



HAL
open science

Analyse psychophysique et apprentissage statistique pour l'aide au diagnostic en imagerie du cancer

Carole Lartizien

► **To cite this version:**

Carole Lartizien. Analyse psychophysique et apprentissage statistique pour l'aide au diagnostic en imagerie du cancer. Imagerie médicale. INSA de LYON; Université Claude Bernard Lyon 1, 2015. tel-02062047

HAL Id: tel-02062047

<https://hal.science/tel-02062047v1>

Submitted on 12 Mar 2019

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.

HABILITATION A DIRIGER DES RECHERCHES

présentée devant

l'Institut National des Sciences Appliquées de Lyon
et l'Université Claude Bernard LYON I

**Analyse psychophysique et apprentissage statistique pour l'aide au
diagnostic en imagerie du cancer**

SPECIALITE :

par
Carole Lartizien-Ménière

Soutenue le 18 juin 2015 devant la Commission d'examen

Bloch Isabelle	Professeur, LTCI, Paris
Comtat Claude	Ingénieur, CEA, Paris
Garcia Christophe	Professeur, LIRIS, Villeurbanne
Rousseau David	Professeur, CREATIS, Villeurbanne
Ruan Su	Professeur, LITIS, Rouen

Table des matières

A. Parcours professionnel	1
A.1 Etat Civil.....	1
A.2 Diplômes.....	1
A.3 Emplois occupés	2
A.4 Expériences d'enseignement	2
A.5 Responsabilités d'encadrement	3
A.5.1 Post-doctorants.....	3
A.5.2 Ingénieurs.....	4
A.5.3 Doctorants	4
A.5.4 Stagiaires.....	5
A.6 Participation à des contrats de recherche	6
A.6.1 Projets de recherche financés	7
A.6.2 Contrats industriels.....	9
A.7 Animation et management de la recherche.....	10
A.8 Production scientifique	10
B. Synthèse des travaux de recherche	11
B.1 Introduction	11
B.2 Base de données pour l'évaluation en imagerie médicale	14
B.2.1 Quelle vérité terrain ?	14
B.2.2 Simulation Monte Carlo pour l'imagerie TEP	14
B.2.3 Simulation d'images TDM.....	20
B.2.4 Limitations de la simulation et alternatives	21
B.3 Analyse psychophysique en imagerie nucléaire par émission de positons	26
B.3.1 Contexte	26
B.3.2 Etat de l'art à la fin des années 90	26
B.3.3 Evaluation des performances diagnostiques en imagerie TEP corps entier	28
B.3.4 Optimisation des protocoles d'imagerie TEP pour la détection des lésions oncologiques 33	
B.3.5 Les observateurs algorithmiques comme alternative au lecteur humain	40
B.4 Diagnostic assisté par ordinateur pour l'imagerie du cancer	50
B.4.1 Contexte scientifique	50
B.4.2 Cadre méthodologique des développements réalisés	50
B.4.3 Contributions dans le domaine des systèmes CADx d'aide au diagnostic	55
B.4.4 Contributions dans le domaine des systèmes CADe d'aide à la détection	72

B.5	Apprentissage statistique en imagerie médicale	84
B.5.1	Extension des SVM au cas des étiquettes incertaines.....	84
B.6	Conclusion et perspectives.....	92
B.6.1	Bilan scientifique.....	92
B.6.2	Encadrement et administration de la recherche	94
B.6.3	Perspectives de recherche	95
B.7	Annexes.....	104
B.7.1	Notations et rappel de définitions	104
B.7.2	Théorie de la décision, analyse ROC et dérivées	107
B.7.3	Les observateurs numériques	115
B.7.4	Apprentissage supervisé	121
B.8	Références bibliographiques	132
B.8.1	Bibliographie	133
B.8.2	Liste de mes publications	143
B.8.3	Quelques papiers.....	152

Remerciements

Préambule

Ma motivation pour la recherche en imagerie médicale est née à l'occasion d'un stage que j'ai effectué à l'Argonne National Laboratory de Chicago aux Etats-Unis à la fin de ma formation d'ingénieur. J'ai été, à cette occasion, impliquée dans l'étude de nouveaux détecteurs pour les applications médicales en imagerie nucléaire par émission de positons (TEP) et commencé à travailler dans le milieu très interdisciplinaire de la recherche en santé. De retour en France, j'ai choisi de poursuivre mon cursus d'ingénieur par un DEA en génie biologique et médical qui m'a permis d'effectuer un premier stage dans le domaine de l'imagerie tomographique par rayonnement synchrotron à l'ESRF de Grenoble. J'ai ensuite rejoint le service hospitalier Frédéric Joliot du CEA à Orsay pour mener un projet de thèse dans le domaine de l'imagerie TEP oncologique qui était en pleine émergence à la fin des années 1990. J'ai poursuivi par un post-doctorat à l'Université de Pittsburgh aux Etats-Unis, au sein de l'équipe de David Townsend et Paul Kinahan, alors pionnière dans le domaine de l'imagerie multi-modalités avec la conception du premier système d'imagerie couplée TEP/TDM. De retour en France, sur la plate-forme d'imagerie du petit animal du CERMEP, j'ai approfondi mes connaissances sur les différentes modalités d'imagerie et précisé les contours du projet de recherche que je souhaitais développer. Ce projet m'a permis d'intégrer le CNRS comme chargée de recherche en 2004.

L'ensemble de mes travaux de recherche au cours de ces années s'est structuré autour de l'imagerie diagnostique du cancer.

Les recherches que j'ai menées dans le cadre de mon doctorat puis de mes post-doctorats étaient motivées par la volonté d'améliorer les performances diagnostiques intrinsèques de l'imagerie TEP. Ma contribution principale dans ce domaine a été de proposer un cadre théorique basé sur l'analyse psychophysique et la théorie de la décision statistique permettant d'évaluer quantitativement l'impact des paramètres de formation de l'image sur la qualité diagnostique. Dans ces travaux, la tâche diagnostique concernait la détection et la caractériser *in vivo* des pathologies cancéreuses.

Depuis 2006, mes travaux se situent s'inscrivent dans le domaine des systèmes d'aide au diagnostic (CAD en anglais pour « computer aided diagnosis ») basés sur des méthodes de classification par apprentissage. Mon ambition est de proposer des méthodes d'analyse statistique de l'image permettant d'extraire l'information pertinente pour le diagnostic. Ces recherches ciblent plus particulièrement les verrous méthodologiques liés à la fusion d'informations hétérogènes (multimodales, multiparamétriques, multidimensionnelles).

Ce mémoire présente la synthèse des travaux de recherche que j'ai menés depuis près de 15 ans d'abord dans le domaine de l'analyse psychophysique puis de l'apprentissage statistique pour l'aide au diagnostic en imagerie du cancer.

Le premier chapitre présente une photographie de mes activités d'enseignement, d'encadrement et de valorisation de la recherche au travers de collaborations scientifiques et montage de projets.

Le cœur de ce manuscrit est ensuite consacré à la synthèse de mes activités de recherche. Une première partie retrace les grandes lignes de mon travail dans le domaine de l'analyse psychophysique. Je ne reviens pas, dans cette présentation, sur la discussion précise des résultats mais sur l'impact de ces travaux dans la communauté scientifique au cours des années 2010. La seconde partie de ce chapitre décrit les travaux plus récents que j'ai réalisés dans le domaine des systèmes d'aide au diagnostic pour l'imagerie du cancer. Je présente ensuite une étude plus récente dans le domaine de l'apprentissage statistique. Cette partie illustre mes centres d'intérêt scientifiques actuels. Elle propose également de les mettre en perspective à la fois avec mes travaux plus anciens et mes projets à venir afin de souligner la continuité de ma démarche scientifique. Mon travail de chercheur ciblé sur une activité scientifique individuelle au cours de la thèse s'est progressivement enrichi par le travail de formation par la recherche que je réalise depuis une dizaine d'années en encadrant des étudiants en thèse, ou en master. Je précise dans cette synthèse la nature de ma contribution, soit comme investigateur principal soit comme superviseur et propose une analyse critique de ces résultats et complémentaire aux discussions présentées dans les articles.

Enfin, la dernière partie contient quelques annexes rappelant des notions de base dans le domaine de l'évaluation en imagerie médicale et de la théorie statistique de la décision. Une sélection d'articles discutés dans le mémoire est également jointe.

A. Parcours professionnel

A.1 Etat Civil

Nom, Prénoms	LARTIZIEN-MENIERE, Carole
Nationalité	Française
Age et situation familiale	41 ans, mariée, 2 enfants
Laboratoire	CREATIS, UMR CNRS 5220, U1044 Inserm INSA-Lyon, Université Lyon 1
Equipe de recherche	Images et Modèles
Adresse professionnelle	INSA- 7 rue Jean Capelle bat. Blaise Pascal, 4ème étage F-69621 Villeurbanne cedex, France
Téléphone professionnel	04 72 43 81 48 (laboratoire)
Email	carole.lartizien@creatis.insa-lyon.fr
web	www.creatis.insa-lyon.fr/site/~lartizien
Statut actuel	Chargée de Recherche CNRS, classe 1, section 7 Responsable adjointe de l'équipe 'images et modèles'
Domaine de recherche	Analyse d'images médicales multi-modalités: analyse psychophysique, apprentissage supervisé. Application à l'aide au diagnostic en imagerie du cancer et en neuroimagerie.

A.2 Diplômes

Mars 2001 **Thèse de Sciences de l'Université Paris XI**

Spécialité Imagerie Médicale

Mention Très Honorable avec les Félicitations du Jury

Titre Optimisation des protocoles d'imagerie corps entier en tomographie par émission de positons pour la détection des hyperfixations du 18F-FDG en oncologie.

Laboratoire Service Hospitalier Frédéric Joliot - CEA, Orsay

Jury

A. Syrota	Président, Professeur, SHFJ, Orsay
I. Buvat	Rapporteur, CR1 CNRS, INSERM U 494, Paris
C. Michel	Rapporteur, Professeur, Université de Louvain-la-Neuve, Belgique
D. Le Guludec	Examineur, Professeur, Service de Médecine Nucléaire de l'hôpital Bichat, Paris
R. Trébossen	Examineur, Ingénieur CEA, SHFJ, Orsay
K. Blohm	Membre invité, Ingénieur, chef de projet Siemens France
R. Di Paola	Directeur de thèse, DR INSERM, INSERM U494, Paris

1997 **DEA de Génie Biomédical** de l'Université Joseph Fourier, Grenoble

Spécialité Imagerie et Rayonnements

Mention Bien

Titre Mise en œuvre d'une méthode d'imagerie en tomographie X à deux énergies appliquée à la détection et à la quantification intra-tissulaire du potassium cérébral.

Laboratoire Equipe, Rayonnement Synchrotron et Recherche Médicale de l'Université Joseph Fourier

Direction Jean-François Le Bas (PuPh, CHU Grenoble)

1996 **Diplôme d'ingénieur de l'École Nationale Supérieure de Physique de Grenoble (ENSPG)** rattachée à L'INPG, Grenoble

Spécialité Instrumentation Nucléaire

Mention Bien

A.3 Emplois occupés

Octobre 2008 – Présent :

Chargée de Recherche CNRS, classe 1, section 7, équipe 'Images et Modèles' du laboratoire CREATIS

Octobre 2004 – Octobre 2008 :

Chargée de Recherche CNRS, classe 2, section 7, équipe 'Imagerie Volumique' du laboratoire CREATIS

Janvier 2002-Septembre 2004:

Post-doctorat au Centre d'Exploration et de Recherche Médicale par Emission de Positons (CERMEP), Lyon. Financement : Bourse de la Région Rhône-Alpes puis bourse européenne (Projet EUMORPHIA FP5).

Mai-Décembre 2001:

Post-doctorat au centre d'imagerie TEP de l'Université de Pittsburgh (University of Pittsburgh Medical Centre), PA, USA, dirigé par le Pr. Dave Townsend et encadré par le Dr Paul Kinahan. Financement : Bourse Lavoisier complétée par une bourse américaine.

A.4 Expériences d'enseignement

J'enseigne régulièrement les bases de l'imagerie médicale nucléaire et de l'analyse d'images en 2nd et 3^{eme} cycle universitaire. Je démarre cette année un enseignement en classification supervisée. La description plus précise de ces enseignements est donnée dans le tableau ci-dessous.

DISCIPLINE	ANNEE	NIVEAU	ETABLISSEMENT	HEURES
Classification supervisée	2014-15	3 ^{ème} cycle : Master II GEGP, parcours : Systèmes et Images	INSA Lyon	CM : 4h/an
Instrumentation et analyse d'images médicales TEP	2014-2015	3 ^{ème} cycle : Master II Ingénierie pour la Santé et le Médicament	Université Claude Bernard Lyon	CM : 9h/an TD : 3h/an
Instrumentation et analyse d'images médicales TEP	Depuis 2009	2 ^{ème} cycle : 5 ^{ème} année Département Génie Electrique	INSA Lyon	CM : entre 4 et 6h/an TD : 2h/an
Instrumentation et analyse d'images médicales TEP	2010 - 2012	3 ^{ème} cycle : Master II Ingénierie pour la Santé et le Médicament, spécialité Physique Médicale	Université Claude Bernard Lyon	CM: 15h/an TD: 2h/an
Instrumentation et analyse d'images médicales TEP	2006-2011	2 ^{ème} cycle : 3 ^{ème} année Option : Sciences Biologiques et Médicales. -	Faculté de Médecine Lyon Nord	CM : 4h/an

J'ai également été invitée à intervenir dans plusieurs cycles de formation. Le détail de ces enseignements ponctuels est donné dans la rubrique « séminaires invités » de ma liste de publications.

A.5 Responsabilités d'encadrement

A.5.1 Post-doctorants

Thibault Frisson, 2007-2009

Financement : projet ANR fGATE (www.fgate.fr). Voir détails dans la section A.6.1.3.

Mission : Développer des modules d'accélération des calculs pour la plateforme de simulation GATE

Statut actuel : Chercheur au CERN

Emilie Niaf, 2012-2013

Financement : projet INCa PAIR Prostate. Voir détails dans la section A.6.1.3.

Mission : Proposer un système d'aide à la détection et diagnostic du cancer de la prostate par fusion de l'information IRM et échographique.

Statut actuel : post-doctorante.

A.5.2 Ingénieurs

Fabrice Prost-Boucle, depuis octobre 2014

Financement : projet INCa LYRIC. Voir détails dans la section A.6.1.3.

Mission : Proposer un système d'aide à la détection et diagnostic du cancer de la prostate par fusion de l'information IRM et échographique

A.5.3 Doctorants

J'ai co-encadré 5 thèses depuis 2006 avec un pourcentage de co-direction supérieur à 75%.

Loredana Baboi, 2004-2007

Etablissement d'inscription : UCBL

Directeur de thèse : Olivier Beuf, CR1 CNRS

Contribution personnelle à l'encadrement : 10%

Sujet de thèse : Méthodologie pour l'IRM abdominale du petit animal à champ magnétique élevé : Application à la caractérisation de lésions hépatiques sur un modèle murin de tumeurs endocrines.

Financement : MESR

Publications associées : [Baboi, MRI 2009], [Baboi, Biomed Imaging Interv J 2007], [Beuf, Gastroenterol Clin Biol 2008], [Lartizien, ISMRM 2009]

Statut actuel : Attachée de Recherche Clinique.

Sandrine Toméï, 2006-2009

Etablissement d'inscription : INSA de Lyon

Directeur de thèse : Christophe Odet, Pr INSA

Contribution personnelle à l'encadrement : 90%

Sujet de thèse : Détection automatique des foyers d'activités dans les images de tomographie par émission de positons (TEP)

Financement : MESR

Publications associées : [Tomei, IEEE TNS 2010], [Tomei, IEEE MIC, 2008], [Lartizien, Int J Comp Assist Radiol Surg, 2009]

Statut actuel : Gérante de société.

Simon Marache-Francisco, 2008-2011

Etablissement d'inscription : INSA de Lyon

Directeur de thèse : Rémy Prost, Pr INSA

Contribution personnelle à l'encadrement : 90%

Sujet de thèse : Apport de la correction du mouvement respiratoire sur la détection en imagerie TEP/TDM oncologique

Financement : CIFRE avec la société Medisys – Philips Healthcare

Publications associées : [Lemaitre, IEEE Proc 2009], [Tomei, IEEE MIC, 2008], [Lartzien, IEEE MIC 2009], [Marache, IEEE MIC 2009], [Marache, IEEE MIC, 2010], [Lartzien, IEEE MIC 2011].

Statut actuel : Ingénieur RetD en informatique dans le secteur privé.

Emilie Niaf, 2009-2012

Etablissement d'inscription : UCBL

Directeur de thèse : Olivier Rouvière, PUPH

Contribution personnelle à l'encadrement : 80%

Sujet de thèse : Aide au diagnostic du cancer de la prostate par IRM multi-séquences : une approche par classification supervisée.

Financement : MESR

Publications associées : [Niaf, SPIE conference, 2011], [Niaf, IEEE statistical signal processing workshop, 2011], [Niaf, PMB, 2012], [Niaf, IEEE TIP, 2014], [Niaf, Radiology, 2014], [Niaf, ICIP, 2014].

Prix dans la catégorie des travaux de jeunes chercheurs à la conférence SPIE de 2011, Prix des amis de l'Université de Lyon en 2013.

Statut actuel : post-doctorante

Jérôme Lehaire, 2012-2015

Etablissement d'inscription : UCBL

Directeur de thèse : Olivier Rouvière, PUPH

Contribution personnelle à l'encadrement : 90%

Sujet de thèse : Cartographie du cancer de la prostate par IRM multi-séquences.

Financement : MESR

Publications associées : [Lehaire, ESMRBM, 2013], [Lehaire, ICIP, 2014].

Meriem El Azami, 2013-2016

Etablissement d'inscription : INSA Lyon

Directeur de thèse : Denis Friboulet, Pr INSA

Contribution personnelle à l'encadrement : 90%

Sujet de thèse : Cartographie du cancer de la prostate par IRM multi-séquences.

Financement : Bourse de thèse du LabeX PRIMES (<http://primes.universite-lyon.fr/>)

Publications associées : [El Azami, IEEE PRNI, 2013], [El Azami, ESANN, 2014].

A.5.4 Stagiaires

Depuis janvier 2006, j'ai encadré 12 stagiaires de dernière année d'école ingénieur et/ou de Master Recherche:

Ludovic Potier, 2006, 5ème de l'INSA de Lyon et Master Recherche 'Informatique, Graphique et Images' (Lyon). Détection en imagerie TEP oncologique. Approche basée sur la théorie de la détection du signal.

Thiago Magalhaes, 2007, 5ème année de l'IUT de Compiègne. Développements méthodologiques pour la caractérisation par imagerie TEP et IRM d'un modèle de tumeurs neuro-endocrines chez la souris.

Simon Marache-Francisco, 2008, 5ème de l'INSA de Rouen et Master Recherche 'Informatique, Génie de l'Information et des Systèmes' (Rouen). Détection de tumeurs par SVM sur des images de Tomographie par Emission de Positons.

Fabien Balezeau, 2008, 3ème année de l'Ecole Nationale Supérieure de Physique de Strasbourg et Master Recherche 'Imagerie, Robotique Médicale et Chirurgicale' (Strasbourg). Développements méthodologiques pour la fusion d'images IRM et TEP du petit animal dans le cadre de diagnostic en oncologie digestive.

Amandine Lemaitre, 2008, ENST Bretagne, Master Signal, Télécommunications Image (STI) de l'université de Rennes 1, Co-encadrement entre les laboratoires LaTIM (Brest) et CREATIS (Lyon). Développement d'un modèle de simulation réaliste représentant la distribution d'activité chez un patient.

Laurent Leduvehat, 2010, Projet de fin d'étude (PFE) de l'ENSEIRB-MATMECA de Bordeaux (<http://www.enseirb-matmeca.fr/>), filière informatique de l'option 'Technologies multimedia. Co-encadrement avec Simon Marache-Francisco, étudiant en 2ème année de thèse sur le sujet : « Développement d'un modèle de simulation réaliste d'images TDM médicales avec SINDBAD ». Durée : 6 mois

Obioma Udobata, 2010, Master international IMESI (Embedded Systems & Medical Image Engineering de l'INSA de Lyon (<http://www.insa-imasters.com/>) sur le sujet : « System of fusion based on Belief theory ». Durée: 3 mois.

Fabien Ricard, 2011, Assistant chef clinique et Master II Ingénierie pour la santé et le médicament (ISM), filière Ingénierie Biomédicale et Pharmaceutique, Parcours Instrumentation et Imagerie Médicale. Stage de Master II sur le sujet « Systèmes CAD pour la TEP-TDM et apport des caractéristiques de texture » Durée : 6 mois à temps partiel.

Matthieu Rogez, 2011, Projet de fin d'étude (PFE) de SUPELEC, Metz, Majeure : Systèmes Interactifs et Robotique sur le sujet : « Développement d'un système d'aide au diagnostic pour l'imagerie TEP/TDM du cancer ». Durée : 5 mois.

Andrei Teodorescu, 2011, 5ème de l'INSA de Lyon, filière Informatique. Co-encadrement avec Emilie Niaf sur le sujet « Implémentation en C/C++/Python d'une librairie logicielle pour la classification d'images médicales : application à la détection du cancer de la prostate en IRM ». Durée : 3 mois.

Adeline Susset, 2012, Master II Traitement des Signaux et des Images, Université de Bordeaux I, sur le sujet « Cartographie du lymphome en imagerie TEP/TDM ». Durée : 6 mois.

Meriem El Azami, 2013, Projet de fin d'étude (PFE) de PHELMA, Grenoble, Filière : Signal Images, Communication, Multimedia (SICOM) Systèmes Interactifs et Robotique sur le sujet : « conception et l'évaluation d'un système d'aide au diagnostic (CAD) pour l'imagerie IRM multi-paramètres de l'épilepsie pharmacorésistante ». Durée : 6 mois.

A.6 Participation à des contrats de recherche

- Projet d'établissement : 2 projets BQR de l'INSA de Lyon
- Projet régional : 1 projet de la Région Rhône-Alpes
- Projet national : 2 projets ANR (programmes CIS 206 et COSINUS 2009) et 2 projets financés par l'INCa, 1 programme du CNRS en 2006.
- Projet international : 1 projet de réseau d'excellence européen

A.6.1 Projets de recherche financés

La liste présentée ci-dessous concerne les projets financés depuis mon recrutement comme CR2 en octobre 2004.

A.6.1.1 Projets d'établissement

Subvention BQR-INSA de Lyon

Octobre 2006- Octobre 2008

Objet : Apport de la tomographie par émission de positons (TEP) pour la détection et la caractérisation de métastases hépatiques sur un modèle de tumeurs neuro-endocrines chez la souris (suite du programme IPA)

Rôle : Porteur du projet

Partenaires : CREATIS : porteur, LRMN.

Montant alloué : 20 keuros.

Retombées : 1 communication en congrès [Lartizien, ISMRM, 2009]

Subvention BQR-INSA de Lyon

Janvier 2014- Décembre 2015

Objet : Nouvelles applications des cellules solaires en bio-imagerie et biométrie. Application à la détection de marqueurs du cancer dans des échantillons biologiques.

Rôle : responsable d'une action de recherche

Partenaires : INL (Lyon) : porteur, CarMeN (Lyon), CREATIS (Lyon).

Montant alloué : 8 keuros.

A.6.1.2 Projets régionaux

Projet SIMED Simulation en Imagerie Médicale pour le Diagnostic et la Thérapie

2006-2010

Objet : Développer, fédérer et promouvoir une recherche commune à l'échelle de la Région Rhône-Alpes autour de la simulation numérique en imagerie médicale et de ses applications pour l'aide au diagnostic et la thérapie. <http://www.creatis.insa-lyon.fr/Simed>

Rôle : participant

Partenaires : CREATIS : porteur, CEA-LETI (Grenoble) : co-porteur, IMTh (Lyon), LIRIS (Lyon), TIMC (Grenoble), ICJ (Lyon), GIPSA (Grenoble)

Montant total alloué: 196 keuros, 5 allocations doctorales.

A.6.1.3 Nationaux

Programme IPA du CNRS-CEA

Octobre 2004- Octobre 2006

Objet : Imagerie multimodale pour la détection et la caractérisation de métastases hépatiques sur un modèle de tumeurs neuro-endocrine chez la souris.

Rôle : participant

Partenaires : CREATIS, LRMN : porteur.

Montant alloué : 6 keuros (budget global : 15 keuros)

Retombées : plusieurs publications : [Baboi, MRI 2009], [Baboi, Biomed Imaging Interv J 2007], [Beuf, Gastroenterol Clin Biol 2008], [Lartizien, ISMRM 2009]

Contrat de recherche ANR- Programme Calcul Intensif et Simulations (CIS) 2006 – Projet fGATE

Janvier 2007- Décembre 2009

Objet : Simulation efficace d'examen précliniques et cliniques en tomographie d'émission. Il s'agissait de consolider le code de la plate-forme de simulation GATE, d'accélérer les temps de calcul et démontrer la faisabilité de générer des images TEP précliniques et cliniques réalistes.

Rôle : Responsable d'une action de recherche

Partenaires : LIF (Orsay) : porteur, LaTIM (Brest), CPPM (Marseille), CREATIS (Lyon) , SHFJ (Orsay), HealthGrid

Montant alloué : 50 keuros (budget global : 578 keuros)

Retombées : les aspects applicatifs qui me concernaient ont été restreints dans le cadre d'une réorganisation du projet visant à consacrer davantage de temps à la consolidation du code.

Contrat de recherche ANR- Programme Cosinus 2009 – Projet VIP Virtual Imaging Platform

Janvier 2010- Décembre 2012

Objet : Développement d'une plate-forme d'imagerie médicale virtuelle. L'objectif est d'intégrer un ensemble de simulateurs d'images médicales (TEP, IRM, US et CT) dans une interface graphique permettant un partage de modèles et l'accès à des ressources de calcul distribué par l'intermédiaire de 'workflow'

Rôle : Participant ; Expert pour le groupe de travail 'sémantique' visant à définir une ontologie adaptée à la simulation d'images médicale et pour la définition des workflows de simulation TEP. Evaluation des fonctionnalités de la plate-forme par la mise en œuvre de quelques applications.

Partenaires : CREATIS (Lyon) : porteur, CEA LETI (Grenoble), I3S (Nice), Visages (Rennes), Société MAAT-G.

Site web : <http://www.creatis.insa-lyon.fr/vip/>

Montant alloué : 271 keuros pour CREATIS sur un budget global de 975 keuros.

Retombées : Plusieurs communications en congrès [Glatard, IEEE ISBI 2012], [Marion, CBMS 2011], deux publications [Glatard, IEEE TMI, 2013], [Gibaud, JBI, 2014]. Ouverture de la plate-forme VIP aux utilisateurs fin 2012.

Contrat de recherche INCa- Programme d'actions intégrées de recherche (PAIR) prostate 2009 – Projet Cartographix

Janvier 2011-Mai 2014

Objet : Evaluation de la position, du volume et de l'agressivité des foyers de cancer de prostate par imagerie multiparamétrique de la prostate. Soutien financier au travail de thèse d'Emilie Niaf et financement d'une année complémentaire de post-doctorat.

Rôle : Responsable d'un groupe de travail sur la conception d'un système d'aide au diagnostic

Partenaires : Labtau (Lyon) : porteur, CREATIS (Lyon), INSERM U703 (Lille), Société SuperSonic Imagine (Aix-en-provence), Société Bracco.

Montant alloué : 72.4 k€ pour CREATIS sur un budget global de 484.7 k€

Retombées : Plusieurs publications en congrès, une publication dans *Physics in Medicine and Biology* [Niaf, PMB 2012] et dans *Radiology* [Niaf, Radiology 2014]

Contrat de recherche INCa- Programme SIRIC (Sites de Recherche Intégrée sur le Cancer) - projet LYRIC

Janvier 2012-Décembre 2017

Objet : ce projet propose 3 programmes de recherche dont le fil conducteur est la progression dans le développement des thérapies ciblées. Le programme 3 concerne le 'Ciblage des tumeurs macroscopiques in vivo : traitement local assisté par imagerie'

Rôle : Participant. Je suis impliquée dans le 3ème programme de recherche dans le cadre du projet de cartographie 3D du cancer de la prostate intégrant des informations extraites des images IRM et US. Recrutement en encadrement d'un ingénieur. Soutien financier au travail de thèse de Jérôme Lehaire.

Partenaires : GCS (groupement de coopération sanitaire) Lyon Cancérologie Universitaire qui associe le Centre Léon Bérard et les Hospices civils de Lyon : porteur, Fondation Synergie Lyon, Centre International de Recherche sur le Cancer (Lyon), Université Claude Bernard, Lyon I.

Montant alloué : 1.3 M€ pour le programme 3 sur un budget global de 9 M€

Contrat de recherche ANR- Programme EQUIPEX de la vague 2 du programme d'investissement d'avenir (PIA) 2011– Projet LILI (Lyon integrated life imaging)

Juin 2012-Décembre 2019

Objet : Installation du premier imageur hybride TEP/IRM. Développement méthodologique et projets applicatifs dans les domaines de la neurologie, oncologie, cardiologie et nutrition et métabolisme.

Partenaires : CRNL (Lyon) : porteur, CERMEP (Lyon), CarMeN (Lyon), CREATIS (Lyon), Hospices Civils de Lyon, Siemens Healthcare

Montant alloué : 4 Meuros + 5 années de post-doctorats

Implication : Mise en place de l'équipement, et suivi scientifique, financier, et opérationnel de LILI. Actions d'information, de formation, et de communication. Recrutement de post-doctorants et ingénieurs. Expertise scientifique en traitement de l'image multimodale.

A.6.1.4 Projets internationaux

Réseau d'excellence européen NoE VPH – Programme ICT du FP 7

Juin 2008-Juin 2011

Objet : Modélisation *in silico* pour le vivant.

Rôle : Participant.

Partenaires : 13 partenaires européens. Projet piloté par le CNRS pour la France.

Montant alloué : 188 keuros/8 Meuros

Retombées : 1 présentation orale au congrès VPH2010

A.6.2 Contrats industriels

Contrat de collaboration avec la Société Philips Healthcare Medisys

Novembre 2008-Février 2011

Objet : Soutien financier à la thèse CIFRE de Simon Marache sur l'évaluation de l'influence du mouvement respiratoire en TEP sur la détection des lésions pulmonaires et hépatiques

Montant alloué : 30 keuros

Rôle : responsable scientifique

Retombées : Plusieurs communications en congrès [Marache IEEE MIC 2010], [Marache IEEE MIC 2009], un article [Lemaitre IEEE Proc 2009]

A.7 Animation et management de la recherche

- Responsable adjointe de l'équipe 'Images et Modèles' de CREATIS depuis 2013.
- Membre des comités de pilotage de structures nationales financées dans le cadre des programmes d'investissement d'avenir (PIA) :
 - groupe de travail WP4 'analyse d'images' de l'infrastructure d'excellence FLI (France Life Imaging) depuis 2013. Participation aux comités de pilotage, organisation de séminaires scientifiques, appels d'offres interne au groupe de travail
 - EQUIPEX LILI (Lyon integrated Life Imaging) depuis 2012 concernant l'installation d'un imageur hybride TEP/IRM simultané à Lyon. Participation aux comités de pilotage, à l'appel d'offre et installation de la machine, expert pour la rédaction de projets scientifique autour de cette machine.
 - Membre des sociétés savantes : SFGBM, IEEE.
- Activité éditoriale pour des revues scientifiques de rang A et des colloques internationaux
- Expertise scientifique auprès de l'ANR et l'ANRT

A.8 Production scientifique

- 30 publications dans des revues scientifiques avec comité de lecture depuis 2002.
- 58 communications dans des colloques avec actes
- H-index= 16 (source : ISI web of knowledge dec 2014)

-Sélection de 5 publications récentes :

C. Lartzien, M. Rogez, E. Niaf, and F. Ricard. Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. IEEE Journal of Biomedical and Health informatics, 18(3) : 946-955, 2014.

E. Niaf, R. Flamary, O. Rouvière, C. Lartzien, and S. Canu. Kernel-based learning from both qualitative and quantitative labels: Application to prostate cancer diagnosis based on multiparametric MR imaging. IEEE Transactions on Image Processing, 23(3) 979- 991, 2014.

E. Niaf, O. Rouvière, F. Mege-Lechevallier, F. Bratan, and C. Lartzien. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Physics in Medicine and Biology, 57(12): 3833-3851, 2012.

C. Lartzien, S. Marache-Francisco, R. Prost. Automatic detection of lung and liver lesions in 3D positron emission tomography images: a pilot study. IEEE Transactions on Nuclear Science, 59(1): 102-112, 2012.

T. Glatard, C. Lartzien, B. Gibaud, R. Ferreira da Silva, G. Forestier, F. Cervenansky, M. Alessandrini, H. Benoit-Cattin, O. Bernard, S. Camarasu-Pop, N. Cerezo, P. Clarysse, A. Gaignard, P. Hugonnard, H. Liebgott, S. Marache, A. Marion, J. Montagnat, J. Tabary, and D. Friboulet. A Virtual Imaging Platform for multi-modality medical image simulation. IEEE Transactions on Medical Imaging, 32(1): 110-118, 2013.

B. Synthèse des travaux de recherche

B.1 Introduction

Ce rapport présente la synthèse des travaux de recherche que j'ai menés depuis près de 15 ans dans le domaine de l'analyse psychophysique et de l'apprentissage statistique pour l'aide au diagnostic en imagerie du cancer.

Sur le plan méthodologique, j'ai décliné la notion de performance diagnostique à différents niveaux d'utilisation de l'image médicale, soit au cours de sa formation, soit dans son utilisation finale pour l'interprétation réalisée par le médecin.

Les recherches que j'ai menées dans le cadre de mon doctorat puis de mes post-doctorats ont été motivées par la volonté d'améliorer la qualité diagnostique intrinsèque de l'image, c'est-à-dire à maximiser l'information disponible dans l'image et exploitable par le médecin pour détecter et caractériser *in vivo* les pathologies cancéreuses. Cette question m'a conduite à apporter une contribution scientifique originale dans le domaine de l'évaluation quantitative des performances diagnostiques et de l'analyse psychophysique en imagerie nucléaire TEP.

Le questionnement récurrent au cours de ces travaux sur la notion de performance diagnostique et d'interaction forte entre le contenu de l'image et l'observateur humain m'a ensuite conduite à m'intéresser à la notion de systèmes d'aide au diagnostic (CAD en anglais pour « computer aided diagnosis »). La motivation scientifique qui guide mes travaux dans ce domaine depuis 2006 est de proposer des stratégies permettant d'exploiter au mieux l'information contenue dans l'image afin d'aider le clinicien dans sa tâche diagnostique. Cette activité se décline dans le cadre formel des méthodes par apprentissage et s'intéresse plus particulièrement aux verrous liés à la fusion d'informations hétérogènes (multimodales, multiparamétriques, multidimensionnelles) et à la cartographie 3D des zones suspectes dans l'image. Les caractéristiques propres de l'imagerie médicale (vérité terrain, taille des échantillons etc) stimulent également des recherches plus théoriques dans le domaine de l'apprentissage statistique.

Je me suis d'abord intéressée à l'imagerie nucléaire, plus particulièrement la tomographie par émission de positons (TEP) en raison des challenges à relever par cette modalité dans le domaine de la santé à la fin des années 1990. La place croissante de la multi-modalité en imagerie du cancer m'a ensuite naturellement orientée à élargir mon champ d'applications à d'autres modalités, en particulier l'IRM.

L'une des originalités de ma démarche méthodologique au cours de ces années a été de baser ces études sur des images virtuelles. Ce type d'image permet d'accéder à la 'vérité terrain', dans notre cas, la localisation des lésions cancéreuses, et de fournir ainsi une évaluation quantitative fiable des performances. J'ai ainsi consacré une part importante de mon activité jusqu'en 2011 à la création de bases d'images simulées les plus réalistes possibles et adaptées aux contraintes statistiques des analyses psychophysiques.

Mon travail de chercheur ciblé sur une activité scientifique individuelle au cours de la thèse s'est progressivement enrichi par le travail de formation par la recherche que je réalise depuis une dizaine d'années auprès d'étudiants en thèse ou master. La recherche de collaborations scientifiques et le montage de projets constituent également une part importante et cruciale de mon activité. J'ai développé des collaborations avec plusieurs chercheurs du laboratoire au sein de mon équipe. J'entretiens également des collaborations avec des partenaires des domaines de la clinique, des statistiques ou des mathématiques appliquées, au niveau national et international et dans le cadre de projets contractualisés ou non. Ces différentes collaborations sont précisées dans l'exposé de mes travaux de recherche.

Dans la suite de cette section sont présentées les contributions principales que j'ai réalisées depuis 15 ans, en commençant par la constitution de bases de données pour l'évaluation. Je retrace ensuite les grandes lignes de mon travail de thèse et de post-doctorat aux Etats-Unis dans le domaine de l'évaluation en imagerie médicale appliquée à l'optimisation des protocoles d'imagerie TEP corps entier pour l'oncologie. La troisième partie de ce chapitre décrit les travaux plus récents que j'ai

réalisés dans le domaine des systèmes d'aide au diagnostic pour l'imagerie du cancer ainsi qu'une étude dans le domaine plus théorique de l'apprentissage statistique. Cette dernière partie illustre mes centres d'intérêt scientifiques actuels. Elle propose également de les mettre en perspective à la fois avec mes travaux plus anciens et mes projets à venir afin de souligner la continuité de ma démarche scientifique

Base de données pour l'évaluation en imagerie médicale

B.2 Base de données pour l'évaluation en imagerie médicale

B.2.1 Quelle vérité terrain ?

La qualité des études, dans le domaine de l'évaluation des performances diagnostiques, dépend fortement de la maîtrise des données sur lesquelles elles reposent. Il est important de disposer de données réalistes sur le plan clinique et de grande dimension en raison des contraintes statistiques de l'analyse psychophysique. Il faut enfin connaître la 'vérité terrain', c'est-à-dire la localisation de la pathologie que l'on cherche à détecter.

Les images cliniques permettent d'optimiser le réalisme. En revanche, il est souvent très difficile de maîtriser la vérité terrain et de disposer de gros volumes de données, ou d'acquisitions répétées chez le même patient pour des paramètres différents.

L'utilisation d'un fantôme physique permet de s'affranchir des contraintes cliniques et de répéter les acquisitions expérimentales en maîtrisant la vérité terrain et les paramètres d'acquisition. Cependant, les images produites sont souvent peu réalistes car les modèles anthropomorphiques les plus aboutis ne reproduisent ni la complexité anatomique, ni l'hétérogénéité de fixation physiologique normale observée chez les patients comme le soulignent les illustrations de l'article présenté dans l'annexe B.8.3.1.

L'utilisation de données simulées s'avère une alternative intéressante si l'on peut disposer d'un modèle mathématique anthropomorphique suffisamment détaillé et d'un simulateur qui reproduise avec précision le processus de formation de l'image ou modélise correctement la structure du bruit. La simulation permet, en effet, de contrôler l'ensemble des paramètres influençant la formation et la qualité de l'image et de générer de gros volumes de données. Reste la recherche du réalisme, à laquelle j'ai consacré une part importante de mon activité de recherche entre 2004 et 2012. La synthèse de ces contributions ainsi qu'une analyse critique est présentée dans la suite de ce chapitre.

B.2.2 Simulation Monte Carlo pour l'imagerie TEP

Cette thématique a constitué une composante importante de mon activité de recherche jusqu'en 2012. Mon travail a porté sur la modélisation de distributions complexes et réalistes de radioactivité chez l'homme et le petit animal. J'ai décliné ce thème de recherche en deux axes. Le premier axe consistait à développer et évaluer des protocoles de simulation de données TEP qui reproduisent de manière précise les performances de caméras TEP réelles. Le second axe concernait la création de bases de données d'images de synthèse TEP représentant la distribution du fluorodeoxyglucose marqué au fluor 18 (18F-FDG), traceur clinique de l'activité métabolique du glucose utilisé en oncologie. L'objectif était de créer une base d'images réunissant, d'une part, la variabilité anatomique et fonctionnelle caractéristique d'une population standard et d'autre part, une variabilité de cas pathologiques et normaux.

B.2.2.1 Modélisation de scanners TEP

Les recherches dans le domaine des méthodes de simulation d'images de médecine nucléaire ont connu un essor important à la fin des années 80. Cet essor s'explique d'abord par l'émergence de la TEP dans le domaine clinique, portée par les applications en oncologie. Comme nous l'avons souligné, la simulation s'impose comme une option intéressante pour produire des données utilisées, d'une part, pour la conception de nouveaux scanners, mais aussi pour l'évaluation de méthodes de traitement et d'analyse d'images. L'arrivée de l'imagerie préclinique au début des années 2000 vient confirmer le besoin d'algorithmes de simulation qui permettent de modéliser le plus précisément possible les processus impliqués dans la formation du signal TEP.

Les premiers algorithmes sont basés sur des méthodes analytiques qui consistent à calculer des projections de la distribution du radiotracer à partir des spécifications géométriques des objets émetteurs et atténuateurs constituant le fantôme simulé. Les termes de coïncidences parasites diffusées et aléatoires sont obtenus à l'aide d'un modèle paramétrique ajusté sur des données expérimentales. Ces méthodes sont intéressantes car elles sont rapides. Elles ne modélisent

cependant pas tous les phénomènes physiques qui contribuent à expliquer la texture du bruit et peuvent ainsi ne pas reproduire correctement les propriétés statistiques des images.

Une activité de recherche importante se développe parallèlement autour des méthodes de simulation de type Monte Carlo (MC). Ces méthodes semblent particulièrement bien adaptées à la simulation des processus physiques intervenant dans la formation de l'image TEP, majoritairement de nature probabilistes. Plusieurs simulateurs TEP MC étaient disponibles à la fin des années 1990 mais tous ne disposaient pas du même niveau de développement ou de validation et aucun n'offrait de performance calculatoire compatible avec la simulation de bases d'images corps entier. On peut, par exemple, citer SimSET [Harrisson et al., 1993], EIDOLON [Zaidi et al., 1998], PET-EGS [Castiglioni et al., 1999]. Une comparaison des codes de simulation en termes de réalisme, de temps de calcul et de flexibilité a été réalisée par [Buvat et al., 2005] et dans la thèse de [Reilhac, 2007].

Cet état de l'art, dressé au démarrage de ma thèse, m'a conduite à choisir le simulateur analytique ASIM développé par Claude Comtat et Paul Kinahan. Cet algorithme proposait, en effet, un modèle validé pour la simulation de données cliniques pour la caméra TEP Siemens ECAT/EXACT HR+ (CTI Siemens) installée au SHFJ d'Orsay et des temps de calcul compatibles avec la simulation de volumes d'image corps entier. Cet algorithme a ainsi été utilisé pour construire la base de données ARSENE présentée dans la section B.2.2.2(i). J'ai bénéficié pour ce travail d'une collaboration forte avec C. Comtat, alors post-doctorant au SHFJ.

A l'issue de ma thèse, je me suis à nouveau intéressée aux méthodes de simulation MC et j'ai choisi de travailler avec le simulateur PET-SORTEO (<http://sorteo.cermep.fr>) développé par Anthonin Reilhac dans le cadre d'une thèse CIFRE avec SIEMENS-France [Reilhac et al., 2004] [Reilhac, 2007]. Ce choix était motivé par le niveau de validation de ce logiciel ainsi que par les temps de calcul affichés pour la simulation d'acquisition de modèles géométrique complexes. En 2005, seul PET-SORTEO pouvait en effet simuler des images TEP corps entier en un temps raisonnable (environ 1 jour.cpu pour une acquisition corps entier).

Le concurrent direct de SORTEO au moment de ce choix était la plate-forme de simulation GATE (<http://opengatecollaboration.healthgrid.org>) développée depuis 2002 dans le cadre d'une collaboration internationale [Jan et al., 2004]. GATE se base sur la librairie GEANT4 du CERN pour la physique des particules [Agostinelli, 2003]. Il inclut la modélisation de tous les phénomènes physiques et présente une grande flexibilité pour la modélisation des sources et du détecteur. Ce projet, initié à l'EPFL de Lausanne par l'équipe de Christian Morel, a pour ambition d'aider à la conception d'un système préclinique, et s'intéresse à la simulation de sources radioactives de forme simple et de faible activité. Si la transposition directe de GEANT 4 est possible pour cette application, elle n'est cependant pas envisageable, dans sa version originale, pour la simulation de plusieurs millions de particules générées lors d'une acquisition clinique. On estime, en effet, un rapport de 3 à 4 ordres de grandeur entre les temps de simulation obtenus avec GATE et SORTEO. Cet écart s'explique par une modélisation simplifiée de certains phénomènes physiques dans SORTEO et l'implémentation de techniques de réduction de variance. De très nombreux efforts de la communauté GATE française, financés dans le cadre de plusieurs appels d'offre de l'ANR (fGATE, hGATE) ont permis depuis de surmonter cet inconvénient majeur. La plate-forme GATE dispose également de l'appui d'une large communauté internationale d'utilisateurs et de développeurs qui permet d'assurer son évolution et sa pérennité. Elle apparait aujourd'hui comme la solution la plus adaptée pour la simulation de bases de données cliniques similaires à celles que nous présentons dans la suite.

Mes deux contributions principales ont concerné l'adaptation de la chaîne de simulation avec PET-SORTEO pour une caméra clinique, la caméra TEP Siemens ECAT/EXACT HR+ (CTI Siemens) [Reilhac et al., 2004] et deux caméras pour l'imagerie préclinique [Lartizien et al., 2007]. Ces deux études ont fait l'objet de collaborations avec deux centres TEP internationaux, le McConnell Brain Imaging Centre de Montréal (Canada) et l'Austrian Research Center GmbH-ARC de Seibersdorf (Autriche). Cet article faisait l'objet d'un taux de citation de 10 et d'un h-index de 2 en septembre 2014.

Ces travaux ont mobilisé une part importante de mon temps mais constitué une étape indispensable pour assurer la qualité des données simulées sur lesquelles s'est appuyée la majorité des projets de recherche que j'ai menés dans le domaine de l'analyse psychophysique puis de l'apprentissage statistique. Ce travail nécessitait de comprendre la physique de formation du signal TEP, depuis les

principes physiques de la désintégration radioactive jusqu'aux mécanismes responsables des pertes de signal (temps mort) dans les circuits électroniques de mise en coïncidence des événements détectés. Il impliquait également d'interagir avec les constructeurs afin d'accéder aux caractéristiques géométriques et électroniques des systèmes TEP modélisés puis de construire un plan d'expérience permettant d'ajuster les paramètres des différents modèles intégrés au simulateur (temps mort, etc) en confrontant des jeux de données simulés et expérimentaux pour différents types de fantôme radioactif.

Ces travaux ont eu des retombées intéressantes comme le suggère le taux de citation reportés ci-dessus. Je ne souhaite cependant pas analyser plus avant leur impact et me concentrer davantage sur la présentation des produits dérivés que sont les bases de données simulées ARSENE, OncoPET_DB et OncoPET_respi présentées dans la section suivante.

B.2.2.2 Création de bases de données simulées

Ces travaux avaient comme objectif de créer une base d'images de synthèse TEP corps entiers la plus réaliste possible et qui intègre, d'une part, une variabilité interindividuelle anatomique et fonctionnelle et d'autre part, une variabilité de cas pathologiques (cancer) et normaux. Ils ont conduit à la construction de plusieurs bases de données de complexité croissante, les bases de données simulées ARSENE, OncoPET_DB et OncoPET_respi présentées dans la section suivante.

(i) La base ARSENE (1998-2001)

J'ai construit cette première base d'images au cours de ma thèse à partir d'un modèle simplifié du corps humain adapté du modèle MCAT [Tsui et al., 1993] représenté sur la Figure B.2-1. Un ensemble d'images saines (sans tumeur) et pathologiques (avec tumeurs) a été généré à l'aide d'un code de simulation analytique [Comtat et al., 1999] développé et validé pour la géométrie du scanner Siemens/CTI Ecat Exact HR+, comme mentionné dans la section précédente. Cette base d'images contient 50 modèles pathologiques (contenant chacun 7 lésions sphériques de 10 mm de diamètre) et 50 modèles sains. Le modèle de tumeurs correspond à un signal Gaussien d'approximativement 2 voxels de largeur à mi-hauteur dans l'image reconstruite. Leur activité a été calibrée pour les besoins des études psychophysiques que nous souhaitons réaliser. Cette étude est rappelée dans la section B.3.3.1(ii). Une description plus fine des caractéristiques de cette base d'images est présentée dans l'article [Lartzien et al., 2004] reporté dans l'annexe B.8.3.3.

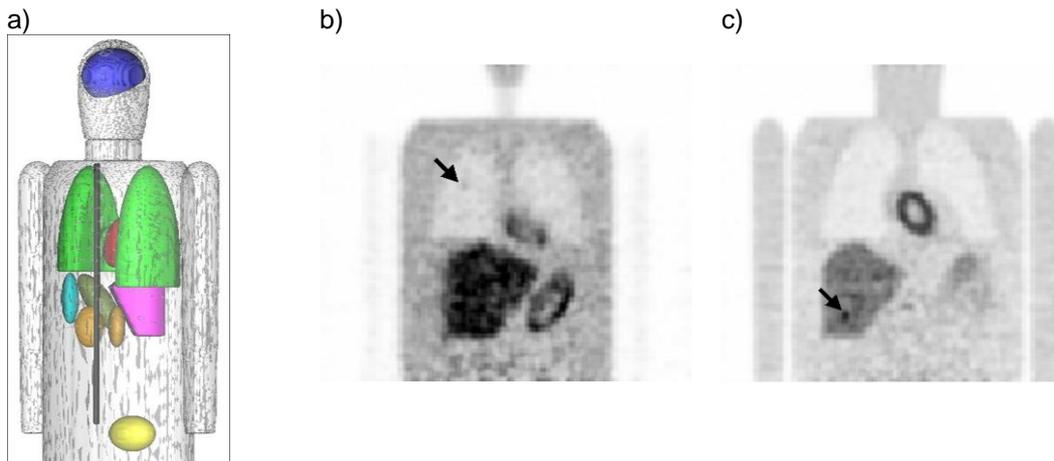


Figure B.2-1 a) Vue 3D du fantôme MCAT, b) coupe coronale présentant une tumeur dans les poumons b), coupe coronale présentant une tumeur dans le foie. Les images b) et c) ont été reconstruites à partir des données de simulation avec l'algorithme itératif AW-OSEM. Les volumes d'images résultants sont constitués de 128 x 128 x 225 voxels de dimension 5.0625 mm x 5.0625 mm x 2.4250 mm.

Cette base de données a été exploitée dans différentes études que j'ai réalisées depuis ma thèse, dans le domaine de l'optimisation des protocoles d'imagerie corps entier [Lartizien et al., 2000, Lartizien et al., 2002, Lartizien et al., 2003, Lartizien et al., 2004], des observateurs numériques [Lartizien et al., 2004], et encore récemment pour l'évaluation d'un système d'aide à la décision pour l'imagerie oncologique [Lartizien et al., 2012]. Ces données ont été diffusées dans le cadre de collaboration, en particulier à l'équipe de Paul Kinahan et Claude Comtat [Janeiro et al., 2006] ainsi que celle de Howard Gifford [Gifford, 2012].

Comme le suggère la Figure B.2-1, la principale limitation de cette base concerne le réalisme du modèle anthropomorphique basé sur une description géométrique des organes par ellipses ou cylindres tronqués. J'ai par la suite cherché à améliorer ce point en m'orientant vers un modèle anthropomorphique plus réaliste. La base oncoPET_DB créée sur le modèle de Zubal est présentée dans la section suivante.

(ii) *La base OncoPET_DB (2006-2009)*

Afin d'améliorer le réalisme des images simulées, j'ai choisi de travailler avec le modèle anatomique proposé par Zubal et al et obtenu par segmentation manuelle de l'acquisition tomodensitométrique (TDM) corps entier d'un patient [Zubal et al., 1994]. J'ai également profité des avancées réalisées en simulation MC et de l'accès privilégié du laboratoire CREATIS à des ressources de calcul distribué pour basculer vers ce type d'algorithme et choisir le logiciel SORTEO. Le modèle de Zubal se présente sous la forme d'une matrice 3D dans laquelle chacun des organes segmentés est représenté par une valeur entière comme illustré sur la Figure B.2-2. Une étude réalisée sur une série de données cliniques en association avec l'équipe médicale du service de médecine nucléaire du Centre Léon Bérard (CLB) de Lyon a permis de déterminer les rapports de fixation du traceur dans les principaux organes sains. Cette série de fixation moyenne par organe a été utilisée en entrée du simulateur pour associer chaque valeur de label du modèle de Zubal à un niveau de captation du traceur. Par ailleurs, une étude de calibration a été réalisée de manière à sélectionner des valeurs de contraste des tumeurs adaptées au besoin des études psychophysiques.

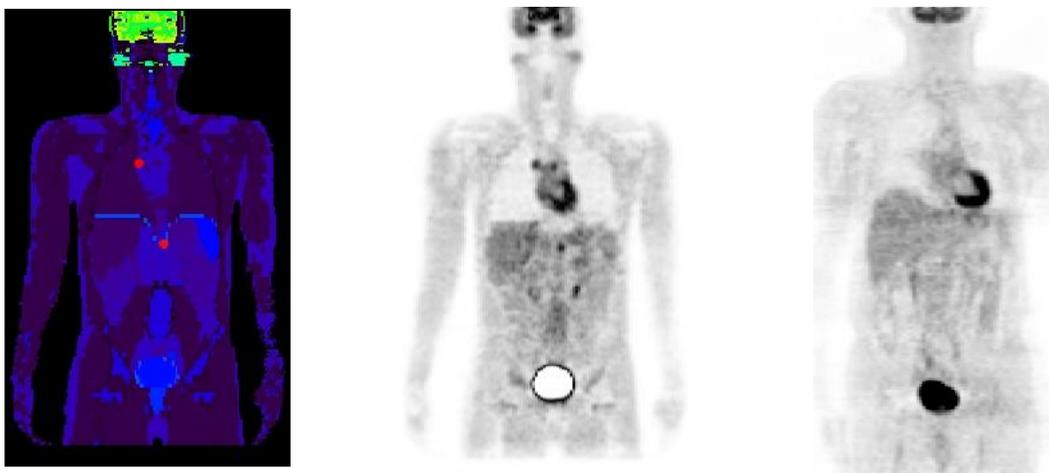


Figure B.2-2. Coupes coronales a) du Zubal, b) d'une image simulée du Zubal avec PET-SORTEO et c) d'une acquisition clinique.

Une première base d'images, nommée OncoPET_DB, a été produite. Cette base est constituée de 50 modèles sains et de 25 modèles pathologiques contenant chacun 5 lésions de taille et contraste variables réparties de manière à reproduire les caractéristiques réalistes de dissémination métastatique du lymphome. Ces simulations ont été réalisées principalement sur la ferme du centre de calcul de l'IN2P3 à Lyon (<http://cc.in2p3.fr>). La simulation d'un volume 3D corps entier nécessitait environ 70h de calculs sur 20 processeurs.

Ce travail a été réalisé dans le cadre de la thèse Sandrine Toméï que j'ai co-encadrée de 2006 à 2009. Il a fait l'objet d'un article publié dans IEEE Transactions on Nuclear Sciences en 2010 [Toméï et al., 2010] qui totalisait un taux de 11 citations en septembre 2014 et un h-index de 3. La base oncoPET est également diffusée librement au sein de la communauté scientifique à partir d'un accès sur la page www.creatis.insa-lyon.fr/oncoPET_DB. Elle totalise une centaine d'utilisateurs enregistrés depuis 2010.

La Figure B.2-2 démontre que la base oncoPET_DB permet d'atteindre un très bon niveau de réalisme. Cependant, cette base ne modélise pas la variabilité anatomique. De plus, à la fin des années 2000, la communauté TEP s'intéresse particulièrement à l'influence du mouvement respiratoire sur la qualité des images TEP, les centres cliniques sont tous équipés de machines TEP couplées à de l'imagerie tomодensitométrique par rayons X (TDM) et commencent à s'intéresser au couplage avec l'imagerie par résonance magnétique (IRM). Je décide donc de continuer à complexifier le modèle de simulation TEP en intégrant la variabilité anatomique et le mouvement physiologique et de m'intéresser parallèlement à la simulation d'autres modalités d'imagerie, en particulier la TDM.

(iii) La base de données OncoPET_respi (2008-2011)

La base de données OncoPET_respi est finalisée fin 2011. Elle est constituée de 15 patients simulés et intègre, d'une part, la variabilité anatomique et d'autre part, la respiration qui contribue à dégrader la qualité de l'image TEP par l'introduction d'artefacts de mouvement. OncoPET_respi est construite à l'aide du logiciel SORTEO et du modèle anthropomorphique XCAT [Segars et al., 2010] qui associe une description fine de l'enveloppe de chaque organe par une surface lisse de type NURBS (Non uniform radial basis surface) paramétrée par un ensemble de points de contrôle. Ce formalisme offre la possibilité d'ajuster le modèle générique de manière à introduire la variabilité anatomique souhaitée comme l'indique la Figure B.2-3. Dans cette étude, nous avons utilisé une série de 15 acquisitions cliniques TDM et procédé à l'ajustement manuel des points de contrôle. Une loi de mouvement de ces différents points a également été introduite de manière à simuler le mouvement des différents organes sous l'effet de la respiration. Ce travail s'est inscrit dans le cadre de la thèse de Simon Marache en contrat CIFRE avec le groupe Medisys de Philips. Le sujet portait sur l'influence des méthodes de correction du mouvement respiratoire sur la détection des lésions en imagerie TEP au FDG.

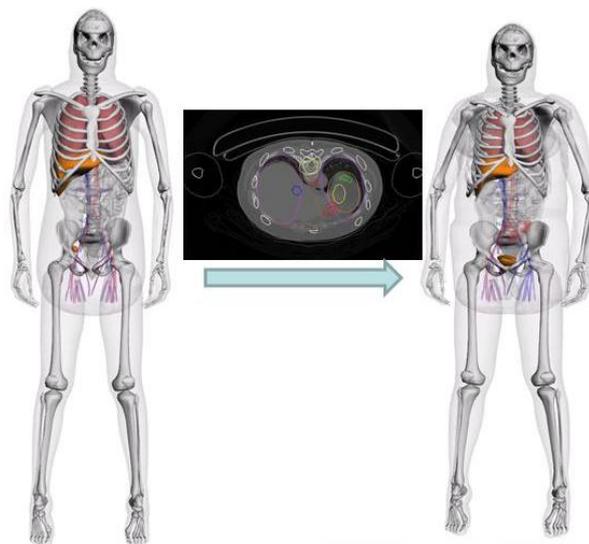


Figure B.2-3. Adaptation manuelle du modèle XCAT sur des données cliniques afin de créer une base de modèles de différente morphologie.

La Figure B.2-4 illustre quelques-uns des modèles générés par cette méthode tandis que la Figure B.2-5 montre un exemple de coupes coronales d'acquisitions 3D simulées et reconstruites avec différentes méthodes de correction du mouvement respiratoire. Les flèches indiquent la position d'anomalies sphériques simulées. Ces lésions ont été calibrées en taille et contraste afin de représenter un ensemble de cas difficiles pour le diagnostic.

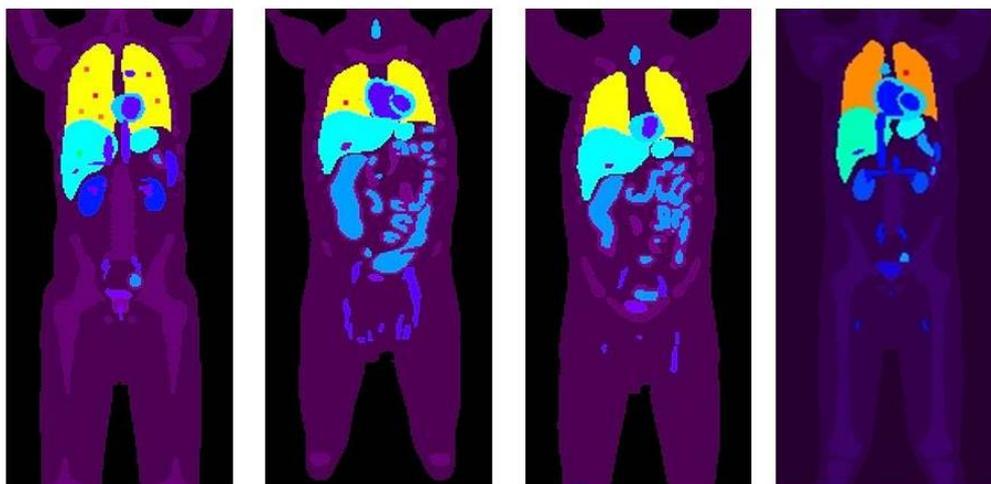


Figure B.2-4. Exemple de modèles anatomiques générés à l'aide du modèle XCAT.

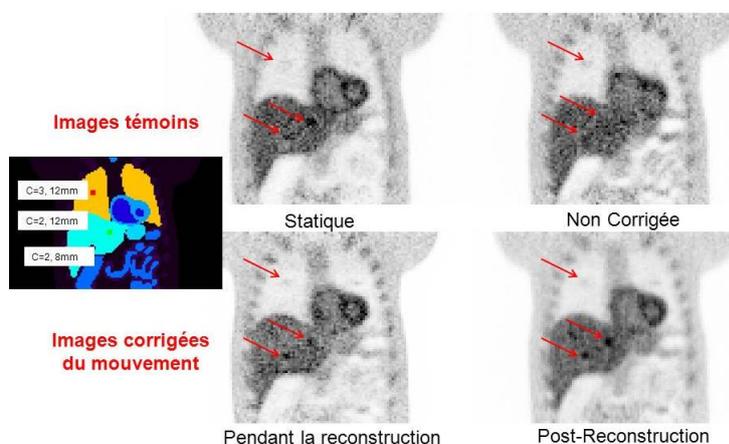


Figure B.2-5. Exemple d'une acquisition TEP simulée à l'aide du modèle représenté à gauche. Les deux images du haut correspondent aux cas extrêmes de données reconstruites sans correction de mouvement (droite) ou en supposant une correction parfaite (gauche). Les deux images du bas représentent les images obtenues en utilisant deux méthodes différentes de correction du mouvement respiratoire.

Ces simulations ont été réalisées en partie sur le cluster de PC du laboratoire CREATIS et principalement sur la ferme du centre de calcul de l'IN2P3 à Lyon (<http://cc.in2p3.fr>). La durée de simulation d'un volume 3D corps entier était d'environ **300h.cpu** sur 20 cœurs.

Une première étude de faisabilité avait donné lieu à un article publié dans la revue Proceedings of the IEEE en 2009 [Le Maitre et al., 2009]. Ce travail a fait l'objet d'une présentation en conférence [Marache-Francisco et al., 2011] mais n'a pas été diffusé comme la base oncoPET_DB. Nous discutons de ce point dans la conclusion de cette section.

B.2.3 Simulation d'images TDM

Comme souligné dans la section précédente, l'évolution de la pratique clinique vers l'association systématique d'une image anatomique TDM à l'image TEP pour le diagnostic m'a conduite à m'intéresser à la simulation de ce type d'imagerie.

Ce travail a été initié en 2009 dans le cadre de la thèse CIFRE de Simon Marache que j'ai présentée dans la section précédente. Nous avons envisagé, pour les besoins de cette thèse, de générer une base d'images TEP et TDM simulant la respiration, afin de reproduire les conditions de l'analyse clinique basée sur l'examen simultané des deux types d'images.

Après une étude de l'état de l'art, nous nous sommes orientés vers le simulateur SINDBAD développé par le LETI. J'ai initié une première collaboration avec ce laboratoire, qui s'est structurée par le financement, en 2010, du projet ANR VIP (Virtual imaging platform) dont l'objectif était de développer une plate-forme de simulation de différentes modalités accessibles par une interface web et connectée à des plates-formes calculs distribués (voir détails dans la section A.6.1.3). L'une des tâches identifiées de VIP concernait l'intégration du logiciel SINDBAD à la plate-forme VIP et sa validation à grande échelle.

Comme pour la simulation d'images nucléaires, la première étape consistait à modéliser la géométrie du système de détection et ajuster les paramètres du simulateur de façon à reproduire le signal mesuré. Nous avons proposé de réaliser une première validation en comparant des données simulées de fantôme géométrique (cylindre de plexiglass) avec des données acquises à l'hôpital cardiologique de Lyon sur un scanner Philips Brilliance 64. Cette étude s'est rapidement heurtée à l'impossibilité d'accéder aux données brutes du scanner, qui correspondent au signal directement enregistré par les détecteurs. Les constructeurs, en effet, ne donnent accès qu'aux images reconstruites ou aux données corrigées des différents types de phénomènes physiques, tels que la diffusion. Or, les données brutes sont cruciales pour contrôler les différents paramètres du modèle de simulation.

Nous avons parallèlement réalisé quelques simulations préliminaires de ce modèle anthropomorphe. La Figure B.2-6 a) représente une projection du fantôme XCAT correspondant à l'image d'atténuation du rayonnement direct (sans prise en compte du diffusé). La Figure B.2-6 b) représente une coupe coronale du modèle XCAT tandis que la Figure B.2-6 c) représente la même coupe coronale extraite de l'image reconstruite à partir d'une acquisition tomographique de ce modèle pour une géométrie de scanner proche de la machine Philips Brilliance 64. Cette illustration démontre que le modèle XCAT est suffisamment précis pour simuler correctement des structures fines telles que l'arbre bronchique. Cependant, l'hétérogénéité physiologique de l'atténuation des différents tissus n'est pas décrite dans ce modèle et pénalise le réalisme des images simulées qui sont trop 'lisses' et ne modélisent pas correctement la texture du bruit anatomique. Elles représentent donc un cas 'facile' pour la tâche de détection.

Ces deux points bloquants nous ont conduits à abandonner la partie TDM dans le cadre de la thèse de Simon Marache. Néanmoins, la structuration réalisée autour du projet VIP permet aujourd'hui de faciliter le travail de simulation par l'intermédiaire d'un accès facilité à SINDBAD au moyen d'une interface graphique conviviale et la production de données de grande dimension grâce aux ressources de calcul distribué. Ce travail a donné lieu à une publication en congrès [Tabary et al., 2009]. Il est également intégré à l'article consacré à VIP publié dans IEEE TMI en 2013 [Glatard et al., 2013].

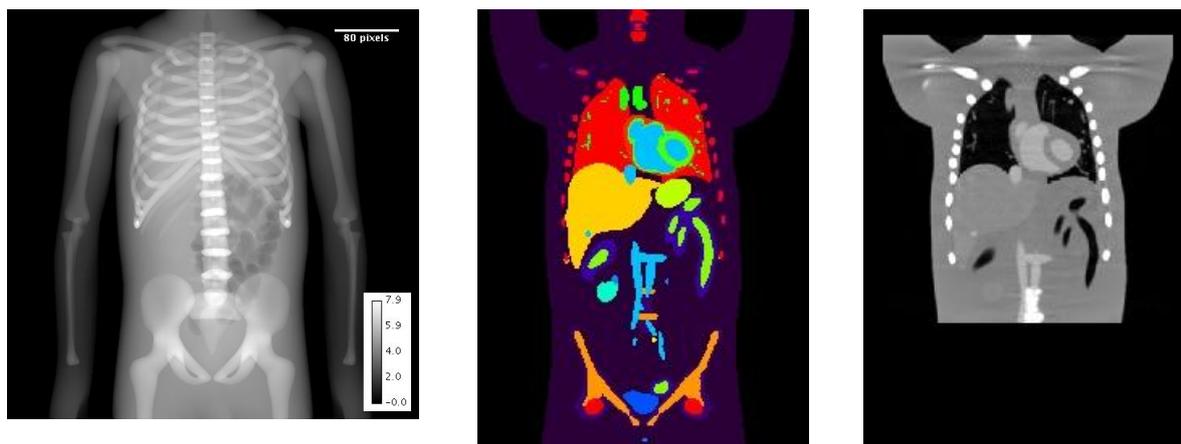


Figure B.2-6. a) Projection simulée du fantôme XCAT correspondant à une source de 110 kVp et un détecteur de $100 \times 100 \text{ cm}^2$ (pixels de 2 mm). b) Modèle de simulation XCAT et c) image reconstruite correspondant à une acquisition cone-beam du modèle b).

B.2.4 Limitations de la simulation et alternatives

B.2.4.1 Limites de la simulation et perspectives

La synthèse des différents projets réalisés au cours des années 2000 souligne une progression forte du réalisme des données TEP simulées comme l'illustrent la comparaison des figures Figure B.2-1, Figure B.2-2 et Figure B.2-5. Une perspective de ce travail proposait de continuer à complexifier le modèle de simulation afin de reproduire les caractéristiques de texture observées sur les images médicales et liées à du bruit physiologique. L'approche que nous avons envisagée reposait sur des méthodes de synthèse de texture. L'idée consistait à paramétrer des modèles de texture sur des images cliniques puis à les plaquer sur le modèle anatomique afin de modéliser une distribution de radioactivité et de densité tissulaire non uniforme pour chaque organe. Cette idée était développée dans un projet déposé en 2012 dans le cadre de l'appel d'offres 'Modèle Numérique' de l'ANR mais il n'a pas été retenu.

Comme souligné dans la section précédente, l'introduction en clinique de scanners couplant la TEP et la TDM m'a conduite, dans un premier temps, à m'intéresser à la simulation d'images tomodensitométriques TDM, l'objectif étant de compléter les bases d'images OncoPET_DB et oncoPET_respi par les images TDM associées. Le bilan des travaux réalisés au cours de l'ANR VIP a rapidement montré que la précision du modèle anatomique XCAT, bien que satisfaisante pour l'imagerie TEP dominée par le bruit de mesure, n'était en revanche pas suffisante pour la simulation d'images anatomiques réalistes, dominées, quant à elles, par le bruit physiologique. Les efforts de recherche dans le domaine de la simulation doivent donc, à mon sens, se concentrer sur la modélisation précise des structures anatomiques et de la variabilité des paramètres physiques que l'on essaie de simuler (hétérogénéité d'atténuation tissulaire en TDM, de fixation radioactive physiologique en TEP, de paramètres de relaxation en IRM etc..). Des avancées dans ce domaine permettront à nouveau à la simulation multi-modalités de tenir une place prépondérante dans le domaine de l'évaluation. Cet axe de recherche s'éloignait trop de mon centre d'intérêt scientifique, aussi j'ai décidé de recentrer mes travaux plus récents sur des bases de données cliniques, après avoir évalué différentes alternatives présentées dans la section suivante.

B.2.4.2 Alternatives à la simulation

(i) *Approches bootstrap pour la génération de données TEP répliquées*

Compte-tenu de la difficulté d'accès à des bases de données cliniques de grande dimension, j'avais envisagé de partir d'une petite population de cas cliniques et d'augmenter artificiellement sa taille par des méthodes de réplique de données de type bootstrap. Cette technique, appliquée à l'imagerie TEP, consiste à générer un jeu de données brutes (sinogramme ou données mode-liste avant reconstruction) par tirage aléatoire avec remise au sein d'une petite population d'images originales statistiquement équivalentes. J'ai proposé différentes stratégies d'échantillonnage de ces données et étudié l'influence du nombre d'échantillons de la population originale sur les propriétés statistiques des données rééchantillonnées. Ces travaux, réalisés à l'aide de données simulées, ont donné lieu à une publication dans IEEE TMI [Lartizien et al., 2010]. Je renvoie à la lecture de cet article pour la description de ce travail. Ces résultats démontrent la possibilité de répliquer des données en conservant les propriétés statistiques d'ordre 1 et 2 du bruit dans les images reconstruites avec un algorithme analytique ou itératif.

Les retombées scientifiques de ce travail sont intéressantes dans plusieurs domaines. Concernant la problématique l'analyse psychophysique développée dans la section B.3.3, l'utilisation de données répliquées par bootstrap permettrait

1) dans le cas d'études ROC classiques chez l'homme, d'augmenter la taille des bases d'échantillon et ainsi réduire la variabilité de mesure des performances diagnostiques (nous renvoyons à la section B.7.2 pour le rappel des métriques de performance de l'analyse psychophysique)

2) dans le cas de l'utilisation de modèles d'observateurs numériques, d'améliorer la précision de l'information statistique mesurée sur les données et ainsi d'estimer de manière plus précise les performances diagnostiques.

Cette méthode par bootstrap s'avère par ailleurs très intéressante pour améliorer la précision de l'information quantitative que l'on peut extraire de l'image TEP, et en particulier pour estimer l'erreur sur la mesure d'un paramètre.

L'article que nous avons publié dans la revue IEEE TMI faisait l'objet d'un taux de 9 citations totalisant un h-index égal à 3 en septembre 2014. L'intérêt majeur de ce travail portait sur les retombées potentielles dans le domaine de la quantification TEP.

Cette étude soulignait cependant quelques limites, la première étant l'impossibilité de générer des échantillons répliqués à partir d'une seule acquisition. Cela supposait donc de pouvoir disposer d'un jeu de données contenant plusieurs répliqués statistiquement équivalents de l'examen TEP de chaque patient. De plus, ce type de méthode permet d'augmenter la précision sur la mesure du bruit statistique mais ne renseigne pas sur les propriétés statistiques du bruit physiologique qui ne sont mesurables qu'en augmentant la variété de cas (i. e. de données). Ces différents points m'ont conduit à ne pas poursuivre cette piste plus avant.

(ii) Bases de données hybrides

L'une des difficultés principales liée à l'utilisation d'images cliniques pour l'évaluation concerne la définition de la vérité terrain. Il est en effet très difficile d'obtenir une annotation précise des zones pathologiques de l'image. L'équipe de G. El Fakhri et al à Boston a proposé une alternative intéressante consistant à fusionner des données brutes acquises sur des volontaires sains avec des acquisitions de fantôme (sphère de plexiglass) reproduisant les zones pathologiques [El Fakhri et al., 2007]. Le positionnement précis des sphères dans le patient était réalisé grâce à un support en plexiglass peu atténuant et calibré en position. Ces bases de données hybrides ont été exploitées dans de nombreuses études publiées par cette équipe pour l'évaluation. Nous discutons de ces résultats dans la section B.3.4. Ce choix méthodologique se heurte néanmoins à certaines limitations. La première concerne l'enregistrement des coïncidences diffusées. L'environnement tissulaire des sphères influence en effet directement la distribution de ce type de coïncidences parasites. Il est donc probable que le profil des coïncidences diffusées produites par les sources radioactives sphériques dans l'air soit différent du profil mesuré lorsque ces mêmes sources sont placées à l'intérieur du patient. On peut donc s'interroger sur l'impact de ces différences sur la détection. De plus, les acquisitions de sources externes ne reproduisent pas le mouvement respiratoire et physiologique des patients, de sorte que les images reconstruites ne sont pas affectées par ce terme de dégradation.

L'impact de ces deux phénomènes sur les performances diagnostiques n'ayant pas été clairement évalué, j'ai préféré ne pas suivre cette approche et fait résolument le choix de travailler avec des données cliniques.

(iii) Création de bases de données cliniques annotées

Une piste intéressante pour constituer une base de données cliniques de taille acceptable consiste à mutualiser les efforts de plusieurs centres d'imagerie pour collecter, annoter et mettre à disposition des jeux de données auprès de la communauté scientifique.

Ce travail de collecte a été réalisé pour différentes modalités d'imagerie, en particulier en TDM pulmonaire et en mammographie avec la constitution de bases d'images cliniques annotées distribuées :

- la Digital Database For Screening Mammography (DDSM)
 - o <http://marathon.csee.usf.edu/Mammography/Database.html>,
- la Mammographic Image Analysis Society (MIAS)
 - o <http://peipa.essex.ac.uk/info/mias.html>
- la Lung Image Database Consortium (LIDC) pour la TDM.
 - o <http://www.cancerimagingarchive.net/>

Cette piste tend à se développer pour d'autres applications cliniques sous l'impulsion, d'une part des études cliniques multicentriques, et d'autre part, de la communauté du traitement et de l'analyse des images médicales qui organise depuis quelques années, au sein de conférences internationales du domaine, des compétitions (challenge) ciblées sur une problématique précise (e.g. segmentation du ventricule en imagerie IRM cardiaque) permettant d'évaluer conjointement plusieurs algorithmes sur la même base de données. Ces challenges nécessitent ainsi de collecter des bases d'images rigoureusement annotées et mises ensuite à la disposition de la communauté.

L'analyse régulière de la littérature et de ces différentes manifestations montre l'absence de bases d'images cliniques TEP ou TEP/TDM oncologiques répondant aux critères statistiques requis pour les méthodes par apprentissage. J'ai donc, à partir de 2009, comme auparavant pour les données simulées, consacré une part de mon activité à la collecte de bases d'images cliniques adaptées au besoin des systèmes d'aide au diagnostic.

Analyse psychophysique en imagerie nucléaire par émission de positons

B.3 Analyse psychophysique en imagerie nucléaire par émission de positons

B.3.1 Contexte

Les premières images de TEP corps entier pour le diagnostic et le suivi thérapeutique des patients en oncologie ont été publiées au milieu des années 90. Les progrès réalisés au cours des années 2000 ont ensuite permis de démontrer le rôle majeur de cette modalité dans les protocoles de soin en cancérologie. Ils ont ainsi contribué à la dissémination des scanners TEP puis TEP/TDM en milieu hospitalier avec la prise en charge des examens au FDG pour différents types de pathologies cancéreuses dont le cancer pulmonaire.

Bien que prometteuses, les premières images publiées en 1997 n'étaient pas d'une qualité suffisante pour être exploitées en clinique. Le challenge était d'arriver à réaliser une image de l'ensemble du patient qui soit de qualité statistique suffisante malgré un temps d'acquisition réduit par rapport aux protocoles d'imagerie classiques centrés sur un organe et une dose de radioactivité injectée équivalente.

Mon projet de thèse, financé par Siemens France, l'un des trois industriels leader dans le domaine de l'imagerie médicale avec Philips et General Electric, avait comme ambition de proposer des protocoles d'acquisition et de reconstruction adaptés à l'imagerie corps entier pour l'oncologie. J'ai proposé d'aborder cette problématique sous l'angle de l'analyse psychophysique, qui permet de quantifier la notion de qualité diagnostique et étudie la relation entre le contenu de l'image et l'observateur chargé de réaliser une tâche donnée. Dans la suite de cette section sont présentées mes contributions principales dans ce domaine pour l'optimisation des protocoles d'imagerie TEP corps entier. Je conclus ce chapitre par l'exposé synthétique de mes travaux sur les observateurs algorithmiques dont l'objectif est de remplacer le lecteur humain par un modèle algorithmique dans les études psychophysiques. Ces derniers travaux ont marqué le changement d'orientation de mes recherches vers le domaine des systèmes d'aide au diagnostic.

Ces travaux réalisés au cours de ma thèse et de mon post-doctorat constituent un travail de recherche propre, réalisé toutefois en collaboration forte avec l'équipe de David Townsend et Paul Kinahan de l'Université de Pittsburgh, USA, qui m'ont accueillie au cours de ma thèse puis lors de mon post-doctorat. J'ai également profité de collaborations avec des partenaires cliniciens américains et français pour la définition des objectifs et les analyses psychophysiques.

B.3.2 Etat de l'art à la fin des années 90

Mon projet de thèse est initié par Bernard Bendriem, responsable en 1997 de l'équipe d'instrumentation du service hospitalier Frédéric Joliot (SHFJ) à Orsay, et Karl Blohm, responsable du Département 'Médecine Nucléaire' de Siemens France. Le SHFJ d'Orsay dispose alors, avec le Centre CYCERON de Caen et le CERMEP Lyon, de l'une des trois premières machines TEP installées en France. Le SHFJ a initié des protocoles de recherche avec différents instituts et hôpitaux parisiens pour évaluer l'apport de l'imagerie TEP corps entier en oncologie. En 1998, la faisabilité a déjà été démontrée, les protocoles d'acquisition dédiés sont implémentés sur les scanners, il a été prouvé que le fluorodeoxyglucose marqué au fluor 18 (18F-FDG) est un traceur intéressant pour le diagnostic et le suivi thérapeutique de nombreux types de cancer mais la qualité des données est encore assez médiocre et très variable en fonction de la morphologie des patients.

L'importance des enjeux cliniques de l'imagerie TEP pour l'oncologie suscite de nombreux développements académiques et industriels à la fois dans le domaine de l'instrumentation des caméras, mais aussi dans le domaine de la formation et du traitement des images. Sur ce dernier point, deux problématiques assez distinctes et complémentaires se développent en parallèle. Une partie de la communauté s'intéresse particulièrement à la quantification des images, qui recouvre l'ensemble des méthodes de reconstruction du signal brut et de correction prè- ou post-reconstruction permettant d'obtenir une mesure précise et reproductible de la concentration radioactive du traceur à l'échelle du voxel. Cette information est particulièrement importante dans le cadre du suivi thérapeutique pour évaluer la réponse au traitement. Elle permet également d'envisager des études

plus fines basées sur une modélisation de la cinétique du radiotracteur, permettant *in fine* d'extraire des paramètres biologiques que l'on souhaite corrélérer à l'agressivité tumorale, par exemple. La seconde partie de la communauté scientifique s'intéresse à l'amélioration des performances diagnostiques. Dans le contexte de l'imagerie TEP corps entier, la tâche diagnostique consiste à rechercher des anomalies de fixation radioactive de petite dimension et de faible contraste, disséminées dans un volume d'images 3D de 3 à 5 millions de voxels.

Mon projet de thèse propose d'adresser cette seconde problématique. Il s'intitule '*Optimisation des protocoles d'imagerie corps entier en tomographie par émission de positons pour la détection des hyperfixations du 18F-FDG en oncologie*'. Mes encadrants me laissent une grande liberté sur le choix des méthodes, me donnent accès à des données cliniques et à la possibilité de réaliser des expérimentations.

L'analyse bibliographique montre que quelques équipes s'intéressent à cette problématique au moyen d'acquisitions expérimentales sur des modèles simplifiés (cylindre) et à l'aide de métriques caractérisant les performances physiques des scanners (SNR) [Badawi et al., 1996]. Les discussions que j'entretiens avec les médecins nucléaires collaborateurs du SHFJ, soulignent, cependant, l'importance d'intégrer la notion de performance diagnostique dans le processus d'optimisation, i.e. le choix des paramètres de formation de l'image doit être guidé par la qualité finale du diagnostic établi à l'aide de l'image. Ce problème m'oriente naturellement vers l'analyse psychophysique de la détection en imagerie médicale. Contrairement aux métriques dérivées uniquement des caractéristiques physiques de l'image, cette méthodologie intègre l'influence de l'observateur et permet de mesurer un critère objectif et quantitatif des performances diagnostiques à partir de la mesure de courbes ROC, de l'anglais 'Receiver Operating Characteristics curves' ou dérivées (FROC, LROC, AFROC) basées sur des modèles dérivés de la théorie de la décision. Une synthèse des éléments de base de cette méthode est présentée dans l'annexe B.7.2 de ce mémoire.

A la fin des années 90, la plupart des études psychophysiques en imagerie nucléaire concerne la tomographie par émission monophotonique (TEMP), essentiellement pour l'imagerie cérébrale ou cardiaque. Les applications en TEP sont plus marginales mais leur nombre augmente entre 1994 et 1999. Les équipes utilisent soit des images cliniques annotées [Lowe et al., 1994, Cutler and Xu, 1996, Farquhar et al., 1999, Gifford et al., 1999], soit des images hybrides combinant des données cliniques sur lesquelles ont été superposées des anomalies simulées [Llacer et al., 1993], soit des données simulées [Gifford et al., 1999]. Elles s'intéressent à différents paramètres, tels que le choix de la méthode de reconstruction ou de la correction d'atténuation. La totalité d'entre elles sont réalisées dans le cadre de l'analyse ROC classique, qui propose un paradigme restrictif et peu représentatif de la pratique clinique en TEP.

Cette analyse de l'état de l'art m'a conduite à proposer un paradigme expérimental original permettant d'évaluer les performances de détection dans les images TEP en condition quasi-clinique et dans un cadre statistique adapté à une analyse de type AFROC. Les éléments principaux de cette chaîne d'évaluation et de son application sont présentés dans les chapitres B.3.3 et B.3.4.

Comme je le souligne dans la conclusion du chapitre B.3.4, les études dérivées de l'analyse ROC sont très coûteuses en temps humain (lecture des images) et en données (approche statistique). L'une des alternatives permettant d'alléger les contraintes expérimentales consiste à remplacer l'observateur humain par un algorithme capable de reproduire les performances de détection du système visuel humain. Le principe de ces algorithmes est présenté dans l'annexe B.7.3 et je consacre le chapitre B.3.5 à la description des travaux que j'ai réalisés dans ce domaine ainsi qu'à leurs retombées.

B.3.3 Evaluation des performances diagnostiques en imagerie TEP corps entier

B.3.3.1 Un outil de mesure des performances diagnostiques en conditions quasi-clinique

L'état de l'art présenté dans la section précédente montre que l'approche psychophysique, s'impose progressivement pour l'évaluation des protocoles d'imagerie nucléaire car elle fournit un cadre méthodologique bien formalisé et une mesure quantitative des performances de détection.

L'objectif de ma thèse était de mesurer de manière objective et reproductible les performances de détection d'un observateur humain dans une tâche diagnostique de détection d'anomalies de fixation radioactive de petite taille et de contraste variable.

Comme souligné dans l'annexe B.7.2.1, l'analyse ROC (Receiver Operating Characteristics curves) classique, initialement développée pour les besoins de l'imagerie radiologique, n'est pas adaptée à l'imagerie TEP. Elle s'applique en effet à des tâches de détection simples : les images de classe 'positive' contiennent une seule anomalie dont la localisation supposée est repérée. La tâche de détection consiste uniquement à reporter un degré de confiance sur la localisation pointée dans l'image. Ce paradigme est très éloigné de la pratique clinique en imagerie TEP, qui consiste à visualiser les images à l'aide d'un outil de visualisation tridimensionnel (3D), contrairement au mode de visualisation planaire utilisé en radiologie, et laisse toute la flexibilité d'ajustement de la dynamique de l'image. De plus, la tâche diagnostique ne consiste pas uniquement à classer l'image comme pathologique ou normale mais à localiser précisément les anomalies de fixation radioactive.

La stratégie AFROC [Chakraborty and Winter, 1990, Swensson, 1996], de l'anglais 'Alternate free receiver operating curve', développée parallèlement par Richard Swensson [Swensson, 1996] et Dave Chakraborty [Chakraborty and Winter, 1990] au cours des années 90, repose sur un paradigme expérimental proche du contexte clinique de l'imagerie TEP. Cette méthode, que nous présentons dans le paragraphe B.7.2.2, permet en effet d'évaluer les performances diagnostiques dans le contexte d'une tâche de détection et localisation d'anomalies multiples. L'observateur doit ainsi localiser toutes les anomalies présentes dans l'image, sans limite de nombre, et indiquer, pour chacune, un niveau de confiance λ_s . La courbe AFROC reporte alors le taux de vrais positifs pour un seuil λ_s en fonction de la fraction d'images analysées dans lesquelles l'observateur a reporté au moins un faux positif avec un niveau de confiance supérieur ou égal à λ_s . Swensson et Chakraborty ont montré que ces courbes peuvent être représentées par le modèle binormal utilisé pour l'ajustement des courbes ROC classiques. Ce formalisme permet de calculer des paramètres statistiques tels que l'aire sous la courbe (AUC) ainsi que les intervalles de confiance associés.

Ce choix du modèle AFROC a été conforté par les travaux menés parallèlement par Wells et al montrant les limites du système d'analyse ROC classique pour les images nucléaires [Wells et al., 1999, Wells et al., 2000]. Cette étude montre que l'observateur humain s'appuie sur l'analyse des plans de coupe adjacents pour déterminer la présence d'une lésion, vraisemblablement en raison du degré élevé de corrélations entre les voxels de l'image dans les 3 directions.

Les contraintes liées aux études psychophysiques concernant la nécessité de disposer d'une base d'échantillons représentatifs de la variabilité de la population et parfaitement caractérisés (contraste, localisation, taille..) m'ont décidée à construire la base d'images simulée ARSENE présentée au paragraphe B.2.2.2(i).

Le cadre de cette stratégie d'évaluation étant fixé, il a ensuite été nécessaire d'étudier la validité des hypothèses du modèle AFROC dans le contexte particulier de l'imagerie TEP puis de dimensionner l'étude pour assurer une puissance statistique suffisante. J'ai eu la chance d'être épaulée dans cette démarche par le Pr Swensson que j'ai rencontré lors d'une visite à l'Université de Pittsburgh au cours de ma thèse. Ce spécialiste de l'analyse psychophysique, et particulièrement des modèles LROC et AFROC, m'a guidée dans le dimensionnement de l'étude statistique sur des points importants que je rappelle ci-dessous. Cette collaboration nous a conduits à proposer un paradigme d'analyse AFROC pour l'imagerie TEP oncologique en condition quasi clinique [Lartzien et al., 2000, Lartzien et al., 2000]. J'ai parallèlement développé une interface graphique intégrée à l'outil de visualisation utilisé alors en routine clinique sur les imageurs TEP Siemens [Lartzien et al., 2000]. Ce module permettait

à l'observateur de manipuler les images dans l'environnement clinique standard et de reporter les lésions suspectes par un simple clic de la souris qui ouvrait alors une fenêtre permettant de reporter le critère décision associé à ce report.

(i) Validation des hypothèses du modèle statistique de l'analyse AFROC

Le choix du modèle AFROC était original à la fois dans le cadre de son application à l'imagerie TEP mais aussi parce que j'avais souhaité mener l'étude en 3D, c'est-à-dire réaliser une tâche de détection dans le volume d'images TEP. Ceci nécessitait de vérifier la validité des hypothèses du modèle statistique AFROC proposé par Swensson. Ce modèle mathématique suppose en effet que le choix de la lésion de plus forte suspicion est indépendant du nombre de lésions présentes dans l'image. Ceci revient à supposer que l'observateur maintienne un critère de décision stable au cours de la tâche de détections multiples. Il est en effet possible d'observer une augmentation du seuil de décision au cours de la tâche en fonction du degré de satisfaction de l'observateur. Cela se traduit par un déplacement vers la droite de l'axe vertical λ_s sur la Figure B.7-2 de l'annexe. Un moyen simple de vérifier cette hypothèse consiste à calculer les variations des fractions de vrais positifs et de faux positifs en fonction du nombre de lésions insérées dans l'image 3D. La condition de stationnarité est vérifiée si le nombre de vrais positifs est une fonction croissante du nombre de lésions effectivement présentes, et si la fraction de faux positifs est indépendante du nombre de lésions réelles. Les résultats expérimentaux présentés sur la Figure B.3-1 ont permis de confirmer cette hypothèse et de valider l'utilisation de la méthodologie AFROC dans notre contexte expérimental et pour les différents types d'images que nous avons utilisés.

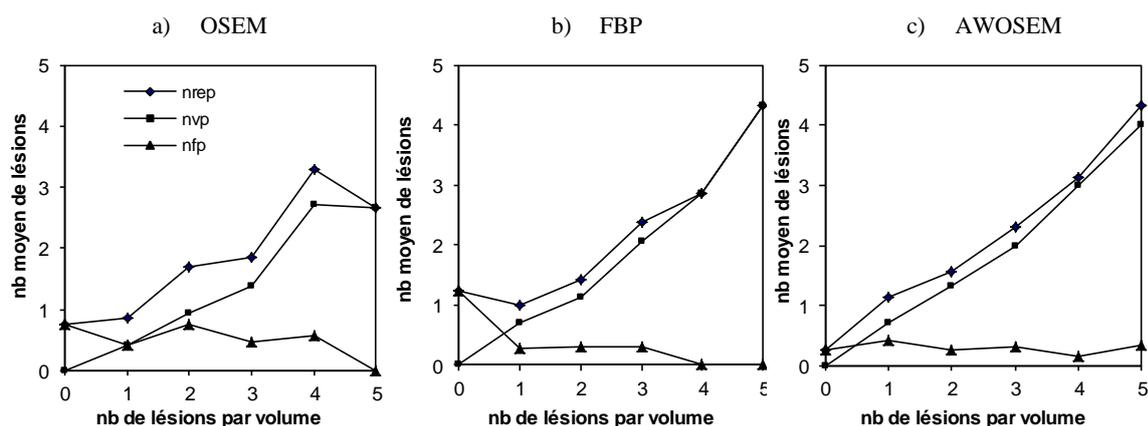


Figure B.3-1: Variations du nombre de vrais et de faux positifs reportés par un observateur pour des lésions insérées dans les poumons et pour les trois méthodes de reconstruction évaluées dans [Lartizien et al., 2003]. Le nombre de lésions totales reportées (nrep) est égal à la somme des vrais (nvp) et des faux positifs (nfp).

(ii) Choix et dimensionnement de la base de données

Les données simulées pour l'ensemble de ces études reposent sur le modèle anthropomorphe ARSENE présenté au paragraphe B.2.2.2(i). Nous avons fait le choix de simuler des lésions de petite taille, soit 10 mm de diamètre, correspondant à une taille critique pour le diagnostic.

Une part importante du travail a consisté à dimensionner les paramètres de l'étude de manière à garantir la puissance du test statistique de comparaison des différents protocoles d'imagerie étudiés. La méthode idéale pour définir les valeurs de ces paramètres consiste à réaliser une étude préliminaire afin de déterminer l'ordre de grandeur des aires sous les courbes ROC. Cette information permet ensuite, à partir de formules empiriques, de déterminer les paramètres de l'étude tels que le nombre d'échantillons, le nombre de lecteurs et de répétition des lectures, etc afin d'obtenir une puissance statistique suffisante [Hanley and McNeil, 1982]. Etant donnée la lourdeur expérimentale du

protocole de lecture, que je rediscute dans la suite, j'ai choisi de ne pas réaliser cette étude ROC préliminaire. Le nombre de lecteurs a été déterminé par la disponibilité et la bonne volonté des collaborateurs médecins que nous avons pu mobiliser pour cette tâche. Les deux études présentées dans la section B.3.4 sont basées sur les lectures de 5 observateurs, qui ont lu chaque série d'images sans répétition.

J'ai ensuite calibré le contraste des lésions pour échantillonner toute la gamme de difficulté de détection, de manière à obtenir une courbe de performance située dans le quart supérieur gauche (entre la ligne de chance et la courbe idéale en échelon de la Figure B.7-3). Une calibration séparée a été réalisée en fonction de la localisation de la lésion dans l'un des trois organes ciblés : tissu mou, foie et poumon. Cette étude a été réalisée à l'aide d'une base de données indépendante, simulée sur le modèle ARSENE. Une étude de détection simplifiée, i.e. sans critère de décision a permis de mesurer des courbes psychométriques qui mesurent la variation du taux de vrais positifs en fonction du contraste des lésions sur le modèle de la Figure B.3-2.

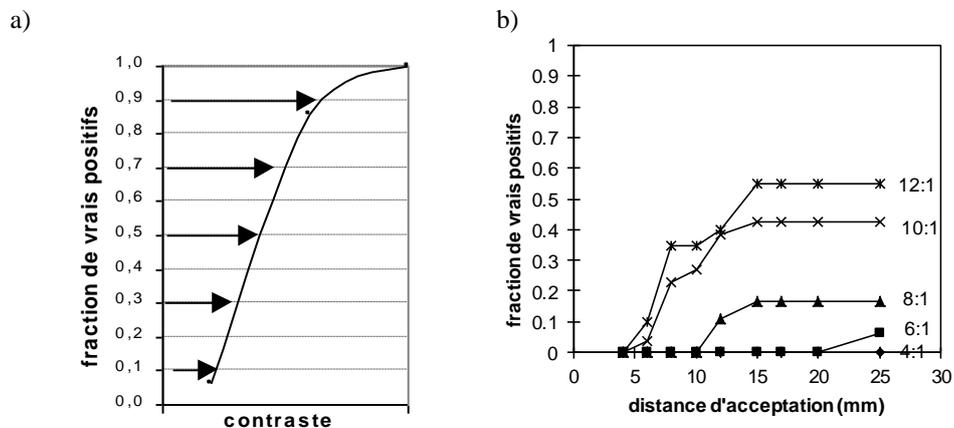


Figure B.3-2. a) Courbe psychométrique et échantillonnage des contrastes et b) fractions de vrais positifs en fonction de la distance d'acceptation pour des lésions simulées dans les poumons et différentes valeurs de contraste

Concernant le dimensionnement du nombre d'échantillons, j'ai suivi les recommandations du Pr Swensson préconisant de fixer le nombre d'échantillons de façon à ce que l'incertitude sur un point de mesure situé au milieu de la courbe psychométrique (i.e. pour une fraction de vrais positifs égale à 0.5) soit d'environ 10%. La fraction de vrais positifs vp étant une variable aléatoire binomiale, l'incertitude sur sa mesure est donnée par

$$\sigma(vp) = \sqrt{\frac{vp(1-vp)}{n}} \quad (\text{B.3-1})$$

où n est le nombre d'échantillons. Pour $\sigma = 0.1$ et $vp = 0.5$, on obtient une valeur de n égale à 25. Par conséquent, nous avons conçu la base de données de cette étude de manière à obtenir environ 25 lésions pour chacune des catégories de contraste.

Le dernier paramètre à ajuster est le degré d'incertitude entre la position exacte de la lésion et la réponse de l'observateur. Il est peu probable en effet que l'observateur positionne le curseur de la souris sur la position exacte du centre de la lésion. La précision de la réponse peut être évaluée par une distance, qualifiée de distance d'acceptation, liée à l'acuité visuelle de l'observateur, à la résolution de l'image et au contraste de la lésion. Les résultats de cette étude ont permis de mesurer l'ensemble des distances séparant chaque réponse à la position de la lésion la plus proche. J'ai ensuite calculé les fractions de vrais positifs correspondant à différentes valeurs du seuil

d'acceptation. La distance d'acceptation correspond à la distance minimale en dessous de laquelle la fraction de vrais positifs décroît. Sur l'exemple de la Figure B.3-2 b), cette distance est de 15 mm, correspondant à une tolérance de 3 voxels dans l'image reconstruite.

B.3.3.2 Impact

L'ensemble de ces développements, depuis la simulation de la base de données jusqu'au traitement automatisé des réponses de l'observateur humain, a permis de construire un outil d'évaluation global de la détectabilité en imagerie TEP corps entier, permettant d'analyser de manière systématique l'influence de différents paramètres d'acquisition et de traitement des images TEP sur la détectabilité. Une synthèse des principales applications de cette méthodologie est présentée dans le paragraphe B.3.4.

Ces travaux ont contribué à imposer l'approche psychophysique comme un standard de l'évaluation en imagerie TEP. Les principales contributions basées sur cette méthode sont les articles [Lartizien et al., 2003] et [Lartizien et al., 2004] que nous présentons dans les annexes B.8.3.2 et B.8.3.3. Ces articles ont été référencés respectivement 44 fois (h-index= 15) et 47 fois (h-index =12) d'après les statistiques de ISI Web of Knowledge en janvier 2015). Des travaux basés sur des métriques classiques, telles que le rapport-signal-sur bruit (SNR), continuent néanmoins à être publiés [Lodge et al., 2006, Brambilla et al., 2007, Park et al., 2008]. Cependant, leur nombre reste limité et le choix de ces métriques est souvent justifié pour réaliser une première sélection de paramètres avant une étude psychophysique. Une étude très récente dans le domaine de la détection en TEP démontre à nouveau les limitations de critères physiques simples par rapport aux métriques issues de l'analyse psychophysique [Rahmim and Tang, 2013] dans le contexte de l'évaluation d'algorithmes de reconstruction qui incorporent la modélisation de la réponse impulsionnelle du détecteur (PSF).

L'interface graphique que j'ai développée a également été citée par une équipe suédoise qui a depuis diffusé un logiciel libre de mesure de performances de détection dans des images 3D et en environnement clinique, le logiciel ViewDEX [Börjesson et al., 2005, Håkansson et al., 2010]. Ce logiciel, aujourd'hui dédié à l'imagerie TDM inclut, en outre, une analyse automatique des réponses des lecteurs (calcul des courbes ROC et dérivées).

Parmi les éléments du dimensionnement que j'ai présentés, la méthodologie permettant d'estimer la distance d'acceptation a été reprise par Kadrmass et al [Kadrmass and Christian, 2002, Kadrmass et al., 2009] dans une étude comparant les performances de différents types d'algorithmes de reconstruction (nous discutons les résultats de ce papier dans la section B.3.4.2.). J'avais également souligné l'importance de la calibration des contrastes des lésions. L'échantillonnage de la gamme de contraste présenté sur la figure Figure B.3-2 a) permet de garantir que la courbe psychophysique traduise les performances moyennes du système que l'on cherche à évaluer. Dans son étude, Kadrmass adopte une position un peu différente. Il souligne que le choix des contrastes détermine la valeur absolue de l'AUC pour un protocole donné mais ne doit pas avoir d'incidence sur le classement des différents protocoles que l'on souhaite comparer. Il tente ainsi de justifier que la calibration rigoureuse des contrastes n'est pas cruciale, contrairement à la méthodologie que j'ai proposée. Cette position semble assez discutable. Comme je le rappelle dans le paragraphe B.7.2.1(iii) de l'annexe et particulièrement dans l'expression (B.7-15), l'erreur standard sur la différence entre deux courbes ROC dépend de plusieurs paramètres dont le terme de variance lié à l'échantillonnage des cas. Le choix des exemples aura donc une incidence sur la valeur absolue de l'aire sous la courbe ROC mais également sur la puissance du test statistique de comparaison. Aussi, il me semble important de choisir un échantillonnage de cas traduisant les performances de la méthode sur un échantillonnage de cas représentant différents niveaux de difficulté.

Quelques groupes ont suivi le formalisme que nous avons introduit en publiant d'autres études psychophysiques avec des observateurs humains. Nous avons déjà cité les travaux d'une équipe de l'Université de l'Utah aux USA qui a proposé un protocole basé sur une étude LROC à partir de données expérimentales acquises sur des modèles anthropomorphiques dans lesquels sont insérées des lésions [Kadrmass and Christian, 2002, Kadrmass et al., 2009]. Leurs études consécutives se sont principalement concentrées sur l'amélioration du modèle expérimental de lésions, allant d'une sphère de plastique creuse en 2002, à une sphère en silicone en 2009, l'objectif étant de s'affranchir de la

capsule externe qui ne fixe pas la radioactivité. Une équipe de l'université du Piedmont oriental en Italie [Matheoud et al., 2011] a aussi contribué dans ce domaine à partir de données expérimentales du même type en proposant une étude ROC classique en 2 dimensions. Les contraintes expérimentales lourdes liées à l'acquisition de ces données limitent la variabilité de localisation et contraste des lésions simulées. Elles sont donc difficilement compatibles avec une étude de type AFROC telle que nous l'avons proposée. Les équipes de Boston et Philadelphie ont proposé une solution alternative basée sur des données hybrides construites à partir d'acquisitions cliniques dans lesquelles sont insérées des lésions acquises séparément [Surti et al., 2011] et insérées artificiellement dans le foie et les poumons (cette méthode est discutée dans la section B.2.4.2(ii)). Cette méthode offre plus de flexibilité et de réalisme que l'acquisition de données fantômes. Cependant, Surti et al se sont limités à une étude LROC qui requiert un volume de données plus petit et des séances de lecture plus courtes qu'une étude AFROC.

Cette analyse souligne les contraintes fortes du formalisme proposé, liées principalement à la disponibilité de lecteurs humains. Ce constat m'a conduit à choisir une voie alternative consistant à remplacer l'observateur humain par un modèle algorithmique. Ces travaux sont présentés dans le chapitre B.3.5. Le chapitre suivant présente les retombées de nos travaux sur le plan applicatif de l'optimisation des protocoles d'imagerie TEP corps entier pour l'oncologie.

B.3.4 Optimisation des protocoles d'imagerie TEP pour la détection des lésions oncologiques

B.3.4.1 Synthèse des résultats

J'ai consacré une part importante de mon travail de thèse à évaluer l'influence des différents paramètres de formation de l'image TEP sur les performances diagnostiques à partir des bases de données présentées dans la section B.2.2.2(i) et de la méthode d'analyse psychophysique résumée dans la section précédente.

L'une des questions d'intérêt à la fin des années 90 concernait le choix du mode d'acquisition des données. Les systèmes TEP permettait alors de réaliser soit des acquisitions en mode bidimensionnel (2D) c'est-à-dire dans des plans perpendiculaires à l'axe du détecteur, soit en mode tridimensionnel (3D), autorisant l'acquisition du signal dans des plans obliques. Le gain significatif de sensibilité apporté par le mode 3D en neuroimagerie motivait les industriels à simplifier l'architecture des machines (et réduire ainsi leur coût) en supprimant le mode 2D. L'avantage du mode 3D n'était cependant pas démontré en oncologie. Certaines études soulignaient, en effet que des sources de radioactivité parasites générées par la concentration du traceur dans certains organes (foie, vessie..) situés en dehors du champ de vue venaient contaminer le signal utile enregistré dans le champ de vue et dégrader ainsi la qualité de l'image reconstruite.

Dans ce contexte, la première question que j'ai étudiée a concerné l'influence du mode d'acquisition et de la dose injectée au patient sur la détectabilité des foyers de fixation radioactive en oncologie.

Une étude rigoureuse aurait consisté à créer autant de bases de données simulées que de couples de paramètres (mode d'acquisition, dose injectée) à évaluer et à faire une étude AFROC pour chacune de ces bases. La lourdeur des études psychophysique a imposé de limiter le nombre d'études. J'ai donc choisi d'utiliser un autre critère physique d'évaluation de la qualité des données pour estimer le paramètre de dose injectée. Ce critère, appelé NEC pour 'Noise Equivalent Count rate' permet d'évaluer la qualité statistique du signal brut mesuré avant reconstruction [Strother et al., 1990].

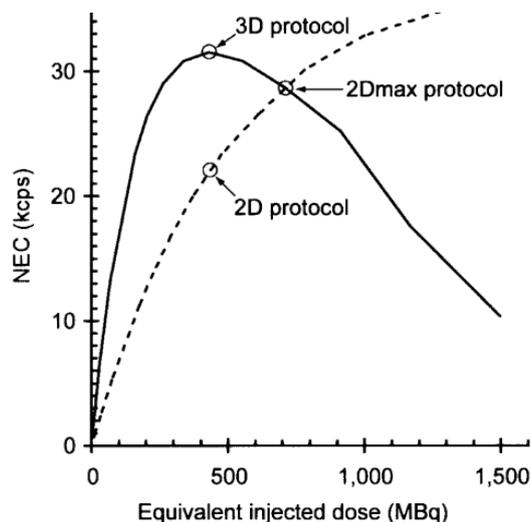


Figure B.3-3. Variation de l'indice NEC an fonction de la dose injectée et du mode d'acquisition pour une acquisition centrée sur l'abdomen.

Dans une première phase, j'ai évalué comment la qualité statistique des données acquises, résumée par l'indice NEC, était reliée à la dose injectée au patient en considérant les deux modes d'acquisition possibles des données TEP (2D vs. 3D). Cette étude expérimentale a été réalisée à l'aide d'un fantôme reproduisant la distribution de radioactivité dans un thorax humain [Lartzien et al., 2002]. Les données mesurées à l'aide du fantôme ont ensuite été corrélées à une série de données cliniques acquises sur des patients afin de valider les résultats obtenus avec le modèle de thorax humain. La description détaillée de cette étude est présentée dans la publication présentée dans l'annexe B.8.3.1. La Figure B.3-3 reprend le résultat principal correspondant à la variation de l'indice NEC pour les deux modes d'acquisitions en fonction de l'organe situé au centre du champ de vue.

Ces courbes montrent que la qualité statistique des données acquises en 3D est supérieure à celles des données acquises en 2D dans la gamme des faibles doses injectées, i.e. inférieure à 500 MBq (~13 mCi). Ces courbes ont permis de sélectionner un échantillonnage de 3 conditions d'acquisition représentées sur la Figure B.3-3 et restreindre ainsi le nombre de paramètres à étudier par l'approche psychophysique. Le point 3D correspond à la dose de radioactivité injectée permettant d'atteindre le maximum du NEC. Nous avons également considéré le point 2D correspondant à une acquisition pour la même valeur de dose injectée, ainsi que le point 2D max correspondant à une acquisition 2D au maximum de dose admissible, soit 740 MBq (20 mCi).

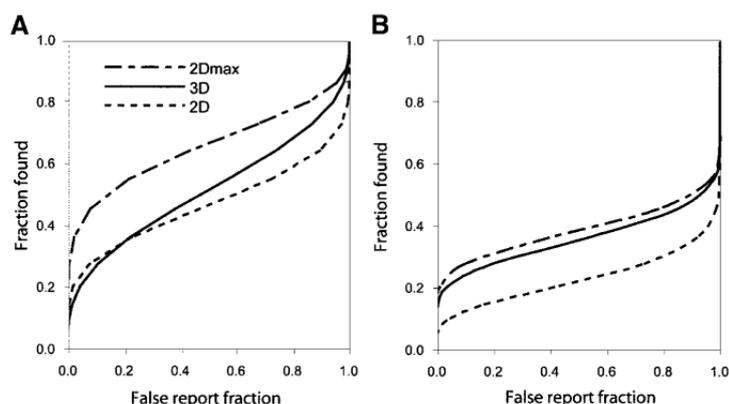


Figure B.3-4: Courbes AFROC moyennées sur 4 lecteurs pour des lésions situées dans les poumons (A) et dans le foie (B)

Trois bases de données simulées sur le modèle ARSENE ont été générées pour chacun de ces trois points de mesure et une étude AFROC basée sur les données de cinq observateurs humains a été réalisée à l'aide de l'outil d'évaluation décrit en B.3.3.1.

Le résultat majeur de cette étude a été de montrer que les données acquises en 3D, dans la gamme de dose injectée habituellement en clinique, permettaient d'atteindre des performances de détection supérieures ou équivalentes à celles obtenues à partir de données en mode 2D pour la même activité injectée [Lartzien et al., 2004]. La publication relative à ce travail est présentée dans la section B.8.3.3. La Figure B.3-4 présente les courbes AFROC moyennées sur 3 lecteurs pour la détection de lésions situées dans les poumons et le foie. On notera également un résultat annexe intéressant montrant que les lésions situées dans le poumon sont mieux détectées avec le mode 2Dmax. Ce résultat tendait donc à justifier le maintien d'acquisitions en mode 2D à forte dose injectée. Je le rediscute dans le paragraphe suivant.

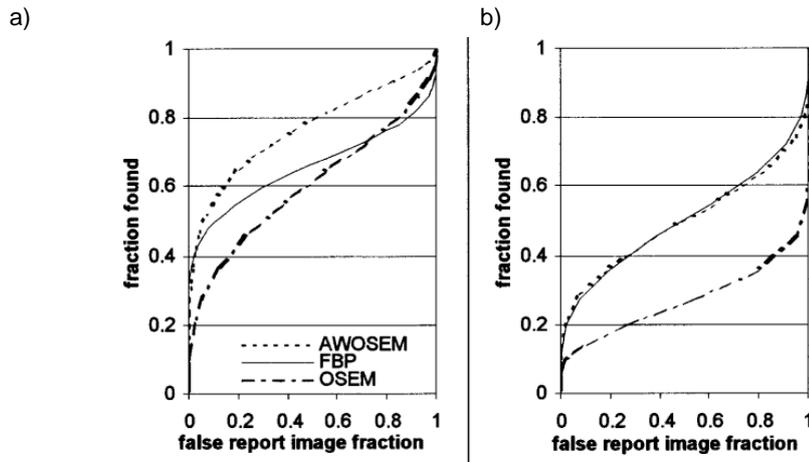


Figure B.3-5: Courbes AFROC moyennées sur les 5 observateurs pour tous les organes et tous les algorithmes de reconstruction.

L'autre approche pour maîtriser les effets du bruit statistique en imagerie oncologique concernait le développement de méthodes de reconstruction statistiques permettant de contrôler la propagation du bruit au cours de la reconstruction contrairement aux méthodes analytiques par rétroprojection filtrée (FBP). J'ai proposé d'étudier l'impact de ces techniques sur les performances diagnostiques. Le jeu de données correspondant au protocole d'acquisition 3D de la Figure B.3-3 a été reconstruit à l'aide de 3 algorithmes de reconstruction, l'algorithme FBP (Filtered backprojection) qui est la méthode analytique de référence en imagerie TEP [Kinahan and Rogers, 1989], l'algorithme OSEM (Ordered-subsets expectation-maximisation) qui est une version de l'algorithme itératif ML-EM dont la convergence a été accélérée [Hudson and Larkin, 1994, Kinahan et al., 1996] et l'algorithme AWOSEM (Attenuation-weighted OSEM) qui est une variation de OSEM incorporant la correction d'atténuation dans le processus itératif de reconstruction [Comtat et al., 1998, Michel et al., 1998].

La publication relative à ce travail est présentée dans l'annexe B.8.3.2 [Lartzien et al., 2003]. Le résultat principal, reproduit sur la Figure B.3-5, montre que la méthode de reconstruction AWOSEM présentant une modélisation plus réaliste du bruit statistique améliore significativement les performances de détection par rapport à l'algorithme OSEM standard et à l'algorithme de rétroprojection filtrée. Nous discutons de l'impact de ces résultats dans la section suivante.

B.3.4.2 Analyse des retombées

(i) Retombées industrielles

La majorité des résultats présentés dans le chapitre précédent ont été obtenus dans le cadre d'une collaboration CIFRE avec la société SIEMENS-France. J'ai ainsi pu bénéficier de contacts et échanges privilégiés avec la société CTI basée à Knoxville, USA, qui développait et produisait les systèmes d'imagerie TEP commercialisés par SIEMENS.

Les résultats concernant les faibles performances de l'algorithme OSEM pour l'imagerie TEP corps entier ont été rapidement pris en compte par SIEMENS qui a intégré l'algorithme AW-OSEM dans sa plate-forme logicielle et préconisé son utilisation en imagerie oncologique clinique. Cette transition vers l'utilisation de méthodes statistiques adaptées n'a cependant pas été immédiate dans tous les centres cliniques. Ainsi en témoigne, par exemple, une publication de 2011 [Matheoud et al., 2011] qui continue de considérer la méthode OSEM comme la méthode de référence pour l'oncologie. Je discute cet article dans la suite de cette section.

Les résultats de l'étude comparative sur le mode d'acquisition ont également eu un impact important sur le développement des nouvelles générations de tomographes, puisqu'ils ont conforté l'orientation

des constructeurs vers les systèmes d'acquisition purement 3D qui constituent aujourd'hui la quasi-totalité des systèmes cliniques. On peut cependant discuter l'interprétation un peu rapide de ces résultats. J'ai en effet montré que, pour des doses classiques de radioactivité injectée, et pour un patient de corpulence moyenne, le mode d'acquisition 3D permettait d'obtenir des performances de détection supérieures ou équivalentes au mode 2D. Cependant, cette étude montrait également que le mode 2D, combiné à une dose injectée plus élevée, permettait de surpasser significativement les performances du mode 3D (dans le cas de lésions pulmonaires). Ce deuxième résultat aurait nécessité une étude plus approfondie. D'autres équipes ont poursuivi ce travail en s'intéressant à l'influence de la morphologie du patient. Leurs résultats, que je discute dans la section suivante, confirment l'intérêt du mode 2D pour les patients de forte corpulence. Ces résultats n'ont cependant pas permis de contrebalancer l'orientation industrielle visant à abandonner la fabrication des systèmes proposant les deux modes d'acquisition.

(ii) Retombées scientifiques

Ces travaux ont été valorisés par des publications dans des revues à fort impact dans le domaine de l'imagerie nucléaire (JNM IF 5.78 en 2015).

(a) Choix de l'algorithme de reconstruction

Les publications concernant le choix de l'algorithme de reconstruction ont été référencés 43 fois (h-index= 15) d'après les statistiques de Web of Science Core Collection en janvier 2015. Une analyse plus fine des publications qui citent cette étude permet de mieux évaluer la portée de ce travail.

La majorité de ces articles s'appuie sur nos travaux pour argumenter le choix d'une méthode de reconstruction itérative par rapport à l'algorithme analytique de rétroprojection filtrée. On peut remarquer que certaines citations ne sont pas appropriées, soit parce que ces études utilisent finalement OSEM [Vogel et al., 2005] alors que nous avons souligné les faibles performances de cet algorithme pour les tâches de détection, comme l'indique la Figure B.3-5, soit parce que l'objectif de ces études est de faire de la quantification [Boellaard et al., 2004, Lubberink et al., 2004].

L'équipe de G. El Fakhri à Boston s'est intéressé à l'optimisation des protocoles d'imagerie TEP corps entier et a développé une approche expérimentale proche de la nôtre, basée sur l'analyse psychophysique. Leurs travaux, cependant, s'appuient rapidement sur l'utilisation de modèles algorithmiques pour remplacer le lecteur humain. Nous développons et discutons ce point méthodologique dans la section suivante. En 2007, El Fakhri et al proposent une étude reposant sur une série de données cliniques acquises sur la base une caméra HR+ et dans lesquelles sont insérées des lésions (petites sphères radioactives) acquises séparément [El Fakhri et al., 2007] (discutée dans la section B.2.4.2(ii)). Ils utilisent un modèle algorithmique du CHO pour évaluer les performances de détection des trois méthodes de reconstruction que nous avons comparées, et aboutissent à des conclusions équivalentes, en classant la méthode AWOSEM devant FBP et OSEM. Les résultats de l'étude [Lartzien et al., 2003] sont ainsi confirmés, et leur portée renforcée puisqu'ils ont été obtenus à partir d'une base de données qui intègre le bruit physiologique et la variabilité interindividuelle, absents de la base de données simulées sur le modèle ARSENE.

(b) Mode d'acquisition 2D ou 3D

Les deux études sur le choix du mode d'acquisition et l'optimisation de la dose injectée ont également obtenu des taux de citation importants (71 citations, h-index = 21 pour le papier sur le NEC et 46 citations, h-index = 12 pour l'étude AFROC (source Web of Science Core Collection, janvier 2015).

Lodge et al, en 2006, s'intéressent à la même problématique sur un scanner au LSO¹ capable de faire des acquisitions 2D et 3D [Lodge et al., 2006]. Ils réalisent une étude sur 27 patients basée sur la mesure de critères quantitatifs tels que des rapports contraste sur bruit mais pas d'étude psychophysique. Ils obtiennent des contrastes similaires (rapport d'activité entre une cible et le fond) entre les 2 modes d'acquisition mais des niveaux de bruit plus faibles en 3D. Cependant le gain du 3D n'est pas significatif.

L'étude de El Fakhri de 2007, que nous avons présentée un peu plus haut, s'intéresse aussi à la comparaison des modes d'acquisition 2D et 3D pour un scanner au BGO [El Fakhri et al., 2007]. Elle met en évidence une supériorité du mode d'acquisition 2D pour les patients 'forts' (avec un indice masse corporelle IMC >31, pour information l'IMC normal se situe entre 18 et 25) et un avantage pour le mode 3D chez les patients de plus faible corpulence (IMC <31). Le modèle ARSENE utilisé dans nos travaux correspond à un patient de morphologie standard, i.e. d'IMC inférieur à 30. Nous pouvons donc conclure que ces travaux confortent nos conclusions, en pondérant cependant la comparaison puisque les conditions expérimentales, en particulier la dose et le temps d'acquisition, sont différentes, de même que la métrique d'évaluation utilisée (CHO vs observateur humain).

D'autres études suivent sur la même problématique du mode d'acquisition entre 2006 et 2009, d'abord sur des scanners au BGO puis sur des nouveaux modèles de scanner LSO. Ces études s'intéressent aux performances diagnostiques mais ne mettent pas en œuvre d'études psychophysiques, vraisemblablement pour les raisons que nous avons évoquées à plusieurs reprises (contraintes de temps et de mobilisation de lecteurs). Visvikis et al obtiennent des conclusions assez similaires sur un autre type de scanner au BGO (GE Advance) [Visvikis et al., 2005]. Leur étude comprend des critères quantitatifs (SUV) et qualitatifs (évaluation de la qualité des images et du degré de confiance dans le report des lésions), mais pas de véritable analyse ROC. Cette étude sur 70 patients montre des performances similaires en 2D et 3D pour des durées d'examen équivalentes (5 min d'acquisition par position de lit). Les auteurs constatent cependant que leur base de données contient peu de patients très corpulents (poids >85 kg). Strobel et al, en 2007, proposent une étude assez similaire à celle de Visvikis, basée sur des données cliniques et un scanner GE Discovery ST au BGO [Strobel et al., 2007]. Les durées d'acquisition sont plus courtes que celles de Visvikis et pénalisent le mode d'acquisition 3D (3 min en 2D contre 1.5 min en 3D). Les auteurs montrent qu'il n'y a pas de dégradation significative de qualité entre les deux modes d'acquisition ; cependant on pourra noter que leurs images sont reconstruites avec l'algorithme OSEM standard, ce qui pénalise les conclusions de leur étude, étant donné les mauvaises performances de cet algorithme discutées précédemment. La même analyse s'applique à l'étude proposée en 2007 par Bettinardi et al sur un scanner GE Discovery STE PET/CT BGO avec des données expérimentales acquises sur un fantôme anthropomorphe [Bettinardi et al., 2007].

En 2009, Komar et al propose une étude proche de celle de Strobel et Visvikis sur 33 patients acquis en mode 2D et 3D sur un scanner GE Discovery STE au BGO [Komar et al., 2009]. Les données acquises en mode liste sont ensuite réorganisées en examens de différentes durées comprise entre 2 et 4 min. Komar et al évaluent les performances d'une variante de l'algorithme OSEM, la méthode VUE, qui intègre différents type de correction. La comparaison de différents critères quantitatifs et qualitatifs souligne une amélioration des performances en mode 3D, même chez des patients à fort BMI.

La synthèse de ces différentes études est difficile à réaliser car les conditions expérimentales (acquisitions sur fantôme ou patient, dose injectée, durée d'acquisition) sont très différentes. On observe cependant des résultats assez reproductibles, tendant à souligner que le mode d'acquisition

¹ cristal scintillant utilisé en remplacement du BGO à partir du milieu des années 2000.

3D permet de reproduire des images de qualité équivalente ou supérieure à celles produites en mode 2D pour le diagnostic, pour des patients de morphologie normale. De meilleures performances sont observées en 2D chez les patients à forte corpulence, cependant aucune étude psychophysique sur des données cliniques n'a clairement démontré de gain significatif permettant de justifier le maintien des systèmes d'acquisition 2D. Comme nous l'avons souligné au début de cette analyse, les industriels ont ainsi fait le choix de simplifier l'architecture des systèmes d'imagerie TEP, qui ne proposent désormais plus l'option 2D.

(c) Durée d'acquisition

Plusieurs études successives ont ensuite essayé de définir le temps d'acquisition optimal en mode 3D. L'objectif est de trouver un compromis acceptable entre la qualité d'image requise pour le diagnostic et la durée d'examen. La réduction du temps d'examen est un enjeu important en pratique clinique car elle permet, sur le plan économique, d'augmenter le rendement de la machine, mais également sur le plan pratique, d'améliorer le confort du patient, et réduire également le risque d'artefacts liés au mouvement du patient pendant l'examen.

Halpern et al, en 2004, [Halpern et al., 2004] réalisent une étude qualitative sur une base de patients acquis sur un scanner LSO Reveal RT PET/CT (CPS innovation) avec différentes durées d'examen. Ils évaluent la qualité diagnostique des images à partir d'indices qualitatifs subjectifs attribués par 3 médecins nucléaires et ne mettent pas en évidence de différences significatives de performance entre les acquisitions de 1 et 4 minutes par position de lit. Leur conclusion préconise des temps d'acquisition de l'ordre de 1 min pour des patients de faible corpulence. Cette étude est reprise et critiquée par Mattheoud en 2011 [Matheoud et al., 2011]. Cette équipe propose une étude ROC basée sur trois lecteurs et une série d'acquisitions sur un fantôme anthropomorphique contenant des lésions non sphériques pour le Biograph Hi-REZ PET/CT (Siemens). Elle montre que les performances de détection sont significativement dégradées dans le cas d'acquisitions de durée inférieure à 2 minutes. Il faut cependant noter que cette étude utilise l'algorithme de reconstruction OSEM qui reste la référence clinique de leur institution. Cet algorithme, comme nous l'avons souligné précédemment, n'est pas performant dans le cas de données très bruitées, et peut expliquer les faibles performances. Les auteurs soulignent d'ailleurs cette limite de leur étude.

Brambilla et al réalisent en 2007 une étude sur fantôme pour le système 3D Biograph16HI-REZ PET/CT au LSO. Ils s'intéressent à l'influence de quatre paramètres (dose injectée, temps d'acquisition, contraste de la lésion, et surface de la lésion) sur le rapport contraste-sur-bruit (CNR) à partir d'un modèle de régression logistique. Cette étude met en évidence l'influence prédominante du temps d'acquisition sur le CNR, alors que la dose injectée n'apparaît qu'en dernière position. Ces conclusions semblent indiquer, comme nous l'avons supposé dans notre étude, que la comparaison des modes d'acquisition doit être réalisée à temps d'acquisition équivalent.

Plus récemment, les études basées sur des critères de performances diagnostiques se sont intéressées à l'évaluation des machines à temps de vol (TOF en anglais pour 'time of flight') qui permettent d'améliorer le rapport signal sur bruit par rapport aux scanners TEP classiques. Dans une étude récente basée sur un modèle d'observateur algorithmique (NPWMF cf B.7.3.2(ii)) et un fantôme anthropomorphique, Kadrmaz et al montrent, sur un scanner Biograph mCT TOF PET/CT (Siemens), que la prise en compte du TOF dans la reconstruction permet de réduire les durées d'acquisition de 40% (145 sec contre 240 sec) pour les mêmes performances de détection (AUC), même chez des patients de forte corpulence [Kadrmaz et al., 2012]. Des résultats similaires ont été montrés par l'équipe de El Fakhri et Surti sur une machine Gemini TF PET/CT (Philips) à l'aide de bases de données hybrides mentionnées précédemment et en utilisant soit une étude ROC classique sur des observateurs humains [Surti et al., 2011], soit un modèle algorithmique de Hotelling (CHO) [El Fakhri et al., 2011]. Leur étude conclut que l'on peut obtenir une qualité équivalente en réduisant la durée d'acquisition de 3 min sur une machine clinique 3D standard à 1 min avec l'option temps de vol.

Comme pour le choix de l'algorithme de reconstruction, l'analyse de ces différents résultats est difficile en raison de leur caractère multiparamétrique. On pourra cependant noter une tendance visant à préconiser des temps d'acquisition supérieurs à deux minutes sur des machines cliniques de type

LSO. Les résultats plus récents sur des machines temps de vol tendent à valider des acquisitions courtes, de l'ordre de 1 minute.

(d) Digression sur l'optimisation de la dose injectée

Nos travaux basés sur la mesure de l'indice NEC [Lartizien et al., 2002] ont également fait l'objet d'un taux élevé de citations (taux de 71 citations en janvier 2015 correspondant à un h-index de 21 d'après le Web of Science Core Collection). J'ai fait le choix de restreindre à la présentation des travaux dans le domaine de l'évaluation et de l'analyse psychophysique, aussi je ne détaillerai pas l'impact de cette étude de manière exhaustive. On peut toutefois souligner que cette métrique est encore très utilisée à l'heure actuelle pour optimiser la dose injectée en fonction de la morphologie du patient. La plupart des études récentes dans ce domaine s'appuient sur les résultats présentés dans l'étude [Lartizien et al., 2004]. Nous avons en effet montré, pour une même courbe ROC correspondant au mode d'acquisition 2D, que les performances de détection semblaient corrélées à la valeur du NEC. Le point d'acquisition 2Dmax de la Figure B.3-3 conduit à de meilleures performances diagnostiques que le point 2D, comme l'indique la Figure B.3-4. Toutefois, cette analyse ne faisait pas l'objet de notre étude et nous n'avons pas été plus avant dans cette comparaison. En 2005, Watson et al ont proposé un modèle permettant d'extrapoler les courbes NEC individuelles à partir de l'enregistrement d'un seul examen statique [Watson et al., 2005]. L'analyse retrospective de ces courbes NEC permet de mieux comprendre l'influence des paramètres tels que la taille ou le poids du patient sur la qualité d'images et de définir ainsi des abaques permettant d'optimiser la dose injectée sur le modèle des travaux que nous avons proposés avec Paul Kinahan [Kinahan et al., 2001]. Les travaux de Watson ont été poursuivis par l'étude de Walker et al en 2009 [Walker et al., 2009]. Une approche basée sur la simulation Monte-Carlo a également été publiée récemment [Karakatsanis et al., 2014].

B.3.5 Les observateurs algorithmiques comme alternative au lecteur humain

B.3.5.1 Introduction

La synthèse de l'état de l'art et des travaux que nous avons présentés dans la section B.3.4 souligne une prise de conscience de l'intérêt de l'analyse psychophysique en imagerie TEP contrebalancée par son utilisation encore limitée. Cette frilosité s'explique par les contraintes expérimentales fortes liées au nombre de cas et de lecteurs nécessaires pour atteindre une puissance statistique suffisante. Dans nos études, par exemple, le temps moyen d'analyse d'un examen TEP, comprenant l'annotation de 7 lésions, était d'environ 15 min, ce qui correspond à une durée de l'étude de 30 à 40 heures pour l'annotation de 3 séries de 50 volumes d'images 3D.

Je me suis donc assez naturellement intéressée aux travaux sur les observateurs algorithmiques proposés comme une alternative à l'observateur humain pour l'évaluation des performances diagnostiques. Ces algorithmes réalisent un test statistique de discrimination, sur le modèle de la décision statistique, consistant d'abord à estimer une variable statistique à partir d'informations extraites de l'image (statistiques d'ordre 1 et 2) puis à la comparer à un seuil pour décider si l'image contient ou non un signal pathologique. Les modèles algorithmiques les plus courants sont décrits dans l'annexe B.7.3. Nous renvoyons également à la description des conventions de notations présentée dans l'annexe B.7.1. Dans ce chapitre, nous nous intéressons aux tâches de classification binaire entre deux classes $\{-1, +1\}$ correspondant aux cas normaux et pathologiques

La synthèse des travaux réalisés dans ce domaine est présentée dans le chapitre suivant, suivie d'une analyse des retombées et des orientations scientifiques récentes dans ce domaine.

B.3.5.2 Proposition d'un observateur adapté à l'imagerie TEP

Ces travaux ont été réalisés au cours du post-doctorat que j'ai effectué à l'Université de Pittsburgh aux Etats-Unis. L'objectif était de concevoir un observateur numérique capable de reproduire les performances de détection humaines pour l'imagerie TEP clinique. Ceci afin d'accélérer les études d'optimisation des protocoles d'imagerie par la méthode développée au cours de ma thèse en s'affranchissant de l'observateur humain.

Comme je le rappelle dans l'annexe B.7.3, il existe deux grandes classes d'observateurs algorithmiques. Les observateurs *idéaux* ou *bayésiens* reposent sur une connaissance *a priori* forte sur les propriétés statistiques de l'image et permettent de mesurer les performances intrinsèques du système de détection, correspondant à la limite supérieure de performance mesurable; les observateurs *linéaires* reposent sur des informations statistiques d'ordre 1 et 2 uniquement et sont généralement utilisés pour reproduire les performances de détection humaines. Je me suis orientée vers ces modèles linéaires, que l'on qualifie également de modèles *anthropomorphiques*.

L'état de l'art réalisé à la fin de l'année 2000 soulignait que ce domaine était exploré depuis quelques années pour l'imagerie nucléaire monophotonique (TEMP), exclusivement en 2D, et principalement dans le domaine de l'imagerie cardiaque. En imagerie TEP, l'équipe du Pr Steeve Moore à Boston venait de présenter un premier travail comparant des protocoles d'imagerie TEP 2D et 3D à l'aide d'un filtre adaptatif non blanchisseur (NPWMF) (voir description dans la section B.7.3.2(ii)) lors de la conférence IEEE MIC [Moore et al., 2000]. La double originalité de son approche était, d'une part de proposer un modèle numérique et d'autre part, de l'adapter à la spécificité de l'imagerie TEP en proposant une implémentation 3D.

J'ai choisi de développer cette approche en proposant des modèles originaux basés sur une analyse volumique (3D) de l'image, au lieu des modèles planaires (2D) habituellement utilisés, ceci afin de reproduire le mode de visualisation tridimensionnelle des images TEP. Ces modèles prennent en compte l'information disponible dans les trois dimensions de l'image TEP, au lieu de l'information du seul plan de coupe centré sur le signal, utilisée par les modèles numériques classiques.

L'année suivante, j'ai présenté une étude [Lartizien et al., 2001] sur la base de données ARSENE décrites en B.2.2.2(i) en proposant un modèle numérique 3D basé sur le modèle de Hotelling (CHO, voir description dans l'annexe B.7.3.2(v)). L'originalité de ce travail était double. Il proposait un filtre purement 3D et intégrait une première analyse comparative des performances humaines avec celles de ce modèle numérique en se basant sur les résultats de l'étude ROC publiée dans [Lartizien et al., 2004].

En 2002, l'équipe de El Fakhri et Moore a présenté de nouveaux résultats basés sur le CHO, mais en revenant sur une implémentation classique 2D [El Fakhri et al., 2002]. J'ai poursuivi la validation du modèle 3D en étudiant plus précisément les corrélations entre les performances d'observateurs humains et de différents modèles algorithmiques volumiques. Ce travail a été conclu par une publication dans la revue *Academic Radiology* en 2004 [Lartizien et al., 2004] qui est reproduite dans la section B.8.3.4 et dont je rappelle quelques résultats ci-après.

Les caractéristiques utilisées généralement par le CHO dans les tâches d'évaluation sont des caractéristiques fréquentielles [Barrett et al., 1993]. Elles correspondent aux contributions du signal sur certaines plages de fréquences. Ce choix est motivé par la volonté de reproduire le processus visuel humain qui traite l'information par un ensemble de filtres avant de l'interpréter [Sachs et al., 1971, Barrett et al., 1993]. J'ai choisi de conserver cette approche en adaptant le processus classique de décomposition fréquentielle au cas 3D. La technique la plus courante, représentée sur la Figure B.7-5 de l'annexe, consiste à intégrer le contenu fréquentiel de l'image sur des plages de fréquences juxtaposées, déterminées à partir d'une fréquence de coupure f_c et de largeur déterminée par une constante Q. Pour cette étude, j'ai choisi un modèle de filtrage à 3 bandes, et réalisé le passage au mode 3D en utilisant des sphères concentriques dans l'espace de Fourier comme représenté sur la Figure B.3-6 correspondant à 3 filtres passe-bande symétriques par rotation sans recouvrement de fréquence.

L'extraction des caractéristiques a été réalisée à partir d'une imagerie 3D cubique centrée sur la lésion d'intérêt, de 24 voxels de côté dans notre étude. Cette imagerie a été convoluée par chacun des 3 filtres passe-bandes pour donner un vecteur de caractéristiques \mathbf{v}_i de dimension 3 calculé au centre de l'imagerie, en suivant les notations de la section B.7.3.2(v). Cette étape permet de se rapprocher du modèle visuel humain, mais également de réduire la dimension du problème par rapport à un modèle de Hotelling classique qui considérerait un vecteur \mathbf{v}_i de 24^3 (13824) caractéristiques!

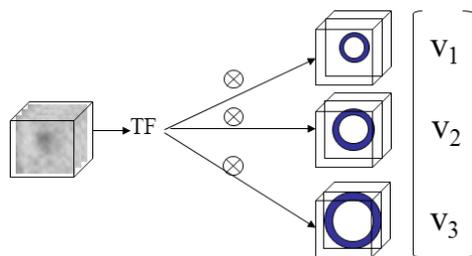


Figure B.3-6 Extraction du vecteur de caractéristiques \mathbf{v}_i du modèle CHO

Les résultats principaux de cette étude sont reproduits sur la Figure B.3-7 qui représente la corrélation entre les indices de détectabilité (défini dans l'annexe B.7.2.1) mesurés avec un observateur humain moyen et avec deux modèles numériques, le CHO et le filtre adaptatif non blanchisseur (NPWMF). J'ai également indiqué la corrélation avec une mesure de contraste-sur-bruit (CNR). Ces résultats ont permis de valider l'utilisation des modèles NPWMF (proposé par Moore et al) et CHO pour l'évaluation des performances diagnostiques sur la base de données ARSENE. Ils indiquent également que la mesure du contraste-sur-bruit peut constituer une métrique intéressante pour l'analyse de ce type d'images.

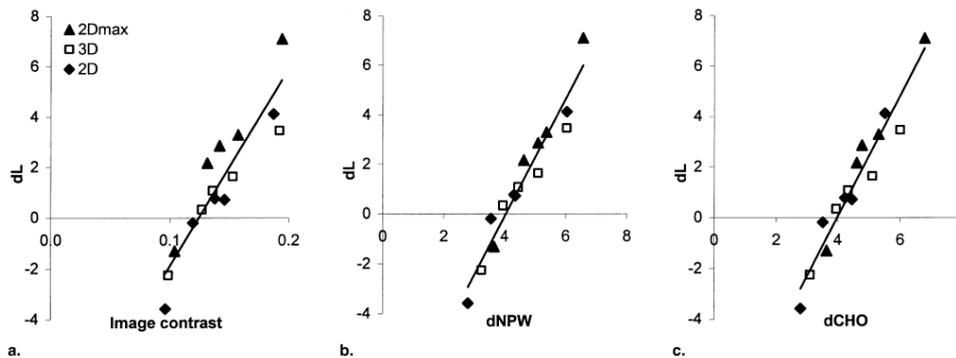


Figure B.3-7. Comparaison des performances de détection réalisées en utilisant le rapport contraste-sur-bruit (a) ainsi que deux modèles algorithmiques, le CHO (b) et le NPWMF (c) par rapport aux performances moyennes de détection humaine mesurées sur 5 lecteurs.

B.3.5.3 Retombées et orientations scientifiques récentes

(i) Retombées

Les résultats principaux de ces travaux, publiés dans la revue *Academic Radiology* en 2004, totalisaient un taux de 26 citations correspondant à un h-index de 8, d'après les statistiques de Web of Science Core Collection en janvier 2015.

Parallèlement à l'augmentation significative des travaux basés sur l'analyse psychophysique que j'ai commentés dans la section précédente, les modèles numériques ont également largement remplacé les lecteurs humains en raison des contraintes statistiques sur le nombre de lecteurs et le nombre d'échantillons. Quelques études se sont basées sur nos travaux pour valider l'utilisation de modèles numériques.

Thoen et al, par exemple, utilisent l'implémentation volumique du modèle NPWMF pour optimiser les paramètres physique d'un scanner PET/MR par simulation Monte Carlo [Thoen et al., 2013]. De Même Wu et al, utilisent une mesure 3D du rapport contraste-sur-bruit (CNR), en s'appuyant sur nos résultats, pour optimiser les paramètres d'acquisitions d'un scanner petit animal GE VISTA à partir de données simulées [Wu et al., 2013].

De nombreux travaux sur l'optimisation des protocoles d'imagerie cliniques en TEP corps entier s'appuient également sur des modèles algorithmiques. La plupart de ces travaux ont été cités et commentés dans la section B.3.4.2 consacrée à l'analyse de l'état de l'art sur ces protocoles. Ils reposent uniquement sur des modèles algorithmiques 2D.

Les principales équipes actives dans ce domaine sont celles de G. El Fakhri à Boston et de Kadrmass et al en Utah. El Fakhri et al utilisent un protocole d'évaluation standardisé basé sur l'utilisation d'une base de données hybrides (dont le principe est décrit dans la section B.2.4.2(ii)) couplée à un modèle CHO 2D. Ce protocole a été utilisé pour étudier l'influence sur les performances diagnostiques du mode d'acquisition [El Fakhri et al., 2007], du temps de vol [El Fakhri et al., 2011] et plus récemment de différentes méthodes de reconstruction prenant en compte le temps de vol et/ou la réponse impulsionnelle du détecteur [Schaefferkoetter et al., 2013]. Ils ont également utilisé le CHO sur des données expérimentales de lapin sur le système PET/MR pour évaluer une méthode de correction du mouvement [Chun et al., 2012].

L'équipe de Kadrmass utilise préférentiellement un protocole basé sur des données expérimentales acquises sur un fantôme anthropomorphe et associé à modèle de filtre adaptatif non-blanchisseur (voir section B.7.3.2(ii) de l'annexe). Ce protocole a été appliqué pour mesurer l'influence de la

modélisation de différents paramètres dans l'algorithme de reconstruction, tels que le temps de vol et/ou la réponse impulsionnelle du détecteur [Kadrmas et al., 2009 , Kadrmas et al., 2009 , Kadrmas et al., 2012].

La plupart des études basées sur un modèle algorithmique pour l'optimisation de protocoles d'imagerie TEP, TEP/TDM ou TEP/MR se basent sur quelques références dont [Lartzien et al., 2004] pour justifier la corrélation des performances du modèle numérique avec celles que l'on mesurerait chez l'homme. Cette corrélation, comme nous l'avons souligné, dépend cependant fortement du type d'images et de la tâche de détection. On peut ainsi reprocher à certains auteurs de justifier un peu rapidement la validité du modèle algorithmique employé pour leur application spécifique. Wu et al, que nous venons de citer, par exemple, citent nos travaux pour motiver l'utilisation du CNR dans le cadre d'une tâche de détection très éloignée du paradigme expérimental que nous avons évalué [Wu et al., 2013].

Dans une étude récente, Schaefferkoetter et al rappellent la nécessité de bien définir les limites de validité du modèle numérique comme substitut de l'observateur humain [Schaefferkoetter et al., 2013]. Leur étude est un exemple de bonne pratique lorsque l'on souhaite utiliser un modèle algorithmique. L'objectif scientifique de cette étude est de comparer les performances de détection de quatre méthodes de reconstruction TEP. Les auteurs commencent par réaliser une étude de type LROC sur trois lecteurs puis évaluent les performances de 8 modèles algorithmiques sur cette même série d'images. Ces 8 modèles 2D comprennent le NPWMF et le CHO (incluant un terme de bruit interne) présentés dans l'annexe B.7.3.2, tous deux combinés avec 4 types de filtre DOG (comme illustré sur la Figure B.7-5 de l'annexe). Les résultats présentés sur la Figure B.3-8 soulignent, pour cette application, que les modèles algorithmiques reproduisent correctement les performances humaines en respectant l'ordre de classement des 4 méthodes de reconstruction. Les auteurs utilisent alors les modèles numériques pour optimiser les paramètres intrinsèques de chacune des méthodes de reconstruction et évaluer l'apport de chaque méthode en fonction de la morphologie du patient.

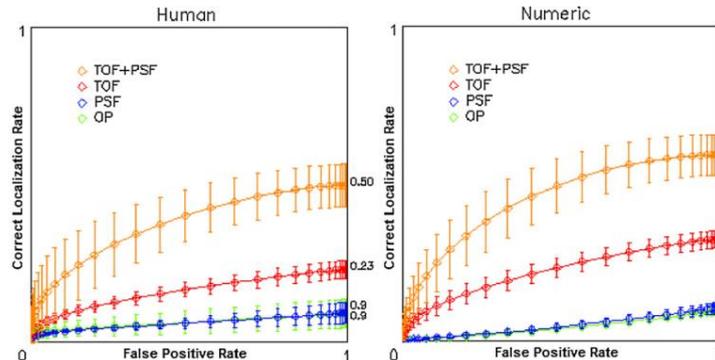


Figure 8. LROC curves showing mean localization performance of liver lesions for three human observers (left) and eight numerical models (right).

Figure B.3-8 Extrait de [Schaefferkoetter et al., 2013] soulignant les corrélations entre les performances de détection mesurées à l'aide de 3 observateurs humains (gauche) et 8 modèles numériques (droite).

Une étude ROC classique réalisée par l'équipe de El Fakhri avait auparavant montré une bonne corrélation entre les performances humaines et celles du CHO en 2D sur des données TEP acquises sur un système TOF [El Fakhri et al., 2011 , Surti et al., 2011]. L'équipe de Kadrmas et al a également vérifié la corrélation entre le classement de certains protocoles obtenus par le modèle NPWMF et des observateurs humains, en soulignant toutefois, que la puissance statistique des comparaisons multiples entre les différents protocoles différait en fonction du type d'observateurs [Kadrmas et al., 2009].

(ii) Orientations de recherche sur les modèles algorithmiques

Une observation précise de la Figure B.3-7 indique que les deux métriques comparées (abscisse et ordonnées) sont différentes. La mesure effectuée chez les observateurs humains est dérivée de l'aire sous la courbe A_z mesurée à partir d'une analyse de type AFROC basée sur une tâche de détection et de localisation des anomalies (voir annexe B.7.2.2 pour une définition plus précise des différents protocoles dérivés de l'analyse ROC). La mesure de l'indice de détectabilité des différents modèles algorithmiques, en revanche, est dérivée de l'aire sous la courbe ROC classique, comme l'indique l'équation (B.7-13) de l'annexe B.7.2.1. Elle correspond à l'évaluation d'une tâche de détection simple sans localisation ni détection multiple. Le modèle algorithmique est calculé à partir d'une image centrée sur la lésion, comme schématisé sur la Figure B.3-6. Dans cette étude, nous avons pu montrer que ces modèles anthropomorphiques, bien que modélisant un paradigme simplifié par rapport à la tâche diagnostique de l'observateur humain, permettait de reproduire correctement l'ordre de classement des différents types d'images analysés par des observateurs humains. Ce résultat n'est cependant pas généralisable. Les recherches récentes dans ce domaine s'intéressent au développement de modèles plus complexes, qui intègrent, par exemple, la tâche de localisation, ou une tâche d'estimation. Un état de l'art exhaustif sur ce sujet a récemment été proposé par Zhang et al [Zhang et al., 2014]. Nous donnons quelques exemples de recherches avancées dans deux types d'applications, qui incluent, soit des tâches de localisation et/ou estimation à partir de modèles 2D, soit le développement de modèles d'observateurs algorithmiques 3D.

(a) Détection et localisation

Le premier axe de recherche concerne le développement de modèles incorporant la tâche de localisation. Cette étude a fait l'objet de nombreux travaux dans l'équipe de Howard Gifford de l'Université du Massachusetts (USA) depuis la fin des années 1990, principalement pour l'imagerie TEMP [Gifford et al., 1999]. J'ai également ponctuellement collaboré avec cette équipe pour l'imagerie TEP [Gifford et al., 2007]. Gifford a proposé et comparé différents modèles algorithmiques intégrant une tâche de localisation. Ces modèles sont basés sur le modèle NPWMF avec et sans préfiltrage. La tâche de localisation est modélisée comme suit : la variable statistique λ est calculée pour chaque pixel de la zone analysée. Le pixel sélectionné est celui qui correspond à la valeur maximale de la variable statistique. Nous avons évalué ce modèle en 2D sur la base de données ARSENE présentée à la section B.2.2.2(i) et montré des taux de corrélation intéressants avec les performances visuelles humaines mesurées dans une tâche LROC [Gifford et al., 2007]. Gifford a poursuivi dans cette direction en proposant récemment un nouveau modèle basé sur une analyse séquentielle de l'image [Gifford, 2012]. Une première phase vise à sélectionner des points suspects de l'image à partir d'une méthode de clustering basée sur l'image du gradient. Le modèle du NPWMF est ensuite appliqué uniquement sur ces points saillants, permettant ainsi de diminuer la détection des faux positifs.

La Figure B.3-9 extraite de l'article [Gifford et al., 2007] montre la corrélation entre les aires sous les courbes A_{LROC} mesurées chez des observateurs humains (abscisse) et algorithmique (ordonnée). Contrairement à la Figure B.3-7, les indices utilisés pour les deux axes sont identiques. Dans cette étude, Gifford avait comme objectif de trouver le modèle algorithmique qui reproduise le plus fidèlement les performances diagnostiques humaines. Cet objectif est un peu différent de celui que je m'étais fixé dans l'étude présentée dans la section B.3.5.2 et qui est également celui de Schaefferkoetter. Il s'agissait en effet de reproduire les performances relatives de l'homme, en respectant notamment, l'ordre de classement des différents protocoles que l'on souhaite comparer. Je reviens sur ce point dans la dernière section de ce chapitre.

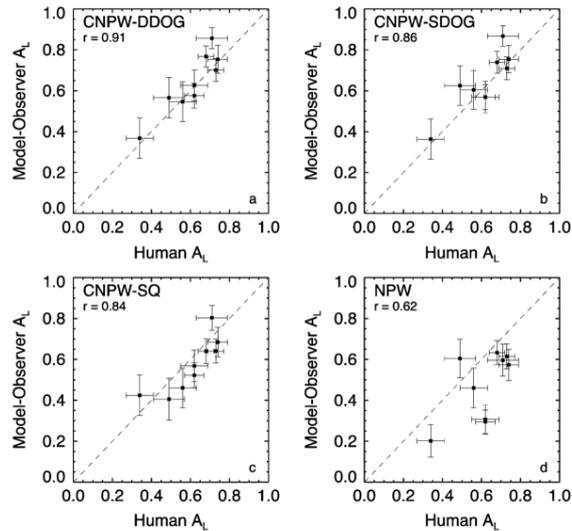


Figure B.3-9. Corrélation entre les performances de détection mesurées chez un observateur humains lors d'une tâche de détection et localisation (LROC) et à l'aide de différents modèles algorithmiques modélisant la tâche de localisation. Extrait de [Gifford et al., 2007].

(b) Détection et estimation

D'autres équipes, et en particulier celle de l'Université de Ghent en Belgique, s'intéressent à la mise au point de modèles permettant de résoudre une tâche conjointe de détection et d'estimation de paramètres du signal détecté tels que la taille et le contraste de l'anomalie, son orientation, le niveau de bruit. Bart Goossens et al [Goossens et al., 2013] ont récemment proposé le modèle CJO pour 'channelized joint observer' qui est basé sur la modélisation du signal par une fonction Gaussienne 2D paramétrée par un vecteur α comprenant la forme (variance différente selon la direction), l'amplitude et l'orientation. L'idée est d'estimer conjointement le vecteur α et la classe l_k de l'image analysée ($\mathbf{x}_i \in \mathbb{R}^N$, avec N la taille de l'image), à partir de l'expression de la probabilité conjointe $\mathbb{P}(Y = l_k, A = \alpha | X = \mathbf{x}_i)$ en faisant des hypothèses fortes sur les formes de la densité de probabilité conditionnelle $\mathbb{P}(X = \mathbf{x}_i | Y = l_k, A = \alpha)$, supposée suivre une loi de distribution Gaussienne. Une fois le vecteur α estimé, l'observateur idéal est utilisé pour déterminer la classe de l'image (hypothèse de classification en deux classes l_0 et l_1). Ce modèle est restreint au cas d'une détection 2D, il permet d'extraire des paramètres sur la forme du signal (à partir d'un modèle gaussien) et ne réalise pas la tâche de localisation. Goossens et al obtiennent de bonnes performances sur des images radiologiques pulmonaires dans lesquelles le niveau de bruit et le contraste des lésions (sphériques) insérées sont contrôlés.

Ce sujet intéresse également une équipe française qui a très récemment repris et validé ce modèle de manière plus exhaustive en collaboration avec Goossens sur des données simulées puis sur une base d'images cérébrales IRM T2 FLAIR de patients atteints de sclérose en plaques (SEP) [Zhang et al., 2013]. Zhang et al ont également proposé une extrapolation du modèle de Goossens par la prise en compte de la tâche de localisation et de détection de plusieurs lésions [Zhang et al., 2012]. Le modèle PCJO pour 'perceptually relevant channelized joint detection and estimation' effectue un traitement séquentiel en deux étapes. La première étape détecte la localisation de lésions suspectes à partir d'un modèle du système visuel humain (SVH). La deuxième étape applique le modèle CJO sur chacune des lésions détectées dans la première étape. Ce modèle a été évalué sur la même base d'images IRM T2 FLAIR de patients SEP. Les performances obtenues en termes d'aire sous la courbe A_{FROC} sont mieux corrélées aux performances humaines mesurées sur une série de 6 radiologues que

les performances mesurées à l'aide d'un modèle de CHO classique (i. e. sans tâche de localisation). Comme pour l'étude de Gifford, dont nous avons parlé précédemment, Zhang et al, s'intéressent ici aux performances absolues de détection.

(c) Modèles numériques volumétriques ou multicoupes ?

Nos travaux sur le modèle de CHO 3D basé sur des filtres sphériques dans l'espace de Fourier ont été approfondis par une autre équipe de l'Université de Ghent. Platisa et al ont en effet proposé de confronter deux approches de modèles algorithmiques pour l'analyse des images 3D [Platiša et al., 2011]. La première correspond aux modèles volumétriques dérivés du CHO que nous avons proposés dans [Lartizien et al., 2004] ; ces modèles sont baptisés vCHO pour 'volumetric CHO'. Dans cette étude, les filtres sont modélisés par des fonctions de Laguerre-Gauss 2D ou 3D. La seconde approche, appelée CHO multi-coupes ou msCHO, consiste à analyser chaque coupe de l'image 3D par un modèle algorithmique 2D puis à combiner les informations. Les trois modèles de CHO multi-coupes proposés sont reportés sur la Figure B.3-10.

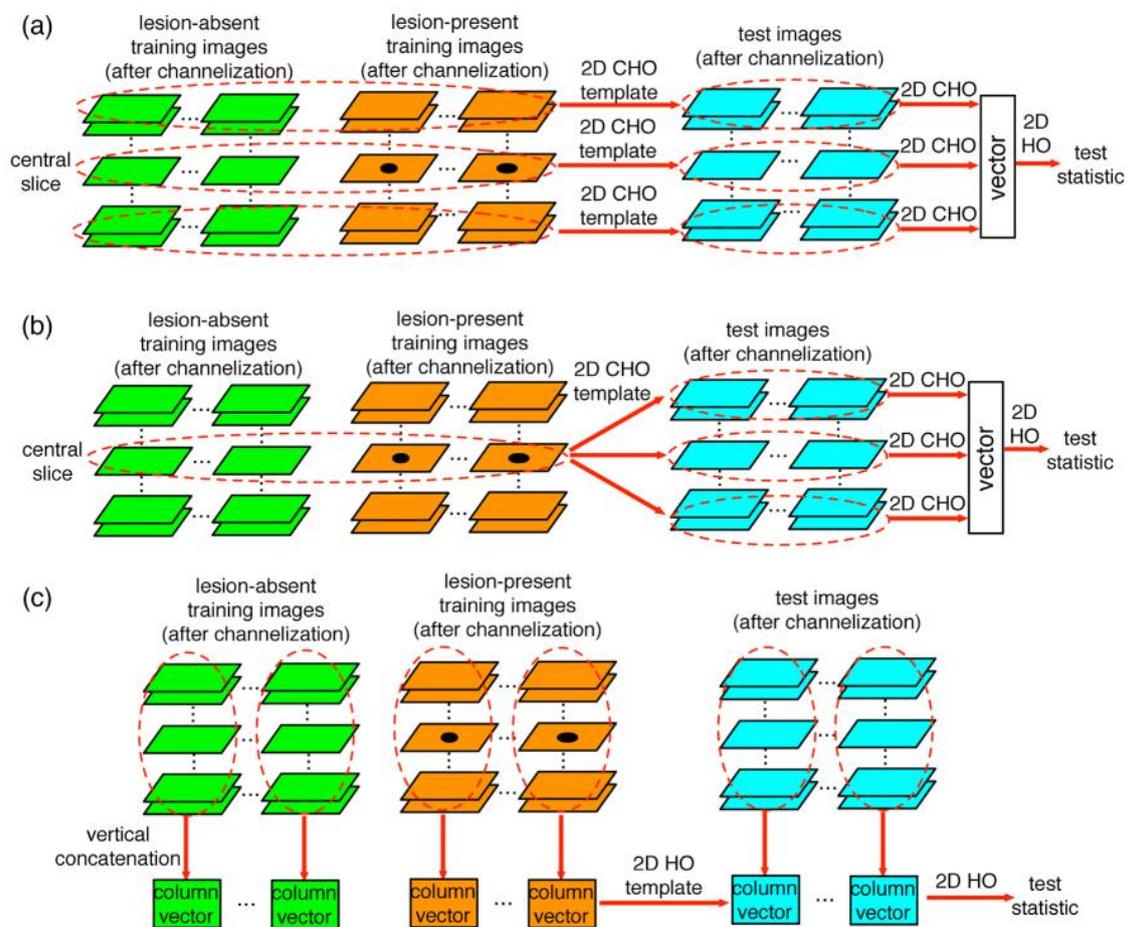


Fig. 9. The mschannelized hotelling observer (CHO) illustration: a: msCHOa illustration; b: msCHO b illustration; c: msCHO c illustration.

Figure B.3-10. Illustration des trois types de modèles CHO multi-coupes. Extrait de [Zhang et al., 2014]

Ces différents modèles ont été évalués sur des données simulées contenant un signal gaussien centré dans l'image et plongé dans 4 types de bruit de fond simulé de complexité croissante (White noise background WNB, correlated noise background CND, lumpy background LB et clustered lumpy background CLB). Les résultats montrent que les performances du vCHO sont supérieures à celles des différents schémas de msCHO et qu'elles se rapprochent de celle de l'observateur idéal (dans le

cas où il est possible de les mesurer, i. e. pour des bruits de type WNB et CNB). Platisa conclut que le vCHO semble mieux indiqué pour mimer l'observateur idéal. Les modèles multi-coupes peuvent, en revanche, être des candidats intéressants de modèle anthropomorphique évolué.

Parallèlement, Ljiljana Platisa a réalisé une étude psychophysique humaine pour comparer l'influence du mode de visualisation des données 3D sur les performances de détection dans différentes conditions de bruit [Platiša et al., 2012]. Elle a ainsi comparé un mode de visualisation mono-coupe (centré sur le signal à détecter) au mode de visualisation multi-coupe. Cette étude sur des données simulées met en évidence une amélioration des performances dans l'approche multi-coupes. L'étude suivante consistant à corréler les observateurs numériques et algorithmiques est annoncée mais n'a pas été encore publiée.

Dans le cadre sa thèse, Lu Zhang a également proposé une extrapolation du modèle PCJO développé conjointement avec l'équipe de Ghent et décrit dans le paragraphe précédent, au cas de l'analyse multi-coupes d'images 3D. Le principe du modèle msPCJO est similaire à celui du PCJO. Une première étape consiste à extraire des localisations suspectes dans l'image 3D à partir d'une méthode basée sur un mélange de Gaussienne. Une imagerie 2D centrée sur le centroïde de chaque lésion présumée est ensuite extraite du plan 2D. On se ramène ainsi à un problème 2D pour lequel l'application du modèle PCJO décrit précédemment est immédiate. Ce modèle n'a pas été évalué à notre connaissance.

(d) La quête de l'observateur universel ?

L'analyse de l'état de l'art récent montre qu'il existe deux tendances concernant l'évaluation des modèles algorithmiques. La première, illustrée par l'exemple de la Figure B.3-9, recherche la corrélation parfaite entre les observateurs numériques et humains. Dans cet exemple, Gifford montre par exemple, que pour le type d'images analysées, le modèle basé sur des différences de gaussienne (en haut à gauche) permet d'obtenir une très bonne corrélation. Il est cependant impossible d'extrapoler les performances de ce modèles à d'autres images TEMP, reconstruites, par exemple, avec un algorithme de reconstruction qui produirait une texture différente de l'image. On peut alors s'interroger sur l'intérêt, in fine, de ce type de modèle très spécifique, qui nécessite une étude de validation chez l'homme à chaque fois que les paramètres des images étudiées sont modifiés.

Jovan Brankov s'est également beaucoup intéressé à la recherche du modèle anthropomorphique idéal pour l'imagerie TEMP cardiaque. Comme Gifford, ses travaux se basent principalement sur le modèle de Hotelling associé à différents type de filtres fréquentiels et différents modèles de bruit interne qu'il essaie d'ajuster de manière à reproduire le bruit inhérent au processus visuel humain. Une étude récente [Brankov, 2013] compare les performances de dix modèles d'observateurs et défend une approche méthodologique intéressante. Afin de 'généraliser' au mieux les performances du modèles algorithmique et éviter le piège de la validation systématique contre une étude ROC humaine, Brankov propose d'entraîner le modèle sur différents types d'image, par exemple des images reconstruites avec différents paramètres de reconstruction ou de post-lissage [Brankov et al., 2009]. La validation de cette approche n'est cependant pas complètement réalisée à ma connaissance mais il serait intéressant de creuser cette idée.

L'approche que j'ai poursuivie, comme l'équipe de Boston, s'intéresse plutôt à reproduire les performances diagnostiques relatives. La plupart des études que nous avons référencées pour l'imagerie TEP dans le paragraphe B.3.5.2 montrent que les performances du CHO et du NPWMF sont généralement bien corrélées aux performances humaines (sans toutefois reproduire les performances absolues). Comme illustré par l'étude de Schaefferkoetter [Schaefferkoetter et al., 2013], une étude ROC préliminaire sur quelques images des différents types considérés permet d'évaluer si le modèle algorithmique reproduit bien l'ordre de classement de l'observateur humain. On peut alors utiliser ce modèle pour optimiser les paramètres des différentes images. Finalement, une étude plus avancée, du type LROC ou AFROC est réalisée avec des lecteurs humains pour la comparaison finale des protocoles retenus.

Diagnostic assisté par ordinateur pour l'imagerie du cancer

B.4 Diagnostic assisté par ordinateur pour l'imagerie du cancer

B.4.1 Contexte scientifique

Le projet que j'ai présenté au CNRS en 2004 s'inscrivait dans la continuité de ma démarche méthodologique ancrée sur la théorie de la décision statistique mais marquait une transition dans les objectifs. Il proposait en effet de s'intéresser au domaine des systèmes d'aide au diagnostic (CAD en anglais) dont la finalité est d'extraire l'information pertinente de l'image pour assister le clinicien dans sa tâche diagnostique en le guidant vers les zones suspectes. Le cadre formel proposé pour ces développements était celui des méthodes de classification par apprentissage.

La première étude que j'ai réalisée dans ce domaine reposait d'ailleurs sur l'analyse discriminante linéaire (ADL) dont le formalisme est similaire au modèle de Hotelling que j'ai beaucoup utilisé dans mes travaux de post-doctorat. Le lien théorique entre l'ADL et le modèle de Hotelling est explicité dans les sections B.7.3.2(iv) et B.7.4.3 de l'annexe.

Les projets que j'ai initiés après mon recrutement étaient pilotés par l'application médicale dans le domaine de l'imagerie du cancer. L'objectif était de proposer des systèmes d'aide à la décision qui répondent à un besoin clinique identifié et intègrent l'expertise des collaborateurs cliniciens. Le savoir-faire acquis au cours de ces différentes études m'a permis d'identifier des verrous spécifiques à l'apprentissage sur des données médicales et proposer des contributions théoriques originales dans ce domaine. Cet axe de recherche occupe aujourd'hui une place importante dans mon activité de recherche comme l'atteste le projet que de recherche que je souhaite mener dans les prochaines années.

Le chapitre B.4.2 revient sur la description synthétique du formalisme des systèmes CAD. Les chapitres B.4.3 et B.4.4 présentent ensuite une synthèse de mes contributions dans le domaine des systèmes CAD pour l'imagerie du lymphome en TEP/TDM et pour l'imagerie du cancer de la prostate par IRM multi-séquences. Je présente enfin dans le chapitre B.5 un exemple de contribution dans le domaine de l'apprentissage pour l'imagerie médicale.

Je souligne, dans l'exposé de ces différentes études, le cadre des collaborations scientifiques que j'ai construites avec des partenaires cliniciens, d'une part, et plus récemment avec des spécialistes du domaine de l'apprentissage supervisé et des mathématiques appliquées. Je précise également la contribution des doctorants et/ou post-doctorants que j'ai encadrés.

B.4.2 Cadre méthodologique des développements réalisés

B.4.2.1 Principe et évaluation des systèmes d'aide au diagnostic

(i) Les systèmes d'aide au diagnostic (CADx) et à la détection (CADE)

On distingue deux catégories de systèmes CAD, les systèmes d'aide au diagnostic, ou CADx et les systèmes d'aide la détection ou CADE, dont le principe est illustré sur la Figure B.4-1. Les CADE réalisent, à partir d'une image ou d'une série d'images recalées du patient, une analyse de chaque voxel, permettant de définir une carte paramétrique d'un indice (score, probabilité) caractérisant la présence de la pathologie dans l'image. Ils réalisent ainsi une tâche de localisation et de classification des anomalies. Les systèmes CADx proposent la même analyse pour une région d'intérêt préalablement délimitée par le médecin sur l'image. Ces systèmes n'intègrent donc pas la tâche de détection, l'objectif étant de proposer un second avis sur une cible suspectée par le médecin. Ces deux approches reposent sur le même formalisme que nous détaillons dans le paragraphe suivant.

■ Les systèmes d'aide au **diagnostic**



■ Les systèmes d'aide à la **détection**

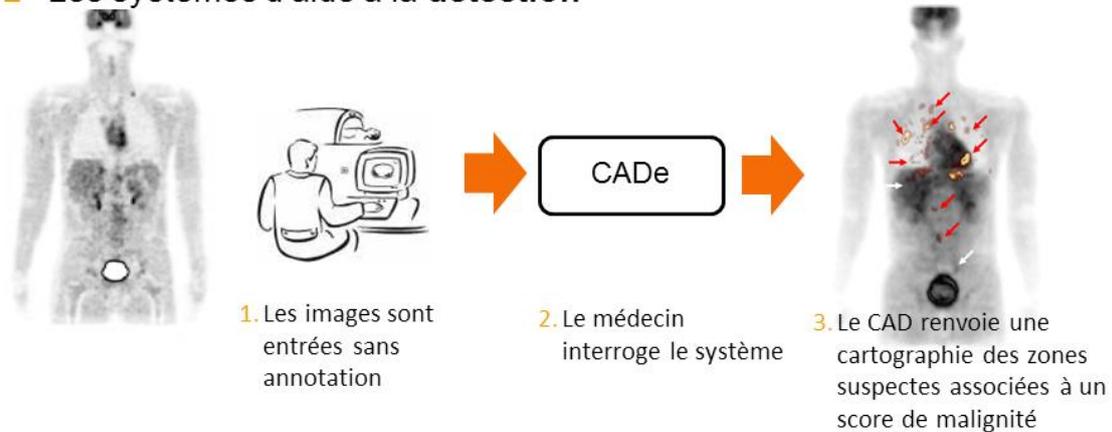


Figure B.4-1. Schéma comparé des systèmes d'aide au diagnostic (CADx) et à la détection (CADe)

(ii) **Principe des systèmes CAD**

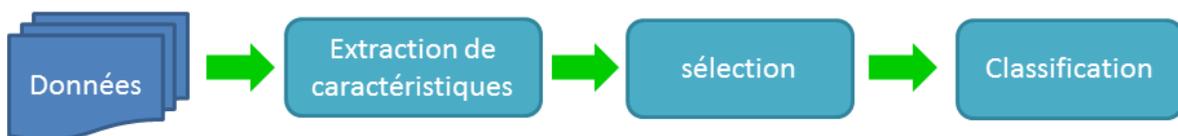


Figure B.4-2. schéma de principe d'un système CAD

Le principe général d'un système d'aide au diagnostic est représenté sur la Figure B.4-2 et résumé dans la suite de ce chapitre.

(a) **Prétraitement**

Le prétraitement des images inclut toutes les méthodes du traitement d'images permettant d'améliorer l'image brute afin d'optimiser l'information pertinente que l'on extrait par la suite pour réaliser la tâche de classification. Il s'agit, par exemple, de recalculer différentes images d'un même patient. Cette étape est particulièrement importante dans le cas de l'analyse par voxel réalisée par les CADe. Le choix de la chaîne de prétraitement dépend de la tâche de classification considérée (CADe, CADx), des

spécificités de chaque modalité d'imagerie et de connaissances a priori éventuelles sur des traitements permettant de rehausser la pathologie que l'on cherche à détecter.

(b) Extraction de caractéristiques

Cette étape consiste à extraire des données descriptives de l'image, appelées caractéristiques, descripteurs ou attributs, qui traduisent l'information pertinente pour la tâche de classification considérée. La littérature dans le domaine des systèmes CAD recense différentes familles de caractéristiques parmi lesquelles

- les caractéristiques géométriques :
 - Elles quantifient les détails visuels de la pathologie utilisés par les médecins pour leur diagnostic (sphéricité, compacité, symétrie etc..).
- Les caractéristiques statistiques :
 - Les statistiques de premier ordre (moyenne, écart-type..) caractérisent l'organisation moyenne des niveaux de gris de l'image sans prendre en compte leur distribution spatiale. Les caractéristiques d'ordre 2, telles que les paramètres de texture, peuvent traduire des différences d'homogénéité entre le signal normal et le signal pathologique. Elles traduisent l'apparence locale de l'image, son motif, c'est à dire l'organisation des détails d'une petite partie de l'image telle que le système visuel humain la perçoit.
- Les caractéristiques fréquentielles :
 - Le système visuel humain traite l'information à travers des canaux sélectifs en fréquence. Il est donc possible de se rapprocher d'une modélisation assez fiable de la vision humaine en se basant sur ces caractéristiques, extraites, par exemple, d'une analyse en ondelettes.
- Les caractéristiques 'biologiques' :
 - Elles quantifient des paramètres modélisant les phénomènes biologiques, tels que le taux de captation d'un traceur radioactif ou les constantes cinétiques d'un agent de contraste.
- Les caractéristiques textuelles ou sémantiques :
 - Elles regroupent généralement des informations issues du dossier médical du patient.

Le choix des caractéristiques définit la dimensionnalité du problème de classification et impacte les performances de classification. Deux approches coexistent pour cette étape.

La première approche consiste à effectuer l'extraction exhaustive d'un maximum de descripteurs, sans connaissance a priori. L'utilisation directe de ces attributs n'est cependant pas recommandée car elle peut conduire au phénomène de surapprentissage lorsque le nombre d'échantillons d'apprentissage est limité ou à une dégradation des performances intrinsèques de classification lorsque ces attributs sont corrélés. On réalise alors généralement une sélection des caractéristiques les plus pertinentes, soit préalablement à l'étape de classification, par des méthodes de type *filtre*, comme indiqué sur la Figure B.4-2, soit intégrée à certains algorithmes de classification, par des méthodes dites *enveloppantes*.

La deuxième approche consiste à sélectionner des caractéristiques descriptives sur la base de connaissances médicales a priori. On sait, par exemple, que la mesure de SUV en imagerie TEP, peut être liée à l'agressivité tumorale, de même que certains paramètres cinétiques des courbes de produits de contraste en imagerie IRM. Dans les travaux présentés dans la suite, j'ai proposé une approche hybride combinant l'extraction d'un maximum de descripteurs statistiques, en particulier les paramètres de texture, ainsi que certaines caractéristiques biologiques sélectionnées en fonction de leur pertinence clinique.

(c) Classification

Les algorithmes de classification ont pour objectif de regrouper des objets (données, région d'intérêt ou voxel en imagerie) en sous-ensembles, appelés *classes* partageant des caractéristiques

descriptives communes. Une première catégorie de méthodes dites 'non-supervisées', ou 'clustering' en anglais, vise à partitionner le groupe de données sans informations a priori sur les caractéristiques des classes (par exemple, sain ou pathologique). Ces méthodes reposent sur une analyse statistique des distributions de densité de probabilité des différentes caractéristiques descriptives et sur la définition d'une fonction de proximité entre les différents éléments de la population. On peut citer, par exemple, l'algorithme de classification des k-moyennes ou les cartes adaptatives.

J'ai privilégié le second type de méthodes, dites *supervisées* ou *par apprentissage*. Le principe de ces méthodes consiste à élaborer un modèle de prédiction à partir des caractéristiques descriptives extraites d'une base de données d'apprentissage dont l'étiquette de classe est connue. Contrairement à l'approche non supervisée, ces méthodes nécessitent de définir au préalable le nombre de classes et de disposer, pour chacune, d'un ensemble d'échantillons de référence. Ce modèle empirique permet ensuite de classer une image test à partir de ce même jeu de caractéristiques. L'annexe B.7.4 décrit plus précisément ces méthodes, particulièrement l'analyse discriminante linéaire (ADL) et les séparateurs à vaste marge (SVM), utilisés dans ces travaux.

(d) Post-traitement

Cette étape concerne principalement les systèmes CADe produisant des cartes paramétriques de même dimension que l'image originale et quantifiant un score de suspicion de présence de la pathologie. Comme pour l'étape de prétraitement, l'objectif du post-traitement est de 'nettoyer' les cartes brutes issues de la classification afin, par exemple, d'éliminer le bruit correspondant à des fausses détections. Les travaux présentés dans le chapitre B.4.4 illustrent le type de méthodes que l'on peut mettre en œuvre.

(iii) *Evaluation des systèmes CAD par apprentissage*

Les systèmes d'aide au diagnostic sont caractérisés par deux niveaux de performance. Le premier niveau concerne la performance diagnostique intrinsèque, qui mesure, par exemple, le taux de bonne classification, à partir d'une base de données test indépendante des données d'apprentissage. L'objectif est de mesurer, à partir de bases de données de taille limitée, les performances que l'on obtiendrait si le modèle prédictif était construit sur une base de dimension infinie représentative de la population de cas. On parle dans ce cas de performance en généralisation. Le second niveau mesure l'impact du système CAD sur les performances diagnostiques des médecins. La majorité des études évalue uniquement le premier niveau de performance.

La recherche de méthodes d'évaluation non biaisées des performances en généralisation des systèmes CAD, corrélée à la recherche de recommandations sur le dimensionnement des études (nombre de cas, influence du type de caractéristiques descriptives etc..) est un thème de recherche actif, en particulier dans l'équipe de Sahiner à l'Université de Ann Harbor aux Etats-Unis [Sahiner et al., 2008]. Une publication de l'AAM (American Association of Medical Physicist) regroupant les principaux compétiteurs dans ce domaine a récemment dressé un ensemble de recommandations sur la mesure de ces différents indices de performances et un bilan des différentes méthodes d'évaluation pour les systèmes basés sur une méthode de classification supervisée [Petrick et al., 2013]. Ces méthodes diffèrent principalement par la manière de partitionner la base de données en une base d'apprentissage permettant d'estimer le modèle prédictif et une base de test permettant de mesurer ses performances en généralisation.

La méthode d'évaluation idéale, dite de *partitionnement* (*hold-out* en anglais) consiste à partitionner la base de données en trois parties constituées d'une base d'apprentissage et d'une base dite de validation servant à optimiser les paramètres du modèle de classification, complétée enfin d'une base de test complètement indépendante des deux premières pour la mesure des performances intrinsèques sur le système optimal. Cette méthode est peu adaptée à l'évaluation des systèmes CAD pour l'imagerie médicale en raison des contraintes lourdes qu'elle implique sur la taille des bases de cas, problématique récurrente à l'ensemble des études statistiques que nous avons décrites dans ce rapport.

Une alternative opposée à la méthode de partitionnement consiste à apprendre et évaluer le modèle de classification sur la même base de données. Cette méthode de *resubstitution* (*resubstitution* en anglais) fournit une mesure très optimiste des performances en généralisation et doit donc être évitée en pratique.

La méthode par *validation croisée* (*k-fold cross validation* en anglais), enfin, consiste à découper la base en k sous-ensembles de même dimension. Un premier modèle prédictif est estimé sur $(k-1)$ sous-ensembles puis évalué sur le sous-ensemble restant. Cette procédure est répétée k fois en modifiant à chaque itération le sous-ensemble test, de sorte que chaque exemple de la base sert à $k-1$ apprentissages et est testé une seule fois. La méthode par *leave-one-out* (LOO) est un cas particulier de la validation croisée pour lequel le nombre de sous-ensembles est égal au nombre d'éléments de la base de données. Chaque exemple est ainsi étiqueté une seule fois avant d'être utilisé pour déterminer la métrique d'intérêt, par exemple, l'aire sous la courbe ROC (AUC) (voir annexe B.7.2). Cette méthode est connue pour donner une estimation faiblement biaisée (biais pessimiste) des performances intrinsèques. D'autres méthodes, telles que la méthode de Fukunaga-Hayes ou de bootstrap sont également intéressantes à considérer; nous les avons utilisées dans la référence [Lartzien et al., 2014] et renvoyons à la lecture de cet article pour une description précise.

J'ai porté une attention particulière à cette étape cruciale que constitue l'évaluation des performances du système CAD. Les choix méthodologiques sont discutés dans chacune des études présentées dans la suite de ce chapitre.

B.4.3 Contributions dans le domaine des systèmes CADx d'aide au diagnostic

B.4.3.1 Aide au diagnostic du lymphome par TEP/TDM

(i) *Problématique et contexte scientifique*

Le lymphome malin est une maladie diffuse se caractérisant généralement par un premier site ganglionnaire atteint qui s'étend ensuite de proche en proche à d'autres organes par voie lymphatique. Des atteintes extra-ganglionnaires surviennent aux stades avancés de la maladie, principalement dans la rate, le système ostéomédullaire, le foie, le poumon, le système cutané, le sein, le système digestif et la sphère ORL. De nombreuses études cliniques ont montré que la TEP permettait de diagnostiquer et caractériser les lésions cancéreuses à des stades plus précoces que l'imagerie anatomique conventionnelle et d'évaluer plus rapidement la réponse au traitement. Le protocole d'évaluation actuel se base sur le suivi d'un nombre réduit de cibles résultant d'un compromis entre le temps d'analyse visuelle des images TEP corps entier et les contraintes cliniques. Cette tâche est délicate et problématique, d'une part parce que le protocole actuel ne permet pas le suivi de toutes les lésions et d'autre part, parce qu'il est très difficile de discriminer une fixation résiduelle normale liée à l'inflammation, par exemple, d'une fixation pathologique liée à l'activité cancéreuse résiduelle.

Ce projet est né de la rencontre en 2011 avec Fabien Ricard, médecin nucléaire et assistant chef clinique au Centre Hospitalier Lyon Sud. Fabien Ricard s'intéressait, dans le cadre de sa thèse de médecine, à l'apport de l'imagerie TEP/TDM pour la prise en charge diagnostique et thérapeutique des patients atteints de lymphome. Sa thèse basée sur l'étude d'une série de données cliniques avait, en premier lieu, permis de confirmer la difficulté d'interprétation de certaines localisations suspectes en imagerie TEP au FDG. Il est souvent difficile, pour certaines pathologies cancéreuses, et en particulier pour le lymphome, de discriminer un processus inflammatoire ou physiologiques normal (Figure B.4-3) d'une fixation maligne (Figure B.4-4). Cette étude démontrait ensuite l'apport significatif de l'interprétation conjuguée de l'image anatomique fournie par la TDM afin, d'une part, de mieux localiser les zones pathologiques mais surtout, de mieux discriminer les zones de fixation suspectes en TEP. Cette étude reportait enfin les difficultés liées à l'analyse visuelle des deux modalités, et en particulier la durée allongée de l'analyse par rapport à l'interprétation de l'image TEP seule ainsi que la grande variabilité de performance inter-observateur en fonction du niveau d'expertise.

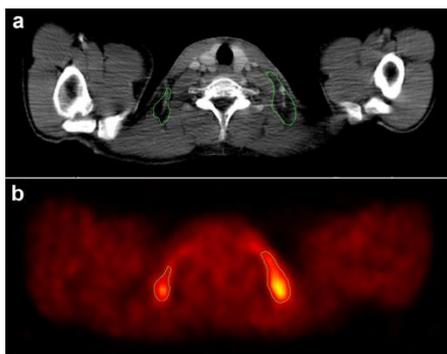


Figure B.4-3. Coupes transverses d'une acquisition TDM (a) et TEP (b) d'un patient présentant une captation de FDG bénigne dans la graisse brune supraclaviculaire.

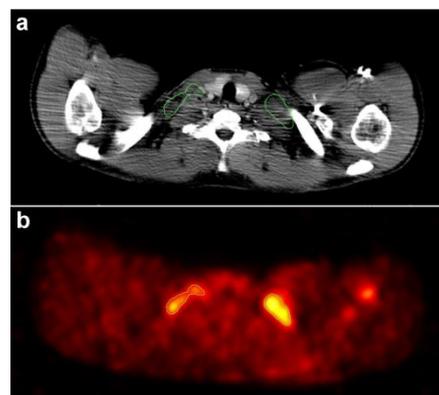


Figure B.4-4. Coupes transverses d'une acquisition TDM (a) et TEP (b) d'un patient présentant une captation de FDG maligne dans un ganglion supraclaviculaire.

Fabien Ricard souhaitait poursuivre ses travaux en développant une collaboration avec des spécialistes du traitement d'images dans l'objectif de produire une image paramétrique des zones suspectes qui synthétise l'information complémentaire des deux modalités TEP et TDM. L'enjeu médical prioritaire était de proposer un système d'aide à la décision qui fournisse un score de malignité fiable pour toute zone de l'image pointée par le médecin. La tâche de localisation des lésions suspectes n'était pas prioritaire dans le cahier des charges défini ensemble.

J'ai proposé de structurer ce projet dans le cadre de deux sujets de stages de 6 mois que j'ai encadrés. Fabien Ricard, dans le cadre d'un stage de master II (Ingénierie pour la santé et le médicament) s'est intéressé à la constitution de la base de données et l'étude statistique univariée des paramètres extraits comme potentiel biomarqueurs d'imagerie du lymphome. Matthieu Rogez, ingénieur à l'école centrale de Metz, dans le cadre de son projet de fin d'études, a travaillé sur l'extraction des caractéristiques discriminantes et l'analyse multivariée par classification supervisée pour la caractérisation des lésions. Ce projet a donné lieu à plusieurs communications dont une communication à la conférence IEEE ISBI en 2012 complétée par un article dans la revue IEEE JBHI en 2014. Le détail méthodologique et les résultats de cette étude sont présentés dans l'annexe B.8.3.5.

La suite de ce chapitre revient sur la discussion de certains points concernant la méthodologie et les résultats puis présente quelques éléments d'analyses complémentaires à la discussion proposée dans l'article IEEE JBHI en 2014 [Lartizien et al., 2014].

(ii) **Choix méthodologiques**

(a) **Système de classification**

Le système d'aide au diagnostic CADx a été construit sur le modèle classique présenté sur la Figure B.4-1. J'ai proposé d'utiliser un séparateur à vaste marge (SVM) dont le formalisme est présenté dans l'annexe B.7.4.2(iii); ce classifieur est en effet généralement bien positionné pour les tâches d'apprentissage sur des bases de données de petite dimension car il est, par construction (minimisation du risque structurel), robuste au phénomène du fléau de la dimension ; par ailleurs, il peut traiter le cas de données non linéairement séparables. Comme indiqué dans la section B.5.1, le modèle prédictif du SVM permet d'estimer, pour tout échantillon test, un score de malignité correspondant à la distance de cet échantillon à la frontière de décision. Cette mesure non bornée peut être *a posteriori* convertie en probabilité d'appartenir à la classe pathologique au moyen de méthodes de calibration, telle que l'algorithme de Platt, utilisé dans cette étude.

Les performances du SVM ont été comparées à celles d'une autre approche supervisée, les forêts aléatoires de décision (RDF pour random decision forest), dont les performances sont également bien positionnées dans le cas d'apprentissage à faible nombre de cas. Les résultats présentés à la conférence IEEE ISBI en 2012 [Lartizien et al., 2012] n'ont pas mis en évidence de différences significatives de performances entre les deux méthodes, je n'ai donc pas poursuivi l'investigation des RDF pour des raisons de coût calculatoire.

Pour cette étude, nous avons utilisé un SVM non linéaire et sélectionné le noyau gaussien défini dans l'annexe B.7.4.2(iii)(b). Les paramètres C et σ ont été optimisés par une méthode de recherche exhaustive sur grille.

(b) **Caractéristiques discriminantes**

L'originalité de notre approche a été de combiner des attributs utilisés dans la pratique clinique (par exemple, la mesure de l'indice SUV en TEP) à une large gamme de paramètres extraits de l'image de manière exhaustive. Cent cinq caractéristiques ont été extraites des images TEP et TDM comprenant des paramètres statistiques d'ordre 1 (e.g. coefficient de kurtosis) et 2 (e. g. entropie) calculés à partir de 4 types de matrice de texture dont le principe est illustré sur la Figure B.4-5:

- la matrice de cooccurrence (Gray level cooccurrence matrix, GLC) mesure les relations spatiales entre les niveaux de gris de voxels voisins dans l'image. Cette matrice est paramétrée par la distance d entre les couples de voxels considérés et l'orientation θ . Chaque élément de la matrice (i,j) contient

la probabilité empirique d'observer des couples de voxels (séparés d'une distance d selon θ) d'intensité respective i et j . Elle permet d'extraire les coefficients de Haralick.

-La matrice des longueurs de segments (Gray level run length Matrix, GLRL) effectue le dénombrement des longueurs des segments de voxels de même niveau d'intensité pour une direction donnée θ . Chaque élément de la matrice (i,j) contient la probabilité empirique d'observer des segments de j voxels et d'intensité i selon θ .

- La matrice des tailles de zones d'intensité homogène (Gray level intensity size zone, GLISZ) représente le dénombrement régions de l'image de même taille et de même niveau d'intensité. Chaque élément de la matrice (i,j) contient la probabilité empirique d'observer des zones homogènes constituées de j voxels et d'intensité i .

- La matrice de différences d'intensité sur le voisinage (Gray level difference matrix, GLD) compare l'intensité d'un voxel avec l'intensité moyenne de ses voisins. Chaque élément de la matrice (i,j) contient le nombre de voxels d'intensité i et dont le voisinage a une intensité moyenne j .

Le choix des paramètres statistiques d'ordre 2, dits de texture, a été guidé par les résultats d'études récentes montrant le pouvoir discriminant de certains d'entre eux pour la caractérisation par imagerie TEP au FDG de plusieurs types de cancers [El Naqa et al., 2009, Yu et al., 2009]. Un autre type d'études soulignait parallèlement l'existence de corrélations entre certaines de ces mesures en TEP et le type de réponse au traitement (répondeur, stable, non répondeur) pour certains cancers [Tixier et al., 2011]. Ces travaux supposent que les variations de certains paramètres physiologiques à l'échelle cellulaire (signature moléculaire), tels que la microvascularisation, peuvent induire des variations de la texture de l'image (marqueurs image) et ainsi permettre de discriminer des zones pathologiques et saines. Cette hypothèse intéresse aujourd'hui la communauté scientifique et fait l'objet de controverses que nous abordons à la fin de ce chapitre. Dans cette étude, nous n'avons pas cherché à investiguer la validité cette hypothèse mais uniquement à répondre à la double question suivante: 1) les paramètres de texture permettent-ils d'améliorer la spécificité de détection dans le cas du lymphome ? et 2) l'imagerie TEP au FDG du lymphome est-elle caractérisée par une signature textuelle spécifique ?

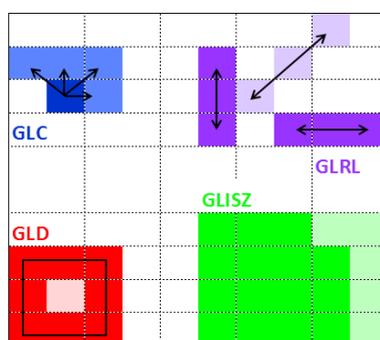


Figure B.4-5 Illustration du principe de calcul des quatre matrices de texture

Le système CADx a été développé à partir de la base de données cliniques collectée par Fabien Ricard, constituée de 25 acquisitions TEP/TDM de patients atteints de lymphome avant traitement. Ces données ont été annotées manuellement et ont permis d'extraire 156 volumes d'intérêt (VOI) (empilement de régions d'intérêt 2D tracées sur des coupes axiales) tracées sur des zones cancéreuses et 32 VOIs tracées sur des zones inflammatoires. Rappelons que notre objectif était de discriminer des zones de fixations suspectes mais non cancéreuses de zones cancéreuses. Nous n'avons donc pas utilisé de régions de référence saines. Une originalité de notre approche a été d'implémenter le calcul des matrices de texture en 3 dimensions sur les formes irrégulières des volumes d'intérêt, contrairement à l'approche classique basée sur le calcul de ces matrices sur des formes régulières et dans des images de dimension 2.

La taille limitée de la base de données est une caractéristique fréquente en imagerie médicale, et dont il faut tenir compte dans les problématiques d'apprentissage supervisé. Il faut en particulier éviter l'écueil du fléau de la dimension ('curse of dimensionality') caractérisé par le phénomène de sur-apprentissage du modèle lorsque la dimension de l'espace des caractéristiques (p) est grande devant le nombre d'échantillons n ($p \gg n$) (cf annexe B.7.4.2).

Une méthode classique pour s'affranchir du fléau de la dimension consiste à réduire la dimension des données par des méthodes de sélection ou d'extraction de caractéristiques. Nous avons, dans cette idée, proposé une approche hybride originale de sélection de caractéristiques combinant six méthodes standard de sélection par filtre [Guyon and Elisseeff, 2003]. L'ensemble des 105 caractéristiques a été classé par ordre décroissant de pouvoir discriminant selon chacun des 6 critères préalablement définis et incluant, entre autre, une mesure d'aire sous la courbe ROC. Chacune des caractéristiques est ainsi caractérisée par un ensemble de 6 valeurs correspondant à son ordre de classement pour chaque critère. La sélection est ensuite effectuée en retenant uniquement les attributs pour lesquels l'ensemble des 6 rangs contient au moins deux valeurs inférieures à 10. Autrement dit, la méthode sélectionne les attributs classés au moins 2 fois sur 6 dans les dix paramètres plus discriminants. Cette approche, référencée par le terme TOP10, a été comparée à une autre méthode filtre, l'analyse en composantes principales (PCA) et à une méthode enveloppante basée sur un algorithme génétique (GA).

(c) Evaluation des performances

Le système de classification par SVM renvoie, pour chaque région test, caractérisé par un vecteur de caractéristique discriminante de dimension 105 (sans sélection) ou inférieure (après sélection), une probabilité d'appartenance à la classe pathologique. Les performances quantitatives de détection de ce système ont été évaluées à l'aide d'une méthode classique de type ROC, comme détaillée dans l'annexe B.7.2.1. La courbe ROC a été tracée à partir des 188 échantillons constitués de 132 cas positifs (ROI tracées sur des lésions cancéreuses) et 56 cas négatifs (ROI tracées sur des zones de fixation suspectes). Nous avons utilisé pour cette évaluation la méthode du leave-one-out (LOO) décrite dans la section B.4.2.1(iii) qui permet d'obtenir une mesure faiblement biaisée des performances que l'on pourrait obtenir à l'aide d'une base de données test de dimension infinie (généralisation des performances) [Sahiner et al., 2008]. L'aire sous la courbe ROC, AUC_{LOO} a été utilisée comme métrique de performance. La mesure de AUC_{LOO} a été comparée à la mesure obtenue par une autre approche proposée par Fukunaga et Hayes qui fournit, quant à elle, une mesure pessimiste des performances en généralisation, AUC_{FH} [Sahiner et al., 2008]. Nous renvoyons à la description synthétique de ces deux méthodes dans l'article [Lartzien et al., 2014]. L'aire sous la courbe ROC a été estimée par la méthode non-paramétrique de Wilcoxon et nous avons choisi l'approche de Dorfman, Berbaum et Metz (DBM) pour la recherche de différences significatives entre les différentes aires sous les courbes ROC [Dorfman et al., 1992].

Différentes combinaisons de caractéristiques TEP et TDM ont été évaluées afin de quantifier l'apport de l'approche bimodale, d'une part et l'apport des paramètres de texture, d'autre part. Les cinq combinaisons suivantes ont ainsi été analysées : 1) caractéristiques TEP et TDM avec sélection par la méthode TOP10 2) caractéristiques TEP et TDM sans sélection, 3) caractéristiques TEP seules avec sélection, 4) caractéristiques TEP seules sans sélection, 5) caractéristiques TEP et TDM d'ordre 1 uniquement (sans prise en compte des paramètres de texture).

(iii) Résultats

Cette section souligne les résultats principaux de cette étude. La méthode de sélection de caractéristiques TOP10 appliquée à la série de 105 caractéristiques a permis de sélectionner un sous-ensemble des 9 caractéristiques TEP et 3 caractéristiques TDM les plus discriminantes. Ces caractéristiques comprennent les valeurs de SUV utilisées classiquement en routine clinique ainsi que le coefficient de kurtosis mesuré sur la TEP et 5 paramètres de texture. La même méthode appliquée à la série des 53 caractéristiques TEP a permis de sélectionner 12 caractéristiques TEP comprenant à nouveau la SUV et le coefficient de Kurtosis ainsi que 6 paramètres extraits des 4 types de matrice de

texture considérés. Ces deux séries sont reportées dans le Tableau B.4-1. Les acronymes utilisés pour les paramètres de texture sont explicités dans l'annexe B.8.3.5.

Tableau B.4-1 Caractéristiques sélectionnées par la méthode hybride TOP sur les séries de 105 paramètres TEP et TDM (gauche) et de 53 paramètres TEP (droite)

PET/TDM	PET
PET kurtosis	SUV min
PET cluster prominence	SUV std
TDM RL LRLGE	Kurtosis
SUV min	Cluster prominence
SUV mean	Sum entropy
SUV peak	SUV max
PET sum entropy	SUV mean
SUV max	SUV peak
TDM mean	ISZ GLN
SUV std	RL HGRE
TDM ISZ LRLGE	ISZ SGLGE
PET ISZ LRHGE	ISZ LRHGE

Tableau B.4-2 Performances quantifiées par l'AUC estimée par validation croisée (LOO) avec les différents types de caractéristiques

	AUC	95% IC*	Nb of Caract.	Durée (min)
PET/TDM	0.878	(0.799,0.955)	105	37
PET/TDM-TOP10	0.911	(0.844,0.978)	12	10
PET	0.894	(0.827,0.961)	53	21
PET-TOP10	0.894	(0.821,0.965)	12	10
PET/CT 1 st order	0.871	(0.797,0.944)	15	12
PET 1 st order	0.825	(0.724,0.926)	8	12
PET/TDM GLC _{d=3}	0.881	(0.788,0.975)	105	37
PET/TDM GLC _{d=3} -TOP10	0.860	(0.761,0.950)	14	11
PET GLC _{d=3}	0.880	(0.802,0.957)	53	21
PET GLC _{d=3} -TOP10	0.893	(0.819,0.967)	12	10

*IC = intervalle de confiance

Le Tableau B.4-2 reporte les valeurs d'aire sous la courbe ROC mesurée par la méthode LOO pour les différentes combinaisons de caractéristiques. Ces valeurs, majoritairement supérieures à 0.89, traduisent un bon niveau de performances intrinsèques du système CADx. Notons que les résultats obtenus par la méthode d'évaluation de Fukunaga et Hayes sont du même ordre de grandeur, ce qui renforce la fiabilité de cette mesure (cf article [Lartzien et al., 2014]).

Les meilleures performances, caractérisées par la valeur de l'aire sous la courbe (AUC_{LOO}), sont obtenues avec l'ensemble des 12 caractéristiques TEP et TDM sélectionnées par la méthode TOP10 ($AUC=0.91$). Elles surpassent les performances obtenues en considérant uniquement les caractéristiques TEP ($AUC=0.894$) ce qui justifie l'intérêt de la fusion d'informations extraite des deux modalités. On peut également constater que la prise en compte des caractéristiques de texture améliore les performances (AUC entre 0.825 et 0.871 en considérant les caractéristiques d'ordre 1 uniquement). L'analyse statistique par la méthode DBM n'a cependant pas permis de mettre en évidence de différence significative de performance entre ces différentes stratégies de CADx.

(iv) Discussion – Perspectives

Comme souligné dans l'introduction de ce chapitre, la manipulation des méthodes de classification supervisée est délicate, principalement en raison du risque de surapprentissage. Les études multiparamétriques sur données synthétiques menées principalement par les équipes de l'Université du Michigan aux Etats-Unis visent à fixer des règles de dimensionnement des études pour la mesure des performances intrinsèques en généralisation en fonction de nombreux paramètres tels que la taille de la base d'apprentissage, la dimension de l'espace des attributs, le degré de corrélation entre

attributs etc..[Sahiner et al., 2008]. Ces recommandations ne sont pas facilement transposables en clinique car elles reposent sur des a priori forts sur les propriétés statistiques des caractéristiques et ne sont pas toujours compatibles avec la réalité pratique, notamment concernant la taille des bases d'apprentissage. Dans cette étude, nous avons donc cherché à définir un cadre méthodologique fiable permettant de mesurer une estimation faiblement biaisée des performances en généralisation.

Sur le plan applicatif, notre étude s'intéressait à proposer un système d'aide au diagnostic pour l'imagerie du lymphome. L'optimisation des performances intrinsèques de cet algorithme nous a parallèlement conduits à évaluer l'apport de la bimodalité TEP/TDM et des caractéristiques de texture.

Je reprends dans la suite quelques points spécifiques de discussion sur les limites de l'étude et les pistes possibles pour enrichir ce travail.

(a) Cadre méthodologique

(i) Evaluation des performances intrinsèques

Cette étude a permis de mesurer les performances intrinsèques d'un système original CADx pour la caractérisation des lésions lymphomateuses en imagerie TEP/TDM. Ce travail constitue la première étape définie dans le protocole d'évaluation recommandé par le groupe d'expert de l'AAM (American Association of Medical Physicist) chargé de ce dossier [Petrick et al., 2013]. Le niveau de performance obtenu sur une base de 25 patients est très encourageant (AUC = 0.91). Comme nous l'avons souligné à plusieurs reprises, ce système n'est pas conçu pour remplacer le clinicien, mais pour l'assister dans son interprétation diagnostique. Ainsi, seule une étude ROC humaine basée sur la lecture, avec ou sans l'aide du système CAD, d'une série d'images clinique équivalente pourrait évaluer son impact réel sur la pratique clinique. Nous n'avons pas réalisé cette étude qui nécessite de mobiliser un ensemble de relecteurs du domaine médical et peut s'avérer lourde et complexe à mettre en œuvre. Cette étape permettrait de valider l'intérêt du système CAD et d'envisager une stratégie de valorisation, s'il était démontré.

(ii) Sur la généralisation des performances

Nous avons choisi d'évaluer les performances de détection à l'aide de la méthode LOO qui correspond au cas extrême des méthodes par validation croisée pour lequel le nombre de partitionnements de la base (K fold) est égal au nombre N d'échantillons. Cette méthode est une référence dans le domaine de l'évaluation des systèmes CAD et est utilisée dans une vaste majorité d'article. Elle fournit théoriquement une approximation faiblement biaisée des performances en généralisation [Sahiner et al., 2000, Sahiner et al., 2008]. Cependant, certaines études théoriques ont montré qu'elle pouvait introduire un biais positif de performances lorsque la sélection de caractéristiques et l'optimisation des paramètres du classifieur n'étaient pas intégrées au processus itératif [Varma and Simon, 2006, Tibshirani and Tibshirani, 2009]. Idéalement, et d'après le formalisme des nested CV, il aurait donc fallu, pour chaque sous-ensemble d'apprentissage A_k de dimension $N-1$, avec $k=1\dots N$:

- appliquer la méthode de sélection de caractéristiques sur A_k
- imbriquer une boucle de validation croisée permettant de sélectionner le meilleur couple de paramètres (C_k, γ_k) pour A_k
- Apprendre le modèle M_k correspondant au couple (C_k, γ_k) sur A_k
- Appliquer le modèle M_k sur l'échantillon de test T_k

Cet algorithme fait immédiatement apparaître la contrainte calculatoire rédhibitoire à son application. Il faudrait, en effet, réaliser N fois la recherche exhaustive des paramètres (C, γ) du SVM. Afin d'estimer le biais potentiel dans la mesure réalisée par validation croisée, j'ai proposé d'utiliser une méthode d'évaluation alternative, la méthode de Fukunaga et Hayes. Concernant la sélection des caractéristiques, j'ai étudié la reproductibilité de la méthode de sélection TOP10 en réalisant plusieurs tirages aléatoire de populations de caractéristiques. Les résultats positifs de ces deux tests ont permis de fiabiliser la mesure de performances intrinsèques publiées dans cet article.

(b) Analyse des résultats

(i) Apport de la bimodalité

Cette étude souligne que les caractéristiques TEP classiques utilisées par le médecin (SUV) restent les plus discriminantes (cf Tableau B.4-1). La combinaison des deux modalités permet d'améliorer les performances globales de détection, nous n'avons cependant pas pu démontrer un gain statistiquement significatif.

(ii) Apport des paramètres de texture pour la classification

Nous avons choisi d'extraire un large spectre de caractéristiques statistiques sans être guidé par aucun a priori clinique ni résultat de la littérature. Notre étude démontre que la prise en compte des caractéristiques de texture dans une analyse multivariée permet d'améliorer les performances intrinsèques du système CADx (AUC=0.894 avec prise en compte des paramètres statistiques d'ordre 2 contre 0.825 sans prise en compte). Nous n'avons cependant pas pu démontrer de différence significative de performance.

La quantification de la texture est un sujet sensible en imagerie TEP ces cinq dernières années. Comme je l'ai souligné en introduction, les enjeux actuels concernent essentiellement leur rôle potentiel de marqueur de l'agressivité tumorale et de la réponse au traitement pour différentes pathologies, en particulier le cancer de l'œsophage [Tixier et al., 2011, Orhac et al., 2014, Tan et al., 2013], le cancer du sein [Willaime et al., 2013], et le cancer du poumon [Cook et al., 2013]. Ces études reposent sur l'analyse statistique univariée du pouvoir pronostic de chaque paramètre pour un type de cancer spécifique. Certaines d'entre elles produisent des résultats controversés, les unes montrant le rôle de certains paramètres comme marqueur de la réponse thérapeutique, d'autres soulignant leur faible reproductibilité et mettant en garde contre l'utilisation abusive de ces indices en clinique.

Notre étude se place dans une perspective différente. Comme les études de El Naqa [El Naqa et al., 2009] et Yu [Yu et al., 2009], l'objectif était de proposer une analyse statistique multivariée pour mieux caractériser les lésions visibles en imagerie au stade du diagnostic. Il ne s'agissait donc pas de sélectionner un ou plusieurs marqueurs de texture mais d'évaluer si certaines de ces caractéristiques permettaient d'améliorer la discrimination entre les lésions suspectes mais non malignes et les lésions lymphomateuses. Cela nous a conduits à considérer un panel large de 92 paramètres issus des 4 matrices de texture illustrées sur la Figure B.4-5 et décrites dans l'article répertorié dans l'annexe B.8.3.5. Tixier et al considèrent que l'on peut classer ces paramètres en deux catégories, la première traduisant l'hétérogénéité locale (GLC et GLD), la seconde considérant les paramètres régionaux (GLISZ et GLRL).

Ces matrices sont régies par différents paramètres tels que l'échelle de discrétisation en niveau de gris qui définit la dimension de la matrice (voir section B.4.3.1(ii)(b)). Guidée par la méthode empirique proposée par Tixier et al, j'ai considéré qu'une discrétisation de la dynamique des images TEP et TDM en 16 niveaux permettait de conserver une précision satisfaisante des variations locales de fixation de FDG et d'atténuation tissulaire, respectivement.

Les matrices GLC et GLD sont également paramétrées par la distance d caractérisant la taille du voisinage local de chaque voxel. Dans cette étude, j'ai privilégié des valeurs de d faibles qui caractérisent les variations très locales d'homogénéité. On s'attend, en effet, à obtenir un 'désordre' plus important dans les zones tumorales. Il n'y a donc pas de réelle motivation à calculer les paramètres de Haralick pour des voxels spatialement très espacés. J'ai néanmoins rajouté dans le tableau les performances obtenues en augmentant sensiblement cette distance ($d = 3$ correspondant à $GLC_{d=3}$ dans le tableau Tableau B.4-2). Les résultats indiquent des performances identiques ou légèrement inférieures (sans différence significative), ce qui confirme notre motivation première de considérer des valeurs faibles.

Florent Tixier, dans le cadre de sa thèse, a évalué la reproductibilité de la mesure de ces paramètres pour les images TEP uniquement en fonction de l'échelle de discrétisation des niveaux de gris de l'image et de la méthode de segmentation utilisée. Il montre que les paramètres locaux extraits de la matrice de cooccurrence de Haralick sont les plus stables. Les métriques issues des matrices GLISZ

et GLRL, qui contiennent l'information régionale de texture sont en revanche beaucoup plus sensibles. Des résultats récents reproduisant le paradigme proposé par Tixier pour d'autres types de cancer confirment ces observations [Orlhac et al.]. Les résultats du Tableau B.4-1 montrent que les coefficients dérivés des matrices GLSZ et GLR sont présents dans la liste des caractéristiques sélectionnées par la méthode TOP 10. On peut cependant noter, dans cette liste organisée par ordre décroissant d'importance, que ces paramètres se classent dans les derniers rangs, après certaines métriques de Haralick telles que l'entropie ou le paramètre de 'cluster prominence' et ne figurent donc pas parmi les plus significatifs. On peut ainsi supposer que la variabilité de mesure de ces paramètres n'impacte pas significativement les performances du CAD, contrairement aux études de marqueur pronostique basées sur une analyse univariée. Une étude exhaustive permettrait cependant de tester la validité de cette hypothèse.

Bien que notre étude ne cible pas la recherche de biomarqueurs image, nous nous sommes intéressés à l'analyse comparative des caractéristiques mises en lumière par les différents types d'études, qu'elles soient basées sur une analyse multivariée [El Naqa et al., 2009, Yu et al., 2009] ou univariée [Tixier et al., 2011, Orhac et al., 2014]. Nous renvoyons à la discussion de l'article [Lartzien et al., 2014] sur ce point. Cette discussion souligne que les paramètres discriminants ne sont pas similaires d'une étude à l'autre. On peut donc s'interroger sur leur sens biologique. Une étude plus poussée serait intéressante mais très difficile à mettre en œuvre. L'idée serait de corréliser les 'marqueurs de l'image' avec des marqueurs biologiques du cancer. L'accès à ces informations est cependant très difficile et nécessiterait de développer des collaborations avec des équipes de biologistes et de statisticiens. A la résolution de la TEP (entre 2 et 5 mm), on ne peut cependant pas s'attendre à visualiser un processus physiologique, tel que la néoangiogenèse par exemple, mais on peut espérer caractériser l'hétérogénéité de ce processus au niveau tumoral, puis rechercher des corrélations avec différents marqueurs de l'agressivité tumorale. En 2007, une étude intéressante sur le cancer du foie a montré que certains paramètres radiologiques étaient corrélés aux profils d'expression génétique de tissus hépatiques sains et cancéreux [Segal et al., 2007]. On peut suspecter que la lourdeur de mise en œuvre de telles études constitue un frein significatif justifiant la rareté des publications de ce type.

Quoiqu'il en soit, les résultats de cette étude montrent que les paramètres de texture se retrouvent parmi les caractéristiques les plus discriminantes pour la problématique de classification binaire 'complexes' (tissu normal mais suspect contre tissu sain) considérée. Ces mêmes coefficients de Haralick se retrouvent également très bien positionnés dans la seconde application présentée dans le chapitre suivant. De mon point de vue, ces attributs ne doivent donc pas être écartés dans le cadre d'analyse statistique multivariée en imagerie du cancer.

B.4.3.2 Cartographie du cancer de prostate par IRM multi-modalités

(i) Problématique et contexte scientifique

Le cancer de prostate est le cancer le plus fréquent et la troisième cause de mortalité par cancer chez l'homme dans les pays développés. Alors que l'imagerie s'est longtemps concentrée sur le bilan d'extension local du cancer (franchissement capsulaire, envahissement ganglionnaire), la communauté médicale et scientifique s'intéresse aujourd'hui au développement de méthodes d'imagerie permettant de disposer d'informations précises sur la position exacte des foyers tumoraux dans la glande, ainsi que sur leur agressivité. Cette cartographie permettrait de proposer une solution alternative ou complémentaire à la biopsie en aveugle, qui reste à l'heure actuelle la référence diagnostique malgré son caractère invasif et sa faible sensibilité. Ces cartes permettraient également à moyen terme de pouvoir réaliser la surveillance active de cancers non agressifs et seraient utiles pour guider les thérapies locales.

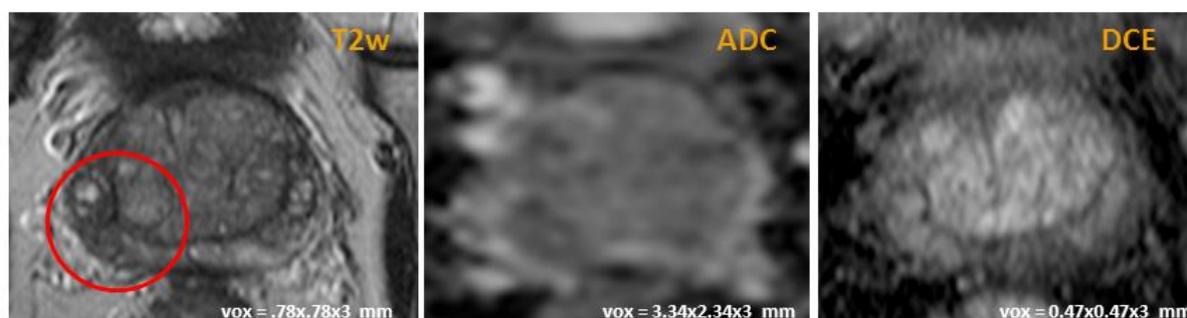


Figure B.4-6. Coupes axiales d'acquisitions IRM d'un patient obtenues à l'aide de 3 séquences IRM : une séquence T2 pondérée à gauche, une séquence de diffusion au milieu et une séquence T1 dynamique avec agent de contraste à droite. Le cercle rouge indique la présence d'une anomalie dans la zone périphérique.

L'IRM est actuellement la méthode qui permet le mieux de localiser le cancer dans la prostate. Des études récentes ont montré que la combinaison de plusieurs séquences d'imagerie permettrait d'améliorer les performances diagnostiques. D'où le concept récent d'IRM multiparamétrique qui consiste à associer, dans un même examen, plusieurs types d'acquisition, comme, par exemple, une acquisition pondérée en T2, une acquisition T1 dynamique avec injection d'un agent de contraste au gadolinium (DCE pour dynamic contrast enhanced, en anglais), et une imagerie de diffusion (ADC pour apparent diffusion coefficient). Reste la problématique souvent complexe pour le radiologue de fusionner ces informations au cours du diagnostic visuel, comme illustré sur la Figure B.4-6. Le Pr Olivier Rouvière, PUPH et chef du service d'imagerie urologique et vasculaire de l'hôpital Edouard Herriot à Lyon, est un spécialiste de cette problématique. Ce domaine fait également l'objet de recherches actives dans différents centres dont les principaux compétiteurs sont l'équipe de Viswanath et Madabushi à la Rutgers University aux Etats-Unis [Viswanath et al., 2009], l'équipe de la Radboud University Medical Centre Nijmegen Pays-Bas [Vos et al., 2008, Vos et al., 2010, Hambrock et al., 2013], l'équipe de l'Université de Chicago [Peng et al., 2013] et celle de l'Université de Toronto [Langer et al., 2009, Ozer et al., 2010].

Ce projet de système d'aide au diagnostic pour l'imagerie IRM multiparamétrique de la prostate a été initié en novembre 2009 par le démarrage de la thèse d'Emilie Niaf que j'ai co-encadrée avec Olivier Rouvière. Il marque un point remarquable de mon parcours scientifique puisqu'il n'est plus relié à l'imagerie TEP. Il s'est parallèlement structuré grâce à deux financements obtenus dans le cadre de l'appel d'offres PAIR Prostate de l'INCa (2010-13) et du programme SIRIC (2012-17). Ces programmes sont présentés dans la section A.6.1. L'objectif *in fine* de cette collaboration est d'établir, à partir des données de l'IRM multiparamétrique, une cartographie précise de la position, du volume et de l'agressivité des foyers cancéreux dans la prostate. La thèse de Mlle Niaf constituait un première avancée dans cette direction avec comme objectif de proposer un système CADx qui permette de

discriminer les lésions bénignes suspectes des lésions cancéreuses, sur le modèle des travaux présentés dans le chapitre B.4.3.1 pour l'imagerie TEP/TDM du lymphome. Nous avons ainsi suivi un formalisme classique basé sur le modèle de la Figure B.4-2 et extrait plus de 100 caractéristiques à partir de trois types de séquences IRM : une acquisition pondérée en T2, une acquisition T1 dynamique avec injection d'un agent de contraste au gadolinium (DCE), et une imagerie de diffusion (ADC). Ce travail reposait sur une base d'images acquises sur des patients atteints d'un cancer de la prostate et subissant une prostatectomie. La base CLARA-P pour 'Corrélation Anatomoradiologiques IRM de Prostate' a été initiée en septembre 2008 par O. Rouvière après déclaration au comité de protection des personnes (CPP) sud-est IV sous la référence L 09-04 et à la commission nationale de l'informatique et des libertés (CNIL) sous la référence N° 08-06. Elle recueille, pour chaque patient, la description de toutes les zones suspectes sur toutes les séquences IRM (position, taille, forme, signal, etc...) ainsi que le résultat de l'analyse histologique de la zone correspondante sur la pièce de prostatectomie (cancer ou non, score de Gleason, etc...). Cette base comptait un quarantaine d'examen cliniques réalisés sur une IRM 1.5 T (GE Healthcare), au démarrage de la thèse d'Emilie Niaf.

Le premier point fort de l'étude menée par Emilie Niaf a été de réaliser une analyse exhaustive des performances de différents schémas de systèmes CAD combinant 6 méthodes d'extraction de caractéristiques et 4 classifieurs, sur la même base de données. Cette étude a permis de fournir une référence de comparaison fiable par rapport aux publications plus hétérogènes sur le sujet. Je reviens sur l'analyse comparative de ces résultats dans la discussion. Les résultats de ce travail ont été publiés dans la revue *Physics in Medicine and Biology* en 2012 [Niaf et al., 2012]. L'autre point fort de ce travail de thèse a été de proposer une évaluation complète des performances du CADx incluant les performances intrinsèques et l'analyse de l'impact clinique de cet outil sur les performances de radiologues juniors et seniors. Ce travail a nécessité la mobilisation à la fois de l'équipe clinique pilotée par Olivier Rouvière mais aussi d'une équipe de statisticiens des HCL pour nous aider à paramétrer l'étude et analyser les performances des lecteurs. Ce travail ambitieux a fait l'objet d'un article récent dans la revue *Radiology* [Niaf et al., 2014].

La suite de ce chapitre discute certains points méthodologiques ainsi que les principaux résultats.

(ii) Choix méthodologiques

(a) Design et évaluation des performances intrinsèques

Comme souligné dans le paragraphe précédent, la portée scientifique de cette étude repose en partie sur la base de données CLARA-P constituée d'images annotées et mises en correspondance avec une référence histologique. L'évaluation des performances du CADx a été réalisée sur une base de 30 patients. Un total de 42 lésions malignes, 46 lésions suspectes et 124 régions saines ont été contourées sur les images de prostate par deux radiologues après mise en correspondance des images IRM avec les lames histologiques produites à partir des pièces de prostate prélevées. L'analyse comparative des résultats de la littérature dans ce domaine souligne que les caractéristiques de cette base ne sont pas égalées si l'on considère le nombre et la précision des annotations. Les travaux de l'équipe de Vos et al, par exemple, s'appuient sur des bases de plus grande taille mais ne disposent pas de l'analyse anatomopathologique comme vérité terrain [Vos et al., 2008, Vos et al., 2010]. L'étude exhaustive réalisée par Emilie Niaf a tout d'abord consisté à assister le travail des radiologues, particulièrement pendant les séances de relecture et de mise en correspondance des pièces anatomo-radiologiques.

A partir d'une synthèse des données publiées par les différentes équipes citées dans le paragraphe précédent, Emilie Niaf a implémenté l'extraction de 145 caractéristiques descriptives extraites des 3 séquences d'images IRM citées précédemment, T2, DCE et ADC. Comme pour l'étude TEP/TDM du lymphome, ces caractéristiques incluent des statistiques du 1^{er} et 2^{ème} ordre, ainsi que des mesures de gradients. Notons que seuls les coefficients de Haralick extraits de la matrice GLC ont été évalués. Elle inclut également un ensemble de paramètres dynamiques extraits de l'analyse par modélisation compartimentale des données T1 dynamiques (volume fractionnel, coefficients de transfert) ainsi que des mesures semi-quantitatives telles que la pente des courbes cinétiques, l'aire sous la courbe de distribution cinétique du traceur. La Figure B.4-7 a) illustre un exemple de coupes transverses de

prostate présentant deux lésions cancéreuses (a) ainsi que les cartes paramétriques dérivées des images dynamiques (b, d et e) et l'image histologique de référence (c).

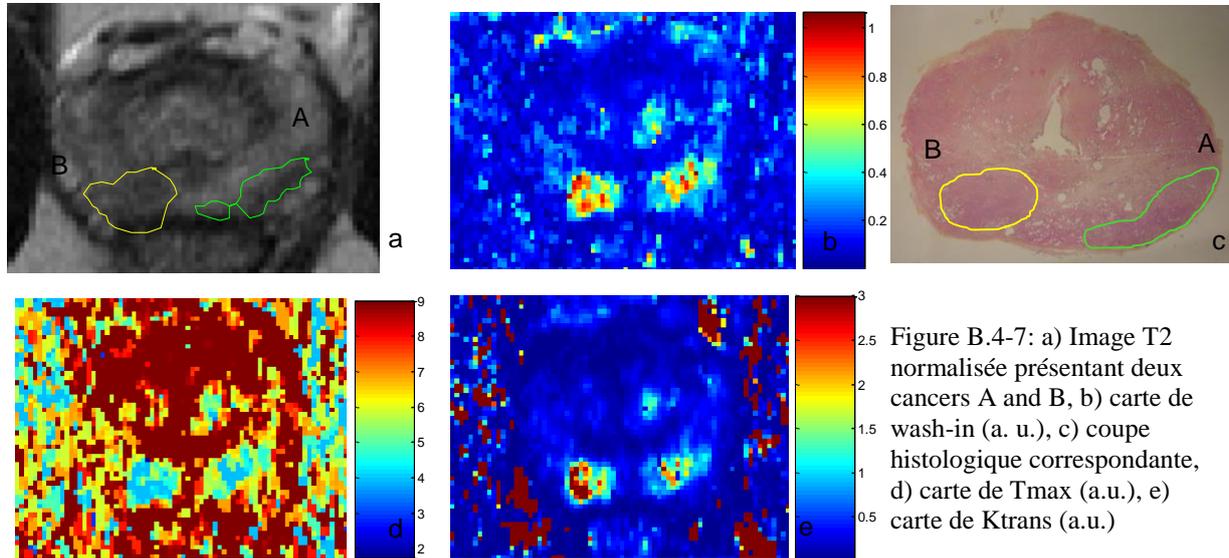


Figure B.4-7: a) Image T2 normalisée présentant deux cancers A and B, b) carte de wash-in (a. u.), c) coupe histologique correspondante, d) carte de Tmax (a.u.), e) carte de Ktrans (a.u.)

Emilie Niaf avait souhaité initialement privilégier les méthodes de sélection de caractéristiques de type par filtre, qui sont peu chronophages car découplées du classifieur (voir section B.4.2.1(ii)(b)). Ce type de méthodes consiste, dans un premier temps, à réaliser un classement des caractéristiques les plus discriminantes, selon un critère prédéterminé et indépendamment du classifieur, puis à sélectionner un sous-ensemble de ces caractéristiques en fixant un seuil sur ce critère. Le problème de ces méthodes est de fixer un seuil, par exemple sur la valeur du seuil de significativité p si le critère de classement retenu est un test de Student. Afin de s'affranchir d'un seuillage arbitraire, Emilie Niaf s'est alors tournée vers une méthode hybride qui consiste à fixer le seuil en fonction d'un critère intégrant les performances du classifieur. Cette méthode proposée par Peng et al [Peng et al., 2005] apparaît comme une alternative intéressante entre les méthodes filtres et les méthodes enveloppantes (wrapper) qui nécessitent une mise en œuvre lourde, comme nous le discutons dans le paragraphe B.4.3.2(iv). L'idée est de déterminer les performances de classification obtenues par différents sous-groupes de caractéristiques obtenus pour différentes valeurs de seuil sur le critère de classement. La Figure B.4-8 montre, par exemple, les performances obtenues avec un SVM à partir de caractéristiques ordonnées suivant la valeur du critère mRMRq qui mesure le rapport entre la redondance (mesure de corrélation aux autres caractéristiques) et la pertinence (mesure de corrélation de chaque caractéristique avec l'étiquette de classe d'appartenance). Le critère de performance utilisé dans cette illustration est l'aire sous la courbe ROC (AUC). Elle atteint sa valeur maximale en sélectionnant les 10 premières caractéristiques ordonnées suivant le critère mRMRq.

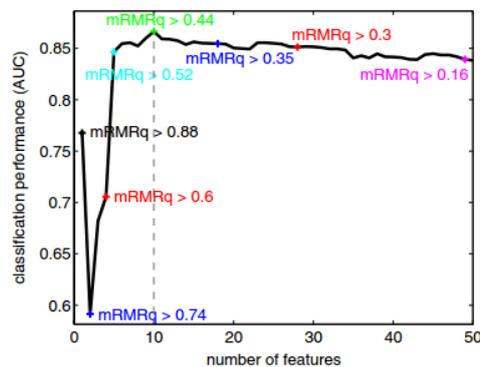


Figure B.4-8. Performances de classification par SVM en fonction du nombre de caractéristiques sélectionnées par le critère du mRMRq (minimum redundancy-maximum relevancy).

Pour cette étude, Emilie Niaf a considéré les 4 méthodes filtres suivantes : le test de Student, l'information mutuelle ainsi que deux métriques combinant des mesures de redondance et de pertinence calculées à partir de l'information mutuelle et qui permettent de tenir compte des corrélations entre les caractéristiques [Peng et al., 2005]. Quatre types de classifieurs ont été évalués: SVM, ADL, k plus proches voisins (KNN) et le classifieur naïf de Bayes (NB). Les principes du SVM et de l'ADL sont décrits dans l'annexe B.7.4. Nous renvoyons à la lecture de l'article [Niaf et al., 2012] pour une présentation plus précise des deux autres algorithmes.

Comme pour l'étude du système CADx pour l'imagerie TEP/TDM du lymphome, les performances des différentes configurations de CAD ont été mesurées à l'aide de courbes ROC en utilisant une approche par validation croisée de type LOO. Une analyse des différences significatives entre les aires sous les différentes courbes ROC expérimentale a été réalisée à l'aide du logiciel ROCKIT (version 1.1-beta, CE Metz, Université de Chicago) qui repose sur un modèle de distribution binormale de la variable de décision.

(b) Evaluation de l'impact clinique du système CADx

Comme nous l'avons souligné en introduction, le point fort de cette étude a été de mener une évaluation complexe du système CAD en mesurant son impact clinique sur une série de douze radiologues de différents niveaux d'expertise, un groupe de 7 lecteurs avec moins de 1 an d'expérience et un groupe de 5 lecteurs avec un niveau d'expertise entre 1 et 7 ans. Trois séries de lecture des images IRM ont été réalisées. Pour chacune, les lecteurs visualisaient simultanément les trois séquences d'images (à l'aide du logiciel OSIRIX) qui reproduit l'environnement des consoles cliniques. Chaque lecteur devait attribuer un score compris entre 0 et 4 (Echelle de Lickert) à chacune des 88 régions d'intérêt contourées sur les trois séquences des 30 patients de l'étude. Une première lecture (R1) des 30 examens a été réalisée sans l'aide du système CAD. Après 5 semaines d'intervalle, délai estimé pour s'affranchir d'un éventuel effet mémoire, une nouvelle séance de lecture était menée en 2 phases. La première phase de lecture (R2) reproduisait la séance de lecture initiale sans l'aide du CAD, ceci afin de quantifier un éventuel effet d'apprentissage. Elle était immédiatement suivie d'une troisième lecture (R3) au cours de laquelle le lecteur était informé du score attribué par le CAD ainsi que de son score de seconde lecture (R2), le laissant libre de modifier cette valeur.

L'analyse statistique des performances du groupe de lecteurs a été réalisée par Murielle Rabilloud et Laurent Roche du service de Biostatistique des Hospices Civils de Lyon. Une première analyse des courbes ROC a été réalisée en considérant d'abord une modélisation individuelle des courbes de chacun des douze lecteurs suivant un modèle de régression proposé par Alonzo et Pepe qui permet de paramétrer un modèle de courbe ROC en fonction de différentes covariables [Alonzo and Pepe, 2002]. Dans notre étude, nous avons considéré le lecteur et le type de lecture comme des effets fixes. Dans un deuxième temps, un modèle similaire a été utilisé en considérant cette fois le lecteur comme un effet aléatoire et en incluant l'expérience du lecteur comme une variable dichotomique. Ce mécanisme a permis d'obtenir des courbes moyennes pour l'ensemble des 7 lecteurs juniors et des 5

lecteurs expérimentés. Une méthode de rééchantillonnage par bootstrap sur les patients a été réalisée (1000 échantillons) afin d'estimer une mesure des intervalles de confiance sur les courbes ROC par une méthode non biaisée (bias-corrected percentile method).

Les statisticiens ont également étudié les variations de distribution des scores attribués aux lésions malignes et bénignes entre les différentes lectures à l'aide d'un modèle de régression ordinaire mixte permettant de prendre en compte des facteurs fixes et aléatoires. Dans cette étude, les lecteurs ont été considérés comme des facteurs aléatoires.

(iii) Résultats

(a) Performances intrinsèques du CADx

La Figure B.4-9 représente les courbes ROC expérimentales obtenues pour les quatre classifieurs évalués avec et sans sélection de caractéristiques (gauche) et avec une sélection préalable des caractéristiques par le test t . Ces performances correspondent à la tâche de discrimination des 42 lésions cancéreuses et 46 lésions suspectes mais bénignes. La meilleure combinaison de paramètres correspondant à un classifieur SVM associé à une sélection des caractéristiques par la méthode du test t permet d'obtenir une aire sous la courbe de 0.82 [0.73-0.90]. On pourra noter (Cf Table 1 de l'article PMB), comme pour l'étude TEP/TDM, que les 34 caractéristiques sélectionnées correspondent à des paramètres extraits des trois types de séquence IRM et contiennent des paramètres de Haralick. Comme nous le détaillons dans la discussion, ces performances sont comparables à celles de la littérature et nous permettent d'avoir une base de référence pour développer une stratégie plus performante.

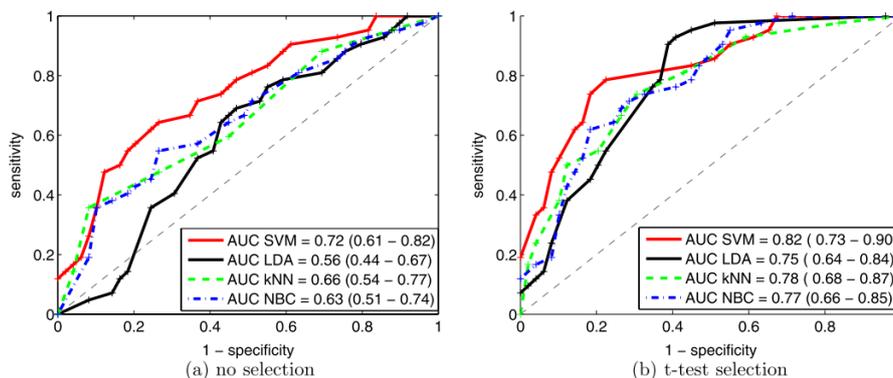


Figure B.4-9. Courbes ROC pour les différents classifieurs évaluées a) sans sélection de caractéristiques et b) avec sélection par la méthode du t-test.

L'analyse des performances reportées dans le tableau 4 du papier PMB montre que la sélection de caractéristiques permet d'augmenter significativement les performances de classification pour le LDA, le KNN et le classifieur de Bayes. Les performances du SVM sont, en revanche, sensiblement meilleures après sélection, mais pas de manière significative. On observe parallèlement que ce classifieur obtient de meilleures performances que les 3 autres algorithmes lorsque l'on considère l'ensemble des 145 caractéristiques. Cela traduit sa capacité intrinsèque à mieux prendre en compte le phénomène du fléau de la dimension.

En l'absence de différence significative de performance pour le SVM avec et sans sélection de caractéristiques, le choix s'est porté sur la solution basée sur un nombre réduit de caractéristiques, afin de réduire les temps de calcul.

(b) Impact clinique du CADx

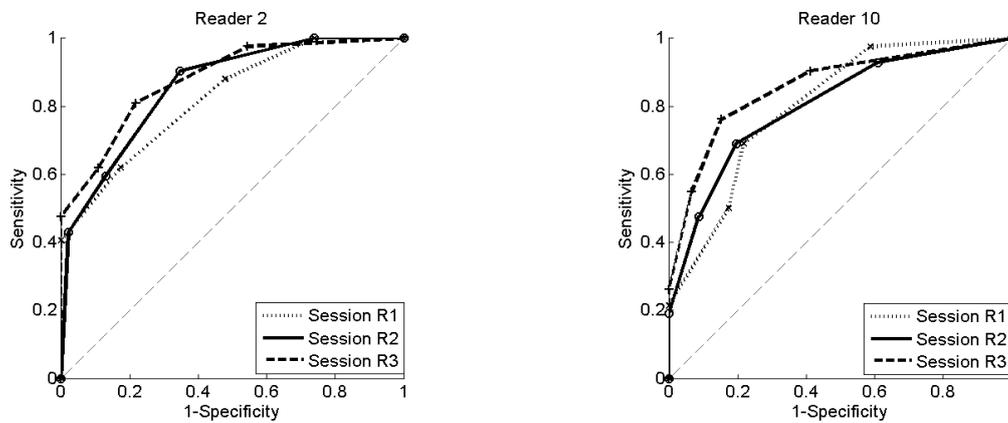


Figure B.4-10. Exemple d'évolution des courbes ROC au cours des trois séances de lecture (R1, R2 et R3) pour un radiologue junior (gauche) et un radiologue expérimenté (droite).

La Figure B.4-10 présente deux exemples de courbes ROC expérimentales mesurées chez un lecteur junior (6 mois d'expérience) et un lecteur senior (2 ans d'expérience) au cours des trois séances de lecture. Les courbes moyennes sur les deux populations de lecteurs et pour les trois lectures (R1, R2 et R3) sont reportées sur la Figure B.4-11 et résumées par les performances en terme d'AUC du Tableau B.4-3. Ces résultats montrent, d'une part, une bonne reproductibilité des performances entre les lectures R1 et R2, soulignant ainsi l'absence d'effet mémoire. Cette tendance est confirmée par l'analyse statistique qui conclut à l'absence de différence significative entre les performances mesurées pour les lectures R1 et R2 ($p=0.69$ et $p=0.9$ pour les groupes juniors et seniors respectivement). D'autre part, l'augmentation systématique des performances avec le système CAD (lecture R3), quel que soit le niveau d'expertise des lecteurs, permet de souligner l'impact positif de ce système en pratique clinique. Cependant, malgré une augmentation comprise entre +0.4% à 6% , selon les lecteurs, des mesures d'AUC entre les lectures R2 et R3, l'étude n'a pas permis de conclure à une différence significative ($p=0.08$).

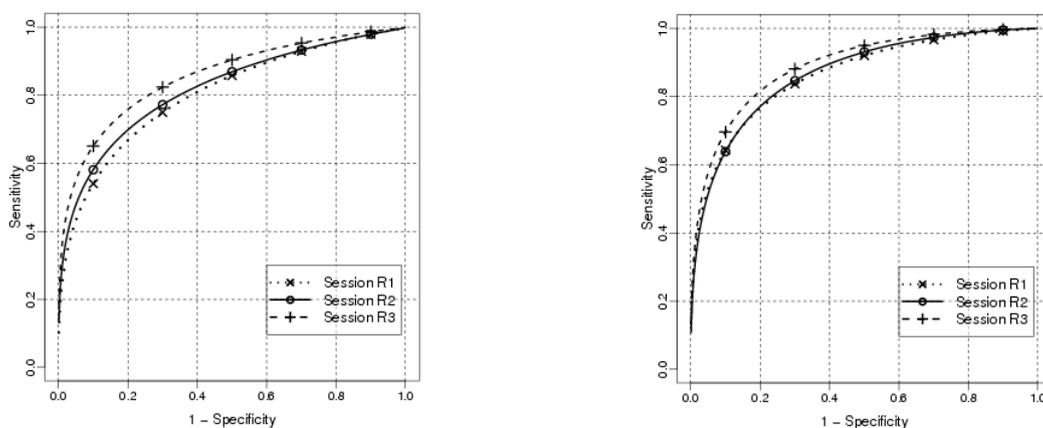


Figure B.4-11. Courbes ROC modélisées pour les trois séries de lecture pour l'ensemble des 7 lecteurs juniors (gauche) et des 5 lecteurs expérimentés (droite). La séance de lecture R3 correspond aux performances mesurées avec le CAD.

	R1		R2		R3	
<i>Juniors</i>	80.3	[74.5;85.6]	81.9	[76.0;87.3]	85.6	[78.4;90.9]
<i>Seniors</i>	86.3	[80.8;90.5]	86.7	[81.2;91.6]	89.1	[83.2;93.8]
<i>Tous</i>	83.0	[77.9-88]	84.1	[78.1-88.7]	87.2	[81-92]

Tableau B.4-3. Aire sous la courbe ROC (AUC) moyenne pour les 7 lecteurs juniors, et les 5 lecteurs seniors ainsi que pour l'ensemble des 12 lecteurs. Cette mesure est exprimée en %.

	Score	R1		R2		R3	
Tous	0	18.6	[11.2;26.4]	19.7	[12.1;27.9]	40.6	[30.1;50.6]
	1	31.3	[25.4;37.6]	33.5	[26.7;41.2]	26.0	[20.1;32]
	2	28.8	[24.7;33.3]	25.8	[20.4;32.6]	19.5	[13.6;26.1]
	3	19.5	[12.1;26.9]	18.6	[12.1;25.3]	11.4	[5.5;18.8]
	4	1.8	[0.9;2.9]	2.4	[1;4.7]	2.4	[0.9;5.7]

Tableau B.4-4. Distribution des scores attribués aux lésions bénignes à partir du modèle de régression ordinale mixte pour les différentes lectures et les deux groupes de lecteurs.

Le Tableau B.4-4 et le Tableau B.4-5 reportent les distributions modélisées des scores moyens au cours des 3 lectures pour l'ensemble des 12 lecteurs. Le Tableau B.4-4, qui concerne la notation des 46 lésions bénignes mais suspectes (NS), souligne à nouveau la similarité des distributions de score entre les lectures R1 et R2 et montre une diminution, en moyenne, des scores au cours de la lecture R3. On observe, en effet, une diminution significative du nombre de zones NS notées 2 ou 3 et une augmentation du nombre de score 0. Cette tendance peut se traduire par une augmentation de la spécificité (en considérant un seuil de positivité de 3). Le Tableau B.4-5, qui correspond à la notation des 42 lésions cancéreuses (M), souligne une augmentation du nombre de lésions M notées avec un score de 4 et une baisse des scores intermédiaires (score de 2 ou 3), ce qui peut se traduire par une augmentation de la sensibilité. Cependant, cette variation n'est pas significative entre les lectures R2 et R3. On observe, de plus, une augmentation du pourcentage de score égal à 0, ce qui peut traduire une limitation du CAD se traduisant ici par une tendance à induire des faux positifs.

	Score	R1		R2		R3	
Tous	0	0.3	[0.0;1.0]	1.7	[0.8;2.7]	4.4	[1.3;8.5]
	1	11.8	[7.4;17.3]	9.4	[5.1;16.0]	10.4	[4.8;17.6]
	2	23.8	[17.0;31.3]	20.5	[14.7;26.6]	13.7	[8.9;19.5]
	3	26.9	[20.3;34.2]	26.8	[20.6;33.2]	25.6	[18.9;32.9]
	4	37.1	[27.1;49.3]	41.6	[29.9;54.2]	45.8	[32.2;59.1]

Tableau B.4-5. Distribution des scores attribués aux lésions cancéreuses à partir du modèle de régression ordinale mixte pour les différentes lectures et les deux groupes de lecteurs.

(iv) Discussion-Perspectives

Ces travaux ont fait l'objet d'un nombre important de communications aussi bien dans des congrès médicaux nationaux (JFR) et internationaux que dans des conférences internationales en traitement d'images médicales (SPIE). On pourra d'ailleurs noter qu'Emilie Niaf a obtenu un prix dans la catégorie des travaux de jeunes chercheurs à la conférence SPIE de 2011. Ces études ont également fait l'objet de deux publications dans des revues de rang A dans le domaine. Les publications relativement récentes (fin 2012 pour l'article PMB [Niaf et al., 2012] et février 2014 pour l'article de Radiology [Niaf et al., 2014]) ne permettent pas de donner une mesure objective de l'impact de ce travail, on pourra cependant noter un taux de citation de 21 pour l'article PMB en janvier 2015 (h-index= 2, d'après le ISI web of knowledge).

Je reviens dans la suite sur la discussion de quelques limites de cette étude et sur les perspectives à donner.

(a) Performances intrinsèques

Comme souligné, les point forts de l'étude publiée dans la revue PMB [Niaf et al., 2012] sont d'avoir pu appuyer cette étude sur la base de données CLARA-P et d'avoir réalisé une étude exhaustive comparant différentes stratégies de systèmes de classification associé à un large panel de caractéristiques descriptives.

J'ai rapporté dans cette synthèse les performances de la tâche de discrimination la plus complexe qui consiste à séparer les lésions bénignes mais suspectes des lésions cancéreuses. Cette tâche est moins souvent abordée dans la littérature. Les travaux les plus proches de ce paradigme sont ceux de l'équipe néerlandaise de l'Université de Nijmegen [Vos et al., 2008, Vos et al., 2010]. L'étude de Vos et al publiée en 2008, par exemple, évalue les performances d'un système CADx basé sur un SVM combiné à des paramètres semi-quantitatifs (pente de rehaussement) et quantitatifs (Ve, Ktrans) mesurés sur des acquisitions DCE. Elle inclut 39 lésions cancéreuses et 21 lésions suspectes extraites de 30 patients. Vos et al obtiennent une AUC de 0.83 [0.75-0.92]. En 2010, cette équipe publie des performances en AUC de 0.85 [0.73-0.90] avec un SVM associant des caractéristiques extraites d'images T2 et une AUC de 0.89 [0.81-0.95] en combinant des caractéristiques issues d'acquisition T2 et dynamique T1. Cette étude repose sur une série de 39 lésions cancéreuses, 19 lésions bénigne et 29 zones normales. Les performances publiées dans cette dernière étude sont supérieures aux performances que nous avons mesurées dans [Niaf et al., 2012] (AUC=0.82 [0.73-0.90]). Cependant, une lecture attentive des deux articles met en évidence une variation du protocole d'évaluation. Dans [Vos et al., 2008], comme dans notre étude, les auteurs évaluent les performances de discrimination entre les lésions normales suspectes (NS) et les lésions cancéreuses (M). Dans [Vos et al., 2010], en revanche, ils prennent en compte des zones de tissu sain prostatique (N) dans la classe dite 'normale' et effectuent ainsi une tâche de discrimination entre les groupe {N, NS} et les lésions cancéreuses {M}. Cette tâche est donc plus simple puisqu'elle incorpore les zones normales de l'image. Cette modification pourrait expliquer l'écart de performance observée. Nous avons contacté, en 2012, l'équipe de Vos pour les interroger sur ce changement de protocoles mais n'avons jamais obtenu de réponses.

Comme pour l'étude lymphome, l'évaluation des performances intrinsèques a été réalisée à l'aide de la méthode de type leave-one-out. Nous renvoyons à la lecture de la discussion sur les limites de cette méthode dans la section B.4.3.1(iv)(a)(ii). Elle peut en effet introduire un biais positif des performances dans le cas où la sélection des caractéristiques et l'optimisation des paramètres du classifieur sont réalisées en dehors de la boucle de validation croisée. Contrairement à l'étude lymphome, nous n'avons pas tenté de quantifier ce biais en utilisant une méthode d'évaluation telle que la méthode de Fukunaga et Hayes. Les performances intrinsèques publiées dans cet article doivent ainsi être modérées. Notons, cependant, la méthode LOO est la référence utilisée dans la majorité des études sur les systèmes CADx.

La sélection de caractéristiques aurait pu faire l'objet d'analyses complémentaires. Une des limitations concerne la charge calculatoire requise par la méthode hybride. Dans son implémentation, Emilie Niaf a, en effet, réalisé l'optimisation des paramètres du classifieur (recherche exhaustive dans l'espace

des paramètres) pour chaque sous-ensemble de caractéristiques ordonnés selon la méthode filtre choisie (test de Student etc.), i.e. par exemple pour chaque point du graphe de la Figure B.4-8. La méthode TOP10 présentée pour l'étude lymphome en TEP/TDM pourrait être une alternative plus efficace en terme de temps de calcul car elle reste une approche de type filtre. Nous avons montré qu'elle se comparait favorablement à des méthodes de sélection enveloppantes telles que l'algorithme génétique. Il aurait été intéressant de réaliser le même genre de comparaison dans cette étude. De plus, la comparaison des performances du SVM sans sélection de caractéristiques (discuté dans la section précédente) ne permet pas de justifier la mise en œuvre de cette méthode de sélection. Dans la suite des travaux menés autour de ce projet, nous avons privilégié des méthodes de sélection plus efficaces sur le plan calculatoire.

(b) Impact clinique

L'étude publiée dans Radiology [Niaf et al., 2014] montre que le système CADx que nous avons développé permet aux radiologues de diagnostiquer plus efficacement les cibles suspectes mais bénignes, i.e. de réduire le taux de fausses détections. En revanche, nous n'avons pas pu démontrer que le CAD permettait d'augmenter le niveau de confiance dans les cibles cancéreuses. Il aurait été intéressant de faire une analyse plus fine du gain en sensibilité et spécificité en fonction du score de Gleason, qui caractérise l'agressivité du cancer. L'un des challenges cliniques, en effet est de mieux détecter les cancers dont l'indice de Gleason est proche de 6. Ces cancers peu agressifs et peu évolutifs s'apparentent facilement à des lésions normales mais suspectes. Leur discrimination fiable et reproductible par des méthodes d'imagerie non-invasive permettrait d'envisager des protocoles dits de 'surveillance active' et de limiter ainsi le surdosage thérapeutique [Somford et al., 2012]. Nous revenons sur ce point de discussion dans la présentation de perspectives discutées dans la partie B.6.3.

Une autre limite de l'étude clinique que nous mentionnons dans la discussion de l'article [Niaf et al., 2014] concerne le protocole de lecture utilisé. Comme indiqué en B.4.3.2(ii)(b), les lecteurs visualisaient simultanément les trois séries d'images correspondant aux trois séquences sur lesquelles étaient superposées les régions d'intérêt à évaluer. Or cette série de régions était identique à celle utilisée pour entraîner le système CAD, ce qui constitue donc une configuration optimale. Il aurait été préférable, comme nous l'avons mentionné précédemment, d'utiliser des bases d'entraînement et de test séparées. Une alternative intéressante aurait été de réaliser un second contourage des zones suspectes. Sans augmenter le nombre d'examen, cette étape aurait permis d'évaluer les performances du CAD sur une série de régions légèrement différentes et de mesurer parallèlement l'influence de la précision du contourage sur les performances.

Les bonnes performances obtenues par ce système nous ont encouragés à initier la seconde phase du projet consistant à intégrer la tâche de localisation automatique des lésions suspectes dans l'image afin de produire une cartographie 3D des zones suspectes. Parallèlement, nous avons entamé une discussion avec les partenaires cliniciens pour évaluer les différentes options de valorisation conjointe du système CAD et de la base CLARA-P.

B.4.4 Contributions dans le domaine des systèmes CADe d'aide à la détection

B.4.4.1 Un système d'aide à la détection pour l'imagerie TEP corps entier

(i) Problématique et contexte scientifique

Le contexte de l'apport de l'imagerie TEP pour la prise en charge des patients atteints de lymphome a été précisé dans la section B.4.3.1(i). J'ai, en particulier, souligné l'intérêt de cette modalité d'imagerie dans le cadre du suivi thérapeutique, pour la détection des fixations radioactives résiduelles qui indiquent une mauvaise réponse au traitement.

Le protocole de prise en charge diagnostique et de suivi du lymphome a fait l'objet, en 2007, de recommandations d'un groupe de travail de l'IHP (international harmonization project) [Juweid et al., 2007]. Il se base sur la sélection et le suivi d'un groupe de six lésions cibles majeures dans l'image TEP corps entier au FDG. Ces cibles sont repérées et quantifiées sur le plan anatomique (taille) et fonctionnel (SUV). Leur évolution est ensuite suivie au cours du traitement par chimiothérapie et permet de définir, à partir d'une nomenclature, si le patient répond ou non au traitement. La présence de fixation résiduelle pathologique de FDG à la fin de la chimiothérapie conditionne le protocole thérapeutique et doit donc être détectée avec précision. La sélection de 6 cibles les plus significatives résulte d'un compromis entre le temps d'analyse visuelle des images TEP corps entier et les contraintes cliniques. Elle est cependant considérée comme un point faible du protocole par les cliniciens puisqu'elle exclut le suivi d'autres lésions éventuelles.

Ce premier projet a été initié avant l'installation de machines couplées TEP/TDM et se base donc uniquement sur l'imagerie TEP. Il est né de discussions avec les partenaires cliniciens lyonnais sur les limites du protocole défini par l'IHP, en particulier avec Thomas Mognetti, aujourd'hui chef du service de médecine nucléaire du centre Léon Bérard (CLB) et le Pr Francesco Giammarile, chef du service de médecine nucléaire du CHU Lyon Sud (CHLS). Les médecins nucléaires lyonnais étaient intéressés par des outils d'analyse automatique des images TEP au FDG permettant de détecter, puis de quantifier, de manière automatique, l'ensemble des lésions cancéreuses de l'image. L'état de l'art réalisé en 2006 et actualisé en 2009 soulignait l'intérêt de plusieurs équipes internationales pour cette problématique et l'absence de systèmes finalisés d'aide à la prise de décision. Une analyse exhaustive est dressée dans l'introduction de l'article annexé dans la section B.8.3.6.

Le verrou scientifique consistait à définir une stratégie qui soit à la fois sensible, c'est-à-dire permette de détecter un maximum de petites lésions cancéreuses mais aussi très spécifique en limitant le nombre de fausses détections qui constituent des pièges pour le médecin et compléxifient le diagnostic. L'apport du CAD est, dans cette situation, négatif.

Nous avons tenté de structurer ce projet par deux demandes de financement auprès de l'ANR en 2007 et 2008 dans le cadre de l'appel d'offre Tecsan avec la société Philips (voir section A.6.1). Malgré l'échec de ces tentatives, notre collaboration s'est organisée par l'encadrement de plusieurs étudiants de master ainsi que par les thèses de Sandrine Toméi puis de Simon Marache financé dans le cadre d'un contrat CIFRE avec la société Philips.

Un premier prototype a été réalisé dans le cadre de la thèse de Sandrine Toméi que j'ai co-encadrée entre octobre 2006 et décembre 2009 avec le Pr Christophe Odet. Ce travail a été présenté en congrès en 2007 (SPIE et GRETSI) et en 2008 (IEEE Medical Imaging Conference). J'ai ensuite poursuivi ce travail en proposant un système original basé sur la combinaison de plusieurs classifieurs. Cette étude a été publiée en 2012 dans la revue IEEE Transactions on Nuclear Sciences annexée en section B.8.3.6 [Lartizien et al., 2012].

Les choix méthodologiques de ce travail ainsi que les principaux résultats sont discutés dans la suite de ce chapitre. Nous renvoyons à la lecture de l'article [Lartizien et al., 2012] pour la description précise des détails méthodologiques.

(ii) Choix méthodologiques

Comme souligné en introduction, notre problématique était celle de la détection de petites fixations radioactives de faible contraste. Le cadre méthodologique que j'avais défini s'inspirait du formalisme des observateurs numériques développé entre 2002 et 2006 et se basait donc initialement sur une analyse discriminante linéaire (dont le formalisme est proche du modèle de Hotelling comme nous le précisons dans l'annexe B.7.4.3). L'idée à l'origine de ce projet était d'ajouter une tâche de localisation au modèle de Hotelling canalisé (CHO) en calculant la variable de décision en chaque voxel de l'image 3D considéré. Cette idée était parallèlement explorée au même moment par d'autres équipes, en particulier dans les travaux de Gifford et Platasa présentés dans le chapitre B.3.5. Les limites du modèle de Hotelling m'ont ensuite conduite à m'intéresser à d'autres méthodes de classification supervisée, dont les SVM. L'objectif était d'établir une carte paramétrique de score de malignité dans des images TEP-FDG corps entier (matrice de dimension $128 \times 128 \times 108$ voxels avec un voxel isotrope de 5.0625 mm de côté).

(a) Choix des classifieurs

Le principe de l'analyse discriminante linéaire (ADL) est rappelé dans l'annexe B.7.4.3. Le choix de l'ADL était motivé par les deux points suivants. Il s'agit d'une part, d'un classifieur linéaire, rapide et par conséquent bien adapté à l'analyse de bases de données de grande dimension. De plus, j'avais montré que le modèle de Hotelling était bien corrélé aux performances humaines dans le cas d'une tâche de détection simple (cf résultats de la Figure B.3-7). L'extrapolation de ces performances dans le contexte plus complexe de la tâche de localisation considérée dans cette application n'était cependant pas évidente. Je me suis donc parallèlement orientée vers d'autres méthodes de classification supervisée, en particulier les séparateurs à vaste marge (SVM) dont le principe est rappelé dans l'annexe B.7.4.2(iii). Ce choix a été guidé par sa capacité à résoudre des problèmes non linéaires grâce à l'utilisation de fonctions noyau.

(b) Choix des caractéristiques descriptives

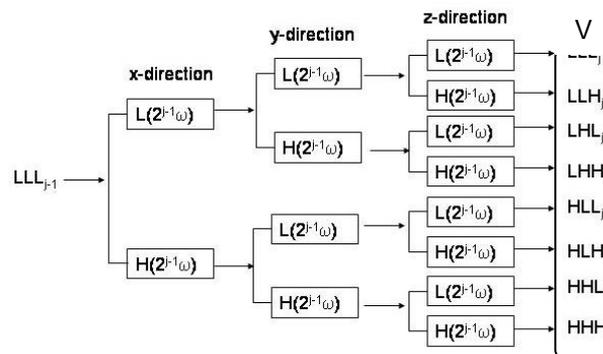


Figure B.4-12. Vecteur de caractéristiques extrait de la transformée en ondelettes discrète non décimée. j est le niveau de décomposition

Comme présenté dans l'annexe B.7.3.2(iv), les filtres classiquement utilisés avec le CHO sont des filtres DOG ou SQR (cf Figure B.7-5). L'utilisation des caractéristiques fréquentielles est généralement motivée par la volonté de reproduire le processus visuel humain qui traite l'information par un ensemble de filtres avant de l'interpréter. J'ai proposé de remplacer cette composante par les coefficients issus d'une décomposition en ondelettes, en m'inspirant d'une étude portant sur la détection de microcalcifications en mammographie [Strickland and Hahn, 1996]. Les ondelettes autorisent une plus grande flexibilité sur le choix de la fonction support, et ainsi une meilleure adaptation au motif que l'on cherche à détecter. Par ailleurs, les ondelettes permettent un paramétrage plus fin du niveau de décomposition fréquentielle de l'image que l'on souhaite analyser. Enfin, le choix d'une ondelette biorthogonale permet de réaliser un traitement simple des images tridimensionnelles, en réalisant une décomposition 1D successive dans les trois directions comme

indiqué sur la Figure B.4-12. Pour cette étude, j'ai sélectionné l'ondelette biorthogonale B-spline 4.4 qui satisfait les propriétés attendues par notre application (détection de sources sphériques).

Le vecteur V de caractéristiques considéré dans cette étude était constitué des coefficients de la transformée en ondelettes discrète non décimée (TOD) (cf Figure B.4-12) qui respecte la condition d'invariance par translation. Cette TOD permet d'obtenir par l'application successive d'un filtre passe-haut (H) ou passe-bas (L), sur les lignes (x), les colonnes (y) et la profondeur (z), huit sous-bandes de détails par niveau de décomposition. Le vecteur V calculé en chaque voxel d'intérêt de l'image est donc constitué des $8 \times j$ coefficients d'ondelettes, où j est le niveau de décomposition, choisi pour cette étude dans l'ensemble $\{2,3,4,5\}$.

(c) Principe du CADe

Le système CADe que nous avons proposé est schématisé sur la Figure B.4-13. Les caractéristiques discriminantes issues de la décomposition en ondelettes de l'image TEP originale constituent les données d'entrée des 2 systèmes de classification supervisée (ADL et SVM), dont l'apprentissage est réalisé parallèlement. Ces deux classifieurs produisent deux cartes de score (degré de malignité), que l'on peut ensuite seuiller pour extraire des cartes binaires de détection. Ces cartes (score or binaire) sont ensuite traitées de manière à réduire les fausses détections.

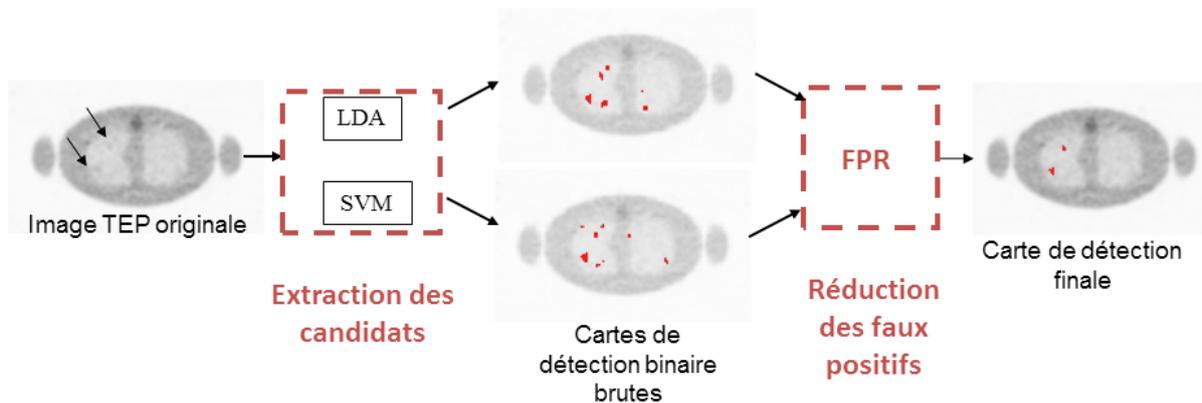


Figure B.4-13. Schéma bloc du prototype de système CADe basé sur la combinaison de deux classifieurs. Dans cet exemple, la carte de détection finale est obtenue en associant les cartes binaires de chaque classifieur par vote à la majorité.

Sandrine Toméï, au cours de sa thèse, avait démontré que chacun des classifieurs entraînés séparément, permettait d'obtenir une très bonne sensibilité (détection des vrais positifs) contrebalancée par des taux élevés de fausses détections (>20 ou 30 faux positifs par images). L'analyse comparative des cartes de scores produites par l'ADL et le SVM mettait en évidence des différences dans la distribution spatiale des fausses détections. Cette observation m'a incitée à combiner les cartes produites par chacun des classifieurs au moyen de deux méthodes de fusion tardive. La première consistait à combiner les cartes binaires de chaque classifieur par une méthode de vote à la majorité. Dans cette méthode, une lésion candidate est détectée dans la carte de détection finale uniquement si elle a été détectée par chacun des classifieurs. Elle est donc optimale lorsque la distribution des vrais positifs est corrélée et celle des faux positifs décorrélée. La sensibilité résultante est ainsi obligatoirement inférieure ou égale à la sensibilité minimale des deux classifieurs. La deuxième stratégie, inspirée d'une méthode proposée en 2011 par l'équipe de l'Université de Nijmegen a été de combiner les cartes de score avant seuillage [Niemeijer et al., 2011]. Cette méthode a pour intérêt de mieux préserver la sensibilité. Il est en effet possible qu'une lésion non détectée par l'un des classifieurs après seuillage de l'image de score 'réapparaisse' après combinaison si la valeur du score combiné est supérieure au seuil appliqué.

(d) Evaluation des performances

L'optimisation puis l'évaluation du système CAD a été réalisée sur la base de données ARSENE présentée dans le chapitre B.2.2.2(i) et utilisée pour les études psychophysiques présentées en B.3.4. Un exemple d'image simulée est présenté sur la Figure B.4-15. Ce choix permettait de comparer les performances du modèle CAD développé avec les performances diagnostiques humaines mesurées sur les mêmes données dans [Lartizien et al., 2004]. Nous avons analysé séparément chacune des catégories de lésions correspondant à un contraste donné. Ceci afin de mieux comprendre le comportement du CADe sur les cas de détection difficiles.

Dans cette étude, les échantillons de la classe positive sont constitués d'un ensemble de N vecteurs de caractéristiques V_i , $i=1..N$, de dimension $8 \times j$. Chaque vecteur V_i est constitué des coefficients d'ondelettes mesurés au centre (voxel central c_i) de chacune des N lésions simulées. La classe négative est également constituée de N vecteurs contenant chacun les $8 \times j$ coefficients d'ondelettes mesurés dans les voxels c_i des cartes de décomposition fréquentielle calculées à partir des images simulées sans lésion.

J'ai mis en œuvre une méthode d'évaluation basée sur un partitionnement des données de type 'hold-out', présentée dans la section B.4.2.1(iii). Comme souligné dans cette section, cette méthode d'évaluation permet, en théorie, de mesurer les performances en généralisation non biaisées, car la base de données test est indépendante de la base d'apprentissage. J'ai choisi de réaliser un apprentissage séparé pour la détection des lésions du foie et des poumons. Les bases d'apprentissage contenaient respectivement 1060 (530 positifs et 530 négatifs) éléments pour le poumon et 1400 (2 fois 700 lésions) pour le foie. Le nombre d'exemples N était ainsi bien supérieur à la dimension du vecteur de caractéristiques (comprise entre 16 et 40 éléments), réduisant fortement le risque de surapprentissage.

La base de validation contenait 25 volumes d'images 3D contenant 53 lésions pulmonaires et 70 lésions hépatiques. Enfin, la base de test était constituée de 25 autres volumes contenant 68 lésions pulmonaires et 63 lésions hépatiques.

(i) Analyse par cluster

L'analyse visuelle des premières cartes binaires obtenues après seuillage des cartes de score (Figure B.4-13 b) m'a conduite à m'interroger sur le choix de la métrique à utiliser pour quantifier les performances diagnostiques. La majorité des travaux publiés proposait une métrique dérivée de l'analyse ROC pour laquelle l'élément de base est le voxel. Ainsi, pour un seuil considéré, chaque voxel détecté comme positif (i.e. dont le score est supérieur au seuil) est compté comme un vrai positif (VP) si il se situe dans une zone de référence étiquetée comme pathologique. La Figure B.4-14 et le Tableau B.4-6 illustrent ce principe sur un exemple représentant une matrice de 8×7 pixels isotropes sur laquelle une région d'intérêt représentant la zone anormale à détecter est contourée en noir (croix noire). On suppose que le CAD produit les cartes binaires représentées sur la Figure B.4-14 pour deux valeurs de seuil λ_1 et λ_2 de la carte de score (on note que $\lambda_1 < \lambda_2$).

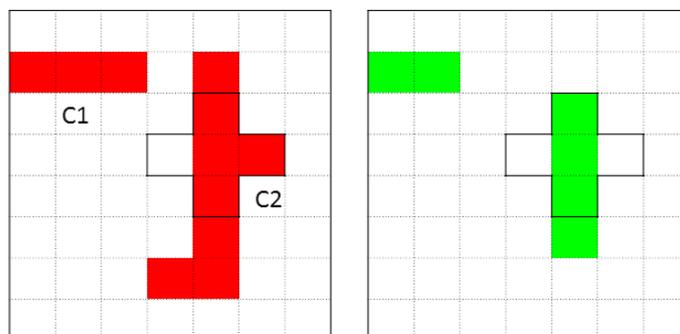


Figure B.4-14. Cartes binaires correspondant à deux valeurs de seuils λ_1 (gauche) et λ_2 (droite). La lésion à détecter est la croix noire composée de 5 voxels.

Supposons que l'on cherche à définir le seuil optimal pour ce CAD. Dans le cas d'une évaluation par voxel, l'analyse des cartes correspondant aux seuils λ_1 et λ_2 donne les résultats reportés dans le Tableau B.4-6. On observe que le passage du seuil λ_1 au seuil λ_2 se traduit par une augmentation de 8% de spécificité compensée cependant par une baisse de 20% de sensibilité. Le choix de seuil optimal ne semble ainsi pas évident.

	TP	FP	Sensibilité	Spécificité
λ_1	4	7	80%	86%
λ_2	3	3	60%	94%

Tableau B.4-6 Évaluation des performances en considérant une analyse par voxel

Une solution alternative consiste à faire une analyse par agrégats (clusters) de voxels voisins. On cherche alors à évaluer si le cluster est caractéristique de la lésion à détecter. Cette méthode suppose tout d'abord d'identifier l'ensemble des clusters de l'image. On peut utiliser pour cela une règle de voisinage et considérer, par exemple, que deux pixels sont connectés s'ils appartiennent au même voisinage en considérant une règle de k-connexité et ont la même étiquette. En fixant $k=8$ pour l'exemple de la Figure B.4-14, on trouve deux clusters repérés par les lettres C1 et C2. La seconde étape consiste à définir un ensemble de règles permettant de caractériser ces clusters afin de décider s'ils décrivent bien la lésion que l'on cherche à détecter. On peut commencer par 'nettoyer' les cartes en éliminant, par exemple, tous les clusters de taille trop petite. Dans cet exemple, nous fixons une taille minimale de 3 voxels. Avec ce critère, le cluster C1 dans la carte correspondant au seuil λ_2 est éliminé. On peut ensuite définir un ensemble de règles sur les vrais positifs intégrant, par exemple, des critères de forme et de taille. Supposons pour cet exemple, qu'un cluster soit considéré comme un vrai positif si l'indice de DICE (rapport de l'intersection des aires du cluster et de la région de référence sur l'union de ces aires) est supérieur à 0.5. Avec ce rapport, le cluster C2 est alors considéré comme un faux positif avec le seuil λ_1 ($DICE=4/9=0.44$) et comme un vrai positif avec le seuil λ_2 ($DICE=3/6=0.5$). Pour le seuil λ_1 , les deux clusters C1 et C2 sont comptés comme des faux positifs, alors qu'avec le seuil λ_2 , le cluster C1 est éliminé en raison de sa taille tandis que le cluster C2 est totalisé comme VP. Ces résultats, reportés dans le Tableau B.4-7, conduisent à choisir λ_2 comme seuil optimal. Il faut noter que cette analyse n'est pas compatible avec une analyse ROC classique mais avec une analyse de type FROC ou AFROC (voir annexe B.7.2.2).

	TP	FP	Sensibilité	FPI
λ_1	0	2	0%	2
λ_2	1	0	100%	0

Tableau B.4-7 Évaluation des performances en considérant une analyse par cluster

L'analyse des deux cartes binaires de la Figure B.4-14 montre que la carte la plus pertinente pour faciliter la tâche de détection d'un observateur humain est la carte verte (à droite). Elle correspond bien au choix du seuil λ_2 que l'on aurait fixé avec la méthode d'analyse par cluster.

Cette schématisation sur un exemple de synthèse est favorable au mode d'analyse par cluster. Il faut cependant modérer ces conclusions et discuter le point concernant la construction des « règles » définissant si un cluster est un vrai positif. Nous renvoyons à la lecture de l'article de revue de Petrick et al publié en 2013 [Petrick et al., 2013] pour appuyer cette discussion. Cet article, rédigé par un groupe de travail de l'AAPM (American Association of Physicists in Medicine), fait une synthèse récente des travaux les plus aboutis sur l'évaluation des CAD et tente de proposer un protocole

standardisé. Si les auteurs semblent favorables à l'approche par cluster, ils soulignent cependant qu'il n'y a pas de consensus sur la notion de « règles », ce qui constitue un frein à la généralisation de cette méthode. Il est ainsi recommandé de mentionner explicitement les règles utilisées dans les publications en essayant de fournir une justification par rapport au contexte médical. Cet article référence une quinzaine de publications basées sur ce mode d'évaluation exclusivement dans le contexte de l'imagerie TDM pour la mammographie ou la détection de tumeurs pulmonaires et de polypes.

Cette analyse par cluster constitue ainsi l'une des contributions originales de l'étude publiée dans [Lartizien et al., 2012]. La soumission initiale de cette étude en 2011 détaillait les règles de sélection définies. Ce point a suscité un commentaire négatif de la part d'un relecteur, considérant qu'il s'agissait d'une approche trop empirique. L'article de Petrick n'était alors pas encore publié, j'ai préféré retirer cette partie du texte qui essayait pourtant de justifier le choix des paramètres. Ces règles sont détaillées ci-dessous.

La détection des agrégats (clusters) a été obtenue en réalisant un parcours lexicographique de la carte binaire de détection et en considérant un voisinage de connexité 26. J'ai fixé la taille minimum d'un cluster à 7 voxels. Les règles permettant de caractériser un VP prenaient en compte 1) la taille du cluster, 2) la taille de la zone d'intérêt (ROI) à détecter, 3) la taille de l'intersection entre le cluster et la ROI ainsi que 4) le rapport des scores moyens sur la lésion et la ROI de référence. Un cluster était totalisé comme VP s'il remplissait les conditions suivantes :

$$\left\{ \begin{array}{l} \text{taille(intersection)/taille(ROI)} > 5\% \\ \text{taille(intersection)/taille(cluster)} > 25\% \\ \text{score_moyen(intersection)/score_moyen(cluster)} > 5 \end{array} \right. \quad (\text{B.4-1})$$

Cette règle a été établie de manière empirique, à partir de l'analyse des cartes de score seuillées. Les deux premiers critères permettaient d'éliminer les clusters de forme trop excentriques, tels que des clusters de forme très allongée (souvent détectés dans les zones de transitions, à l'interface entre le foie et poumon, par exemple). Le dernier critère permettait de vérifier que le centre de masse du cluster détecté était bien positionné sur la lésion à détecter.

L'ensemble de ces règles et des paramètres associés a été ajusté de manière à optimiser l'information pertinente contenue dans les cartes binaires pour guider le lecteur vers les zones d'intérêt de l'image. J'ai également vérifié que les performances n'étaient pas trop sensibles aux variations des différentes valeurs de seuil.

(iii) Résultats

(a) Evaluation du CAD

Je présente ici une synthèse des résultats de l'article annexé en section B.8.3.6.

Le Tableau B.4-8 présente l'évolution de la sensibilité et du taux moyen de fausse détection par image (FPI) pour les différentes stratégies évaluées. Les deux premières lignes indiquent les performances individuelles de l'ADL et du SVM et les deux suivantes les performances obtenues en combinant soit les cartes binaires (PRBin), soit les cartes de scores (PRSco). La méthode de combinaison PRSco permet d'obtenir le meilleur compromis sensibilité-spécificité pour les lésions situées dans le poumon ou dans le foie et réduit d'un facteur 2 à 3 le nombre de fausses détection tout en maintenant une sensibilité de détection très élevée (De l'ordre de 90 à 95%).

Tableau B.4-8 : Evolution des performances des différents schémas de classification

	Poumons		Foie	
	Sensibilité	FPI	Sensibilité	FPI
LDA	0.90	40	0.92	30
SVM	0.97	34	0.92	11
PRBin	0.91	20	0.76	7
PRSco	0.91	18	0.94	10

La Figure B.4-15 permet de visualiser l'amélioration impliquée par les méthodes de fusion tardive pour un ensemble de lésions situées dans le poumon. Elle représente une vue 3D MIP d'un modèle contenant 5 tumeurs superposée aux cartes binaires correspondantes pour l'ADL (a), le SVM (b) et les combinaisons des cartes binaires (c) ou des cartes de score (d).

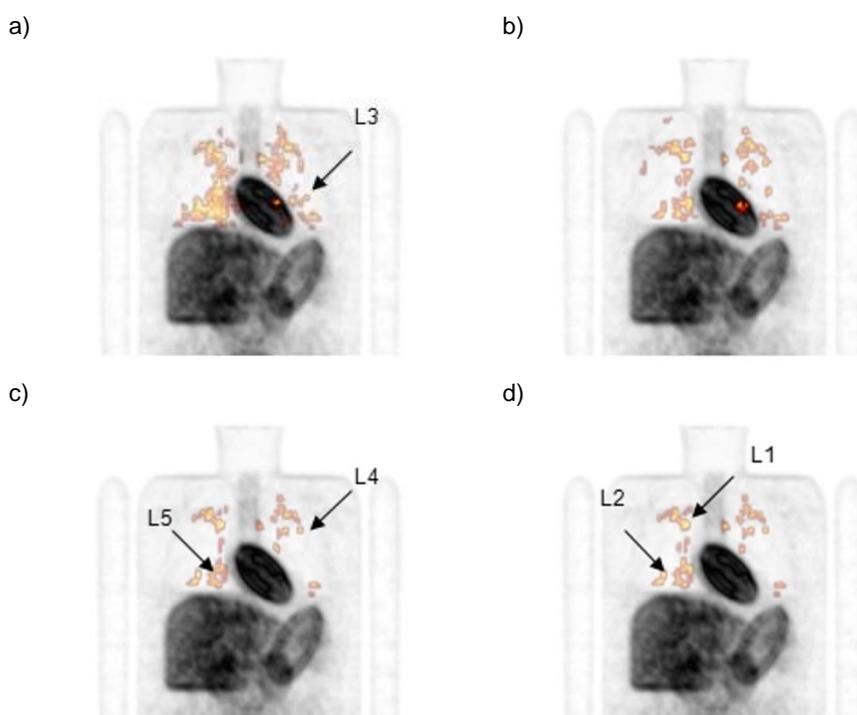


Figure B.4-15. Vue 3D en rendu MIP d'une image TEP superposée à la carte binaire de détection obtenue à l'aide des 4 schémas de classification reconstruite (a) LDA, (b) SVM, (c) PRBin, (d) PRSco. Cinq lésions de contraste variable : L1: 90%, L2: 30%, L3: 10%, L4: 10%, et L5: 90%.

Ces résultats soulignent les faibles performances des systèmes basés sur un classifieur unique, principalement en raison du taux très élevé de fausses détections. Les méthodes de combinaison de classifieurs permettent d'obtenir des performances satisfaisantes, mais néanmoins toujours trop peu spécifiques (10 FPI en moyenne par images).

(b) Comparaison aux performances humaines

Le Tableau B.4-9 et le Tableau B.4-10 présentent les sensibilités obtenues par les différentes stratégies pour chaque niveau de contraste des lésions du poumon et du foie. Les sensibilités humaines de référence (10%, 30%, 50%, 70% et 90%) correspondent aux sensibilités moyennes mesurées chez l'homme. Ces tableaux montrent que le CAD surpasse la sensibilité de détection

humaine. Il est particulièrement important de constater les performances remarquables obtenues pour les très faibles contrastes (10 et 30%). Le CAD détecte, en effet, entre 71 et 100% des lésions détectées par l'homme dans 10% et 30% des cas seulement.

Tableau B.4-9

SENSIBILITE DES DIFFERENTS SCHEMAS DE CAD EN FONCTION DU CONTRASTE DES LESIONS PULMONAIRES

Detection humaine	10 %	30 %	50 %	70 %	90 %
LDA	82%	91%	92%	100%	89%
SVM	88%	100%	100%	100%	100%
PRBin	76%	91%	100%	100%	94%
PRSco	76%	82%	100%	100%	94%

Tableau B.4-10:

SENSIBILITE DES DIFFERENTS SCHEMAS DE CAD EN FONCTION DU CONTRASTE DES LESIONS HEPATIQUES

Detection humaine	10 %	30 %	50 %	70 %	90 %
LDA	93%	88%	80%	95%	100%
SVM	86%	100%	80%	100%	92%
PRBin	71%	75%	40%	84%	100%
PRSco	93%	88%	90%	95%	100%

(iv) *Discussion – Perspectives*

(a) **Performances intrinsèques et impact clinique**

Les résultats présentés dans le Tableau B.4-9 et le Tableau B.4-10 montrent que les différents systèmes CAD étaient très efficaces pour détecter des tumeurs de faible contraste, mais trop peu spécifiques pour envisager une évaluation clinique, malgré le gain significatif de performance apporté par la combinaison des classifieurs. Comme je l'ai discuté dans la section B.4.3.1(i), la dissémination, au même moment, de machines bimodales TEP/TDM dans les centres hospitaliers, a permis de relancer ce projet ; nous avons en effet supposé que l'ajout d'attributs anatomiques mesurés dans les images TDM permettrait de pallier à la faible spécificité du système basé sur la TEP seule. J'ai alors choisi de continuer à travailler à l'aide de données simulées et me suis orientée vers la simulation de données réalistes en TDM.

J'ai structuré cette action dans le cadre du projet ANR VIP (voir section A.6.1.3) et du projet de thèse CIFRE de Simon Marache (voir section A.5.3). La conclusion de cette étude est résumée sur la Figure B.2-6 et discutée plus précisément dans le chapitre B.2.3. Elle montrait que les modèles simulés n'étaient pas suffisamment réalistes pour les besoins de l'étude CAD et identifiait les pistes de recherche à développer pour se rapprocher de la complexité des images TDM cliniques.

Ces contraintes m'ont poussée à reconsidérer l'intérêt des données cliniques. Cette idée a été confirmée par ma rencontre, au même moment, avec Fabien Ricard, chef de clinique au CHLS. Comme détaillé dans la section B.4.3.1(i), Fabien Ricard avait constitué une base d'images TEP et TDM de patients atteints de lymphome et s'intéressait aux méthodes d'aide au diagnostic. Il m'a encouragé à temporiser le travail sur la détection des zones de fixation résiduelle et à réorienter l'objectif clinique de ces développements vers la discrimination des zones de fixations bénignes et pathologiques. Ceci nous a conduits à initier la seconde phase de ce projet en 2011, détaillée dans la partie B.4.3.1.

(a) Impact scientifique

Les résultats du Tableau B.4-9 et du Tableau B.4-10 montrent que le CAD est plus sensible que l'observateur humain pour le type d'image considéré dans cette étude. Nous n'avons pas pu comparer précisément la spécificité car les protocoles des études psychophysiques différaient. L'étude psychophysique humaine imposait le report d'un nombre maximal de 7 cibles par volume et indiquait au lecteur le nombre moyen de cibles par organe (2.5 lésions dans le foie et les poumons et 2 lésions dans le fond), alors que le système CAD n'était pas contraint par cette information a priori. Cette différence peut contribuer à expliquer le taux plus faible de fausses alarmes chez les observateurs humains.

Nous avons proposé d'utiliser le système CAD basé sur le SVM comme un observateur algorithmique anthropomorphique capable d'effectuer une tâche de localisation. Les deux études réalisées dans ce sens s'inscrivaient dans la continuité des travaux que j'ai présentés dans le chapitre B.3.5, et plus particulièrement dans le cadre des modèles algorithmiques complexes que j'ai discutés dans la section B.3.5.3. Je renvoie à la lecture de ce chapitre qui détaille l'état de l'art et les orientations de recherche dans ce domaine.

Une première étude a été réalisée pour évaluer l'impact de la correction de l'effet de volume partiel (EVP) sur la détection des petites tumeurs en TEP. Cette étude a été réalisée en collaboration avec le LaTIM (INSERM U650) de Brest. Elle a donné lieu à une communication en congrès en 2009 [Lartizien et al., 2009].

Une seconde étude de même type a été également réalisée, en parallèle, pour évaluer l'apport de la correction du mouvement respiratoire en TEP. Cette étude s'inscrivait dans le cadre de la thèse CIFRE de Simon Marache en partenariat avec le groupe Medisys de la société Philips Healthcare. La Figure B.4-16 illustre un exemple de courbes de type FROC obtenues à partir de la base d'images simulées oncoPET_respi décrite dans la section B.2.2.2(iii). Ces courbes représentent la sensibilité de détection obtenue en fonction du nombre moyen de fausses détections dans l'organe considéré pour quatre types d'images : les courbes tracées en violet et rouge correspondent respectivement au cas d'une correction de mouvement parfaite ou de l'absence de correction tandis que les courbes intermédiaires de couleur verte et bleue indiquent les performances obtenues par deux méthodes de correction du mouvement respiratoire évaluées dans cette étude. Ces performances sont bien situées dans l'intervalle de valeurs borné par les cas extrêmes, soulignant ainsi qu'elles arrivent à restaurer partiellement les performances diagnostiques que l'on obtiendrait avec un patient en apnée pendant toute la durée de l'acquisition.

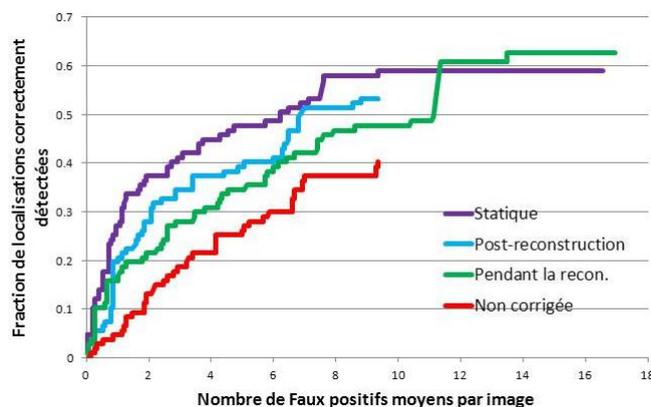


Figure B.4-16. Courbes FROC mesurées pour la détection des lésions hépatiques avec les différents types d'images.

Fort des résultats obtenus à l'aide de l'observateur numérique, nous avons ensuite entrepris de les comparer aux performances diagnostiques humaines, appliquant ainsi les recommandations discutées dans la section B.3.5.3(i) sur la nécessité de vérifier que le système algorithmique et l'homme réagissent de manière similaire (augmentation ou diminution des performances) lorsque l'on modifie les paramètres de l'image. Les résultats préliminaires que nous avons obtenus avec un (seul) observateur humain n'ont pas été concluants. Ils montraient, en effet, que les performances humaines ne semblaient pas améliorées par la correction de mouvement, contrairement à celles du système CAD. Il aurait fallu enrichir l'étude ROC humaine, i. e. la compléter par d'autres observateurs afin de dégager une tendance plus nette.

L'étude développée dans la section B.3.5.2 avait montré une très bonne corrélation entre les performances de détection moyennes d'une série d'observateurs humains et celles d'un modèle de Hotelling dont le formalisme est très proche de l'analyse discriminante linéaire. Cependant, les données ainsi que le système de classification étaient différents. De plus cette étude n'incluait pas la tâche de 'localisation' des lésions suspectes puisqu'il s'agissait de calculer un score de détection pour chaque lésion supposée connue. Ces différences méthodologiques peuvent sans doute expliquer les différences de comportement des observateurs numériques et souligner la complexité de cette problématique, qui constitue un domaine d'étude à part entière.

Apprentissage statistique en imagerie médicale

B.5 Apprentissage statistique en imagerie médicale

Ce chapitre concerne la présentation de mes contributions plus récentes dans le domaine des méthodes de classification à noyau. Cet axe de recherche est beaucoup moins abouti que les thématiques présentées dans les précédents chapitres. J'ai néanmoins choisi de les mentionner ici pour souligner leur importance et ma motivation à poursuivre dans cette voie.

Cette activité s'est développée dans le cadre d'une collaboration avec des spécialistes des machines à noyaux, principalement avec Stéphane Canu du LITIS de Rouen et Rémi Flamary de l'OCA Lagrange à Nice.

La motivation d'approcher et manipuler les machines à noyaux est née du constat de performances un peu décevantes des systèmes de classification classiques. Comme nous l'avons souligné dans le chapitre B.4.3.2, consacré au système CADx pour l'imagerie multiparamétrique IRM du cancer de la prostate, la comparaison exhaustive des performances de différents classifieurs combinés à un large panel de caractéristiques descriptives n'a pas permis d'extraire une configuration significativement plus performante que les autres. Ce constat m'a incité à étudier les mécanismes des méthodes de classification par apprentissage afin de proposer des solutions plus adaptées aux spécificités de l'imagerie médicale. Parmi ces spécificités, on peut citer la taille généralement faible des bases de données d'apprentissage, ainsi que le bruit et l'incertitude sur les annotations, qui constituent un obstacle à la création de données de référence fiables.

Les machines à noyaux reposent sur un fondement théorique solide, elles sont puissantes car elles permettent de résoudre des problèmes linéaires et non linéaires, adaptées à la gestion de données de grande dimension par leur propriété de parcimonie et résistent bien au fléau de la dimension car elles reposent sur la minimisation du risque structurel. Nous avons également observé que ces méthodes se comportaient bien dans les différents cas d'étude exposés dans les exemples précédents. Ces différents arguments ont motivé le choix de ce formalisme comme base de ces développements.

La démarche que nous avons adoptée consiste à identifier les limites du modèle SVM classique pour la problématique médicale d'intérêt puis à essayer de modéliser un facteur correctif dans le cadre du problème d'optimisation standard, sous forme d'un terme de régularisation ou d'une contrainte supplémentaire. Cette démarche est illustrée par une étude récente que nous avons récemment adressée dans le cadre du projet d'imagerie du cancer de la prostate. D'autres pistes sont évoquées dans les perspectives présentées dans la partie B.6.3.

B.5.1 Extension des SVM au cas des étiquettes incertaines.

Les systèmes d'aide au diagnostic que nous avons présentés dans le chapitre B.4.3 ont été évalués à l'aide de données cliniques pour lesquels peut se poser le problème de la vérité terrain. Nous ne disposons, par exemple, d'aucune preuve histologique des zones annotées comme saines ou pathologiques sur les images TEP/TDM. De même, la mise en correspondance des images IRM de prostate avec les données de l'anatomopathologie n'est pas parfaite non plus. Se pose alors le problème de la fiabilité de l'évaluation des performances du système CAD, problématique que nous avons longtemps réglée en utilisant des données simulées (voir par exemple, le système de CADe pour l'imagerie TEP présenté en B.4.4.1). J'ai proposé d'aborder cette question en étudiant comment intégrer cette incertitude dans le système de classification par SVM. L'objectif était de pouvoir apprendre un modèle de prédiction (fonction de décision du SVM) fiable et généralisable en présence de données dont l'étiquette de classe est incertaine.

En collaboration avec le LITIS, nous avons proposé une extension de la méthode de classification par SVM, permettant de traiter les incertitudes sur les étiquettes de certains points de la base de données d'apprentissage. Notre proposition consiste à rajouter au problème classique des SVM une contrainte sur la précision des probabilités estimées par le modèle pour les points d'apprentissage incertains. Ce formalisme, baptisé P-SVM pour SVM 'probabiliste' est détaillé dans la section suivante. Nous l'avons évalué sur des données de synthèse dans un premier temps, puis sur les données d'imagerie par IRM multiparamétrique de la base CLARA-P présentée dans le chapitre B.4.3. Ces données contiennent,

en plus d'une référence histologique précise de chaque lésion, une série d'annotations réalisée en aveugle par plusieurs radiologues en utilisant une échelle de Lickert (score entre 0 et 4). Les modèles SVM et P-SVM ont été entraînés sur les données d'apprentissage annotées par les médecins mais sans connaissance de la vérité histologique. Les performances intrinsèques de ces modèles ont ensuite été mesurées sur un ensemble de données test dont l'étiquette de classe était fixée par la vérité histologique. Ce paradigme a ainsi permis de sélectionner le modèle le plus solide en présence de données d'apprentissage incertaines.

Ce travail a donné lieu à deux communications en congrès [Niaf et al., 2011], [Niaf et al., 2012] et à un article publié dans la revue IEEE TIP [Niaf et al., 2014]. Nous présentons une synthèse de cette étude.

B.5.1.1 Formalisme

Le formalisme classique des SVM est présenté dans l'annexe B.7.4.2(iii)(a). Nous rappelons ici les idées et résultats principaux.

On se place dans un problème de classification en 2 classes '-1' et '1'. Soit un ensemble de n individus $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ notés $\mathbf{x}_i = (x_i(1), \dots, x_i(p))^T$ de centroïde $\bar{\mathbf{x}} \in \mathcal{X}$. Chaque individu appartient à une classe $y_i \in \mathcal{Y} = \{-1, 1\}$. On note n_{-1} et n_1 le nombre d'individu dans chaque classe et $\bar{\mathbf{x}}_{-1} \in \mathcal{X}$ et $\bar{\mathbf{x}}_1 \in \mathcal{X}$ respectivement les centres de gravité des deux nuages de points. L'objectif est de rechercher une fonction discriminante linéaire $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ qui minimise le risque empirique approximé par le coût charnière de Hinge et la norme ℓ_2 de \mathbf{w} . L'expression résultante du problème d'optimisation sous contrainte, dans le cas d'un problème linéairement séparable avec marge souple, s'écrit comme

$$\begin{cases} \min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ s.c. y_i (\mathbf{w}^T \mathbf{x}_i + b) > 1 - \xi_i, \quad i=(1, 2, \dots, n) \\ \xi_i > 0 \end{cases} \quad (\text{B.5-1})$$

Ce problème peut être résolu par la minimisation du Lagrangien par rapport à \mathbf{w} et b . La fonction discriminante qui renvoie la distance à l'hyperplan séparateur s'écrit

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b$$

Où α_i sont les coefficients de Lagrange. L'ensemble des points d'apprentissage caractérisé par un coefficient α_i non nul constituent les vecteurs supports du modèle de prédiction.

Dans le cas que nous proposons de traiter, nous disposons d'un ensemble d'apprentissage dont les m premiers exemples sont annotés sans ambiguïté, i.e. ont une classe d'appartenance $y_i = l_i \in \{-1, +1\}$, tandis que les $(n-m)$ exemples suivants sont caractérisés par leur probabilité $y_i = p_i \in [0, 1]$ d'appartenance à la classe +1.

Le formalisme des P-SVM que nous avons proposé consiste à contraindre l'estimation des probabilités estimées \hat{p}_i à se rapprocher de la probabilité empirique p_i d'appartenir à la classe +1. La probabilité \hat{p}_i est estimée à partir de la fonction logistique proposée par Platt pour convertir les distances $f(\mathbf{x})$ à l'hyperplan en probabilité \hat{p}_i (cf détails donnés dans l'annexe B.7.4.2(iv)), soit

$$\hat{p}_i = \varphi(f(\mathbf{x})) = \frac{1}{1 + \exp(-A f(\mathbf{x}) + B)} \quad (\text{B.5-2})$$

où A et B sont appris sur les exemples d'apprentissage. Dans cette étude, nous avons fixé $B=0$.

Le terme de contrainte prend alors la forme

$$p_i - \eta \leq \varphi(\mathbf{w}^T \mathbf{x}_i + b) \leq p_i + \eta \quad i=(1,2,..n) \quad (\text{B.5-3})$$

Cette relation exprime seulement que la probabilité a posteriori \hat{p}_i doit appartenir à l'intervalle de valeurs centré sur p_i et de largeur η paramétrant la précision de l'annotation.

On peut montrer que l'équation (B.5-3) peut s'écrire sous la forme

$$a \cdot z_i^- \leq \mathbf{w}^T \mathbf{x}_i + b \leq a \cdot z_i^+$$

$$\text{Avec } a = 1/A \text{ et } z_i^\pm = \log\left(\frac{1}{p_i \pm \eta} - 1\right)$$

En introduisant de nouvelles variables ressorts ξ_i^- et ξ_i^+ , on se ramène alors au problème suivant

$$\left\{ \begin{array}{l} \min_{\mathbf{w}, b, a, \xi_i^-, \xi_i^+} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i + C' \sum_i (\xi_i^- + \xi_i^+) \\ y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = (1, 2, \dots, m) \\ a \cdot z_i^- - \xi_i^- \leq \mathbf{w}^T \mathbf{x}_i + b \leq a \cdot z_i^+ + \xi_i^+, i = (m+1, \dots, n) \\ \xi_i \geq 0, \quad i = (1, 2, \dots, m) \\ \xi_i^- \geq 0 \text{ et } \xi_i^+ \geq 0, \quad i = (m+1, \dots, n) \end{array} \right. \quad (\text{B.5-4})$$

L'expression (B.5-4) est un problème de minimisation d'une fonction objectif de 6 variables vérifiant un ensemble de contraintes de boîtes. Dans sa thèse, Emilie Niaf a montré que ce problème pouvait s'exprimer sous forme d'un programme quadratique similaire à celui du SVM classique [Niaf et al., 2014].

B.5.1.1 Evaluation des performances

Une étude des performances du P-SVM a été réalisée sur des données synthétiques puis sur les données d'image IRM multiparamétriques de prostate. Ces performances ont été comparées à celle du SVM classique et à une forme dérivée appelé *Fuzzy SVM* (F-SVM) [Lin and Wang, 2002] dont le problème de minimisation est formulée de la manière suivante

$$\left\{ \begin{array}{l} \min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i m_i \xi_i \\ \text{s.c. } y_i (\mathbf{w}^T \mathbf{x}_i + b) > 1 - \xi_i, \quad i=(1,2,..n) \\ \xi_i > 0 \end{array} \right. \quad (\text{B.5-5})$$

Où $0 \leq m_i \leq 1$ pondère l'appartenance de l'exemple \mathbf{x}_i à la classe +1. Le principe de cet algorithme est de pondérer différemment le coût de mauvaise classification des exemples d'apprentissage en fonction de leur probabilité d'appartenance à la classe +1. Le facteur de pondération m_i est d'autant plus faible que l'incertitude sur l'étiquette de classe de l'échantillon est grande. Dans cette étude, nous avons utilisé $m_i = |2p_i - 1|$. Le problème (B.5-5) se ramène très simplement à la formulation classique des SVM et peut donc se résoudre par les mêmes algorithmes.

Les résultats de la Figure B.5-1 illustrent les performances de classification obtenues sur un exemple de données d'apprentissage simulées à partir de 100 points d'apprentissage tirés aléatoirement de deux distributions gaussiennes bidimensionnelles centrées en $(-0.3, -0.5)$ et $(0.3, 0.5)$ et de même variance $\sigma^2 = 0.7$. Les probabilités *a posteriori* d'appartenance à la classe +1 des points sont déterminées par la formule de Bayes (cf équation (B.7-54)) à partir des fonctions de distributions des deux classes et en supposant des probabilités *a priori* des deux classes $\mathbb{P}(Y=1)$ et $\mathbb{P}(Y=-1)$ de 0.5. On introduit alors un bruit additif constant δ sur les 100 points d'apprentissage, de telle sorte que les probabilités bruitées $\hat{\mathbb{P}}(Y=1|X=\mathbf{x}_i)$ sont estimées par

$$\hat{\mathbb{P}}(Y=1|X=\mathbf{x}_i) = \mathbb{P}(Y=1|X=\mathbf{x}_i) + \delta$$

Les performances sont mesurées sur une grille de 1000 points espacés régulièrement. L'algorithme P-SVM renvoie directement une probabilité d'appartenance à la classe +1 tandis les distances à l'hyperplan extraits des algorithmes F-SVM et SVM sont converties en probabilité par l'algorithme de Platt (cf B.7.4.2(iv)).

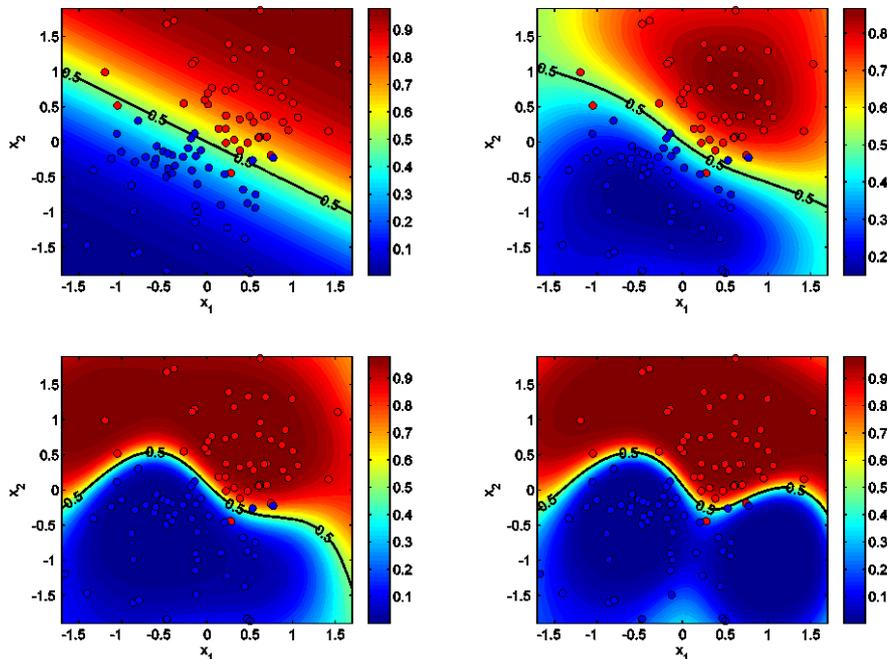


Figure B.5-1. (a) Distribution de probabilités d'appartenance à la classe +1 et comparaison des performances de classification de (b) P-SVM, (c) F-SVM et (d) SVM. Dans cet exemple $\delta = 0.15$

La Figure B.5-1 reporte les isocontours correspondant aux probabilités d'appartenance à la classe +1 ainsi que les fonctions discriminantes calculées par les trois méthodes SVM, P-SVM et F-SVM. Cet

exemple souligne que le P-SVM se comporte mieux que SVM et F-SVM dans cet exemple. Ce résultat visuel est confirmé par les métriques reportées dans le Tableau B.5-1 correspondant à l'aire sous la courbe ROC (AUC), la précision (Acc) et la divergence de Kullback-Leibler (KL). Notons que la dernière métrique, KL, est une mesure de similarité entre la distribution théorique, \mathbb{P} , et la distribution estimée \mathbb{Q} , d'après la formule

$$D_{KL}(\mathbb{P}\|\mathbb{Q}) = \sum_{i=1}^n \mathbb{P}(Y=1|X=\mathbf{x}_i) \log \left(\frac{\mathbb{P}(Y=1|X=\mathbf{x}_i)}{\mathbb{Q}(Y=1|X=\mathbf{x}_i)} \right)$$

Cette mesure tend vers 0 lorsque les deux distributions sont équivalentes.

Tableau B.5-1

Performance des différents algorithmes sur les données de synthèse

	P-SVM	SVM	F-SVM
AUC	1	0.97	0.99
Acc	0.98	0.93	0.96
KL	23	255	166

Nous avons ensuite évalué l'intérêt de la méthode P-SVM sur la base de données cliniques CLARA-P. Comme précisé en introduction, nous avons supposé que l'étiquette exacte des cibles n'était pas connue pour l'apprentissage. Dans cette hypothèse, chaque exemple de la base d'apprentissage a été étiqueté par la valeur moyenne p_i du score attribué par 5 lecteurs avec une incertitude η_i correspondant à l'écart-type du score expérimental. L'évaluation des performances est alors réalisée à partir de l'étiquette correcte de chaque exemple de la base de test déterminé à partir de la référence histologique. Le Tableau B.5-2 montre que les performances obtenues par le P-SVM sont significativement supérieures aux performances du SVM et légèrement supérieures aux performances obtenues avec la méthode F-SVM.

Tableau B.5-2

Performance des différents algorithmes sur les données cliniques de la base CLARA-P

	P-SVM	SVM	F-SVM
AUC	0.862	0.847	0.861
Acc	0.863	0.857	0.860

B.5.1.2 Discussion

Ce projet est un exemple intéressant de formalisation d'un problème spécifique en imagerie médicale dans le cadre théorique de l'apprentissage supervisé. Il souligne également l'efficacité de la recherche multidisciplinaire puisqu'il est le résultat d'une collaboration avec une équipe de cliniciens et une équipe de spécialistes des machines à noyaux.

La solution que nous avons proposée, le P-SVM, obtient de très bonnes performances par rapport au SVM et au F-SVM sur des données de synthèse. Ceci se traduit notamment par la mesure de divergence de Kullback-Leibler qui est significativement inférieure à celle mesurée avec le F-SVM dans le Tableau B.5-2. Ce résultat quantitatif se traduit également visuellement sur les isocontours de la Figure B.5-1, mais il est moins évident sur les mesures de performance de classification (AUC et Acc) qui sont moins sensibles à une erreur d'étiquetage. La différence de performance entre P-SVM

et F-SVM est d'autant plus importante que le niveau de bruit augmente dans les données. Nous renvoyons à la lecture de l'article [Niaf et al., 2014] pour plus de détails sur l'analyse comparative des deux méthodes.

Les performances de P-SVM sur les données cliniques sont également significativement meilleures que celles obtenues avec une version classique du SVM. En revanche, nous n'avons pas pu mettre en évidence de différence nette de performance avec la méthode F-SVM sur ce jeu de données. Comme évoqué pour les données de synthèse, les métriques de performances AUC et Acc sont moins sensibles aux erreurs sur l'estimation des étiquettes. Il est possible que le bruit d'étiquetage des données de la base CLARA-P soit faible et ne constitue donc pas un cas difficile permettant de différencier les deux algorithmes, F-SVM et P-SVM. Il aurait été intéressant, comme pour les données de synthèse, de modifier le bruit sur l'étiquetage des données expérimentales afin d'étudier le comportement des deux algorithmes dans des situations plus complexes. Par ailleurs, un des points limitants de cette étude concerne l'absence de comparaison avec la régression logistique qui permet d'estimer directement la probabilité *a posteriori*, ici, $\mathbb{Q}(Y = 1 | X = x_i)$. Il aurait été intéressant de rajouter cet élément de comparaison à l'étude.

Conclusion et perspectives

B.6 Conclusion et perspectives

B.6.1 Bilan scientifique

J'ai proposé, dans ce mémoire, une analyse critique des trois axes principaux de mon travail de recherche constitués par la simulation d'images médicales réalistes, l'évaluation des performances diagnostiques en imagerie nucléaire TEP et enfin, l'apprentissage statistique pour l'aide au diagnostic en imagerie multi-modalités. J'ai choisi d'apporter un éclairage différent sur mes contributions scientifiques dans ces trois thématiques. Pour les deux premiers axes, qui correspondent à des travaux plus anciens, je me suis intéressée à l'impact des principales contributions sur la communauté scientifique. J'ai souhaité, en revanche, décrire plus précisément les travaux plus récents dans le domaine de l'apprentissage statistique et des systèmes d'aide au diagnostic, en complétant les discussions proposées dans les publications par une analyse critique et des perspectives à considérer à moyen terme. Je reprends dans ce bilan, les points majeurs de chacun de ces trois axes et leur positionnement relatif dans mes perspectives de recherche.

B.6.1.1 Simulation

J'ai contribué activement au domaine de la simulation d'images TEP entre 1999 et 2009. Mon objectif était de valider des protocoles de simulations d'images corps entier réalistes afin de construire des bases d'images de synthèse pour mes travaux dans le domaine de l'analyse psychophysique puis de l'apprentissage. J'ai travaillé, d'une part, sur la modélisation réaliste d'imageurs TEP cliniques et précliniques, l'objectif étant d'ajuster les paramètres du simulateur afin de reproduire correctement les taux de comptage des différents types d'évènements simulés, que ce soit le signal utile (coïncidences vraies) ou le signal parasite (coïncidences diffusés, aléatoire) ainsi que les pertes liées aux limitations de l'électronique de détection (temps morts, etc). Cette contribution n'est pas détaillée dans ce mémoire mais a donné lieu à plusieurs publications dont les principales sont [Lartizien et al., 2007], [Reilhac et al., 2004]. J'ai également travaillé sur la construction d'un modèle anthropomorphique pour l'imagerie TEP en proposant des modèles de plus en plus réalistes comme rappelé dans la section B.2.2.2. Ce travail a été publié principalement dans les références [Tomeï et al., 2010], [Le Maitre et al., 2009]. Il a également donné lieu à la mise en ligne de la base de données simulées OncoPET_db https://www.creatis.insa-lyon.fr/oncoPET_DB/ qui suscite encore aujourd'hui l'intérêt de la communauté, principalement dans le domaine du traitement d'images.

Comme souligné dans la section B.2.4, l'orientation de mon activité vers la prise en compte d'informations multimodales a contribué au ralentissement de mes activités dans cette thématique. Malgré le réalisme des données TEP simulées, le modèle anthropomorphique XCAT de la base oncoPET_DB ne modélise pas les processus physiologiques à l'échelle cellulaire. Par conséquent, il n'est pas possible aujourd'hui de représenter la complexité du bruit physiologique, que l'on souhaite ensuite caractériser à l'aide des paramètres de texture. Par ailleurs et pour les mêmes raisons, la simulation réaliste d'images TDM nécessite également de disposer d'un modèle complexe de l'anatomie. Les travaux en simulation TDM que j'ai réalisés dans le cadre de l'ANR VIP ont en effet démontré la difficulté d'obtenir des simulations réalistes en imagerie anatomique TDM et/ou IRM, le point bloquant concernant la précision de la description du modèle anatomique.

J'ai essayé de poursuivre ces travaux en proposant une méthode de simulations hybrides qui consisterait à plaquer des modèles de texture extraits d'images cliniques sur les modèles géométriques de type XCAT. Les demandes de financement sur ce projet n'ont pas abouti, j'ai choisi de ne pas démarrer ce projet sans support humain car il s'éloigne trop de mon activité principale. Il serait cependant très utile de pouvoir disposer de bases d'images multimodalités corps entier. Je n'ai pas connaissance de projets parallèles en imagerie multi-modalités pour l'oncologie.

B.6.1.2 Evaluation et analyse psychophysique

J'ai consacré mon travail de thèse et de post-doctorat à la mise en place d'un cadre théorique pour l'évaluation des performances diagnostiques en imagerie nucléaire TEP corps entier pour l'oncologie puis à l'optimisation des protocoles d'imagerie sur la base de ce formalisme.

Les travaux formalisant l'analyse psychophysique en imagerie nucléaire ont eu un impact modéré, malgré un taux de citations honorable. Les équipes qui citent nos travaux reconnaissent la rigueur du formalisme mais soulignent la difficulté à mettre en œuvre ces protocoles qui imposent de lourdes contraintes: mobilisation de lecteurs, constitution d'une base d'échantillons de grande dimension pour garantir un minimum de puissance statistique. On peut cependant noter que quelques équipes, dont celles de El Fahkri à l'Université de Boston et Kardmas à l'Université de Salt Lake City publient régulièrement des études basées sur ce formalisme dans des revues de rang A dans le domaine de la médecine nucléaire.

Les résultats de l'optimisation des protocoles d'imagerie ont eu des retombées intéressantes, que ce soit sur le choix du mode d'acquisition ou de l'algorithme de reconstruction. Leur portée est cependant aujourd'hui plus limitée en raison des évolutions technologiques, en particulier avec l'introduction du temps de vol ou des machines hybrides, qui nécessitent une réactualisation constante des recommandations. Ce travail est aujourd'hui principalement réalisé par les deux équipes citées dans le paragraphe précédent.

Les travaux sur les modèles numériques 3D ont eu un impact important, à la fois dans la communauté scientifique qui s'intéresse au formalisme de ces modèles, indépendamment du domaine applicatif, et dans celle de l'imagerie nucléaire. La communauté de recherche dans ce domaine est très active, comme nous l'avons souligné dans la section B.3.5.3 et s'intéresse principalement à la quête de « l'observateur universel ».

B.6.1.3 Apprentissage statistique

J'ai souligné à plusieurs reprises le lien théorique entre les modèles numériques et l'apprentissage statistique. On peut cependant observer une séparation assez nette des communautés travaillant dans ces deux domaines, liée, selon mon interprétation, aux objectifs différents qu'elles poursuivent. Les premiers, en effet, se concentrent sur la recherche d'un modèle anthropomorphique permettant de s'affranchir de l'observateur humain pour 'alléger' les études psychophysiques. Les seconds visent à proposer un système complémentaire à l'observateur humain et capable de l'assister dans sa tâche diagnostique en le guidant vers des zones suspectes de l'image ou en lui fournissant un second avis.

J'ai choisi au milieu des années 2000, de m'intéresser à ces systèmes d'aide au diagnostic, encore très marginaux pour ne pas dire inexistant en imagerie nucléaire. Cette transition s'est effectuée assez naturellement grâce au bagage méthodologique de l'analyse psychophysique. Après quelques études avec différents algorithmes de classification dont l'analyse discriminante linéaire et les arbres de décision, j'ai choisi de privilégier les séparateurs à vaste marges, qui offrent un cadre théorique bien formalisé et ouvert à l'intégration de certaines spécificités de l'imagerie médicale. Je reviens également sur ce point dans la section suivante consacrée aux perspectives de mes travaux de recherche.

Les travaux en imagerie TEP/TDM du lymphome ont permis d'aboutir à un système CADx présentant de bonnes performances intrinsèques pour discriminer les lésions cancéreuses des lésions suspectes mais bénignes. La poursuite de ce projet a été stoppée par le départ de Fabien Ricard. Nous n'avons ainsi pas pu aller jusqu'à la mesure de l'impact clinique de ce système. Une suite intéressante à donner à ce projet serait également de considérer l'apport prédictif de ces modèles, en particulier dans le cadre du suivi thérapeutique. J'ai discuté dans la section 0, les travaux visant à extraire des marqueurs de la réponse au traitement dans des images TEP dites 'précoces', c'est-à-dire réalisées juste après la mise en place d'un traitement. Ceci afin de pouvoir détecter les patients non-répondeurs pour lesquels une réorientation thérapeutique rapide est souhaitable. L'analyse multivariée que nous avons développée dans un objectif d'aide au diagnostic, peut facilement être adaptée à ce type d'application. Il faudrait, dans ce cas, constituer une base d'apprentissage constitué de patients

disposant d'un examen TEP/TDM précoce et regroupés suivant le type de réponse au traitement (répondeur, non-répondeur, répondeur partiel etc..).

Les travaux en imagerie IRM multiparamétrique de la prostate illustrent bien la démarche globale dans la gestion de ce type de projet. Nous avons tout d'abord souligné l'importance de s'appuyer sur une base de données annotée, puis d'évaluer les performances intrinsèques d'un système complet et enfin d'aller jusqu'à la démonstration de l'intérêt clinique. La réussite de ce projet est liée, d'une part, à la bonne orchestration des tâches entre les équipes des différentes disciplines, mais aussi aux efforts de structuration des projets par des demandes de financements, qui ont permis, contrairement au projet précédent, de recruter deux étudiants en thèse et un post-doctorant. Concernant le projet de système CADx développé ces trois dernières années, il serait intéressant de le finaliser par une étude de type 'preuve du concept' préalable à la réalisation d'un produit logiciel propre. Cette étude devrait, entre autres, s'intéresser aux questions suivantes : sensibilité à la taille de la base de données, auto-adaptation de l'apprentissage avec rajouts de nouveaux exemples, sensibilité aux types de données (différents imageurs, différentes séquences). Ce type d'étude nécessite l'appui d'un industriel. Nous réfléchissons au moyen de formaliser ce projet de valorisation industrielle. L'étape suivante, engagée avec la thèse de Jérôme Lehaire, consiste à développer un système d'aide à la détection (CADE). Je précise ce projet dans la section suivante.

Ces différents projets ont permis de réaliser des avancées méthodologiques intéressantes dans le domaine des CAD, de mettre en place un protocole d'évaluation rigoureux pour approcher au mieux les performances intrinsèques en généralisation et d'aboutir enfin à des algorithmes adaptés au besoin des cliniciens. L'interaction forte avec la problématique clinique constitue également un terrain riche pour comprendre les limites des systèmes proposés et stimuler des travaux plus théoriques dans l'objectif d'adapter le formalisme de l'apprentissage statistique aux spécificités de l'imagerie médicale. J'ai illustré cette démarche dans le contexte des travaux sur l'incertitude d'étiquetage des données d'IRM présentés dans la section B.5.1. C'est résolument dans cette voie que je compte poursuivre mes travaux à moyen terme.

B.6.2 Encadrement et administration de la recherche

Les résultats de recherche que j'ai obtenus depuis mon recrutement au CNRS en 2004 sont le fruit d'une recherche propre, en particulier dans le domaine des CAD pour l'imagerie nucléaire du cancer et de travaux réalisés par les étudiants en thèse que j'ai encadrés au cours de ces années. J'ai pris soin de préciser les acteurs de chacun des projets détaillés dans la section A.6, une présentation plus quantitative de cette activité d'encadrement est donnée dans la section A.5.

Ces travaux ont été publiés dans des revues internationales de rang A et reconnus par la communauté scientifique comme l'atteste le taux de référencement élevé de mes publications les plus anciennes. Les indicateurs positifs concernent également le prix poster de la conférence SPIE en 2012 pour les travaux de thèse de Mlle Niaf, par exemple.

J'ai légèrement réorienté ma thématique de recherche à partir de 2008 à la fois sur le plan méthodologique, en m'intéressant plus spécifiquement à la classification supervisée, et sur le plan applicatif en considérant la notion d'imagerie multimodalités, multiparamétrique et multidimensionnelle. La visibilité scientifique de ces travaux s'accroît grâce aux publications récentes dans le domaine des systèmes d'aide au diagnostic. J'ai introduit et stimulé l'émergence de ce thème de recherche au sein de l'équipe 'images et modèles' de CREATIS. Il s'inscrit aujourd'hui comme l'un des thèmes affichés de l'équipe dans le cadre du prochain contrat quinquennal (renouvellement de l'unité en janvier 2016).

J'anime depuis deux ans un groupe de travail au sein de l'équipe constitué par les étudiants en thèse, les ingénieurs et les stagiaires que je co-encadre. Au cours de ces travaux, j'ai co-encadré 5 doctorants (avec un taux de co-encadrement > 75%) ainsi qu'un à deux stagiaires de master par an.

Mes travaux de recherche sont interdisciplinaires et reposent sur un partenariat avec des collaborateurs cliniciens qui guident les développements vers des problématiques d'intérêt clinique fort. J'ai toujours cherché à structurer des collaborations au niveau régional et national et à proposer des contributions à la fois originales sur le plan théorique et pertinentes sur le plan applicatif. Je souhaite à présent aller plus loin dans la proposition de stratégies de classification adaptées aux

spécificités de l'imagerie médicale et me rapprocher d'équipes de statisticiens ou mathématiciens appliqués dans le domaine de l'apprentissage supervisé. Ce point est développé dans l'exposé de mes perspectives de recherche.

Je mène parallèlement un travail d'animation de la recherche au niveau régional, dans le cadre du WP4 du labex PRIMES au sein duquel j'ai encouragé la reconnaissance d'un groupe de travail sur la classification. Cette activité a été soutenue récemment par le financement de la thèse de Mlle El Azami. Je suis également responsable adjointe de l'équipe « Images et modèles » de CREATIS et m'implique de manière croissante dans l'animation et le management de la recherche au niveau régional (labeX PRIMES et EquipeX LILI) et au niveau national (FLI).

B.6.3 Perspectives de recherche

Mon projet de recherche vise à améliorer les performances diagnostiques des systèmes d'aide à la décision en proposant des méthodes par apprentissage adaptées aux spécificités des données d'imagerie médicale. L'objectif *in fine* est de pouvoir fournir au clinicien une information précise et propre à chaque patient sur la **localisation** (cartographie) et la **caractérisation** (stadification) de la pathologie.

Mes travaux de recherche, au cours des cinq prochaines années, dans le domaine de l'apprentissage supervisé vont s'articuler autour des trois thématiques principales que je détaille dans la suite de cette section. Je présente, pour conclure, les projets de recherche plus cliniques sur lesquels je souhaite adosser ces développements.

B.6.3.1 Axes de recherche

Mes perspectives de recherche à moyen terme dans le domaine de l'apprentissage statistique se déclinent autour des axes suivants :

(i) ***Vers une cartographie probabiliste et multiclasse de la pathologie***

Les travaux présentés dans la section B.4.3.1 sur le système CADe pour l'imagerie TEP du lymphome ont souligné la nécessité de développer des stratégies permettant de réduire le taux de fausses détections (faux positifs). L'état de l'art, dans différentes applications médicales, confirme ce constat ; les systèmes actuels, qui réalisent une analyse automatique de l'image et produisent une carte paramétrique de la pathologie ne sont pas assez spécifiques pour aider les cliniciens dans leur tâche diagnostique.

L'une des solutions discutée dans la section B.4.3.1 consiste à enrichir la base d'attributs discriminants, par exemple, en intégrant des paramètres extraits de différentes modalités d'imagerie. L'autre approche, que je souhaite développer, consiste à tenir compte des relations spatiales dans l'image au cours de l'apprentissage du modèle de décision. On peut, en effet raisonnablement supposer que deux voxels proches spatialement, appartiennent à la même classe. Il s'agit alors de formuler cet a priori dans le modèle de classification. Le formalisme des SVM offre un cadre méthodologique bien adapté pour cela.

Nous avons réalisé une première étude dans ce domaine, basée sur l'ajout d'un terme de régularisation spatiale de la fonction objectif des SVM [Niaf et al., 2014], conduisant à l'équation (B.6-1)

$$\min_{f \in H} \frac{1}{n} \sum_{i=1}^n L(y_i, f(\mathbf{x}_i)) + \lambda \Omega(f) + \lambda_s \sum_{i,j=1}^N S_{i,j} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 \quad (\text{B.6-1})$$

Où $\mathbf{S} \in \mathcal{M}_N(\mathbb{R})$ est la matrice définissant les règles de connexions entre les N voxels de l'image.

$$\begin{cases} S_{i,j} = 1 \text{ si } \mathbf{x}_i \text{ et } \mathbf{x}_j \text{ sont connectés} \\ S_{i,j} = 0 \text{ sinon} \end{cases}$$

Ce terme force les classes prédites $f(\mathbf{x}_i)$ et $f(\mathbf{x}_j)$ des points d'apprentissage \mathbf{x}_i et \mathbf{x}_j à être proches si ces échantillons sont liés par une relation d'adjacence définie par la matrice S .

Nous avons montré que ce problème pouvait être reformulé comme un programme quadratique similaire au problème SVM standard. Dans cette étude, présentée à la conférence IEEE ICIP à l'automne 2014, nous avons modélisé une relation de connectivité simple entre les 8 voxels plus proches voisins, sans tenir compte de contraintes anatomiques. Les résultats préliminaires présentés sur la Figure B.6-1 soulignent l'apport de cette méthode sur la qualité visuelle des cartes probabilistes générées.

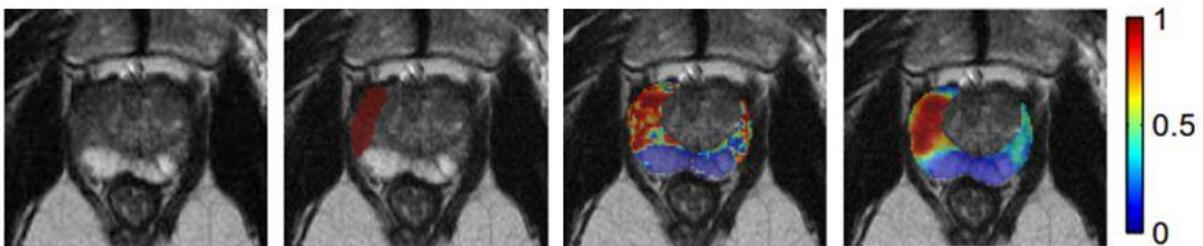


Figure B.6-1. Illustration des performances du modèle SVM avec a priori sur les relations entre voxels. De gauche à droite, coupe transverse de prostate en IRM T2-w, coupe similaire sur laquelle a été superposé un masque indiquant la présence d'une région cancéreuse, carte de probabilité estimée par un SVM classique, carte de probabilité estimée par un SVM avec a priori sur les voxels 8-connexes.

Ces premiers résultats sont encourageants, même si le gain en performance diagnostique intrinsèque n'est pas significatif pour cette application. La prise en compte de relations plus réalistes devrait permettre d'améliorer ces résultats. Il faudra cependant veiller à proposer des modèles d'a priori compatibles avec la mise en forme d'une fonction objectif que l'on puisse reformuler sous forme d'un problème quadratique standard. Outre la complexité de mise en forme du problème d'optimisation, ces travaux nécessitent également de considérer les contraintes calculatoires imposées par ces modèles d'apprentissage sur des données de grande dimension. Les collaborations que j'entretiens à la fois avec des équipes spécialistes des machines à noyaux (LITIS, OCA) et du calcul haute performance (CREATIS) offrent un environnement adapté au déploiement de ce type de projet.

Je souhaite également développer des méthodes de classification permettant d'enrichir l'information renvoyée par le système d'aide au diagnostic, qui repose jusqu'à présent principalement sur un système de classification binaire estimant un score (parfois traduit en terme de probabilité) d'appartenance à la classe pathologique. L'une des pistes que nous étudions en imagerie du cancer, serait par exemple, de pouvoir compléter la réponse du système CAD avec un indice corrélé avec un marqueur biologique de l'agressivité tumorale, tel que le score de Gleason, par exemple. Certaines pistes de recherche dans ce domaine, en cours d'étude dans le cadre de la thèse de Jérôme Lehaire, sont détaillées dans la section B.6.3.2(i).

(ii) *Apprentissage sur des données d'imagerie hétérogènes*

La base de mon projet de recherche est d'exploiter la complémentarité des informations extraites de différentes modalités d'imagerie pour enrichir les performances diagnostiques, en particulier la spécificité, des systèmes d'aide à la décision. La problématique de l'apprentissage sur des données multimodales n'est pas triviale, surtout lorsque l'on envisage, par exemple, d'associer des données de nature hétérogène, telles que des volumes matricielles d'images TEP ou IRM associés à des signaux

de magnétoencéphalographie (MEG) représentés sous forme d'un maillage. Il faut alors s'interroger sur la manière de fusionner ces données.

Plusieurs stratégies sont à considérer. La stratégie la plus simple, que nous avons adoptée jusqu'à présent dans les projets présentés dans la section B.4.3, consiste à concaténer les différents attributs, quel que soit le type d'image. Cette approche, dite de fusion précoce, semble acceptable lorsque les caractéristiques sont de même nature, dans ces études, principalement des paramètres de texture. On peut s'interroger cependant sur l'impact de cette fusion dans le projet d'imagerie de la prostate qui mélange des caractéristiques extraites de l'image avec des paramètres biologiques issus de modèles pharmacocinétiques. Une alternative à la fusion précoce consiste à réaliser une classification séparée pour chaque classe d'attribut et à fusionner ensuite les résultats de chaque classifieur. On parle, dans ce cas, de fusion tardive. Une autre piste intéressante dans le cadre spécifique des SVM, concerne l'apprentissage de noyaux multiples ou multiple kernel learning (MKL) en anglais [Lanckriet et al., 2004] qui permet d'utiliser un type de noyau par classe d'attribut au sein du même problème d'optimisation. Ces différentes pistes seront explorées en s'adossant aux projets d'application cliniques présentés dans la section B.6.3.2.

(iii) Apprentissage avec des données déséquilibrées

Le troisième axe que je souhaite développer concerne la prise en compte d'une autre spécificité des bases d'apprentissage en imagerie médicale, qui concerne la gestion de bases de données déséquilibrées, c'est-à-dire ne présentant pas une distribution homogène des échantillons de chaque classe.

Une piste intéressante pour aborder ce problème consiste à considérer le problème de classification comme un *problème de détection de changement*, appelé aussi détection d'anomalies (outlier detection, en anglais). Contrairement à la problématique de classification en plusieurs classes, ces algorithmes apprennent uniquement une représentation de la classe normale puis évaluent l'appartenance d'un échantillon test à cette distribution. Les algorithmes de type one-class SVM [Schölkopf et al., 2001] ou SVDD pour support vector data description [Tax and Duin, 2004] offrent une cadre dérivé des SVM pour ce type de problématique. Ce projet est mené par Meriem El Azami dans le cadre de sa thèse sur laquelle je reviens dans la section suivante.

L'apprentissage de la normalité peut être une alternative intéressante dans la gestion des bases de données déséquilibrées (imbalanced datasets) telles que nous les rencontrons en imagerie médicale. Dans le cas d'une analyse par voxel, en effet, la grande majorité des voxels appartiennent à la classe normale, les zones pathologiques représentant uniquement une fraction faible du volume totale de l'image. En imagerie IRM du cancer de la prostate, la série de 35 patients de la base CLARA-P décrite dans la section est constituée de 80% de voxels sains, contre 7% de voxels correspondant à des zones suspectes mais bénignes et 3% de voxels correspondant à des lésions cancéreuses. Les méthodes de type détection de changement peuvent ainsi être intéressantes pour séparer, dans une première phase, la majorité des échantillons normaux des échantillons pathologiques et réduire ainsi le déséquilibre des différentes classes avant d'appliquer une stratégie de classification binaire ou multiclasse. D'autres approches pour la gestion de données déséquilibrées sont également envisageables, dont les méthodes de type AdaBoost [Freund and Schapire, 1997] qui permettent de sélectionner les échantillons d'intérêt.

B.6.3.2 Projets

Les différents axes de recherche que j'ai évoqués seront déclinés à court et moyen terme dans le domaine de l'imagerie du cancer, dans la continuité du projet que je mène depuis cinq ans avec l'équipe du Pr Rouvière sur la cartographie du cancer de prostate. Je souhaite également m'investir au cours des prochaines années dans le domaine de la neuroimagerie. Cette orientation est déjà prise dans le cadre d'une collaboration naissante avec une équipe du Centre de Recherche en Neurosciences de Lyon (CRNL) autour d'un projet sur l'imagerie multimodale de l'épilepsie pharmacorésistante portée par le travail de thèse de Meriem El Azami que je résume dans la suite de ce paragraphe.

(i) **Cartographie 4D du cancer de prostate basée sur l'imagerie IRM et échographique**

Ce projet s'inscrit dans la continuité du projet de thèse d'Emilie Niaf et du projet CARTOGRAPHIX financé par l'INCa entre 2012 et 2014 (voir section A.6.1). Il s'organise à court terme autour de la thèse de Jérôme Lehaire (voir section A.5.3) et du projet LYRIC (voir section A.6.1).

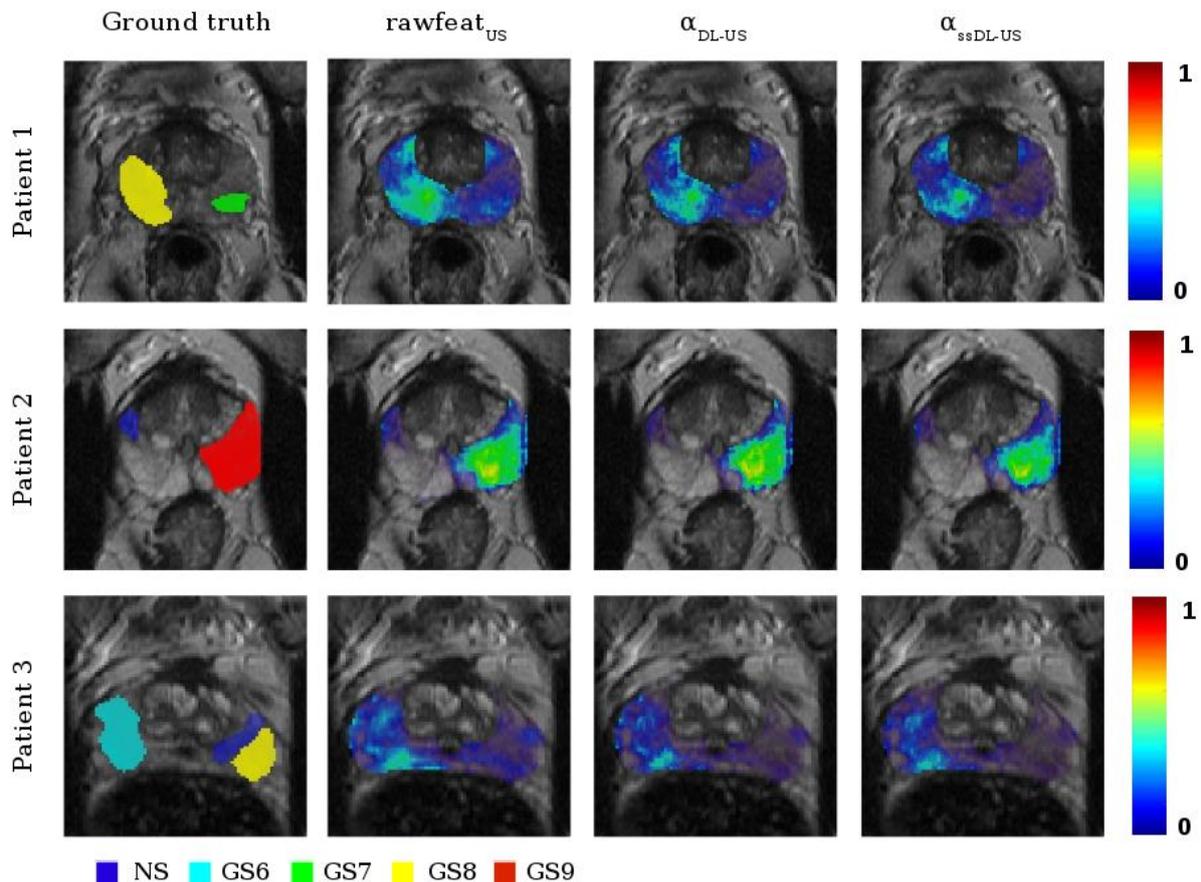


Figure B.6-2. Coupes transverses d'acquisitions IRM T2-w de prostate de 3 patients présentant différents types de lésions : NS = suspect mais bénin, GS6 = Gleason 6 etc. Les images situées dans la colonne de gauche indiquent la localisation et le type de lésions. Les images des colonnes suivantes reportent les cartes de probabilités estimées par un SVM linéaire associé à trois types de caractéristiques : $rawfeat$, α_{DL} et α_{SSDL} .

Le travail de thèse de Jérôme Lehaire, que j'ai évoqué dans la section précédente, porte sur la cartographie 4D du cancer de prostate. Jérôme Lehaire, a dans cet objectif abordé les problèmes spécifiques du prétraitement des données, et en particulier du recalage impératif des données IRM multiparamétriques, puisque l'apprentissage est réalisé au niveau du voxel. Ce travail a donné lieu à une communication orale à la conférence ESMRMBM en 2013 [Lehaire et al., 2013]. Ce sujet est également l'occasion de s'intéresser à la problématique de l'apprentissage sur des données déséquilibrées et sur les stratégies d'apprentissage multiclasse permettant de raffiner la classification probabiliste bi-classe (pathologique versus sain) en fournissant un indice d'agressivité tumorale. Des premiers résultats de cartographies probabilistes du cancer de prostate ont été présentés lors de la conférence ICIP 2014. Les cartes de la Figure B.6-2 illustrent les performances visuelles de plusieurs stratégies de classification combinant des SVM linéaires ou non linéaires avec différents types de caractéristiques discriminantes. Les caractéristiques $rawfeat$ correspondent à une sélection de 70 caractéristiques décrites dans la section B.4.3.2(ii) et le papier [Niaf et al., 2012]. Ces attributs sont

extraits des trois types de séquences IRM (T2-w, ADC et DCE) et comprennent des statistiques de 1^{er} et 2nd ordre ainsi que des mesures de gradients et des paramètres semi-quantitatifs. Les deux autres types de caractéristiques ont été extraits à partir de méthodes de codage parcimonieux (sparse coding) qui permettent de représenter le vecteur de caractéristiques comme une combinaison linéaire de quelques éléments d'un dictionnaire. Les coefficients utilisés pour le problème de classification sont les coefficients de pondération de chaque élément du dictionnaire. La méthode que nous avons utilisée consiste non seulement à estimer ces coefficients mais également à apprendre les éléments du dictionnaire à partir de données d'entraînement, de manière non-supervisée (α_{DL}) ou semi-supervisée (α_{SSDL}) [Mairal et al., 2012]. Nous renvoyons à la lecture de la référence [Lehaire et al., 2014] pour plus de détails méthodologiques. Ces cartes soulignent les bonnes performances visuelles observées. On remarquera, par exemple que la lésion suspecte mais non cancéreuse (NS) est bien étiquetée avec une probabilité de malignité inférieure à 0.5, et que la probabilité moyenne d'agressivité est une fonction croissante du score de Gleason. Comme évoqué dans la section précédente, ces premiers résultats constituent une base solide pour proposer de nouvelles stratégies permettant de mieux évaluer et affiner la corrélation entre les scores du CAD et l'indice d'agressivité du cancer.

Le soutien du programme LYRIC permet également de nous intéresser à un autre aspect du projet consistant à intégrer des attributs extraits de données échographiques au système CAD actuel développé pour l'IRM. Le recrutement en octobre 2014 d'un ingénieur sur ce projet va nous permettre, à court terme, de disposer d'une chaîne de traitement de l'information permettant de recalibrer l'image ultrasonore sur les données IRM multiparamétrique et d'extraire ensuite les attributs des deux modalités.

(ii) Cartographie 3D des foyers épileptogènes dans le bilan préopératoire des épilepsies pharmacorésistantes

J'ai initié, il y a deux ans, un projet visant à développer un système CADe pour la neuroimagerie de l'épilepsie pharmacorésistante. Ce projet est réalisé avec une équipe du Centre de Recherche en Neurosciences de Lyon (CRNL), et l'assistance de Nicolas Costes du CERMEP et d'Alexander Hammers du Kings College à Londres. Il est fédéré par le travail de thèse de Meriem El Azami, financé dans le cadre du labex PRIMES. Cette thèse vise proposer une méthode d'analyse statistique multivariée qui détecte et caractérise de manière automatique les anomalies cérébrales responsables de crises d'épilepsie pharmacorésistantes, à partir de la combinaison de différentes modalités d'imagerie dont l'IRM anatomique, la TEP et la Magnétoencéphalographie (MEG). L'ambition de ce projet est de pouvoir assister à moyen terme l'équipe médicale dans l'établissement du planning chirurgical. A plus long terme, et compte-tenu des performances intrinsèques que nous pourrions mesurer, nous espérons que ce système non invasif puisse remplacer l'examen invasif de stéréoelectroencéphalographie (SEEG) basé sur l'enregistrement du signal électrique cérébral à l'aide d'électrodes implantées en profondeur, et qui constitue aujourd'hui la méthode de référence.

Les anomalies que nous souhaitons détecter ne présentent pas un motif spatial reproductible d'un patient à l'autre, et n'ont pas de localisation privilégiée. Par ailleurs, il est très difficile d'avoir accès à une vérité terrain fiable. Les patients opérés ont un suivi sur plusieurs mois permettant de quantifier le niveau de réussite de l'opération. On utilise une échelle qualitative allant de 'libre de crise' lorsque le patient n'est plus sujet à aucune crise à 'état dégradé' lorsque les crises n'ont pas diminué et sont accompagnées d'une dégradation de l'état général liée aux séquelles opératoires (perte de certaines fonctions suite à l'opération etc). Le problème de la vérité terrain que nous avons discuté à plusieurs reprises dans ce rapport est donc particulièrement marqué dans cette application. Nous nous sommes par conséquent orientés vers des méthodes de détection de changement, qui permettent de discriminer les individus dits 'normaux' de l'ensemble des 'autres' ('outliers') rassemblant différentes classes de pathologie. Mlle El Azami a proposé un premier prototype de système CADe basé sur un modèle oneClass SVM associé à des caractéristiques extraites d'acquisitions IRM pondérées en T1. Les performances du modèle entraîné sur une base de 38 témoins ont été comparées aux performances obtenues par une analyse statistique univariée de type SPM (statistical parametric mapping) qui est aujourd'hui la méthode de référence en neuroimagerie. Cette comparaison a montré que, pour le type de lésions considérées, le système proposé était plus sensible et plus spécifique,

comme l'illustre, par exemple, la Figure B.6-3. Ce travail a donné lieu à une présentation à la conférence IEEE Pattern Recognition in Neuroimaging (PRNI) en juin 2013 [El Azami et al., 2013].

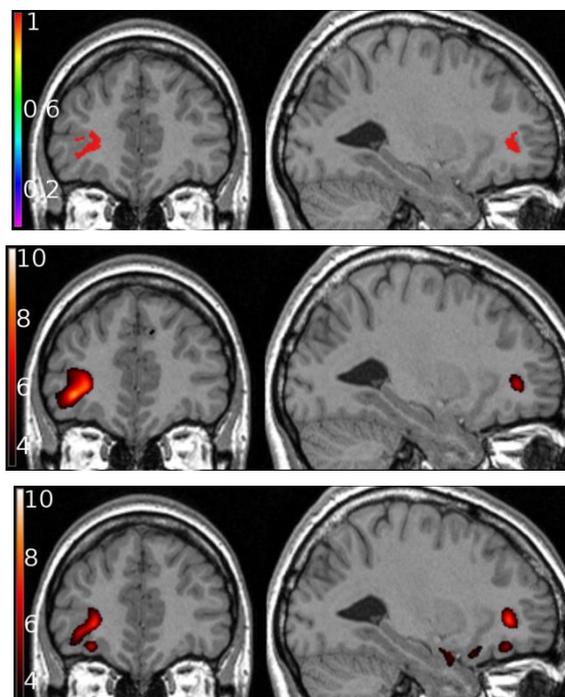


Figure B.6-3. Cartographie probabiliste estimée à partir d'un modèle de type OC-SVM basée sur 6 caractéristiques extraites de l'image IRM T1 (haut). Cartes de score Z estimées par une analyse SPM en utilisant la mesure d'extension (milieu) ou de jonction (bas) comme variable explicative.

Les challenges méthodologiques générés par cette application sont multiples. Mlle El Azami s'est, par exemple, intéressée à la stabilité des méthodes de détection de changement en présence de bruit dans la classe d'apprentissage, fréquent en imagerie médicale. Cette problématique a donné lieu à la proposition d'un algorithme original dérivée de la méthode SVDD, dont les premiers résultats ont été présentés dans une conférence internationale en intelligence artificielle [El Azami et al., 2014]. Ce sujet donne également l'opportunité de s'intéresser à la prise en compte des corrélations entre voxels voisins, et des données hétérogènes.

(iii) *Recherche propre et collaborations*

Ma contribution dans les deux projets précédents consiste principalement à structurer la collaboration scientifique, veiller à l'interaction entre les différentes communautés, définir et formaliser les axes de recherche en apprentissage et encadrer le travail des étudiants et/ou postdoctorants. Je souhaite, parallèlement à ces projets, conserver un axe de recherche propre. Comme évoqué précédemment, j'aimerais m'intéresser au problème d'apprentissage sur des données hétérogènes, en particulier dans le domaine de la neuroimagerie. Je m'intéresse, par exemple, à la connectomique qui permet d'étudier l'existence de réseaux fonctionnels entre les différentes structures cérébrales et leur altération dans le cadre de pathologies neurologiques. J'espère ainsi développer mon réseau de collaboration régional en neuroimagerie et stimuler de nouveaux projets dans ce domaine, par exemple autour de la plate-forme d'imagerie hybride TEP/IRM LILI.

Une autre piste serait également de stimuler des projets autour de la recherche de modèles prédictifs discutée. Comme je l'ai discuté dans la section B.4.3.1(iv) pour l'imagerie oncologique, la communauté scientifique s'intéresse aujourd'hui à la recherche d'un ou plusieurs biomarqueurs précoces de la réponse au traitement. Il me semble qu'une approche intégrative, basée sur des

méthodes d'analyses multivariées, s'avère plus pertinente pour répondre à ce challenge. La notion de modèle prédictif est également très pertinente dans le domaine de l'imagerie neurologique, en particulier pour le suivi et la prédiction de l'évolution de certaines maladies neurodégénératives, telles que la maladie d'Alzheimer ou la sclérose en plaques.

Annexes

B.7 Annexes

B.7.1 Notations et rappel de définitions

B.7.1.1 Formulation

On se place dans un problème de classification en L classes indexées par l'étiquette $l_k \in \mathbb{Z}, k \in \mathbb{N}$. Soit un ensemble de n individus dans un espace de dimension p , $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ notés $\mathbf{x}_i = (x_i(1), \dots, x_i(p))^t$ de centroïde $\bar{\mathbf{x}} \in \mathcal{X}$. Chaque individu appartient à une classe $y_i \in \mathcal{Y} = \{l_0, \dots, l_{L-1}\}$. On note n_k le nombre d'individus dans la classe l_k et $\bar{\mathbf{x}}_k \in \mathcal{X}$ le centre de gravité de l'ensemble des individus de la classe l_k .

On suppose que les individus sont regroupés par classe, soit

$$\forall \mathbf{x}_i \in \{\mathbf{x}_{n_{k-1}+1}, \dots, \mathbf{x}_{n_{k-1}+n_k}\}, \text{ alors } y_i = l_k$$

Dans le cas où $L=2$, on pose $l_0=-1$ et $l_1=1$.

B.7.1.2 Rappel de définitions

Soit \mathcal{M}_p l'ensemble des matrices carrées de dimension p .

Soit $\mathbf{S} \in \mathcal{M}_p(\mathbb{R})$, $\mathbf{u} \in \mathbb{R}^p$ et $\mathbf{v} \in \mathbb{R}^p$.

On peut montrer que

$$\mathbf{u}^t \mathbf{S} \mathbf{u} = tr(\mathbf{U} \mathbf{S}) \text{ avec } \mathbf{U} \equiv \mathbf{u} \mathbf{u}^t \in \mathcal{M}_p(\mathbb{R})$$

et

$$(\mathbf{u}^t \mathbf{v})^2 = tr(\mathbf{V} \mathbf{U}) \text{ avec } \mathbf{V} \equiv \mathbf{v} \mathbf{v}^t \in \mathcal{M}_p(\mathbb{R})$$

La matrice de variance-covariance totale $\mathbf{V} \in \mathcal{M}_p(\mathbb{R})$ est définie par

$$\mathbf{V} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t \quad (\text{B.7-1})$$

La matrice de variance intra-classe $\mathbf{S}_w \in \mathcal{M}_p(\mathbb{R})$ est définie par

$$\mathbf{S}_w = \frac{1}{n} \sum_{k=1}^L n_k \mathbf{S}_w^k \quad (\text{B.7-2})$$

avec $\mathbf{S}_w^k \in \mathcal{M}_p(\mathbb{R})$ la matrice de variance-covariance pour la classe l_k définie par

$$\mathbf{S}_w^k = \frac{1}{n_k} \sum_{i=n_{k-1}+1}^{n_{k-1}+n_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T \quad (\text{B.7-3})$$

La matrice de variance interclasse, $\mathbf{S}_b \in \mathcal{M}_p(\mathbb{R})$ est définie par

$$\mathbf{S}_b = \frac{1}{n} \sum_{k=1}^L n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^t \quad (\text{B.7-4})$$

Dans le cas ou $L=2$, on peut montrer que

$$\mathbf{S}_b = \frac{n_0 n_1}{n^2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)^t \quad (\text{B.7-5})$$

Et que la variance totale $\mathbf{S} \in \mathcal{M}_p(\mathbb{R})$ est égale à

$$\mathbf{V} = \mathbf{S}_w + \mathbf{S}_b \quad (\text{B.7-6})$$

B.7.2 Théorie de la décision, analyse ROC et dérivées

L'évaluation de la 'qualité' en imagerie médicale est un problème complexe et encore sujet à de multiples interprétations subjectives. Jusqu'au début des années 80, les études portent principalement sur les caractéristiques physiques de l'image, telles que la fonction de modulation de transfert (MTF) ou le spectre de puissance du bruit (NPS). Pourtant, dès la fin des années 60, certaines équipes soulignent l'importance de considérer l'influence de l'observateur humain [Rossmann, 1966, Biberman, 1973]. Cette prise de conscience est une étape fondamentale qui se formalise vingt ans plus tard par l'adaptation des principes psychophysiques établis dans le domaine de la détection du signal [Wagner and Brown, 1985].

Les études psychophysiques consistent à évaluer les performances d'un observateur dans une tâche de détection. Contrairement aux critères dérivés uniquement des caractéristiques physiques de l'image, cette méthodologie intègre l'influence de l'observateur dans le processus de décision. Ainsi, il est possible d'obtenir un critère objectif de détectabilité à partir de la méthodologie par l'intermédiaire d'études psychophysiques appelées études ROC, de l'anglais 'Receiver Operating Characteristics curves' ou dérivées (LROC, FROC, AFROC) basées sur des modèles dérivés de la théorie de la décision.

B.7.2.1 Théorie de la décision statistique et métriques associées

La théorie de la décision tente de déterminer, à partir de données psychologiques, économiques, sociologiques, etc., le comportement optimal d'un agent face à une situation de choix à l'aide, notamment, de modèles mathématiques [Van Trees, 1968]. Cette théorie est à la base de l'analyse conceptuelle de la détection du signal, et permet de prédire ou d'optimiser l'attitude d'un individu ou d'un système programmé, appelé observateur, ou décisionnaire, et chargé de classer un ensemble d'éléments (ex : images cliniques) dans un ensemble de catégories (ex : « présence d'anomalies », « absence d'anomalies ») [Green and Swets, 1966].

La théorie de la décision statistique propose un modèle hiérarchique de la prise de décision. Le premier niveau de ce modèle consiste à extraire un paramètre des données, appelé **variable de décision** λ à partir de l'image $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$. Ce processus est modélisé par la fonction f

$$\begin{aligned} f : \mathbb{R}^p &\rightarrow \mathbb{R} \\ \mathbf{x}_i &\mapsto \lambda_i \end{aligned} \tag{B.7-7}$$

Cette variable $\lambda_i = f(\mathbf{x}_i)$ caractérise la probabilité subjective de l'existence d'un signal. La seconde étape du mécanisme de prise de décision consiste à comparer la valeur de λ_i déterminée par l'observateur à une valeur seuil λ_s , appelée **critère** ou **seuil de décision**, par l'intermédiaire d'un test statistique binaire. Le cas étudié est alors considéré comme positif si λ_i est supérieur à λ_s .

La valeur du seuil λ_s dépend, entre autres, des conditions expérimentales et des caractéristiques de l'observateur. Le modèle de décision suppose que l'ensemble des valeurs de λ_i évaluées sur des cas positifs suit une distribution de probabilité $\mathbb{P}(f(\mathbf{x}_i) = \lambda_i | y_i = I_1)$ de valeur moyenne $\bar{\lambda}_1$ et d'écart type σ_1 . De même, les valeurs de λ_i évaluées sur des cas négatifs se répartissent selon une distribution de probabilité $\mathbb{P}(f(\mathbf{x}_i) = \lambda_i | y_i = I_0)$ de valeur moyenne $\bar{\lambda}_0$ et d'écart type σ_0 . On admet également que les valeurs de λ_i sont en moyenne plus élevées pour les cas positifs, c'est-à-dire $\bar{\lambda}_1 \geq \bar{\lambda}_0$.

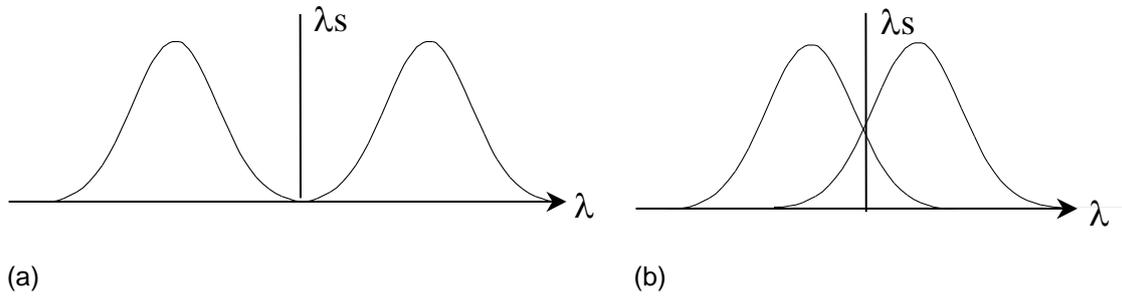


Figure B.7-1 : Distribution de probabilité de la variable de décision dans le cas positif et négatif (a) sans recouvrement et (b) avec recouvrement

La Figure B.7-1 représente deux configurations possibles des distributions positives et négatives. Nous prenons pour cette illustration le cas particulier de distributions gaussiennes mais le modèle général ne fait pas d'hypothèse sur la forme de ces distributions. L'axe horizontal est appelé axe de décision. La Figure B.7-1 (a) illustre le cas idéal de deux distributions qui ne se superposent pas. La ligne verticale séparant les deux distributions représente une valeur possible λ_s du critère de décision. Cette valeur est optimale pour cet exemple car elle permet d'obtenir une séparation parfaite. En réalité, la répartition des deux profils de la variable de décision présente souvent une zone de recouvrement, illustrée par la Figure B.7-1 (b). Ce recouvrement permet de modéliser l'erreur de décision. Dans cet exemple, toute valeur λ supérieure à λ_s et appartenant à la distribution des cas négatifs sera classée par erreur dans la catégorie des cas positifs.

Ce modèle de mécanisme de décision permet mesurer la capacité de l'observateur à différencier les deux états (positifs et négatifs) en évaluant le degré de séparation des deux distributions. Celui-ci est mesuré par l'indice de détectabilité d

$$d = \frac{\lambda_1 - \lambda_0}{\sqrt{P_1\sigma_1^2 + P_0\sigma_0^2}} \quad (\text{B.7-8})$$

Où P_i est la probabilité a priori de la classe l_i avec $\sum_{i=0}^1 P_i = 1$

En pratique, on fixe généralement $P_i = 1/2$ soit

$$d = \frac{\lambda_1 - \lambda_0}{\sqrt{\frac{\sigma_1^2 + \sigma_0^2}{2}}} \quad (\text{B.7-9})$$

On définit également le rapport signal-sur-bruit (SNR) comme

$$SNR^2 = P_1 P_0 d^2 = P_1 P_0 \frac{(\lambda_1 - \lambda_0)^2}{P_1 \sigma_1^2 + P_0 \sigma_0^2} \quad (\text{B.7-10})$$

Dans la suite, nous nous intéressons aux applications de la théorie de la décision dans le contexte de la détection d'un signal pathologique en imagerie médicale. Nous qualifions de *cas positif* toute image contenant un signal avéré anormal. A l'inverse, les images non pathologiques sont dénommées *cas négatifs*.

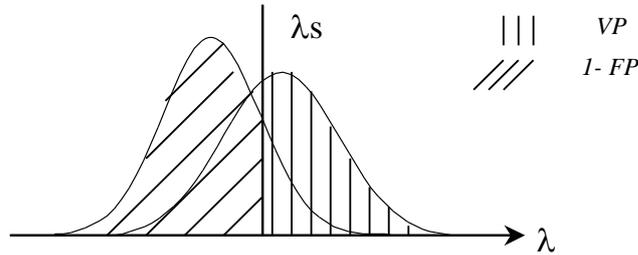


Figure B.7-2 : Distribution de probabilité et représentation du critère de décision

La ligne verticale repérant la valeur du niveau de confiance λ_s sur la Figure B.7-2 définit 4 types de diagnostics possibles.

- Un vrai positif (VP) est un cas positif correctement détecté. Il appartient au sous-ensemble
$$VP = \{ \lambda_i = f(\mathbf{x}_i) | y_i = I_1, \lambda_i \geq \lambda_s \}.$$
- Un vrai négatif (VN) correspond à un cas négatif dont la variable de décision est inférieure au seuil, c'est-à-dire qu'il appartient à
$$VN = \{ \lambda_i = f(\mathbf{x}_i) | y_i = I_0, \lambda_i < \lambda_s \}.$$

Ces deux types de diagnostic sont des réponses correctes. On définit également deux types de réponses incorrectes :

- Un faux négatif (FN) est un cas positif classé comme négatif, soit
$$FN = \{ \lambda_i = f(\mathbf{x}_i) | y_i = I_1, \lambda_i < \lambda_s \}$$
- Un faux positif (FP), correspond à un cas négatif diagnostiqué comme positif, soit
$$FP = \{ \lambda_i = f(\mathbf{x}_i) | y_i = I_0, \lambda_i \geq \lambda_s \}.$$

On définit également la sensibilité du système comme la fraction d'examens positifs correctement identifiés:

$$\text{sensibilité} = \frac{\int_{\lambda_s}^{\infty} \mathbb{P}(f(\mathbf{x}_i) = \lambda | y_i = I_1) d\lambda}{\int_{-\infty}^{\infty} \mathbb{P}(f(\mathbf{x}_i) = \lambda | y_i = I_1) d\lambda} \quad (\text{B.7-11})$$

De même la spécificité est définie par la fraction de vrais négatifs (VN) correctement labelisés:

$$\text{spécificité} = \frac{\int_{-\infty}^{\lambda_s} \mathbb{P}(f(\mathbf{x}_i) = \lambda | y_i = I_0) d\lambda}{\int_{-\infty}^{\infty} \mathbb{P}(f(\mathbf{x}_i) = \lambda | y_i = I_0) d\lambda} \quad (\text{B.7-12})$$

B.7.2.1 Analyse ROC

(i) Principe

L'analyse ROC (de l'anglais, Receiver Operating Characteristic) repose sur le modèle de la décision que nous venons d'énoncer. Les références méthodologiques majeures dans ce domaine sont [Green and Swets, 1966, Egan, 1975, Swets and Pickett, 1982, Metz, 1986].

Comme nous venons de le définir, la sensibilité et la spécificité mesurent l'efficacité du système d'imagerie pour les deux types de cas. Cependant, leur valeur est entièrement contrôlée par le seuil de décision adopté λ_s . On peut donc difficilement comparer deux systèmes sur la base d'un seul couple de mesure de (sensibilité, spécificité) par système. La différence estimée dans ce cas ne traduirait pas forcément les différences de performances intrinsèques des systèmes mais celles des critères de décision utilisés.

L'approche de l'analyse ROC consiste à caractériser le système en évaluant ce couple de paramètre pour différentes valeurs du seuil de décision. La courbe ROC représente les variations de la fraction de vrais positifs en fonction de la fraction de faux positifs. Chaque point de mesure correspond à une valeur différente du critère de décision.

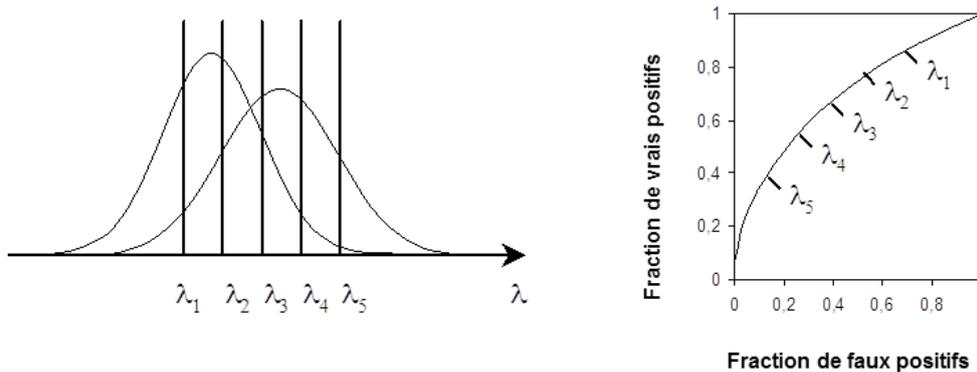


Figure B.7-3 : Principe de construction d'une courbe **ROC**

La surface délimitée par les axes de ce graphique est appelée **carré unité**. Les points situés dans le quart supérieur droit du carré unité correspondent à des valeurs faibles du seuil de décision ; cette zone de fonctionnement du système est caractérisée par une sensibilité forte et une spécificité faible. Le seuil de décision est de plus en plus strict lorsque l'on se déplace sur la courbe en se rapprochant de l'origine, comme l'indique la Figure B.7-3.

La comparaison des courbes ROC de deux systèmes d'imagerie permet d'éliminer l'influence du critère de décision et constitue ainsi une méthode objective d'évaluation des performances de détection.

Dans l'hypothèse d'un modèle de distribution binormale des cas positifs et négatifs (i.e. approximé par des gaussiennes, comme sur l'exemple de la Figure B.7-2), on montre que l'aire A_z est égale à l'index de détectabilité d , défini au paragraphe précédent,

$$A_z = d = \frac{\lambda_1 - \lambda_0}{\sqrt{\frac{\sigma_1^2 + \sigma_0^2}{2}}} \quad (\text{B.7-13})$$

(ii) Mesure expérimentale

La fonction f modélisant le processus de décision n'est pas mesurable directement pour un observateur humain. Le paradigme expérimental proposé par l'analyse ROC permet cependant d'échantillonner des couples de points (sensibilité, spécificité) de la courbe ROC correspondant à différentes valeurs du seuil de décision. Cette méthode repose sur une base de données contenant des échantillons de cas des deux classes (sain/négatif et pathologique/positif). Les cas positifs contiennent une seule anomalie par image de forme et localisation fixée. L'approche expérimentale classique [Metz, 1986] consiste à faire lire la série de données à un observateur en lui demandant de quantifier son degré de confiance sur la présence de l'anomalie dans l'image au moyen d'une échelle numérique et/ou qualitative du critère de décision, la plus répandue étant l'échelle discrète à 5 niveaux de Lickert :

- λ_1 : l'image est certainement ou presque certainement négative,
- λ_2 : l'image est probablement négative,
- λ_3 : l'image contient peut-être un signal,
- λ_4 : l'image est probablement positive,
- λ_5 : l'image est certainement ou presque certainement positive.

Cette méthode conduit ainsi à segmenter l'axe du critère de décision et à échantillonner les points de la courbe ROC correspondant ainsi aux couples $(FVP_{\lambda_i}, FFP_{\lambda_i})$ de fractions de vrais positifs et de faux positifs correspondant à un niveau de confiance supérieur ou égal à λ_i .

(iii) Ajustement des courbes ROC et analyse statistique

L'ajustement des courbes ROC à partir de mesures expérimentales a fait l'objet de nombreuses publications. L'objectif est d'estimer l'aire sous la courbe ainsi que l'intervalle de confiance associé à cette mesure tenant compte de différents facteurs dont la taille et la variabilité de l'échantillonnage de cas, la variabilité inter- et intra observateur etc. Ces intervalles de confiance permettent de comparer, par un test statistique, si la différence d'aire entre deux courbes (et donc de performance diagnostique) mesurées pour les deux protocoles que l'on souhaite comparer, est significative. Nous renvoyons à l'ouvrage de référence de Swets and Pickett [Swets and Pickett, 1982] pour le détail des fondements statistiques de ces mesures et des modèles d'estimation des différents termes de variabilité.

Il existe deux grands types de méthodes d'ajustement des courbes ROC. La première correspond aux méthodes paramétriques qui reposent sur l'hypothèse de distribution normale des deux distributions [Pesce and Metz, 2007]. Le second type comprend l'ensemble des méthodes non paramétriques basées sur l'estimation de différentes variables statistiques telles que le test de Wilcoxon [Gallas and Pesce, 2009].

Plusieurs facteurs contribuent à augmenter la variance des points de mesure de la courbe ROC. Le premier est lié au niveau de complexité de l'échantillonnage des cas. Il peut s'avérer en effet qu'une série de données contiennent plus de cas « difficiles », c'est-à-dire par exemple, des lésions situées dans des régions très inhabituelles ou très faiblement contrastées. Il est alors probable que l'on mesure des performances de détection inférieures à celles que l'on obtiendrait avec une série de cas plus faciles. Les deux autres termes de variance sont liés aux variabilités intra- et inter- individuelles. La composante intra-individuelle peut être estimée en mesurant les performances d'un observateur à qui l'on demande d'interpréter plusieurs fois la même série d'images. La composante inter-individuelle est facilement accessible à partir des performances de différents observateurs mesurées pour le même échantillonnage de cas.

La méthode la plus générale de comparaison de deux courbes ROC consiste à calculer le paramètre statistique z égal au rapport de la différence des aires A_z sur l'erreur standard de la différence des aires, soit

$$z = \frac{A_{z1} - A_{z2}}{\sqrt{\sigma^2 (A_{z1} - A_{z2})}} \quad (\text{B.7-14})$$

On peut alors évaluer l'existence d'une différence significative en testant l'hypothèse nulle selon laquelle z appartient à une distribution normale centrée. Une valeur de probabilité faible permet de démontrer l'existence d'une différence significative entre les courbes.

Swets et Pickett [Swets 1982] montrent que l'erreur standard sur la différence des aires peut être estimée par la relation :

$$\sigma^2(A_{z1} - A_{z2}) = 2 \left[\sigma_{c+intra-obs}^2 (1 - r_{c-intra-obs}) + \frac{\sigma_{inter-obs+intra-obs}^2}{l} (1 - r_{inter-obs-intra-obs}) - \sigma_{intra-obs}^2 \right] \quad (\text{B.7-15})$$

où

l est le nombre de lecteurs,

$\sigma_{c+intra-obs}^2$ est la variance que l'on mesurerait en faisant lire à un observateur plusieurs séries de cas différents,

$\sigma_{inter-obs+intra-obs}^2$ est la variance que l'on mesurerait en faisant lire à plusieurs observateurs une même série de cas.

$r_{c-intra-obs}$ et $r_{inter-obs-intra-obs}$ sont des termes de corrélation. Le terme $r_{c-intra-obs}$ est non nul si on utilise les mêmes cas dans les deux modalités. On peut alors l'estimer en partitionnant la base d'échantillons et en mesurant le coefficient de corrélation entre les aires A_z obtenues par les différents observateurs sur chaque partition de cas. Le terme $r_{inter-obs-intra-obs}$ est non nul si on utilise le même groupe de lecteurs pour lire les cas des deux modalités. On peut estimer sa valeur en calculant le coefficient de corrélation entre les aires A_z obtenues par chaque lecteur dans les deux modalités comparées.

L'expression (B.7-15) souligne que l'on peut améliorer la puissance du test statistique en augmentant le nombre de lecteurs (l), et en introduisant des corrélations entre les cas ($r_{c-intra-obs}$) et les lecteurs ($r_{inter-obs-intra-obs}$), de même qu'en augmentant le nombre de cas (diminution du terme $\sigma_{c+intra-obs}^2$).

(iv) **Logiciels d'analyse**

Plusieurs logiciels permettent de réaliser l'ajustement des courbes ROC à partir de modèles paramétriques ou non paramétriques et de comparer les performances diagnostiques de protocoles à partir d'analyse psychophysiques associant plusieurs lecteurs et plusieurs bases de cas (MRMC en anglais pour 'multiple readers multiple cases'). Nous citerons les logiciels les plus connus :

- ROC-kit développé par l'Université de Chicago :
<http://metz-roc.uchicago.edu/MetzROC/>
- OR-DBM MRMC, développé par Kevin Berbaum de l'Université de l'Iowa :
<http://perception.radiology.uiowa.edu/>
- OBUMRM, développé par Nancy A. Obuchowski, Ph.D., The Cleveland Clinic Foundation:
<http://www.lerner.ccf.org/qhs/software/obumrm.php>

B.7.2.2 Dérivées de l'analyse ROC

La méthodologie ROC, telle que nous l'avons décrite au paragraphe précédent, est adaptée à des problématiques simples et peu représentatives des conditions cliniques (une seule anomalie par image, absence de tâche de localisation). En pratique, le médecin recherche la présence d'une ou plusieurs anomalies de taille et de contraste variables et situées dans différentes régions de l'image. La tâche de perception est donc plus complexe que celle des études ROC classiques. En particulier, le risque de commettre un faux positif ou négatif est aggravé par la présence d'ambiguïtés relatives au bruit physiologique et au processus de formation de l'image. De nouvelles méthodes, dérivées de l'analyse ROC, ont été proposées dans le but de se rapprocher des conditions réalistes d'interprétation des images. Une revue récente de cette thématique est présentée dans [Metz, 2006, He and Frey, 2009]. Nous présentons ci-dessous une synthèse des méthodes disponibles et renvoyons à la Figure B.7-4 qui présente la forme des différents types de courbes.

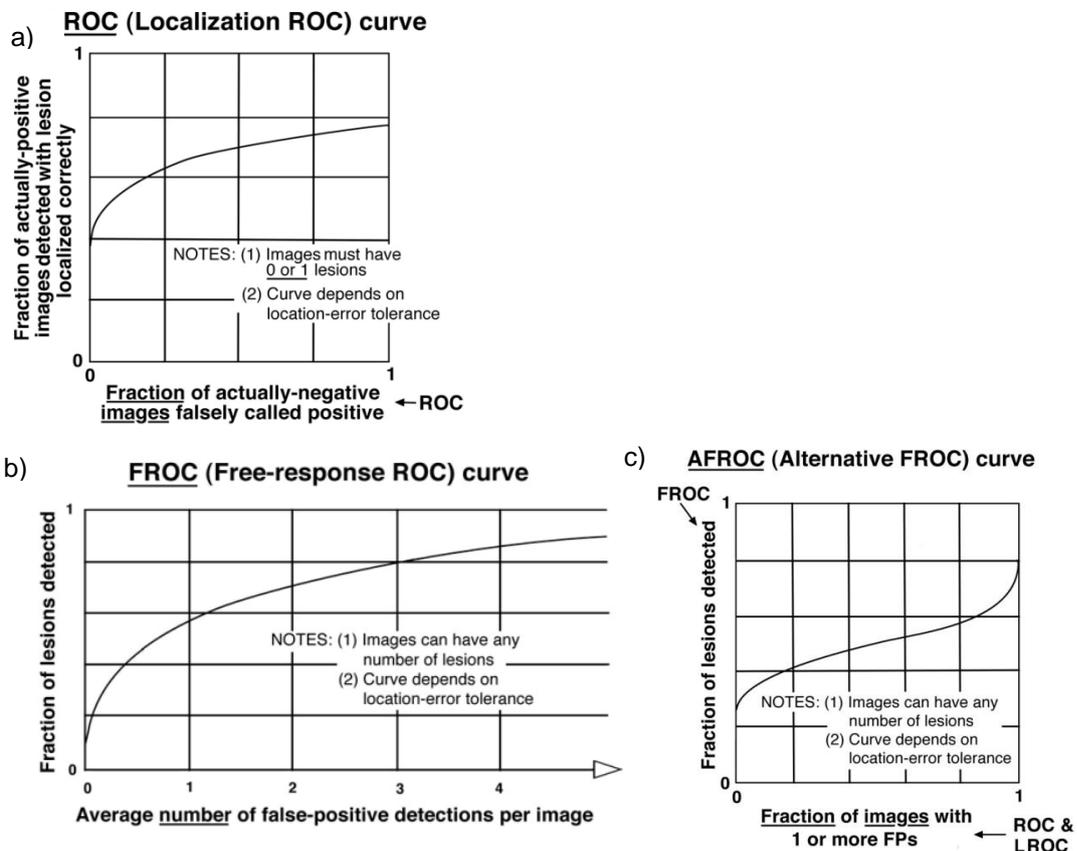


Figure B.7-4. Illustration des principales analyses dérivées de l'analyse ROC pour des tâches de détection et localisation. Extrait de [Metz, 2006]

Les courbes LROC (Localization Response Operating Characteristics) (a) mesurent les performances du système diagnostique à *détecter* et *localiser* les anomalies réelles de l'image. Contrairement à la méthodologie ROC, chaque image positive contient une seule anomalie localisée aléatoirement dans l'image. L'observateur doit indiquer la position la plus suspecte, si il pense que l'image est positive ainsi qu'un niveau de confiance. Celle-ci est comptée comme un événement *VP* si la distance entre la position suspectée et la position réelle ne dépasse pas une limite de tolérance fixée préalablement. La courbe LROC représente les variations des vrais positifs correctement localisés en fonction des faux positifs mesurés sur une série d'images négatives, i.e. ne contenant pas d'anomalies, pour différentes valeurs du critère de décision λ_s .

Cette méthodologie a un connu un succès limité avant l'introduction par Swensson, à la fin des années 90, du formalisme statistique permettant d'ajuster les paramètres de la courbe LROC et d'en extraire des métriques telle que l'aire sous la courbe A_{LROC} [Swensson, 1996]. Elle reste cependant éloignée du paradigme clinique (seule l'anomalie la plus suspecte est reportée).

L'analyse FROC (Free Response Operating Characteristics) ((b)) propose un protocole de mesure plus proche de la réalité clinique puisque les cas positifs contiennent plusieurs anomalies localisées aléatoirement dans l'image. L'observateur n'est pas informé du nombre de cibles à détecter par image. Il doit localiser et évaluer toutes les anomalies présentes dans l'image sans limite de nombre et en indiquant un niveau de confiance pour chacune. La courbe FROC reporte alors la fraction de vrais positifs (identifiés, comme pour l'analyse LROC, par rapport à une distance entre la position reportée et la position exacte de l'anomalie) en fonction du nombre moyen de fausses détections par image, pour différentes valeurs du critère de décision λ_s .

Les courbes FROC ne peuvent pas être ajustées par une méthode statistique simple car l'axe des abscisses n'est pas borné. On préfère alors se rapporter à une variante de ces courbes, appelées **courbes AFROC** pour 'Alternate Free Response Operating Characteristics' ((c)) qui représentent le taux de vrais positifs pour un seuil λ_s en fonction de la fraction d'images analysées dans par lesquelles l'observateur a reporté au moins un faux positif avec un niveau de confiance supérieur ou égal à λ_s . Cette méthode permet de retrouver dans un espace borné. Les travaux de Swensson [Swensson, 1996] et Chakraborty [Chakraborty and Winter, 1990] ont permis de définir un modèle paramétrique général permettant d'extraire des mesures quantitatives telles que l'aire sous la courbe AFROC, A_{AFROC} .

Différents logiciels permettent de modéliser les études LROC et AFROC, dont le plus connu est le logiciel JAFROC développé par Dev Chakraborty de l'Université de Pittsburgh : <http://www.devchakraborty.co>

B.7.3 Les observateurs numériques

Un observateur algorithmique est un test statistique de discrimination basé sur le modèle de la décision statistique.

Le principe consiste à calculer un paramètre statistique appelé variable de décision $\lambda_i \in \mathbb{R}$ à partir de l'observation \mathbf{x}_i et d'une fonction f définie dans l'équation (B.7-7). On peut alors, comme pour les observateurs humains, mesurer des performances de détection à partir de métriques telles que l'aire A_z sous la courbe ROC ou le SNR (cf B.7.2.1). Dans la suite nous utilisons la notation $\lambda(\mathbf{x}_i) = \lambda_i = f(\mathbf{x}_i)$ ou $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ est la représentation lexicographique d'une image.

Il existe deux grandes classes d'observateurs algorithmiques, les observateurs **idéaux** ou **bayésiens** et les observateurs **linéaires**.

D'une manière générale, les observateurs bayésiens servent à estimer la limite supérieure des performances intrinsèques du système de détection. Les observateurs linéaires, en revanche, se comportent de manière sous-optimale, sauf sous certaines conditions de bruit pour lesquelles ils sont équivalents aux observateurs idéaux. On les utilise préférentiellement pour essayer de reproduire les performances humaines en rajoutant des termes de bruit modélisant les mécanismes de perception visuelle humaine. On utilise alors le terme d'**observateurs anthropomorphiques**.

La majorité des observateurs que nous décrivons par la suite ont été développés pour répondre à la tâche de détection classique de l'analyse ROC, c'est-à-dire la détection d'un signal pathologique dans une image 2D sans tâche de localisation. Des travaux plus récents, que nous commentons dans la section B.3.5.3(ii), se sont intéressés à la conception de modèles plus complexes permettant d'inclure la tâche de localisation d'une ou plusieurs anomalies ainsi que des modèles adaptées à l'analyse 3D.

B.7.3.1 L'observateur idéal ou bayésien

Dans le cas d'une tâche de détection binaire (voir notations de la section B.7.1.1) entre deux classes l_1 et l_0 , on montre que le coût d'appartenance à la classe l_1 , la variable de décision de l'observateur idéal $\lambda_{id}(\mathbf{x}_i)$ est estimée par le logarithme du rapport des vraisemblances [Helstrom, 1995]

$$\lambda_{id}(\mathbf{x}_i) = \ln \left(\frac{\mathbb{P}(X = \mathbf{x}_i | Y = l_1)}{\mathbb{P}(X = \mathbf{x}_i | Y = l_0)} \right) \quad (\text{B.7-16})$$

Où $\mathbb{P}(X = \mathbf{x}_i | Y = l_k)$ est définie comme la probabilité d'observer l'image \mathbf{x}_i sous l'hypothèse d'appartenance à la classe l_k .

Cette expression est justifiée dans la section B.7.4.2(ii). Ce test statistique minimise le risque de Bayes et permet d'évaluer les performances intrinsèques du système de mesure grâce à une utilisation optimale de l'information restituée par l'image. Il suppose cependant de connaître les distributions de probabilité statistique de l'observation \mathbf{x}_i dans les deux hypothèses. En pratique, cette information est difficilement accessible, sauf dans le cas de tâches simples et peu représentatives des problématiques cliniques.

En supposant une distribution gaussienne de la variable de décision dans les deux conditions, il est possible de dériver une expression linéaire de $\lambda_{id}(\mathbf{x}_i)$ équivalente à celle du modèle de Hotelling présenté dans le paragraphe suivant. Cette expression est dérivée dans le paragraphe B.7.3.2(iv). Plus récemment, Barrett et al ont proposé une méthode d'approximation du rapport de vraisemblance à partir d'un développement en série de Taylor des fonctions de probabilité. Cette approche permet d'étendre le champ d'application de l'observateur bayésien à des tâches plus complexes, telles que la détection d'un signal positionné aléatoirement dans l'image [Barrett and Abbey, 1997]. D'autres

approches basées sur des méthodes de Monte Carlo par chaînes de Markov ont également été proposées [Kupinski et al., 2003].

B.7.3.2 Les observateurs linéaires

(i) Définitions

Pour les observateurs linéaires, la variable λ du test statistique est une fonction linéaire de l'image $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$, soit

$$\lambda(\mathbf{x}_i) = \mathbf{u}' \mathbf{x}_i \quad (\text{B.7-17})$$

Où $\mathbf{u} \in \mathcal{X} = \mathbb{R}^p$ est un filtre utilisant les informations statistiques d'ordre 1 et 2 de l'image.

Ces modèles algorithmiques n'imposent pas de connaître les distributions de probabilités, contrairement à l'approche bayésienne. Ils peuvent être utilisés comme alternative (dégradée) à l'observateur idéal, mais sont principalement développés pour reproduire les performances humaines. Cette dernière application suppose de comprendre le mécanisme de la perception visuelle, et surtout d'analyser comment l'observateur humain est influencé par les corrélations du bruit, et s'il adapte une stratégie de perception en fonction de la nature du bruit. La méthode employée pour la mise au point de ces observateurs, dits anthropomorphiques, consiste à générer des images dont les propriétés de bruit sont parfaitement maîtrisées, contrairement aux images acquises sur des patients. On utilise alors les performances d'un observateur humain comme référence et l'on cherche à adapter les paramètres du modèle algorithmique de manière à ajuster les performances de détection de l'algorithme sur celles de l'homme.

On peut montrer que l'expression du SNR défini en (B.7-10) s'écrit dans le cas d'un modèle linéaire (LO pour observateur linéaire) [Barrett, 1990]

$$SNR_{LO}^2 = \frac{\text{tr}(\mathbf{U}\mathbf{S}_b)}{\text{tr}(\mathbf{U}\mathbf{S}_w)} \quad (\text{B.7-18})$$

avec \mathbf{S}_b et \mathbf{S}_w les matrices de covariance inter- et intra-classes définies en B.7.1.2 et $\mathbf{U} \equiv \mathbf{u}\mathbf{u}' \in \mathcal{M}_p(\mathbb{R})$

Démonstration :

D'après l'expression (B.7-10) du SNR dans le cas d'un problème de classification binaire, $\bar{\lambda}_i \in \mathbb{R}$ et $\sigma_i^2 \in \mathbb{R}$ sont la valeur moyenne et la variance respectivement de la variable de décision calculée sur les échantillons de la classe l_i . σ_i^2 est donc la projection de la covariance intra-classe \mathbf{S}_w^i pour la classe l_i sur le plan de vecteur directeur \mathbf{u} , soit

$$\sigma_i^2 = \mathbf{u}' \mathbf{S}_w^i \mathbf{u} \quad (\text{B.7-19})$$

De même, $\bar{\lambda}_i$ est la projection du centroïde des échantillons de la classe l_i sur le plan de vecteur directeur \mathbf{u} , soit,

$$\bar{\lambda}_i = \mathbf{u}' \bar{\mathbf{x}}_i \quad (\text{B.7-20})$$

L'expression du SNR s'écrit alors

$$\begin{aligned} SNR_{LO}^2 &= \frac{n_1 n_0}{n^2} \frac{(\mathbf{u}' \bar{\mathbf{x}}_1 - \mathbf{u}' \bar{\mathbf{x}}_0)^2}{\frac{1}{n} (n_1 \mathbf{u}' \mathbf{S}_w^1 \mathbf{u} + n_0 \mathbf{u}' \mathbf{S}_w^0 \mathbf{u})} \\ &= \frac{n_1 n_0}{n^2} \frac{(\mathbf{u}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0))^2}{\mathbf{u}' \mathbf{S}_w \mathbf{u}} \end{aligned} \quad (\text{B.7-21})$$

En utilisant les propriétés définies en B.7.1.2 pour les matrices de rang 1, on peut écrire

$$(\mathbf{u}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0))^2 = \text{tr}((\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)' \mathbf{U})$$

et

$$\mathbf{u}' \mathbf{S}_w \mathbf{u} = \text{tr}(\mathbf{U} \mathbf{S}_w)$$

Soit en reprenant l'expression de la matrice inter-classe \mathbf{S}_b définie en (B.7-5) dans le cas où $L=2$, on obtient

$$SNR_{LO}^2 = \frac{\text{tr}(\mathbf{U} \mathbf{S}_b)}{\text{tr}(\mathbf{U} \mathbf{S}_w)} \quad (\text{B.7-22})$$

Dans ce paragraphe, nous donnons une description des principaux observateurs linéaires.

(ii) Filtre adaptatif non blanchisseur (NPWMF)

La variable de décision de l'observateur adaptatif non blanchisseur (de l'anglais, 'nonprewhitening matched filter' (NPWMF) est le produit de corrélation de l'image $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ par un masque proportionnel au profil du signal $\mathbf{s} \in \mathcal{X} = \mathbb{R}^p$ que l'on souhaite détecter:

$$\lambda_{NPWMF}(\mathbf{x}_i) = \mathbf{s}' \mathbf{x}_i \quad (\text{B.7-23})$$

On peut montrer facilement, d'après l'expression générale du SNR pour les modèles linéaires (B.7-18) que

$$SNR_{NPWMF}^2 = \frac{\text{tr}(\mathbf{S} \mathbf{S}_b)}{\text{tr}(\mathbf{S} \mathbf{S}_w)} \quad (\text{B.7-24})$$

Avec $\mathbf{S} \equiv \mathbf{s} \mathbf{s}' \in \mathcal{M}_p(\mathbb{R})$

Ce filtre est basé sur la détection du pic d'intensité du signal et n'utilise aucune information sur les propriétés de bruit de l'image. Il est optimal, c'est-à-dire équivalent à un observateur idéal, dans le cas d'un bruit blanc gaussien et d'un signal de position fixe et connue [Burgess et al., 1981]. En revanche, cette stratégie est sous-optimale dans le cas d'un bruit corrélé ou en présence de bruit anatomique [Myers et al., 1985].

(iii) Filtre adaptatif blanchisseur (PWMF)

La variable de décision de l'observateur adaptatif blanchisseur (de l'anglais, 'prewhitening matched filter' (PWMF)) est le produit de corrélation de l'image $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$, préalablement blanchie par la matrice \mathbf{S}_w^0 de covariance de la classe l_0 , par un masque proportionnel au profil du signal $\mathbf{s} \in \mathcal{X} = \mathbb{R}^p$

$$\lambda_{PWMF} = \mathbf{s}^t (\mathbf{S}_w^0)^{-1} \mathbf{x}_i \quad (\text{B.7-25})$$

D'après l'équation (B.7-18), on montre que le SNR de filtre adaptatif blanchisseur s'exprime comme

$$SNR_{PWMF}^2 = \text{tr} \left((\mathbf{S}_w^0)^{-1} \mathbf{S}_b \right) \quad (\text{B.7-26})$$

Ce filtre est optimal dans le cas d'un bruit gaussien corrélé et d'un signal de position fixe et connue. Les résultats publiés dans la littérature montrent que les performances humaines sont inférieures à celles du filtre adaptatif blanchisseur en présence de bruit corrélé.

(iv) Modèle de Hotelling

Le test de Hotelling utilise un masque capable de s'adapter aux corrélations spécifiques du bruit de l'image. On peut montrer qu'il optimise ainsi les performances de détection des observateurs linéaires [Hotelling, 1931]. La valeur de la variable de décision est donnée par :

$$\lambda_{Hot}(\mathbf{x}_i) = \mathbf{u}^t \mathbf{x}_i \quad (\text{B.7-27})$$

avec

$$\mathbf{u} = (\mathbf{S}_w)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0) \quad (\text{B.7-28})$$

où \mathbf{S}_w est la matrice de covariance intra-classe définie au paragraphe B.7.1.2

Soit, d'après l'expression (B.7-28) de \mathbf{u} , le dénominateur de (B.7-21) s'écrit

$$\mathbf{u}^t \mathbf{S}_w \mathbf{u} = \mathbf{u}^t \mathbf{S}_w (\mathbf{S}_w)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0) = \mathbf{u}^t (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)$$

D'où

$$\begin{aligned} SNR_{Hot}^2 &= \frac{n_1 n_0}{n^2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)^t \mathbf{u} \\ &= \frac{n_1 n_0}{n^2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0)^t (\mathbf{S}_w)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0) \end{aligned} \quad (\text{B.7-29})$$

Le test de Hotelling prend en compte la matrice de covariance totale (incluant S_w^1 calculée sur les images de la classe positive) et améliore ainsi les performances de classification des observateurs NPWMF et PWMF dans le cas de signaux localisés aléatoirement.

Lien avec l'observateur idéal de Bayes

Supposons que la distribution de probabilité $\mathbb{P}(X = \mathbf{x}_i | Y = l_k)$ suive une loi Gaussienne de centroïde $\bar{\mathbf{x}}_k$ et de covariance \mathbf{S}_w^k , soit

$$\mathbb{P}(X = \mathbf{x}_i | Y = l_k) = \frac{1}{(2\pi)^{p/2} |\det(\mathbf{S}_w^k)|} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_k)^t (\mathbf{S}_w^k)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_k)\right)$$

D'après l'équation (B.7-16), l'expression de la variable de décision de l'observateur idéal de Bayes devient, dans le cas d'un problème de classification binaire entre les classes négative l_0 et positive l_1 ,

$$\lambda_{id}(\mathbf{x}_i) = \log \left(\frac{\frac{1}{(2\pi)^{p/2} |\det(\mathbf{S}_w^1)|} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_1)^t (\mathbf{S}_w^1)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_1)\right)}{\frac{1}{(2\pi)^{p/2} |\det(\mathbf{S}_w^0)|} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_0)^t (\mathbf{S}_w^0)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_0)\right)} \right)$$

Après simplification du terme constant, on obtient

$$\begin{aligned} \lambda_{id}(\mathbf{x}_i) &= -\frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_1)^t (\mathbf{S}_w^1)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_1) + \frac{1}{2}(\mathbf{x}_i - \bar{\mathbf{x}}_0)^t (\mathbf{S}_w^0)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_0) \\ &= \frac{1}{2} \mathbf{x}_i^t (\mathbf{S}_w^0 - \mathbf{S}_w^1)^{-1} \mathbf{x}_i + \left(\bar{\mathbf{x}}_0 (\mathbf{S}_w^0)^{-1} - \bar{\mathbf{x}}_1 (\mathbf{S}_w^1)^{-1} \right)^t \mathbf{x}_i + \frac{1}{2} \left(\bar{\mathbf{x}}_0^t (\mathbf{S}_w^0)^{-1} \bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1^t (\mathbf{S}_w^1)^{-1} \bar{\mathbf{x}}_1 \right) \end{aligned}$$

Dans le cas de deux distributions de covariance égale, soit $\mathbf{S}_w^1 = \mathbf{S}_w^0$, on obtient

$$\lambda_{id}(\mathbf{x}_i) = (\mathbf{S}_w^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0))^t \mathbf{x}_i + \frac{1}{2} \left(\bar{\mathbf{x}}_0^t (\mathbf{S}_w^0)^{-1} \bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1^t (\mathbf{S}_w^1)^{-1} \bar{\mathbf{x}}_1 \right)$$

Soit, d'après l'équation (B.7-28)

$$\lambda_{id}(\mathbf{x}_i) = \lambda_{Hor}(\mathbf{x}_i) + k \tag{ B.7-30}$$

$$\text{Avec } k = \frac{1}{2} \left(\bar{\mathbf{x}}_0^t (\mathbf{S}_w^0)^{-1} \bar{\mathbf{x}}_0 - \bar{\mathbf{x}}_1^t (\mathbf{S}_w^1)^{-1} \bar{\mathbf{x}}_1 \right)$$

(v) Observateur linéaire canalisé

Dans ce paragraphe, nous utilisons une représentation matricielle de l'image définie par $\mathbf{g}_i \in \mathbb{R}^{p_x \times p_y}$ ou p_x et p_y sont les dimensions de l'image selon les deux directions. On note que $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ est la représentation lexicographique de \mathbf{g}_i .

L'état de l'art sur les modèles anthropomorphiques montre que les performances de détection humaines sont supérieures à celles du filtre NPWMF en présence de bruit corrélé mais inférieures à celles du filtre PWMF. Une alternative proposée consiste ainsi à effectuer un filtrage partiel de l'information avant d'appliquer l'un des tests statistiques décrits précédemment. Cette approche semble d'ailleurs cohérente avec le modèle de la perception visuelle humaine selon lequel l'information visuelle est traitée par un ensemble de filtres avant d'être interprétée [Sachs et al., 1971].

Les observateurs développés sur ce modèle proposent de réaliser une décomposition spectrale de l'image sur un ensemble de c filtres passe-bande choisis de manière à couvrir l'ensemble du spectre fréquentiel de l'image.

Chaque image \mathbf{x}_i (ou \mathbf{g}_i) est alors représentée par un vecteur $\mathbf{V}_i \in \mathcal{V} = \mathbb{R}^c$, $\mathbf{V}_i = (v_i(1), \dots, v_i(c))^t$ de dimension égale au nombre c de filtres et de composante $v_i(k)$ définie par

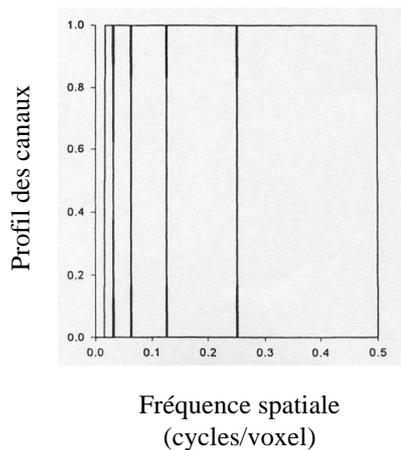
$$v_i(k) = \iint_{R_k} G_i(v_x, v_y) dv_x dv_y \quad (\text{B.7-31})$$

où $G(v_x, v_y)$ est la transformée de Fourier de l'image \mathbf{g}_i et R_k est le support fréquentiel du filtre.

La variable de décision des différents modèles linéaires est alors calculée sur le vecteur \mathbf{V}_i .

Différents modèles de décomposition spectrale ont été envisagés et décrits dans la littérature. En particulier, le modèle des filtres SQR (de l'anglais 'square channels') est un ensemble de filtres passe-bande non superposables et symétriques par rotation. Ils correspondent à une subdivision annulaire du domaine de Fourier. King et al [King et al., 1997] ont également proposé une sous-décomposition en secteur angulaire, permettant de prendre en compte l'asymétrie potentielle des anomalies. Un modèle de filtres utilisant des fonctions de Gabor, définies comme la différence de deux gaussiennes (DOG pour 'difference of Gaussian') a également été proposé ([Yao and Barrett, 1992]). D'autres équipes essaient également de se rapprocher du modèle de perception visuelle de Sachs [Eckstein and Whiting, 1995].

(a) SQR



(b) DOG

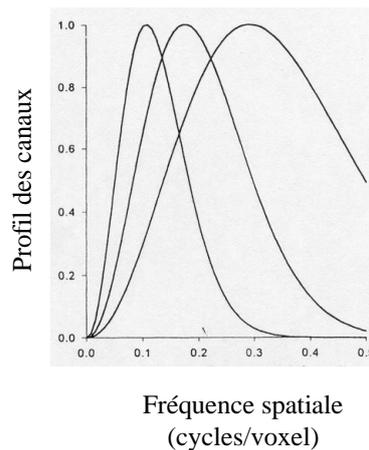


Figure B.7-5 : Filtrés SQR et DOG (extrait de [King et al., 1997])

B.7.4 Apprentissage supervisé

B.7.4.1 Définitions

Le terme d'apprentissage supervisé recouvre l'ensemble des méthodes qui permettent de prédire l'étiquette d'un objet, i. e. son appartenance à une classe prédéfinie, à partir d'une base d'apprentissage pré-étiquetée. Cette étiquette peut être continue, on parle alors de *régression*, ou discrète; on parle alors de *classement* ou *classification*.

Ces méthodes consistent à estimer un modèle ou une fonction de décision permettant de séparer les objets dans différentes classes.

Dans la suite, nous utilisons les conventions définies dans la section B.7.1.1. Ici, $\mathbf{x}_i \in \mathcal{X} = \mathbb{R}^p$ est soit la représentation lexicographique d'une image, soit un ensemble de caractéristiques de cette image.

Le problème de classification peut être séparé en deux étapes :

- 1) L'inférence consiste à apprendre un modèle ou *fonction de décision* $f : \mathcal{X} \rightarrow \mathbb{R}$ à partir de données d'apprentissage
- 2) La décision réalisée à partir d'une règle n individus dans un espace de dimension p , $D(\mathbf{x}_i)$ appliquée à la fonction de décision f . Cette étape est souvent assez simple et consiste souvent à fixer un seuil sur f ou prendre la fonction $D(\mathbf{x}_i) = \text{signe}(f(\mathbf{x}_i))$.

La majorité des méthodes proposées pour résoudre ce problème reposent sur la notion de *minimisation du risque de mauvaise classification*. Ces méthodes sont organisées selon deux types, les modèles *génératifs* et les modèles *discriminants*.

Dans *les modèles génératifs*, apprennent les densités de probabilité conditionnelles $\mathbb{P}(X = \mathbf{x}_i | Y = l_k)$ de chaque classe à partir des données d'apprentissage et d'un modèle *a priori* sur la forme de ces distributions. Le théorème de Bayes permet ensuite d'estimer les probabilités postérieures $\mathbb{P}(Y = l_k | X = \mathbf{x}_i)$ qui constitue la fonction de décision

$$f : \mathcal{X} \rightarrow [0,1] \quad (\text{B.7-32})$$

$$\mathbf{x}_i \mapsto \mathbb{P}(Y = l_1 | X = \mathbf{x}_i)$$

L'exemple le plus connu est le classifieur bayésien naïf basé sur l'hypothèse de distribution gaussienne multidimensionnelle des probabilités conditionnelles.

Dans *les modèles discriminants*, on ne s'intéresse pas directement à l'estimation des probabilités conditionnelles. On cherche à construire un modèle prédictif à l'aide d'une méthode d'approximation. Il existe deux types de méthodes :

Le premier type est basé sur une fonction de décision f similaire à celle du modèle génératif (B.7-32). Cependant, contrairement au modèle génératif, on ne cherche pas à modéliser les probabilités conditionnelles mais directement les probabilités postérieures. On ajuste les paramètres du modèle sur les données d'apprentissage (généralement par minimisation du logarithme de la fonction de vraisemblance). L'exemple le plus courant est la régression logistique.

Le second type de méthode consiste à estimer directement la fonction de décision f sans passer par l'estimation des densités de probabilités mais en estimant les paramètres de l'hyperplan séparateur à partir des exemples d'apprentissage. La fonction f renvoie alors la distance du point à l'hyperplan. L'approche la plus classique repose sur une modélisation linéaire de l'hyperplan séparateur dans l'espace des caractéristiques \mathbf{x}_i .

$$\begin{aligned}
 f : \mathcal{X} &\rightarrow \mathbb{R} \\
 \mathbf{x}_i &\mapsto \mathbf{w}'\mathbf{x}_i + b
 \end{aligned}
 \tag{B.7-33}$$

L'apprentissage des paramètres de l'hyperplan \mathbf{w} et b peut s'exprimer sous forme d'un problème d'optimisation.

La deuxième classe de méthode de classification supervisée repose sur une approche géométrique, le formalisme le plus connu étant l'analyse discriminante linéaire (ADL).

Dans la suite, nous donnons le principe du classifieur idéal de Bayes et décrivons un modèle de chaque famille de classifieurs, le séparateur à vaste marge (SVM) et l'ADL.

B.7.4.2 Approche statistique basée sur la minimisation du risque

(i) Fondements statistiques

La théorie statistique de l'apprentissage repose sur la notion de *risque* R aussi appelé *erreur de prédiction*. Les paramètres de la fonction de décision f pour un problème donné sont ceux qui minimisent l'espérance de l'erreur de prédiction R entre l'ensemble des étiquettes prédites $f(\mathbf{x}_i)$ et des étiquettes réelles y_i , soit

$$R(f) = \mathbb{E} \left[L(\mathbf{Y}, f(\mathbf{X})) \right] = \int_{\mathcal{X} \times \mathcal{Y}} L(y_i, f(\mathbf{x}_i)) \mathbb{P}(\mathbf{x}_i, y_i) d\mathbf{x}_i dy_i
 \tag{B.7-34}$$

Où $L(\cdot, \cdot) : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ est une fonction de perte qui quantifie le coût de l'erreur de prédiction commise. La minimisation de l'expression (B.7-34) permet d'obtenir le classifieur de Bayes. En pratique, nous avons souligné la difficulté d'accéder à la loi de probabilité conjointe $\mathbb{P}(\cdot, \cdot)$. On peut cependant estimer *le risque empirique* R_{emp} à partir des n exemples d'apprentissage, par la relation

$$R_{emp}(f) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(\mathbf{x}_i))
 \tag{B.7-35}$$

Le problème d'optimisation consiste alors à déterminer la fonction de décision f qui minimise R_{emp} , soit

$$\min_{f \in H} \frac{1}{n} \sum_{i=1}^n L(y_i, f(\mathbf{x}_i))$$

Cependant, la taille de la base d'apprentissage étant limitée, il est toujours possible de trouver une fonction f qui minimise le risque sur les exemples d'apprentissage mais ne présente pas de bonne performance avec des exemples tests indépendants. Vapnik et Chervonenkis ont en effet montré que

$$R(f) - R_{emp}(f) \leq \sqrt{\frac{p(\ln(2n/p) + 1) - \ln(\eta - 4)}{n}}
 \tag{B.7-36}$$

Avec la probabilité $1 - \eta$

Où n et p sont respectivement la taille de la base de données d'apprentissage et la dimension du problème.

Cette relation souligne que la minimisation du risque empirique ne garantit pas que l'on obtienne un minimum du risque R . On peut alors contraindre la solution f en rajoutant au problème de minimisation un terme de régularisation $\Omega(\cdot)$ mesurant la complexité de la fonction f , soit

$$\min_{f \in H} \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \lambda \Omega(f) \quad (\text{B.7-37})$$

On parle alors d'un problème de minimisation du **risque structurel**.

Dans la suite, nous considérons un problème de classification binaire de n individus $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathcal{X} \times Y$ avec $\mathcal{X} = \mathbb{R}^p$ et $\mathcal{Y} = \{l_{-1}, l_1\}$.

(ii) Le classifieur de Bayes

Dans le cas d'un problème de classification binaire, et en utilisant les conventions de notation de la section B.7.1, l'expression du risque pour chaque classe est donné par

$$R(l_1, \mathbf{x}_i) = L(l_1, l_1) \mathbb{P}(Y = l_1 | X = \mathbf{x}_i) + L(l_1, l_{-1}) \mathbb{P}(Y = l_{-1} | X = \mathbf{x}_i)$$

$$R(l_{-1}, \mathbf{x}_i) = L(l_{-1}, l_1) \mathbb{P}(Y = l_1 | X = \mathbf{x}_i) + L(l_{-1}, l_{-1}) \mathbb{P}(Y = l_{-1} | X = \mathbf{x}_i)$$

avec

$\mathbb{P}(Y = l_j | X = \mathbf{x}_i)$, la probabilité *a posteriori* de classer l'échantillon \mathbf{x}_i dans la classe l_j

$L(l_k, l_j)$, le coût correspondant au classement d'une observation de la classe l_j dans la classe l_k .

L'étiquette de \mathbf{x}_i sera l_1 si et seulement si

$$R(l_1, \mathbf{x}_i) < R(l_{-1}, \mathbf{x}_i), \text{ soit}$$

$$(L(l_1, l_1) - L(l_{-1}, l_1)) \mathbb{P}(Y = l_1 | X = \mathbf{x}_i) < (L(l_1, l_{-1}) - L(l_{-1}, l_{-1})) \mathbb{P}(Y = l_{-1} | X = \mathbf{x}_i)$$

En supposant que $L(l_1, l_1) - L(l_{-1}, l_1) < 0$, on obtient

$$\frac{\mathbb{P}(Y = l_1 | X = \mathbf{x}_i)}{\mathbb{P}(Y = l_{-1} | X = \mathbf{x}_i)} > \frac{L(l_1, l_{-1}) - L(l_{-1}, l_{-1})}{L(l_1, l_1) - L(l_{-1}, l_1)}$$

Soit, d'après le théorème de Bayes (B.7-54),

$$\mathbb{P}(Y = l_1 | X = \mathbf{x}_i) > \frac{L(l_1, l_{-1}) - L(l_{-1}, l_{-1})}{L(l_1, l_1) - L(l_{-1}, l_1)} \mathbb{P}(Y = l_{-1} | X = \mathbf{x}_i)$$

$$\Leftrightarrow \mathbb{P}(X = \mathbf{x}_i | Y = l_1) \mathbb{P}(Y = l_1) > \frac{L(l_1, l_{-1}) - L(l_{-1}, l_{-1})}{L(l_1, l_1) - L(l_{-1}, l_1)} \mathbb{P}(X = \mathbf{x}_i | Y = l_{-1}) \mathbb{P}(Y = l_{-1})$$

$$\Leftrightarrow \frac{\mathbb{P}(X = \mathbf{x}_i | Y = l_1)}{\mathbb{P}(X = \mathbf{x}_i | Y = l_{-1})} > \frac{L(l_1, l_{-1}) - L(l_{-1}, l_{-1})}{L(l_1, l_1) - L(l_{-1}, l_1)} \frac{\mathbb{P}(Y = l_{-1})}{\mathbb{P}(Y = l_1)}$$

La fonction de décision du modèle de Bayes s'écrit ainsi comme le logarithme du rapport des vraisemblances,

$$f(\mathbf{x}_i) = \ln \left(\frac{\mathbb{P}(X = \mathbf{x}_i | Y = 1)}{\mathbb{P}(X = \mathbf{x}_i | Y = -1)} \right) \quad (\text{B.7-38})$$

Et la règle de décision correspondante,

$$D(\mathbf{x}_i) = \begin{cases} l_1 & \text{si } f(\mathbf{x}_i) \geq k \\ l_{-1} & \text{sinon} \end{cases}$$

avec $k = \frac{L(l_1, l_{-1}) - L(l_{-1}, l_{-1})}{L(l_1, l_1) - L(l_{-1}, l_1)} \frac{\mathbb{P}(Y = -1)}{\mathbb{P}(Y = 1)}$

(iii) Formalisme du SVM

Le modèle des séparateurs à vaste marge, SVM, a été proposé à la fin des années 90 à partir de la théorie statistique de l'apprentissage de Vapnik-Chervonenkis [Vapnik, 1995]. Ils a rapidement été adopté pour sa capacité à travailler avec des données de grande dimension, le faible nombre d'hyper paramètres et le fait qu'ils soient bien fondés théoriquement. Le principe du SVM consiste à trouver l'hyperplan qui optimise la séparation entre les classes d'objets considérées, ie qui maximise la marge (la distance) entre les échantillons et l'hyperplan séparateur (cf Figure B.7-6)

(a) Formulation dans le cas de données linéairement séparables

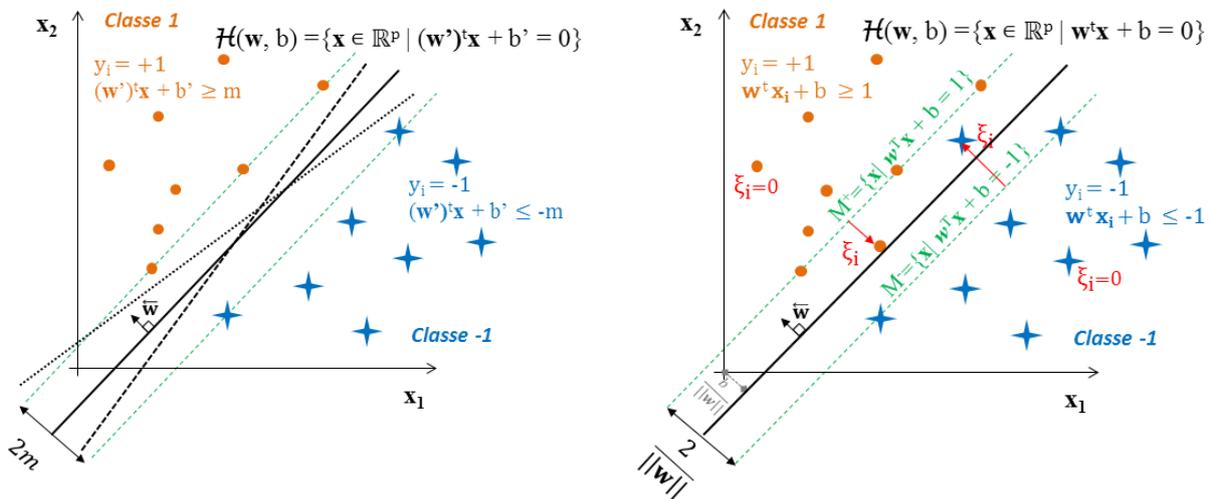


Figure B.7-6. Illustration du principe de classification par SVM.

Comme discuté dans la section précédente, le SVM est un modèle discriminant basé sur l'hypothèse d'une fonction de décision linéaire d'équation $f(\mathbf{x}_i) = \mathbf{w}'\mathbf{x}_i + b$ avec \mathbf{w} un vecteur normal à l'hyperplan

et b un réel tel $\frac{b}{\|\mathbf{w}\|}$ soit la distance de l'origine à l'hyperplan.

On rappelle que la distance d'un point \mathbf{x} à l'hyperplan est donnée par $dist(\mathbf{x}, \Delta) = \frac{\mathbf{w}'\mathbf{x} + b}{\|\mathbf{w}\|}$

Dans le cas de données linéairement séparables, la minimisation du risque empirique de mauvaise classification, que l'on peut formuler comme $y_i(\mathbf{w}'\mathbf{x}_i + b) > 0, \forall i=(1, \dots, n)$, n'est pas suffisante pour conduire à la détermination d'une fonction de décision unique et qui se généralise bien, comme indiqué sur la Figure B.7-6 (gauche).

La proposition de Vapnik consiste à contraindre ce problème en recherchant l'hyperplan qui maximise la marge m définie comme la plus petite distance d'un point de la base d'apprentissage à la frontière de décision.

On définit l'hyperplan intermédiaire d'équation $f(\mathbf{x}_i) = (\mathbf{w}')^t \mathbf{x}_i + b'$. Le formalisme des SVM consiste ainsi à trouver les paramètres \mathbf{w}' et b' solutions du problème d'optimisation suivant

$$\begin{cases} \max_{\mathbf{w}', b'} m \\ \text{s.c. } y_i \left((\mathbf{w}')^t \mathbf{x}_i + b' \right) > m, \quad i=(1, 2, \dots, n) \end{cases}$$

Ce problème est mal posé. En effet, si (\mathbf{w}', b') est une solution, alors $(k \mathbf{w}', kb')$, $\forall k > 0$, l'est aussi. On peut alors effectuer le changement de variable suivant

$$\mathbf{w} = \frac{\mathbf{w}'}{m \|\mathbf{w}'\|} \quad \text{soit } \|\mathbf{w}\| = \frac{1}{m}$$

$$b = \frac{b'}{m \|\mathbf{w}'\|}$$

et se ramener à la forme canonique suivante

$$\begin{cases} \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.c. } y_i \left(\mathbf{w}'\mathbf{x}_i + b \right) > 1, \quad i=(1, 2, \dots, n) \end{cases}$$

En pratique, il est souvent difficile de trouver une séparatrice linéaire parfaite. On introduit alors la notion de 'marge souple' en autorisant certains exemples à se trouver dans l'espace situé entre les deux marges. On cherche alors la fonction de décision f qui maximise la marge et minimise le coût des mauvaises classifications estimée par la somme pondérée des distances ξ_i de chaque échantillon \mathbf{x}_i à la marge par le terme de coût $C > 0$, comme illustré sur la Figure B.7-6 (droite), soit

$$\begin{cases} \min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i \\ \text{s.c. } y_i \left(\mathbf{w}'\mathbf{x}_i + b \right) > 1 - \xi_i, \quad i=(1, \dots, n) \\ \xi_i > 0 \end{cases} \quad (\text{B.7-39})$$

Le problème d'optimisation est équivalent au modèle de l'équation (B.7-37) avec

$$L(y_i, f(\mathbf{x}_i)) = \max(0, 1 - y_i f(\mathbf{x}_i)) \text{ et } \Omega(f) = \|\mathbf{w}\|^2$$

Le terme de coût $L(\cdot, \cdot)$ est appelé 'coût charnière' ou coût 'Hinge' en anglais et le terme de régularisation $\Omega(\cdot)$ est la norme ℓ_2 .

Ce problème est un programme quadratique de taille p , la dimension de l'espace des observations. On peut le résoudre directement dans l'espace primal à l'aide d'algorithmes d'optimisation numérique tel que l'algorithme de gradient conjugué.

Il peut également être intéressant de le reformuler dans l'espace dual. La recherche de la solution passe alors par l'écriture du Lagrangien défini comme la somme de la fonction objectif et d'une combinaison linéaire des contraintes pondérées par les variables duales

$$\mathcal{L}(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2} \|\mathbf{w}\|^2 - C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w}' \mathbf{x}_i + b) - (1 - \xi_i)) - \sum_{i=1}^n \beta_i \xi_i \quad (\text{B.7-40})$$

avec $\xi = [\xi_1, \dots, \xi_n]$

$\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_n]$ et $\boldsymbol{\beta} = [\beta_1, \dots, \beta_n]$, l'ensemble des variables duales, ou multiplicateurs de Lagrange.

D'après le théorème de Karush, Kuhn et Tucker, la solution (\mathbf{w}, b, ξ) du problème de minimisation sous contrainte défini en (B.7-40) correspond à un point selle du Lagrangien défini comme

$$\begin{cases} \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \min_{\mathbf{w}, b, \xi} \mathcal{L}(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{s.c. } \alpha_i > 0, \quad \beta_i > 0, \quad i = 1, \dots, n \end{cases} \quad (\text{B.7-41})$$

sous les contraintes suivantes appelées *conditions de complémentarité*

$$\begin{cases} \alpha_i (y_i (\mathbf{w}' \mathbf{x}_i + b) - 1 + \xi_i) = 0 & i = (1, \dots, n) \\ \beta_i \xi_i = 0 & i = (1, \dots, n) \end{cases} \quad (\text{B.7-42})$$

Le point selle du lagrangien $\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}, \boldsymbol{\beta})$ est atteint lorsque les dérivées par rapport à chaque variable primaire (\mathbf{w}, b, ξ) s'annulent. On obtient alors *les conditions de stationnarité* suivantes :

$$\begin{cases} \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ C - \alpha_i - \beta_i = 0 \quad i = 1, \dots, n \end{cases} \quad (\text{B.7-43})$$

La dernière condition permet d'éliminer les termes β_i car

$$\beta_i > 0 \text{ et } C - \alpha_i - \beta_i = 0 \Rightarrow 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n$$

Après, simplification, l'expression du Lagrangien devient alors

$$\mathcal{L}(\mathbf{w}, b, \xi, \boldsymbol{\alpha}, \boldsymbol{\beta}) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i' \mathbf{x}_j + \sum_{i=1}^n \alpha_i \quad (\text{B.7-44})$$

On se ramène ainsi à la résolution d'un problème de minimisation quadratique sous contrainte de taille n qui ne dépend plus que de α .

$$\begin{cases} \min_{\alpha} -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j + \sum_{i=1}^n \alpha_i \\ \text{s.c. } 0 \leq \alpha_i \leq C \quad i=(1, \dots, n) \\ \text{et } \mathbf{y}^t \boldsymbol{\alpha} = 0 \end{cases} \quad (\text{B.7-45})$$

A partir de la condition d'optimalité $\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i$ définie dans l'équation (B.7-43), et de la matrice $\mathbf{G} \in \mathcal{M}_n(\mathbb{R})$ appelée matrice de Gram ou matrice des influences définie par

$$G_{ij} = y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (\text{B.7-46})$$

On obtient

$$\mathbf{w}^t \mathbf{w} = \boldsymbol{\alpha}^t \mathbf{G} \boldsymbol{\alpha} = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^t \mathbf{x}_j$$

Le problème (B.7-45) peut alors se formuler comme

$$\begin{cases} \min_{\alpha} \frac{1}{2} \boldsymbol{\alpha}^t \mathbf{G} \boldsymbol{\alpha} - \mathbf{e}^t \boldsymbol{\alpha} \\ \text{s.c. } 0 \leq \alpha_i \leq C \quad i=(1, \dots, n) \\ \text{et } \mathbf{y}^t \boldsymbol{\alpha} = 0 \end{cases} \quad (\text{B.7-47})$$

avec \mathbf{e} un vecteur unité de dimension n .

Ce problème peut être résolu au moyen de différentes méthodes de résolution de problème quadratique et permet donc d'obtenir le vecteur de paramètres α . On peut alors en déduire l'expression de la fonction de décision comme

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i^t \mathbf{x} + b \quad (\text{B.7-48})$$

Les conditions de complémentarité du théorème de Kuhn-Tucker (cf (B.7-42)) permettent d'établir que les points dont le coefficient de Lagrange α_i est non nul doivent vérifier la condition $y_i (\mathbf{w}^t \mathbf{x}_i + b) = 1 - \xi_i$, c'est-à-dire être situés sur les marges. Ces points sont appelés *vecteurs support*. D'après l'expression (B.7-48), ce sont eux qui déterminent la forme de la fonction de décision f .

Le coefficient b peut être obtenu par l'expression $b = y_{\text{sup}} - \mathbf{w}^t \mathbf{x}_{\text{sup}}$ ou \mathbf{x}_{sup} est un vecteur support.

(b) Formulation généralisée dans le cas non linéaire

Dans le cas de données non linéairement séparables, la méthode consiste à projeter les données de l'espace $\mathcal{X} = \mathbb{R}^p$ vers un espace $\mathcal{Z} = \mathbb{R}^q$ avec $q \gg p$ de dimension supérieure et doté d'un produit scalaire $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$, dans lequel une séparatrice linéaire peut être trouvée.

Soit la fonction de projection $\phi(\mathbf{x})$

$$\phi: \mathbb{R}^p \rightarrow \mathbb{R}^q, \quad q \gg p$$

et $f(\mathbf{x}_i) = \mathbf{w}^T \phi(\mathbf{x}_i) + b$ l'hyperplan séparateur dans l'espace de redescription \mathcal{H}

On peut alors suivre un cheminement similaire à celui du cas linéaire et aboutir au problème d'optimisation

$$\begin{cases} \max_{\alpha} -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_H + \sum_{i=1}^n \alpha_i \\ \text{avec } 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n \\ \text{et } \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (\text{B.7-49})$$

qui permet de dériver l'expression de la fonction de décision

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle_H + b \quad (\text{B.7-50})$$

L'expression (B.7-53) montre que la fonction de décision ne dépend que des produits scalaires entre les couples de points d'apprentissage $\phi(\mathbf{x})$ projetés dans l'espace de redescription \mathcal{H} .

On considère alors l'ensemble des fonctions $\phi(\mathbf{x})$ appartenant à un espace de Hilbert à noyau reproduisant (EHNR). Dans cet espace, on peut définir une fonction noyau $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ symétrique telle que

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad (\text{B.7-51})$$

On montre alors que l'expression (B.7-49) peut s'écrire sous forme d'un problème d'optimisation similaire à (B.7-47) dans l'espace dual avec $G_{ij} = y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$. La frontière de décision s'exprime alors comme

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \quad (\text{B.7-52})$$

Les fonctions noyaux permettent ainsi d'effectuer les calculs dans l'espace d'origine \mathcal{X} de dimension inférieure à l'espace de redescription \mathcal{H} sans toutefois imposer de limites sur la complexité de la transformation $\phi(\mathbf{x})$ et la dimension de \mathcal{H} .

Dans les cas pratique, les noyaux les plus utilisés sont

Le noyau linéaire :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

Le noyau gaussien :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (\text{B.7-53})$$

(iv) SVM et probabilités

L'utilisation classique du SVM consiste à considérer la fonction de décision $D(\mathbf{x}_i) = \text{signe}(f(\mathbf{x}_i))$ pour la tâche de classification.

Il peut cependant être intéressant d'estimer une probabilité d'appartenance à la classe l_1 (ou l_{-1}). Les modèles génératifs dont nous avons parlé en introduction (Classifieur de Bayes, régression logistique) estiment directement cette probabilité. Dans le cas des modèles discriminants, on peut modéliser cette probabilité à partir de la mesure de la distance du point d'intérêt \mathbf{x}_i à l'hyperplan séparateur à partir de la distance du point à l'hyperplan. On modélise la probabilité a posteriori de la classe l_1 , $\mathbb{P}(Y = l_1 | X = \mathbf{x}_i)$ à partir de la formule de Bayes,

$$\mathbb{P}(Y = l_1 | X = \mathbf{x}_i) = \frac{\mathbb{P}(X = \mathbf{x}_i | Y = l_1) \mathbb{P}(Y = l_1)}{\mathbb{P}(X = \mathbf{x}_i | Y = l_1) \mathbb{P}(Y = l_1) + \mathbb{P}(X = \mathbf{x}_i | Y = l_{-1}) \mathbb{P}(Y = l_{-1})} \quad (\text{B.7-54})$$

que l'on peut formuler à l'aide d'une fonction sigmoïde

$$\mathbb{P}(Y = l_1 | X = \mathbf{x}) = \frac{1}{1 + \exp(-\varphi(\mathbf{x}))} \quad (\text{B.7-55})$$

en posant

$$\varphi(\mathbf{x}) = \ln\left(\frac{\mathbb{P}(X = \mathbf{x}_i | Y = l_{-1}) \mathbb{P}(Y = l_{-1})}{\mathbb{P}(X = \mathbf{x}_i | Y = l_1) \mathbb{P}(Y = l_1)}\right) \quad (\text{B.7-56})$$

Le modèle proposé par Platt [Platt, 1999] suppose que $\varphi(\mathbf{x})$ est une fonction linéaire de la fonction discriminante $f(\mathbf{x})$ soit $\varphi(\mathbf{x}) = -Af(\mathbf{x}) + B$. Les paramètres A et B sont appris sur l'ensemble d'apprentissage.

B.7.4.3 Analyse discriminante linéaire

Comme pour le SVM, nous présentons le formalisme de l'analyse discriminante linéaire dans le cas d'un problème de classification binaire de n individus $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathcal{X} \times Y$ avec $\mathcal{X} = \mathbb{R}^p$ et $\mathcal{Y} = \{-1, 1\}$.

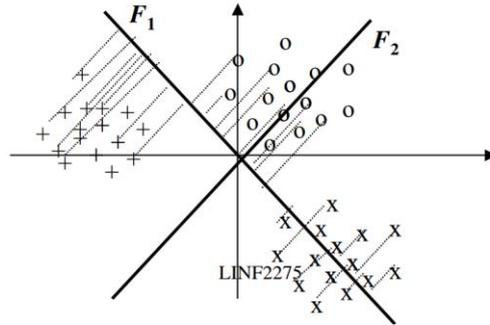


Figure B.7-7. Illustration du principe de classification par LDA.

L'analyse de Fisher repose sur l'hypothèse d'une fonction de séparation linéaire des classes $f(\mathbf{x}) = \mathbf{w}'\mathbf{x} + b$ et cherche l'axe de projection \mathbf{w} qui maximise la séparation des classes dans le repère de l'axe de projection comme représenté sur la Figure B.7-7.

Cette condition s'exprime par la maximisation du critère de Fisher représentant le rapport de la projection de la variance inter-classe sur la projection de la variance intra-classe

$$J(\mathbf{w}) = \frac{\mathbf{w}'S_b\mathbf{w}}{\mathbf{w}'S_w\mathbf{w}} \quad (\text{B.7-57})$$

On peut remarquer que le critère de Fisher est invariant par rapport à la transformation $\mathbf{w} \rightarrow \alpha\mathbf{w}$. On peut donc considérer l'ensemble des solutions \mathbf{w} correspondant à $\mathbf{w}'S_w\mathbf{w} = 1$, ce qui permet de reformuler le problème sous forme du problème suivant d'optimisation sous contrainte

$$\begin{cases} \max_{\mathbf{w}} \mathbf{w}'S_b\mathbf{w} \\ \text{avec } \mathbf{w}'S_w\mathbf{w} = 1 \end{cases} \quad (\text{B.7-58})$$

La recherche de la solution passe alors par l'écriture du Lagrangien

$$\mathcal{L}(\mathbf{w}) = \mathbf{w}'S_b\mathbf{w} - \lambda(\mathbf{w}'S_w\mathbf{w} - 1)$$

L'annulation de la dérivée de $\mathcal{L}(\mathbf{w})$ par rapport à \mathbf{w} permet d'aboutir à l'expression

$$(S_w)^{-1}S_b\mathbf{w} = \lambda\mathbf{w}$$

\mathbf{w} est donc un vecteur propre de la matrice $(S_w)^{-1}S_b$ associée à la valeur propre λ .

Or S_w est une matrice de rang 1 car elle s'écrit comme le produit de deux vecteurs colonnes (cf equation (B.7-3)). La matrice $S_w^{-1}S_b$ est donc également une matrice de rang 1 car $\text{rang}((S_w)^{-1}S_b) \leq \min(\text{rang}((S_w)^{-1}), \text{rang}(S_b))$. On peut ainsi en déduire que \mathbf{w} est le seul vecteur propre de cette matrice associé à la valeur propre non nulle λ .

On peut alors montrer que

$$\mathbf{w} = (\mathbf{S}_w)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{-1}) \quad (\text{B.7-59})$$

Et

$$\lambda = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{-1})^t (\mathbf{S}_w)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{-1}) \quad (\text{B.7-60})$$

L'expression de \mathbf{w} est ainsi directement reliée à celle de la variable de décision du modèle de Hotelling présenté en B.7.3.2(iv).

Dans le cas de deux distributions gaussiennes de même covariance ($\mathbf{S}_w^{-1} = \mathbf{S}_w^1$), on peut montrer que l'expression du biais b permettant une séparation optimale (au sens de Bayes) est donnée par

$$b = \frac{1}{2} \left(\bar{\mathbf{x}}_{-1}^t (\mathbf{S}_w^{-1})^{-1} \bar{\mathbf{x}}_{-1} - \bar{\mathbf{x}}_1^t (\mathbf{S}_w^1)^{-1} \bar{\mathbf{x}}_1 \right) \quad (\text{B.7-61})$$

B.8 Références bibliographiques

B.8.1 Bibliographie

- S. Agostinelli (2003). Geant4 - A Simulation Toolkit. *Nuclear Instruments and Methods in Physics Research Section A*, A506: 250-303.
- T. A. Alonzo et M. S. Pepe (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, 3(3): 421-432.
- R. D. Badawi, P. K. Marsden, B. F. Cronin, J. L. Sutcliffe et M. N. Maisey (1996). Optimization of noise-equivalent count rates in 3D PET. *Physics in Medicine and Biology*, 41: 1755-1776.
- H. H. Barrett (1990). Objective assessment of image quality: Effects of quantum noise and object variability. *Journal of the Optical Society (A)*, 7: 1266-1278.
- H. H. Barrett et C. K. Abbey (1997). *Bayesian detection of random signals on random backgrounds*. Information Processing in Medical Imaging.
- H. H. Barrett, J. Yao, J. P. Rolland et K. J. Myers (1993). Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences*, 90(21): 9758-9765.
- M. P. Bettinardi V, Danna M, Giovacchini G, Landoni C, Picchio M, Gilardi MC, Savi A, Castiglioni I, Lecchi M, Fazio F. (2007). Two-dimensional vs three-dimensional imaging in whole body oncologic PET/CT: a Discovery-STE phantom and patient study. *Quarterly Journal of Nuclear Medicine and Molecular Imaging*, 51(3): 214-223.
- L. M. Biberman (1973). *Perception of displayed information*. New York, Plenum.
- R. Boellaard, N. C. Krak, O. S. Hoekstra et A. A. Lammertsma (2004). Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *Journal of Nuclear Medicine*, 45(9): 1519-1527.
- S. Börjesson, M. Håkansson, M. Båth, S. Kheddache, S. Svensson, A. Tingberg, A. Grahn, M. Ruschin, B. Hemdal, S. Mattsson et L. G. Månsson (2005). A software tool for increased efficiency in observer performance studies in radiology. *Radiation Protection Dosimetry*, 114(1-3): 42-52.
- M. Brambilla, R. Matheoud, C. Secco, G. Sacchetti, S. Comi, M. Rudoni, A. Carriero et E. Inglese (2007). Impact of target-to-background ratio, target size, emission scan duration, and activity on physical figures of merit for a 3D LSO-based whole body PET/CT scanner. *Medical Physics*, 34(10): 3854-3865.
- J. G. Brankov (2013). Evaluation of the channelized Hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection. *Physics in Medicine and Biology*, 58(20): 7159-7182.
- J. G. Brankov, Y. Yang, L. Wei, I. El Naqa et M. Wernick (2009). Learning a channelized observer for image quality assessment. *IEEE Transactions on Medical Imaging*, 28: 991-999.
- A. E. Burgess, R. F. Wagner, R. J. Jennings et H. B. Barlow (1981). Efficiency of human visual signal detection. *Science*, 214: 93-94.
- I. Buvat, I. Castiglioni, J. Feuardent et M. C. Gilardi (2005). Unified description and validation of Monte Carlo simulators in PET. *Physics in Medicine and Biology*, 50: 329-346.
- I. Castiglioni, Cremonesi O, M. C. Gilardi, V. Bettinardi, G. Rizzo, A. Savi, E. Bellotti et F. Fazio (1999). Scatter correction techniques in 3D PET: a Monte Carlo evaluation. *IEEE Transactions on Nuclear Sciences*, 46: 2053-2058.
- D. P. Chakraborty et L. H. Winter (1990). Free-response methodology: Alternate analysis and a new observer-performance experiment. *Radiology*, 174: 873-881.
- S. Y. Chun, T. G. Reese, J. Ouyang, B. Guerin, C. Catana, X. Zhu, N. M. Alpert et G. El Fakhri (2012). MRI-based nonrigid motion correction in simultaneous PET/MRI. *Journal of Nuclear Medicine*, 53(8): 1284-1291.

- C. Comtat, P. E. Kinahan, M. Defrise, C. Michel, C. Lartzien et D. W. Townsend (1999). *Simulating whole-body PET scanning with rapid analytical methods*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Seattle, WA.
- C. Comtat, P. E. Kinahan, M. Defrise, C. Michel et D. W. Townsend (1998). Fast reconstruction of 3D PET data with accurate statistical modeling. *IEEE Transactions on Nuclear Science*, 45(3): 1083-1089.
- G. J. Cook, C. Yip, M. Siddique, V. Goh, S. Chicklore, A. Roy, P. Marsden, S. Ahmad et D. Landau (2013). Are Pretreatment 18F-FDG PET Tumor Textural Features in Non–Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy? *Journal of Nuclear Medicine*, 54(19-26)
- P. D. Cutler et M. Xu (1996). Strategies to improve 3D whole-body PET image reconstruction. *Physics in Medicine and Biology*,(41): 1453-1467.
- D. D. Dorfman, K. S. Berbaum et C. E. Metz (1992). ROC characteristic rating analysis: Generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*, 27: 723-731.
- M. P. Eckstein et J. S. Whiting (1995). Lesion detection in structured noise. *Acad. Radiol.*, 2(3): 249-253.
- J. P. Egan (1975). *Signal Detection Theory and ROC analysis*. New York, Academic Press.
- M. El Azami, A. Hammers, N. Costes et C. Lartzien (2013). *Computer aided diagnosis of intractable epilepsy with MRI imaging based on textural information*. 3rd International Workshop on Pattern Recognition in NeuroImaging (PRNI), Philadelphia, USA.
- M. El Azami, C. Lartzien et S. Canu (2014). *Robust outlier detection with L0-SVDD*. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), bruges (Belgium).
- G. El Fakhri, C. Holdsworth, R. D. Badawi, P. A. Santos, S. C. Moore, A. D. Van den Abbeele et M. F. Kijewski (2002). *Impact of acquisition geometry and patient habitus on lesion detectability in whole-body FDG-PET: a channelized Hotelling observer study*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Norfolk, VA.
- G. El Fakhri, P. A. Santos, R. D. Badawi, C. H. Holdsworth, A. D. Van Den Abbeele et M. F. Kijewski (2007). Impact of acquisition geometry, image processing, and patient size on lesion detection in whole-body F-18-FDG PET. *Journal of Nuclear Medicine*, 48(12): 1951-1960.
- G. El Fakhri, S. Surti, C. M. Trott, J. Scheuermann et J. S. Karp (2011). Improvement in Lesion Detection with Whole–Body Oncologic TOF-PET. *Journal of Nuclear Medicine*, 52(3): 347–353.
- I. El Naqa, P. W. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, W. L. Thorstad et J. O. Deasy (2009). Exploring feature-based approaches in PET imaging for predicting cancer treatment outcomes. *Pattern Recognition*, 42: 1162-1171.
- T. H. Farquhar, J. Llacer, C. K. Hoh, J. Czernin, S. S. Gambhir, M. A. Seltzer, D. H. Silverman, J. Qi, C. Hsu et E. J. Hoffman (1999). ROC and localization ROC analyses of lesion detection in whole-body FDG PET: Effects of acquisition mode, attenuation correction and reconstruction algorithm. *Journal of Nuclear Medicine*, 40(12): 2043-2052.
- Y. Freund et R. Schapire (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1): 119-139.
- B. D. Gallas et L. L. Pesce (2009). *Comparison of ROC methods for partially paired data*. SPIE Medical Imaging.
- H. C. Gifford (2012). *Performance characteristics of a visual-search human-model observer with sparse PET image data*. Society of Photo-Optical Instrumentation Medical Imaging Conference, San Diego, CA, SPIE.

- H. C. Gifford, P. E. Kinahan, C. Lartizien et M. A. King (2007). Evaluation of Multiclass Model Observers in PET LROC Studies. *IEEE Transactions on Nuclear Science*, 54(1): 116-123.
- H. C. Gifford, R. G. Wells et M. A. King (1999). *A comparison of human observer LROC and numerical observer ROC for tumor detection in SPECT images*. IEEE Nuclear Science Symposium and Medical Imaging Conference.
- H. C. Gifford, R. G. Wells et M. A. King (1999). *Human detection performance in PET and time-of-flight PET*. International Meeting on Fully Three-dimensional Image Reconstruction in Radiology and Nuclear Medicine.
- T. Glatard, C. Lartizien, B. Gibaud, R. Ferreira da Silva, G. Forestier, F. Cervenansky, M. Alessandrini, H. Benoit-Cattin, O. Bernard, S. Camarasu-Pop, N. Cerezo, P. Clarysse, A. Gaignard, P. Hugonnard, H. Liebgott, S. Marache, A. Marion, J. Montagnat, J. Tabary et D. Friboulet (2013). A Virtual Imaging Platform for multi-modality medical image simulation. *IEEE Transactions on Medical Imaging*, 32(1): 110-118.
- B. Goossens, H. Luong, L. Platiša et W. Philips (2013). *Objectively measuring signal detectability, contrast, blur and noise in medical images using channelized joint observers* SPIE Medical Imaging, Orlando.
- D. M. Green et J. A. Swets (1966). *Signal Detection Theory and Psychophysics*. New York.
- I. Guyon et A. Elisseeff (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3: 1157-1182.
- M. Håkansson, S. Svensson, S. Zachrisson, A. Svalkvist, M. Båth et L. G. Månsson (2010). VIEWDEX: an efficient and easy-to-use software for observer performance studies. *Radiation Protection Dosimetry*, 139(1-3): 42-51.
- B. S. Halpern, M. Dahlbom, A. Quon, C. Schiepers, C. Waldherr, D. H. Silverman, O. Ratib et J. Czernin (2004). Impact of patient weight and emission scan duration on PET/CT image quality and lesion detectability. *Journal of Nuclear Medicine*, 45(5): 797-801.
- T. Hambrock, P. C. Vos, C. A. Hulsbergen-van de Kaa, J. O. Barentsz et H. J. Huisman (2013). Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging--effect on observer performance. *Radiology*, 266(2): 521-530.
- J. A. Hanley et B. J. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143: 29-36.
- R. L. Harrison, S. D. Vannoy, D. R. Haynor, S. B. Gillipsie, M. S. Kaplan et T. K. Lewellen (1993). *Preliminary experience with the photon history generator module for a public-domain simulation system for emission tomography*. IEEE Nuclear Science Symposium and Medical Imaging Conference, San Francisco.
- X. He et E. Frey (2009). ROC, LROC, FROC, AFROC: an alphabet soup. *Journal of the American College of Radiology*, 6: 652-655.
- C. W. Helstrom (1995). *Elements of Signal Detection and Estimation*, Prentice Hall.
- H. Hotelling (1931). The generalisation of Student's ratio. *Ann. Math. Stat.*, 2: 360-378.
- H. Hudson et R. Larkin (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13: 601-609.
- S. Jan, G. Santin, D. Strul, S. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P. M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A. F. Chatzioannou, Y. Choi, Y. H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S. J. Glick, C. J. Groiselle, D. Guez, P. F. Honore, S. Kerhoas-Cavata, A. S. Kirov, V. Kohli, M. Koole, M. Krieguer, D. J. Van der Laan, F. Lamare, G. Largeron, C. Lartizien, D. Lazaro, M. C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. Perez, U. Pietrzyk, F. R. Rannou, M. Rey, D. R. Schaart, C. R. Schmidlein, L. Simon, T. Y. Song, J. M. JVieira, D. Visvikis, R. Van de Walle, E. Wieërs et C. Morel (2004). GATE: a simulation toolkit for PET and SPECT. *Physics in Medicine and Biology*: 4543-4561.

- L. Janeiro, C. Comtat, C. Lartizien, P. E. Kinahan, M. Defrise, C. Michel, R. Trébossen et P. Almeida (2006). Numerical observer studies comparing FORE+AWOSEM, FORE+NECOSEM and NEC based fully 3D OSEM for 3D whole-body PET imaging. *IEEE Transactions on Nuclear Science*, 53(3): 1194-1199.
- M. E. Juweid, S. Stroobants, O. S. Hoekstra, F. M. Mottaghy, M. Dietlein, A. Guerhazi, G. A. Wiseman, L. Kostakoglu, K. Scheidhauer, A. Buck, R. Naumann, K. Spaepen, R. J. Hicks, W. A. Weber, S. N. Reske, M. Schwaiger, L. H. Schwartz, J. M. Zijlstra, B. A. Siegel et B. D. Cheson (2007). Use of Positron Emission Tomography for Response Assessment of Lymphoma: Consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *Journal of Clinical Oncology*, 25(5): 571-578.
- D. J. Kadrmas, M. Bugrahan Oktay, M. E. Casey et J. J. Hamill (2012). Effect of Scan Time on Oncologic Lesion Detection in Whole-Body PET. *IEEE Trans. Nuc. Sci.*, 59(5): 1940-1947.
- D. J. Kadrmas, M. E. Casey, N. F. Black, J. J. Hamill, V. Y. Panin et M. Conti (2009). Experimental Comparison of Lesion Detectability for Four Fully-3D PET Reconstruction Schemes. *IEEE Transactions on Medical Imaging*, 28(4): 523-534.
- D. J. Kadrmas, M. E. Casey, M. Conti, B. W. Jakoby, C. Lois et D. W. Townsend (2009). Impact of time-of-flight on PET tumor detection. *Journal of Nuclear Medicine*, 50(8): 1315-1323.
- D. J. Kadrmas et P. E. Christian (2002). Comparative Evaluation of Lesion Detectability for 6 PET Imaging Platforms Using a Highly Reproducible Whole-Body Phantom with ^{22}Na Lesions and Localization ROC Analysis. *Journal of Nuclear Medicine*, 43(11): 1545-1554
- N. A. Karakatsanis, G. Loudos, A. Rahmim et K. Nikita (2014). Monte-Carlo based characterization of the counting rate (NECR) response for personalized optimization of the administered activity in clinical PET imaging. *Frontiers in Biomedical Technology*, 1: 14-34.
- P. E. Kinahan, C. C. Meltzer, B. McCook, S. Chander, C. Lartizien et R. G. Swensson (2001). Image quality as a function of acquisition protocol for whole-body PET oncology imaging. *Journal of Nuclear Medicine*, 42(5): 201P.
- P. E. Kinahan, C. Michel, M. Defrise, D. W. Townsend, M. Sibomana, M. Lonnew, D. F. Newport et J. D. Luketich (1996). *Fast iterative image reconstruction of 3D PET data*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Anaheim, CA.
- P. E. Kinahan et J. G. Rogers (1989). Analytic 3D image reconstruction using all detected events. *IEEE Transactions on Nuclear Science*, 36: 964-968.
- M. A. King, D. J. De Vries et E. J. Soares (1997). *Comparison of the channelized Hotelling and human observers for lesion detection in hepatic SPECT imaging*. Society of Photo-Optical Instrumentation Medical Imaging Conference, San Diego, CA.
- G. Komar, M. Teräs, M. Seppänen, J. Hirvonen, T. Vahlberg, J. Bergman et H. Minn (2009). Comparison of 2D and 3D performance for FDG PET with different acquisition times in oncological patients. *Nuclear Medicine Communications*, 30(1): 16-24.
- M. Kupinski, J. W. Hoppin, E. Clarkson et H. H. Barrett (2003). Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," 20, 430-438 (2003). *Journal of the Optical Society of America A*, 20: 430-438.
- G. Lanckriet, N. Cristianini, L. El Ghaoui, P. Bartlett et M. Jordan (2004). Learning the Kernel Matrix with Semi-Definite Programming. *Journal of Machine Learning Research*, 5: 27-72.
- D. L. Langer, T. H. van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson et M. A. Haider (2009). Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of magnetic resonance imaging*, 30(2): 327-334.
- C. Lartizien, J.-B. Aubin et I. Buvat (2010). Comparison of bootstrap resampling methods for 3-D PET imaging. *IEEE Transactions on Medical Imaging*, 29(7): 1442-1454

- C. Lartzien, C. Comtat, P. E. Kinahan, N. Ferreira, B. Bendriem et R. Trébossen (2002). Optimization of injected dose based on noise equivalent count rates for 2- and 3-dimensional whole-body PET. *Journal of Nuclear Medicine*, 43(9): 1268-1278.
- C. Lartzien, O. De Dreuille, D. Brasse, J. F. Gaillard et R. Trébossen (2000). NEC performances of 3 whole-body PET scanners in clinical imaging : a correlation between anthropomorphic phantom measurements and patient studies. *J. Nucl. Med.*, 41: 180P.
- C. Lartzien, P. E. Kinahan et C. Comtat (2004). A lesion detection observer study comparing 2D versus fully-3D whole-body PET imaging protocols. *Journal of Nuclear Medicine*, 45: 714-723.
- C. Lartzien, P. E. Kinahan et C. Comtat (2004). Volumetric model and human observer comparisons of tumor detection for whole-body PET. *Academic Radiology*, 11(6): 637-648.
- C. Lartzien, P. E. Kinahan, C. Comtat, M. Lin, R. Swensson, R. Trebossen et B. Bendriem (2000). *Evaluating Image Reconstruction Methods for Tumor Detection Performance in Whole-Body PET Oncology Imaging*. Society of Photo-Optical Instrumentation Medical Imaging Conference, San Diego, CA, SPIE.
- C. Lartzien, P. E. Kinahan, C. Comtat, M. Lin, R. Swensson, V. Villemagne, R. Trebossen et B. Bendriem (2000). Performing ROC studies with the same volume visualization software tools used in whole-body 3D PET oncology imaging. *Journal of Nuclear Medicine*, 41(5): P102 (abstract).
- C. Lartzien, P. E. Kinahan, R. Swensson, C. Comtat, M. Lin, V. Villemagne et R. Trébossen (2003). Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging. *Journal of Nuclear Medicine*, 44(2): 276-290.
- C. Lartzien, P. E. Kinahan, R. Swensson, C. Comtat et R. Trebossen (2001). *Numerical and human observer comparison of 2D versus 3D whole-body PET imaging protocols for tumor detection*. IEEE Nuclear Science Symposium and Medical Imaging Conference, San Diego, CA.
- C. Lartzien, C. Kuntner, A. L. Goertzen, A. C. Evans et A. Reilhac (2007). Validation of PET-SORTEO Monte Carlo simulations for the geometries of the MicroPET R4 and Focus 220 PET scanners. *Physics in Medicine and Biology*, 52: 4845-4862.
- C. Lartzien, S. Marache-Francisco et R. Prost (2012). Automatic detection of lung and liver lesions in 3D positron emission tomography images: a pilot study. *IEEE Transactions on Nuclear Science*, 59(1): 102-112.
- C. Lartzien, M. Rogez, E. Niaf et F. Ricard (2014). Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. *IEEE Journal of Biomedical and Health Informatics*, 18(3): 946-955.
- C. Lartzien, M. Rogez, A. Susset, F. Giammarile, E. Niaf et F. Ricard (2012). *Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information*. IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain.
- C. Lartzien, S. Tomei, A. Le Pogam, S. Marache-Francisco et D. Visvikis (2009). *Impact of Partial Volume Correction in Whole-Body PET Imaging : a Computer-Aided Detection Study*. IEEE Nuclear Science Symposium and Medical Imaging Conference, San Diego, CA.
- A. Le Maitre, W. P. Segars, S. Marache-Francisco, A. Reilhac, M. Hatt, S. Tomei, C. Lartzien et D. Visvikis (2009). Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications. *Proceedings of the IEEE*, 97(12): 2026-2038.
- J. Lehaire, R. Flamary, O. Rouviere et C. Lartzien (2014). *Computer aided diagnostic for prostate cancer detection and characterization combining learned dictionaries and supervised classification*. IEEE international conference on image processing (ICIP), Paris, France.
- J. Lehaire, C. Lartzien, E. Niaf, M. Baumann, R. Souchon, F. Bratan et O. Rouviere (2013). *Evaluation of a 3D MR/MR elastic registration method for prostate cancer imaging*. European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Toulouse, France.

- C. Lin et S. Wang (2002). Fuzzy support vector machine. *IEEE transactions on neural networks*, 13(2): 464–471.
- J. Llacer, E. Veklerov, L. R. Baxter, S. T. Grafton, L. K. Griffeth, R. A. Hawkins, C. K. Hoh, J. C. Mazziotta, E. J. Hoffman et C. E. Metz (1993). Results of clinical receiver operating characteristics study comparing filtered backprojection and maximum likelihood estimator images in FDG PET studies. *Journal of Nuclear Medicine*, 1993: 1198-1203.
- M. A. Lodge, R. D. Badawi, R. Gilbert, P. E. Dibos et B. R. Line (2006). Comparison of 2-dimensional and 3-dimensional acquisition for 18F-FDG PET oncology studies performed on an LSO-based scanner. *Journal of Nuclear Medicine* 47(1): 23-31.
- V. J. Lowe, J. M. Hoffman, D. M. DeLong, E. F. Patz et R. E. Coleman (1994). Semiquantitative and visual analysis of FDG-PET images in pulmonary abnormalities. *J. Nucl. Med.*, 35: 1771-1776.
- M. Lubberink, R. Boellaard, A. P. van der Weerd, F. C. Visser et A. A. Lammertsma (2004). Quantitative comparison of analytic and iterative reconstruction methods in 2- and 3-dimensional dynamic cardiac 18F-FDG PET. *Journal of Nuclear Medicine*, 45(12): 2008-2015.
- J. Mairal, F. Bach et J. Ponce (2012). Task-driven dictionary learning. *IEEE transactions on pattern analysis and machine intelligence*, 34(4): 791–804.
- S. Marache-Francisco, R. Prost, J.-M. Rouet et C. Lartzien (2011). *Incorporating Patient-Specific Variability in the OncoPET_DB Database*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Valencia, Spain.
- R. Matheoud, L. Leva, C. Secco, G. Sacchetti, P. D. Monica, M. Brambilla et E. Inglese (2011). Small lesions detectability with the Biograph 16 Hi-Res PET/CT scanner and fast imaging protocols: performance evaluation using an anthropomorphic thoracic phantom and ROC analyses. *Annals of Nuclear Medicine*, 25(3): 179-188.
- C. E. Metz (1986). ROC methodology in radiologic imaging. *Invest. Radiol.*, 21: 720-733.
- C. E. Metz (2006). Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *Journal of the American College of Radiology*, 3(6): 413-422.
- C. Michel, M. Sibomana, A. Bol, X. Bernard, M. Lonneaux, M. Defrise, C. Comtat, P. E. Kinahan et D. W. Townsend (1998). *Preserving Poisson characteristics of PET data with weighted OSEM reconstruction*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Toronto, Ontario, IEEE.
- S. C. Moore, G. El Fakhri, R. Badawi, A. D. Van den Abbeele et R. E. Zimmerman (2000). *Relative lesion detectability in 3D vs. 2D dedicated multi-ring PET*. IEEE Nuclear Science Symposium and Medical Imaging Conference, Lyon, France.
- K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton et G. W. Seeley (1985). Effect of noise correlation on detectability of disk signals in medical imaging. *J. Opt. Soc. Am.*, A2: 1752-1759.
- E. Niaf, R. Flamary, S. Canu, O. Rouviere et C. Lartzien (2012). *Handling learning samples uncertainties in SVM classification: application to MRI-based prostate cancer computer-aided diagnosis*. IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain.
- E. Niaf, R. Flamary, C. Lartzien et S. Canu (2011). *Handling uncertainties in SVM classification*. IEEE Workshop on Statistical Signal Processing, Nice, Fr.
- E. Niaf, R. Flamary, A. Rakotomamonjy, O. Rouviere et C. Lartzien (2014). *SVM with feature selection and smooth prediction in images : application to CAD of prostate cancer*. IEEE International conference on image processing, Paris, France.
- E. Niaf, R. Flamary, O. Rouviere, C. Lartzien et S. Canu (2014). Kernel-Based Learning From Both Qualitative and Quantitative Labels: Application to Prostate Cancer Diagnosis Based on Multiparametric MR Imaging. *IEEE Transactions on Image Processing*, 23(3): 979-991.

- E. Niaf, C. Lartzien, F. Bratan, L. Roche, M. Rabilloud, F. Mège-Lechevallier et O. Rouvière (2014). Prostate Focal Peripheral Zone Lesions: Characterization at Multiparametric MR Imaging—Influence of a Computer-aided Diagnosis System *Radiology*, 271(3): 761-769.
- E. Niaf, O. Rouvière, F. Mege-Lechevallier, F. Bratan et C. Lartzien (2012). Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine and Biology*, 57(12): 3833-3851.
- M. Niemeijer, M. Loog, M. D. Abramoff, M. A. Viergever, M. Prokop et B. Van Ginneken (2011). On combining computer-aided detection systems. *IEEE Transactions on Medical Imaging*, 30(2): 215-223.
- F. Orlhac, M. Soussan, J. A. Maisonobe, C. A. Garcia, B. Vanderlinden et I. Buvat (2014). Tumor texture analysis in 18F-FDG-PET: relationships between texture parameters, histogram indices, SUVs, metabolic volumes and total lesion glycolysis. *Journal of Nuclear Medicine*, 55: 414-422.
- S. Ozer, D. L. Langer, X. Liu, M. A. Haider, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick et I. S. Yetik (2010). Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*, 2010(37): 4.
- S. J. Park, D. Ionascu, J. Killoran, M. Mamede, V. H. Gerbaudo, L. Chin et R. Berbeco (2008). Evaluation of the combined effects of target size, respiratory motion and background activity on 3D and 4D PET/CT images. *Physics in Medicine and Biology*, 53(13): 3661-3679.
- H. Peng, F. Long et C. Ding (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on pattern analysis and machine intelligence*, 27(8): 1226-1238.
- Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener et A. Oto (2013). Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score--a computer-aided diagnosis development study. *Radiology*, 267(3): 787-796.
- L. L. Pesce et C. E. Metz (2007). Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. . *Academic Radiology*, 14: 814-829.
- N. Petrick, B. Sahiner, S. G. Armato, A. Bert, L. Correale, S. Delsanto, M. T. Freedman, D. Fryd, D. Gur, L. Hadjiiski, Z. Huo, Y. Jiang, L. Morra, S. Paquerault, V. Raykar, F. W. Samuelson, R. M. Summers, G. Tourassi, H. Yoshida, B. Zheng, C. Zhou et H. P. Chan (2013). Evaluation of computer-aided detection and diagnosis systems. *Medical Physics*, 40(8): 087001-087001.
- L. Platiša, B. Goossens, E. Vansteenkiste, S. C. Park, B. D. Gallas, A. Badano et W. Philips (2011). Channelized Hotelling observers for the assessment of volumetric imaging data sets. *J Opt Soc Am A Opt Image Sci Vis.*, 28(6): 1145-1163.
- L. Platiša, A. Kumcu, M. Platisa, E. Vansteenkiste, K. Deblaere, A. Badano et W. Philips (2012). *Volumetric detection tasks with varying complexity : Human observer performance*. SPIE Medical Imaging.
- J. Platt (1999). Probabilistic outputs for support vector machines. *Advances in large margin classifiers*. M. Press. Cambridge, MA.
- A. Rahmim et J. Tang (2013). Noise propagation in resolution modeled PET imaging and its impact on detectability. *Physics in Medicine and Biology*, 58: 6945-6968.
- A. Reilhac (2007). Validation et exploitation d'un simulateur TEP de Monte Carlo, INSA Lyon.
- A. Reilhac, C. Lartzien, N. Costes, S. Sans, C. Comtat, R. Gunn et A. C. Evans (2004). PET-SORTEO: A Monte Carlo-based simulator with high count rate capabilities. *IEEE Transactions on Nuclear Science*, 51: 46-52.
- K. Rossmann (1966). Comparison of several methods for evaluating image quality of radiographic screen-film systems. *Am. J. Roentgenol. Radium. Ther. Nucl. Med.*, 97(3): 772-775.

- M. B. Sachs, J. Nachmias et J. G. Robson (1971). Spatial-Frequency Channels in Human Vision. *Journal of the Optical Society of America*, 61(9): 1176-1186.
- B. Sahiner, H. P. Chan et L. Hadjiiski (2008). Classifier performance prediction for computer-aided diagnosis using a limited dataset. *Medical Physics*, 35: 1559-1570.
- B. Sahiner, H. P. Chan, N. Petrick, R. F. Wagner et L. Hadjiiski (2000). Feature selection and classifier performance in computer-aided diagnosis: the effect of finite sample size. *Medical Physics*, 27: 1509-1522.
- J. Schaefferkoetter, M. Casey, D. W. Townsend et G. El Fakhri (2013). Clinical impact of time-of-flight and point response modeling in PET reconstructions: a lesion detection study. *Physics in Medicine and Biology*, 58(5): 1465-1478.
- B. Schölkopf, J. Platt, J. C. Shawe-Taylor, A. J. Smola et R. C. Williamson (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7): 1443-1471.
- E. Segal, C. B. Sirlin, C. Ooi, A. S. Adler, J. Gollub, X. Chen, B. K. Chan, G. R. Matcuk, C. T. Barry, H. Y. Chang et M. D. Kuo (2007). Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nature Biotechnology*, 25(6): 675-680.
- W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes et B. M. W. Tsui (2010). 4D XCAT phantom for multimodality imaging research. *Medical Physics*, 37(9): 4902-4915.
- D. M. Somford, T. Hambroek, C. A. Hulsbergen-van de Kaa, J. J. Fütterer, I. M. van Oort, J. P. van Basten, H. F. Karthaus, J. A. Witjes et J. O. Barentsz (2012). Initial experience with identifying high-grade prostate cancer using diffusion-weighted MR imaging (DWI) in patients with a Gleason score $\leq 3 + 3 = 6$ upon schematic TRUS-guided biopsy: a radical prostatectomy correlated series. *Investigative Radiology*, 47(3): 153-158.
- R. N. Strickland et H. Hahn (1996). Wavelet transform for detecting microcalcifications in mammograms. *IEEE Transactions on Medical Imaging*, 15(2): 218-229.
- K. Strobel, M. Rüdy, V. Treyer, P. Veit-Haibach, C. Burger et T. F. Hany (2007). Objective and subjective comparison of standard 2-D and fully 3-D reconstructed data on a PET/CT system. *Nuclear Medicine Communications*, 28(7): 555-559.
- S. C. Strother, M. E. Casey et E. J. Hoffman (1990). Measuring PET scanner sensitivity: Relating count rates to image signal-to-noise ratios using noise equivalent counts. *IEEE Transactions on Nuclear Science*, 37: 783-788.
- S. Surti, J. Scheuermann, G. El Fakhri, M. E. Daube-Witherspoon, R. Lim, N. Abi-Hatem, E. Moussallem, F. Benard, D. Mankoff et J. S. Karp (2011). Impact of time-of-flight PET on whole-body oncologic studies: a human observer lesion detection and localization study. *Journal of Nuclear Medicine*, 52: 712-719.
- R. G. Swensson (1996). Unified measurement of observer performance in detecting and localizing target objects on images. *Medical Physics*, 23: 1709-1725.
- J. A. Swets et R. M. Pickett (1982). *Evaluation of diagnostic systems : methods from signal detection theory*. New York, Academic Press.
- J. Tabary, S. Marache-Francisco, S. Valette, W. P. Segars et C. Lartzien (2009). *Realistic X-Ray CT Simulation of the XCAT Phantom with SINDBAD*. IEEE Nuclear Science Symposium and Medical Imaging Conference, San Diego, CA, USA.
- S. Tan, S. Kligerman, W. Chen, M. Lu, G. Kim, S. Feigenberg, W. D. D'Souza, M. Suntharalingam et W. Lu (2013). Spatial-temporal [^{18}F]FDG-PET features for predicting pathologic response of esophageal cancer to neoadjuvant chemoradiation therapy. *International Journal of Radiation Oncology . Biology . Physics* 85(5): 1375-1382.
- D. M. J. Tax et R. P. W. Duin (2004). Support vector data description. *Machine Learning*, 54(1): 45-66.

- H. Thoen, V. Keereman, P. Mollet, R. Van Hoken et S. Vandenberghe (2013). Influence of detector pixel size, TOF resolution and DOI on image quality in MR-compatible whole-body PET. *Physics in Medicine and Biology*, 58(18): 6459-6479.
- R. J. Tibshirani et R. Tibshirani (2009). A bias correction for the minimum error rate in cross-validation. *Annals of Applied Statistics*, 3(2): 822-829.
- F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos et D. Visvikis (2011). Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *Journal of Nuclear Medicine*, 52(3): 369-378.
- S. Tomei, A. Reilhac, D. Visvikis, N. Boussion, C. Odet, F. Giammarile et C. Lartzien (2010). OncoPET_DB: a freely distributed database of realistic simulated whole body [18F]FDG PET images for oncology. *IEEE Transactions on Nuclear Science*, 57(1): 246-255.
- B. M. W. Tsui, J. A. Terry et G. T. Gullberg (1993). Evaluation of cardiac cone-beam SPECT using observer performance experiments and ROC analysis. *Investigative Radiology*, 28(12): 1101-11012.
- H. L. Van Trees (1968). *Detection, Estimation and Modulation Theory*. New York, Wiley.
- V. Vapnik (1995). *Statistical learning theory*. New York, Wiley.
- S. Varma et R. Simon (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7: 91.
- D. Visvikis, D. Griffiths, D. C. Costa, J. B. Bomanji et P. J. Ell (2005). Clinical evaluation of 2D versus 3D whole-body PET image quality using a dedicated BGO PET scanner. *European Journal of Nuclear Medicine and Molecular Imaging*, 32(9): 1050-1056.
- S. Viswanath, B. Nicolas Bloch, M. Rosen, J. Chappelow, R. Toth, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur et A. Madabhushi (2009). *Integrating Structural and Functional Imaging for Computer Assisted Detection of Prostate Cancer on Multi-Protocol In Vivo 3 Tesla MRI*. SPIE Medical Imaging Conference.
- W. V. Vogel, B. M. Wensing, J. A. van Dalen, P. F. Krabbe, F. J. van den Hoogen et W. J. Oyen (2005). Optimised PET reconstruction of the head and neck area: improved diagnostic accuracy. *European Journal of Nuclear Medicine and Molecular Imaging*, 32(11): 1276-1282.
- P. C. Vos, T. Hambroek, J. O. Barentsz et H. J. Huisman (2010). Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Physics in Medicine and Biology*, 55(6): 1719-1734.
- P. C. Vos, T. Hambroek, C. A. Hulsbergen-van de Kaa, J. J. Fütterer, J. O. Barentsz et H. J. Huisman (2008). Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. *Medical Physics*, 35(3): 888-899.
- R. F. Wagner et D. G. Brown (1985). Unified SNR analysis of medical imaging systems. *Phys. Med. Biol.*, 30(6): 489-518.
- M. D. Walker, J. C. Matthews, M.-C. Asselin, A. Saleem, C. Dickinson, N. Charnley, P. J. Julyan, P. M. Price et T. Jones (2009). Optimization of the Injected Activity in Dynamic 3D PET : A Generalized Approach Using Patient-Specific NECs as Demonstrated by a Series of 15O-H₂O Scans. *Journal of Nuclear Medicine and Biology*, 50: 1409–1417.
- C. C. Watson, M. Casey, B. Bendriem, J. P. Carney, D. W. Townsend, S. Eberl, S. R. Meikle et F. P. DiFilippo (2005). Optimizing Injected Dose in Clinical PET by Accurately Modeling the Counting-Rate Response Functions Specific to Individual Patient Scans. *Journal of Nuclear Medicine and Biology*, 46: 1825–1834.
- R. G. Wells, M. A. King, H. C. Gifford et P. H. Pretorius (1999). *Single-slice versus multi-slice display for human-observer lesion-detection studies*. IEEE Nuclear Science Symposium and Medical Imaging Conference.

- R. G. Wells, M. A. King, H. C. Gifford et P. H. Pretorius (2000). Single-slice versus multi-slice display for human-observer lesion-detection studies. *IEEE Transactions on Nuclear Science*, 47(3): 1037-1044.
- J. M. Willaime, F. E. Turkheimer, L. M. Kenny et E. O. Aboagye (2013). Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of 18F fluorothymidine-positron emission tomography. *Physics in Medicine and Biology*, 58(2): 187-203.
- T.-H. Wu, J.-A. Yu et C.-C. Yang (2013). An Investigation of PET Scan Parameters for Lesion Detection With GE VISTA Small-Animal Scanner Using Monte Carlo Simulation and Multivariate Analysis. *IEEE Transactions on Nuclear Science*, 60(2): 774-782.
- J. Yao et H. H. Barrett (1992). *Predicting human performance by a channelized Hotelling observer model*. SPIE.
- H. Yu, C. Caldwell, C. Mah et D. Mozeg (2009). Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Transactions on Medical Imaging*, 28(3): 374-383.
- H. Zaidi, C. Labbe et C. Morel (1998). EIDOLON: implementation of an environment for Monte Carlo simulation of fully 3D positron tomography on a high-performance parallel platform. *Parallel Comput.*, 24: 1523-1536.
- L. Zhang, C. Cavaro-Ménard et P. Le Callet (2014). An overview of model observers. *IRBM*, 35(4): 214-224.
- L. Zhang, C. Cavaro-Ménard, P. Le Callet et J. Y. Tanguy (2012). A perceptually relevant channelized joint observer (PCJO) for the detection-localization of parametric signals. *IEEE Transactions on Medical Imaging*, 31(10): 1875-1888.
- L. Zhang, B. Goossens, C. Cavaro-Ménard, P. Le Callet et D. Ge (2013). Channelized model observer for the detection and estimation of signals with unknown amplitude, orientation, and size. *Journal of the Optical Society of America (A)*, 30(11): 2422-2432.
- I. G. Zubal, C. R. Harrell, E. O. Smith, Z. Rattner, G. Gindi et P. B. Hoffer (1994). Computerized three-dimensional segmented human anatomy. *Medical Physics*, 21: 299-302.

B.8.2 Liste de mes publications

B.8.2.1 Journaux internationaux à comité de lecture

30 publications dans des revues scientifiques avec comité de lecture

E. Niaf, C. Lartizien, F. Bratan, L. Roche, M. Rabilloud, F. Mège-Lechevallier, and O. Rouvière. *Characterizing prostate focal peripheral zone lesions at multiparametric MR imaging: Influence of a computer-aided diagnosis system*. **Radiology**, 271(3) : 761-769, 2014.

C. Lartizien, M. Rogez, E. Niaf, and F. Ricard. *Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information*. **IEEE Journal of Biomedical and Health Informatics**, 18(3) : 946-955, 2014.

E. Niaf, R. Flamary, O. Rouvière, C. Lartizien, and S. Canu. *Kernel-based learning from both qualitative and quantitative labels: Application to prostate cancer diagnosis based on multiparametric MR imaging*. **IEEE Transactions on Image Processing**, 23(3) 979- 991, 2014.

B. Gibaud, G. Forestier, H. Benoit-Cattin, F. Cervenansky, P. Clarysse, D. Friboulet, A. Gaignard, P. Hugonnard, C. Lartizien, H. Liebgott, et al. *OntoVIP: an ontology for the annotation of object models used for medical image simulation*. **Journal of Biomedical Informatics**, 52 : 279-292, 2014.

T. Glatard, C. Lartizien, B. Gibaud, R. Ferreira da Silva, G. Forestier, F. Cervenansky, M. Alessandrini, H. Benoit-Cattin, O. Bernard, S. Camarasu-Pop, N. Cerezo, P. Clarysse, A. Gaignard, P. Hugonnard, H. Liebgott, S. Marache, A. Marion, J. Montagnat, J. Tabary, and D. Friboulet. *A Virtual Imaging Platform for multi-modality medical image simulation*. **IEEE Transactions on Medical Imaging**, 32(1): 110-118, 2013.

E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien. *Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI*. **Physics in Medicine and Biology**, 57(12): 3833-3851, 2012.

C. Lartizien, S. Marache-Francisco, R. Prost. *Automatic Detection of Lung and Liver Lesions in 3D Positron Emission Tomography Images: a Pilot Study*. **IEEE Transactions on Nuclear Science**, 59(1): 102-112, 2012.

C. Lartizien, J.-B. Aubin, and I. Buvat. *Comparison of Bootstrap Resampling Methods for 3-D PET Imaging*. **IEEE Transactions on Medical Imaging**, 29(7): 1442-1454, 2010.

S. Tomei, A. Reilhac, D. Visvikis, N. Boussion, C. Odet, F. Giammarile, and C. Lartizien. *OncoPET_DB: a freely distributed database of realistic simulated whole body [18F]FDG PET images for oncology*. **IEEE Transactions on Nuclear Science**, 57(1): 246-255, 2010.

L. Baboi, F. Pilleul, L. Milot, C. Lartizien, G. Poncet, C. Roche, J.-Y. Scoazec, O. Beuf. *Magnetic resonance imaging follow-up of liver growth of neuroendocrine tumors in an experimental mouse model*. **Journal of Magnetic Resonance Imaging**, 28(2): 264-272, 2010.

A. Lemaitre, W. Segars, S. Marache, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, D. Visvikis. *Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications*. **Proceedings of the IEEE**, 97(12): 2026-38, 2009.

O. Beuf, C. Lartizien, L. Milot, L. Baboi, C. Roche, J.B. Langlois, J.Y. Scoazec, and F. Pilleul. *Multimodal imaging for the detection and characterization of liver lesions in a mouse model of neuroendocrine tumor*. **Gastroenterologie Clinique et Biologique**, 32(1):32-40, 2008.

C. Lartizien, C. Kuntner, A L Goertzen, AC Evans and A Reilhac. *Validation of PET-SORTEO Monte Carlo simulations for the geometries of the MicroPET R4 and Focus 220 PET scanners*. **Physics in Medicine and Biology**, 52: 4845–4862, 2007.

H.C. Gifford, P.E. Kinahan, C. Lartizien, M.A. King, *Evaluation of Multiclass Model Observers in PET LROC Studies*, **IEEE Transactions on Nuclear Science**. 54(1): 116 – 123, 2007.

- L.M. Riou, A. Broisat, **C. Lartizien**, MC Toufektsian, S. Maitrejean, M. Janier, G. Vanzetto, D. Fagret, C. Ghezzi. *Assessment of Non-Reperfused and Reperfused Myocardial Infarction using Diffusible or Deposited Radiolabelled Perfusion Imaging Agents*. **European Journal of Nuclear Medicine**, 34(3): 330-337, 2007.
- L. Janeiro, C. Comtat, **C. Lartizien**, P.E. Kinahan, M. Defrise, C. Michel, R. Trébossen and P. Almeida, *Numerical observer studies comparing FORE+AWOSEM, FORE+NECOSEM and NEC based fully 3D OSEM for 3D whole-body PET imaging*. **IEEE Transactions on Nuclear Science**. 53(3):1194-1199, 2006.
- D. Brasse, P. E. Kinahan, **C. Lartizien**, C. Comtat, M. Casey, and C. Michel, *Correction methods for random coincidences in fully 3D whole-body PET: impact on data and image quality*. **Journal of Nuclear Medicine**, 46: 859-867, 2005.
- K. Ziemons, E. Auffray, R. Barbier, G. Brandenburg, P. Bruyndonckx, Y. Choi, D. Christ, N. Costes, Y. Declais, O. Devroede, C. Dujardin, A.A. Fedorov, U. Heinrichs, A. Hollendung, M. J. Korjhik, M. Krieguer, C. Kuntner, G. LARGERON, **C. Lartizien**, H. Larue, P. Lecoq, S. Leonard, J. Marteau, C. Morel, J.B. Mosset, C. Parl, C. Pedrini, A. G. Petrosyan, U. Pietrzyk, M. Rey, S. Saladino, D. Sappey-Marinié, M. Janier, L. Simon, M. Streun, S. Tavernier, and J.M. Vieira. *The ClearPET™ project: development of a 2-nd generation high performance small animal PET scanner*. **Nuclear Instruments and Methods in Physics Research Section A**, 537:307-311, 2005.
- S. Jan, G. Santin, D. Strul, S. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P.M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A.F. Chatziioannou, Y. Choi, Y.H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S.J. Glick, C.J. Groiselle, D. Guez, P.-F. Honore, S. Kerhoas-Cavata, A.S. Kirov, V. Kohli, M. Koole, M. Krieguer, D.J. van der Laan, F. Lamare, G. LARGERON, **C. Lartizien**, D. Lazaro, M.C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. Perez, U. Pietrzyk, F.R. Rannou, M. Rey, D.R. Schaart, C.R. Schmittlein, L. Simon, T.Y. Song, J.-M. Vieira, D. Visvikis, R. Van de Walle, E. Wieërs, C. Morel, *GATE: a simulation toolkit for PET and SPECT*. **Physics in Medicine and Biology**, 49: 4543-4561, 2004.
- D. Sappey-Marinié, O. Beuf, C. Billotey, E. Chereul, J. Dupuy, M. Jeandey, D. Grenier, J. Hasserodt, J.B. Langlois, **C. Lartizien**, W. Mai, C. Odet, J. Samarut, D. Vray, L. Zimmer, and M. Janier. *The ANIMAGE project: a multimodal imaging platform for small animal research*. **Nuclear Instruments and Methods in Physics Research Section A**, 527:117-123, 2004.
- E. Auffray, P. Bruyndonckx, O. Devroede, A. Fedorov, U. Heinrichs, M. Korjik, M. Krieguer, C. Kuntner, **C. Lartizien**, P. Lecoq, S. Leonard, Ch. Morel, J. B. Mosset, Ch. Pedrini, A. Petrosian, U. Pietrzyk, M. Rey, S. Saladino, D. Sappey-Marinié, L. Simon, M. Streun, S. Tavernier, J. M. Vieira and K. Ziemons. *The ClearPET project*. **Nuclear Instruments and Methods in Physics Research Section A**, 527:171-174, 2004.
- C. Lartizien**, P.E. Kinahan, C. Comtat. *Volumetric model and human observer comparisons of tumor detection for whole-body PET*. **Academic Radiology**, 11(6): 637-648, 2004.
- J.-S. Kim, P.E. Kinahan, **C. Lartizien**, C. Comtat, T.K. Lewellen, *A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging*. **IEEE Transactions on Nuclear Science**, 51(1): 34–40, 2004.
- A. Reilhac, **C. Lartizien**, N. Costes, S. Sans, C. Comtat, R. Gunn, A. Evans. *PET-SORTEO: A Monte Carlo-based simulator with high count rate capabilities*. **IEEE Transactions on Nuclear Science**, 5: 46-52, 2004.
- C. Lartizien**, P.E. Kinahan, C. Comtat, R. Swensson. *A lesion detection observer study comparing 2-dimensional versus fully 3-dimensional whole-body PET imaging protocols*. **Journal of Nuclear Medicine**, 45: 714-723, 2004.
- C. Lartizien**, A. Reilhac, N. Costes, M. Janier, D. Sappey-Marinié. *Monte Carlo simulation-based design study of a LSO-LuAP small animal PET system*. **IEEE Transactions on Nuclear Science**, 50: 1433-1438, 2003.
- F. Dolle, H. Valette, Y. Bramouille, I. Guenther, C. Fuseau, C. Coulon, **C. Lartizien**, S. Jegham, P. George, O. Curet, J-L Pinquier and M. Bottlaender. *Synthesis and in vivo imaging properties of [C-*

11]befloxatone: A novel highly potent positron emission tomography ligand for mono-amine oxidase-A. **Bioorganic and Medicinal Chemistry Letters**, 13 (10): 1771-1775, 2003.

C. Lartizien, P.E. Kinahan, R. Swensson, C. Comtat, M. Lin, V. Villemagne and R. Trébossen. *Evaluating image reconstruction methods for tumor detection in 3-dimensional whole-body PET oncology imaging*. **Journal of Nuclear Medicine**, 44: 276-290, 2003.

C. Lartizien, C. Comtat, P.E. Kinahan, N. Ferreira, B. Bendriem, R. Trébossen. *Optimization of injected dose based on noise equivalent count rates for 2- and 3-dimensional whole-body PET*. **Journal of Nuclear Medicine**, 43: 1268-1278, 2002.

N.C. Ferreira, R. Trébossen, **C. Lartizien**, V. Brulon, P. Merceron, B. Bendriem. *A hybrid scatter correction for 3D PET based on an estimation of the distribution of unscattered coincidences: implementation on the ECAT EXACT HR+*. **Physics in Medicine and Biology**, 47: 1555-1571, 2002.

B.8.2.2 Conférences internationales avec comité de lecture et publication des actes

58 communications

(i) Année 2014

J. Lehaire, R. Flamary, O. Rouvière, and **C. Lartizien**. *Computer aided diagnostic for prostate cancer detection and characterization combining learned dictionaries and supervised classification*. **IEEE international conference on image processing (ICIP)**, Paris, France, 10/2014.

M. El Azami, **C. Lartizien**, and S. Canu. *Robust outlier detection with LO-SVDD*. **European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)**, Bruges (Belgium), 2014.

E. Niaf, R. Flamary, A. Rakotomamonjy, O. Rouvière, and **C. Lartizien**. *SVM with feature selection and smooth prediction in images: application to CAD of prostate cancer*. **IEEE International Conference on Image Processing (ICIP)**, Paris, 10/2014.

(ii) Année 2013

M. El Azami, A. Hammers, N. Costes, and **C. Lartizien**. *Computer Aided Diagnosis of Intractable Epilepsy with MRI Imaging Based on Textural Information*. International Workshop on Pattern Recognition in Neuroimaging (PRNI), Philadelphia, USA, pp. 90-93, 2013.

J. Lehaire, **C. Lartizien**, E. Niaf, M. Baumann, R. Souchon, F. Bratan, and O. Rouvière. *Evaluation of a 3D MR/MR elastic registration method for prostate cancer imaging*. **European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)**, Toulouse, 10/2013.

E. Hilaire, C. Robert, X. Lojacono, **C. Lartizien**, I. Buvat, and V. Maxim. *Imagerie Compton en protonthérapie : de la simulation GATE à la reconstruction d'images*. **XXIVème Colloque GRETSI - Traitement du Signal et des Images (GRETSI'13)**, Brest, France, pp. 1-4, 09/2013.

E. Hilaire, C. Robert, X. Lojacono, **C. Lartizien**, I. Buvat, and V. Maxim. *Compton imaging in proton beam therapy: from GATE simulations to image reconstruction*. **Interdisciplinary Symposium on Signals and Systems for Medical Applications**, Paris, France, 3-4 juin 2013.

(iii) Année 2012

C. Lartizien, M. Rogez, A. Susset, F. Giammarile, E. Niaf, F. Ricard, "Computer aided staging of lymphoma patients with FDG PET/CT imaging based on textural information". **IEEE International Symposium on Biomedical Imaging (ISBI)**, Barcelona, Spain, pp 118-121, 2012.

E. Niaf, R. Flamary, S. Canu, O. Rouvière, **C. Lartizien**, "Handling learning samples uncertainties in SVM classification: application to MRI-based prostate cancer computer-aided diagnosis". IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2012.

T. Glatard, A. Marion, H. Benoit-Cattin, S. Camarasu-Pop, P. Clarysse, R. Ferreira da Silva, G. Forestier, B. Gibaud, **C. Lartizien**, H. Liebgott, et al., "Multi-modality image simulation with the virtual imaging platform: Illustration on cardiac MRI and echography", IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, pp 98-101, 2012.

(iv) Année 2011

E. Niaf, R. Flamary, **C. Lartizien**, and S. Canu, "Handling uncertainties in SVM classification", *IEEE Workshop on Statistical Signal Processing*, pp. 757-760, 2011.

S. Marache, R. Prost, J. M. Rouet, and **C. Lartizien**, "Incorporating Patient-Specific Variability in the OncoPET_DB Database", IEEE Nuclear Science Symposium and Medical Imaging Conference, Valencia, Spain, pp. 3531-3533, 2011.

A. Marion, G. Forestier, H. Benoit-Cattin, S. Camarasu-Pop, P. Clarysse, R. Ferreira da Silva, B. Gibaud, T. Glatard, P. Hugonnard, **C. Lartizien**, et al., "Multi-modality medical image simulation of biological models with the Virtual Imaging Platform (VIP)", The 24th International Symposium on Computer-Based Medical Systems (IEEE CBMS 2011), Bristol, UK, pp. 1-6, 2011.

G. Forestier, A. Marion, H. Benoit-Cattin, P. Clarysse, D. Friboulet, T. Glatard, P. Hugonnard, **C. Lartizien**, H. Liebgott, J. Tabary, et al., "Sharing object models for multi-modality medical image simulation: a semantic approach", The 24th International Symposium on Computer-Based Medical Systems (IEEE CBMS 2011), Bristol, UK, pp. 1-6, 2011.

E. Niaf, O. Rouvière and **C. Lartizien**. *Computer-aided diagnosis for prostate cancer detection in the peripheral zone via multisequence MRI*. **Proceedings of the 2011 SPIE Medical Imaging Conference**, Vol. 7963, 2011.

O. Udobata, C. Muller and **C. Lartizien**. *Computer-aided tumor detection stemmed from the fuzzification of the Dempster-Shafer Theory*. **Proceedings of the 2011 SPIE Medical Imaging Conference**, Vol. 7963, 2011.

E. Niaf, O. Rouvière, F. Bratan, A. - L. Chesnais, F. Mège-Lechevallier, and **C. Lartizien**, "Diagnostic assisté par ordinateur pour la détection du cancer de la prostate par IRM multi-paramétrique", Journées Françaises de radiologie (JFR) 2011.

(v) Année 2010

S. Marache-Francisco, F. Lamare, H. Fayad, D. Visvikis, R. Prost, J.-M. Rouet, **C. Lartizien**. *Impact of respiratory motion correction on the detection of small lesions in whole-body PET imaging: a simulation study*. **Proceedings of the 2010 IEEE Nuclear Science Symposium and Medical Imaging Conference**, pp. 3531 – 3533, 2010.

(vi) Année 2009

J. Tabary, S. Marache-Francisco, S. Valette, W. P. Segars, **C. Lartizien**. *Realistic X-Ray CT Simulation of the XCAT Phantom with SINDBAD*. **Proceedings of the 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference**, pp 3980-83, 2009.

C. Lartizien, S. Marache, A. Le Pogam, S. Tomei and D. Visvikis. *Impact of Partial Volume Correction in Whole-Body PET Imaging : a Computer-Aided Detection Study*. **Proceedings of the 2009 IEEE Nuclear Science Symposium and Medical Imaging Conference**, pp 3214-17, 2009.

C. Lartizien, S. Tomei, S. Marache. *Improving performance of automated detection systems for 3D PET oncology imaging by use of resampled training images*. **International Journal of Computer Assisted Radiology and Surgery**, vol. 4, sup 1, pp S186-193, 2009.

C. Lartizien, L. Canaple, and O. Beuf. *Development and evaluation of a dedicated setup for co-registered PET/MRI abdominal imaging of the mouse*. In **Proceedings of the 17th Scientific Meeting of the International Society of Magnetic Resonance in Medicine, Hawai, USA**, pp 3092, April 2009.

(vii) **Année 2008**

S. Tomei, S. Marache-Francisco, C. Odet, **C. Lartizien**. *Automatic detection of active nodules in 3D PET oncology imaging using the Hotelling Observer and the Support Vector Machines: a comparison study*. **Proceedings of the 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference**, pp 5314-5319, Oct 2008.

S. Tomei, A. Reilhac, D. Visvikis, C. Odet, F. Giammarile, T. Mognetti, **C. Lartizien**. *Development of a database of realistic simulated whole body [18F]FDG PET images for lymphoma*. **Proceedings of the 2008 IEEE Nuclear Science Symposium and Medical Imaging Conference**, pp 4958-4963, Oct 2008.

(viii) **Année 2007**

C. Lartizien, S. Tomei, V. Maxim and C. Odet. *Combining a wavelet transform with a channelized Hotelling observer for tumor detection in 3D PET oncology imaging*. **Proceedings of the 2007 SPIE Medical Imaging Conference**, 6515 (19):1-10, San Diego, USA, Feb 2007.

L.M. Baboi, L. Milot, **C. Lartizien**, C. Roche, J.Y. Scoazec, F. Pilleul, O. Beuf. *Characterization of neuro-endocrine tumors in an athymic nude mouse model using dedicated synchronization strategies for T2-weighted MR imaging at 7T*, **Proceedings of the 29th IEEE EMBS Annual International Conference**, 1: 2879-82, Lyon, August 2007.

S. Tomei, C. Odet and **C. Lartizien**. *Détection d'hyper signaux en Tomographie par Emission de Positons 3D: approche statistique basée sur la combinaison d'une transformée en ondelettes discrète avec un test de Hotelling*. **Actes du XXIème colloque GRETSI**, Troyes, pp 869-872, Sept 2007.

(ix) **Année 2006**

L. Baboi, L. Milot, F. Pilleul, **C. Lartizien**, and O. Beuf. *Synchronization strategies in T2-weighted MR imaging for detection of mouse liver metastasis*. **Proceedings of the 14th Scientific Meeting of the International Society of Magnetic Resonance in Medicine**, Seattle, Washington, USA, May 2006.

(x) **Année 2005**

Khodaverdi, M.; Weber, S.; Streun, M.; Parl, Ch.; Larue, H.; Brandenburg, G.; Bauer, A.; Dehnhardt, M.; Auffray, E.; Boutemur, M.; Bruyndonckx, P.; Choi, Y.; D'Asseler, Y.; Devroede, O.; Dujardin, C.; Fedorov, A.; Heinrichs, U.; Janier, M.; Jung, J.H.; Korjik, M.; Krieguer, M.; LARGERON, G.; **Lartizien, C.**; Lecoq, P.; Lemaitre, C.; Leonard, S.; Loude, J.-F.; Morel, C.; Mosset, J.-B.; Pedrini, C.; Petrosyan, A.G.; Rey, M.; Sappey-Marinié, D.; Roldan, P.S.; Simon, L.; Song, T.Y.; Staelens, S.; Tavernier, S.; Trummer, J.; Van Holen, R.; Vieira, J.-M.; Wieers, E.; Wisniewska, M.; Wisniewski, D.; Wu, Y.; Pietrzyk, U.; Ziemons, K. *High resolution imaging with ClearPET TM Neuro - first animal images*, **Proceedings of the 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference**, 3 : 1641-1644, Oct 2005.

C. Lartizien et I. Buvat. *Comparison of 3D PET data bootstrap resampling methods for numerical observer studies*. **Proceedings of the 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Puerto-Rico, USA, 4 : 2138 – 2147, Oct 2005.

C. Lartizien, A. Goertzen, A. Reilhac, N. Costes, I. Magnin, G. Gimenez, A. C. Evans. *Validation of SORTEO Monte Carlo simulations for the geometry of the Concorde MicroPET R4 system*. **Proceedings of the 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Puerto-Rico, USA, 5 : 3036 - 3039, Oct 2005.

(xi) Année 2004

H. C. Gifford, P.E. Kinahan, **C. Lartizien**, M.A. King. *Evaluation of multiclass model observers in PET LROC studies*. **Proceedings of the 2004 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Portland, USA, 7 : 4068– 4071, 2004.

K. Ziemons, E. Auffray, P. Bruyndonckx, A. Fedorov, M. Korzhik, **C. Lartizien**, P. Lecoq, C. Morel, C. Pedrini, A. Petrosyan, U. Pietrzyk, D. Sappey-Marinier, M. Streun, and S. Tavernier. The ClearPET LSO/LuYAP Phoswich Camera: A high Performance Small Animal PET System. **Proceedings of the 2004 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Portland, USA, 4 : 2430– 2433, Oct 2004.

P.M. Cheng, P.E. Kinahan, C. Comtat, J.S. Kim, **C. Lartizien**, T.K. Lewellen. Effect of scan duration on lesion detectability in PET oncology imaging. **Proceedings of the 2004 IEEE International Symposium on Biomedical Imaging**, 2: 1432-1435, 2004.

(xii) Année 2003

P.E. Kinahan, **C. Lartizien**, J.S. Kim, C. Comtat, T.K. Lewellen. *Evaluating image quality for whole-body PET oncology imaging*. **Journal of Nuclear Medicine**, 44 (5): 254P, 2003.

J.S. Kim, P.E. Kinahan, **C. Lartizien**, C. Comtat, T.K. Lewellen. *Performance comparisons of planar and volumetric observers for lesion detection in PET scanning*. **Proceedings of the 2003 SPIE Medical Imaging Conference**, San Diego, 5034 : 89-99, 2003.

A. Reilhac, N. Costes, **C. Lartizien**, R. Gunn, A. Evans, A. Dagher. *Motion correction for PET ligand imaging*. Proceedings of the Human Brain Mapping Conference, New-York, 2003.

L. Janeiro, C. Comtat, **C. Lartizien**, P. Kinahan, M. Defrise, C. Michel, R. Trébossen. *A comparison between FORE+NECOSEM and NEC OSEM 3D for whole-body PET imaging*. **Proceedings of the 2003 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Portland. USA. 4: 2831-2835, 2003

Crystal Clear Collaboration. *The ClearPET LSO/LuYAP phoswich scanner : a high performance small animal PET scanner*. **Proceedings of the 2003 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Portland, USA, 3: 1728- 1732, 2003.

(xiii) Année 2002

C. Lartizien, A. Reilhac, N. Costes, M. Janier, D. Sappey-Marinier. *Monte Carlo simulation-based design study of a LSO-LuAP small animal PET system*. **Proceedings of the 2002 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Norfolk, 3 : 1547-1551, 2002.

P.E. Kinahan, **C. Lartizien**, S. Chandler, C.C. Meltzer, B. McCook, F. Torok. *Optimization of whole-body PET FDG oncology scanning protocols*. **Journal of Nuclear Medicine**, 43 : 216P, 2002.

D. Roumenov, M. Ricard, H. Valette, **C. Lartizien**, F. Dolle, J. Pinquier, O. Curet, P. George, R. Trebossen, M. Bottlaender. *Determination of the effective dose delivered to humans after an injection of befloxtone*. **Journal of Nuclear Medicine**, 43 (5): 357P, 2002.

A. Reilhac, **C. Lartizien**, N. Costes, S. Sans, C. Comtat, A. Evans. *Accounting for singles rate related phenomena in PET Monte Carlo-based simulations*. **Proceedings of the 2002 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Norfolk, 3 : 1617-1621, 2002.

P. E. Kinahan, J. S. Kim, **C. Lartizien**, C. Comtat, T.K. Lewellen. *A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging*. **Proceedings of the 2002 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Norfolk, 2 : 1267-1271, 2002.

L. Janeiro, C. Comtat, P. E. Kinahan, **C. Lartizien**, M. Defrise, C. Michel, R. Trébossen, P. Almeida. *NEC-scaling applied to FORE+OSEM*. **Proceedings of the 2002 IEEE Nuclear Science Symposium and Medical Imaging Conference**, Norfolk, 2 : 717-721, 2002.

(xiv) Année 2001

C. Lartizien, P.E. Kinahan, C. Comtat, R. Swensson, M. Lin, R. Trébossen. *Numerical and human observer comparisons of tumor detection for 2D versus 3D whole-body PET imaging protocols. Proceedings of the 2001 IEEE Nuclear Science Symposium and Medical Imaging Conference*, San Diego, 4 : 2243-2247, 2001.

D. Brasse, P.E. Kinahan, **C. Lartizien**, C. Comtat, M. Casey, C. Michel, T. Bruckbauer. *Correction methods for random coincidences in 3D whole-body PET imaging. Proceedings of the 2001 Nuclear Science Symposium and Medical Imaging Conference*, San Diego, 4 : 2080-2084, 2001.

P.E. Kinahan, C.C. Meltzer, B. McCook, S. Chandler, **C. Lartizien**, R. Swensson. *Image quality as a function of acquisition protocol for whole-body PET oncology imaging. Journal of Nuclear Medicine*, 42 : 201P, 2001.

R. Swensson, **C. Lartizien**, P. Kinahan. *Observer performance in finding and localizing simulated tumors within volumetric PET images. Proceedings of the 2001 Medical Image Perception Conference*, Warrenton, VA, 2001.

(xv) Année 2000

C. Lartizien, P.E. Kinahan, C. Comtat, M. Lin, R. Swensson, V. Villemagne, R. Trébossen, B. Bendriem. *Performing ROC studies with the same visualization software tools used in whole-body 3D PET oncology imaging. Journal of Nuclear Medicine*, 41 : 102P, 2000.

C. Lartizien, O. De Dreuille, D. Brasse, J. F. Gaillard, R. Trébossen. *NEC performances of 3 whole-body PET scanners in clinical imaging : A correlation between anthropomorphic phantom measurements and patient studies. Journal of Nuclear Medicine*, 41: 180P, 2000.

O. de Dreuille, **C. Lartizien**, H. Foehrenbach, P. Maszelin, D. Brasse, J.F. Gaillard. *Optimization of clinical PET-FDG acquisitions on the ADAC C-PET scanner: a study based on a whole-body phantom and patient data. Journal of Nuclear Medicine*, 41 : 176P, 2000.

C. Lartizien, P.E. Kinahan, C. Comtat, M. Lin, R. Swensson, R. Trébossen, B. Bendriem. *Evaluating image reconstruction methods for tumor detection performance in whole-body PET oncology imaging. Proceedings of the 2000 SPIE Medical Imaging Conference*, San Diego, 3981 : 49-59, 2000.

P.E. Kinahan, **C. Lartizien**, D. Brasse, C. Comtat, M. Defrise, C. Michel, D. Townsend, M. Lin, C. Bai. *Whole-body oncology imaging with 3D PET. Proceedings of the 2000 NIRS Conference*, Japan, 2000.

C. Lartizien, P.E. Kinahan, C. Comtat, R. Swensson, B. Bendriem, M. Lin, R. Trébossen. *A Tumor detection observer study comparing 2D versus 3D whole-body PET imaging protocols. Proceedings of the 2000 IEEE Nuclear Science Symposium and Medical Imaging Conference*, Lyon, 3 : 81-86, 2000.

N. C. Ferreira, R. Trébossen, **C. Lartizien**, V. Brulon, P. Merceron, B. Bendriem. *A hybrid scatter correction for 3D PET based on an estimation of the distribution of unscattered coincidences: Implementation on the ECAT EXACT HR+. Proceedings of the 2000 IEEE Nuclear Science Symposium and Medical Imaging Conference*, Lyon, 2 : 136-143, 2000.

(xvi) Année 1999

C. Lartizien, C. Comtat, R. Trébossen, N. Ferreira, M.-C. Grégoire, B. Bendriem. *Noise equivalent count measurements for 2D and 3D whole-body PET using an anthropomorphic phantom. Proceedings of the 1999 Nuclear Medicine Physics and Engineering conference*, Chicago, 1999.

C. Comtat, P.E. Kinahan, M. Defrise, C. Michel, **C. Lartizien**, D.W. Townsend. *Simulating whole-body PET scanning with rapid analytical methods. Proceedings of the 1999 IEEE Nuclear Science Symposium and Medical Imaging Conference*, Seattle, 3 : 1260-1264, 1999.

(xvii) Année 1998

A.M. Charvet, **C. Lartizien**, F. Estève et al. *Synchrotron radiation computed tomography applied to the brain: phantom studies at the ESRF medical beamline*. In Ando M, Uyama C, editors. **Medical Applications of Synchrotron Radiation**: Tokyo: Springer, pp. 95-98, 1998.

B.8.2.3 Communications à des congrès, symposium

C. Lartizien, N. Costes, M. Jeandey, A. Reilhac, M. Janier, D. Sappey-Marinié. *Simulated performances of the Lyon ClearPET prototype*. **Proceedings of the International Conference on Imaging Technologies in Biomedical Sciences**, Milos, 2003.

C. Lartizien, N. Costes, A. Reilhac, M. Janier, D. Sappey-Marinié. *Simulated imaging capabilities of the Lyon ClearPET scanner for realistic [18F]-FDG activity distributions in mice*. **Proceedings of the Academy of Molecular Imaging (AMI) Meeting**, Madrid, 2003.

D. Sappey-Marinié, **C. Lartizien**, N. Costes, R. Barbier, J. Marteau, C. Pedrini, C. Dujardin, P. Lecoq, S. Tavernier, M. Janier. *Développement instrumental d'un prototype microTEP pour l'étude du petit animal transgénique*. **Actes du 41^{ème} colloque de Médecine Nucléaire de Langue Française**, Tours, 2003.

C. Lartizien, C. Kuntner and A. Reilhac, *Validation of SORTEO Monte Carlo simulations for the geometry of the microPET Focus 220 system*, **Proceedings of the 1st European Conference on Molecular Imaging Technology EuroMedIm**, Marseille, France, May 2006.

L. Baboi, L. Milot, **C. Lartizien**, C. Roche, J.Y. Scoazec, F. Pilleul et O. Beuf. *Stratégies de synchronisation en imagerie RMN pondérée T2 pour la détection des lésions hépatiques chez la souris*. **Réunion de bilan du programme interdisciplinaire 2001-2004 du CNRS 'Imagerie du petit animal'**, pp 54-55, Marseille, Oct 2006.

L. Baboi, L. Milot, **C. Lartizien**, C. Roche, J.Y. Scoazec, F. Pilleul et O. Beuf. *Caractérisation de métastases hépatiques sur un modèle de tumeurs neuro-endocrines chez la souris*. **Réunion de bilan du programme interdisciplinaire 2001-2004 du CNRS 'Imagerie du petit animal'**, pp 56-57, Marseille, Oct 2006.

L. Baboi, **C. Lartizien**, L. Milot, C. Roche, J.Y. Scoazec, F. Pilleul et O. Beuf. *Imagerie multimodale de métastases hépatiques sur un modèle de tumeurs neuroendocrines chez la souris*. **Réunion de bilan du programme interdisciplinaire 2001-2004 du CNRS 'Imagerie du petit animal'**, pp 58-59, Marseille, Oct 2006.

E. Niaf, O. Rouvière, and **C. Lartizien**, "Aide au diagnostic du cancer de la prostate par IRM multi-sequences: une approche par classification supervisée", RITS - Recherche en Imagerie et Technologies pour la Santé, Rennes, France, 2011.

E. Niaf, O. Rouvière, and **C. Lartizien**, "Computer-Aided Diagnosis system for assisting radiologists in prostate cancer detection with Multiparametric MRI", Focal Therapy and Imaging in Prostate and Kidney Cancer, Noordwijk, Netherlands, 2011.

B.8.2.4 Séminaires invités,

C. Lartizien. Imagerie multimodale dans le diagnostic clinique: enjeux et challenges. Colloque de l'IFSBM, Villejuif, Novembre 2014.

C. Lartizien. Classification supervisée pour l'aide au diagnostic en imagerie multimodale. Colloque Imagerie Bio-Santé organisé par la mission interdisciplinaire du CNRS, Juillet 2014. Paris.

C. Lartizien. Simulation réaliste d'images pour l'aide à l'évaluation en imagerie médicale. Groupe de travail national "[Recherche en imagerie cardiaque](#)" du GDR STIC-SANTE, Juillet 2012.

C. Lartzien. Système d'aide à la détection en imagerie du cancer. Institut Elie Cartan de Nancy (IECN). Octobre 2009.

C. Lartzien. Détection en imagerie TEP pour l'oncologie. Service de Biostatistique du CH Lyon Sud. Mars 2009.

C. Lartzien. *Utilisation de la TEP en oncologie clinique: spécificités et évolutions méthodologiques.* Congrès annuel de la Société Française de Physique Médicale, Grenoble, juin 2002

C. Lartzien. *Simulation of the ClearPET prototype.* 1^{ère} journée scientifique d'imagerie du petit animal, Institut des Sciences Cognitives, Lyon, avril 2002.

C. Lartzien. *Optimisation des protocoles d'imagerie corps entier en tomographie par émission de positons pour la détection des hyperfixations du [18F]-FDG en oncologie.* Institut de Physique de Hautes Energies, Université de Lausanne, mai 2001.

C. Lartzien. *Application de la tomographie par émission de positons en oncologie.* Institut Gustave Roussy (IGR), Paris, décembre 2000.

C. Lartzien. *Optimisation des protocoles d'acquisition et de reconstruction des données en oncologie TEP corps entier.* Centre d'Exploration et de Recherche Médicale par Emission de Positons (CERMEP), Lyon, mai 2000.

C. Lartzien. *Assessing image quality in whole-body PET oncology.* Département de Radiologie-Université de Pittsburgh, PA, mars 1999.

B.8.2.5 Enseignement sur invitation

C. Lartzien. Evaluation des protocoles d'imagerie TEP: quelques outils de l'analyse psychophysique. Enseignement post-universitaire sur la « TEP : aspects physiques et quantitatifs », Paris, juin 2014.

C. Lartzien. Principes physiques de la TEP/IRM. Formation nationale sur la « combinaison TEP-IRM, conception, réalisation et analyse d'expérimentations en neuroimagerie » organisée dans le cadre des JIRFNI (Journées Inter-Régionales de Formation en Neuro-Imagerie), Lyon, juin 2013

C. Lartzien. Principes physiques de la TEP. Formation nationale sur la « TEP conception, réalisation et analyse d'une expérimentation de neuroimagerie » organisée dans le cadre des JIRFNI, Lyon, nov 2010

C. Lartzien. Optimisation des protocoles d'imagerie TEP corps entier pour la détection. Enseignement post-universitaire sur les « aspects physiques et pratiques de la TEP », Rennes, mars 2001.

B.8.2.6 Workshops

C. Lartzien. Evaluation de l'impact du mouvement respiratoire sur le diagnostic en imagerie TEP oncologique. Groupe de travail national "[Quantification en Imagerie Moléculaire](#)" du GDR STIC-SANTE, Mai 2012.

C. Lartzien. Détection automatique de fixations radioactives dans les images TEP oncologiques. Groupe de travail national "[Quantification en Imagerie Moléculaire](#)" du GDR STIC-SANTE, Février 2010.

C. Lartzien. Comparaison de différentes approches bootstrap pour la génération de données répliquées en TEP, Groupe de travail national "[Quantification en Imagerie Moléculaire](#)" du GDR STIC-SANTE, Jan 2006

C. Lartzien. *Developments in nuclear imaging.* Eumorphia Annual Meeting, Londres, octobre 2003.

C. Lartzien. *Developments in small animal oncology imaging.* Eumorphia Annual Meeting, Londres, octobre 2003.

J. Lumbroso, **C. Lartizien**, R. Trébossen. *3D whole-body PET imaging: experience of the Orsay PET facility*. Siemens/CTI ECAT Users Meeting, Barcelone, aout 2000.

C. Lartizien, C. Comtat, R. Trébossen, et al. *Noise equivalent count measurements for 2D and 3D whole-body PET using an anthropomorphic phantom and patient data*. Siemens/CTI ECAT Users Meeting, Amsterdam, avril 1999.

B.8.3 Quelques papiers

B.8.3.1 JNM 2002

B.8.3.2 JNM 2003

B.8.3.3 JNM 2004

B.8.3.4 AR 2004

B.8.3.5 JBHI 2014

B.8.3.6 TNS 2012

