hybride plein	AUC test = 0.928 [0.853 ; 1]		AUC entraînement = 0.942 [0.922 ; 0,962]	
	AUC test univarié	AUC test au retrait du modèle hybride plein	AUC entraînement univarié	AUC entraînement au retrait du modèle hybride plein
Index radiomiques RFE11 seuls (SVM)	0.822 [0.715 ; 0.929] (-11,42%)	0.869 (-6,36%)	0.851 [0.821 ; 0.882] (-9,66%)	0.822 (-12,70%)
Score CNN seul (SVM)	0.782 [0.655 ; 0.910] (-15,73%)	0.881 (-5,06%)	0.819 [0.784 ; 0.854] (-13,06%)	0.857 (-9,02%)
Données cliniques seules (SVM) (âge et sexe)	0.608 [0.458 ; 0.757] (-34,48%)	0.918 (-1,08%)	0.533 [0.488 ; 0.578] (-43,42%)	0.945 (+0,30%)
Données descriptives IRM seules* (SVM)	0.612 [0.468 ; 0.757] (-34,05%)	0.907 (-2,23%)	0.564 [0.518 ; 0.609] (-40,13%)	0.941 (-0,10%)

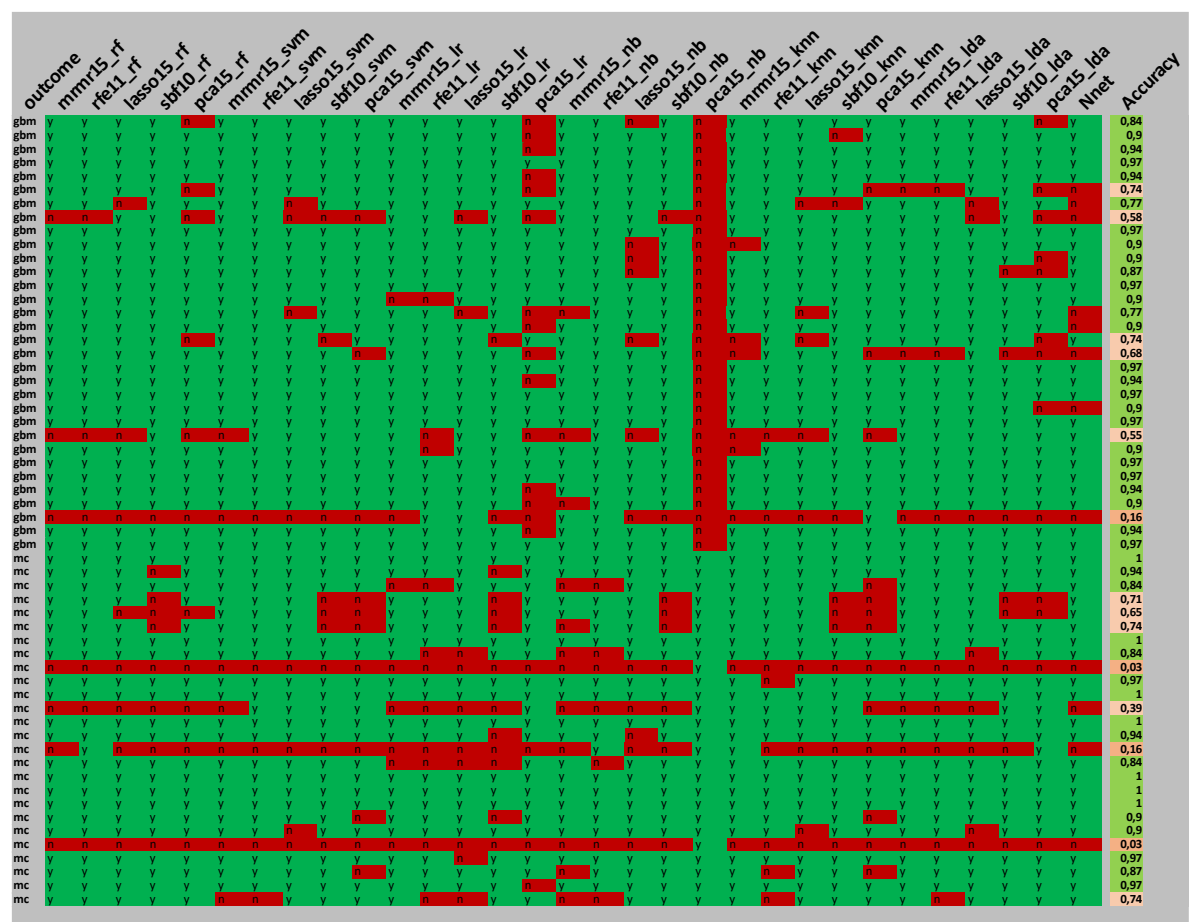
*(atteinte corticale, profonde, sous tentorielle, rehaussement, croisement Flair, nombre de lésions)

Une heatmap des prédictions permettait d'isoler 4/58 (6,9%) lésions présentant une difficulté de diagnostic, puisque moins de 25% des modèles étaient capables de correctement les classer (Tableau 10). Le meilleur modèle RFE11_SVM ne permettait pas de classer justement ces 4 lésions. Pour 44/58 (75,9%) lésions, la difficulté diagnostique était moindre, celles-ci étant correctement classées dans plus de 75% des modèles. 10/58 (17,2%) lésions présentaient une

difficulté intermédiaire, correctement classées pour 25 à 75% des modèles. L'accord global inter classifieurs (Fleiss Kappa) était de 0,45.

Tableau 10. Heatmap des prédictions de chaque lésion test en fonction des modèles.

Chaque ligne représente une lésion test avec classe réelle dans la colonne de gauche. Chaque colonne représente la prédiction du modèle correspondant. Y : prédiction correcte, N : prédiction erronée. Colonne de droite, Exactitude (Accuracy) moyenne globale de l'ensemble des 31 classifieurs. Accord inter classifieurs : Fleiss Kappa = 0,45.



2.2. Modèles Random Forest

2.2.1. Résumé des performances des modèles Random Forest

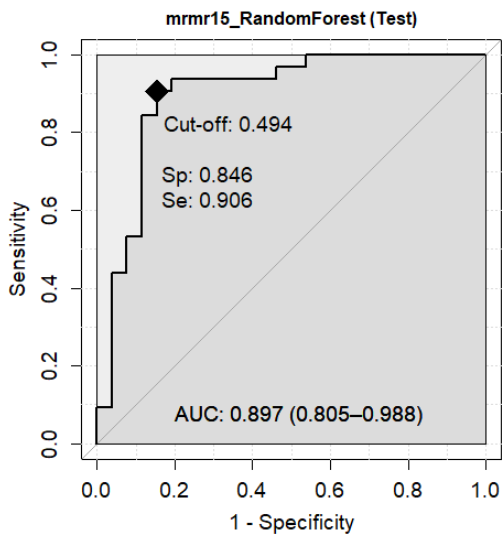
Les performances des modèles Random Forest sont résumés dans le tableau 1.

Tableau 1. Performances des modèles Random Forest sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_RF	RFE11_RF	LASSO15_RF	SBF10_RF	PCA15_RF
Test					
AUC	0.897 [0.805 ; 0.988]	0.912 [0.829 ; 0.996]	0.877 [0.780 ; 0.973]	0.844 [0.733 ; 0.956]	0.852 [0.753 ; 0.950]
Exactitude	0.862 [0.773 ; 0.951]	0.897 [0.818 ; 0.975]	0.810 [0.709 ; 0.911]	0.793 [0.689 ; 0.897]	0.448 [0.320 ; 0.576]
Sensibilité	0.875 [0.790 ; 0.960]	0.906 [0.831 ; 0.981]	0.813 [0.712 ; 0.913]	0.794 [0.690 ; 0.898]	0
Spécificité	0.846 [0.753 ; 0.939]	0.885 [0.802 ; 0.967]	0.808 [0.706 ; 0.909]	0.731 [0.617 ; 0.845]	1
VPP	0.875 [0.790 ; 0.960]	0.906 [0.831 ; 0.981]	0.839 [0.744 ; 0.933]	0.824 [0.725 ; 0.922]	0.794 [0.690 ; 0.898]
VPN	0.846 [0.753 ; 0.939]	0.885 [0.802 ; 0.967]	0.778 [0.671 ; 0.885]	0.792 [0.687 ; 0.896]	0.448 [0.320 ; 0.576]
X ² p	1.10 ⁻⁷	8.10 ⁻⁹	8.10 ⁻⁶	3.10 ⁻⁵	n/a
McNemar p	1	1	1	0.773	n/a
F1-score	0.875 [0.790 ; 0.960]	0.906 [0.831 ; 0.981]	0.825 [0.728 ; 0.923]	0.818 [0.719 ; 0.917]	0
MCC	0.686 [0.567 ; 0.806]	0.756 [0.645 ; 0.867]	0.584 [0.457 ; 0.710]	0.545 [0.417 ; 0.673]	0
Youden J	0.721 [0.543 ; 0.899]	0.791 [0.634 ; 0.948]	0.620 [0.418 ; 0.822]	0.575 [0.367 ; 0.782]	0
Bootstrap p	0.4326	0.6888	0.3419	0.4807	0.9169
Entrainement					
AUC	0.935 [0.918 ; 0.953]	0.929 [0.909 ; 0.950]	0.924 [0.903 ; 0.945]	0.885 [0.857 ; 0.914]	0.808 [0.739 ; 0.876]
Exactitude	0.859 [0.798 ; 0.919]	0.914 [0.790 ; 0.914]	0.868 [0.809 ; 0.927]	0.783 [0.711 ; 0.855]	0.783 [0.711 ; 0.855]
mtry	16	2	2	2	2
ntree	500	500	500	500	500

2.2.2. Courbes ROC et performances de prédiction par modèle RF

mRMR15_RF

Echantillon Test

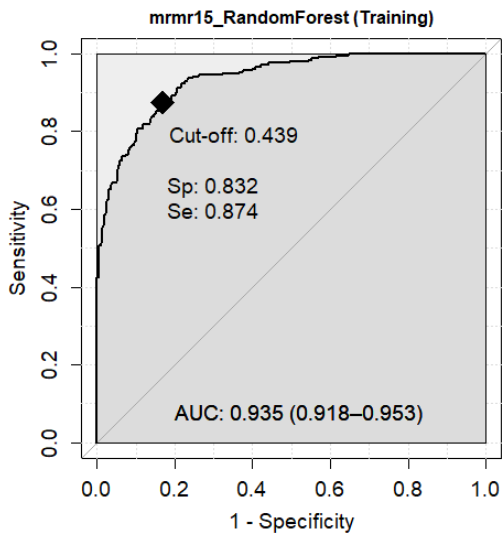


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	4
GBM	4	28

AUC = 0.897 [0.805 ; 0.988]
 Exactitude = 0.862 [0.773 ; 0.951]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.875 [0.790 ; 0.960]
 VPN = 0.846 [0.753 ; 0.939]
 X^2 p-valeur = 1.10^{-7}
 McNemar p = 1
 F1-Score = 0.875 [0.790 ; 0.960]
 MCC = 0.686 [0.567 ; 0.806]
 Youden J = 0.721 [0.543 ; 0.899]

Echantillon Entraînement

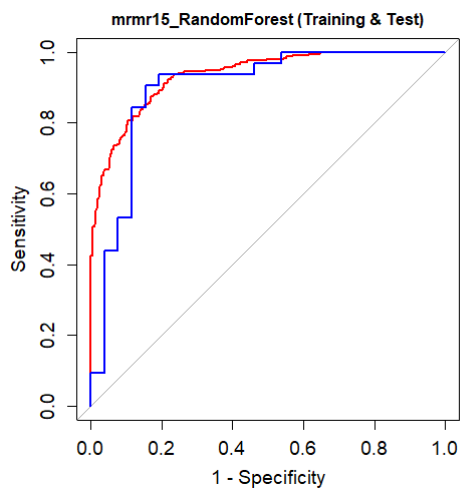


AUC = 0.935 [0.918 ; 0.953]
 Exactitude = 0.859 [0.798 ; 0.919]

Hyperparamètres :

Mtry = 16
 Ntree = 500

Comparaison ROC test et entraînement

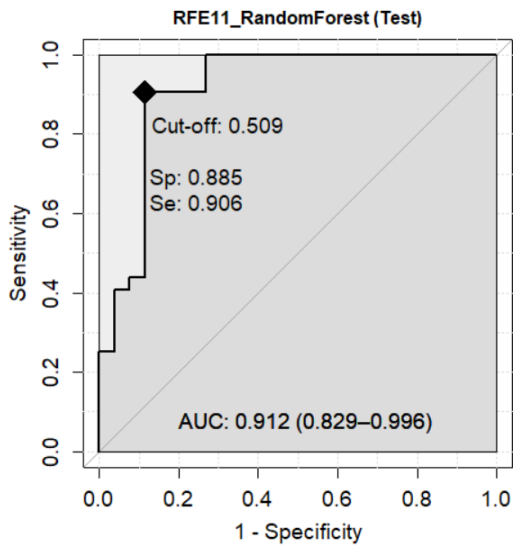


Bootstrap test :

D = 0.78476, boot.n = 2000, boot.stratified = 1, p-value = 0.4326

RFE11_RF

Echantillon Test

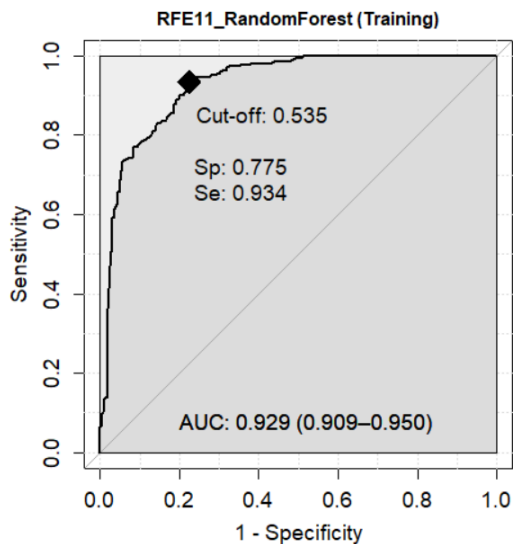


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	23	3
GBM	3	29

AUC = 0.912 [0.829 ; 0.996]
 Exactitude = 0.897 [0.818 ; 0.975]
 Sensibilité = 0.906 [0.831 ; 0.981]
 (GBM est la classe positive)
 Spécificité = 0.885 [0.802 ; 0.967]
 VPP = 0.906 [0.831 ; 0.981]
 VPN = 0.885 [0.802 ; 0.967]
 X^2 p-valeur = $8 \cdot 10^{-9}$
 McNemar p = 1
 F1-Score = 0.906 [0.831 ; 0.981]
 MCC = 0.756 [0.645 ; 0.867]
 Youden J = 0.791 [0.634 ; 0.948]

Echantillon Entraînement

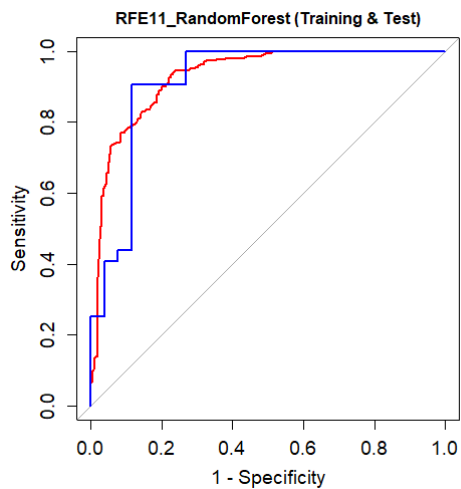


AUC = 0.929 [0.909 ; 0.950]
 Exactitude = 0.914 [0.790 ; 0.914]

Hyperparamètres :

Mtry = 2
 Ntree = 500

Comparaison ROC test et entraînement

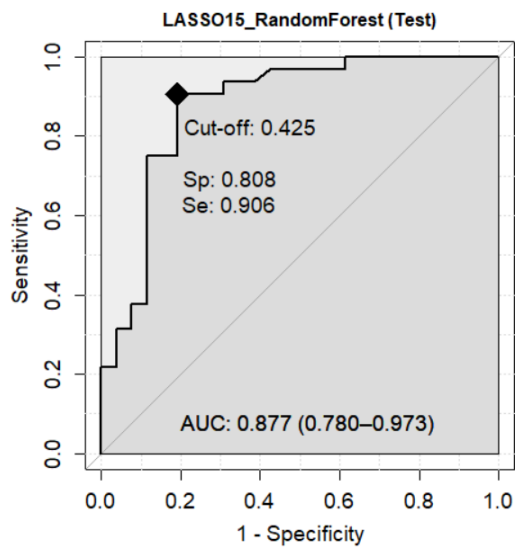


Bootstrap test :

$D = 0.40044$, boot.n = 2000, boot.stratified = 1, p-value = 0.6888

LASSO15_RF

Echantillon Test

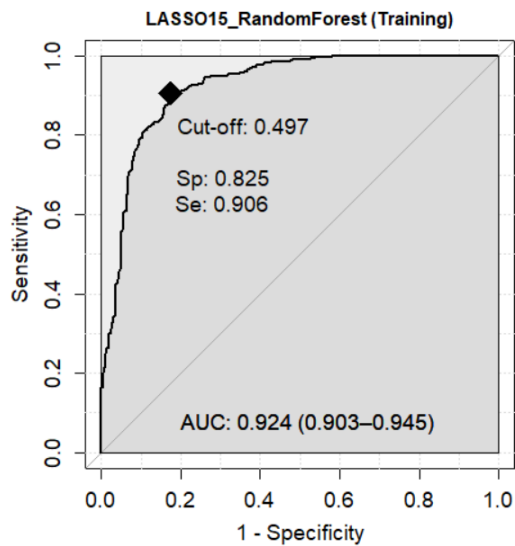


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	21	6
GBM	5	26

AUC = 0.877 [0.780 ; 0.973]
 Exactitude = 0.810 [0.709 ; 0.911]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.808 [0.706 ; 0.909]
 VPP = 0.839 [0.744 ; 0.933]
 VPN = 0.778 [0.671 ; 0.885]
 χ^2 p-valeur = $8 \cdot 10^{-6}$
 McNemar p = 1
 F1-Score = 0.825 [0.728 ; 0.923]
 MCC = 0.584 [0.457 ; 0.710]
 Youden J = 0.620 [0.418 ; 0.822]

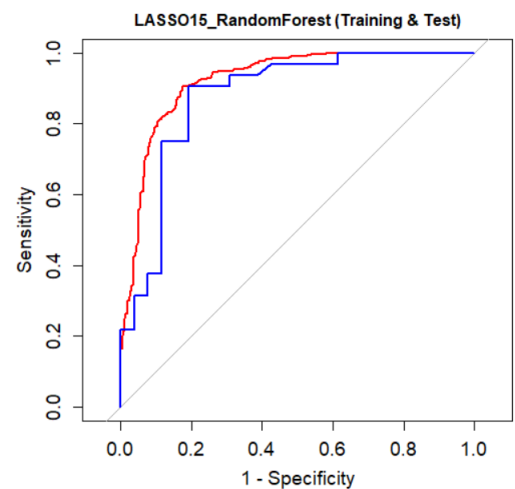
Echantillon Entraînement



AUC = 0.924 [0.903 ; 0.945]
 Exactitude = 0.868 [0.809 ; 0.927]

Hyperparamètres :
 Mtry = 2
 Ntree = 500

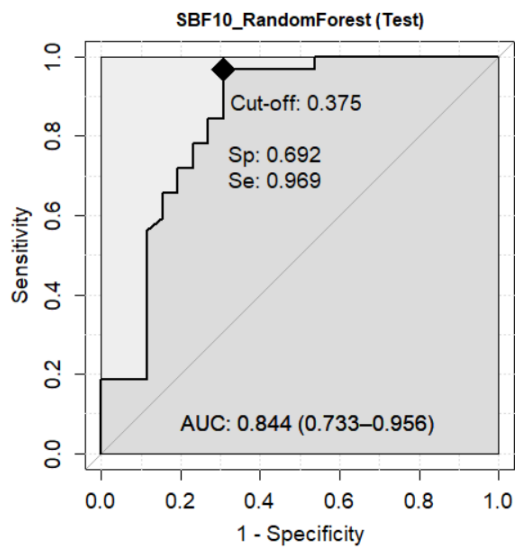
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.95045, boot.n = 2000, boot.stratified = 1, p-value = 0.3419

SBF10_RF

Echantillon Test

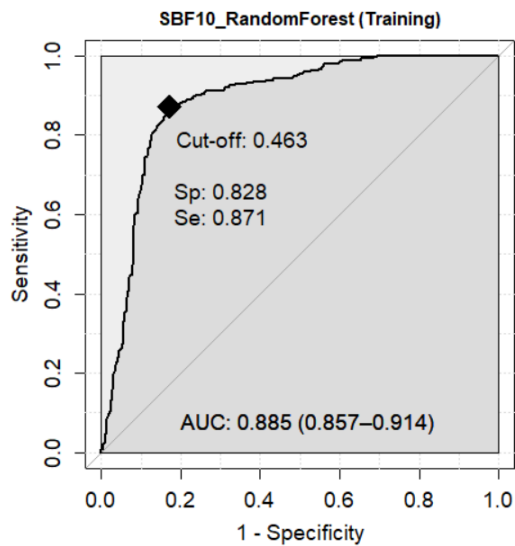


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	5
GBM	7	27

AUC = 0.844 [0.733 ; 0.956]
 Exactitude = 0.793 [0.689 ; 0.897]
 Sensibilité = 0.794 [0.690 ; 0.898]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.824 [0.725 ; 0.922]
 VPN = 0.792 [0.687 ; 0.896]
 X^2 p-valeur = $3 \cdot 10^{-5}$
 McNemar p = 0.773
 F1-Score = 0.818 [0.719 ; 0.917]
 MCC = 0.545 [0.417 ; 0.673]
 Youden J = 0.575 [0.367 ; 0.782]

Echantillon Entraînement

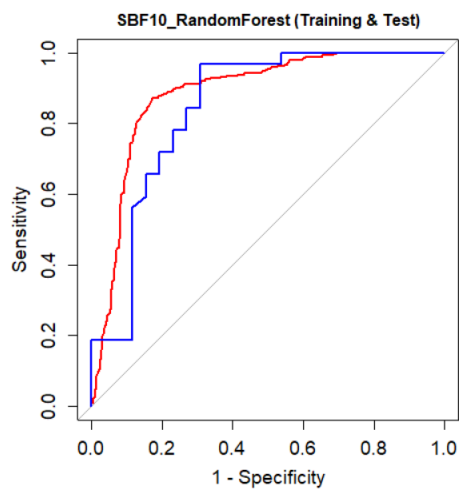


AUC = 0.885 [0.857 ; 0.914]
 Exactitude = 0.783 [0.711 ; 0.855]

Hyperparamètres :

Mtry = 2
 Ntree = 500

Comparaison ROC test et entraînement

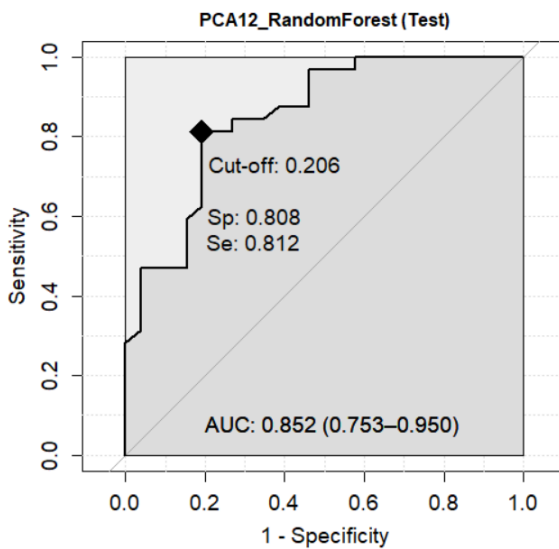


Bootstrap test :

$D = 0.70522$, boot.n = 2000, boot.stratified = 1, p-value = 0.4807

PCA15_RF

Echantillon Test

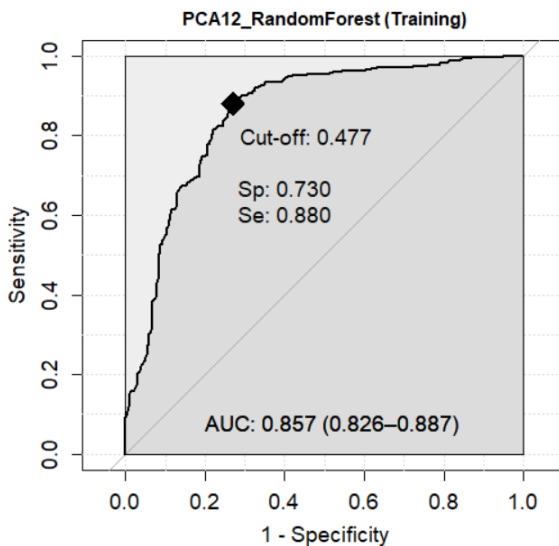


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	26	32
GBM	0	0

AUC = 0.852 [0.753 ; 0.950]
 Exactitude = 0.448 [0.320 ; 0.576]
 Sensibilité = 0
 (GBM est la classe positive)
 Spécificité = 1
 VPP = 0.794 [0.690 ; 0.898]
 VPN = 0.448 [0.320 ; 0.576]
 X^2 p-valeur = n/a
 McNemar p = n/a
 F1-Score = 0
 MCC = 0
 Youden J = 0

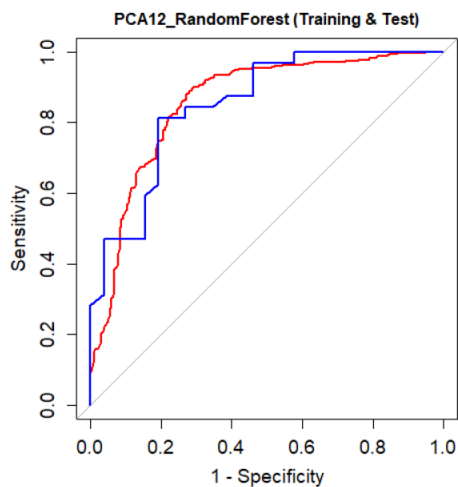
Echantillon Entraînement



AUC = 0.808 [0.739 ; 0.876]
 Exactitude = 0.783 [0.711 ; 0.855]

Hyperparamètres :
 Mtry = 2
 Ntree = 500

Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.10433, boot.n = 2000,
 boot.stratified = 1, p-value = 0.9169

2.3. Modèles Support Vector Machine

2.3.1. Résumé des performances des modèles Support Vector Machine

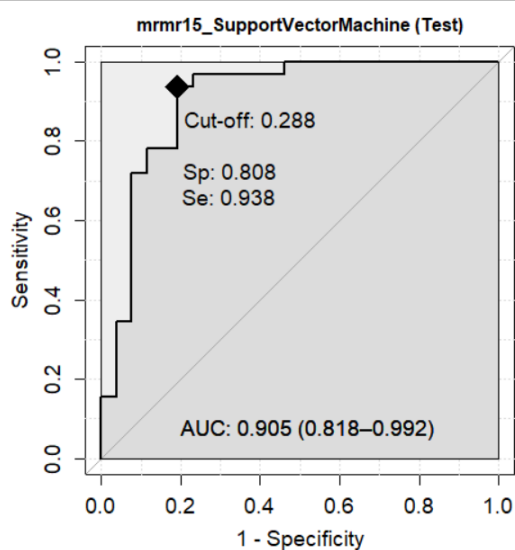
Les performances des modèles SVM sont résumés dans le tableau 1.

Tableau 1. Performances des modèles Support Vector Machine sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_SVM	RFE11_SVM	LASSO15_SVM	SBF10_SVM	PCA15_SVM
Test					
AUC	0.905 [0.818 ; 0.992]	0.928 [0.853 ; 1]	0.891 [0.804 ; 0.977]	0.865 [0.768 ; 0.963]	0.873 [0.784 ; 0.961]
Exactitude	0.828 [0.730 ; 0.925]	0.914 [0.842 ; 0.986]	0.828 [0.730 ; 0.925]	0.828 [0.730 ; 0.925]	0.741 [0.629 ; 0.854]
Sensibilité	0.781 [0.675 ; 0.888]	0.969 [0.924 ; 1]	0.813 [0.712 ; 0.913]	0.875 [0.790 ; 0.960]	0.75 [0.639 ; 0.861]
Spécificité	0.885 [0.802 ; 0.967]	0.846 [0.753 ; 0.939]	0.846 [0.753 ; 0.939]	0.769 [0.661 ; 0.878]	0.731 [0.617 ; 0.845]
VPP	0.893 [0.813 ; 0.972]	0.886 [0.804 ; 0.968]	0.867 [0.779 ; 0.954]	0.824 [0.725 ; 0.922]	0.774 [0.667 ; 0.882]
VPN	0.767 [0.658 ; 0.876]	0.957 [0.904 ; 1]	0.786 [0.680 ; 0.891]	0.833 [0.737 ; 0.929]	0.704 [0.586 ; 0.821]
X ² p	1.10 ⁻⁶	1.10 ⁻⁹	2.10 ⁻⁶	2.10 ⁻⁶	0.0007
McNemar p	0.34	0.037	0.752	0.75	1
F1-score	0.833 [0.737 ; 0.929]	0.925 [0.858 ; 0.993]	0.839 [0.744 ; 0.933]	0.848 [0.756 ; 0.941]	0.762 [0.652 ; 0.872]
MCC	0.628 [0.504 ; 0.752]	0.793 [0.689 ; 0.897]	0.621 [0.496 ; 0.746]	0.615 [0.490 ; 0.741]	0.445 [0.317 ; 0.572]
Youden J	0.666 [0.477 ; 0.854]	0.815 [0.677 ; 0.953]	0.659 [0.465 ; 0.852]	0.644 [0.451 ; 0.838]	0.481 [0.255 ; 0.706]
Bootstrap p	0.4013	0.7158	0.2852	0.5368	0.3441
Entraînement					
AUC	0.943 [0.926 ; 0.960]	0.942 [0.922 ; 0.962]	0.939 [0.918 ; 0.960]	0.897 [0.871 ; 0.923]	0.827 [0.794 ; 0.860]
Exactitude	0.881 [0.824 ; 0.937]	0.890 [0.835 ; 0.944]	0.899 [0.847 ; 0.952]	0.838 [0.774 ; 0.902]	0.792 [0.722 ; 0.863]
Cost C	0.5	1	1	1	1
Sigma	0.0226	0.0246	0.0239	0.031	0.0309
Kernel	Gaussian Radial Basis	Gaussian Radial Basis	Gaussian Radial Basis	Gaussian Radial Basis	Gaussian Radial Basis
Support Vectors	91	82	83	84	90

2.3.2. Courbes ROC et performances de prédiction par modèle SVM

mRMR15_SVM

Echantillon Test

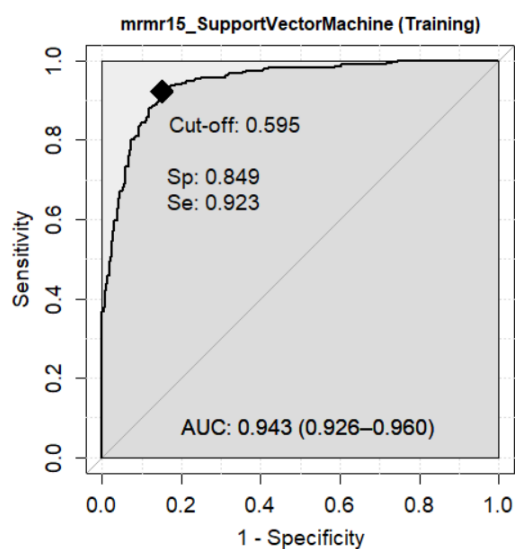


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	23	7
GBM	3	25

AUC = 0.905 [0.818 ; 0.992]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.781 [0.675 ; 0.888]
 (GBM est la classe positive)
 Spécificité = 0.885 [0.802 ; 0.967]
 VPP = 0.893 [0.813 ; 0.972]
 VPN = 0.767 [0.658 ; 0.876]
 χ^2 p-valeur = 1.10^{-6}
 McNemar p = 0.34
 F1-Score = 0.833 [0.737 ; 0.929]
 MCC = 0.628 [0.504 ; 0.752]
 Youden J = 0.666 [0.477 ; 0.854]

Echantillon Entraînement

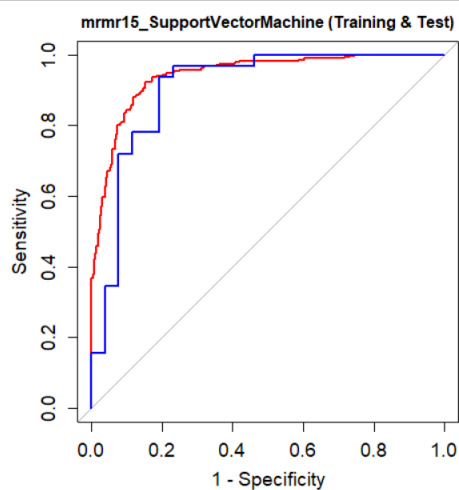


AUC = 0.943 [0.926 ; 0.960]
 Exactitude = 0.881 [0.824 ; 0.937]

Hyperparamètres :

Cost C = 0.5
 Sigma = 0.0226
 Kernel : Gaussian Radial Basis
 Number of support vectors : 91

Comparaison ROC test et entraînement

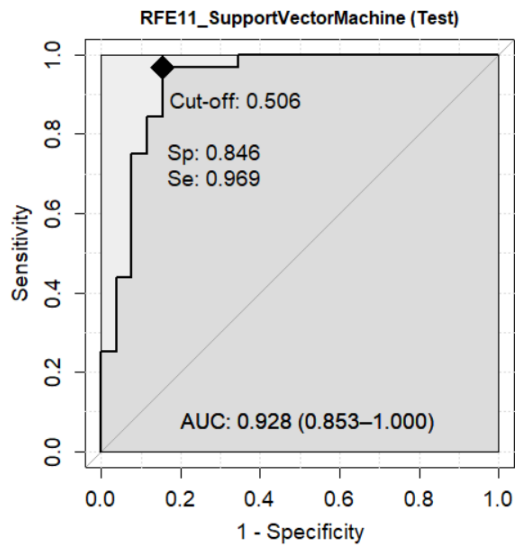


Bootstrap test :

D = 0.83922, boot.n = 2000, boot.stratified = 1, p-value = 0.4013

RFE11_SVM

Echantillon Test

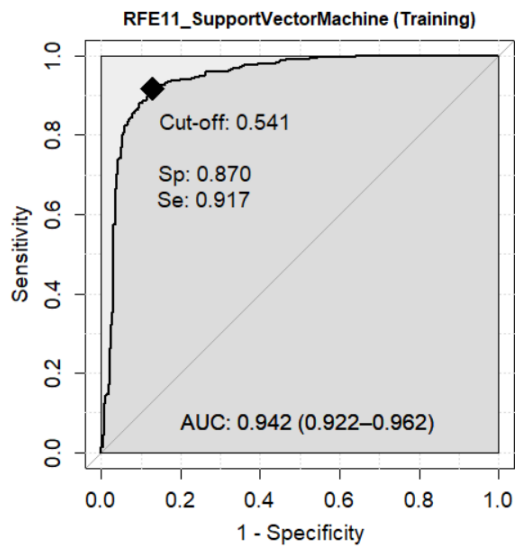


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	1
GBM	4	31

AUC = 0.928 [0.853 ; 1]
 Exactitude = 0.914 [0.842 ; 0.986]
 Sensibilité = 0.969 [0.924 ; 1]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.886 [0.804 ; 0.968]
 VPN = 0.957 [0.904 ; 1]
 X^2 p-valeur = 1.10^{-9}
 McNemar p = 0.037
 F1-Score = 0.925 [0.858 ; 0.993]
 MCC = 0.793 [0.689 ; 0.897]
 Youden J = 0.815 [0.677 ; 0.953]

Echantillon Entraînement

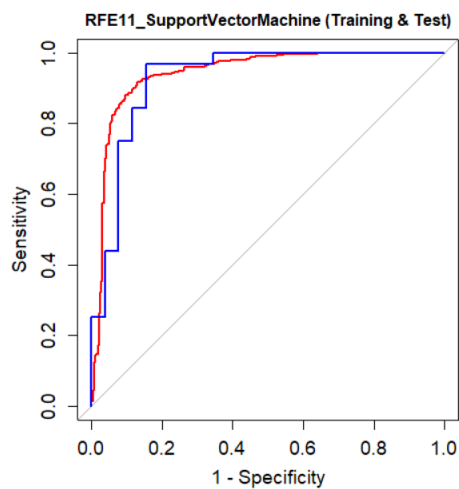


AUC = 0.942 [0.922 ; 0.962]
 Exactitude = 0.890 [0.835 ; 0.944]

Hyperparamètres :

Cost C = 1
 Sigma = 0.0246
 Kernel : Gaussian Radial Basis
 Number of support vectors : 82

Comparaison ROC test et entraînement

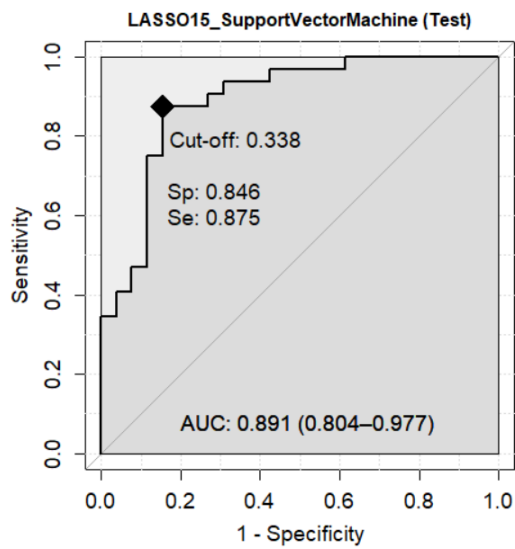


Bootstrap test :

D = 0.36413, boot.n = 2000, boot.stratified = 1, p-value = 0.7158

LASSO15_SVM

Echantillon Test

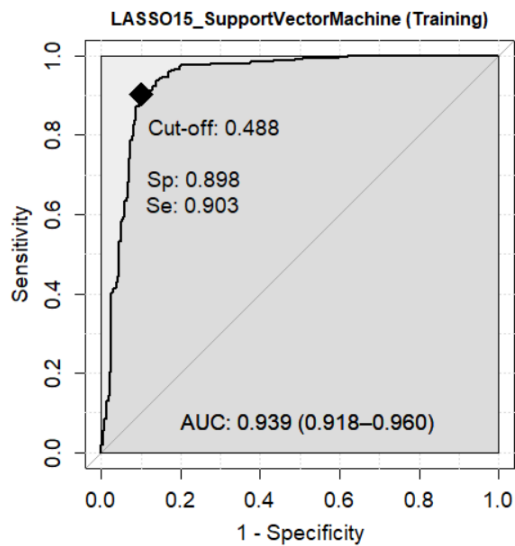


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	6
GBM	4	26

AUC = 0.891 [0.804 ; 0.977]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.867 [0.779 ; 0.954]
 VPN = 0.786 [0.680 ; 0.891]
 X^2 p-valeur = $2 \cdot 10^{-6}$
 McNemar p = 0.752
 F1-Score = 0.839 [0.744 ; 0.933]
 MCC = 0.621 [0.496 ; 0.746]
 Youden J = 0.659 [0.465 ; 0.852]

Echantillon Entraînement

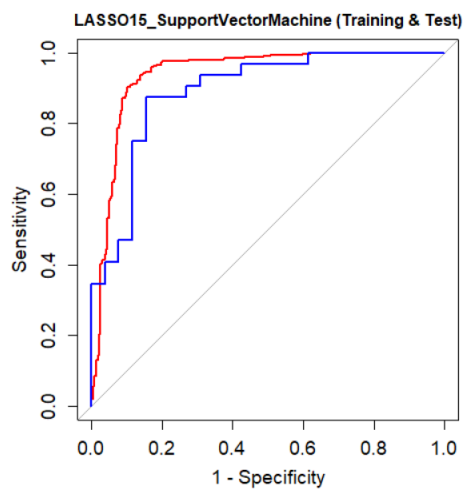


AUC = 0.939 [0.918 ; 0.960]
 Exactitude = 0.899 [0.847 ; 0.952]

Hyperparamètres :

Cost C = 1
 Sigma = 0.0239
 Kernel : Gaussian Radial Basis
 Number of support vectors : 83

Comparaison ROC test et entraînement

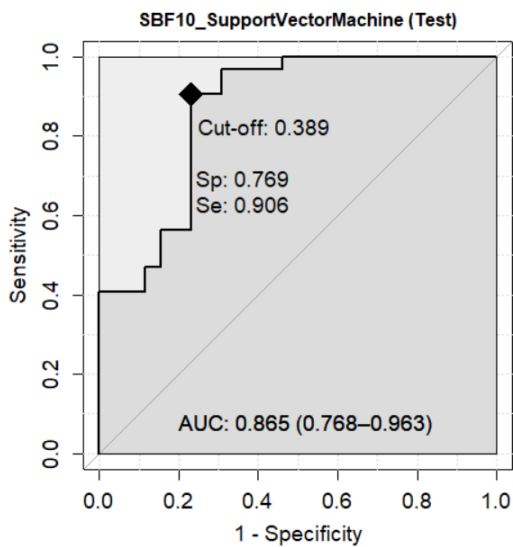


Bootstrap test :

D = 1.0686, boot.n = 2000, boot.stratified = 1, p-value = 0.2852

SBF10_SVM

Echantillon Test

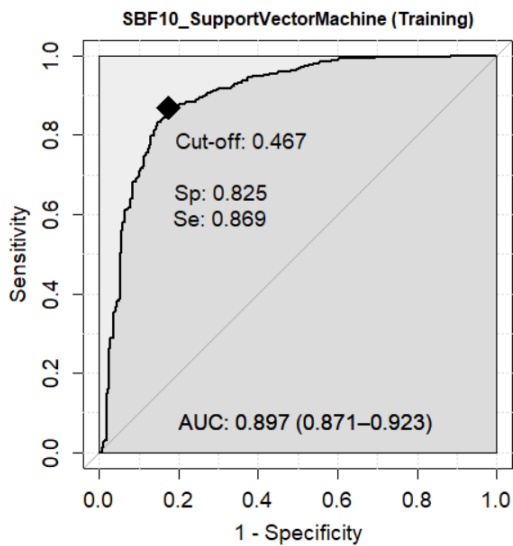


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	20	4
GBM	6	28

AUC = 0.865 [0.768 ; 0.963]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.769 [0.661 ; 0.878]
 VPP = 0.824 [0.725 ; 0.922]
 VPN = 0.833 [0.737 ; 0.929]
 X^2 p-valeur = $2 \cdot 10^{-6}$
 McNemar p = 0.75
 F1-Score = 0.848 [0.756 ; 0.941]
 MCC = 0.615 [0.490 ; 0.741]
 Youden J = 0.644 [0.451 ; 0.838]

Echantillon Entraînement

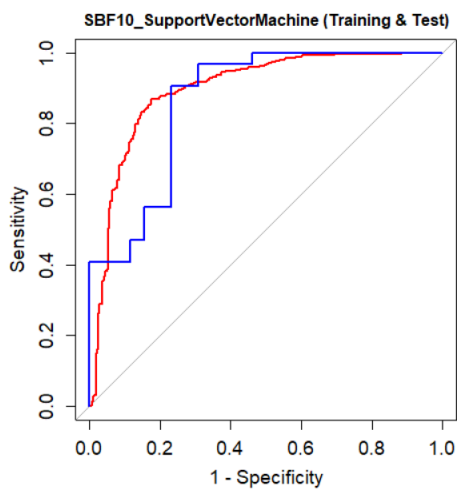


AUC = 0.897 [0.871 ; 0.923]
 Exactitude = 0.838 [0.774 ; 0.902]

Hyperparamètres :

Cost C = 1
 Sigma = 0.031
 Kernel : Gaussian Radial Basis
 Number of support vectors : 84

Comparaison ROC test et entraînement

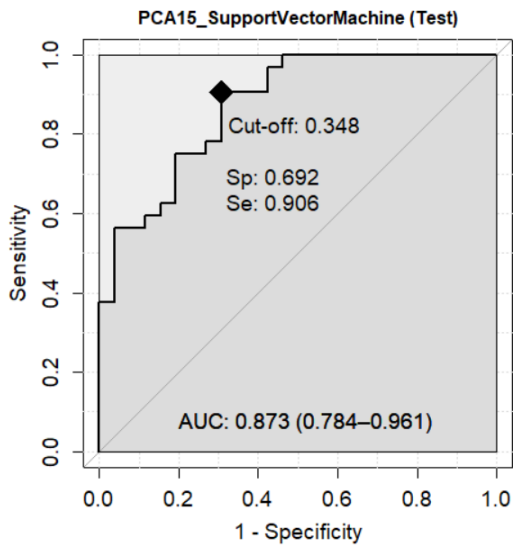


Bootstrap test :

D = 0.61765, boot.n = 2000, boot.stratified = 1, p-value = 0.5368

PCA15_SVM

Echantillon Test

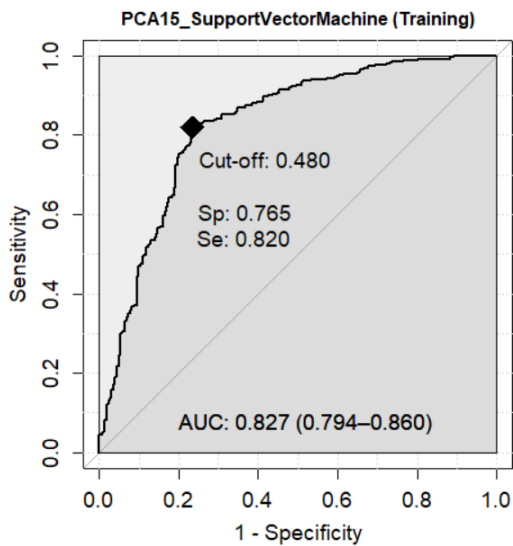


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	8
GBM	7	24

AUC = 0.873 [0.784 ; 0.961]
 Exactitude = 0.741 [0.629 ; 0.854]
 Sensibilité = 0.75 [0.639 ; 0.861]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.774 [0.667 ; 0.882]
 VPN = 0.704 [0.586 ; 0.821]
 χ^2 p-valeur = 0.0007
 McNemar p = 1
 F1-Score = 0.762 [0.652 ; 0.872]
 MCC = 0.445 [0.317 ; 0.572]
 Youden J = 0.481 [0.255 ; 0.706]

Echantillon Entraînement

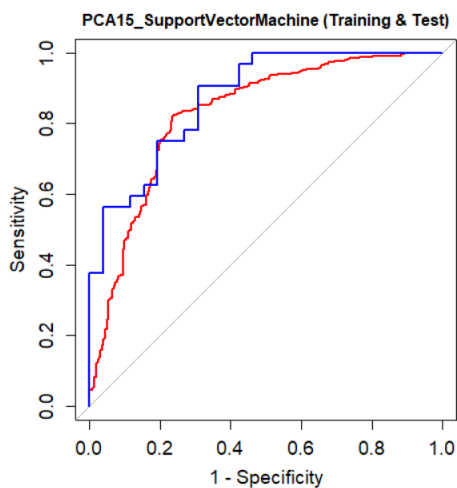


AUC = 0.827 [0.794 ; 0.860]
 Exactitude = 0.792 [0.722 ; 0.863]

Hyperparamètres :

Cost C = 1
 Sigma = 0.0309
 Kernel : Gaussian Radial Basis
 Number of support vectors : 90

Comparaison ROC test et entraînement



Bootstrap test :

D = -0.94609, boot.n = 2000, boot.stratified = 1, p-value = 0.3441

2.4. Modèles Régression Logistique

2.4.1. Résumé des performances des modèles Régression Logistique

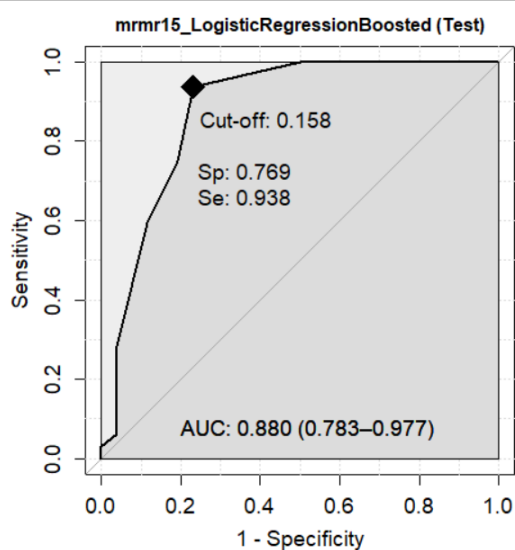
Les performances des modèles LR sont résumés dans le tableau 1.

Tableau 1. Performances des modèles Régression Logistique sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_LR	RFE11_LR	LASSO15_LR	SBF10_LR	PCA15_LR
Test					
AUC	0.880 [0.783 ; 0.977]	0.850 [0.738 ; 0.961]	0.849 [0.741 ; 0.957]	0.791 [0.673 ; 0.910]	0.755 [0.635 ; 0.876]
Exactitude	0.776 [0.669 ; 0.883]	0.793 [0.689 ; 0.897]	0.741 [0.629 ; 0.854]	0.690 [0.571 ; 0.809]	0.672 [0.552 ; 0.793]
Sensibilité	0.75 [0.639 ; 0.861]	0.781 [0.675 ; 0.888]	0.75 [0.639 ; 0.861]	0.688 [0.568 ; 0.807]	0.531 [0.403 ; 0.660]
Spécificité	0.808 [0.706 ; 0.909]	0.808 [0.706 ; 0.909]	0.731 [0.617 ; 0.845]	0.692 [0.574 ; 0.811]	0.846 [0.753 ; 0.939]
VPP	0.828 [0.730 ; 0.925]	0.833 [0.737 ; 0.929]	0.774 [0.667 ; 0.882]	0.733 [0.620 ; 0.847]	0.810 [0.708 ; 0.911]
VPN	0.724 [0.609 ; 0.839]	0.75 [0.639 ; 0.861]	0.704 [0.586 ; 0.821]	0.643 [0.520 ; 0.766]	0.595 [0.468 ; 0.721]
X ² p	7.10 ⁻⁵	2.10 ⁻⁵	0.0007	0.0089	0.0069
McNemar p	0,579	0,77	1	0.81	0,022
F1-score	0.787 [0.681 ; 0.892]	0.806 [0.705 ; 0.908]	0.762 [0.652 ; 0.872]	0.710 [0.593 ; 0.826]	0.642 [0.518 ; 0.765]
MCC	0.520 [0.391 ; 0.649]	0.551 [0.423 ; 0.679]	0.445 [0.317 ; 0.572]	0.343 [0.221 ; 0.466]	0.354 [0.231 ; 0.478]
Youden J	0.558 [0.345 ; 0.771]	0.589 [0.381 ; 0.797]	0.481 [0.255 ; 0.706]	0.380 [0.142 ; 0.618]	0.377 [0.156 ; 0.599]
Bootstrap p	0.678	0.565	0.3151	0.4006	0.3855
Entraînement	mRMR15_LR	RFE11_LR	LASSO15_LR	SBF10_LR	PCA15_LR
AUC	0.902 [0.879 ; 0.924]	0.883 [0.857 ; 0.910]	0.907 [0.884 ; 0.930]	0.844 [0.813 ; 0.874]	0.811 [0.777 ; 0.845]
Exactitude	0.824 [0.757 ; 0.890]	0.814 [0.747 ; 0.882]	0.841 [0.777 ; 0.904]	0.786 [0.715 ; 0.857]	0.761 [0.687 ; 0.835]
Itérations	11	21	21	11	31

2.4.2. Courbes ROC et performances de prédiction par modèle LR

mRMR15_LR

Echantillon Test

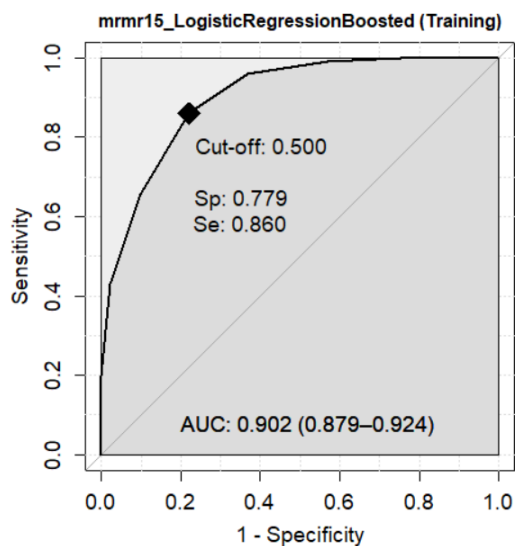


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	21	8
GBM	5	24

AUC = 0.880 [0.783 ; 0.977]
 Exactitude = 0.776 [0.669 ; 0.883]
 Sensibilité = 0.75 [0.639 ; 0.861]
 (GBM est la classe positive)
 Spécificité = 0.808 [0.706 ; 0.909]
 VPP = 0.828 [0.730 ; 0.925]
 VPN = 0.724 [0.609 ; 0.839]
 χ^2 p-valeur = $7 \cdot 10^{-5}$
 McNemar p = 0,579
 F1-Score = 0.787 [0.681 ; 0.892]
 MCC = 0.520 [0.391 ; 0.649]
 Youden J = 0.558 [0.345 ; 0.771]

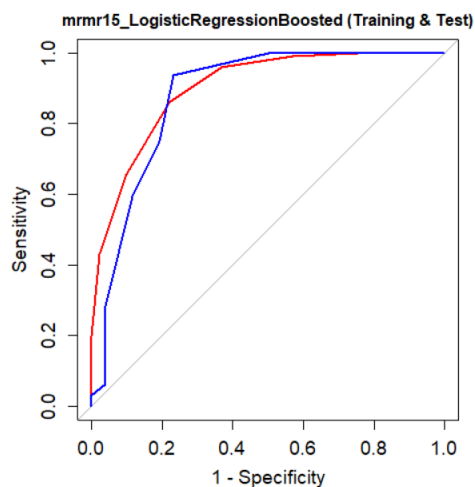
Echantillon Entraînement



AUC = 0.902 [0.879 ; 0.924]
 Exactitude = 0.824 [0.757 ; 0.890]

Hyperparamètres :
 N_iter = 11

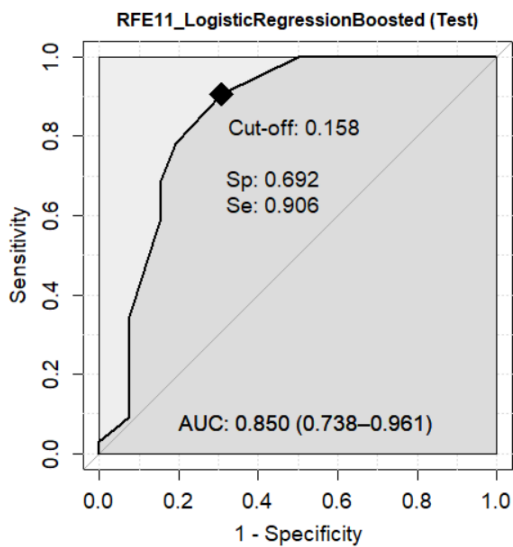
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.41517, boot.n = 2000, boot.stratified = 1, p-value = 0.678

RFE11_LR

Echantillon Test

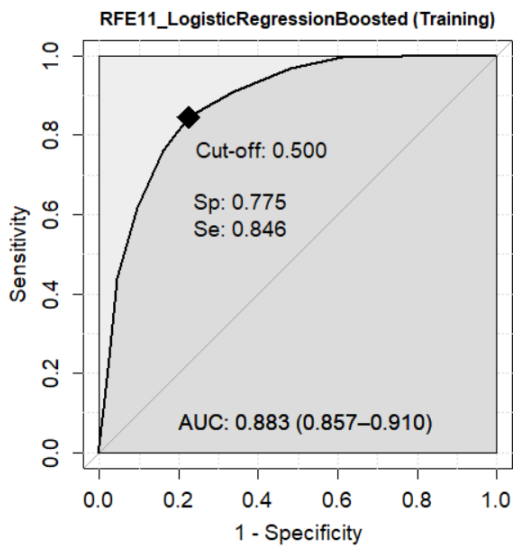


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	21	7
GBM	5	25

AUC = 0.850 [0.738 ; 0.961]
 Exactitude = 0.793 [0.689 ; 0.897]
 Sensibilité = 0.781 [0.675 ; 0.888]
 (GBM est la classe positive)
 Spécificité = 0.808 [0.706 ; 0.909]
 VPP = 0.833 [0.737 ; 0.929]
 VPN = 0.75 [0.639 ; 0.861]
 X^2 p-valeur = $2 \cdot 10^{-5}$
 McNemar p = 0,77
 F1-Score = 0.806 [0.705 ; 0.908]
 MCC = 0.551 [0.423 ; 0.679]
 Youden J = 0.589 [0.381 ; 0.797]

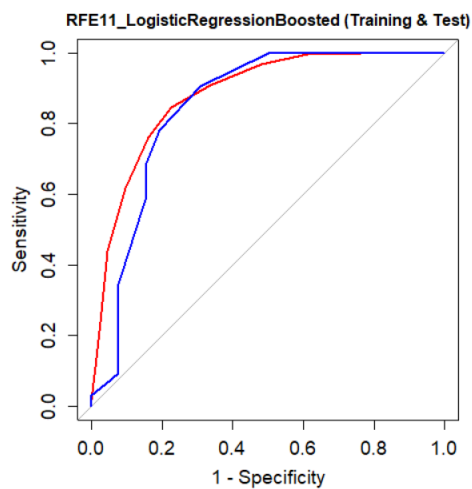
Echantillon Entraînement



AUC = 0.883 [0.857 ; 0.910]
 Exactitude = 0.814 [0.747 ; 0.882]

Hyperparamètres :
 N_iter = 21

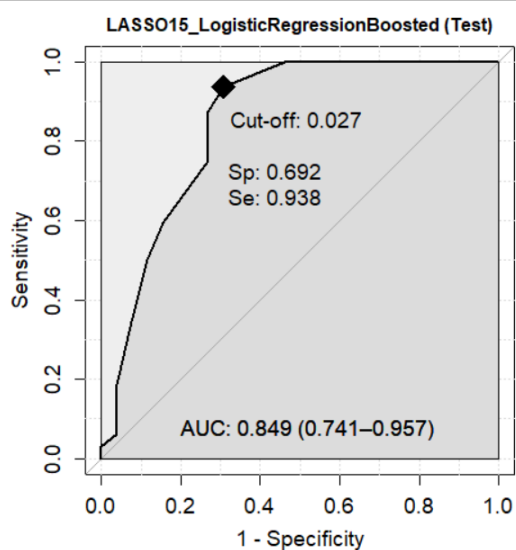
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.5755, boot.n = 2000, boot.stratified = 1, p-value = 0.565
 a

LASSO15_LR

Echantillon Test

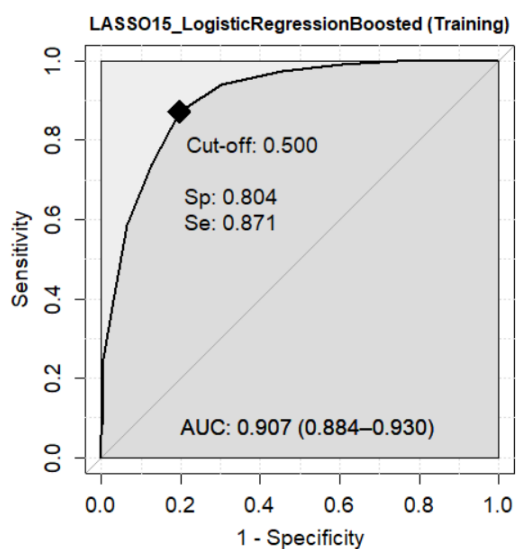


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	8
GBM	7	24

AUC = 0.849 [0.741 ; 0.957]
 Exactitude = 0.741 [0.629 ; 0.854]
 Sensibilité = 0.75 [0.639 ; 0.861]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.774 [0.667 ; 0.882]
 VPN = 0.704 [0.586 ; 0.821]
 χ^2 p-valeur = 0.0007
 McNemar p = 1
 F1-Score = 0.762 [0.652 ; 0.872]
 MCC = 0.445 [0.317 ; 0.572]
 Youden J = 0.481 [0.255 ; 0.706]

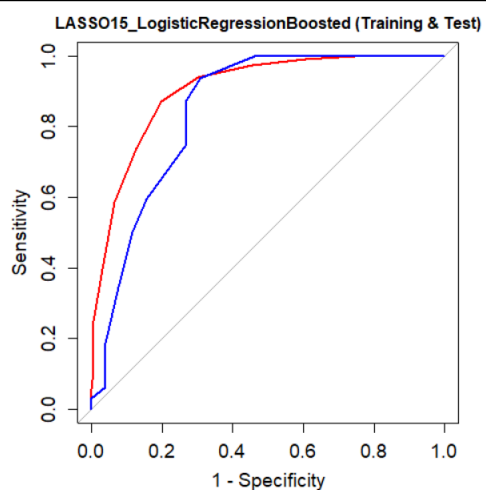
Echantillon Entraînement



AUC = 0.907 [0.884 ; 0.930]
 Exactitude = 0.841 [0.777 ; 0.904]

Hyperparamètres :
 N_iter = 21

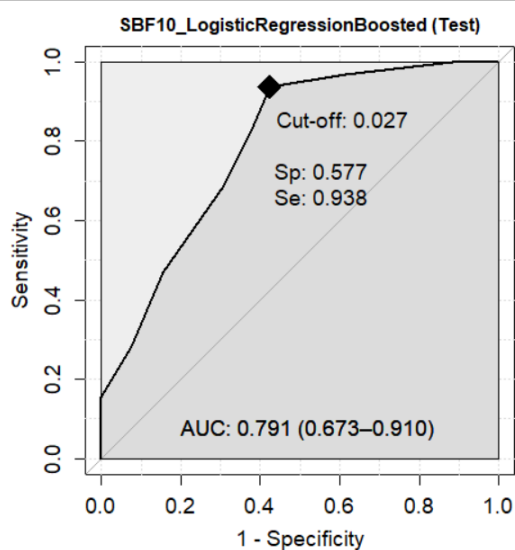
Comparaison ROC test et entraînement



Bootstrap test :
 D = 1.0046, boot.n = 2000, boot.stratified = 1, p-value = 0.3151

SBF10_LR

Echantillon Test

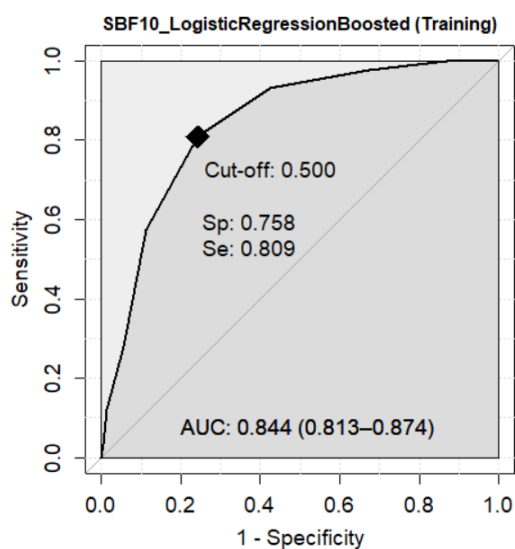


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	18	8
GBM	10	22

AUC = 0.791 [0.673 ; 0.910]
 Exactitude = 0.690 [0.571 ; 0.809]
 Sensibilité = 0.688 [0.568 ; 0.807]
 (GBM est la classe positive)
 Spécificité = 0.692 [0.574 ; 0.811]
 VPP = 0.733 [0.620 ; 0.847]
 VPN = 0.643 [0.520 ; 0.766]
 χ^2 p-valeur = 0.0089
 McNemar p = 0.81
 F1-Score = 0.710 [0.593 ; 0.826]
 MCC = 0.343 [0.221 ; 0.466]
 Youden J = 0.380 [0.142 ; 0.618]

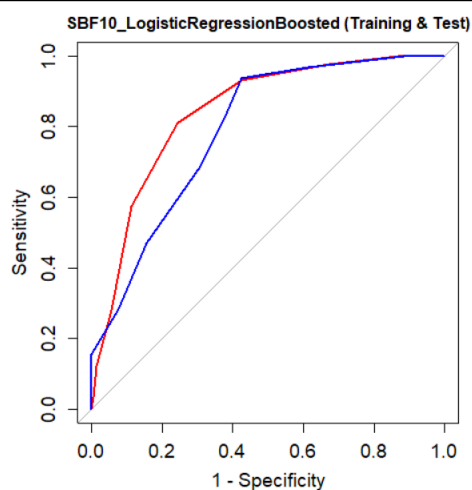
Echantillon Entraînement



AUC = 0.844 [0.813 ; 0.874]
 Exactitude = 0.786 [0.715 ; 0.857]

Hyperparamètres :
 N_iter = 11

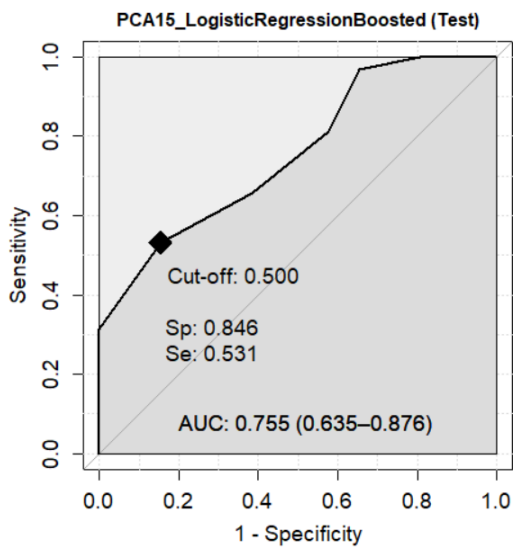
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.84051, boot.n = 2000, boot.stratified = 1, p-value = 0.4006

PCA15_LR

Echantillon Test

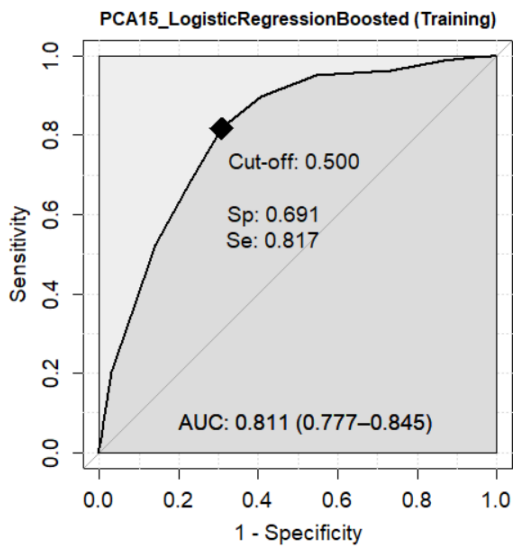


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	15
GBM	4	17

AUC = 0.755 [0.635 ; 0.876]
 Exactitude = 0.672 [0.552 ; 0.793]
 Sensibilité = 0.531 [0.403 ; 0.660]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.810 [0.708 ; 0.911]
 VPN = 0.595 [0.468 ; 0.721]
 χ^2 p-valeur = 0.0069
 McNemar p = 0,022
 F1-Score = 0.642 [0.518 ; 0.765]
 MCC = 0.354 [0.231 ; 0.478]
 Youden J = 0.377 [0.156 ; 0.599]

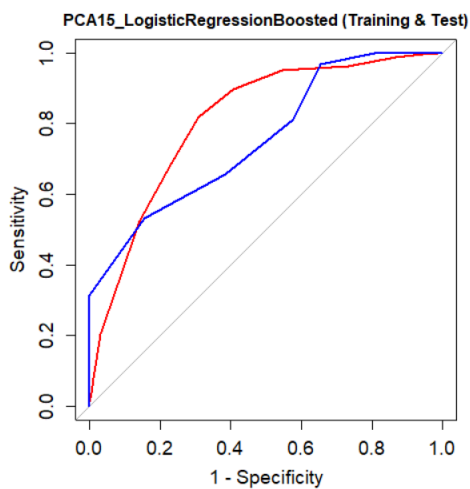
Echantillon Entraînement



AUC = 0.811 [0.777 ; 0.845]
 Exactitude = 0.761 [0.687 ; 0.835]

Hyperparamètres :
 N_iter = 31

Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.8679, boot.n = 2000, boot.stratified = 1, p-value = 0.3855

2.5. Modèles Bayésiens Naïfs

2.5.1. Résumé des performances des modèles Bayésiens Naïfs

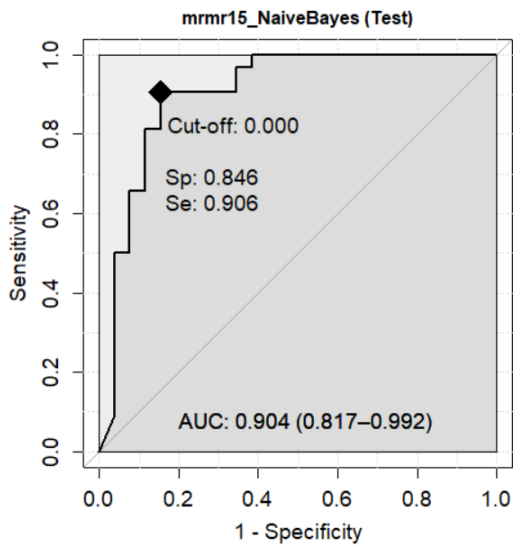
Les performances des modèles NB sont résumés dans le tableau 1.

Tableau 1. Performances des modèles Bayésiens Naïfs sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_NB	RFE11_NB	LASSO15_NB	SBF10_NB	PCA15_NB
Test					
AUC	0.904 [0.817 ; 0.992]	0.847 [0.736 ; 0.959]	0.816 [0.700 ; 0.932]	0.786 [0.644 ; 0.928]	0.793 [0.672 ; 0.914]
Exactitude	0.759 [0.648 ; 0.869]	0.793 [0.689 ; 0.897]	0.759 [0.648 ; 0.869]	0.741 [0.629 ; 0.854]	0.448 [0.320 ; 0.576]
Sensibilité	0.656 [0.534 ; 0.778]	0.844 [0.750 ; 0.937]	0.781 [0.675 ; 0.888]	0.75 [0.639 ; 0.861]	0
Spécificité	0.885 [0.967 ; 802]	0.731 [0.617 ; 0.845]	0.731 [0.617 ; 0.845]	0.731 [0.617 ; 0.811]	1
VPP	0.875 [0.790 ; 0.960]	0.794 [0.690 ; 0.898]	0.781 [0.675 ; 0.888]	0.774 [0.667 ; 0.882]	n/a
VPN	0.676 [0.556 ; 0.797]	0.792 [0.687 ; 0.896]	0.731 [0.617 ; 0.845]	0.704 [0.586 ; 0.821]	0.448 [0.320 ; 0.576]
X ² p	1.10 ⁻⁴	3.10 ⁻⁵	0.00028	0.0007	n/a
McNemar p	0,061	0,772	1	1	4.10 ⁻⁸
F1-score	0.75 [0.639 ; 0.861]	0.818 [0.719 ; 0.917]	0.781 [0.675 ; 0.888]	0.762 [0.652 ; 0.872]	0
MCC	0.511 [0.382 ; 0.640]	0.545 [0.417 ; 0.673]	0.477 [0.349 ; 0.606]	0.445 [0.317 ; 0.572]	n/a
Youden J	0.541 [0.336 ; 0.745]	0.575 [0.367 ; 0.782]	0.512 [0.291 ; 0.733]	0.481 [0.255 ; 0.706]	0
Bootstrap p	0.9657	0.4191	0.8976	0.4974	0.3739
Entraînement					
AUC	0.906 [0.881 ; 0.932]	0.896 [0.870 ; 0.921]	0.808 [0.771 ; 0.845]	0.837 [0.805 ; 0.869]	0.736 [0.696 ; 0.776]
Exactitude	0.831 [0.766 ; 0.897]	0.817 [0.750 ; 0.884]	0.775 [0.703 ; 0.848]	0.788 [0.717 ; 0.859]	0.681 [0.599 ; 0.762]
Kernel	Non	Oui	Oui	Oui	Non
Filtre Laplacien	Non	Non	Non	Non	Non
Probabilités a priori	- mc: 0.4488 - gbm: 0.5512				

2.5.2. Courbes ROC et performances de prédiction par modèle NB

mRMR15_NB

Echantillon Test

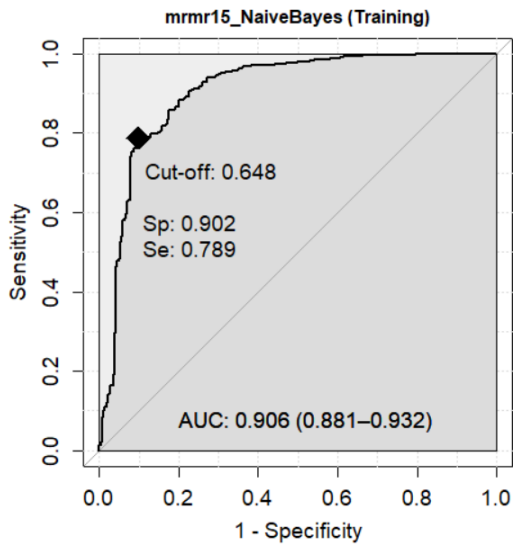


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	23	11
GBM	3	21

AUC = 0.904 [0.817 ; 0.992]
 Exactitude = 0.759 [0.648 ; 0.869]
 Sensibilité = 0.656 [0.534 ; 0.778]
 (GBM est la classe positive)
 Spécificité = 0.885 [0.967 ; 802]
 VPP = 0.875 [0.790 ; 0.960]
 VPN = 0.676 [0.556 ; 0.797]
 X^2 p-valeur = 1.10^{-4}
 McNemar p = 0,061
 F1-Score = 0.75 [0.639 ; 0.861]
 MCC = 0.511 [0.382 ; 0.640]
 Youden J = 0.541 [0.336 ; 0.745]

Echantillon Entraînement



AUC = 0.906 [0.881 ; 0.932]
 Exactitude = 0.831 [0.766 ; 0.897]

Hyperparamètres :

Laplace smoothing = 0

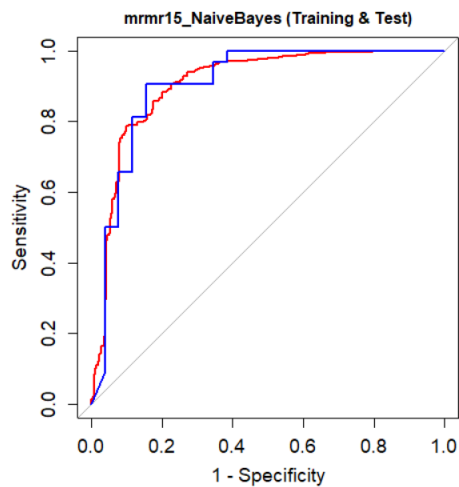
Usekernel = False

Prior probabilities :

- mc: 0.4488

- gbm: 0.5512

Comparaison ROC test et entraînement



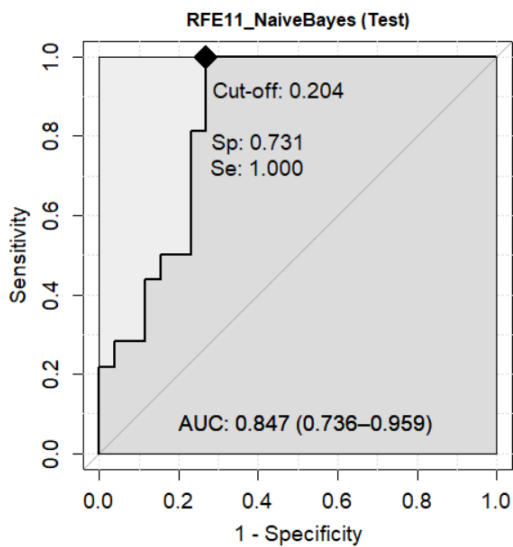
Bootstrap test :

D = 0.043005, boot.n = 2000,

boot.stratified = 1, p-value = 0.9657

RFE11_NB

Echantillon Test

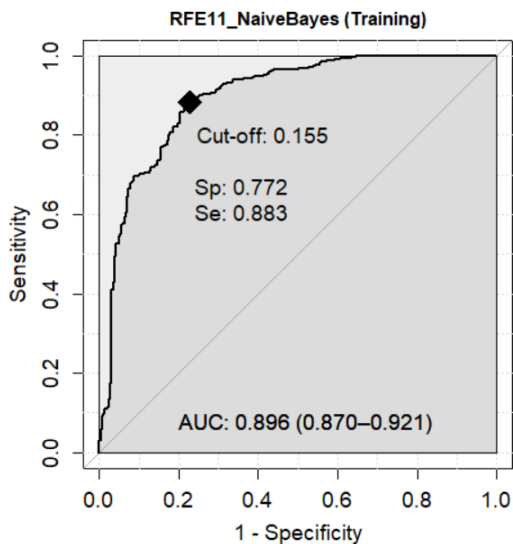


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	5
GBM	7	27

AUC = 0.847 [0.736 ; 0.959]
 Exactitude = 0.793 [0.689 ; 0.897]
 Sensibilité = 0.844 [0.750 ; 0.937]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.794 [0.690 ; 0.898]
 VPN = 0.792 [0.687 ; 0.896]
 X^2 p-valeur = $3 \cdot 10^{-5}$
 McNemar p = 0,772
 F1-Score = 0.818 [0.719 ; 0.917]
 MCC = 0.545 [0.417 ; 0.673]
 Youden J = 0.575 [0.367 ; 0.782]

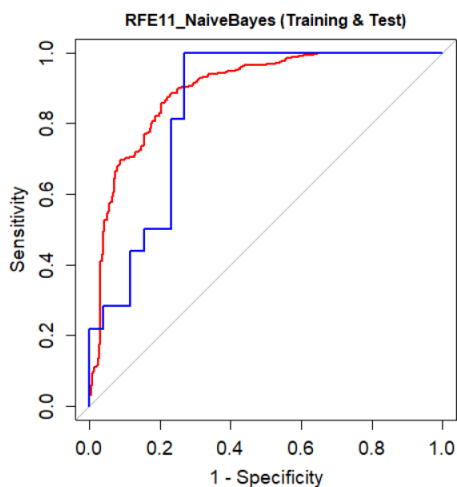
Echantillon Entraînement



AUC = 0.896 [0.870 ; 0.921]
 Exactitude = 0.817 [0.750 ; 0.884]

Hyperparamètres :
 Laplace smoothing = 0
 Usekernel = True
 Prior probabilities :
 - mc: 0.4488
 - gbm: 0.5512

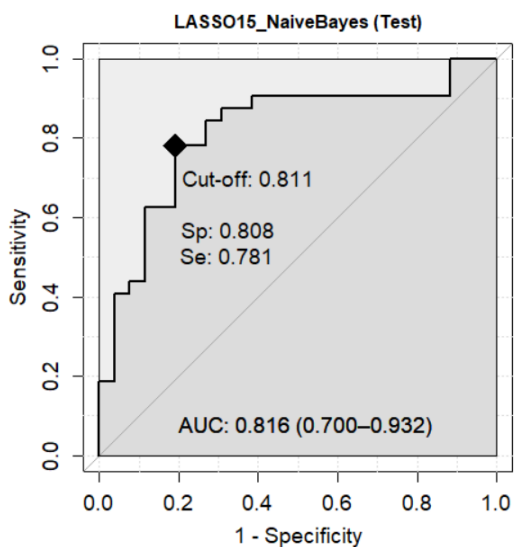
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.80795, boot.n = 2000, boot.stratified = 1, p-value = 0.4191

LASSO15_NB

Echantillon Test

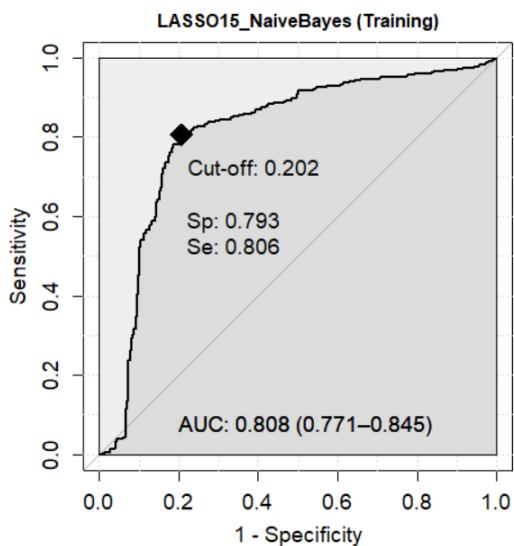


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	7
GBM	7	25

AUC = 0.816 [0.700 ; 0.932]
 Exactitude = 0.759 [0.648 ; 0.869]
 Sensibilité = 0.781 [0.675 ; 0.888]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.781 [0.675 ; 0.888]
 VPN = 0.731 [0.617 ; 0.845]
 χ^2 p-valeur = 0.00028
 McNemar p = 1
 F1-Score = 0.781 [0.675 ; 0.888]
 MCC = 0.477 [0.349 ; 0.606]
 Youden J = 0.512 [0.291 ; 0.733]

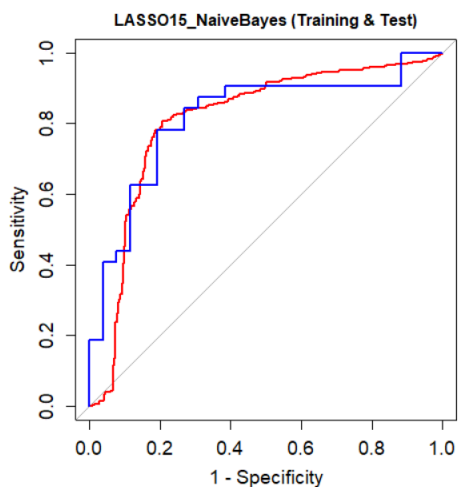
Echantillon Entraînement



AUC = 0.808 [0.771 ; 0.845]
 Exactitude = 0.775 [0.703 ; 0.848]

Hyperparamètres :
 Laplace smoothing = 0
 Usekernel = True
 Prior probabilities :
 - mc: 0.4488
 - gbm: 0.5512

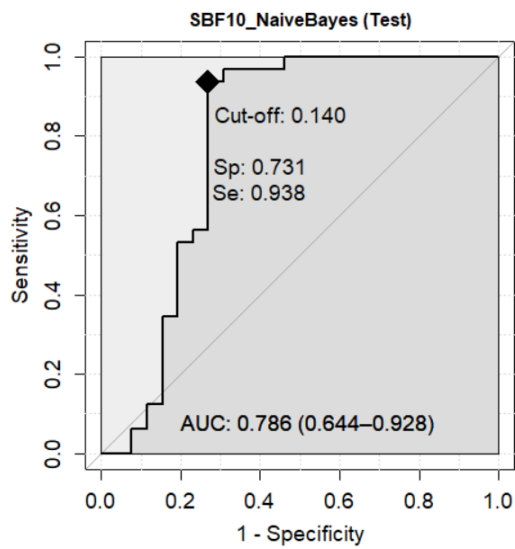
Comparaison ROC test et entraînement



Bootstrap test :
 $D = -0.12868$, boot.n = 2000, boot.stratified = 1, p-value = 0.8976

SBF10_NB

Echantillon Test

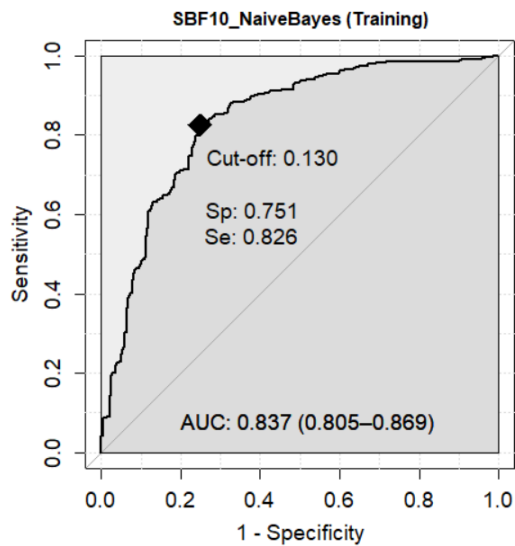


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	8
GBM	7	24

AUC = 0.786 [0.644 ; 0.928]
 Exactitude = 0.741 [0.629 ; 0.854]
 Sensibilité = 0.75 [0.639 ; 0.861]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.811]
 VPP = 0.774 [0.667 ; 0.882]
 VPN = 0.704 [0.586 ; 0.821]
 χ^2 p-valeur = 0.0007
 McNemar p = 1
 F1-Score = 0.762 [0.652 ; 0.872]
 MCC = 0.445 [0.317 ; 0.572]
 Youden J = 0.481 [0.255 ; 0.706]

Echantillon Entraînement



AUC = 0.837 [0.805 ; 0.869]
 Exactitude = 0.788 [0.717 ; 0.859]

Hyperparamètres :

Laplace smoothing = 0

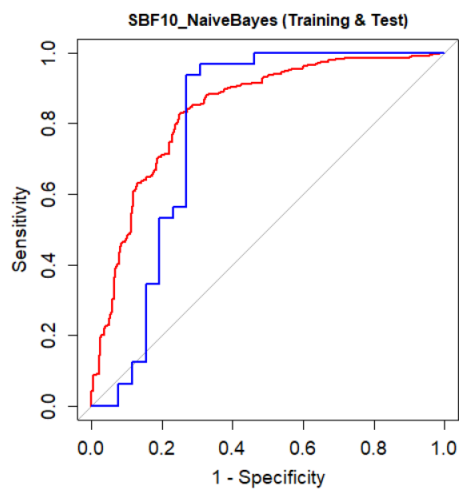
Usekernel = True

Prior probabilities :

- mc: 0.4488

- gbm: 0.5512

Comparaison ROC test et entraînement

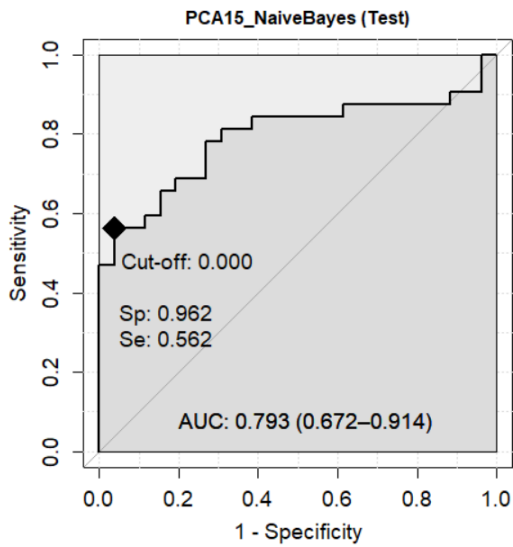


Bootstrap test :

D = 0.67853, boot.n = 2000, boot.stratified = 1, p-value = 0.4974

PCA15_NB

Echantillon Test

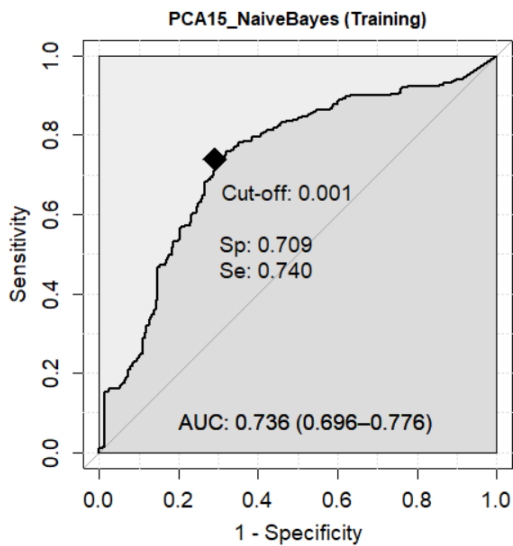


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	26	32
GBM	0	0

AUC = 0.793 [0.672 ; 0.914]
 Exactitude = 0.448 [0.320 ; 0.576]
 Sensibilité = 0
 (GBM est la classe positive)
 Spécificité = 1
 VPP = -
 VPN = 0.448 [0.320 ; 0.576]
 χ^2 p-valeur = -
 McNemar p = 4.10^{-8}
 F1-Score = 0
 MCC = -
 Youden J = 0

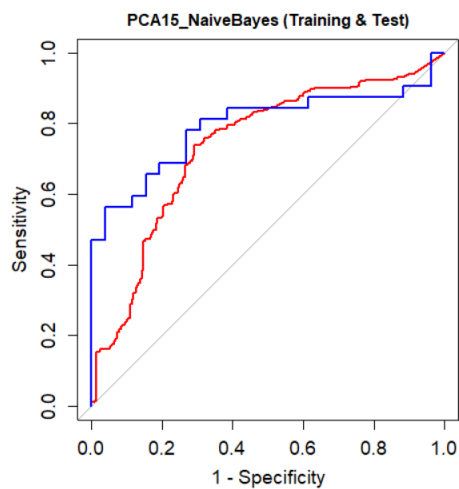
Echantillon Entraînement



AUC = 0.736 [0.696 ; 0.776]
 Exactitude = 0.681 [0.599 ; 0.762]

Hyperparamètres :
 Laplace smoothing = 0
 Usekernel = False
 Prior probabilities :
 - mc: 0.4488
 - gbm: 0.5512

Comparaison ROC test et entraînement



Bootstrap test :
 $D = -0.88919$, boot.n = 2000, boot.stratified = 1, p-value = 0.3739

2.6. Modèles k-plus-proches voisins

2.6.1. Résumé des performances des modèles k-plus-proches voisins

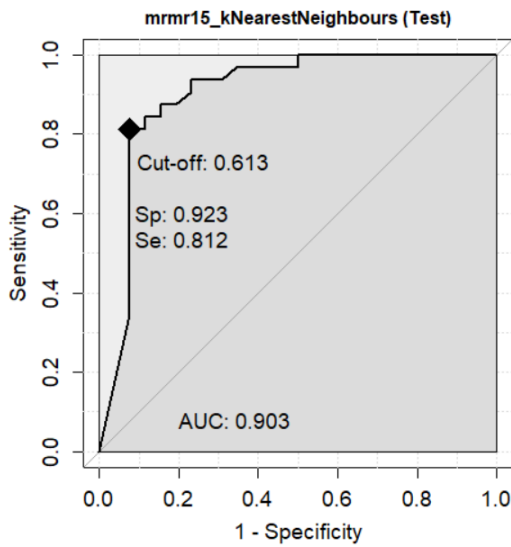
Les performances des modèles k-NN sont résumés dans le tableau 1.

Tableau 1. Performances des modèles k-plus-proches voisins sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_kNN	RFE11_kNN	LASSO15_kNN	SBF10_kNN	PCA15_kNN
Test					
AUC	0.903 [0.826 ; 0.979]	0.887 [0.806 ; 0.968]	0.859 [0.770 ; 0.949]	0.844 [0.750 ; 0.937]	0.815 [0.715 ; 0.915]
Exactitude	0.862 [0.773 ; 0.951]	0.845 [0.752 ; 0.938]	0.828 [0.730 ; 0.925]	0.828 [0.730 ; 0.925]	0.759 [0.648 ; 0.869]
Sensibilité	0.875 [0.790 ; 0.960]	0.875 [0.790 ; 0.960]	0.813 [0.712 ; 0.913]	0.875 [0.790 ; 0.960]	0.781 [0.675 ; 0.888]
Spécificité	0.846 [0.753 ; 0.939]	0.808 [0.706 ; 0.909]	0.846 [0.753 ; 0.939]	0.869 [0.661 ; 0.878]	0.731 [0.617 ; 0.845]
VPP	0.875 [0.90 ; 0.960]	0.848 [0.756 ; 0.941]	0.867 [0.779 ; 0.954]	0.824 [0.725 ; 0.922]	0.781 [0.675 ; 0.888]
VPN	0.846 [0.753 ; 0.939]	0.84 [0.746 ; 0.934]	0.786 [0.680 ; 0.891]	0.833 [0.737 ; 0.929]	0.731 [0.617 ; 0.845]
X ² p	1.10 ⁻⁷	7.10 ⁻⁷	2.10 ⁻⁶	3.10 ⁻⁶	0.0003
McNemar p	1	1	0,752	0,752	1
F1-score	0.875 [0.790 ; 0.960]	0.862 [0.773 ; 0.950]	0.839 [0.744 ; 0.933]	0.848 [0.756 ; 0.941]	0.781 [0.675 ; 0.888]
MCC	0.686 [0.567 ; 0.806]	0.651 [0.528 ; 0.773]	0.621 [0.496 ; 0.746]	0.615 [0.490 ; 0.741]	0.477 [0.349 ; 0.606]
Youden J	0.721 [0.543 ; 0.899]	0.683 [0.496 ; 0.869]	0.659 [0.465 ; 0.852]	0.644 [0.451 ; 0.838]	0.512 [0.291 ; 0.733]
Bootstrap p	0.8287	0.8884	0.9777	0.5703	0.8515
Entraînement					
AUC	0.892 [0.865 ; 0.919]	0.880 [0.851 ; 0.909]	0.861 [0.829 ; 0.893]	0.875 [0.847 ; 0.904]	0.804 [0.769 ; 0.839]
Exactitude	0.833 [0.768 ; 0.898]	0.827 [0.761 ; 0.893]	0.835 [0.770 ; 0.899]	0.828 [0.763 ; 0.894]	0.757 [0.682 ; 0.831]
Best Kernel	optimal	optimal	optimal	optimal	optimal
Best k	6	6	6	8	6
Distance	2	2	2	2	2

2.6.2. Courbes ROC et performances de prédiction par modèle kNN

mRMR15_kNN

Echantillon Test

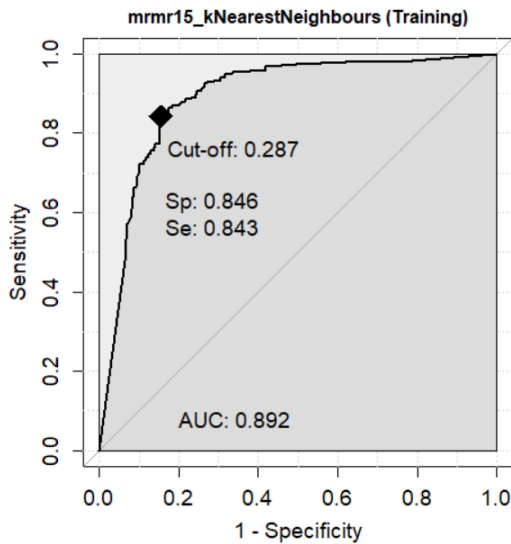


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	4
GBM	4	28

AUC = 0.903 [0.826 ; 0.979]
 Exactitude = 0.862 [0.773 ; 0.951]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.875 [0.90 ; 0.960]
 VPN = 0.846 [0.753 ; 0.939]
 χ^2 p-valeur = 1.10^{-7}
 McNemar p = 1
 F1-Score = 0.875 [0.790 ; 0.960]
 MCC = 0.686 [0.567 ; 0.806]
 Youden J = 0.721 [0.543 ; 0.899]

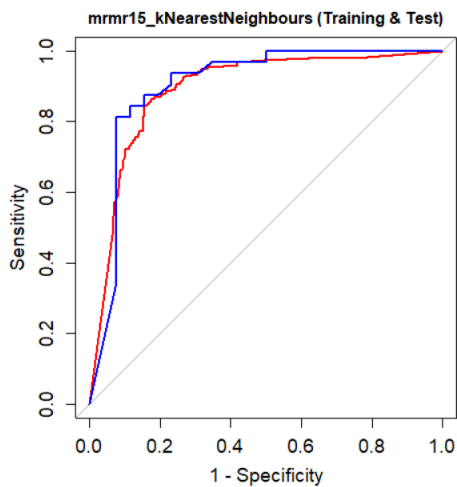
Echantillon Entraînement



AUC = 0.892 [0.865 ; 0.919]
 Exactitude = 0.833 [0.768 ; 0.898]

Hyperparamètres :
 Best kernel : optimal
 Best k = 6
 Distance = 2

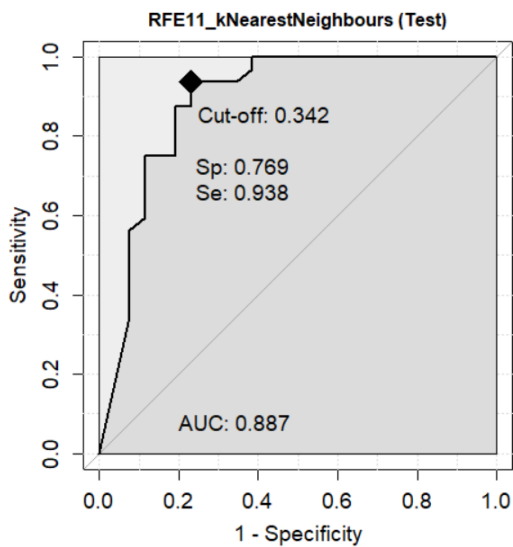
Comparaison ROC test et entraînement



Bootstrap test :
 D = -0.21631, boot.n = 2000, boot.stratified = 1, p-value = 0.8287

RFE11_kNN

Echantillon Test

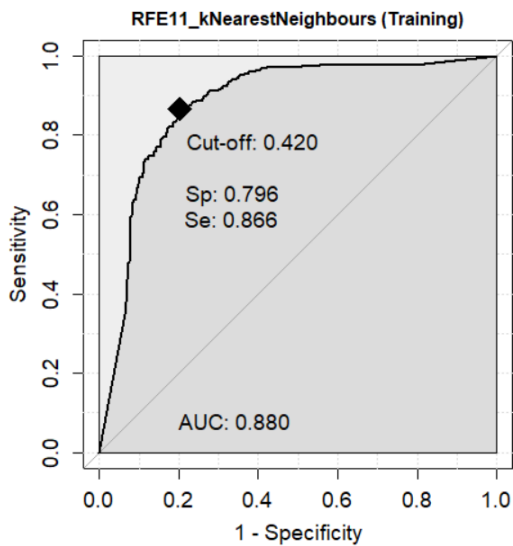


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	21	4
GBM	5	28

AUC = 0.887 [0.806 ; 0.968]
 Exactitude = 0.845 [0.752 ; 0.938]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.808 [0.706 ; 0.909]
 VPP = 0.848 [0.756 ; 0.941]
 VPN = 0.84 [0.746 ; 0.934]
 X^2 p-valeur = 7.10^{-7}
 McNemar p = 1
 F1-Score = 0.862 [0.773 ; 0.950]
 MCC = 0.651 [0.528 ; 0.773]
 Youden J = 0.683 [0.496 ; 0.869]

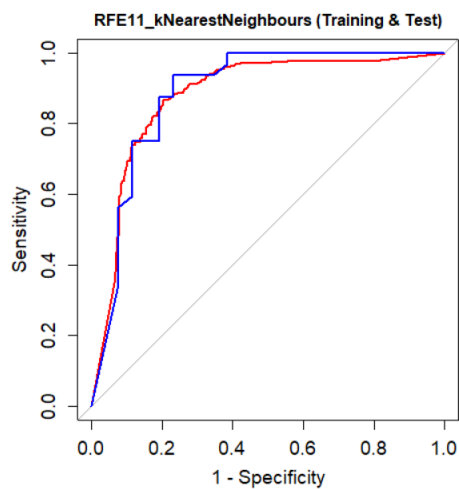
Echantillon Entraînement



AUC = 0.880 [0.851 ; 0.909]
 Exactitude = 0.827 [0.761 ; 0.893]

Hyperparamètres :
 Best kernel : optimal
 Best k = 6
 Distance = 2

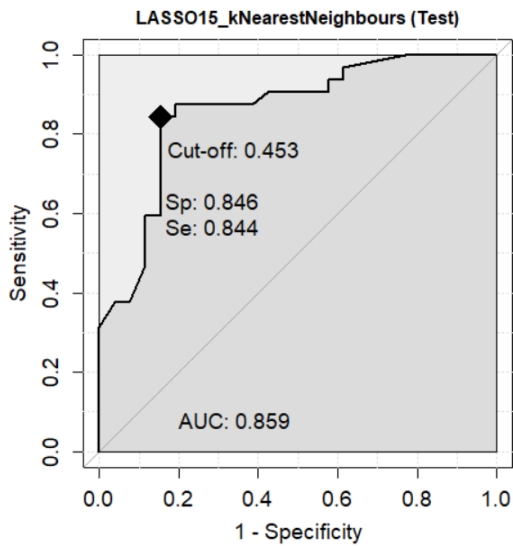
Comparaison ROC test et entraînement



Bootstrap test :
 $D = -0.14036$, boot.n = 2000, boot.stratified = 1, p-value = 0.8884

LASSO15_kNN

Echantillon Test

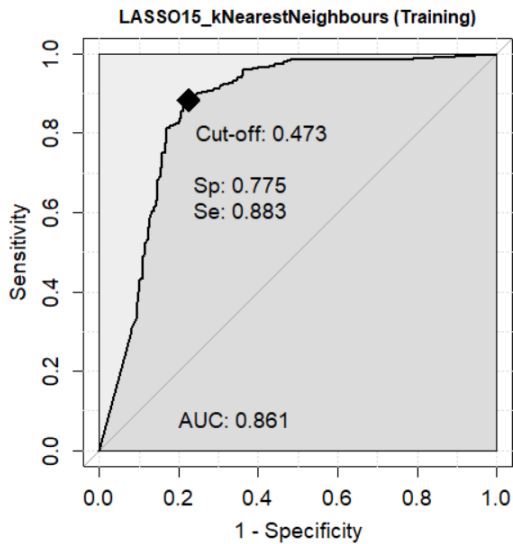


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	6
GBM	4	26

AUC = 0.859 [0.770 ; 0.949]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.867 [0.779 ; 0.954]
 VPN = 0.786 [0.680 ; 0.891]
 χ^2 p-valeur = $2 \cdot 10^{-6}$
 McNemar p = 0,752
 F1-Score = 0.839 [0.744 ; 0.933]
 MCC = 0.621 [0.496 ; 0.746]
 Youden J = 0.659 [0.465 ; 0.852]

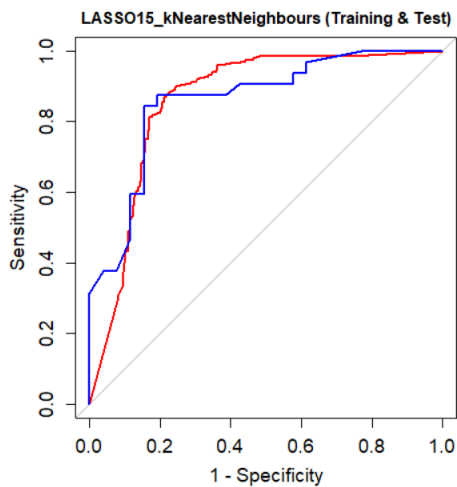
Echantillon Entraînement



AUC = 0.861 [0.829 ; 0.893]
 Exactitude = 0.835 [0.770 ; 0.899]

Hyperparamètres :
 Best kernel : optimal
 Best k = 6
 Distance = 2

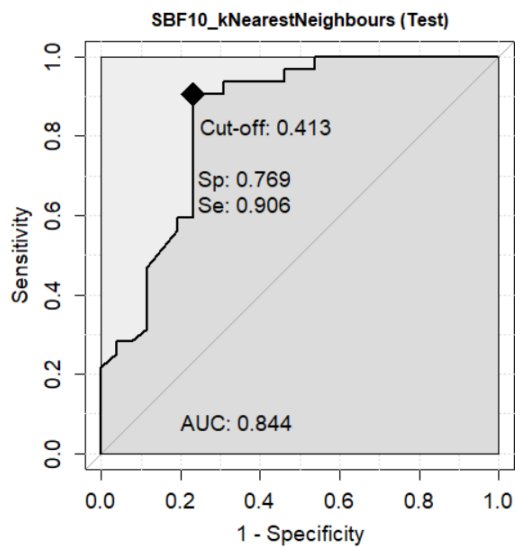
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.027923, boot.n = 2000,
 boot.stratified = 1, p-value = 0.9777

SBF10_kNN

Echantillon Test

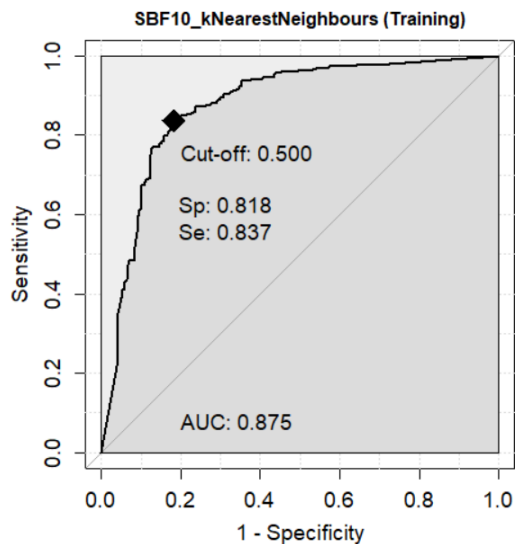


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	20	4
GBM	6	28

AUC = 0.844 [0.750 ; 0.937]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.869 [0.661 ; 0.878]
 VPP = 0.824 [0.725 ; 0.922]
 VPN = 0.833 [0.737 ; 0.929]
 X^2 p-valeur = $3 \cdot 10^{-6}$
 McNemar p = 0,752
 F1-Score = 0.848 [0.756 ; 0.941]
 MCC = 0.615 [0.490 ; 0.741]
 Youden J = 0.644 [0.451 ; 0.838]

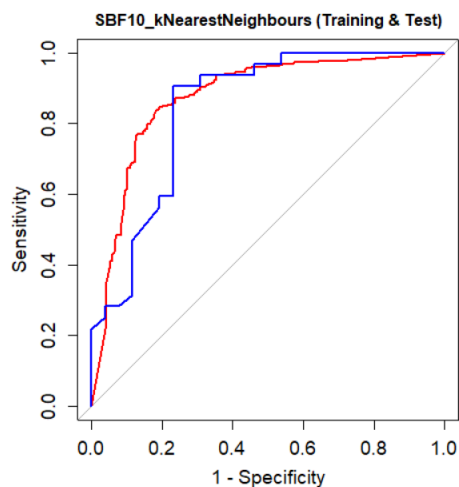
Echantillon Entraînement



AUC = 0.875 [0.847 ; 0.904]
 Exactitude = 0.828 [0.763 ; 0.894]

Hyperparamètres :
 Best kernel : optimal
 Best k = 8
 Distance = 2

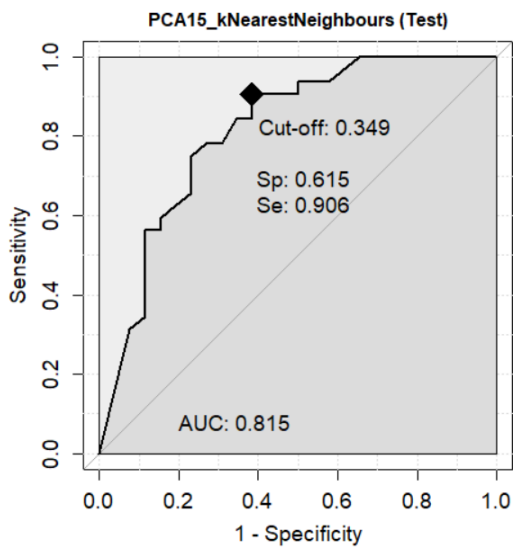
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.56756, boot.n = 2000, boot.stratified = 1, p-value = 0.5703

PCA15_kNN

Echantillon Test

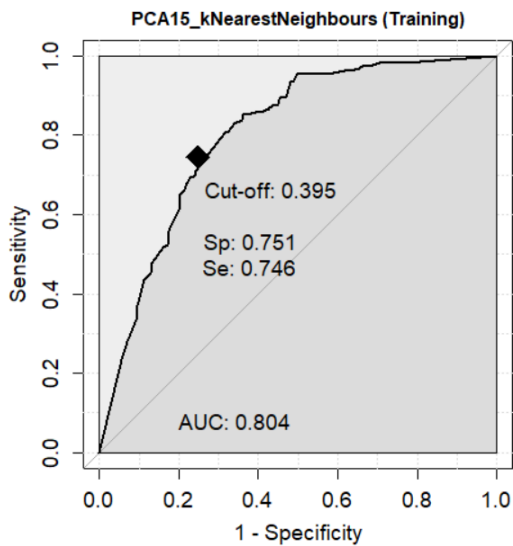


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	19	7
GBM	7	25

AUC = 0.815 [0.715 ; 0.915]
 Exactitude = 0.759 [0.648 ; 0.869]
 Sensibilité = 0.781 [0.675 ; 0.888]
 (GBM est la classe positive)
 Spécificité = 0.731 [0.617 ; 0.845]
 VPP = 0.781 [0.675 ; 0.888]
 VPN = 0.731 [0.617 ; 0.845]
 χ^2 p-valeur = 0.0003
 McNemar p = 1
 F1-Score = 0.781 [0.675 ; 0.888]
 MCC = 0.477 [0.349 ; 0.606]
 Youden J = 0.512 [0.291 ; 0.733]

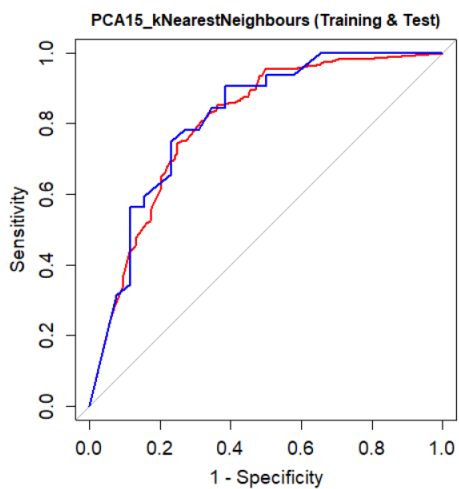
Echantillon Entraînement



AUC = 0.804 [0.769 ; 0.839]
 Exactitude = 0.757 [0.682 ; 0.831]

Hyperparamètres :
 Best kernel : optimal
 Best k = 6
 Distance = 2

Comparaison ROC test et entraînement



Bootstrap test :
 D = -0.18722, boot.n = 2000, boot.stratified = 1, p-value = 0.8515

2.7. Modèles Analyse Discriminante Linéaire

2.7.1. Résumé des performances des modèles Analyse Discriminante Linéaire

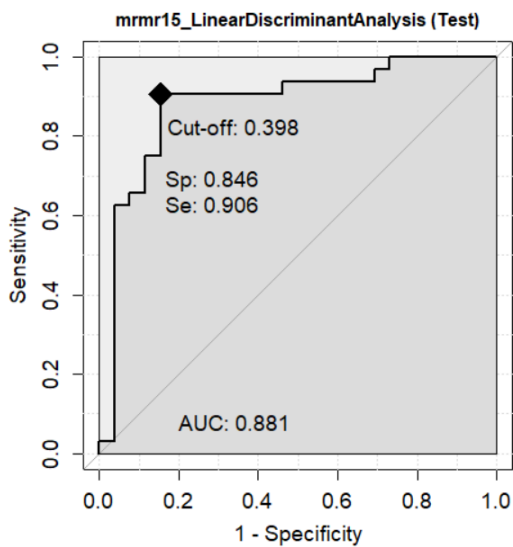
Les performances des modèles LDA sont résumés dans le tableau 1.

Tableau 1. Performances des modèles Analyse Discriminante Linéaire sur les échantillons d'entraînement et de test. Valeurs avec intervalle de confiance à 95%.					
Statistique / Modèle	mRMR15_LDA	RFE11_LDA	LASSO15_LDA	SBF10_LDA	PCA15_LDA
Test					
AUC	0.881 [0.798 ; 0.964]	0.924 [0.856 ; 0.992]	0.893 [0.813 ; 0.973]	0.880 [0.796 ; 0.963]	0.862 [0.753 ; 0.939]
Exactitude	0.862 [0.773 ; 0.951]	0.828 [0.730 ; 0.925]	0.793 [0.689 ; 0.897]	0.828 [0.730 ; 0.925]	0.759 [0.648 ; 0.869]
Sensibilité	0.875 [0.790 ; 0.960]	0.813 [0.712 ; 0.913]	0.813 [0.712 ; 0.913]	0.844 [0.750 ; 0.937]	0.75 [0.639 ; 0.861]
Spécificité	0.846 [0.753 ; 0.939]	0.846 [0.753 ; 0.939]	0.769 [0.661 ; 0.878]	0.808 [0.706 ; 0.909]	0.769 [0.661 ; 0.878]
VPP	0.875 [0.790 ; 0.960]	0.867 [0.780 ; 0.954]	0.742 [0.629 ; 0.855]	0.844 [0.750 ; 0.937]	0.8 [0.697 ; 0.903]
VPN	0.846 [0.753 ; 0.939]	0.786 [0.680 ; 0.891]	0.769 [0.661 ; 0.878]	0.808 [0.706 ; 0.909]	0.714 [0.598 ; 0.831]
X ² p	1.10 ⁻⁷	2.10 ⁻⁶	3.10 ⁻⁵	3.10 ⁻⁶	0.0002
McNemar p	1	0,752	1	1	0,79
F1-score	0.875 [0.790 ; 0.960]	0.776 [0.669 ; 0.883]	0.813 [0.712 ; 0.913]	0.844 [0.750 ; 0.937]	0.774 [0.667 ; 0.882]
MCC	0.686 [0.567 ; 0.806]	0.621 [0.496 ; 0.746]	0.547 [0.419 ; 0.675]	0.617 [0.491 ; 0.742]	0.482 [0.353 ; 0.611]
Youden J	0.721 [0.543 ; 0.899]	0.839 [0.744 ; 0.933]	0.582 [0.373 ; 0.791]	0.651 [0.457 ; 0.846]	0.519 [0.299 ; 0.739]
Bootstrap p	0.6932	0.5707	0.6675	0.2986	0.1555
Entraînement	mRMR15_LDA	RFE11_LDA	LASSO15_LDA	SBF10_LDA	PCA15_LDA
AUC	0.902 [0.877 ; 0.927]	0.901 [0.877 ; 0.926]	0.912 [0.889 ; 0.935]	0.828 [0.794 ; 0.861]	0.768 [0.729 ; 0.807]
Exactitude	0.837 [0.773 ; 0.902]	0.816 [0.748 ; 0.883]	0.846 [0.783 ; 0.908]	0.751 [0.676 ; 0.826]	0.742 [0.666 ; 0.818]

2.7.2. Courbes ROC et performances de prédiction par modèle LDA

mRMR15_LDA

Echantillon Test

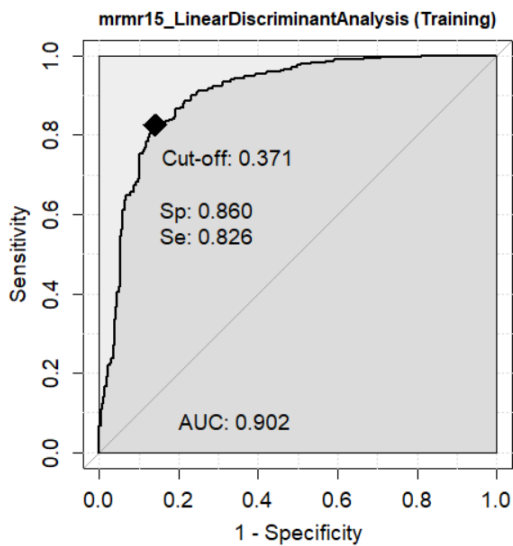


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	4
GBM	4	28

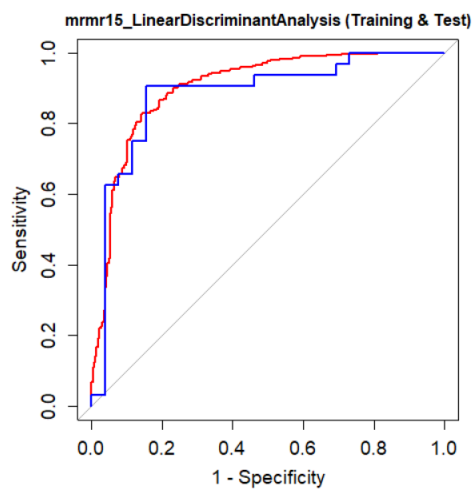
AUC = 0.881 [0.798 ; 0.964]
 Exactitude = 0.862 [0.773 ; 0.951]
 Sensibilité = 0.875 [0.790 ; 0.960]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.875 [0.790 ; 0.960]
 VPN = 0.846 [0.753 ; 0.939]
 X^2 p-valeur = 1.10^{-7}
 McNemar p = 1
 F1-Score = 0.875 [0.790 ; 0.960]
 MCC = 0.686 [0.567 ; 0.806]
 Youden J = 0.721 [0.543 ; 0.899]

Echantillon Entraînement



AUC = 0.902 [0.877 ; 0.927]
 Exactitude = 0.837 [0.773 ; 0.902]

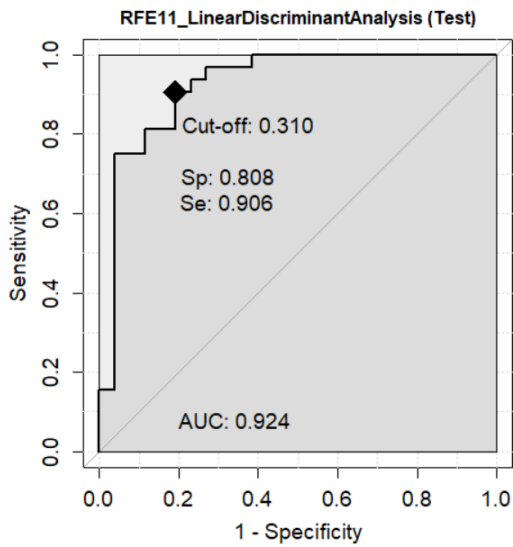
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.39451, boot.n = 2000, boot.stratified = 1, p-value = 0.6932

RFE11_LDA

Echantillon Test

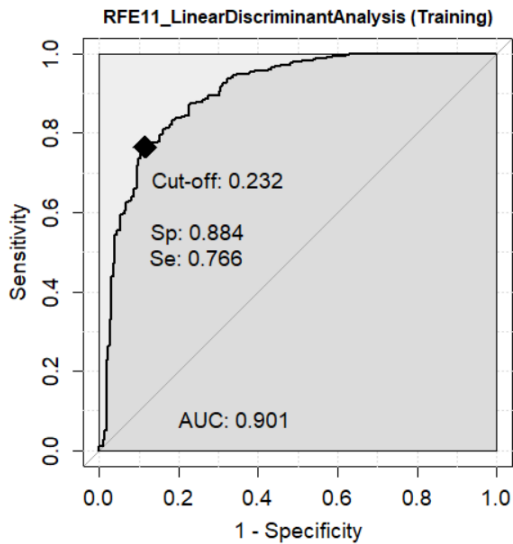


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	22	6
GBM	4	26

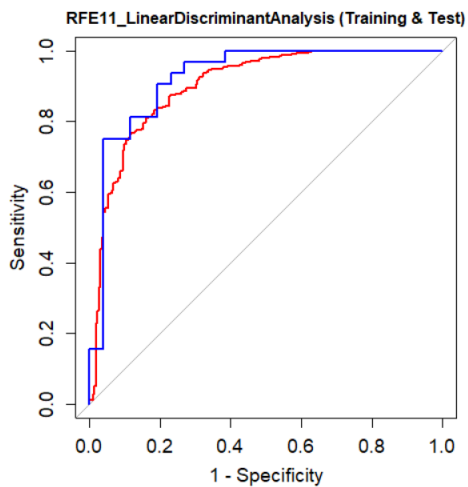
AUC = 0.924 [0.856 ; 0.992]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.846 [0.753 ; 0.939]
 VPP = 0.867 [0.780 ; 0.954]
 VPN = 0.786 [0.680 ; 0.891]
 X^2 p-valeur = $2 \cdot 10^{-6}$
 McNemar p = 0,752
 F1-Score = 0.776 [0.669 ; 0.883]
 MCC = 0.621 [0.496 ; 0.746]
 Youden J = 0.839 [0.744 ; 0.933]

Echantillon Entraînement



AUC = 0.901 [0.877 ; 0.926]
 Exactitude = 0.816 [0.748 ; 0.883]

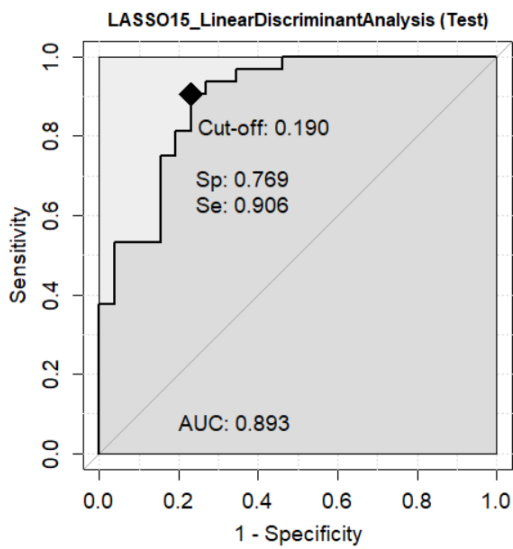
Comparaison ROC test et entraînement



Bootstrap test :
 $D = -0.56695$, boot.n = 2000, boot.stratified = 1, p-value = 0.5707

LASSO15_LDA

Echantillon Test

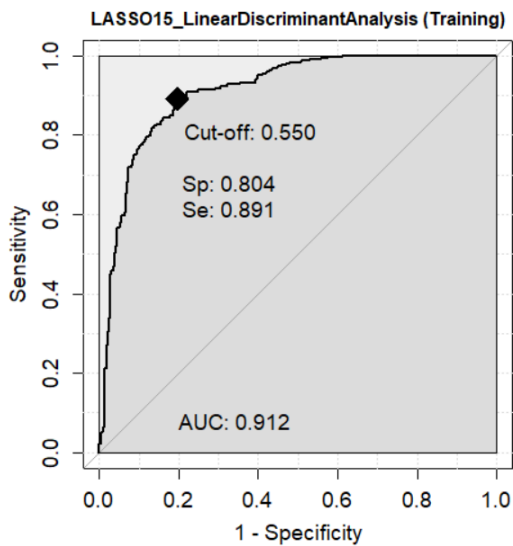


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	20	6
GBM	6	26

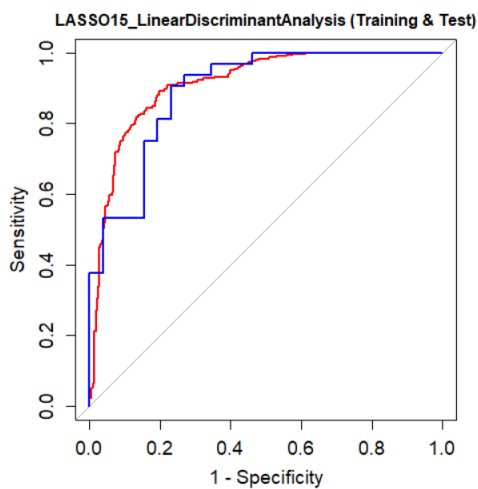
AUC = 0.893 [0.813 ; 0.973]
 Exactitude = 0.793 [0.689 ; 0.897]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.769 [0.661 ; 0.878]
 VPP = 0.742 [0.629 ; 0.855]
 VPN = 0.769 [0.661 ; 0.878]
 X^2 p-valeur = $3 \cdot 10^{-5}$
 McNemar p = 1
 F1-Score = 0.813 [0.712 ; 0.913]
 MCC = 0.547 [0.419 ; 0.675]
 Youden J = 0.582 [0.373 ; 0.791]

Echantillon Entraînement



AUC = 0.912 [0.889 ; 0.935]
 Exactitude = 0.846 [0.783 ; 0.908]

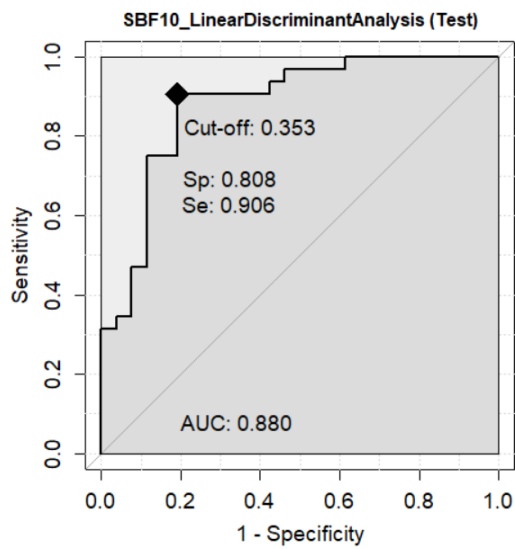
Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.42958, boot.n = 2000, boot.stratified = 1, p-value = 0.6675

SBF10_LDA

Echantillon Test

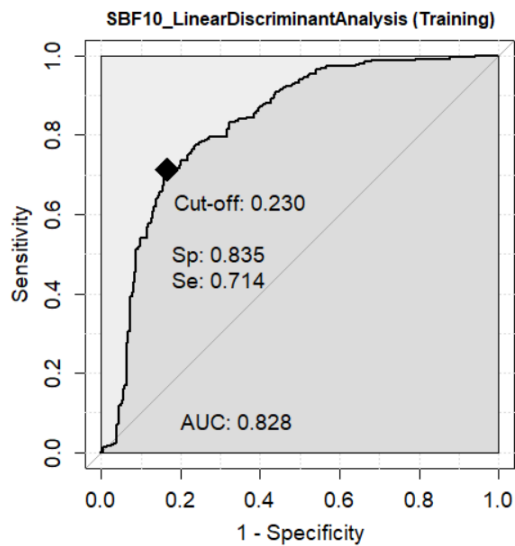


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	21	5
GBM	5	27

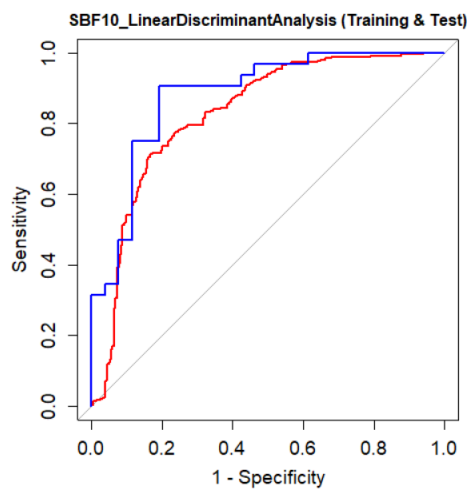
AUC = 0.880 [0.796 ; 0.963]
 Exactitude = 0.828 [0.730 ; 0.925]
 Sensibilité = 0.844 [0.750 ; 0.937]
 (GBM est la classe positive)
 Spécificité = 0.808 [0.706 ; 0.909]
 VPP = 0.844 [0.750 ; 0.937]
 VPN = 0.808 [0.706 ; 0.909]
 X^2 p-valeur = 3.10^{-6}
 McNemar p = 1
 F1-Score = 0.844 [0.750 ; 0.937]
 MCC = 0.617 [0.491 ; 0.742]
 Youden J = 0.651 [0.457 ; 0.846]

Echantillon Entraînement



AUC = 0.828 [0.794 ; 0.861]
 Exactitude = 0.751 [0.676 ; 0.826]

Comparaison ROC test et entraînement

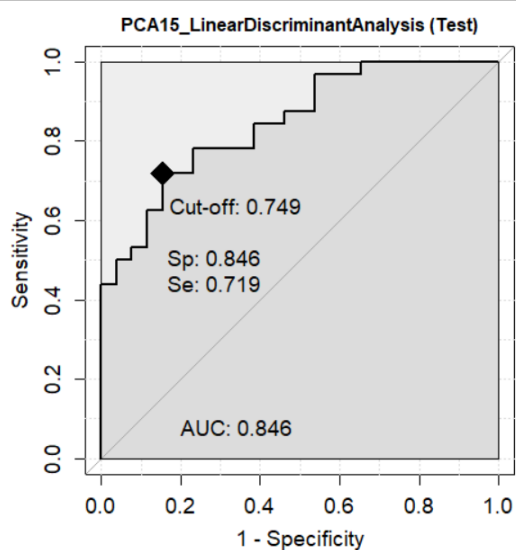


Bootstrap test :

D = -1.0394, boot.n = 2000, boot.stratified = 1, p-value = 0.2986

PCA15_LDA

Echantillon Test

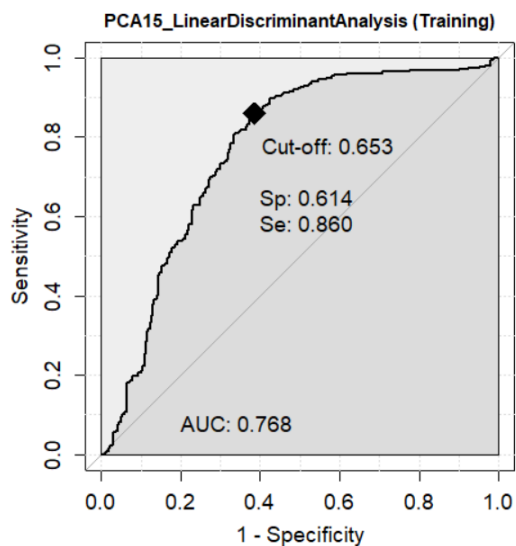


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	20	8
GBM	6	24

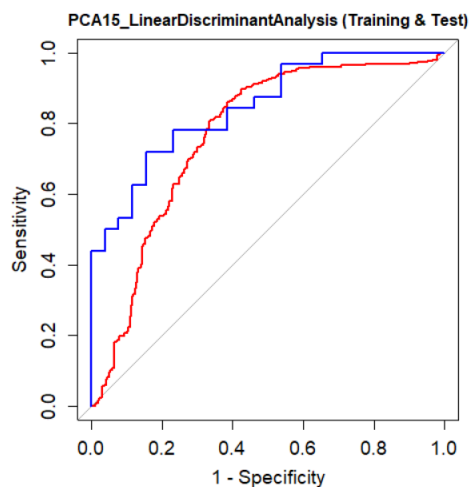
AUC = 0.862 [0.753 ; 0.939]
 Exactitude = 0.759 [0.648 ; 0.869]
 Sensibilité = 0.75 [0.639 ; 0.861]
 (GBM est la classe positive)
 Spécificité = 0.769 [0.661 ; 0.878]
 VPP = 0.8 [0.697 ; 0.903]
 VPN = 0.714 [0.598 ; 0.831]
 χ^2 p-valeur = 0.0002
 McNemar p = 0,79
 F1-Score = 0.774 [0.667 ; 0.882]
 MCC = 0.482 [0.353 ; 0.611]
 Youden J = 0.519 [0.299 ; 0.739]

Echantillon Entraînement



AUC = 0.768 [0.729 ; 0.807]
 Exactitude = 0.742 [0.666 ; 0.818]

Comparaison ROC test et entraînement



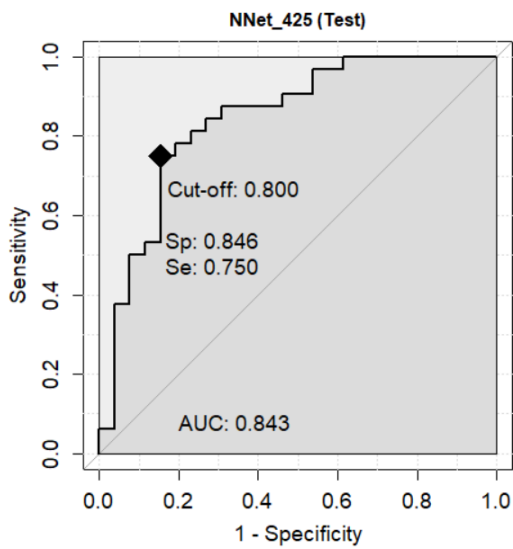
Bootstrap test :
 $D = -1.4202$, boot.n = 2000, boot.stratified = 1, p-value = 0.1555

2.8. Performances et courbe ROC du modèle NNet

Les performances du modèle NNet sont résumés dans le tableau suivant :

NNet

Echantillon Test

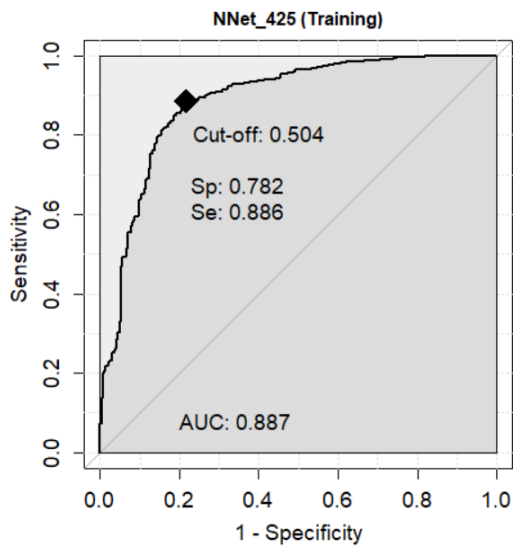


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	20	6
GBM	6	26

AUC = 0.843 [0.736 ; 0.949]
 Exactitude = 0.793 [0.689 ; 0.897]
 Sensibilité = 0.813 [0.712 ; 0.913]
 (GBM est la classe positive)
 Spécificité = 0.769 [0.661 ; 0.878]
 VPP = 0.813 [0.712 ; 0.913]
 VPN = 0.769 [0.661 ; 0.878]
 χ^2 p-valeur = 3.10^{-5}
 McNemar p = 1
 F1-Score = 0.813 [0.712 ; 0.913]
 MCC = 0.547 [0.419 ; 0.675]
 Youden J = 0.582 [0.373 ; 0.791]

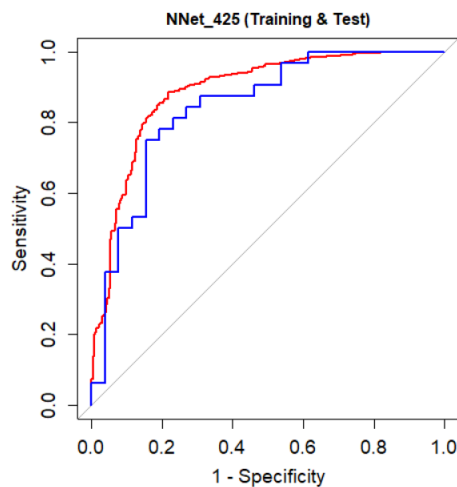
Echantillon Entraînement



AUC = 0.887 [0.860 ; 0.914]
 Exactitude = 0.75 [0.639 ; 0.861]

Paramètres :
 Architecture :
 1 couche, 440-1-2 (443 poids)
 Decay = 0.1
 Entropy fitting
 Softmax

Comparaison ROC test et entraînement



Bootstrap test :
 D = 0.80696, boot.n = 2000, boot.stratified = 1, p-value = 0.4197

2.9. Performances et score du modèle CNN en parallèle

Le réseau de neurones convolutifs 2D (CNN) était entraîné sur 7267 images issues de 82 mesures (44,3%) de l'échantillon d'entraînement, avec validation interne sur 3696 images représentant 45 mesures (24,3%). Puis, les prédictions étaient effectuées sur l'échantillon de test avec 6 images par patient, correspondant aux 6 segmentations, pour 58 lésions test (31,4%).

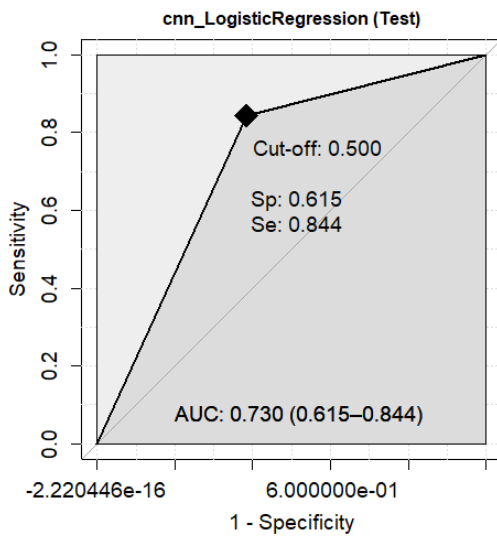
Les performances étaient évaluées par régression logistique sur le score CNN.

L'AUC test était de 0.730 [0.615-0.844] contre 0.713 [0.680 ; 0.746] à l'entraînement. L'exactitude test était de 0.741[0.629 ; 0.854] avec une sensibilité de 0.844 [0.750 ; 0.937] pour le GBM, et une spécificité de 0.615 [0.490 ; 0.741]. Les performances du CNN étaient également testées avec le classifieur SVM correspondant au meilleur modèle.

Les résultats du CNN et ses paramètres sont résumés ci-dessous :

CNN (Régression Logistique)

Echantillon Test

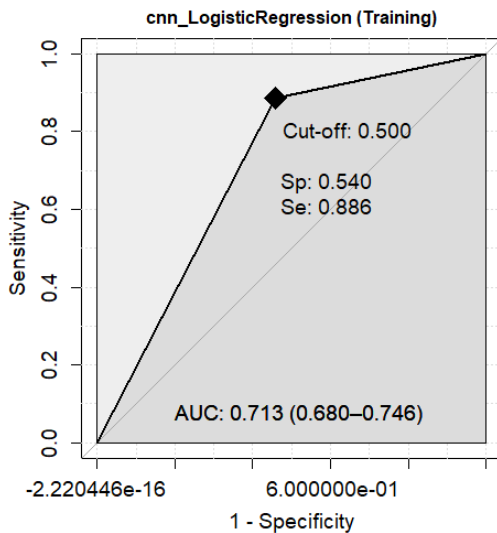


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	16	5
GBM	10	27

AUC = 0.730 [0.615 ; 0.844]
 Exactitude = 0.741 [0.629 ; 0.854]
 Sensibilité = 0.844 [0.750 ; 0.937]
 (GBM est la classe positive)
 Spécificité = 0.615 [0.490 ; 0.741]
 VPP = 0.730 [0.615 ; 0.844]
 VPN = 0.762 [0.652 ; 0.872]
 χ^2 p-valeur = 0.0008
 McNemar p = 1
 F1-Score = 0.783 [0.676 ; 0.889]
 MCC = 0.439 [0.311 ; 0.567]
 Youden J = 0.459 [0.240 ; 0.678]

Echantillon Entraînement



AUC = 0.713 [0.680 ; 0.746]
 Exactitude = 0.680 [0.638 ; 0.721]

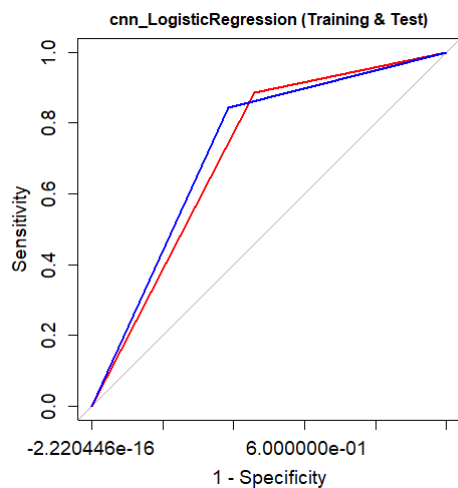
Paramètres :

Architecture Séquentielle 2D : Input,128,3
 → Conv2D,64,3 → MP2 → Conv2D,128,3
 → MP2 → FC,64 → FC,32 → FC,32 →
 FC,10 → Softmax

Hyperparamètres :

- L2 regularization = 0.01
- Learning rate = 0.01
- Decay = 1E-6
- Optimizer = Stochastic Gradient Descent
- Batch size = 32
- Metric = val_loss

Comparaison ROC test et entraînement



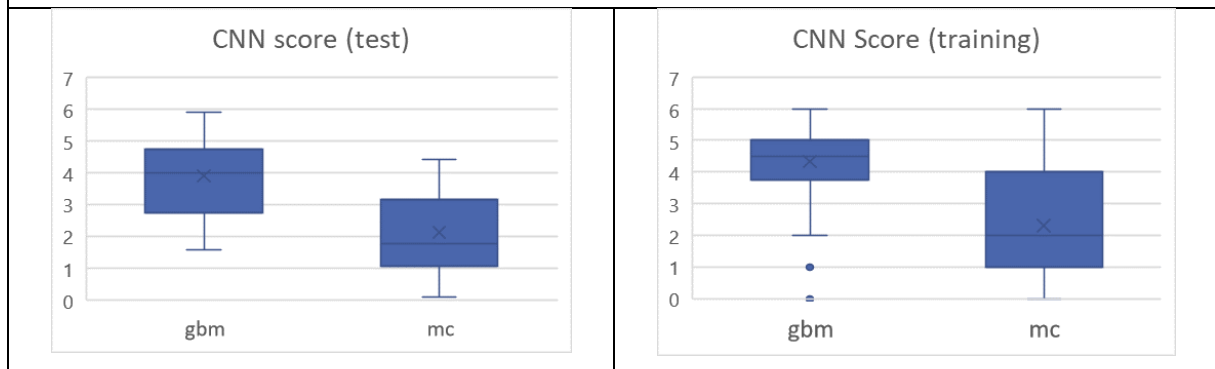
Bootstrap test :

D = -0.28541, boot.n = 2000, boot.stratified = 1, p-value = 0.7753

Le score CNN correspondant à la somme des 6 probabilités (régions (a) (b) et (d) en séquences T1 et FLAIR) était de 3,88 [3.45 ; 4.33] pour les GBM contre 2.11 [1.64 ; 2.59] pour les MC dans l'échantillon test, avec une différence significative ($p=0.0054$) (tableau 1).

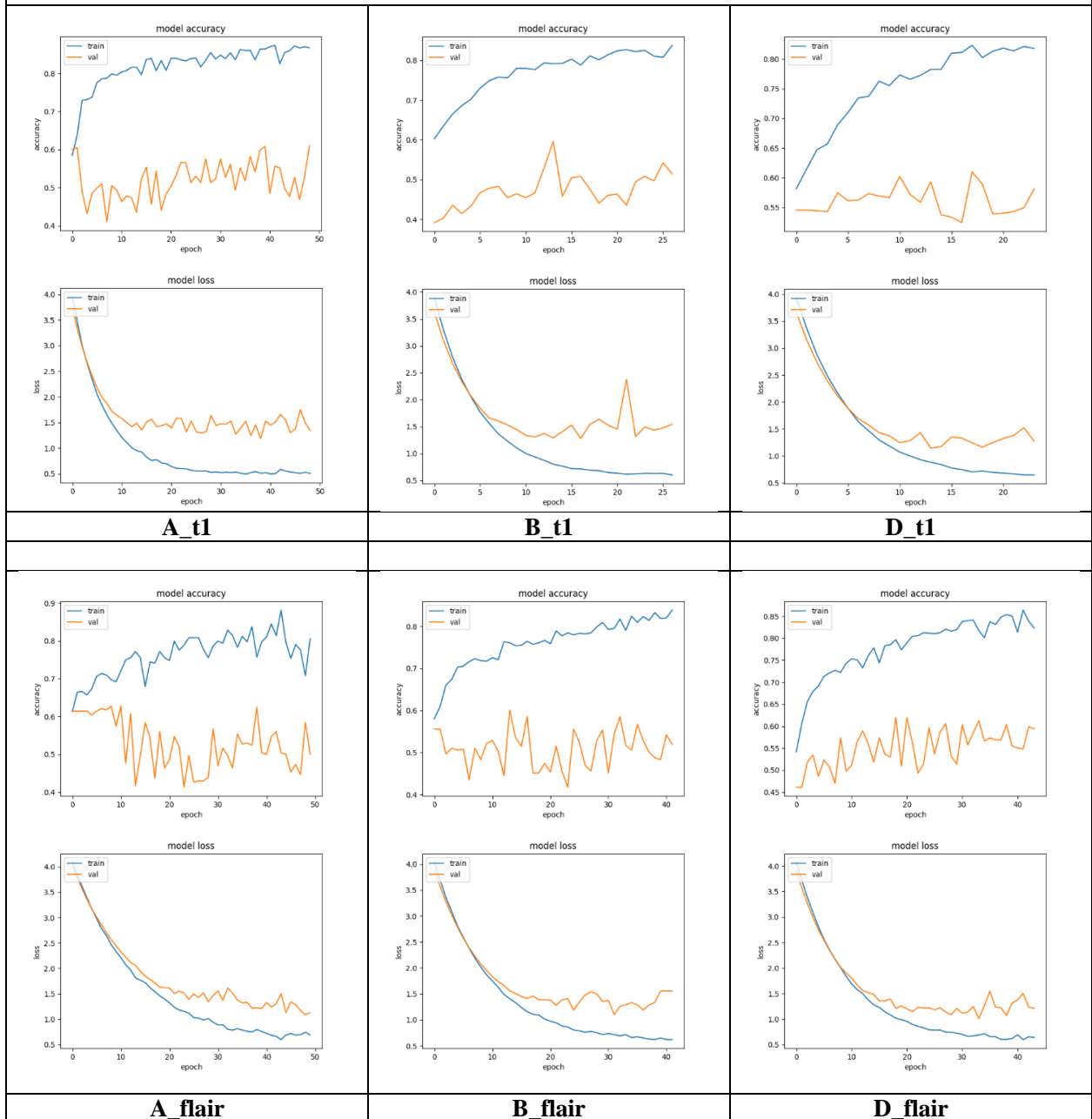
Le score était ensuite rétro-propagé sur les sujets de l'échantillon d'entraînement puis injecté aux données radiomiques et cliniques pour l'objectif principal.

Tableau 1. Représentation du score CNN sur les échantillons de test et d'entraînement. Score CNN test des GBM : 3,88 [3.45 ; 4.33]. Score CNN test des MC : 2.11 [1.64 ; 2.59]. Score CNN entraînement des GBM : 4.10 [3.82 ; 4.38]. Score CNN entraînement des MC : 2.46 [2.18 – 2.74].



Les performances d'entraînement sont résumées dans le tableau 2.

Tableau 2. Evolution de l'exactitude et de la fonction de perte lors de l'entrainement et de la validation interne en fonction des epochs. La divergence entre les fonctions de perte de l'entrainement (loss train) et de la validation interne (loss val) indiquent un risque d'overfitting. L'optimum des poids est défini par le minimum de la fonction de perte lors de la validation interne, lorsqu'un nouveau minimum n'est plus retrouvé depuis 10 cycles (epochs).

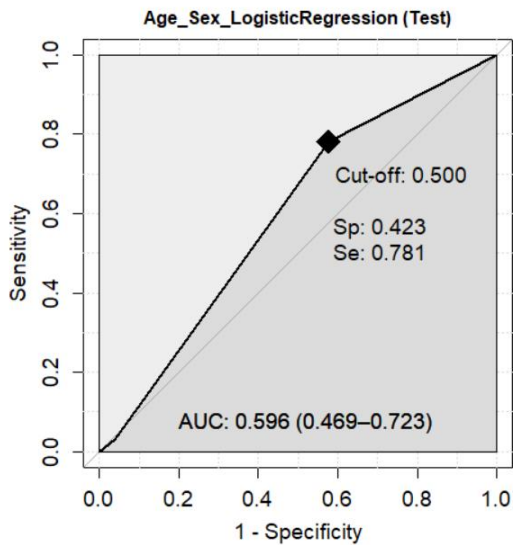


2.10. Performances des modèles cliniques et IRM

Les performances des modèles cliniques et IRM sont résumés dans les tableaux suivants :

Modèle clinique : Age et sexe (Régression Logistique)

Echantillon Test

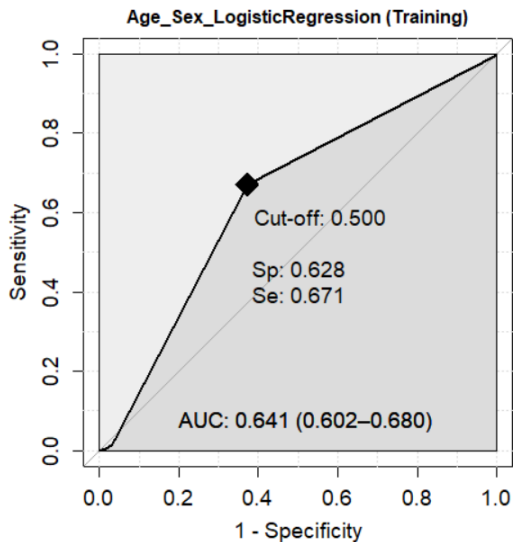


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	11	7
GBM	15	25

AUC = 0.596 [0.469 ; 0.723]
 Exactitude = 0.621 [0.496 ; 0.746]
 Sensibilité = 0.781 [0.675 ; 0.888]
 (GBM est la classe positive)
 Spécificité = 0.423 [0.296 ; 0.550]
 VPP = 0.625 [0.500 ; 0.750]
 VPN = 0.611 [0.486 ; 0.737]
 χ^2 p-valeur = 0.165
 McNemar p = 0.136
 F1-Score = 0.694 [0.576 ; 0.813]
 MCC = 0.182 [0.083 ; 0.282]
 Youden J = 0.204 [0.0 ; 0.438]

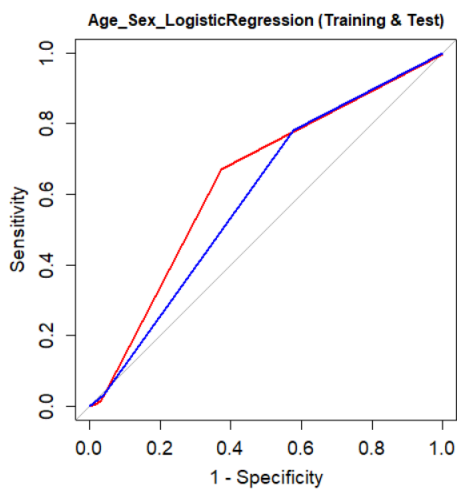
Echantillon Entraînement



AUC = 0.641 [0.605 ; 0.680]
 Exactitude = 0.652 [0.569 ; 0.735]

N Iter = 11

Comparaison ROC test et entraînement

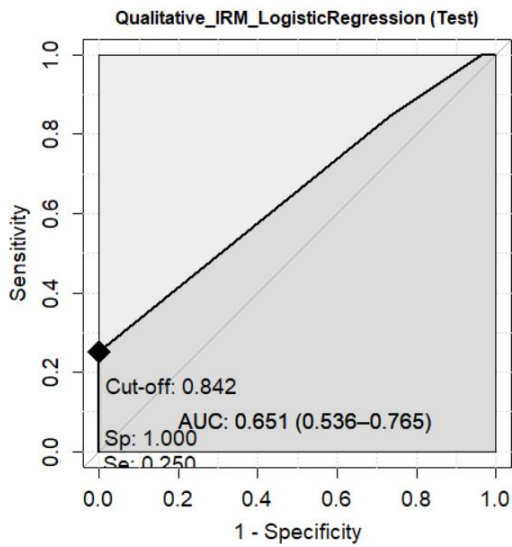


Bootstrap test :

D = 0.653, boot.n = 2000, boot.stratified = 1, p-value = 0.5138

Modèle descriptif IRM (Régression Logistique) :

Echantillon Test

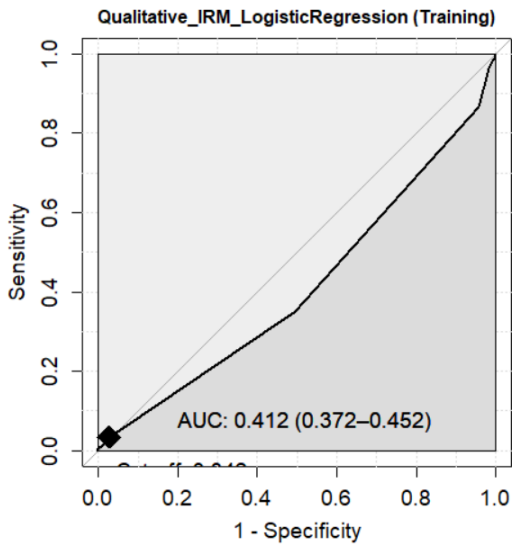


Matrice de confusion :

Prediction	Classe réelle	
	MC	GBM
MC	7	5
GBM	19	27

AUC = 0.596 [0.469 ; 0.723]
 Exactitude = 0.586 [0.459 ; 0.713]
 Sensibilité = 0.587 [0.460 ; 0.714]
 (GBM est la classe positive)
 Spécificité = 0.269 [0.155 ; 0.383]
 VPP = 0.625 [0.500 ; 0.750]
 VPN = 0.583 [0.456 ; 0.710]
 χ^2 p-valeur = 0.465
 McNemar p = 0.008
 F1-Score = 0.692 [0.574 ; 0.811]
 MCC = 0.096 [0.02 ; 0.172]
 Youden J = 0.113 [0.0 ; 0.321]

Echantillon Entraînement



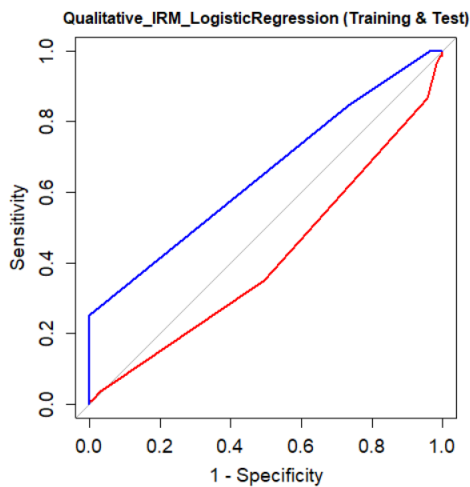
AUC = 0.641 [0.605 ; 0.680]
 Exactitude = 0.580 [0.494 ; 0.666]

N Iter = 21

Variables binaires :

Atteinte profonde / Atteinte corticale /
 Croisement Flair / Nombre de lésions /
 Atteinte sous tentorielle / Rehaussement perceptible

Comparaison ROC test et entraînement



Bootstrap test :

D = -3.8591, df = 71.587, p-value = 0.0002468

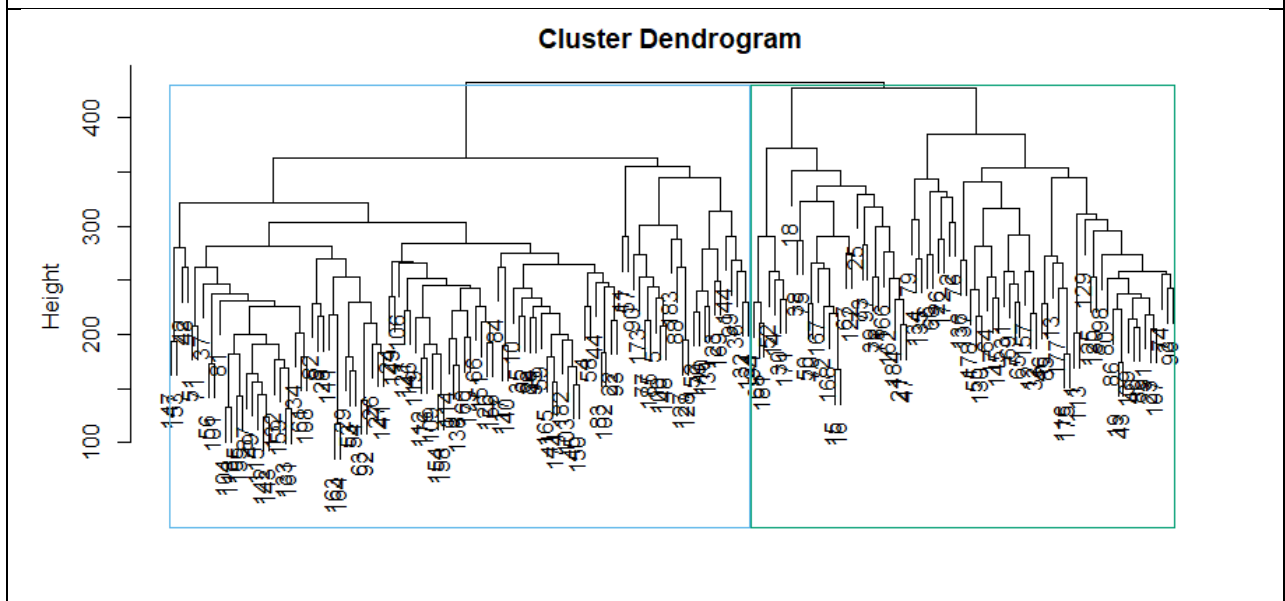
3. Objectifs secondaires

3.1. Importance et profils des index radiomiques

3.1.1. Sélection des index radiomiques

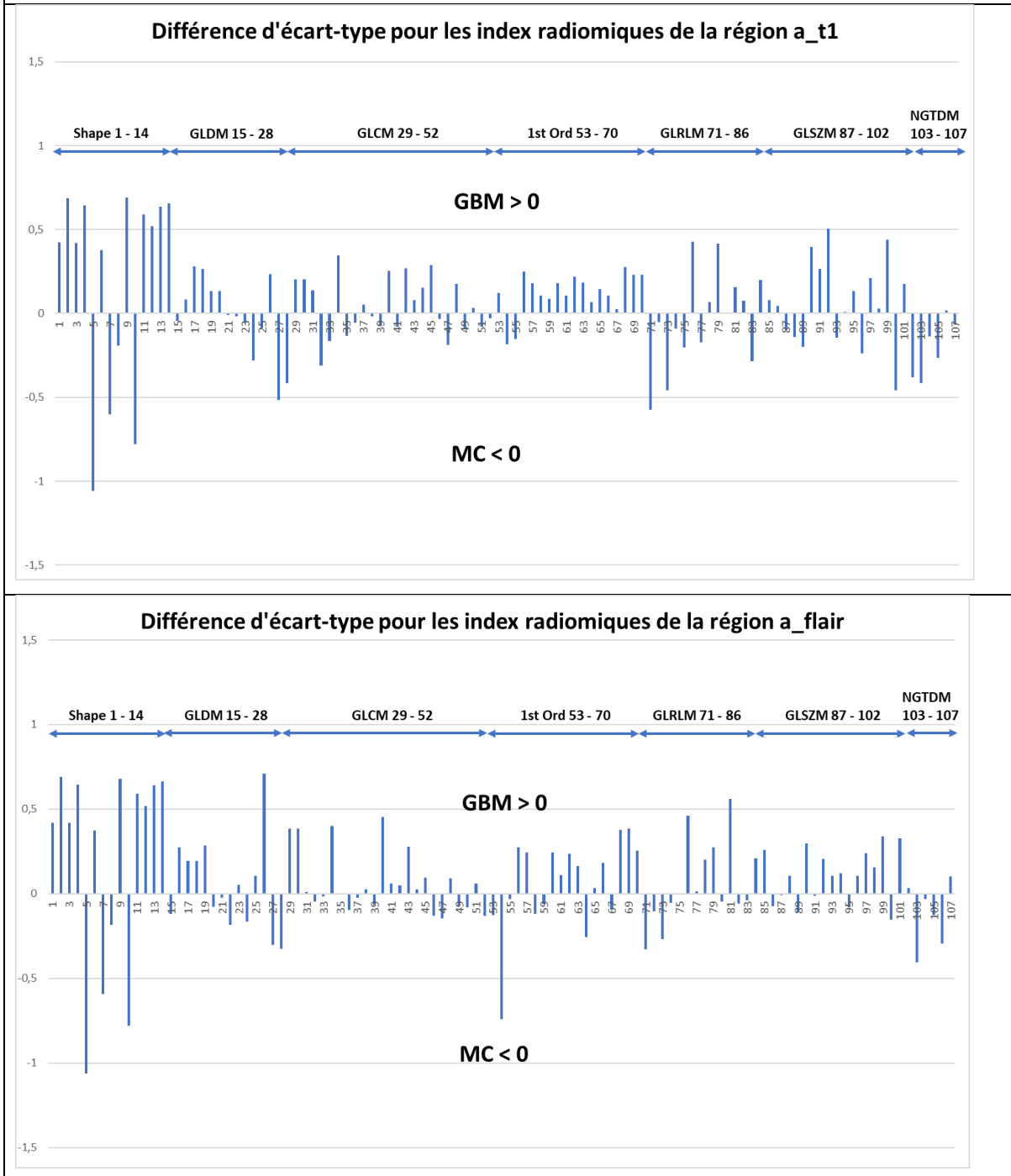
Après calcul et extraction de 642 index radiomiques (107 index par région et par séquence), ceux-ci étaient visualisés à l'aide d'un regroupement hiérarchique (figure 1) permettant d'observer la séparabilité potentielle des deux groupes.

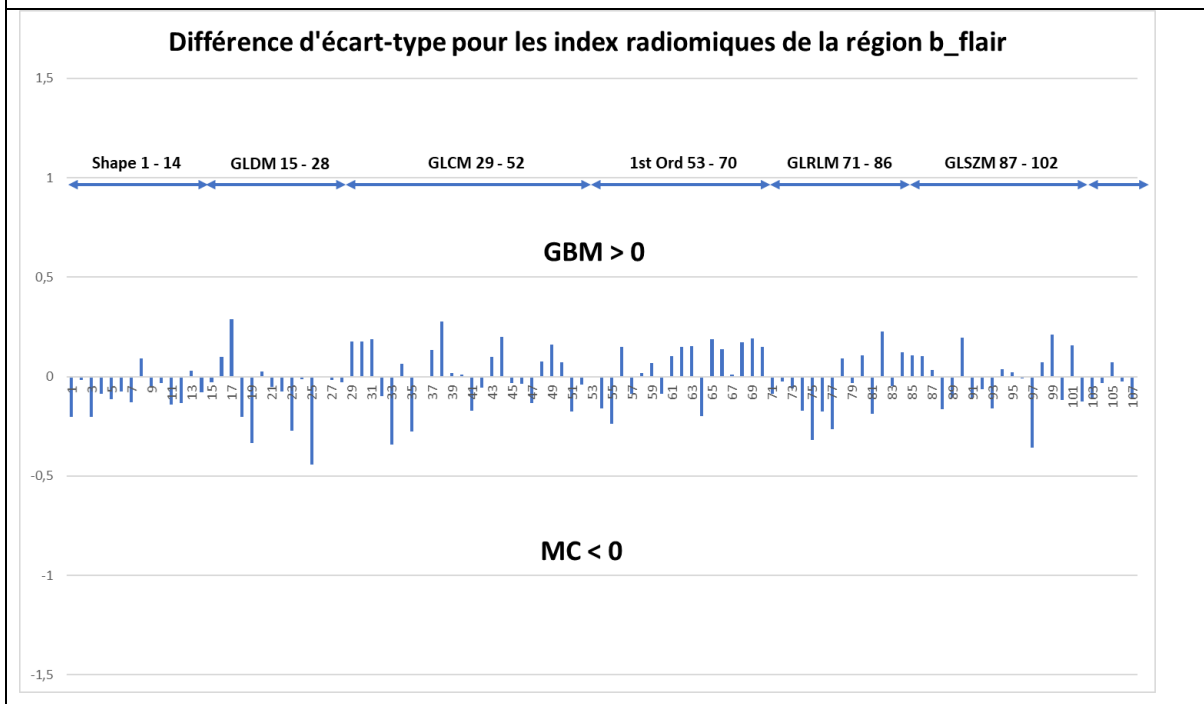
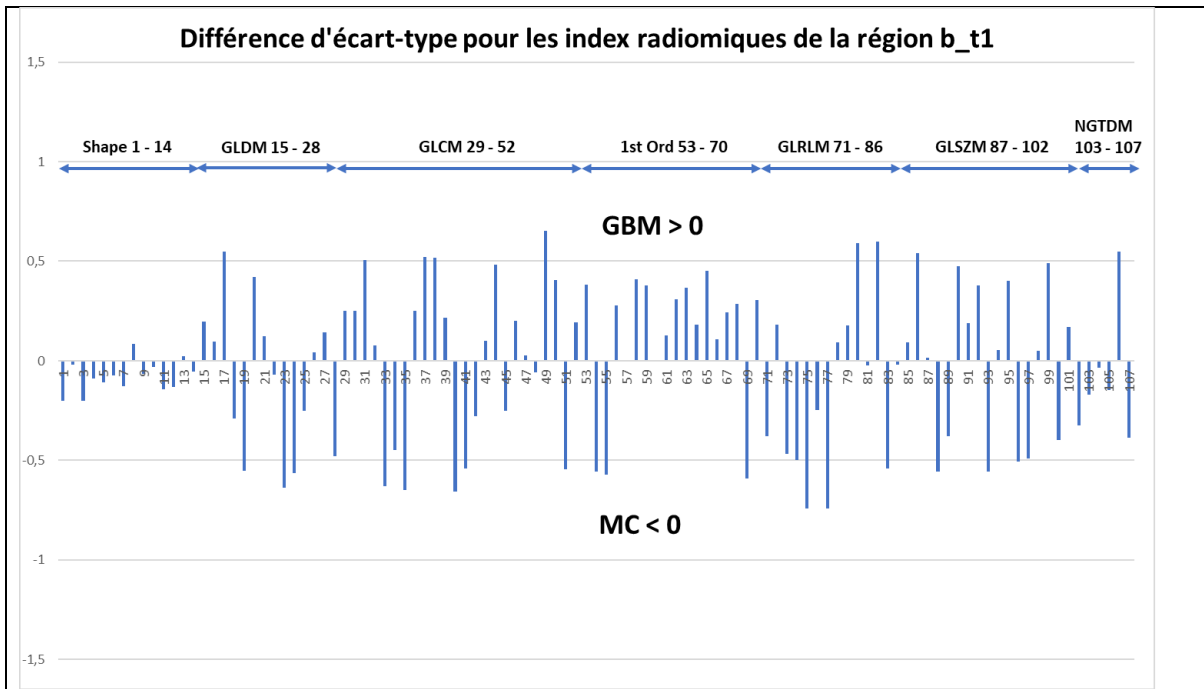
Figure 1. Regroupement hiérarchique des lésions analysées selon les 642 index radiomiques sur l'ensemble de la cohorte.



Après centrage et réduction des 642 index, les différences en écart types des index entre les deux types de lésions étaient représentés dans la figure 2. La moyenne des différences en valeurs absolues était de 0.235 écart-type pour la région (a) T1, 0.231 pour la région (a) flair, 0.121 pour la région (b) T1, 0.121 pour la région (b) flair, 0.216 pour la région (d) T1, et 0.165 pour la région (d) flair.

Figure 2. Représentation des différences d'écart-type entre GBM (positif) et MC (négatif) en fonction des 642 index radiomiques regroupés par type d'index. La correspondance entre le numéro et le nom de l'index est disponible dans le tableau 1 ci-après.





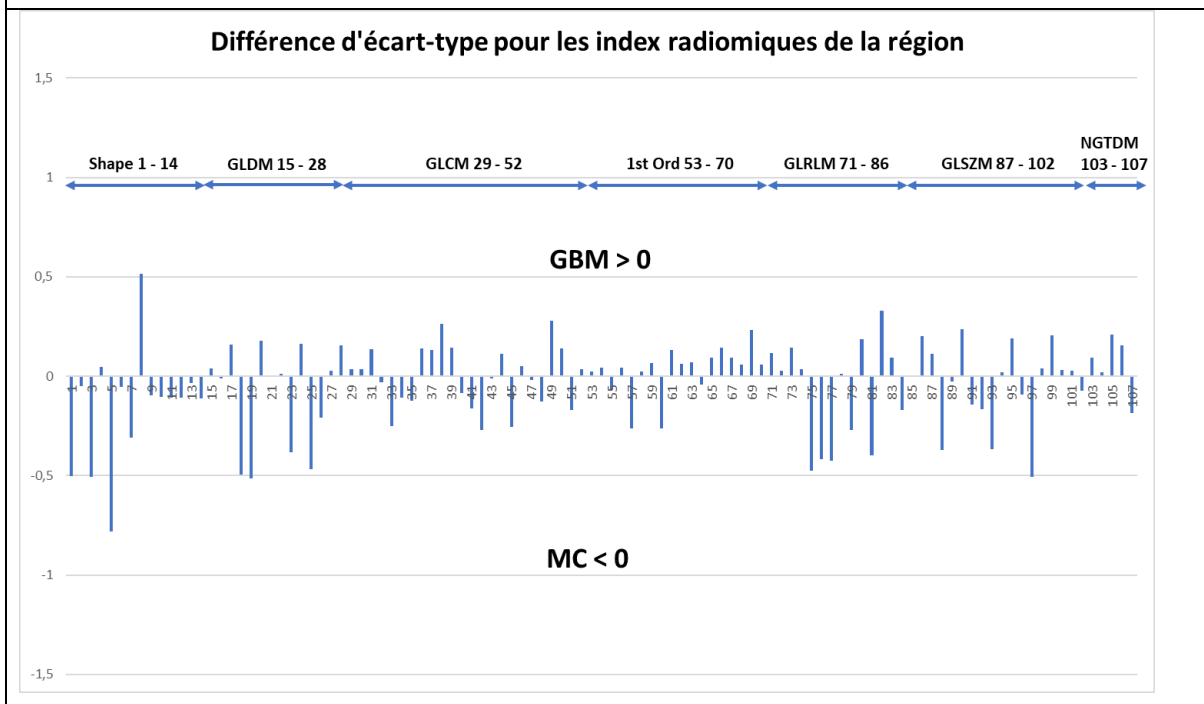
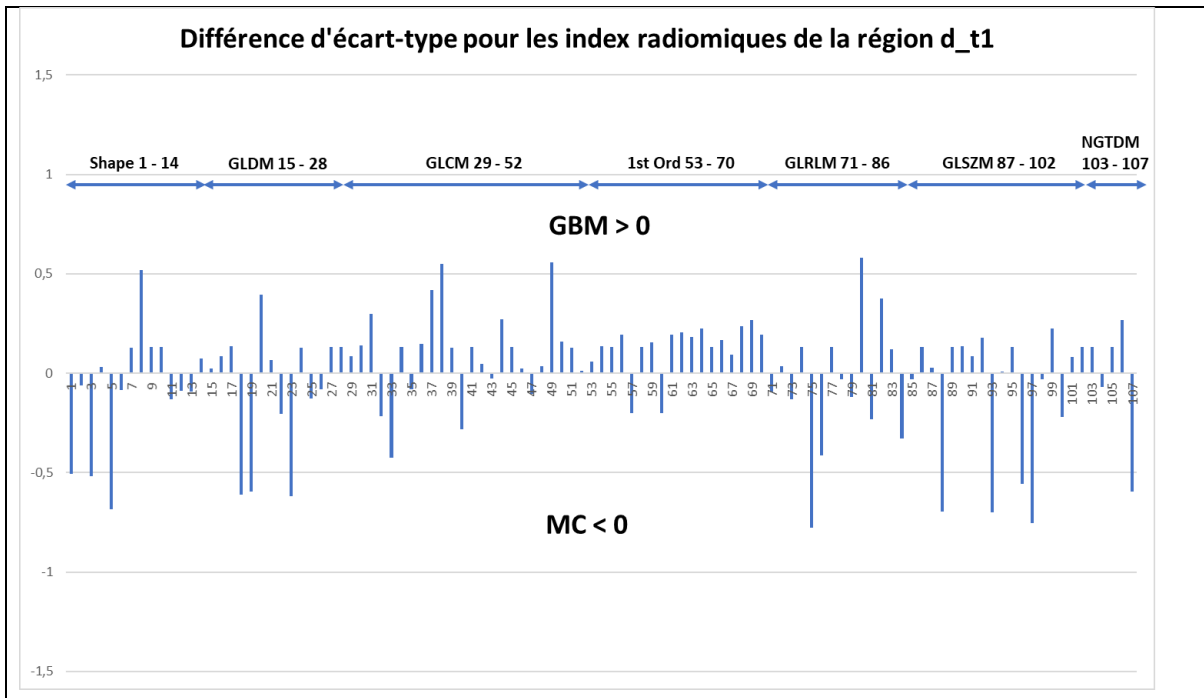


Tableau 1. Liste des index radiomiques extraits par région et par séquence.					
Correspondance entre les numéros d'index et les noms des index.					
Numero	Type	Index	Numero	Type	Index
1	shape	VoxelVolume	55	firstorder	Uniformity
2	shape	Maximum3DDiameter	56	firstorder	Median
3	shape	MeshVolume	57	firstorder	Energy
4	shape	MajorAxisLength	58	firstorder	RobustMeanAbsoluteDeviation
5	shape	Sphericity	59	firstorder	MeanAbsoluteDeviation
6	shape	LeastAxisLength	60	firstorder	TotalEnergy
7	shape	Elongation	61	firstorder	Maximum
8	shape	SurfaceVolumeRatio	62	firstorder	RootMeanSquared
9	shape	Maximum2DDiameterSlice	63	firstorder	90Percentile
10	shape	Flatness	64	firstorder	Minimum
11	shape	SurfaceArea	65	firstorder	Entropy
12	shape	MinorAxisLength	66	firstorder	Range
13	shape	Maximum2DDiameterColumn	67	firstorder	Variance
14	shape	Maximum2DDiameterRow	68	firstorder	10Percentile
15	gldm	GrayLevelVariance	69	firstorder	Kurtosis
16	gldm	HighGrayLevelEmphasis	70	firstorder	Mean
17	gldm	DependenceEntropy	71	glrlm	ShortRunLowGrayLevelEmphasis
18	gldm	DependenceNonUniformity	72	glrlm	GrayLevelVariance
19	gldm	GrayLevelNonUniformity	73	glrlm	LowGrayLevelRunEmphasis
20	gldm	SmallDependenceEmphasis	74	glrlm	GrayLevelNonUniformityNormalized
21	gldm	SmallDependenceHighGrayLevelEmphasis	75	glrlm	RunVariance
22	gldm	DependenceNonUniformityNormalized	76	glrlm	GrayLevelNonUniformity
23	gldm	LargeDependenceEmphasis	77	glrlm	LongRunEmphasis
24	gldm	LargeDependenceLowGrayLevelEmphasis	78	glrlm	ShortRunHighGrayLevelEmphasis
25	gldm	DependenceVariance	79	glrlm	RunLengthNonUniformity
26	gldm	LargeDependenceHighGrayLevelEmphasis	80	glrlm	ShortRunEmphasis
27	gldm	SmallDependenceLowGrayLevelEmphasis	81	glrlm	LongRunHighGrayLevelEmphasis
28	gldm	LowGrayLevelEmphasis	82	glrlm	RunPercentage
29	glcm	JointAverage	83	glrlm	LongRunLowGrayLevelEmphasis
30	glcm	SumAverage	84	glrlm	RunEntropy
31	glcm	JointEntropy	85	glrlm	HighGrayLevelRunEmphasis
32	glcm	ClusterShade	86	glrlm	RunLengthNonUniformityNormalized
33	glcm	MaximumProbability	87	glszm	GrayLevelVariance
34	glcm	Idmn	88	glszm	ZoneVariance
35	glcm	JointEnergy	89	glszm	GrayLevelNonUniformityNormalized
36	glcm	Contrast	90	glszm	SizeZoneNonUniformityNormalized
37	glcm	DifferenceEntropy	91	glszm	SizeZoneNonUniformity
38	glcm	InverseVariance	92	glszm	GrayLevelNonUniformity
39	glcm	DifferenceVariance	93	glszm	LargeAreaEmphasis
40	glcm	Idn	94	glszm	SmallAreaHighGrayLevelEmphasis
41	glcm	Idm	95	glszm	ZonePercentage
42	glcm	Correlation	96	glszm	LargeAreaLowGrayLevelEmphasis

43	glcm	Autocorrelation	97	glszm	LargeAreaHighGrayLevelEmphasis
44	glcm	SumEntropy	98	glszm	HighGrayLevelZoneEmphasis
45	glcm	MCC	99	glszm	SmallAreaEmphasis
46	glcm	SumSquares	100	glszm	LowGrayLevelZoneEmphasis
47	glcm	ClusterProminence	101	glszm	ZoneEntropy
48	glcm	Imc2	102	glszm	SmallAreaLowGrayLevelEmphasis
49	glcm	Imc1	103	ngtdm	Coarseness
50	glcm	DifferenceAverage	104	ngtdm	Complexity
51	glcm	Id	105	ngtdm	Strength
52	glcm	ClusterTendency	106	ngtdm	Contrast
53	firstorder	InterquartileRange	107	ngtdm	Busyness
54	firstorder	Skewness			

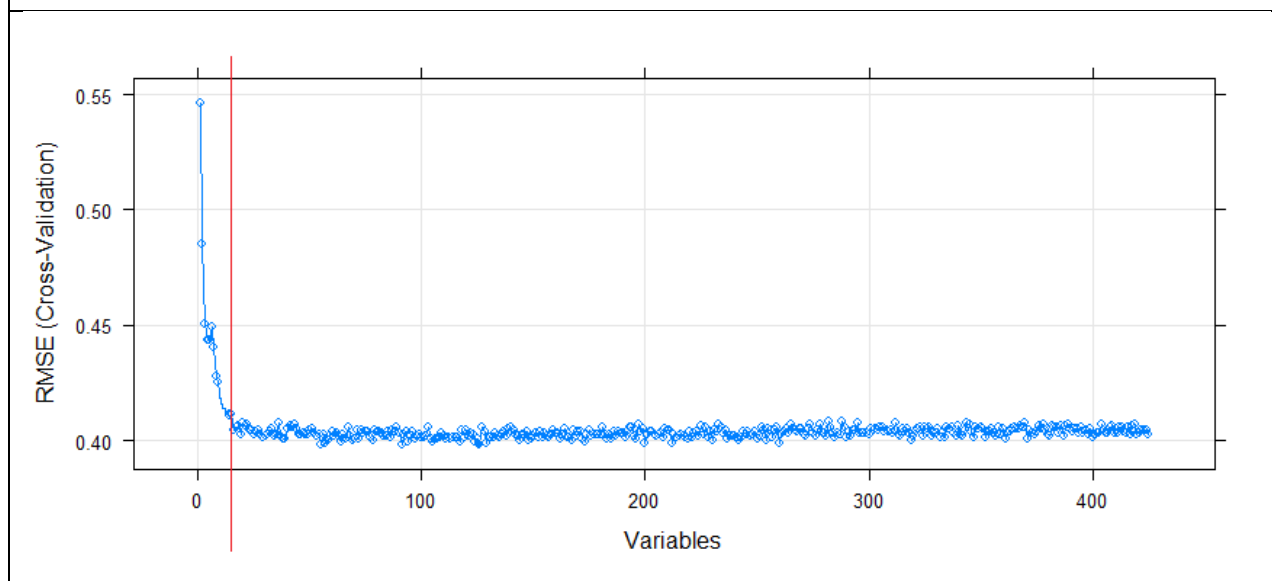
Parmi ces 642 index, 425 étaient conservés par analyse univariée sur l'échantillon d'entraînement ($p < 0,40$). Ces index étaient soumis aux différents algorithmes de sélection dont les résultats sont colligés dans le tableau 2.

Tableau 2. Liste des index radiomiques retenus par chaque algorithme de sélection.

RFE11	Poids relatif	Poids	mRMR15	Poids relatif	Poids
a_flair_Sphericity (Shape)	100	11,41	a_flair_Sphericity (Shape)	100	0,15
a_t1_Sphericity (Shape)	71,95	8,21	d_t1_RunVariance (GLRLM)	56,67	0,085
b_t1_Contrast (NGTDM)	52,15	5,95	a_flair_Skewness (1st.ord)	32,67	0,049
b_t1_Idn (GLCM)	50,66	5,78	d_t1_Imc1 (GLCM)	15,33	0,023
a_flair_Skewness (1st.ord)	43,12	4,92	a_t1_SmallDependenceLowGrayLevelEmphasis (GLDM)	19,33	0,029
b_t1_Kurtosis (1st.ord)	37,51	4,28	d_flair_DependenceNonUniformity (GLDM)	16,67	0,025
b_t1_GrayLevelNonUniformity (GLRLM)	35,67	4,07	d_flair_Sphericity (Shape)	22,67	0,034
b_t1_Imc1 (GLCM)	33,39	3,81	b_t1_SmallAreaEmphasis (GLSZM)	18	0,027
b_t1_LongRunLowGrayLevelEmphasis (GLRLM)	32,95	3,76	b_flair_ClusterShade (GLCM)	16,67	0,025
a_flair_Flatness (Shape)	31,81	3,63	a_flair_LargeDependenceHighGrayLevelEmphasis (GLDM)	10	0,015
b_t1_Median (1st.ord)	31,81	3,63	a_flair_Flatness (Shape)	12	0,018
			d_t1_LargeAreaHighGrayLevelEmphasis (GLSZM)	12,67	0,019
			b_t1_LowGrayLevelRunEmphasis (GLRLM)	10	0,015
			b_t1_Contrast (NGTDM)	8,67	0,013
			b_t1_Skewness (1st.ord)	9,33	0,014
LASSO15	Poids relatif	Poids	SBF10		Rang
a_flair_10Percentile (1st.ord)	100	0,253	a_flair_Coarseness (NGTDM)		1
a_flair_Complexity (NGTDM)	95,65	0,242	a_flair_Elongation (Shape)		2
a_flair_Flatness (Shape)	91,3	0,231	a_flair_Flatness (Shape)		3
a_flair_GrayLevelNonUniformity (GLRLM)	86,96	0,22	a_t1_GrayLevelNonUniformity (GLDM)		4
a_flair_LowGrayLevelZoneEmphasis (GLSZM)	83	0,21	a_flair_GrayLevelNonUniformity (GLRLM)		5
a_flair_Maximum2DDiameterRow (Shape)	79,45	0,201	a_t1_Idmn (GLCM)		6
a_flair_Minimum (1st.ord)	75,89	0,192	a_t1_Idn (GLCM)		7
a_flair_Skewness (1st.ord)	72,33	0,183	a_t1_Kurtosis (1st.ord)		8
a_flair_SmallDependenceLowGrayLevelEmphasis (GLDM)	69,17	0,175	a_flair_LargeAreaHighGrayLevelEmphasis (GLSZM)		9
a_flair_Sphericity (Shape)	66,01	0,167	a_flair_LargeDependenceHighGrayLevelEmphasis (GLDM)		10
a_t1_Maximum2DDiameterSlice (Shape)	62,85	0,159			
a_t1_MCC (GLCM)	60,08	0,152			
a_t1_ShortRunLowGrayLevelEmphasis (GLRLM)	57,31	0,145			
a_t1_SmallDependenceLowGrayLevelEmphasis (GLDM)	54,55	0,138			
b_flair_ClusterShade (GLCM)	52,17	0,132			

L'algorithme Recursive Feature Elimination sélectionnait 11 index radiomiques par validation croisée à 5 blocs, sur un classifieur englobé de type Random Forest, en minimisant l'écart-type des résidus du modèle (RMSE, Root Mean Square Error), en imposant un maximum d'index à 15 (figure 2).

Figure 2. Minimisation de l'écart-type des résidus du modèle par l'algorithme RFE. Un premier minimum est atteint pour 11 index avec RMSE = 0.411.



L'algorithme mRMR permettait d'extraire 15 index en fonction de la matrice d'information mutuelle, le nombre d'index étant choisi par l'opérateur.

L'algorithme LASSO réalisait une régression logistique pénalisée avec validation croisée à 5 blocs. Les paramètres lambda et gamma sont déterminés automatiquement par l'algorithme jusqu'à la convergence permettant d'obtenir 13 index.

La sélection par filtre (SBF) permettait de sélectionner le nombre d'index optimal en les testant préalablement sur un algorithme de type Random Forest avec 500 arbres et 10 variables évaluées à chaque nœud, lors d'une validation croisée à 5 blocs avec le taux d'erreur pour jugement. Cette sélection permettait d'extraire 10 index.

L'algorithme kernel-PCA transformait toutes les variables de la cohorte en 15 composantes principales via une fonction de base radiale, en séparant les mesures d'entraînement

et de test pour éviter une fuite de données. Les composantes principales ainsi calculées ne sont plus explicites.

3.1.2. Importance des index radiomiques pour le meilleur modèle

Les index sont nommés selon le format « segment_séquenceIRM_index (type d'index) ». Par exemple l'index a_flair_Sphericity (Shape) correspond à l'index radiomique « Sphericity » de type Shape, pour la séquence FLAIR, et correspondant à la segmentation de la région (a) c'est-à-dire le rehaussement T1, ici reporté sur la séquence FLAIR.

L'importance des 11 index retenus par l'algorithme RFE pour le classifieur SVM est analysée dans le tableau 3 et représentée dans la figure 3. Les valeurs moyennes de ces index et leurs différences entre GBM et MC sur la cohorte entière sont reportées dans le tableau 4.

Figure 3. Variation d'écart-type des index retenus par RFE entre GBM (positif) et MC (négatif) dans la cohorte entière.

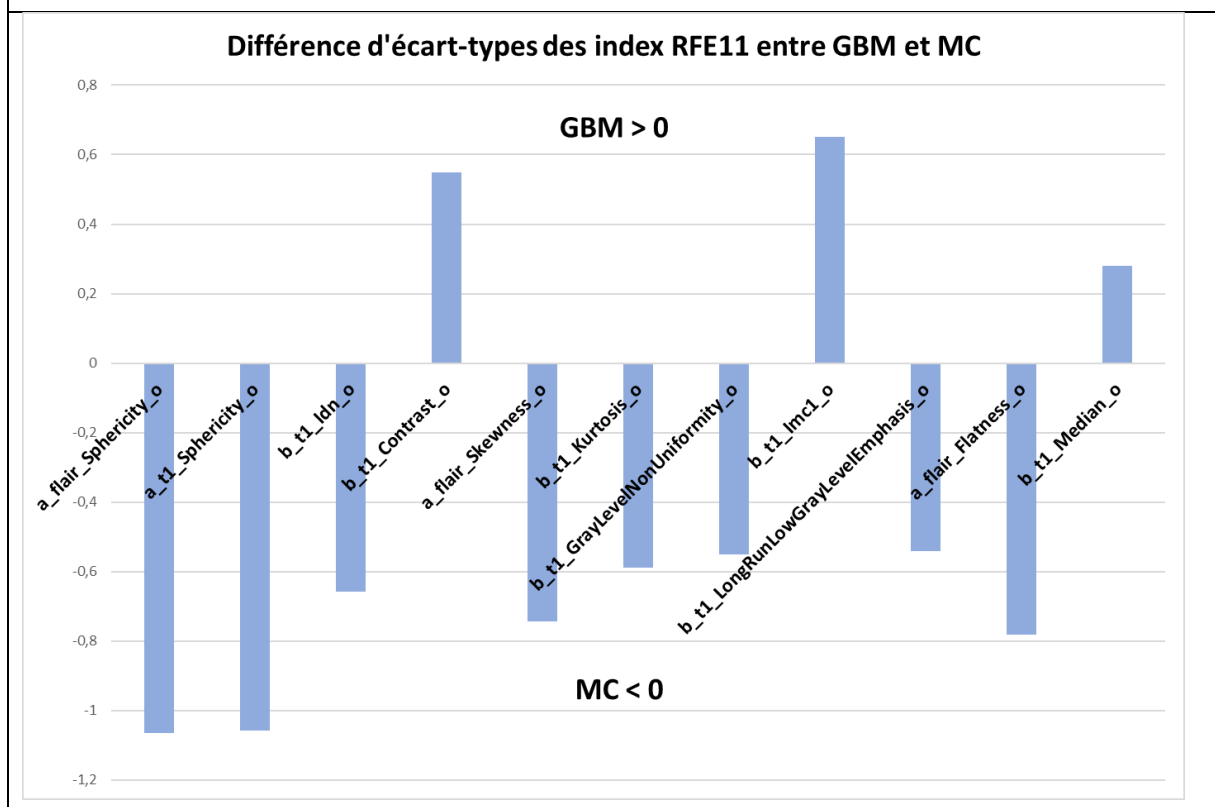


Tableau 3. Importance pour l'AUC d'entraînement des index sélectionnés pour le modèle RFE11_SVM, sur le jeu isolé des index radiomiques. L'AUC de référence est de 0.851. Rang de sélection par l'algorithme RFE. L'AUC1 est calculée lors des retraits itératifs d'un index en partant du modèle avec 11 index radiomiques et par rang RFE décroissant. L'AUC2 est calculée lors du simple retrait de l'index du modèle à 11 index radiomiques. Les données cliniques, IRM et le score CNN sont exclues. Valeurs p du test de Student en analyse univariée bilatérale sur la cohorte entière. Pour 4 index retirés individuellement, l'AUC2 était améliorée de 0.24% à 1.18%. L'AUC sans ces 4 index était améliorée à 0.869 contre 0.851 pour le jeu d'index radiomiques isolé. L'AUC d'entraînement pour le modèle complet restait inchangé.

Index radiomiques	Rang	AUC1 aux retraits itératifs (%)	AUC2 au retrait simple (%)	p
a_flair_Sphericity (Shape)	1	0.851 (n=11 index)	0.847 (-0.47%)	1.10 ⁻⁹
a_t1_Sphericity (Shape)	2	0.847 (-0.47%) (n-1)	0.850 (-0.12%)	1.10 ⁻⁹
b_t1_Contrast (NGTDM)	3	0.796 (-6.46%) (n-2)	0.850 (-0.12%)	0.079
b_t1_Idn (GLCM)	4	0.794 (-6.70%)	0.853 (+0.24%)	5.10 ⁻⁶
a_flair_Skewness (1st.ord)	5	0.792 (-6.93%)	0.818 (-3.88%)	1.10 ⁻⁷
b_t1_Kurtosis (1st.ord)	6	0.762 (-10.46%)	0.848 (-0.35%)	1.10 ⁻⁴
b_t1_GrayLevelNonUniformity (GLRLM)	7	0.742 (-12.81%)	0.851 (0%)	8.10 ⁻⁴
b_t1_Imc1 (GLCM)	8	0.726 (-14.69%)	0.858 (+0.82%)	6.10 ⁻⁶
b_t1_LongRunLowGrayLevel Emphasis (GLRLM)	9	0.692 (-18.68%)	0.861 (+1.18%)	9.10 ⁻⁴
a_flair_Flatness (Shape)	10	0.663 (-22.09%)	0.845 (-0.71%)	4.10 ⁻⁸
b_t1_Median (1st.ord)	11	0.617 (-27.50%)	0.854 (+0.35%)	0.056

Tableau 4. Valeurs moyennes des index radiomiques sélectionnés par RFE pour les GBM et MC. Différences en écart-type entre les groupes GBM et MC, dans le sens GBM moins MC.

Index radiomiques	Moyenne GBM	Z-score GBM	Moyenne MC	Z-score MC	Différence d'écart-type
a_flair_Sphericity (Shape)	0.732 ± 0.085	-0.474σ	0.827 ± 0.064	+0.586σ	-1.06σ
a_t1_Sphericity (Shape)	0.734 ± 0.085	-0.474σ	0.829 ± 0.064	+0.583σ	-1.06σ
b_t1_Contrast (NGTDM)	11.35 ± 20.02	+0.113σ	6.83 ± 14.73	-0.139σ	+0.25σ
b_t1_Idn (GLCM)	0.962 ± 0.012	-0.295σ	0.970 ± 0.012	+0.362σ	-0.66σ
a_flair_Skewness (1st.ord)	-0.755 ± 0.964	-0.333σ	-0.054 ± 0.762	+0.410σ	-0.74σ
b_t1_Kurtosis (1st.ord)	8.72 ± 3.12	-0.265σ	12.07 ± 5.36	+0.325σ	-0.59σ
b_t1_GrayLevelNonUniformity (GLRLM)	9231,57 ± 7247.86	-0.247σ	15671,48 ± 14792.77	+0.304σ	-0.50σ
b_t1_Imc1 (GLCM)	-0.323 ± 0.081	+0.956σ	-0.378 ± 0.080	-0.360σ	+0.65σ
b_t1_LongRunLowGrayLevel Emphasis (GLRLM)	0.210 ± 0.289	-0.242σ	0.830 ± 1.63	+0.298σ	-0.54σ
a_flair_Flatness (Shape)	0.616 ± 0.118	-0.350σ	0.714 ± 0.112	+0.430σ	-0.78σ
b_t1_Median (1st.ord)	352.52 ± 271.16	+0.125σ	279,69 ± 242.90	-0.154σ	+0.28σ

Les deux premiers index retenus par ordre d'importance RFE étaient a_flair_Sphericity (Shape) et a_t1_Sphericity (Shape), tous deux explorant la sphéricité de la lésion rehaussée en T1 et en FLAIR. La sphéricité est le ratio entre le périmètre de la lésion et le périmètre d'un cercle de même surface, variant de 0 à 1, la valeur 1 indiquant un cercle parfait. Cet index variait de manière identique sur les deux séquences, avec une sphéricité plus importante de +1.06 écart-type pour les MC (tableau 4). Le retrait de ces deux index entraînait une baisse de 6,46% de l'AUC du modèle employant uniquement le jeu d'index radiomique.

Un troisième index morphologique arrivait au 10^e rang du RFE : a_flair_Flatness (Shape). Cet index correspond au rapport entre le petit axe et le grand axe du volume, avec une valeur de 1 évoquant un aspect sphérique. Il était augmenté pour les MC de +0.78 écart-type. Le retrait de l'ensemble des index morphologiques du modèle hybride complet était responsable d'une baisse d'AUC de 0.942 à 0.903, soit 4,14% de baisse relative.

Parmi les 8 autres index retenus par RFE, il était compté 3 index de 1^e ordre, et 5 index de texture. Tous exploraient la région (a) rehaussée en T1 ou la région (b) associant le rehaussement T1 et l'anomalie FLAIR péri-rehaussement. Ces 8 index faisaient état de différences concernant la distribution des intensités et la complexité des textures entre les lésions GBM et MC avec, par rang décroissant de sélection :

- L'index `b_t1_Contrast` (NGTDM) fait état du changement local d'intensité en fonction de l'intervalle des intensités. Ainsi le contraste est maximal lors de forte variation sur une courte distance et lorsque les lésions ont une plage d'intensité très grandes. L'index était majoré pour les GBM de +0.25 écart-type.
- L'index `b_t1_Idn` (GLCM) (Inverse Difference normalized) évalue la différence de valeurs d'intensité locale, inversée, et normalisée par le nombre de niveaux de gris. L'index varie dans le sens de l'homogénéité des intensités et du faible nombre de niveaux d'intensités. Il était plus élevé pour les MC avec +0.66 écart-type.
- L'index `a_flair_Skewness` (1st.order) mesure l'asymétrie de la distribution des intensités autour de la moyenne, avec une Skewness négative lorsque les intensités hautes sont plus nombreuses, et inversement pour une Skewness positive. L'index était majoré de +0.74 écart-type pour les MC.
- L'index `b_t1_Kurtosis` (1st.order) est la mesure d'un aspect en pic de la distribution. Ici, une valeur augmentée est en faveur d'une distribution regroupée en pic plus à distance de la moyenne, et une valeur basse pour une densité de la distribution plus aplatie et proche de la moyenne. Kurtosis était augmenté pour les MC de +0.59 écart-type.
- L'index `b_t1_GrayLevelNonUniformity` (GLRLM) mesure la similarité des niveaux de gris des différentes séries de voxels voisins de même niveau de gris. Une faible valeur est en faveur de faibles variations de niveaux de gris. Il était plus important pour les MC de +0.50 écart-type.
- L'index `b_t1_Imc1` (GLCM) (Informational Measure of Correlation) est une mesure de la complexité de l'information de la matrice GLCM en évaluant une matrice de corrélation entre les niveaux d'intensités et le nombre de voxels correspondant à chaque niveau. L'index augmente alors avec l'hétérogénéités des niveaux de gris et de leurs occurrences. Il était plus élevé de +0.65 écart-type pour les GBM.
- L'index `b_t1_LongRunLowGrayLevelEmphasis` (GLRLM) met en avant les séries longues de voxels d'intensités similaires, avec une emphase sur les bas niveaux d'intensités. Il était augmenté pour les MC de +0.54 écart-type.
- L'index `b_t1_Median` (1st.order) correspond à la valeur d'intensité médiane de la distribution. La médiane d'intensité était plus élevée pour les GBM de +0.28 écart-type.

Le retrait de ces 8 index de texture et de premier ordre entraînait une baisse de l'AUC d'entraînement de 0.942 à 0.892, soit -5,31%.

Le commentaire et l'illustration de ces résultats sont réalisés dans le chapitre Discussion.

3.2. Stabilité des index radiomiques et des segmentations

La stabilité de 186 index radiomiques était testée sur deux segmentations d'environ 1cm³ homo et controlatérales à la lésion dans une région de substance blanche d'apparence saine, par défaut frontale supérieure. 93 index étaient évalués en séquence T1 et 93 en séquence T2 FLAIR, après l'exclusion des index morphologiques. Le coefficient de corrélation de concordance (CCC) était en moyenne de $0,76 \pm 0,20$ pour les index T1 et de $0,69 \pm 0,20$ pour les index FLAIR. 51/93 (54,8%) index T1 et 31/93 (33,3%) index FLAIR avaient un CCC supérieur à 0,80.

La stabilité des index dans les régions tumorales était également testée après nouvelle segmentation par le même opérateur pour 30 lésions. Le CCC était en moyenne de $0,92 \pm 0,13$, toute séquence et régions confondues. En séquence T1 injectée, il était de $0,94 \pm 0,09$, et de $0,90 \pm 0,15$ en séquence FLAIR.

Tableau 5. Stabilité des index radiomiques retenus par RFE11.

Index RFE11	Poids relatif	Poids	Stabilité en région saine (CCC)	Stabilité en région lésionnelle (CCC)	Shapiro test p
a_flair_Sphericity (Shape)	100	11,41	-	0,93 [0,86 ; 0,97]	3.56e-06
a_t1_Sphericity (Shape)	71,95	8,21	-	0,93 [0,86 ; 0,97]	2.87e-06
b_t1_Contrast (NGTDM)	52,15	5,95	0,91	0,98 [0,97 ; 0,99]	4.96e-22
b_t1_Idn (GLCM)	50,66	5,78	0,63	0,97 [0,94 ; 0,99]	8.03e-06
a_flair_Skewness (1st.ord)	43,12	4,92	0,46	0,97 [0,94 ; 0,99]	1.67e-09
b_t1_Kurtosis (1st.ord)	37,51	4,28	0,39	0,98 [0,96 ; 0,99]	4.85e-15
b_t1_GrayLevelNonUniformity (GLRLM)	35,67	4,07	0,78	0,98 [0,95 ; 0,99]	3.37e-14
b_t1_Imc1 (GLCM)	33,39	3,81	0,61	0,96 [0,92 ; 0,98]	0.001
b_t1_LongRunLowGrayLevel Emphasis (GLRLM)	32,95	3,76	0,71	0,99 [0,99 ; 1]	1.63e-24
a_flair_Flatness (Shape)	31,81	3,63	-	0,90 [0,81 ; 0,95]	0.002
b_t1_Median (1st.ord)	31,81	3,63	0,99	0,99 [0,94 ; 1]	1.54e-16

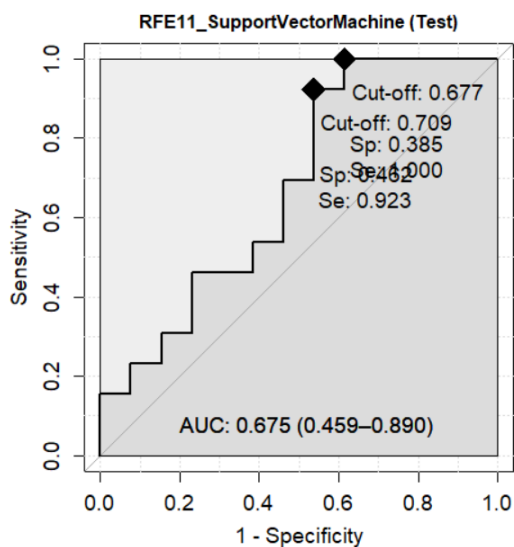
La répétition de la segmentation pour 30 lésions à l'aveugle retrouvait un coefficient de similarité de Dice (DSC) en moyenne à $0,92 \pm 0,06$. Le DSC était de $0,94 \pm 0,05$ pour la segmentation de la région (a) T1 injectée et de $0,90 \pm 0,07$ pour la région (b) FLAIR.

3.3. Classification des métastases cérébrales selon le primitif

Après exclusion des sujets ayant des données manquantes sur l'origine de la lésion primitive, le meilleur modèle RFE11_SVM était entraîné sur la classification binaire des MC d'origine pulmonaire versus les MC d'autre origine, en gardant les mêmes variables précédemment employées pour l'objectif principal. L'échantillon d'entraînement était de 57 lésions MC avec 41 mesures (71,9%) correspondant à une origine pulmonaire et 16 mesures d'autres origines. L'échantillon de test était de 26 lésions MC avec 13 lésions (50%) d'origine pulmonaire et 13 lésions non pulmonaires.

RFE11_SVM (MGMT)

Echantillon Test

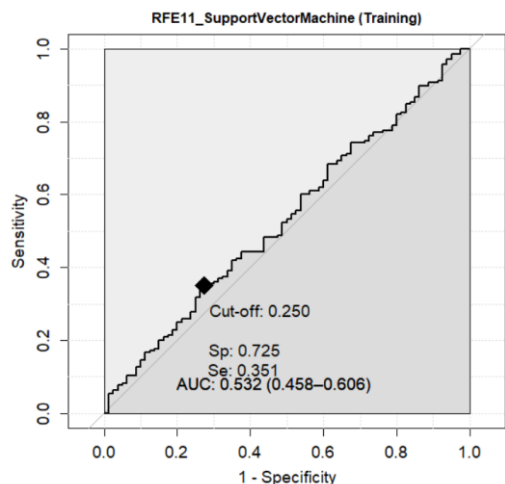


Matrice de confusion :

Prediction	Classe réelle	
	Autre	Pulm
Autre	0	0
Pulm	13	11

AUC = 0.598 [0.362 ; 0.834]
 Exactitude = 0.5 [0.308; 0.692]
 Sensibilité = 1
 (Pulm est la classe positive)
 Spécificité = 0.154 [0.015 ; 0.293]
 VPP = 0.5 [0.308 ; 0.692]
 VPN = -
 χ^2 p-valeur = -
 McNemar p = 0.0009
 F1-Score = 0.667 [0.485 ; 0.848]
 MCC = 0
 Youden J = 0

Echantillon Entraînement

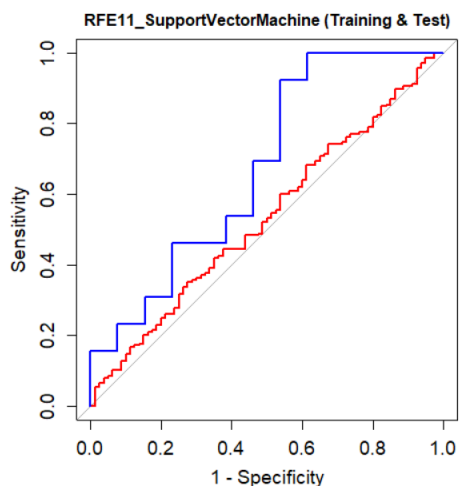


AUC = 0.542 [0.468 ; 0.617]
 Exactitude = 0.717 [0.60 ; 0.834]

Hyperparamètres :

Cost C = 0.5
 Sigma = 0.0194
 Kernel : Gaussian Radial Basis
 Number of support vectors : 48

Comparaison ROC test et entraînement



Bootstrap test :

D = -1.2214, boot.n = 2000, boot.stratified = 1, p-value = 0.2219

3.4. Classification des glioblastomes selon le statut MGMT

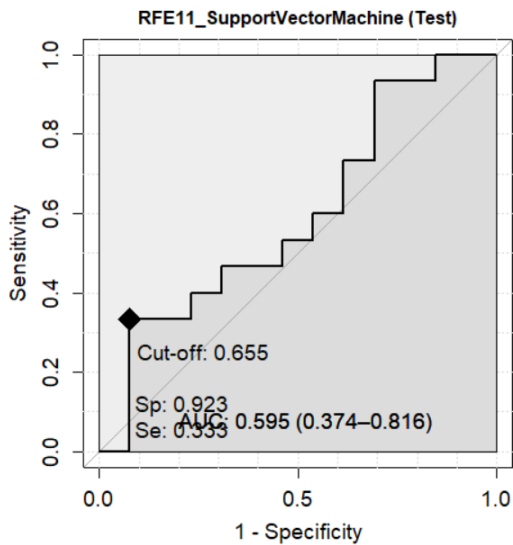
Après exclusion des sujets ayant des données manquantes sur le statut MGMT, le meilleur modèle RFE11_SVM était entraîné sur la classification binaire méthylation / absence de méthylation MGMT en gardant les mêmes variables précédemment employées pour l'objectif principal.

L'échantillon d'entraînement était de 65 lésions GBM avec 34 mesures (52,3%) correspondant à un statut MGMT méthylé et 31 mesures à un statut non méthylé.

L'échantillon de test était de 28 lésions GBM avec 13 lésions (46,4%) MGMT méthylées et 15 lésions non méthylées.

RFE11_SVM (MGMT)

Echantillon Test

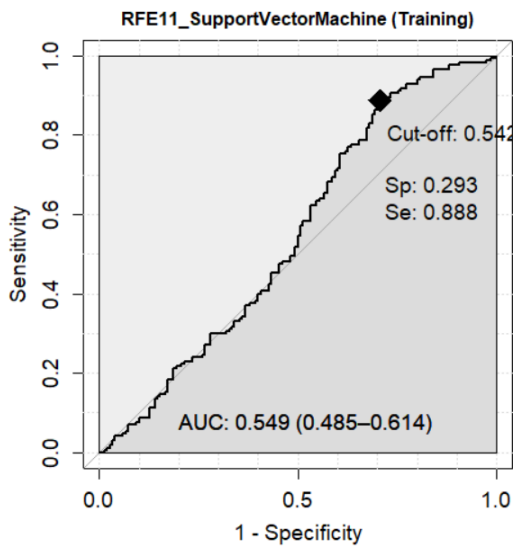


Matrice de confusion :

Prediction	Classe réelle	
	NonMet	Met
NonMet	6	7
Met	7	8

AUC = 0.595 [0.374 ; 0.816]
 Exactitude = 0.50 [0.315 ; 0.685]
 Sensibilité = 0.533 [0.349 ; 0.718]
 (Met est la classe positive)
 Spécificité = 0.462 [0.277 ; 0.646]
 VPP = 0.533 [0.349 ; 0.718]
 VPN = 0.462 [0.277 ; 0.646]
 χ^2 p-valeur = 1
 McNemar p = 1
 F1-Score = 0.533 [0.349 ; 0.718]
 MCC = 0 [0.0 ; 0.101]
 Youden J = 0 [0 ; 0.364]

Echantillon Entraînement

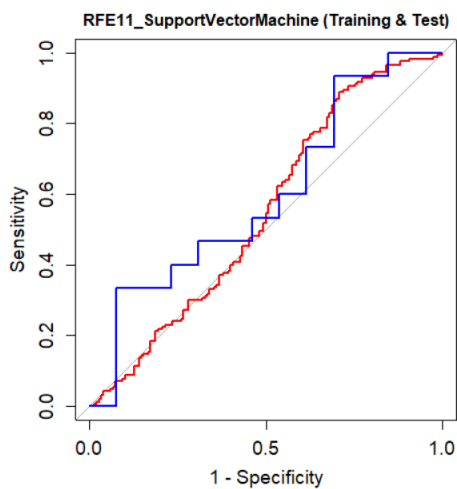


AUC = 0.549 [0.485 ; 0.614]
 Exactitude = 0.584 [0.463 ; 0.705]

Hyperparamètres :

Cost C = 0.5
 Sigma = 0.0230
 Kernel : Gaussian Radial Basis
 Number of support vectors : 63

Comparaison ROC test et entraînement



Bootstrap test :

$D = -0.40001$, boot.n = 2000, boot.stratified = 1, p-value = 0.6891

IV. DISCUSSION

1. Résultats

Ce travail permet de montrer les bonnes performances de modèles hybrides basés sur des algorithmes de classification conventionnels et augmentés par un CNN pour la tâche de classification binaire entre GBM et MC, en exploitant notamment des données radiomiques. Il n’y avait pas de différence significative en âge, sexe ou nombre de lésions entre les groupes GBM et MC ou les échantillons d’entraînement ou de test.

En particulier, le meilleur modèle RFE11_SVM proposait un AUC_{test} de 0.928 [0.853 ; 1], confirmant les résultats de précédentes études ayant le même objectif. Ainsi, Bae et al. rapportent un modèle associant RFE et un SVM linéaire avec un AUC_{test} à 0.886 [0.798 ; 0.927] en combinant les index radiomiques de deux régions : rehaussement T1 et Flair péri-rehaussement sur une imagerie 3T chez 248 sujets (130). Cependant, dans leur étude, le meilleur modèle était un réseau de neurones artificiels avec un AUC à 0,96. Han et al. font état de plusieurs modèles SVM performants avec un AUC_{test} variant de 0.893 à 0.922, en segmentant uniquement la région rehaussée en T1, sur une imagerie exclusivement à 3T chez 471 patients (60). Dans leur étude, un modèle Random Forest semblait toutefois plus discriminant avec un AUC_{test} à 0,94. Qian et al. faisaient état d’un AUC_{test} variant entre 0.79 et 0.90 pour différents modèles SVM, avec une performance maximale pour l’association LASSO et SVM, en séquences T1, T1 injectée et T2, avec segmentation du rehaussement tumoral en imagerie 3T chez 412 sujets (59). Artzi et al. présentaient les performances d’un SVM avec un AUC à 0.96 après réduction des index par PCA, sur une imagerie 1.5 et 3T en séquence T1 injectée, en segmentant la région rehaussée, chez 439 patients (58).

L’originalité de ce travail réside dans l’augmentation des classifieurs conventionnels par un CNN, qui calculait une probabilité de classe par analyse directe de l’imagerie 2D. Testé uniquement sur les données du CNN ou sur les données radiomiques, le classifieur RFE11_SVM montrait en effet des performances acceptables à bonnes, avec des AUC_{test} respectivement de 0,782 IC_{95%}[0,782 ; 0,910] et de 0,822 IC_{95%}[0,715 ; 0,929], tandis que ces performances devenaient très bonnes à excellentes lorsque toutes les informations étaient employées de concert. De plus, ces résultats étaient obtenus avec un recrutement inférieur de 164 sujets et 185 mesures. L’emploi d’une méthode de sélection *wrapper* telle que RFE englobant une fonction Random Forest, suivant un filtre statistique avec test t de Student et $p < 0,40$, associés à un classifieur SVM non linéaire dont le kernel est une fonction de base radiale, a permis d’atteindre

ce niveau de performance, en plus du caractère hybride du modèle. L'imagerie était issue d'un grand nombre de centres, de machines, de modèles et de constructeurs différents, à 3T et 1.5T, permettant de réduire le risque de surajustement à l'entraînement des modèles. En effet, la grande majorité des modèles ne présentaient pas de différence significative de courbes ROC à l'entraînement ou lors du test. Les images exploitées provenaient des séquences de routine clinique, sans nécessiter le recours aux données brutes d'acquisition. Tous les sujets bénéficiaient d'une biopsie permettant de confronter les modèles au standard de référence. Enfin, les données cliniques, et descriptives d'imagerie ne semblaient pas avoir un poids relatif significatif pour les modèles hybrides, comparativement aux données radiomiques ou du CNN qui leur étaient préférées à l'entraînement.

L'exactitude du modèle RFE11_SVM était de 91,4% avec 5 erreurs de prédiction sur 58 lésions test. Une prédiction MC s'avérait être un GBM (figure 1), cette prédiction étant vraisemblablement influencée par un index de sphéricité très augmenté, cet index ayant un poids important et redondant dans le modèle RFE11_SVM, puisqu'il représente les deux premiers index par rang RFE. Quatre prédictions GBM étaient réellement des MC (figure 2), avec des index d'hétérogénéité majorés et une baisse de la sphéricité pour 3 lésions. La dernière lésion présentait quelques artefacts de mouvements, n'entravant toutefois pas l'interprétation de routine, étant de ce fait conservée dans l'analyse. Quatre de ces 5 erreurs de classification correspondaient aux 4 lésions présentant une importante difficulté de classification, puisqu'elles n'étaient correctement prédites que par moins de 25% des modèles.

Figure 1. : Erreur de classification dans l'échantillon test : lésion prédite MC, avec classe réelle GBM. Séquence T1 injectée à gauche, T2 FLAIR à droite. Difficulté de diagnostic : 16% des classifieurs prédisaient correctement cette lésion.

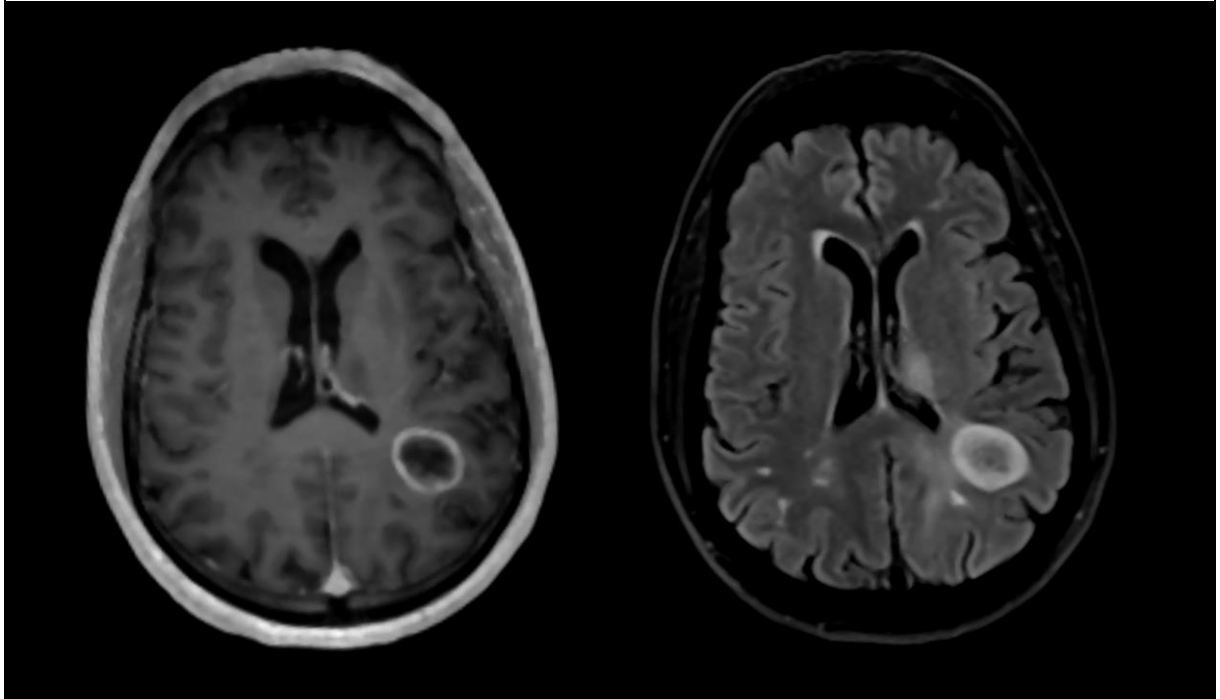
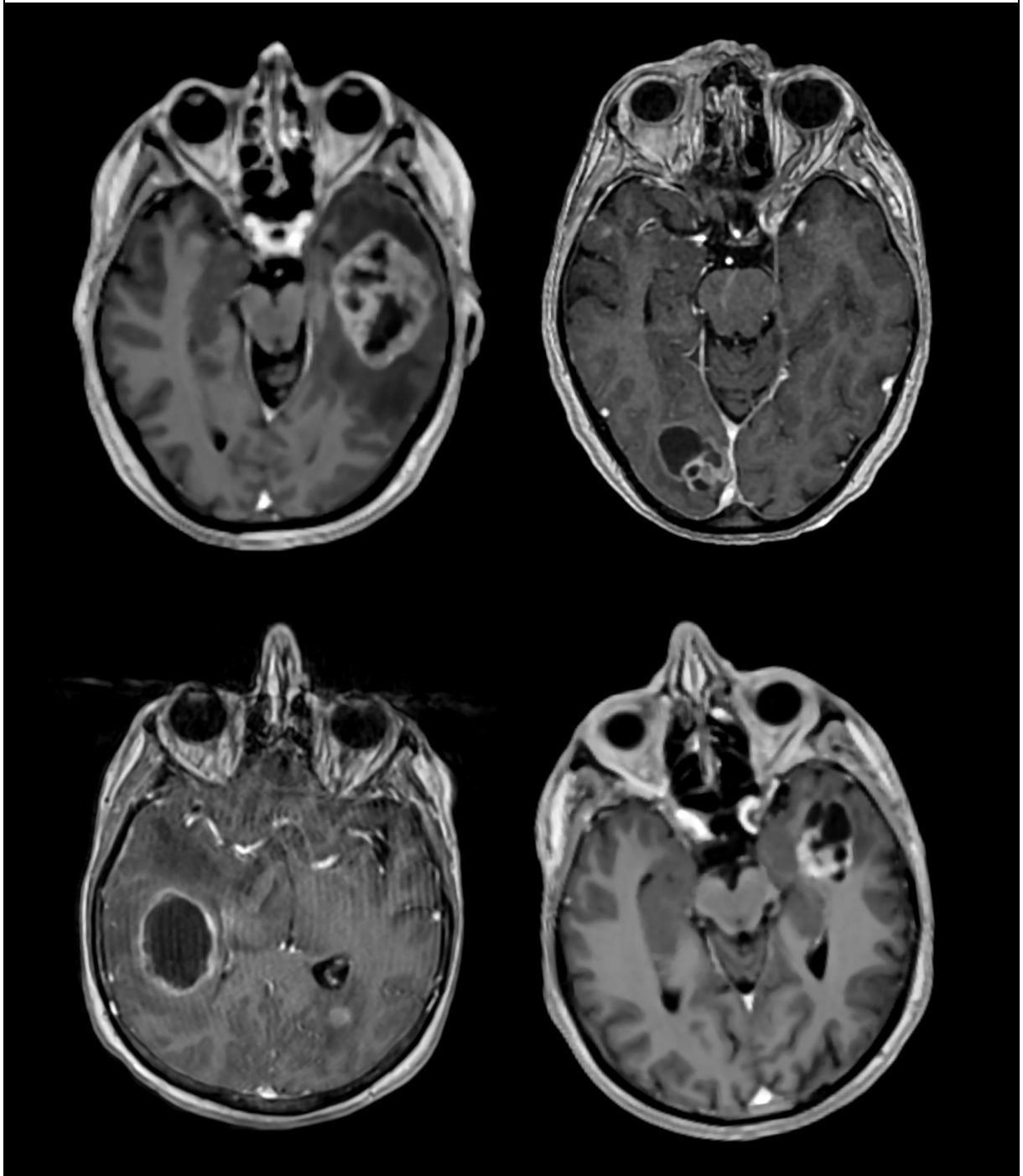


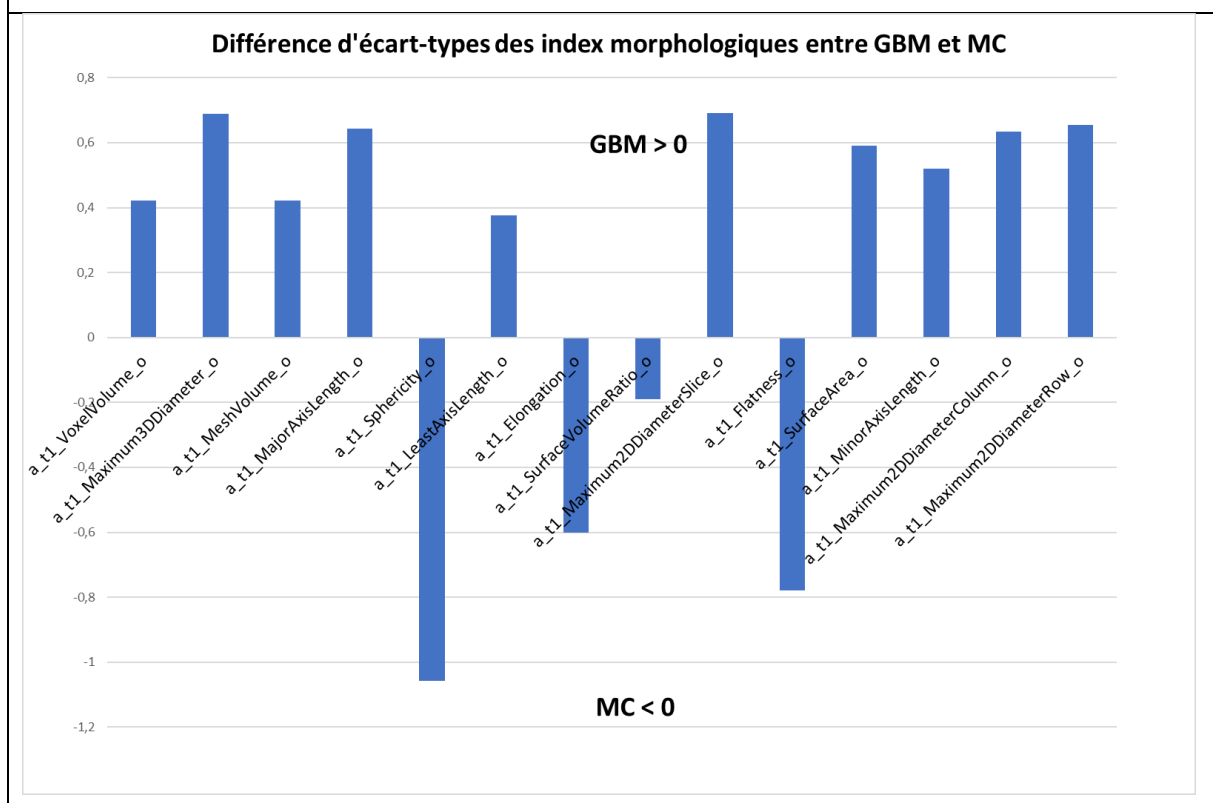
Figure 2. Quatre prédictions erronées, avec le classifieur RFE11_SVM classant ces lésions GBM, la classe réelle étant MC. Séquence T1 injectée. Respectivement, l'exactitude de classement était de 3%, 16%, 3% et 74% des classificateurs.



2. Index radiomiques pour le modèle RFE11_SVM

Pour le meilleur modèle RFE11_SVM, 3 index morphologiques étaient retenus en 1^e, 2^e et 10^e rang d'importance. Ils faisaient état de la sphéricité des lésions de la région (a) plus marquée de +1.06 écart-type pour les MC (figures 3 et 4). Précédemment, cet aspect a été rapporté dans plusieurs études ayant pour objectif de différencier MC et GBM. Yang et al. rapportent une exactitude de classification pour cette question de 95,8% sur 48 lésions en générant des modèles 3D représentatifs de chaque type de lésion, avec 2 index morphométriques 3D de sphéricité en tenseur de diffusion (67). Yang et al. montraient également les bonnes performances de classification d'un réseau de neurones sur 48 lésions en se basant uniquement sur 61 index morphologiques 2D, avec un AUC de 0.975 et une exactitude de 97,9%, en utilisant la décomposition du tenseur de diffusion p:q (131).

Figure 3. Variation d'écart-types des index morphologiques entre GBM (positif) et MC (négatif) dans la cohorte entière.



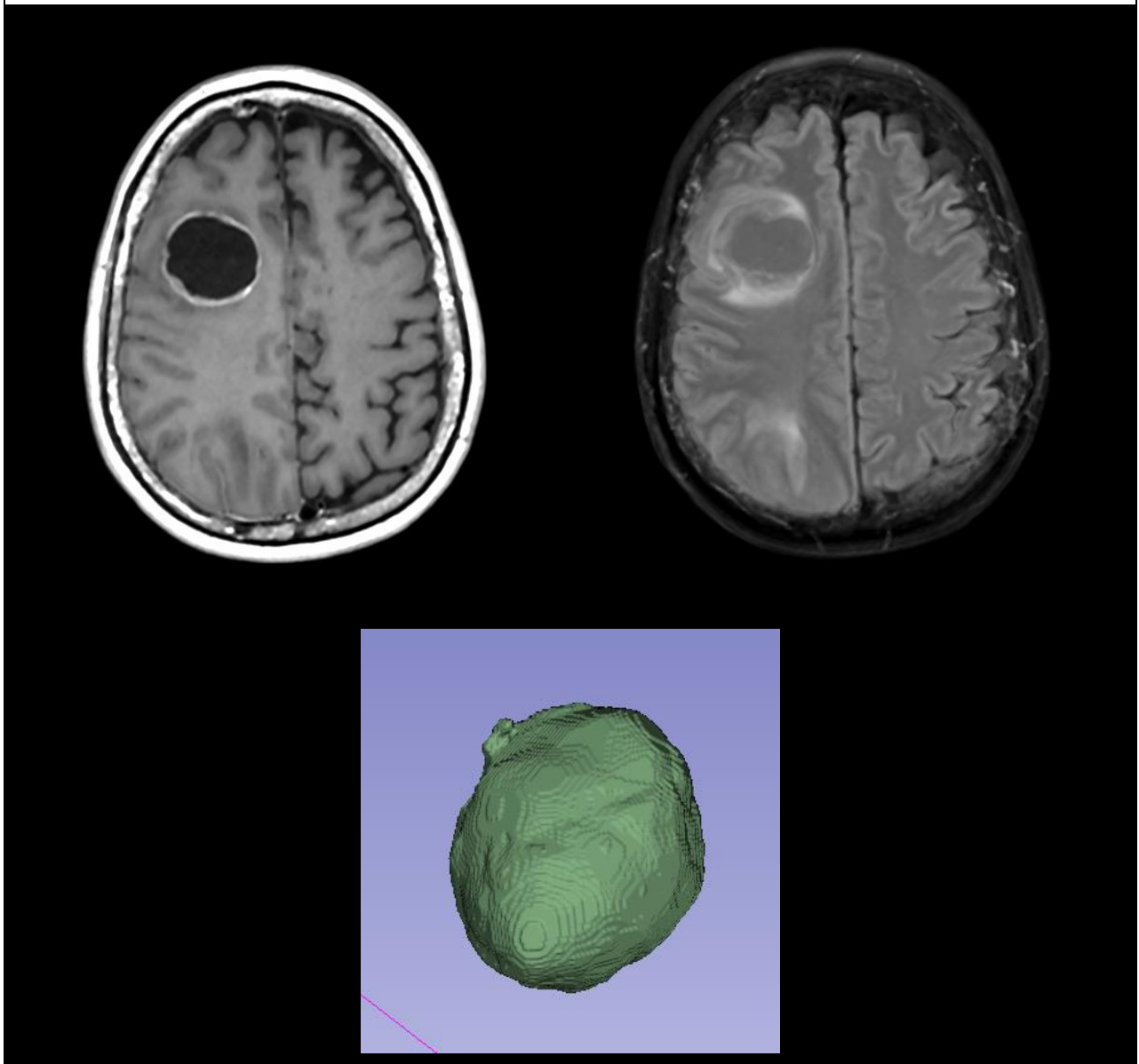
Pour l'étude d'Artzi et al., l'index morphologique du périmètre lésionnel était retenu parmi les 10 plus importants index de leur classifieur SVM (AUC=0.96) (58). Le périmètre tumoral est directement relié à l'index de sphéricité en le comparant au périmètre d'un cercle de même surface. Bae et al. ne relèvent pas d'index de sphéricité de la tumeur elle-même dans les 10 premiers index, mais retient le grand axe et l'index Flatness de l'hypersignal Flair péri-rehaussement (130), ceci pouvant suggérer un aspect plus arrondi et étendu de l'œdème vasogénique par rapport à l'infiltrat glial.

L'aspect plus arrondi, les contours réguliers et bien limités des métastases cérébrales par rapport aux glioblastomes sont des aspects macroscopiques bien connus permettant d'orienter le diagnostic, ce d'autant plus lorsque les lésions sont multiples et qu'un antécédent de lésion primitive est connu (49,132). Cependant, dans ce travail, tous les patients MC de la cohorte bénéficiaient d'une biopsie stéréotaxique ou chirurgicale, témoignant d'une certaine incertitude diagnostique et parfois d'un bilan systémique non productif. Ainsi, il semble intéressant de noter que l'index de sphéricité pourrait être différent entre GBM et MC, lorsque cet aspect n'est pas macroscopiquement évident.

A l'inverse, les glioblastomes présentent une hétérogénéité lésionnelle due à une croissance plus anarchique, suivant notamment les faisceaux de la substance blanche profonde, les structures méningées via un envahissement pial, ou encore le liquide céphalo-rachidien (133). Les différents sous types cellulaires, en fonction des mutations, entre autres d'EGFR, et du temps de doublement, seraient également responsables d'une prolifération inhomogène (134).

Figure 4. Exemple caricatural d'une MC présentant une sphéricité élevée de 0,89 (+1.0 σ).

Séquence T1 injectée à gauche, T2 FLAIR à droite, rendu 3D brut en bas.



Dans la région (b) en séquence T1, l'analyse des index de premier ordre retrouve une valeur d'intensité médiane plus importante pour les GBM, évoquant un potentiel rehaussement de deuxième passage plus intense pour ces lésions. Cet aspect était associé à un Kurtosis élevé pour les GBM mais moins augmenté que pour les MC (figure 7). Un Kurtosis élevé étant en faveur d'un regroupement des occurrences d'intensités autour d'une même valeur, formant un pic, cela pourrait se traduire par un potentiel rehaussement tumoral plus intense et hétérogène des GBM, et un rehaussement moins intense et plus homogène pour les MC (figures 3 et 4). Cela est à fortement pondérer par le fait que la région péri-rehaussement (d) comprise dans la région (b)

des GBM était moins volumineuse que celles des MC (MeshVolume à $43,1\text{cm}^3$ vs $67,2\text{cm}^3$, $p=6.10^{-4}$) engendrant un effet de volume sur l'histogramme. En effet, la valeur d'intensité médiane de la région (a) T1 était inférieure pour les GBM à -7.59 contre -4.26 pour les MC, à la limite de la significativité ($p=0.052$). Toutefois, cet effet de volume peut être considéré comme une information complémentaire comprise dans le même index Kurtosis.

Dans la région (a) en séquence Flair, l'index Skewness était presque nulle pour les MC, évoquant une symétrie de distribution de l'histogramme autour de la valeur moyenne, illustrant le caractère plus homogène du signal FLAIR des MC (figure 5 et 6). Concernant les GBM, la Skewness était nettement négative, cet aspect évoquant un décalage vers la droite de la distribution de l'histogramme avec une queue de distribution gauche plus importante. La valeur moyenne d'intensité Flair dans la région (a) était plus importante pour les GBM ($+0.12\sigma$) par rapport aux MC (-0.15σ), à la limite de la significativité ($p=0.066$) suggérant qu'une approche quantitative du signal FLAIR de ces lésions pourrait permettre de les différencier.

Figure 4. Exemple d'une MC présentant un Kurtosis nettement augmenté à $14,6 (+0,1\sigma)$.

Skewness faiblement négative à $-1,4 (+0,6\sigma)$ en région (b) T1. Skewness a_flair à $-0,20$.

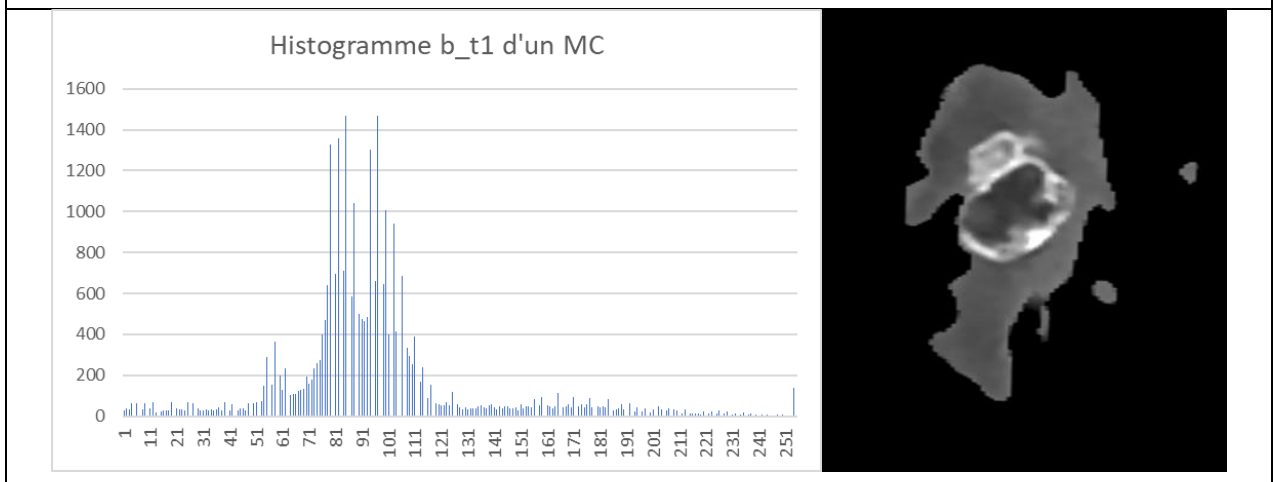


Figure 4. Exemple d'un GBM en b_t1 présentant un Kurtosis faiblement augmenté à 8,2 (-0,03 σ). Skewness faiblement augmenté (+1,8 et +0,7 σ) en région (b) T1. Skewness en région a_flair à -1,74.

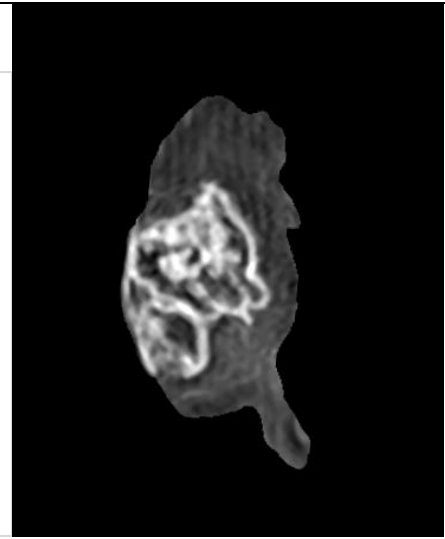
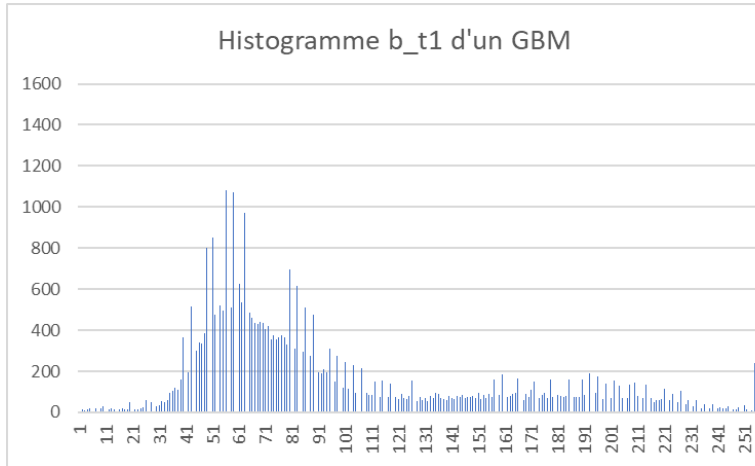


Figure 5. Exemple d'une MC présentant une faible Skewness de -0.15 (+0.20 σ) dans la région (a) FLAIR. Kurtosis +0 σ (3.77).

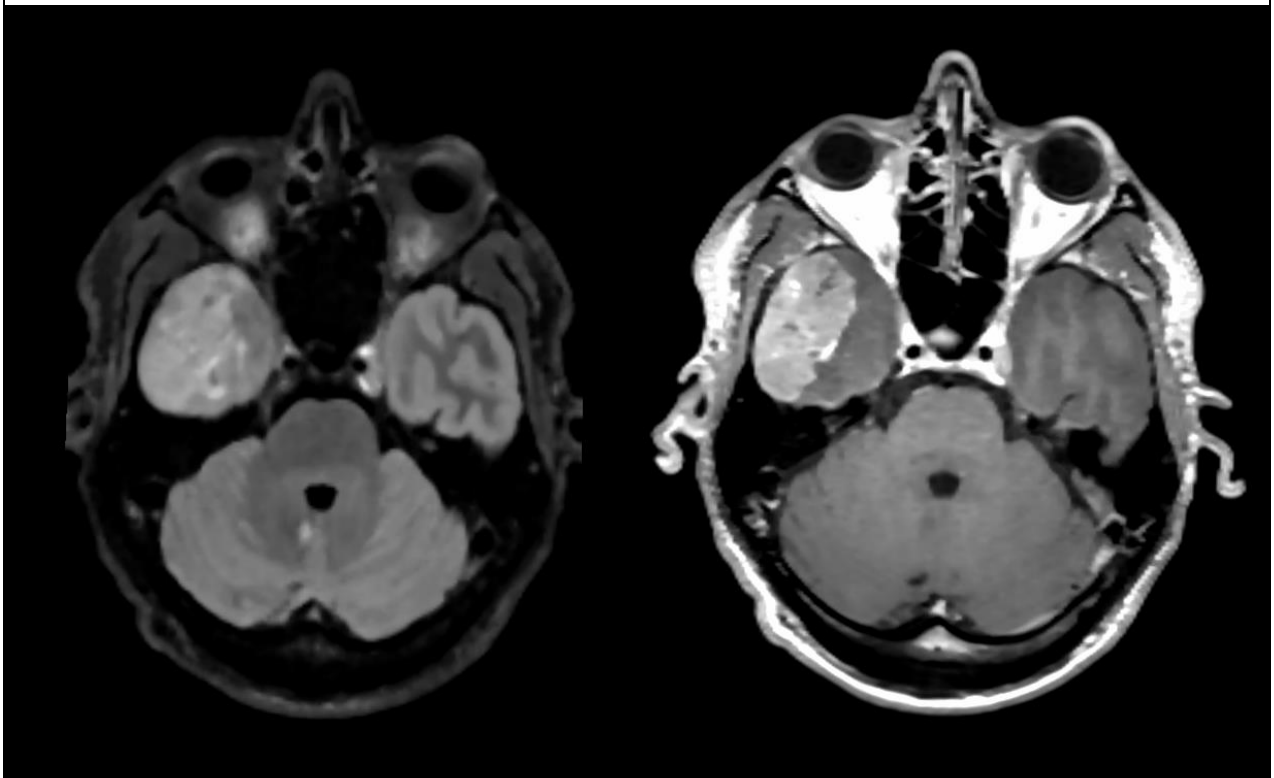


Figure 6. Exemple caricatural d'un GBM présentant une Skewness négative de $-0,79$ (-3.58σ) dans la région (a) FLAIR. Kurtosis à $8,67$ ($+3.96\sigma$).

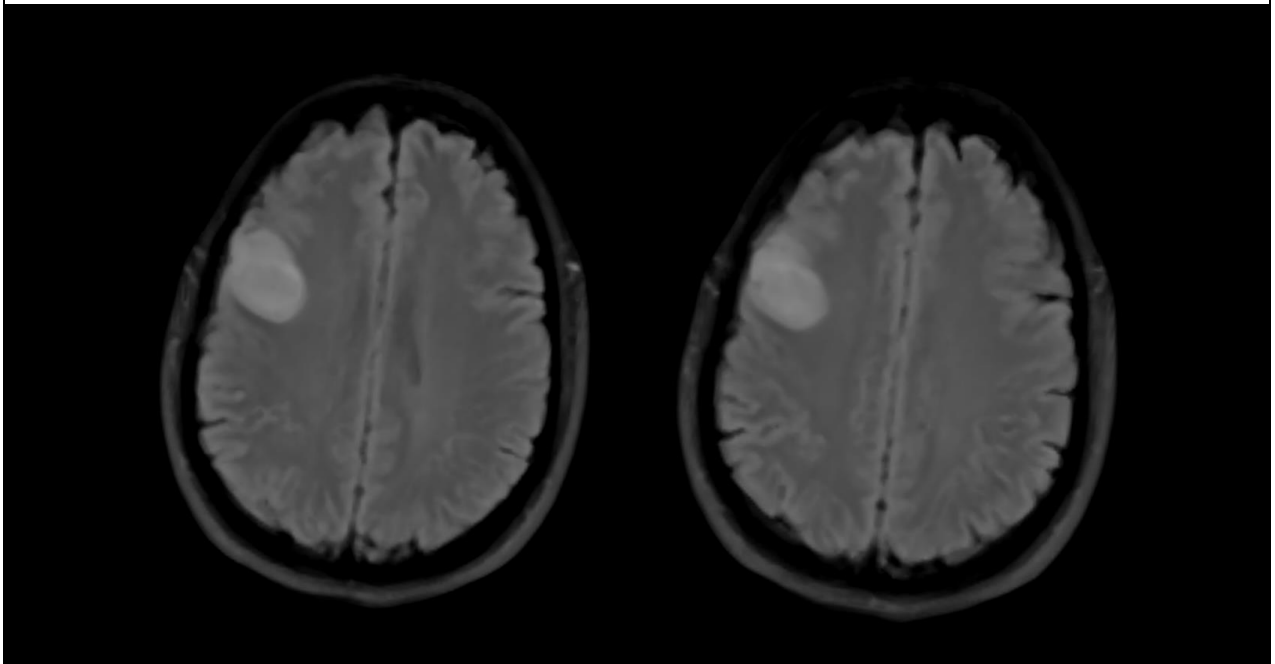
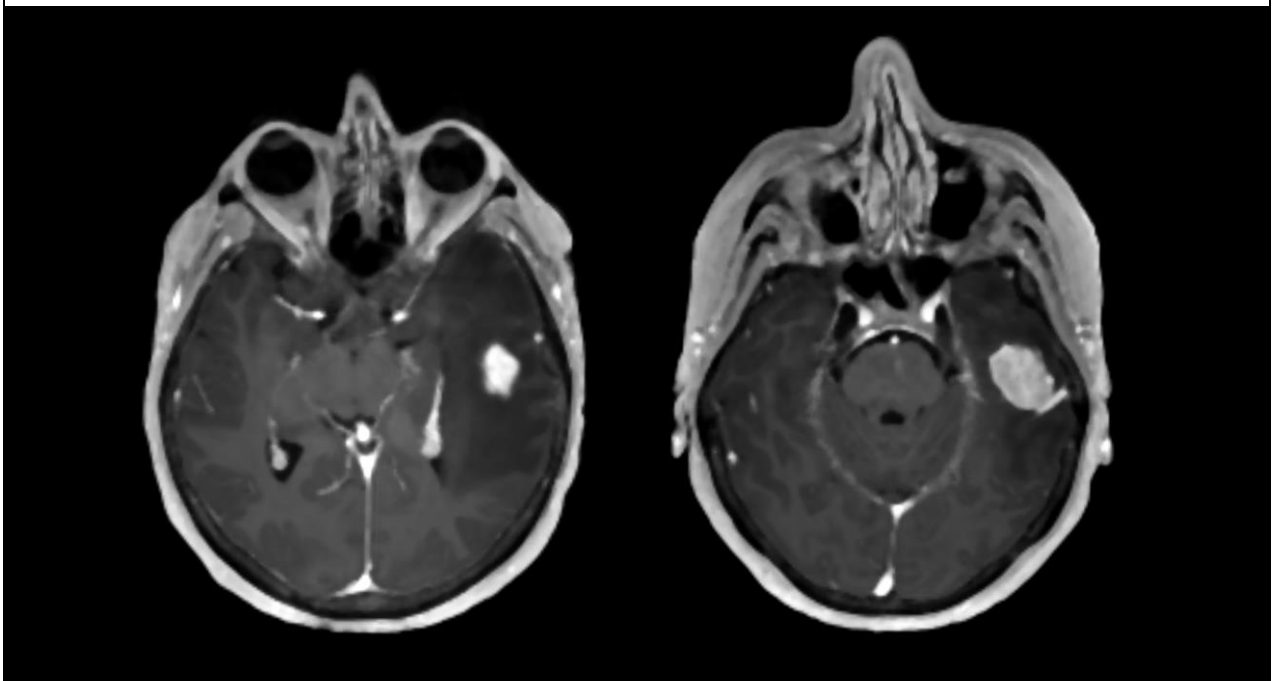


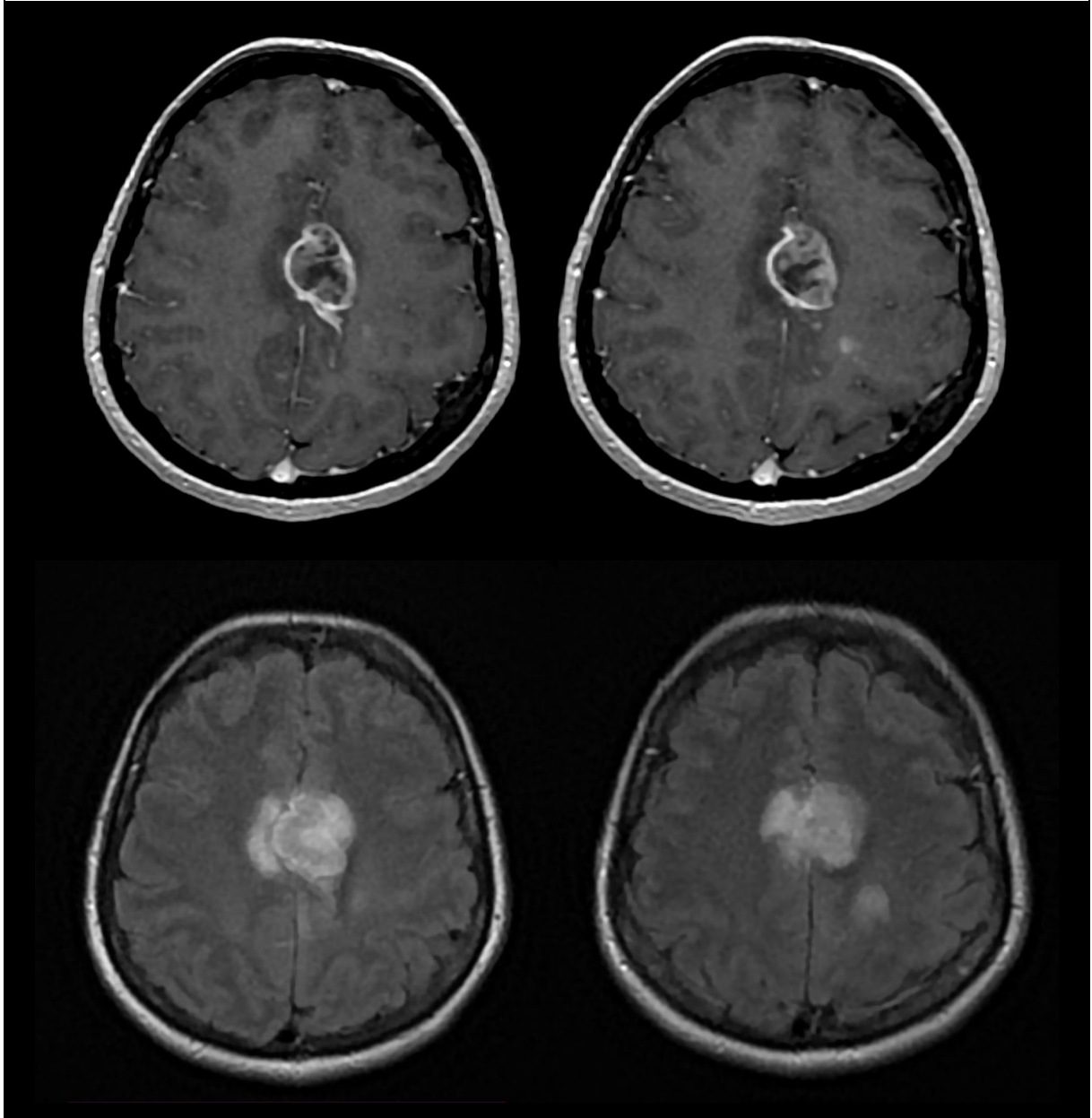
Figure 7. Exemple caricatural d'une MC présentant un Kurtosis élevé à $21,28$ ($+2.60\sigma$) de la région (b) T1. Le pic d'intensité correspond au rehaussement globalement homogène de la lésion.



Les cinq autres index de texture retenus par RFE évoluaient tous dans le sens d'une plus grande hétérogénéité des lésions GBM dans la région (b). Cela se traduisait par une augmentation relative du contraste et des différences de niveaux de gris sur les index `b_t1_Contrast` (NGTDM), `b_t1_Idn` (GLCM) (figure 8). Les nombreux niveaux d'intensité et leurs occurrences respectives majoraient la complexité de la matrice GLCM, augmentant l'index `b_t1_Imc1` (GLCM) pour les GBM.

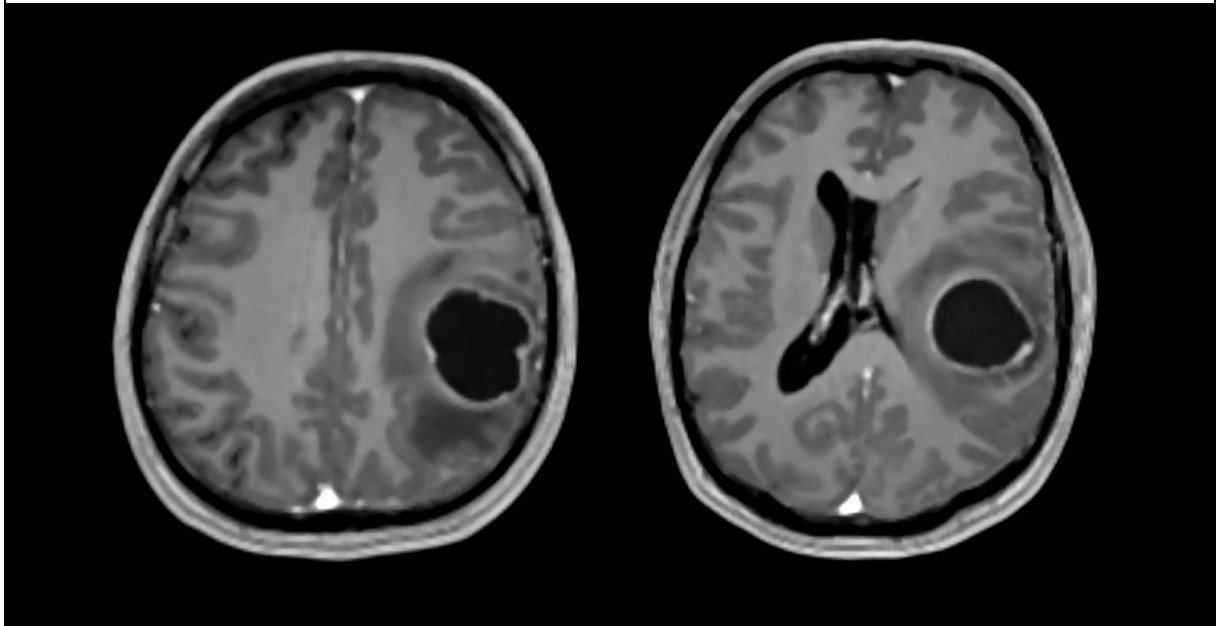
L'index `b_t1_GrayLevelNonUniformity` (GLRLM) variait dans le sens inverse de celui attendu, puisqu'il était inférieur pour les GBM traduisant une plus faible variation des niveaux d'intensités que pour les MC, allant à l'encontre d'une hétérogénéité. Cependant, cet aspect est à nouveau à pondérer avec la région péri-rehaussement (d) explorée par la segmentation (b), plus volumineuse pour les MC et pouvant aboutir à une variation globale plus marquée pour ces dernières. En effet, le même index exploré uniquement dans la région (a) T1 retrouvait une valeur moyenne significativement augmentée pour les GBM par rapport aux MC (7545,9 vs 5173,1, $p=0.027$) confirmant l'effet du volume de la région (d).

Figure 8. Exemple caricatural d'un GBM présentant un $b_t1_Contrast$ (NGTDM) élevé à 29,74 (+5.02 σ) de la région (b) T1. b_t1_Idn (GLCM) nettement abaissé à 0.941 (-1.90 σ) et $b_t1_GrayLevelNonUniformity$ (GLRLM) nettement abaissé à 1329.4 (-0.98 σ).



Enfin, l'index $b_t1_LongRunLowGrayLevelEmphasis$ (GLRLM) faisait état des régions nécrotiques ou kystiques de bas niveaux de gris, homogènes, se traduisant par des séries spatiales plus longues de mêmes niveaux d'intensités, plus nombreuses pour les MC (figure 9).

Figure 9. Exemple caricatural d'une MC présentant un $b_t1_LongRunLowGrayLevelEmphasis$ élevé à $0.50 (+2.16\sigma)$ de la région (b) T1.



Aucun index explorant uniquement la région (d) de l'hypersignal FLAIR péri-rehaussement ne fut sélectionné par la sélection RFE, LASSO ou SBF. Ceci pourrait paraître surprenant, d'autant qu'il est établi que l'hypersignal FLAIR péri-rehaussement constitue un infiltrat glial pour les GBM et un œdème vasogénique pour les MC (49). Les deux lésions sont parfois différenciables dans ces régions en imagerie de perfusion DSC, avec un bolus de premier passage responsable d'une chute maximal du signal en écho de gradient (peak) plus important pour les glioblastomes et lésions gliales de haut grade, en corrélation avec un volume sanguin cérébral augmenté (54). De plus, le retour à la ligne de base après le bolus est plus important pour les GBM, qui conservent une barrière hémato-encéphalique partiellement fonctionnelle, à l'inverse des MC dont les capillaires fenestrés plus importants sont dépourvus de jonctions serrées basales (54). Un pic relatif de choline par rapport à la créatine ou le NAA peut être décelable dans la région péri-rehaussement des GBM en spectroscopie (55).

La moyenne des différences des index radiomiques entre GBM et MC était de 0.216σ en valeur absolue pour le segment (d) de la région T1, suggérant la possibilité de séparer les deux groupes par cette seule région. Seule la réduction de dimension par mRMR retenait 3 index de cette région. mRMR était la seule méthode se basant sur la matrice d'information mutuelle, c'est-à-dire écartant les informations redondantes. En conséquence, une optimisation

supplémentaire du meilleur modèle serait possible, en appliquant cet algorithme en supplément des autres méthodes employées.

Parmi 186 index testés pour leur stabilité, seuls 54,8% en T1 et 33,3% en FLAIR avaient une très bonne répétabilité, supérieure à 0.80. Parmi les index retenus dans le modèle RFE11, à l'exclusion des index morphologiques, la stabilité moyenne des 8 index n'était que de 68,5% avec seulement 4 index supérieurs à 0.70. Cette instabilité a pu causer un biais à l'entraînement des modèles. Toutefois, la méthode d'évaluation de la stabilité n'était pas optimale, puisqu'elle s'appuyait sur une segmentation répétée homolatérale et controlatérale de la substance blanche saine.

3. Limites en design

Concernant la population, l'inclusion était effectuée de manière rétrospective et consécutive pour les GBM et MC. Les types histologiques des MC étaient des adénocarcinomes peu à bien différenciés avec une large majorité d'origine pulmonaire. Les origines mammaires et colorectales étaient également représentées. Cependant, les origines cutanées ou rénales, ainsi que les autres sous-types tels que les carcinomes épidermoïdes et les cancers neuroendocrines à petites cellules pulmonaires n'étaient pas présents dans cette étude (49).

En termes de design d'étude, il aurait été préférable de comparer le modèle hybride à un modèle radiologue pour estimer un éventuel avantage apporté par ce type d'approche. De plus, une double segmentation aurait permis un contrôle par expert et une évaluation plus précise de la stabilité des index radiomique (135). La segmentation semi-automatique par propagation avec contrôle qualitatif était chronophage, robuste et plus rapide qu'une segmentation coupe à coupe. Plusieurs méthodes de segmentation automatique sont envisageables mais doivent atteindre des performances élevées pour éviter un biais, et complétées par une évaluation visuelle (136,137). Le traitement des images n'était pas effectué en aveugle de la classe.

La problématique de différenciation des GBM et MC a été traitée de manière binaire et mutuellement exclusive. En routine clinique, les probabilités a priori ne se limitent pas seulement à cette dichotomie et doivent être confrontées à celles d'un lymphome cérébral primitif, d'un abcès cérébral, ou encore d'une lésion inflammatoire pseudo-tumorale (132).

75,9% des lésions test présentaient une faible difficulté de classification, puisqu'elles étaient bien classées par plus de 75% des modèles. Une approche visant à quantifier la difficulté diagnostique, par exemple devant l'échec des multiples classifieurs ou un accord inter-observateur faible entre radiologues de niveaux d'expertise différents, pourrait permettre d'identifier les lésions problématiques. Cela permettrait ensuite de cibler l'entraînement des classifieurs sur ces lésions si l'on souhaite générer un modèle capable de répondre à une problématique d'expert. Pour une aide au diagnostic plus généraliste, il serait en revanche intéressant d'équilibrer le panel de difficulté des lésions.

Ce travail n'explorait que les séquences T1 injectée et T2 FLAIR. Il est probable qu'une association à d'autres séquences, comme les séquences de perfusion DSC, diffusion, T1 sans injection voire spectroscopie, pourrait permettre d'augmenter les performances (138). Cependant, cette augmentation de données devrait s'accompagner d'une majoration de la cohorte explorée.

Le rééchantillonnage spatial était de 1x1x1mm correspondant à la meilleure résolution de la séquence T1 injectée. De ce fait il y avait une interpolation des séquences Flair, avec une augmentation artificielle des données pour le calcul des index radiomiques.

Le CNN présentait des performances intermédiaires lorsqu'il était utilisé isolément. Cela est probablement lié à un manque de données, avec plus de 7000 images d'entraînement pour seulement 185 lésions analysées, pouvant engendrer un surajustement. Ce dernier était contrôlé par la restauration des poids au minimum atteint par la fonction de perte de l'échantillon de validation. Une augmentation du nombre d'observations et une diminution du ratio images par lésion pourrait optimiser le réseau.

4. Perspectives

Cette étude pourrait être complétée par une confrontation du meilleur modèle RFE11_SVM à un échantillon de test externe, notamment dans un centre avec des machines et méthodes d'acquisition différentes pour en évaluer la robustesse. Une confrontation à un modèle radiologue, qui reste la référence diagnostique avant l'histologie, devrait être réalisée, notamment pour comparer les courbes ROC, et explorer les sujets difficilement classables : sont-

ils les mêmes pour les classifieurs et les radiologues, ou différent-ils ? Le dernier cas échéant, une plus-value diagnostique du modèle d'apprentissage automatique serait envisageable.

Dans ce travail, différents modèles ont été entraînés afin d'identifier le plus performant. Cependant, il est probable que ce processus de sélection ait entraîné une perte de puissance diagnostique par rapport à un modèle idéal en se privant des autres algorithmes. Récemment, et pour palier à la variabilité des algorithmes selon le type de variables questionnées, des méthodes d'apprentissage par ensembles (*ensemble learning*) se sont démocratisées (139). Ainsi, l'apprentissage par empilement (*stacked generalization*) consiste à générer un super classifieur (*super learner*) qui évalue un ensemble d'algorithmes pour une question posée. Ces algorithmes peuvent être mis en compétition, mais aussi être associés séquentiellement, avec le classifieur d'aval exploitant les erreurs du classifieur d'amont. Les algorithmes peuvent aussi être mis à contribution en proportions égales, donnant des résultats de prédiction pondérés (*model averaging*) (139).

L'hétérogénéité intra-tumorale, spatiale et temporelle de différentes lésions néoplasiques a été explorée depuis le développement des nouvelles techniques de séquençage, et témoigne de contingents cellulaires présentant des différences de statuts génétiques ou d'expression génétique (140,141). Concernant l'oncogenèse des GBM, la croissance tumorale passe par une amplification des Récepteurs à la Tyrosine Kinase (RTK) tels qu'EGFR, KIT, VEGFR2, PDGFRA ou MET (140). Le statut de ces RTK pourrait être associé à baisse de l'efficacité de certaines thérapies ciblées (140). L'exploration de corrélations entre les index radiomiques d'hétérogénéité et les différents statuts génétiques pourrait permettre de rendre compte de l'évolution temporelle et de la distribution spatiale des contingents cellulaires, notamment sous traitement ciblé. A ce titre, des cartographies faisant état de l'entropie locale de l'information via des index de texture pourrait être un marqueur de cette hétérogénéité.

L'émergence de modèles hybrides d'apprentissage automatique explorant tout type d'information, telles que des données cliniques, biologiques, génétiques, radiologiques, histologiques etc. et dont ils détermineraient eux-mêmes l'utilité lors de l'entraînement, semble fortement probable. Ces modèles seraient également capables de déterminer l'algorithme le plus efficace à l'entraînement, voire d'associer plusieurs types d'algorithmes en parallèle, chacun étant le plus adapté à un type d'information. En sortie d'algorithme, au-delà d'une classification dichotomique, ces modèles seraient multi-classes, nécessitant toutefois de larges cohortes. Selon l'évolution favorable des algorithmes de détection des lésions et d'auto-segmentation, ainsi que

du traitement automatique du langage en interaction avec les bases de données de compte rendus, des tentatives d'automatisation de bout en bout du processus diagnostique pourrait voir le jour. Il sera alors nécessaire d'en considérer tout l'impact éthique, eu égard au caractère peu explicite d'algorithmes complexes.

V. CONCLUSION

L'objectif principal de ce travail était d'entraîner et de tester des modèles radiomiques hybrides, capables de distinguer les GBM des MC avec preuve histologique, sur une IRM pré-thérapeutique, en séquences T1 injectée et T2 FLAIR.

164 sujets (94 GBM et 70 MC), étaient analysés après inclusion rétrospective et 185 lésions mesurées (102 GBM et 83 MC). Les IRM étaient issues de 27 machines différentes et 20 centres. Trois segmentations étaient analysées : le rehaussement T1, l'hypersignal FLAIR péri-rehaussement et la lésion entière. 642 index radiomiques étaient extraits, transformés par loi centrée réduite (Z-score). La cohorte était divisée en échantillon d'entraînement (68,6%) et de test (31,4%). L'échantillon d'entraînement voyait ses index réduits à 425 par filtre statistique (test de Student, $p < 0.40$).

Puis, une sélection des index par mRMR, RFE, LASSO, SBF ou kPCA était effectuée sur l'échantillon d'entraînement. Les modèles hybrides incluaient les index radiomiques sélectionnés, l'âge, le sexe, une description IRM qualitative et un score de probabilité issu d'un CNN 2D. Ce dernier était entraîné sur les images T1 et FLAIR.

31 modèles furent entraînés. Le modèle hybride RFE11_SVM avec 11 index radiomiques présentait les meilleures performances : $AUC_{\text{test}}=0,928$ $IC_{95\%}[0,853;1]$, exactitude=91,4% $IC_{95\%}[0,842 ; 0,986]$, Sensibilité (GBM)=0,969 $IC_{95\%}[0,924 ; 1]$, Spécificité=0,846 $IC_{95\%}[0,753 ; 0,939]$. Trois index morphologiques montraient une sphéricité augmentée des MC. Cinq index de texture et 3 index de premier ordre étaient en faveur d'une hétérogénéité lésionnelle plus marquée des GBM.

Au total, un modèle radiomique hybride est capable de différencier efficacement GBM et MC, en étant entraîné sur des données radiomiques, cliniques, descriptives IRM et augmenté par un CNN.

VI. BIBLIOGRAPHIE

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. mars 2012;48(4):441-6.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. févr 2016;278(2):563-77.
3. pubmeddev. Home - PubMed - NCBI [Internet]. [cité 12 févr 2020]. Disponible sur: <https://www.ncbi.nlm.nih.gov/pubmed/>
4. Poggio T, Mhaskar H, Rosasco L, Miranda B, Liao Q. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *Int J Autom Comput*. oct 2017;14(5):503-19.
5. Gillies RJ, Anderson AR, Gatenby RA, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin Radiol*. juill 2010;65(7):517-21.
6. Sun Y, Reynolds HM, Parameswaran B, Wraith D, Finnegan ME, Williams S, et al. Multiparametric MRI and radiomics in prostate cancer: a review. *Australas Phys Eng Sci Med*. mars 2019;42(1):3-25.
7. Ather S, Kadir T, Gleeson F. Artificial intelligence and radiomics in pulmonary nodule management: current status and future applications. *Clinical Radiology*. janv 2020;75(1):13-9.
8. Wu W, Pierce LA, Zhang Y, Pipavath SNJ, Randolph TW, Lastwika KJ, et al. Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study. *Eur Radiol*. nov 2019;29(11):6100-8.
9. Jun Wang null, Xia Liu null, Di Dong null, Jiangdian Song null, Min Xu null, Yali Zang null, et al. Prediction of malignant and benign of lung tumor using a quantitative radiomic method. *Conf Proc IEEE Eng Med Biol Soc*. 2016;2016:1272-5.
10. Jia T-Y, Xiong J-F, Li X-Y, Yu W, Xu Z-Y, Cai X-W, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol*. 1 sept 2019;29(9):4742-50.
11. Liu Y, Kim J, Balagurunathan Y, Li Q, Garcia AL, Stringfield O, et al. Radiomic Features Are Associated With EGFR Mutation Status in Lung Adenocarcinomas. *Clinical Lung Cancer*. 1 sept 2016;17(5):441-448.e6.
12. Choi Y, Kim SH, Youn IK, Kang BJ, Park W, Lee A. Rim sign and histogram analysis of apparent diffusion coefficient values on diffusion-weighted MRI in triple-negative breast cancer: Comparison with ER-positive subtype. *PLoS One*. 18 mai 2017;12(5):e0177903.

13. Leithner D, Bernard-Davila B, Martinez DF, Horvat JV, Jochelson MS, Marino MA, et al. Radiomic Signatures Derived from Diffusion-Weighted Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes. *Mol Imaging Biol.* avr 2020;22(2):453-61.
14. Li H, Zhu Y, Burnside ES, Huang E, Drukker K, Hoadley KA, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer.* 2016;2:16012.
15. Waugh SA, Purdie CA, Jordan LB, Vinnicombe S, Lerski RA, Martin P, et al. Magnetic resonance imaging texture analysis classification of primary breast cancer. *Eur Radiol.* févr 2016;26(2):322-30.
16. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. *Sci Rep.* 7 déc 2015;5(1):17787.
17. Nougaret S, McCague C, Tibermacine H, Vargas HA, Rizzo S, Sala E. Radiomics and radiogenomics in ovarian cancer: a literature review. *Abdom Radiol.* juin 2021;46(6):2308-22.
18. Meier A, Veeraraghavan H, Nougaret S, Lakhman Y, Sosa R, Soslow RA, et al. Association between CT-texture-derived tumor heterogeneity, outcomes, and BRCA mutation status in patients with high-grade serous ovarian cancer. *Abdom Radiol.* 1 juin 2019;44(6):2040-7.
19. Vargas HA, Veeraraghavan H, Micco M, Nougaret S, Lakhman Y, Meier AA, et al. A novel representation of inter-site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome. *Eur Radiol.* 1 sept 2017;27(9):3991-4001.
20. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E, et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat Commun.* 15 févr 2019;10(1):764.
21. Jian J, Li Y, Pickhardt PJ, Xia W, He Z, Zhang R, et al. MR image-based radiomics to differentiate type I and type II epithelial ovarian cancers. *Eur Radiol.* 1 janv 2021;31(1):403-10.
22. Qian L, Ren J, Liu A, Gao Y, Hao F, Zhao L, et al. MR imaging of epithelial ovarian cancer: a combined model to predict histologic subtypes. *Eur Radiol.* 1 nov 2020;30(11):5815-25.
23. Song X, Ren J-L, Zhao D, Wang L, Ren H, Niu J. Radiomics derived from dynamic contrast-enhanced MRI pharmacokinetic protocol features: the value of precision diagnosis ovarian neoplasms. *Eur Radiol.* 1 janv 2021;31(1):368-78.
24. Wang G, Sun Y, Chen Y, Gao Q, Peng D, Lin H, et al. Rapid identification of human ovarian cancer in second harmonic generation images using radiomics feature analyses

- and tree-based pipeline optimization tool. *Journal of Biophotonics*. 2020;13(9):e202000050.
25. Zhang H, Mao Y, Chen X, Wu G, Liu X, Zhang P, et al. Magnetic resonance imaging radiomics in categorizing ovarian masses and predicting clinical outcome: a preliminary study. *Eur Radiol*. 1 juill 2019;29(7):3358-71.
 26. Horvat N, Bates DDB, Petkovska I. Novel imaging techniques of rectal cancer: what do radiomics and radiogenomics have to offer? A literature review. *Abdom Radiol*. 1 nov 2019;44(11):3764-74.
 27. Huang Y, Liang C, He L, Tian J, Liang C, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *JCO*. 20 juin 2016;34(18):2157-64.
 28. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? *Eur Radiol*. 1 mai 2018;28(5):2058-67.
 29. Golia Pernicka JS, Gagniere J, Chakraborty J, Yamashita R, Nardo L, Creasy JM, et al. Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol*. 1 nov 2019;44(11):3755-63.
 30. Liang C, Huang Y, He L, Chen X, Ma Z, Dong D, et al. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget*. 22 avr 2016;7(21):31401-12.
 31. Peng Z, Wang Y, Wang Y, Jiang S, Fan R, Zhang H, et al. Application of radiomics and machine learning in head and neck cancers. *Int J Biol Sci*. 1 janv 2021;17(2):475-86.
 32. Wang F, Zhang B, Wu X, Liu L, Fang J, Chen Q, et al. Radiomic Nomogram Improves Preoperative T Category Accuracy in Locally Advanced Laryngeal Carcinoma. *Front Oncol*. 15 oct 2019;9:1064.
 33. Ren J, Tian J, Yuan Y, Dong D, Li X, Shi Y, et al. Magnetic resonance imaging based radiomics signature for the preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. *Eur J Radiol*. sept 2018;106:1-6.
 34. Wu W, Ye J, Wang Q, Luo J, Xu S. CT-Based Radiomics Signature for the Preoperative Discrimination Between Head and Neck Squamous Cell Carcinoma Grades. *Frontiers in Oncology*. 2019;9:821.
 35. Mukherjee P, Cintra M, Huang C, Zhou M, Zhu S, Colevas AD, et al. CT-based Radiomic Signatures for Predicting Histopathologic Features in Head and Neck Squamous Cell Carcinoma. *Radiol Imaging Cancer*. 15 mai 2020;2(3):e190039.
 36. Fernandes C, Costa A, Osório L, Lago RC, Linhares P, Carvalho B, et al. Current Standards of Care in Glioblastoma Therapy. *Exon Publications*. 20 sept 2017;197-241.
 37. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol*. 1 juin 2016;131(6):803-20.

38. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. *New England Journal of Medicine*. 10 mars 2005;352(10):987-96.
39. Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, et al. A Randomized Trial of Bevacizumab for Newly Diagnosed Glioblastoma [Internet]. <http://dx.doi.org/10.1056/NEJMoa1308573>. Massachusetts Medical Society; 2014 [cité 3 juill 2020]. Disponible sur: <https://www.nejm.org/doi/10.1056/NEJMoa1308573>
40. Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, et al. Bevacizumab plus Radiotherapy–Temozolomide for Newly Diagnosed Glioblastoma [Internet]. <http://dx.doi.org/10.1056/NEJMoa1308345>. Massachusetts Medical Society; 2014 [cité 3 juill 2020]. Disponible sur: <https://www.nejm.org/doi/10.1056/NEJMoa1308345>
41. BW S, CP W. World Cancer Report 2014 [Internet]. [cité 5 sept 2021]. Disponible sur: <https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014>
42. Lowery FJ, Yu D. Brain Metastasis: Unique Challenges and Open Opportunities. *Biochim Biophys Acta*. janv 2017;1867(1):49-57.
43. Kaye AH. *Essential Neurosurgery*. Wiley; 2005. 309 p.
44. Kornienko VN, Pronin IN. *Diagnostic Neuroradiology*. Springer Science & Business Media; 2008. 1295 p.
45. Lin X, DeAngelis LM. Treatment of Brain Metastases. *J Clin Oncol*. 20 oct 2015;33(30):3475-84.
46. Schiff null. Single Brain Metastasis. *Curr Treat Options Neurol*. janv 2001;3(1):89-99.
47. Patil CG, Yi A, Elramsisy A, Hu J, Mukherjee D, Irvin DK, et al. Prognosis of patients with multifocal glioblastoma: a case-control study: Clinical article. *Journal of Neurosurgery*. 1 oct 2012;117(4):705-11.
48. Singh G, Mehrotra A, Sardhara J, Das KK, Jamdar J, Pal L, et al. Multiple glioblastomas: Are they different from their solitary counterparts? *Asian Journal of Neurosurgery*. 10 janv 2015;10(4):266.
49. Dietemann J-L. *Neuro-imagerie diagnostique*. 3e édition. Issy-les-Moulineaux: Elsevier Masson; 2018. 808 p.
50. Dziurzynski K, Blas-Boria D, Suki D, Cahill DP, Prabhu SS, Puduvalli V, et al. Butterfly glioblastomas: a retrospective review and qualitative assessment of outcomes. *J Neurooncol*. sept 2012;109(3):555-63.
51. Lasocki A, Gaillard F, Tacey M, Drummond K, Stuckey S. Incidence and prognostic significance of non-enhancing cortical signal abnormality in glioblastoma. *J Med Imaging Radiat Oncol*. févr 2016;60(1):66-73.

52. Lasocki A, Gaillard F, Tacey M, Drummond K, Stuckey S. Multifocal and multicentric glioblastoma: Improved characterisation with FLAIR imaging and prognostic implications. *Journal of Clinical Neuroscience*. sept 2016;31:92-8.
53. Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of Neurosurgery*. 1 août 2001;95(2):190-8.
54. Cha S, Lupo JM, Chen M-H, Lamborn KR, McDermott MW, Berger MS, et al. Differentiation of Glioblastoma Multiforme and Single Brain Metastasis by Peak Height and Percentage of Signal Intensity Recovery Derived from Dynamic Susceptibility-Weighted Contrast-Enhanced Perfusion MR Imaging. *American Journal of Neuroradiology*. 1 juin 2007;28(6):1078-84.
55. Server A, Josefsen R, Kulle B, Mæhlen J, Schellhorn T, Gadmar Ø, et al. Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors: *Acta Radiologica* [Internet]. 1 avr 2010 [cité 4 mars 2020]; Disponible sur: <https://journals.sagepub.com/doi/10.3109/02841850903482901>
56. Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology*. 1 févr 2019;290(2):479-87.
57. Essig M, Nguyen TB, Shiroishi MS, Saake M, Provenzale JM, Enterline DS, et al. Perfusion MRI: The Five Most Frequently Asked Clinical Questions. *AJR Am J Roentgenol*. sept 2013;201(3):10.2214/AJR.12.9544.
58. Artzi M, Bressler I, Bashat DB. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *Journal of Magnetic Resonance Imaging*. 2019;50(2):519-28.
59. Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. *Cancer Letters*. 1 juin 2019;451:128-35.
60. Han Y, Zhang L, Wei J, Zhang S, Chen S, Tian J, et al. Differentiation between Glioblastoma and Metastases from Lung and Other Sites by Combination Clinic/Routine MRI Radiomics [Internet]. Rochester, NY: Social Science Research Network; 2019 oct [cité 5 mars 2020]. Report No.: ID 3475577. Disponible sur: <https://papers.ssrn.com/abstract=3475577>
61. Kong D-S, Kim J, Ryu G, You H-J, Sung JK, Han YH, et al. Quantitative radiomic profiling of glioblastoma represents transcriptomic expression. *Oncotarget*. 5 janv 2018;9(5):6336-45.
62. Kickingreder P, Burth S, Wick A, Götz M, Eidel O, Schlemmer H-P, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. *Radiology*. 20 juin 2016;280(3):880-9.

63. Beig N, Patel J, Prasanna P, Hill V, Gupta A, Correa R, et al. Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma. *Scientific Reports*. 8 janv 2018;8(1):1-11.
64. Lao J, Chen Y, Li Z-C, Li Q, Zhang J, Liu J, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep*. 04 2017;7(1):10353.
65. Lee J, Narang S, Martinez JJ, Rao G, Rao A. Associating spatial diversity features of radiologically defined tumor habitats with epidermal growth factor receptor driver status and 12-month survival in glioblastoma: methods and preliminary investigation. *J Med Imaging (Bellingham)*. oct 2015;2(4):041006.
66. Hu LS, Ning S, Eschbacher JM, Baxter LC, Gaw N, Ranjbar S, et al. Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro Oncol*. janv 2017;19(1):128-37.
67. Yang G, Jones TL, Howe FA, Barrick TR. Morphometric model for discrimination between glioblastoma multiforme and solitary metastasis using three-dimensional shape analysis. *Magnetic Resonance in Medicine*. 2016;75(6):2505-16.
68. Larue RTHM, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncologica*. 2 nov 2017;56(11):1544-53.
69. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-level discretization impacts reproducible MRI radiomics texture features. Fan Y, éditeur. *PLoS ONE*. 7 mars 2019;14(3):e0213459.
70. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic Resonance Imaging*. janv 2004;22(1):81-91.
71. Derclé L, Ammari S, Bateson M, Durand PB, Haspinger E, Massard C, et al. Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: ROI-area, acquisition protocol and tissue site exert substantial influence. *Sci Rep*. 11 2017;7(1):7952.
72. Midya A, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J Med Imaging (Bellingham)*. janv 2018;5(1):011020.
73. Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The Quantitative Imaging Network: NCI's Historical Perspective and Planned Goals. *Transl Oncol*. févr 2014;7(1):1-4.
74. Mulshine JL, Gierada DS, Armato SG, Avila RS, Yankelevitz DF, Kazerooni EA, et al. Role of the Quantitative Imaging Biomarker Alliance in optimizing CT for the evaluation of lung cancer screen-detected nodules. *J Am Coll Radiol*. avr 2015;12(4):390-5.

75. ESR statement on the stepwise development of imaging biomarkers. *Insights Imaging*. 9 févr 2013;4(2):147-52.
76. O'Connor JPB, Aboagye EO, Adams JE, Aerts HJWL, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14(3):169-86.
77. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. arXiv:161207003 [cs, eess] [Internet]. 17 déc 2019 [cité 10 févr 2020]; Disponible sur: <http://arxiv.org/abs/1612.07003>
78. Wu X, Yang Z, Peng J, Zhou J. Global denoising for 3D MRI. *Biomed Eng Online*. 12 mai 2016;15(1):54.
79. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. juin 2010;29(6):1310-20.
80. Kalavathi P, Prasath VBS. Methods on Skull Stripping of MRI Head Scan Images-a Review. *J Digit Imaging*. 2016;29(3):365-79.
81. Cho ZH, Ro YM. Reduction of susceptibility artifact in gradient-echo imaging. *Magnetic Resonance in Medicine*. 1992;23(1):193-200.
82. Brooks RA, Chiro GD. Beam hardening in X-ray reconstructive tomography. *Phys Med Biol*. mai 1976;21(3):390-8.
83. Johnson HJ, Harris G, Williams K. BRAINSFit: Mutual Information Registrations of Whole-Brain 3D Images, Using the Insight Toolkit. *The Insight Journal*. 30 nov 2007;180.
84. Statistical Parametric Mapping: The Analysis of Functional Brain Images - 1st Edition [Internet]. [cité 22 janv 2020]. Disponible sur: <https://www.elsevier.com/books/statistical-parametric-mapping-the-analysis-of-functional-brain-images/penny/978-0-12-372560-8>
85. Hounsfield GN. Computed medical imaging. Nobel lecture, Decemberr 8, 1979. *J Comput Assist Tomogr*. oct 1980;4(5):665-74.
86. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. *PLoS One* [Internet]. 15 juill 2014 [cité 23 janv 2020];9(7). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4098900/>
87. Velazquez ER, Aerts HJWL, Gu Y, Goldgof DB, De Ruyscher D, Dekker A, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol*. nov 2012;105(2):167-73.
88. Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, et al. Reproducibility of dynamic contrast-enhanced MR imaging. Part II. Comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. *Radiology*. mars 2013;266(3):812-21.

89. Velazquez ER, Parmar C, Jermoumi M, Mak RH, van Baardwijk A, Fennessy FM, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep* [Internet]. 18 déc 2013 [cité 23 janv 2020];3. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3866632/>
90. Vezhnevets V, Konouchine V. “GrowCut” - Interactive Multi-Label N-D Image Segmentation By Cellular Automata. 2005;7.
91. Kikinis R, Pieper S. 3D Slicer as a tool for interactive brain tumor segmentation. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society [Internet]. Boston, MA: IEEE; 2011 [cité 23 janv 2020]. p. 6982-4. Disponible sur: <http://ieeexplore.ieee.org/document/6091765/>
92. Thevenaz P, Blu T, Unser M. Image Interpolation and Resampling. In: *Handbook of Medical Imaging, Processing and Analysis*. Academic Press; 2000. p. 393-420.
93. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*. 3 juin 2014;5(1):1-9.
94. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present... any future? *Eur J Nucl Med Mol Imaging*. janv 2017;44(1):151-65.
95. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 1 nov 2017;77(21):e104-7.
96. Brock A, Chang H, Huang S. Non-genetic heterogeneity — a mutation-independent driving force for the somatic evolution of tumours. *Nature Reviews Genetics*. mai 2009;10(5):336-42.
97. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 19 sept 2013;501(7467):338-45.
98. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nature Reviews Cancer*. mai 2012;12(5):323-34.
99. Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: Preliminary evidence of an association with tumour metabolism, stage, and survival. *Clinical Radiology*. 1 févr 2012;67(2):157-64.
100. Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging*. 6 juill 2010;10(1):137-43.
101. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of Primary Colorectal Cancer Heterogeneity by Using Whole-Tumor Texture Analysis: Contrast-enhanced CT Texture as a Biomarker of 5-year Survival. *Radiology*. 1 janv 2013;266(1):177-84.

102. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, et al. Locally Advanced Squamous Cell Carcinoma of the Head and Neck: CT Texture and Histogram Analysis Allow Independent Prediction of Overall Survival in Patients Treated with Induction Chemotherapy. *Radiology*. 1 déc 2013;269(3):801-9.
103. Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep*. 24 mars 2016;6:23428.
104. Bouyer J. Méthodes statistiques : Médecine - Biologie [Internet]. VUIBERT; 2017 [cité 26 sept 2021]. 400 p. Disponible sur: https://www.amazon.fr/M-C3-A9thodes-statistiques-M-C3-A9decine-Biologie-Jean-Bouyer-dp-2311660071/dp/2311660071/ref=dp_ob_title_bk
105. Sun R, Limkin EJ, Dercle L, Reuzé S, Zacharaki EI, Chargari C, et al. Imagerie médicale computationnelle (radiomique) et potentiel en immuno-oncologie. *Cancer/Radiothérapie*. oct 2017;21(6-7):648-54.
106. Hamon J. Optimisation combinatoire pour la sélection de variables en régression en grande dimension : Application en génétique animale [Internet] [phdthesis]. Université des Sciences et Technologie de Lille - Lille I; 2013 [cité 14 févr 2020]. Disponible sur: <https://tel.archives-ouvertes.fr/tel-00920205>
107. Yu L, Liu H. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. :8.
108. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. août 2005;27(8):1226-38.
109. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics*. 15 sept 2013;29(18):2365-8.
110. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. :26.
111. Saghapour E, Kermani S, Sehhati M. A novel feature ranking method for prediction of cancer stages using proteomics data. *PLoS One* [Internet]. 21 sept 2017 [cité 17 févr 2020];12(9). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608217/>
112. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2005;67(2):301-20.
113. Jolliffe I. Principal Component Analysis. In: *Encyclopedia of Statistics in Behavioral Science* [Internet]. American Cancer Society; 2005 [cité 6 nov 2021]. Disponible sur: <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013192.bsa501>
114. Chen L-F, Liao H-YM, Ko M-T, Lin J-C, Yu G-J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*. 1 oct 2000;33(10):1713-26.
115. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *RadioGraphics*. 17 févr 2017;37(2):505-15.

116. Giger ML. Machine Learning in Medical Imaging. *Journal of the American College of Radiology*. 1 mars 2018;15(3, Part B):512-20.
117. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition [Internet]. O'Reilly Media, Inc. 2019 [cité 11 mars 2020]. Disponible sur: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
118. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 1 août 2018;9(4):611-29.
119. Mitchell T. *Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression* [Internet]. 2005 [cité 11 mars 2020]. Disponible sur: <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>
120. Chiu Y-WC. *Machine Learning with R Cookbook*. :997.
121. Günther F, Fritsch S. neuralnet: Training of Neural Networks. *The R Journal*. 2010;2(1):30.
122. LeDell E. useR! Machine Learning Tutorial [Internet]. 2018 [cité 27 avr 2020]. Disponible sur: <https://koalaverse.github.io/machine-learning-in-R/>
123. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, éditeurs. *Advances in Neural Information Processing Systems 25* [Internet]. Curran Associates, Inc.; 2012 [cité 20 avr 2020]. p. 1097-105. Disponible sur: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
124. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis*. 1 déc 2015;115(3):211-52.
125. Papers with Code - ImageNet Benchmark (Image Classification) [Internet]. [cité 15 févr 2021]. Disponible sur: <https://paperswithcode.com/sota/image-classification-on-imagenet>
126. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *International Journal of Radiation Oncology*Biography*Physics*. 15 nov 2018;102(4):1143-58.
127. van Timmeren JE, Leijenaar RTH, van Elmpt W, Wang J, Zhang Z, Dekker A, et al. Test-Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomography*. déc 2016;2(4):361-5.
128. Lin LI-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*. 1989;45(1):255-68.
129. D'souza RN, Huang P-Y, Yeh F-C. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size. *Sci Rep*. 21 janv 2020;10(1):834.

130. Bae S, An C, Ahn SS, Kim H, Han K, Kim SW, et al. Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. *Scientific Reports*. 21 juill 2020;10(1):12110.
131. Yang G, Jones TL, Barrick TR, Howe FA. Discrimination between glioblastoma multiforme and solitary metastasis using morphological features derived from the $p : q$ tensor decomposition of diffusion tensor imaging: DISCRIMINATION OF GBM AND SOLITARY MET USING MORPHOLOGICAL FEATURES OF DTI. *NMR Biomed*. sept 2014;27(9):1103-11.
132. Osborn AG, Hedlund G-L, Salzman K-L. *Osborn's Brain : Imaging, pathology, and anatomy*. 2nd édition. Philadelphia, PA: Elsevier - Health Sciences Division; 2017. 1300 p.
133. Blanchet L, Krooshof PWT, Postma GJ, Idema AJ, Goraj B, Heerschap A, et al. Discrimination between Metastasis and Glioblastoma Multiforme Based on Morphometric Analysis of MR Images. *American Journal of Neuroradiology*. 1 janv 2011;32(1):67-73.
134. Stein AM, Demuth T, Mobley D, Berens M, Sander LM. A Mathematical Model of Glioblastoma Tumor Spheroid Invasion in a Three-Dimensional In Vitro Experiment. *Biophys J*. 1 janv 2007;92(1):356-65.
135. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers—From the Radiology Editorial Board. *Radiology*. 1 mars 2020;294(3):487-9.
136. Shelke SM, Mohod SW. Automated Segmentation and Detection of Brain Tumor from MRI. In: 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2018. p. 2120-6.
137. Charron O, Lallement A, Jarnet D, Noblet V, Clavier J-B, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Computers in Biology and Medicine*. avr 2018;95:43-54.
138. Zhang P, Liu B. Differentiation among Glioblastomas, Primary Cerebral Lymphomas, and Solitary Brain Metastases Using Diffusion-Weighted Imaging and Diffusion Tensor Imaging: A PRISMA-Compliant Meta-analysis. *ACS Chem Neurosci*. 5 févr 2020;11(3):477-83.
139. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. *Eur J Epidemiol*. 1 mai 2018;33(5):459-64.
140. Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, et al. Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell*. 13 déc 2011;20(6):810-7.
141. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol*. févr 2018;15(2):81-94.

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 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 et reconnaissant 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 d'opprobre et méprisé de mes confrères si j'y manque.

RESUME

Objectif : Générer des modèles radiomiques hybrides d'apprentissage automatique capables de classer glioblastomes (GBM) et métastases cérébrales (MC) en imagerie par résonance magnétique (IRM).

Matériel et Méthodes : L'étude rétrospective incluait les sujets souffrant d'un GBM ou de MC avec preuve histologique, explorés en IRM pré-thérapeutique. Après segmentation semi-automatique, 642 index radiomiques étaient extraits par lésion en séquences T1 injectée et T2 FLAIR dans trois régions : rehaussement T1, hypersignal Flair péri-rehaussement, et lésion entière. Sept classifieurs étaient entraînés sur cinq sélections d'index radiomiques, l'âge, le sexe, la description IRM, et un score de probabilité issu d'un réseau de neurones convolutifs (CNN), avec validation croisée à 5 blocs et 5 répétitions. L'AUC-ROC était calculée sur l'échantillon test.

Résultats : 164 sujets (94 GBM et 70 MC) étaient analysés et 185 lésions mesurées (102 GBM et 83 MC). 31 modèles furent entraînés sur 127 lésions (68,6%, 70 GBM et 57 MC) puis testés sur 58 lésions (31,4%, 32 GBM et 26 MC). Le modèle hybride RFE11_SVM (*Recursive Feature Elimination, Support Vector Machine*) avec 11 index radiomiques présentait les meilleures performances : $AUC_{test}=0,928$ $IC_{95\%}[0,853;1]$, exactitude=91,4% $IC_{95\%}[84,2\% ; 98,6\%]$, sensibilité (GBM)=96,9% $IC_{95\%}[92,4\% ; 100\%]$, spécificité=84,6% $IC_{95\%}[75,3\% ; 93,9\%]$. Isolément, les 11 index sélectionnés avaient un $AUC_{test}=0,822$ $IC_{95\%}[0,715 ; 0,929]$ et le score CNN un $AUC_{test}=0,782$ $IC_{95\%}[0,655 ; 0,910]$ par SVM. Trois index morphologiques rapportaient la sphéricité des MC et 8 index de texture l'hétérogénéité des GBM.

Conclusion : Un modèle radiomique hybride augmenté par CNN différencie efficacement un GBM d'une MC.

Mots clefs : glioblastome ; métastase cérébrale ; radiomique ; apprentissage automatique ; IRM ; texture ; intelligence artificielle ; radiologie.