

Rapport bibliographique

Correction du mouvement respiratoire en TEP

Simon Marache-Francisco

Laboratoire CREATIS/LRMN - Philips Medisys

17 octobre 2011



A destination de :

- Jean-Michel Rouet, Philips Medisys
- Carole Lartizien, CREATIS/LRMN
- Rémy Prost, CREATIS/LRMN

Table des matières

I	Introduction	9
II	Imagerie TEP	11
1	Principe Physique	15
1.1	Généralités	15
1.1.1	Détecteur	15
1.2	perturbation trajet du photon	16
1.2.1	Diffusion	16
1.2.2	Déviation	16
1.2.3	Absorbtion	16
2	Déroulement d'une acquisition	17
2.1	2D / 3D	17
2.2	Format des données	17
2.2.1	List-mode	17
2.2.2	Sinogramme	18
3	Algorithmes de reconstruction	19
3.1	Itératifs	19
3.1.1	EM	19
3.1.2	OSEM	19
3.2	Analytiques	19

III Mouvement respiratoire	21
4 Respiration et influence sur les acquisitions TEP/TDM	23
4.1 Mouvement respiratoire	23
4.2 Localisation et Volume	24
4.3 Mesure de l'activité des tumeurs	25
4.3.1 Décalage de la carte d'atténuation	25
4.3.2 Déplacement de la tumeur au cours du cycle	26
4.4 Impact du mouvement respiratoire sur la detection	27
4.4.1 Thielemans 2006	27
4.4.2 Chang 2010	28
5 Processus d'estimation du mouvement	29
5.1 Capteurs externes	29
5.1.1 Spiromètre	29
5.1.2 Ceinture	29
Todo : donner ref de raison utilisation	29
5.1.3 Basés caméras	30
5.1.4 Techniques basées sur les images TEP	30
5.2 Estimation du champ de mouvement	31
Todo : traduire	31
5.2.1 Image TDM 4D	31
5.2.2 Image TEP 4D	32
5.3 Modèle	32
6 Correction du mouvement respiratoire	33
6.1 Synchronisation respiratoire	33
6.2 Synchronisation respiratoire avec recalage	35
6.2.1 Estimation du mouvement respiratoire corps entier	35
Todo : continuer	35
6.3 Correction pré-reconstruction	36

6.4	Correction pendant la reconstruction	37
6.5	Déconvolution de l'image	38
IV	Evaluation des performances de diagnostique	41
7	Performance des outils de détection	43
7.1	Généralités	43
7.2	Méthodologie ROC - Receiver-Operating Curve	44
7.2.1	Courbes Localization ROC (L-ROC)	46
7.3	F-ROC	47
7.3.1	Courbes Free-ROC	47
7.3.2	Courbes Alternative Free-ROC	48
7.3.3	Comparaison des courbes	48
8	Systèmes de détection	51
8.1	Les CAD en TEP	51
8.2	Types de classification	51
8.2.1	supervisée - méthodologie	51
8.2.2	non supervisée - méthodologie	52
8.3	classifieurs	53
8.3.1	SVM (Separateur à Vaste Marge)	53
8.3.2	LDA	53
8.4	Systèmes humain	53
V	Simulation et base de données (MIC 2011)	55
9	Simulations	57
9.1	principe des simulations	57
9.1.1	monte carlo	57
9.1.2	analytiques	57
9.1.3	MC accélérés	57

9.2	simulateurs disponibles	57
9.3	processus de simulation avec SORTEO	57
9.4	Contribution à SORTEO	57
10	Base de donnée	59
10.1	Présentation	59
10.2	Modèles	59
10.3	données clef	59
VI	Résultats	61
11	Méthodes	63
11.1	Système CAD	63
11.1.1	Vecteur de Caractéristiques	64
11.1.2	Génération de la base d'apprentissage	65
11.1.3	Apprentissage de la Machine à Vecteur de Support (SVM)	65
11.1.4	Génération des sites présumés	65
11.1.5	Évaluation du résultat	66
11.1.6	Résumé	66
11.2	Optimisation des paramètres du système CAD	67
11.2.1	Paramètres à optimiser	67
11.2.2	Méthodes d'optimisation	69
12	Analyse des résultats	73
12.1	Optimisation des paramètres	73
12.1.1	Courbe Free-ROC	78
12.1.2	Comparaison des performances JAFROC	78
12.2	Comparaison des performances des différentes méthodes sur le Poumon . .	80
12.2.1	Courbes Free-ROC	81
12.2.2	Comparaison des performances JAFROC	82
12.3	Comparaison des performances des différentes méthodes Foie	86

12.3.1 Comparaison des performances JAFROC	86
12.3.2 Courbes Free-ROC	86
VII Conclusion	91
VIII Bibliographie	93

Première partie

Introduction

Deuxième partie

Imagerie TEP

La Tomographie par Émission de Positons (TEP) est une modalité d'imagerie fonctionnelle utilisant la désintégration d'un traceur radioactif pour mettre en valeur les zones de forte activité métaboliques. Elle est principalement utilisée en imagerie cérébrale, oncologie et cardiologie.

Chapitre 1

Principe Physique

1.1 Généralités

L'imagerie TEP permet de visualiser de manière indirecte les désintégrations de particules survenant dans le corps du patient. Pour cela, on inject un “traceur” contenant une particule radioactive dans le corps du patient. Ce traceur est conçu de manière à se fixer sur les zones du corps que l'on souhaite imager. Pendant toute la durée de l'examen, les particules radioactives vont se désintégrer selon la loi de décroissance radioactive de la formule 1.1.

$$dN = -\lambda N dt \quad (1.1)$$

N représente le nombre de particules radioactives présentes dans le corps du patient. dN représente la variation de ce nombre de particules (le nombre de désintégrations par dt) et λ est une constante dépendant de l'élément radioactif.

Chaque désintégration d'un élément radioactif va déclencher l'émission d'une particule β , aussi appellée positon. En oncologie, on utilise le Fluor ^{18}F qui se désintègre en Oxygène ^{18}O en émettant le positon. Cette particule va parcourir quelques mm avant de s'annihiler avec un électron en émettant 2 photons dans deux directions opposées avec une énergie de 511 KeV.

Ce seront ces photons qui vont être détectés par l'imageur TEP pour reconstituer la position de la désintégration initiale.

1.1.1 Détecteur

Les détecteurs utilisés en TEP sont constitués d'un matériau photomultiplicateur placé devant un capteur. Chaque photon va entraîner un qui va déclencher une émission lumi-

neuse à chaque photon détecté, placé à côté suivi par un détecteur qui va convertir cette émission lumineuse en impulsion électrique.

1.2 perturbation trajet du photon

1.2.1 Diffusion

1.2.2 Déviation

1.2.3 Absorption

Chapitre 2

Déroulement d'une acquisition

2.1 2D / 3D

2.2 Format des données

Les données acquises par une caméra TEP peuvent être stockées sous deux formes principales : Sinogramme et list-mode.

2.2.1 List-mode

Ce format correspond à un enregistrement “brut” des données issues de l’électronique de la caméra.

Ce format de fichier est en fait un enregistrement séquentiel des évènements, dans leur ordre de détection. On peut enregistrer chaque détection indépendamment, ou encore uniquement les coïncidences. Les évènements sont datés, ce qui permet de conserver l’informations temporelles.

Il existe plusieurs formats de fichiers pour le stockage de ces données, notamment le format LMF (List-Mode Format) développé pour le projet ClearPET et le format ROOT développé par le CERN.

L’avantage de ces formats est qu’ils permettent de conserver les informations sur la dynamique de l’acquisition, mais aussi qu’ils permettent le stockage de métadonnées utiles en simulations, notamment le nombre de diffusions, ou de marquer les coincidences fortuites.

2.2.2 Sinogramme

Le sinogramme est une image

Chapitre 3

Algorithmes de reconstruction

3.1 Itératifs

3.1.1 EM

3.1.2 OSEM

3.2 Analytiques

Troisième partie

Mouvement respiratoire

Chapitre 4

Respiration et influence sur les acquisitions TEP/TDM

Le mouvement respiratoire en imagerie TEP engendre plusieurs effets sur les images reconstruites, qui seront détaillés ci-après. Ils occasionnent notamment une diminution de la qualité des images, ce qui peut perturber le travail des praticiens.

4.1 Mouvement respiratoire

Ce mouvement est la succession d'une phase inspiratoire, suivi d'une phase expiratoire. Chacune de ces phases combine plusieurs mouvements élémentaires [servant laval, 2007] :

1. thoracique, avec un déplacement des côtes
2. abdominal, avec un déplacement du diaphragme
3. en cas d'inspiration forcée, action des pectoraux

La variabilité inter et intra-patient de ce mouvement est très importante : le volume d'air inspiré peut varier de 500mL à 1200mL selon que la personne a une respiration normale ou profonde. Pour ces deux extrêmes, la fréquence respiratoire varie de 5 cycles/min. à 20 cycles/min. [Sherwood, 2006].

Or une acquisition TEP a une durée de plusieurs minutes par lit, ce qui amène à la reconstruction d'une image dégradée, notamment au niveau de la localisation de la tumeur, de son activité mesurée et, par extension, de sa détection. De la même manière, des artefacts apparaissent au niveau des zones de forts mouvement dans les images TDM, lorsque les organes bougent pendant une rotation du capteur.

Je vais tout d'abord introduire les effets visibles sur les images, puis je m'attarderai sur les mesures quantitatives utilisées pour mesurer l'apport de la correction du mouve-

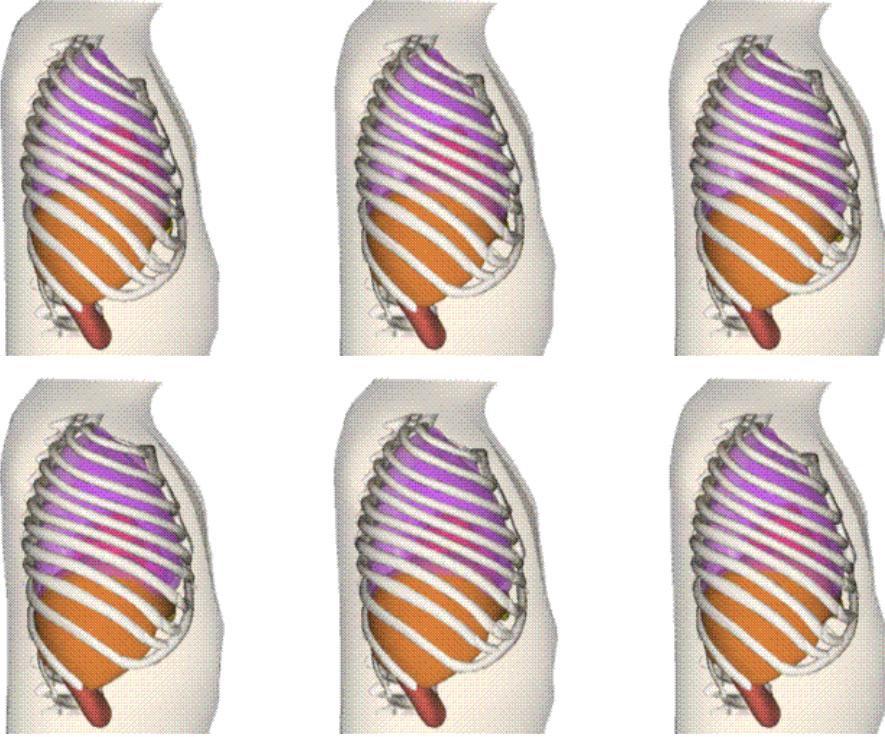


FIG. 4.1: Modélisation de la respiration par le fantôme XCAT.

ment sur les tumeurs. Ensuite, je détaillerais les publications utilisant des critères pouvant être assimilés à de la détection.

4.2 Localisation et Volume

$$\%PRD = \left| \frac{Image\ non\ corrigée - Image\ de\ référence}{Image\ de\ référence} \right| \quad (4.1)$$

La localisation et le volume des tumeurs peuvent être modifiés par le mouvement engendré par la respiration (voir figure 4.2). D'après [Lamare et al., 2007a], réalisé sur des simulations monte-Carlo à l'aide du logiciel de simulation Geant4[Jan et al., 2004] en utilisant le modèle NCAT[Segars, 2001], la largeur à mi-hauteur des lésions peut être modifiée de 48% (équation 4.1 : Percent relative difference) dans le cas d'une lésion de 7mm de diamètre dans la partie basse du poumon. L'imprécision axiale sur le positionnement de la tumeur peut atteindre 9% dans les mêmes conditions.

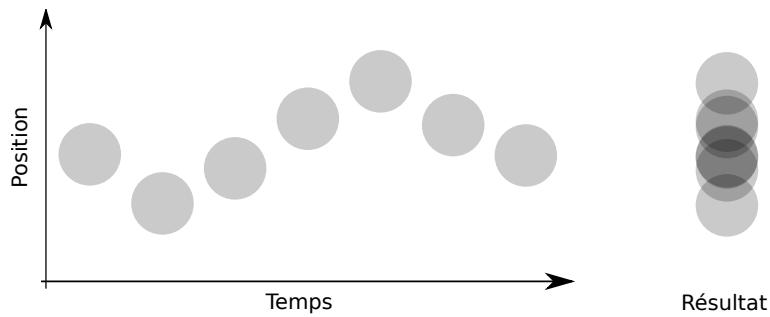


FIG. 4.2: *Effet du déplacement d'une tumeur sur les données acquises. La position de la tumeur change en fonction du temps, ce qui provoque l'acquisition d'une tumeur équivalente présentée à droite.*

4.3 Mesure de l'activité des tumeurs

Le contraste des tumeurs par rapport au fond est un critère important pour déterminer la malignité des tumeurs [Dimitrakopoulou-Strauss et al., 2002] [Krak et al., 2005]. Mais il va aussi directement déterminer si la tumeur est détectable ou non.

L'article de [Chang et al., 2010] est une évaluation clinique réalisée sur 13 patients (21 tumeurs au total).

L'activité peut être influencée par la respiration de deux manières : par un mauvais ajustement de la carte d'atténuation, et par le moyennage de la position de la tumeur.

4.3.1 Décalage de la carte d'atténuation

La carte d'atténuation utilisée pour corriger les images reconstruites est basée sur une image TDM prise à un instant donné du cycle. Or l'atténuation de la zone correspondant à une tumeur peut être différente de celle du tissus environnant.

$$\% \text{ variation } SUV_{max} = 100 \times \frac{|SUV_1 - SUV_2|}{(SUV_1 + SUV_2)/2} \quad (4.2)$$

D'après l'étude [Erdi et al., 2004] réalisée sur cinq patients, on peut observer des variations de SUV_{max} (voir eq. 4.2) allant jusqu'à 24% selon que la correction d'atténuation est réalisée à partir d'une image TDM en fin d'expiration ou d'inspiration. Sur l'ensemble du cycle, la variation peut atteindre 30%.

Artefacts dûs à l'utilisation de l'image TDM

On peut voir sur les images de la figure 4.3 des artefacts présents sur les images TDM utilisées pour la correction d'atténuation. Ces artefacts proviennent de la manière dont les images sont acquises : la caméra tourne autour du sujet dans un mouvement hélicoïdal, et l'algorithme de reconstruction va ensuite utiliser les acquisitions pour reconstruire une image complète. Or en cas de respiration rapide, des incohérences peuvent survenir quand le mouvement du diaphragme est tellement rapide qu'il va plus vite que la caméra. Ce type d'artefacts va créer des incohérences dans les images TEP reconstruites.



FIG. 4.3: Artefacts présents sur des images TDM utilisées pour la correction d'atténuation

4.3.2 Déplacement de la tumeur au cours du cycle

Le mouvement respiratoire va avoir pour effet de déplacer la tumeur pendant l'acquisition, ce qui va moyenner la quantité de radioactivité sur l'ensemble du cycle. Si le déplacement de la tumeur est suffisamment grand par rapport à son diamètre, la réduction de radioactivité va être importante.

L'étude [Boucher et al., 2004] montre que sur des fantomes, un déplacement d'une source radioactive de 6mm sur un cycle respiratoire moyen entraîne une sous-estimation

de l'activité maximale de la tumeur de 41% pour une lésion de 1.2mL et de 21% pour une sphère de 19.4mL.

4.4 Impact du mouvement respiratoire sur la détection

Peu de travaux ont été réalisés sur l'impact du mouvement respiratoire sur la détection des tumeurs. Globalement, les critères utilisés sont principalement des mesures orientées sur quantification des lesions (SUV_{max} , profils de lesions, ...). Très peu d'articles utilisent des critères orientés détection tels que les observateurs.

Voici une liste des critères utilisés dans différentes publications pour évaluer la performances d'algorithmes de correction du mouvement respiratoire :

1. SUV_{max} , contraste : [Chang et al., 2010] [Lamare et al., 2007b] [Nehmeh et al., 2002] [Detorie and Dahlbom, 2008]
2. Line profile : [Chang et al., 2010] [Thielemans et al., 2006] [Lamare et al., 2007b]
3. Volume et position de la lésion : [Chang et al., 2010] [Lamare et al., 2007b] [Nehmeh et al., 2002]
4. Rapport signal sur bruit (SNR) : [Chang et al., 2010]
5. Observateur de Hotelling (CHO) : [Thielemans et al., 2006]

Comme on peut le voir, les deux seuls critères qui pourrait s'approcher d'une étude sur la détection sont le SNR et le CHO, mais ils sont largement sous-représentés.

Je vais me concentrer sur les deux publications qui utilisent un observateur, et présenterai les résultats des autres publications dans la partie suivante.

Un autre document par rahmim arman [Rahmim et al.,] propose d'utiliser le CHO pour évaluer l'amélioration de la détection des défauts dans de l'imagerie cardiaque corrigée du mouvement respiratoire. Ce document n'a pas encore donné lieu à publication.

4.4.1 Thielemans 2006

Dans sa publication, Thielemans [Thielemans et al., 2006] utilise le CHO [Barrett et al., 1993] dans le cas où le signal (activité des lesions) et le fond sont connus (activité du poumon). Cet observateur est un classifieur linéaire, utilisé en conjonction avec des informations fréquentielles.

Cependant ils utilisent le CHO uniquement sur des lésions de fort diamètre et contraste (13mm de diamètre et contraste de 4.25 :1, sur des simulations analytiques). Les résultats présentés (fig. 4.4) montrent une amélioration du score pour les méthodes de correction de l'ordre de 50% dans certains cas.

Mais il est difficile d'évaluer de manière précise l'apport des méthodes de détection à l'aide de ces seuls "scores" car ce sont des résultats qualitatifs.

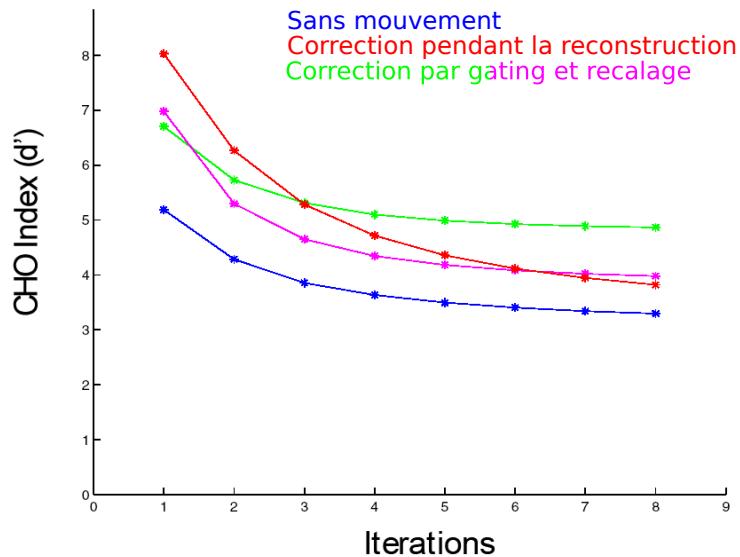


FIG. 4.4: Index de CHO pour différents techniques de correction du mouvement respiratoire en fonction du nombre d'itérations. la méthode utilisant le gating est comparée avec une méthode évolution présentée dans le document.

4.4.2 Chang 2010

Cette publication utilise entre autres le rapport signal sur bruit pour évaluer l'amélioration de la visibilité des lésions. Son intérêt est qu'elle est réalisée sur 13 patients et compare les performances sur 21 tumeurs au total.

Ils ont calculés le rapport signal sur bruit (SNR) en divisant le SUV moyen de la tumeur par l'écart-type d'une zone d'intérêt dans le poumon. On observe une amélioration moyenne de 26% du SNR sur les tumeurs, et pouvant atteindre 66%.

Il est intéressant de constater qu'il y a un point où le SNR diminue de 3.4%, mais que le SUV_{max} et le SUV_{moyen} augmentent tout de même de presque 18%. Cela semble indiquer une erreur car le SUV de la zone d'intérêt n'est pas censé changer de manière importante.

Chapitre 5

Processus d'estimation du mouvement

5.1 Capteurs externes

Ces capteurs fournissent un signal qui sera corrélé avec la respiration. Ils vont permettre par la suite de faire correspondre les données acquises avec une phase particulière du mouvement respiratoire.

5.1.1 Spiromètre

Le spiromètre est un capteur externe placé sur la bouche du patient et qui permet de mesurer les déplacements d'air dans le système respiratoire [Guivarc'h et al., 2004]. Les spiromètres mesurent un débit ou un volume d'air inspiré/expiré (voir illustration figure 5.1). A partir de l'une des grandeurs, il est possible d'estimer l'autre facilement. L'avantage du spiromètre est qu'il permet d'accéder à une mesure caractérisant directement la respiration du patient, et n'est pas sujet à des perturbations externes (mouvements involontaires par exemple). Par contre cela demande un appareillage qui peut être assez invasif pour le patient.

5.1.2 Ceinture

Pour mesurer le signal respiratoire, il est possible d'utiliser un capteur qui va mesurer le périmètre du thorax. L'extension de cette ceinture va correspondre aux mouvements de la cage thoracique et de l'abdomen pendant la respiration du patient. C'est une mesure indirecte de l'amplitude du mouvement respiratoire utilisée couramment en routine clinique ||A Faire : donner ref de raison utilisation|| .

Différentes technologies existent pour mesurer cette information (RespiTrace R250 de Studley Data Systems, Respiratory Belt Transducer de ADInstruments, ...). Elles sont

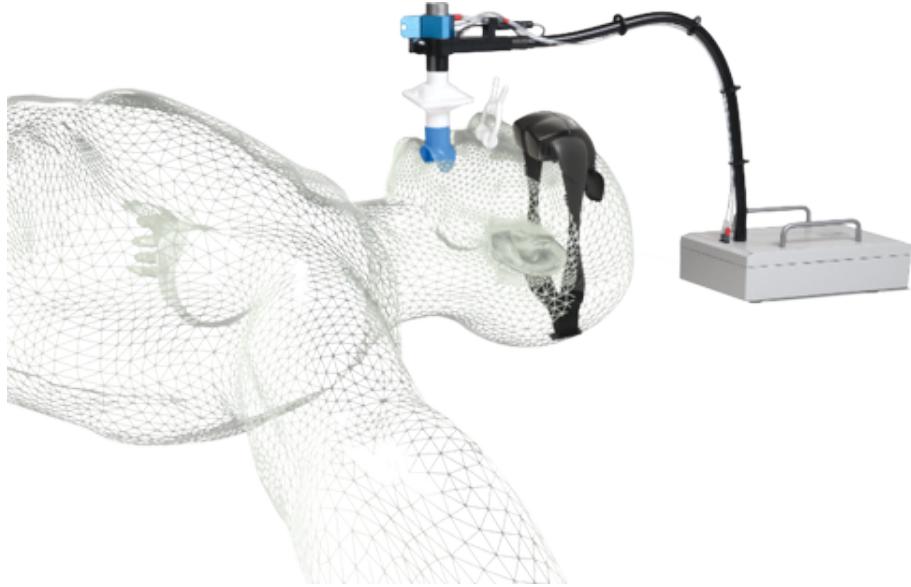


FIG. 5.1: spiromètre *Syn'r* : on peut voir le système de mesure de la respiration ainsi qu'un système de moniteurs implantés dans les lunettes pour aider le patient à contrôler sa respiration

basées sur plusieurs effets (résistif, inductif...) et ont l'avantage d'avoir un faible coût et de ne pas perturber le patient.

5.1.3 Basés caméras

Des caméras peuvent être utilisées pour estimer le mouvement respiratoire. Une des techniques consiste à utiliser des informations surfaciques en reconstruisant en 3D certaines parties du corps à l'aide de plusieurs caméras (avec ou sans marqueurs) ou de caméra temps de vol. Cela permet d'avoir plus d'informations sur la respiration.

Une autre technique consiste à installer un marqueur sur le corps du patient et de relever les déplacements de ce marqueur à l'aide d'une caméra. Un tel système est décrit dans [Nehmeh et al., 2002] : Respiratory Gating System de Varian Medical Systems (voir figure 5.2).

Ces techniques ont l'avantage d'être moins invasives et plus facilement acceptées par le patient. Cependant, elles sont beaucoup plus sensibles aux mouvements parasites.

5.1.4 Techniques basées sur les images TEP

Une publication utilise des images TEP pour déduire le signal respiratoire [Bundschuh et al., 2007] : les images TEP sont reconstruites par pas temporel de 0.5s. La position axiale du ba-

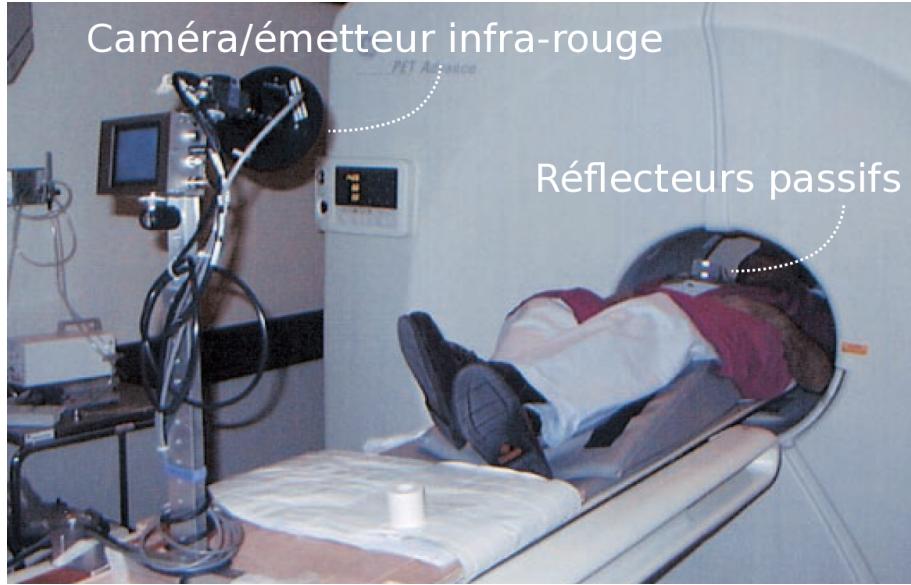


FIG. 5.2: Photographie du système RGS de Varian medical Systems en action : une caméra va détecter le déplacement d'une zone du thorax en mesurant le déplacement de marqueurs placés sur un bloc plastique.

rycentre de chaque image donne une estimation du signal respiratoire. Cette technique donne les meilleurs résultats pour une zone d'intérêt centrée sur une tumeur de forte intensité.

5.2 Estimation du champ de mouvement

Le signal respiratoire acquis par les méthodes précédemment citées est utilisé pour décomposer les données acquises en TDM ou TEP en plusieurs phases, chacune correspondant à un instant du cycle. Ces informations sont utilisées pour assembler les données acquises en **||A Faire : traduire|| “bins”**, reconstruits indépendamment. Ces reconstructions vont être utilisées pour estimer les champ de mouvements à l'aide de techniques de recalage.

5.2.1 Image TDM 4D

Les images TDM peuvent être acquises en mode dynamique de manière à obtenir un ensemble d'image couvrant tout le cycle respiratoire [Lamare et al., 2007b, Qiao et al., 2006]. Les données des

Des algorithmes de recalage sont utilisés de manière à déduire le champ de mouvement. Bien que les images soient de bonne qualité et permettent une estimation précise du

champ de mouvement, cela demande une exposition supplémentaire aux rayonnements X.

Plusieurs publications basées sur des simulations se servent des cartes de labels utilisées pour la simulation pour réaliser les estimation de mouvements. Cela donne une estimation dans le “meilleur des cas”, où l'image TDM est parfaitement en phase avec les images TEP.

5.2.2 Image TEP 4D

Les données regroupées selon l'instant du cycle auquel ils appartiennent [Dawood et al., 2008, Dawood et al., 2006]. Les images sont reconstruites indépendamment sans correction d'atténuation, puis un algorithme de recalage est utilisé pour estimer le champ de mouvement. L'avantage de cette technique est de ne pas nécessiter d'irradiation ni de temps supplémentaire. Cependant, les images reconstruites sont de mauvaises qualité et peuvent réduire la précision du champ de mouvement estimé.

5.3 Modèle

Une autre voie est en cours de développement basée sur la création d'un modèle de respiration généralisé adapté à chaque patient à partir de données réduites. Fayad [Fayad et al., 2010] propose une méthode basée sur l'analyse en composantes principales pour modéliser les mouvements. Ce modèle est ensuite adapté à un patient à partir de deux images TDM prises à des instants différents du cycle. Enfin, une caméra 3D permettant d'obtenir la surface du corps du patient est utilisé pour synchroniser le mouvement respiratoire et affiner le modèle.

L'avantage de ce modèle est qu'il est totalement continu, et permet l'extraction d'un nombre arbitraire de phases.

Chapitre 6

Correction du mouvement respiratoire

6.1 Synchronisation respiratoire

La synchronisation respiratoire correspond à un découpage du cycle respiratoire en plusieurs phases (voir Fig.6.1). Une seule d'entre elles sera utilisée pour la reconstruction. En théorie cela permet d'avoir le meilleur résultat, car il est possible de sélectionner les évènements correspondants à la phase où a été acquise la carte d'atténuation.

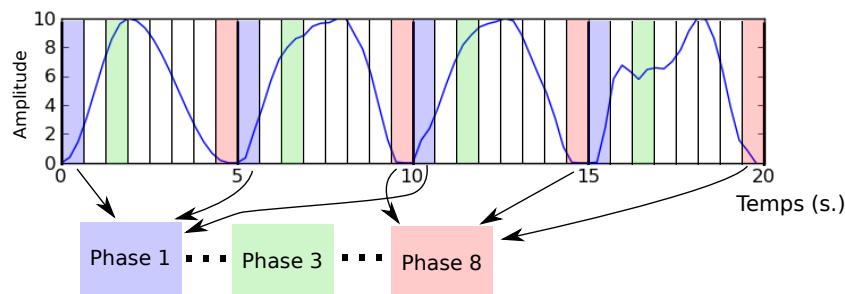


FIG. 6.1: Illustration de la synchronisation respiratoire : Le cycle est découpé en plusieurs phases

Cette technique est notamment présentée dans [Nehmeh et al., 2002], où le signal respiratoire est estimé par une caméra qui suit un marqueur placé sur le torse du patient. L'auteur annonce une réduction du volume des tumeurs pouvant aller jusqu'à 34%, avec une augmentation du SUV_{max} de 160%.

Une autre publication [Boucher et al., 2004] utilise un thermomètre détectant l'air chaud émis en début de cycle respiratoire pour réaliser la synchronisation. Les différentes reconstructions issues de l'expérience sont visibles figure 6.2.

Une variante de cette technique ne nécessitant pas de capteur est décrite dans [Nehmeh et al., 2003]. Un point faiblement radioactif est fixé au-dessus du torse du patient. Les acquisitions de

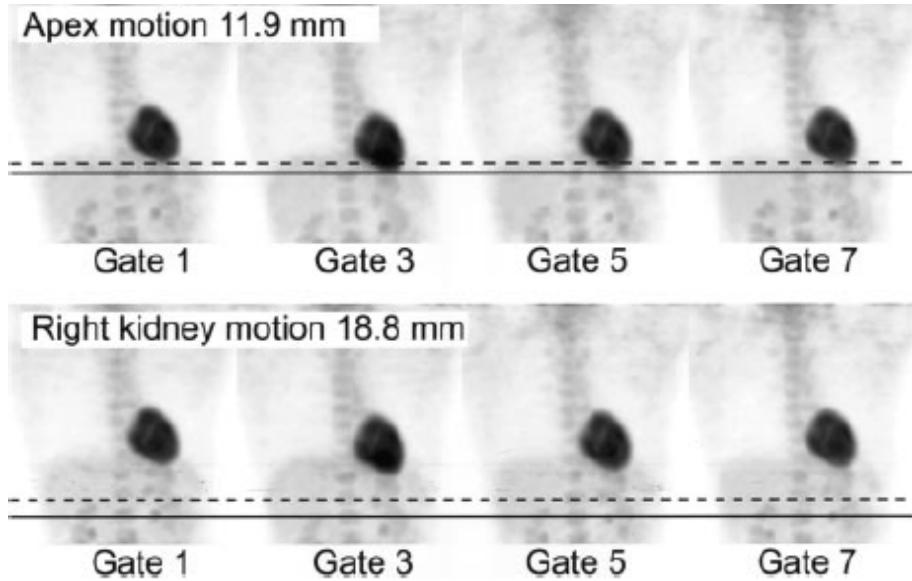


FIG. 6.2: Illustration de l'étendue du mouvement respiratoire sur des images reconstruites après synchronisation respiratoire [Boucher et al., 2004]. La rangée du haut montre l'étendue du mouvement de l'apex du cœur, et celle du bas l'étendue du mouvement du rein

l'imageur sont ensuite enregistrées par blocs temporels de 1 seconde, et une zone d'intérêt est reconstruite dans chacune des images. Les données où le point source est dans cette zone d'intérêt sont sommées et l'image finale reconstruite.

Guoping et al. [Chang et al., 2010] réalisent la carte d'atténuation à partir d'une image TDM réalisée en respiration libre, et reconstruisent les données TEP acquises lorsque l'amplitude respiratoire est proche de celle utilisée pour l'acquisition TDM (voir exemples figure 6.3. Les résultats présentés sur 13 patients (21 tumeurs) montrent une amélioration du rapport signal sur bruit pouvant aller de -3.4 à 81% suivant les tumeurs, avec une amélioration moyenne de 26.3%.

Le principal problème de ces techniques est qu'elles demandent un temps d'acquisition beaucoup plus long à qualité d'image égale. Si l'on ne conserve que 20% des événements détectés, cela signifie qu'il faut augmenter le temps d'acquisition d'un facteur 5 pour obtenir une image d'une qualité égale. Il n'est donc pas envisageable de mettre en place ces protocoles en routine clinique, car le temps disponible n'est pas suffisant. C'est pour cela que de nombreuses équipes se sont mises à travailler sur une évolution de cette technique, où les images sont déformées les unes sur les autres pour prendre en compte toutes les informations de l'acquisition.j jj

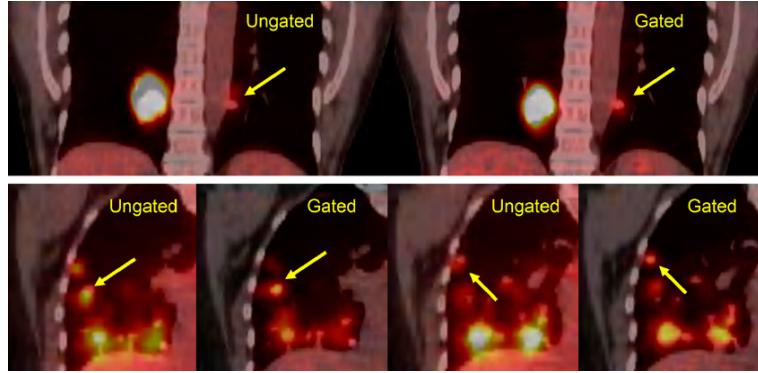


FIG. 6.3: *Images TEP/TDM superposées du poumon reconstruites avec et sans gating respiratoire en utilisant la méthode décrite dans [Chang et al., 2010]. On peut observer que les tumeurs sont mieux définies et correspondent à l'image TDM qui sert de référence.*

6.2 Synchronisation respiratoire avec recalage

Pour réaliser cela, les différentes techniques se basent sur une estimation préalable du mouvement respiratoire. Les images de chaque phase sont reconstruites indépendamment, puis recalées sur une phase de référence grâce au champ de mouvement. Enfin, les images déformées sont sommées. La difficulté se situe dans l'estimation du champ de mouvement interne lors de la respiration, car ce mouvement est complexe.

Les premières publications décrivant cette technique l'utilisaient notamment pour réaliser de l'imagerie cardiaque en TEP [Klein and Huesman, 1997]. Cette publication démontre la faisabilité du procédé sur un animal en utilisant des techniques de flux optique pour estimer le champs de mouvement. En effet le coeur a l'avantage d'avoir une activité métabolique intense, ce qui rend l'estimation de son mouvement aisée même sur des images avec une faible statistique.

6.2.1 Estimation du mouvement respiratoire corps entier

Le mouvement respiratoire corps entier est complexe et nécessite des modèles élastiques
 ||A Faire : continuer|| . cependant, cette complexité nécessite de réaliser l'estimation sur des images 3D du patient, et donc d'estimer les mouvements internes. Pour cela, on peut utiliser les deux modélités disponibles lors d'un examen TEP/CT :

imagerie TEP avec gating

L'acquisition TEP est réalisée en conservant les informations temporelles associées. Les premiers algorithmes étaient utilisés en imagerie cardiaque [Klein et al., 2001] avec

des transformations simples (affines), puis d'autres algorithmes plus adaptés aux images corps entier ont été utilisées, comme les flux optiques [Dawood et al., 2006], ou l'interpolation par B-spline [Bai and Brady, 2009].

imagerie CT 4D

Les images CT 4D peuvent être utilisées pour réaliser l'estimation du mouvement respiratoire. Cela nécessite par contre une dose plus importante et un temps d'acquisition plus long. Dawood a réalisé plusieurs publications sur le sujet en utilisant le flux optique pour l'estimation du champ de mouvement [Dawood et al., 2006, Dawood et al., 2008]. L'algorithme a été étudié sur des images de patients réels. Une autre publication [Thorndyke et al., 2006] indique une amélioration du rapport de contraste sur bruit (CNR) d'un facteur 3 grâce à la correction.

6.3 Correction pré-reconstruction

Les méthodes de correction du mouvement pré-reconstruction modifient les positions des Lignes de réponse (LDR) fournies par le scanner. Ce recalage des LDR correspond à un déplacement des lignes de réponse dans l'espace du détecteur (voir fig. 6.4) en fonction du mouvement respiratoire. La limitation principale de ce type de méthode est que le champ de mouvement ne peut pas être élastique.

Cependant, il a été étudié en imagerie du cerveau [Bloomfield et al., 2003], où il permettait de corriger les mouvements de la tête. Il a été aussi utilisé en imagerie cardiaque TEP [Livieratos et al., 2005] en utilisant un champ de mouvement rigide (rotation suivie d'une translation).

Dans les deux cas, les résultats ont montré une nette amélioration des images (voir fig. 6.5)

Dans le cadre du mouvement respiratoire du thorax, l'approche de recalage par LDR a été expérimentée par Frederic Lamare [Lamare et al., 2007a], mais avec des résultats mitigés.

Cette technique de correction du mouvement a été utilisée pour la correction du mouvement respiratoire du thorax [Lamare et al., 2007a, Lamare et al., 2007b], avec des performances plus limitées. En effet, le champ était approximé par une transformation affine, qui peut difficilement modéliser le mouvement du thorax dans son ensemble.

Ces résultats ont été améliorés par l'utilisation de la technique suivante qui permettait la prise en compte d'un mouvement élastique.

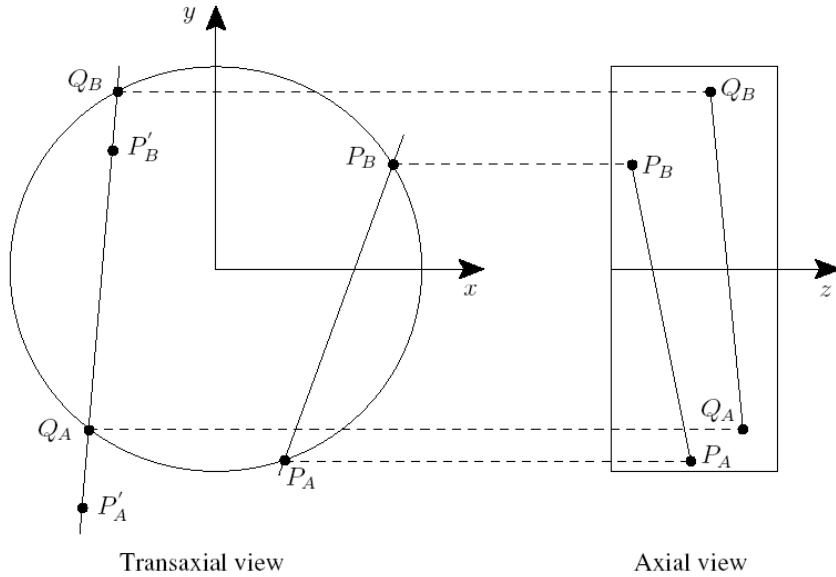


FIG. 6.4: Illustration du recalage des lignes de réponse dans l'espace du détecteur : P_A et P_B représentent les positions des détections, P'_A et P'_B les positions des points corrigé et Q_A et Q_B les détections correspondantes

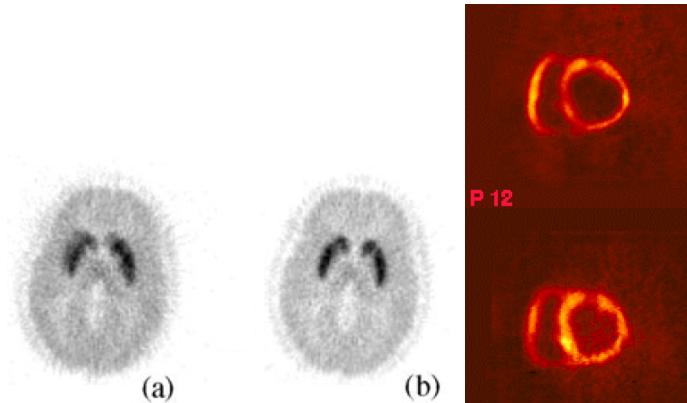


FIG. 6.5: Résultats de l'algorithme de recalage des LOR sur des images de patients utilisant le radiotraceur $[^{11}\text{C}]\text{raclopride}$. (a) montre une image non corrigée du mouvement et (b) une image corrigée. On peut noter que les éléments interne du cerveau sont beaucoup mieux définis. (c) représente une coupe du cœur petit axe non corrigée (en haut) et corrigée (en bas). On peut voir une amélioration de la définition de l'image.

6.4 Correction pendant la reconstruction

Plusieurs auteurs ont présenté des méthodes permettant de réaliser la correction de mouvement pendant la reconstruction. Qiao et al. [Qiao et al., 2006] et Lamare et al. [Lamare et al., 2007b] ont proposés une méthode de correction du mouvement respiratoire basé sur une modi-

fication de la matrice de sensibilité lors de la reconstruction pour prendre en compte le mouvement. Tous les deux utilisent un champ de mouvement élastique estimé en utilisant un champ interpolé par B-splines.

L'algorithme original utilisé est basé sur OPL-EM [Reader et al., 2002] qui organise les données en “sous-ensemble” de la même manière que OS-EM [Hudson, 1994] mais en utilisant les informations list-mode dans l'algorithme de Lamare. Le principe de la reconstruction avec correction du mouvement respiratoire est décrit par la formule suivante :

$$f^{k+1} = \frac{f^k}{S} \sum_{N_{frames}} P_t^T \frac{1}{P_t f^k} \quad (6.1)$$

f^k est l'image à l'itération k ,

T est l'opérateur de transposition

P_t représente la matrice système à l'instant t . Chaque élément p_{ij} de cette matrice indique la probabilité de détecter à la ligne de réponse i un événement généré au voxel j .

S est la matrice de sensibilité :

$$S = \frac{1}{N_{frames}} \sum_{N_{frames}} P_t^T N A_t \quad (6.2)$$

A_t est la matrice permettant de corriger les effets de l'atténuation au temps t et N est la matrice de normalisation qui compense l'inhomogénéité spatiale de la sensibilité.

Dans la publication [Lamare et al., 2007b], deux variantes de cette technique sont comparées avec la correction par synchronisation respiratoire avec recalage présentée précédemment ainsi que la correction pré-reconstruction. Les résultats présentés montrent un clair avantage pour la correction pendant la reconstruction, avec des performances

6.5 Déconvolution de l'image

Cette technique qui est peu décrite en [El Naqa et al., 2006] utilise une connaissance du mouvement respiratoire acquise à partir d'une image TDM 4D pour déduire un filtre appelé TLP (*Tumor Location Probability*) qui correspond à la dégradation dû au mouvement respiratoire.

L'image est ensuite déconvoluée pour corriger les effets du mouvement respiratoire. Cette méthode a été évaluée sur un fantôme physique et des patients réels à l'aide d'un grand nombre de critères provenant pour partie de la TEP (sous-estimation de l'activité de

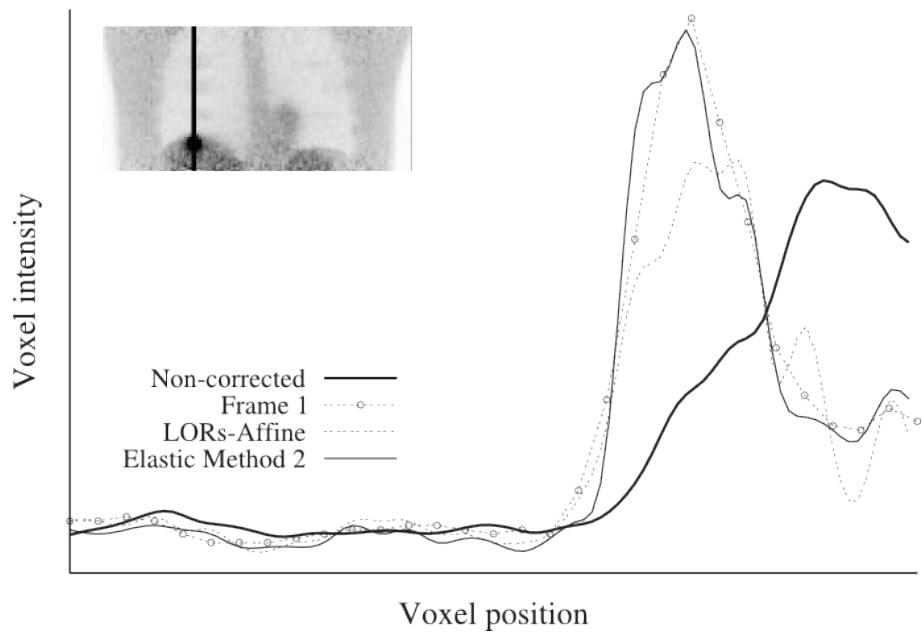


FIG. 6.6: comparaison des performances des différentes techniques de correction du mouvement sur un profil d'image TEP contenant une tumeur placée au niveau du diaphragme.

la tumeur, exemples d'images), et pour partie du domaine de la déconvolution (entropie, "rugosité").

Quatrième partie

Evaluation des performances de diagnostique

Chapitre 7

Performance des outils de détection

7.1 Généralités

En oncologie, la détection des site tumoraux est une étape capitale dans la prise en charge des patients. Elle permet l'évaluation de l'état d'avancement de la maladie, ou encore d'étudier la réponse à un traitement [Dimitrakopoulou-Strauss et al., 2002]. Cette détection se fait actuellement par le médecin qui va observer les images TEP/TDM acquises à la recherche de fixations anormales. Cependant, ces fixations peuvent provenir d'autres facteurs, tels que la "grasse brune", des muscles activés ou encore une inflammation locale [Bordessoule et al., 2006]. Associés aux limites de l'imageur, surtout au niveau du rapport signal sur bruit des volumes reconstruits, il est donc possible qu'il y ait une erreur lors du diagnostique.

Dans cette partie, je vais détailler les techniques qui permettent de comparer les performances de plusieurs observateurs (médecins ou algorithmes) face aux mêmes images, ou alors du même observateur face à plusieurs types d'images différentes.

Pour l'instant le problème va être simplifié au cas où un observateur doit classer un signal en "Sain" (normal, H0) ou "Pathologique" (anormal, H1). Les performances d'un classifieur sont indiquées par la matrice de confusion (table 7.1), qui recense les signaux correctement et incorrectement classés

		Classe estimée	
		Sain	pathologique
Classe réelle	Sain	VN (<i>Vrai Négatif</i>)	FP (<i>Faux Positif</i>)
	Pathologique	FN (<i>Faux Négatif</i>)	VP (<i>Vrai Positif</i>)

TAB. 7.1: Matrice de confusion : donne une vue d'ensemble des performances du classifieur. Elle indique le résultat de la classification de signaux connus.

On utilise habituellement deux grandeurs pour mesurer les performances d'un classi-

fieur :

La *sensibilité* (eq. 7.1) correspond à la proportion d'images correctement évaluées pathologiques par l'observateur par rapport au nombre total d'images réellement pathologiques. Elle donne une information sur la capacité du classifieur à détecter les cas pathologiques.

$$Sensibilite = \frac{VP}{VP + FN} \quad (7.1)$$

La *spécificité* (eq. 7.2) représente le même type de grandeur, mais cette fois ci appliquée aux cas non pathologiques : elle correspond à la capacité du test à donner un résultat négatif lorsque l'image est non pathologique.

$$Specificite = \frac{VN}{VN + FP} \quad (7.2)$$

Ces deux grandeurs sont complémentaires mais ne permettent pas à elles seules de comparer des classifieurs. En effet, un utilisateur va donner des notes, qui vont indiquer son degré de certitude sur la présence de la pathologie (à ne pas confondre avec des notations sur la gravité des lésions, comme les techniques de gradation de [Genestie et al., 1998]).

Les techniques de comparaisons d'organes de décision comme les ROC (Receiver-Operating Curve) permettent de prendre en compte ces incertitudes. Elles proviennent à l'origine du domaine des télécommunications pendant la seconde guerre mondiale, où il fallait une métrique permettant de tester les performances des systèmes RADAR [Zou et al., 2007] pour la détection des avions ennemis. Les courbes ROC servent donc à évaluer la capacité de un ou plusieurs "observateurs" à discriminer des signaux entre deux classes "normal" et "anormal". Les informations de sensibilité et de spécificité se limitent à comparer les performances pour un niveau de détection donné.

7.2 Méthodologie ROC - Receiver-Operating Curve

Les ROC [Swets and Pickett, 1982][Metz, 1986] sont des courbes indiquant la spécificité et la sensibilité du modèle de classification pour différents niveaux de certitudes. Elles fournissent une mesure objective des performances d'un observateur dans une tâche de discrimination entre deux classes.

Elles peuvent être utilisées pour comparer les performances relatives de différents observateurs ou pour déterminer leurs paramètres optimaux.

L'évaluation d'un observateur par la méthode ROC implique de créer un jeu de données de données labellisé en deux classes : Normale (H_0) et Pathologiques (H_1). L'observateur va se voir présenter l'ensemble des images et devra les noter individuellement selon un

barème défini à l'avance (par exemple 0 : pas du tout pathologique, 1 : potentiellement pathologique, 3 :équivoque, 6 : certainement pathologique). Par convention, plus la note (notée λ) sera élevée, plus l'observateur va considérer qu'il est en présence d'un cas pathologique. A l'inverse, une note basse va indiquer un cas présumé sain.

L'observateur peut être un humain ou un algorithme, et les notes peuvent être discrètes ou continues.

Le tracé de la courbe ROC se fait en reportant la sensibilité et la valeur "1-spécificité" du classifieur pour différents seuils. Par construction, la courbe va commencer au point (0,0) (tous les points sont marqués négatifs) et se terminer au point de coordonnée (1,1) (tous les points sont marqués positifs).

Le formalisme ROC considère que les distributions de probabilité des notes des cas H0 et H1 suivent une loi gaussienne (voir fig. 7.2). Ce modèle de décision suppose que l'ensemble des valeurs de λ évaluées sur des cas H0 (sain) suit une distribution de probabilité $P(\lambda_0, \mu_0)$ de valeur moyenne λ_0 et d'écart-type σ_0 . De même, les valeurs de λ évaluées sur des cas H1 (pathologiques), suivent une distribution de probabilité $P(\lambda_1, \mu_1)$. Le mécanisme de décision se base sur le choix d'une valeur de seuil λ_s au-delà de laquelle les observations sont considérées comme pathologiques.

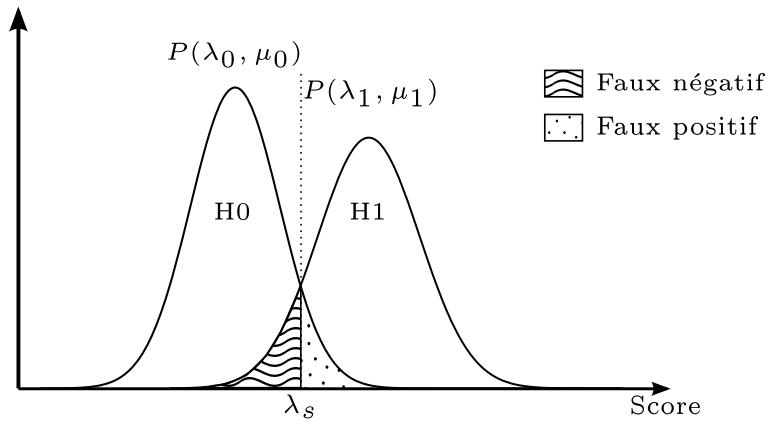


FIG. 7.1: Modèle de la distribution de probabilité de la variable de décision dans pour les populations H0 ($P(\lambda_0, \mu_0)$) et H1 ($P(\lambda_1, \mu_1)$) dans les études ROC. λ_s représente le seuil à partir duquel une observation sera catégorisée H0 ou H1.

Ce seuil permet de modifier de manière dynamique la répartition des observations dans la matrice de confusion. Cela permet d'enrichir la comparaison des observateurs par rapport au couple (sensibilité/spécificité) seul.

Un ensemble d'indicateurs permettent de comparer les performances de classificateurs à partir des courbes ROC. La performance est représentée par une FDM (Figure De Mérite). La FDM la plus simple consiste à choisir un niveau de spécificité (noté α) et à comparer les sensibilités des différents classificateurs. L'avantage de ce système est qu'il permet de comparer les performances dans des conditions proches de la réalité, où l'on cherche à

rester dans un taux de spécificité données. Cependant, les résultats vont dépendre du paramètre α . Une métrique plus globale est l'aire sous la courbe ROC. Étant donné que la courbe sera nécessairement comprise dans un carré unitaire, la valeur de l'aire sera comprise entre 0 (le classifieur donne systématiquement les mauvaises réponses), 0.5 (le classifieur donne des réponses aléatoires) et 1 (le classifieur donne toujours la bonne réponse)[Nie et al., 2006].

IL est important lors du calcul de la FDM d'avoir une estimation de l'erreur. Il est possible de l'estimer en ajustant une courbe théorique (répondant à la loi théorique de la figure 7.2). Plusieurs logiciels ont été développés pour estimer les paramètres, qui ont été comparés dans la publication [Stephan et al., 2003] (AccuROC, Analyse-It, CMDT, GraphROC, MedCalc, mROC, ROCKIT, and SPSS).

Une grandeur souvent utilisée dans la littérature pour évaluer la pertinence d'un résultat est la *p-valeur*. Elle représente la probabilité d'obtenir un résultat au moins aussi extrême que le résultat obtenu (dans notre cas, la courbe ROC), en prenant en compte l'hypothèse selon laquelle le classifieur est aléatoire. Elle permet de vérifier si le test est statistiquement significatif.

Le problème des courbes ROC est que l'observateur ne donne pas d'information de localisation du problème dans l'image. Dans notre cas, nous voulons comparer des classifiants qui détectent les tumeurs dans l'image. Il faut non seulement savoir si des lésions sont présentes, mais aussi avoir leur nombre et leur localisation. Cela est plus proche du travail en routine clinique qui consiste à évaluer l'étendue et le nombre des lésions pour déterminer l'efficacité d'un traitement par exemple.

Pour éviter cette limitation, plusieurs extensions à la méthodologie ROC sont décrites dans la littérature : L-ROC, AF-ROC ou encore F-ROC. Les L-ROC sont décrites ci-après, tandis que les AF-ROC et F-ROC seront décrites dans la section suivante.

7.2.1 Courbes Localization ROC (L-ROC)

L'analyse L-ROC [Farquhar et al., 1999] ajoute l'information de localisation lors de la décision. L'observateur doit indiquer sur l'image qu'il considère comme pathologique la localisation de la lésion la plus probable. Elle est considérée comme un vrai positif si la distance entre la localisation indiquée et la localisation réelle de la lésion est inférieur à une certaine distance.

Cependant, bien que cette technique prenne en compte l'information de localisation, elle ne permet pas de traiter de manière satisfaisante les cas multi-lesions.

7.3 F-ROC

7.3.1 Courbes Free-ROC

les courbes F-ROC [Bunch et al., 1978] sont une généralisation des courbes ROC aux cas où l'on évalue la capacité de l'observateur à détecter un ensemble de lésions dans une série d'images. Chaque image pouvant contenir un nombre indéfini de lésions. L'observateur va donc devoir pointer sur l'image l'ensemble des sites suspects et y associer une note.

Dans ce cas, on ne peut pas utiliser le formalisme ROC car le terme de spécificité n'est pas directement calculable pour chaque niveau de confiance. On utilise à sa place le nombre moyen de faux positifs par image pour un seuil donné (voir fig.7.3.1).

On utilise les termes de LL (Localisation de Lésion) et NL (Non-Lésion) en lieu et place des informations de vrai positifs et faux positifs sur les courbes ROC. De la même manière, la sensibilité et la spécificité sont respectivement FLL (Fraction de localisation de lésion) et FNL (Fraction des Non-Lésions).

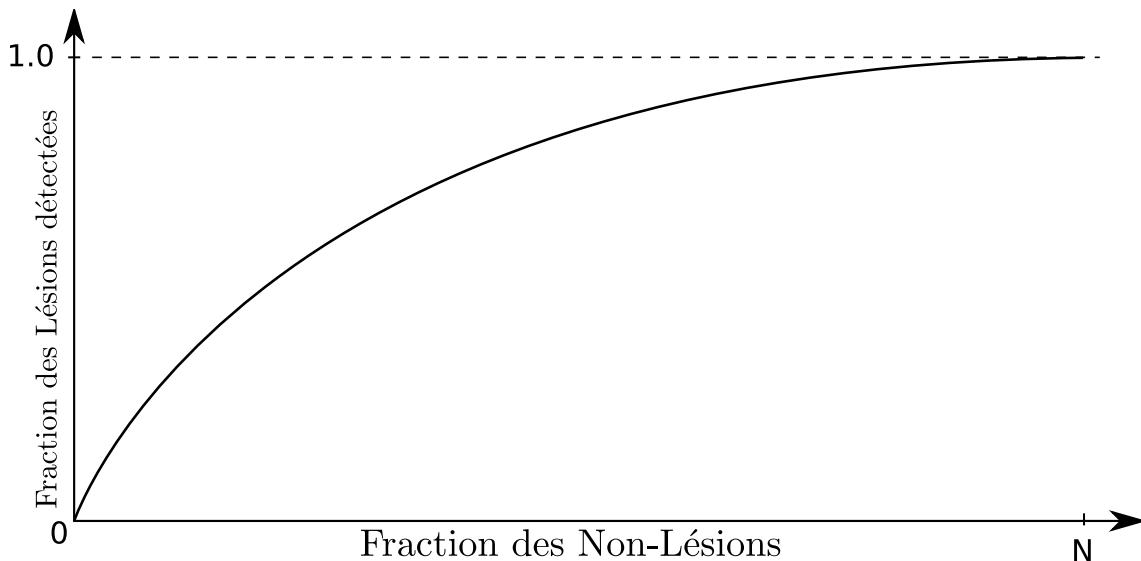


FIG. 7.2: Courbe Free ROC

Les courbes F-ROC n'ayant pas de bornes sur l'axes des abscisses, il est impossible de comparer plusieurs courbes à partir de l'aire sous la courbe. Il reste cependant possible de comparer la sensibilité pour un nombre de faux positifs donnés, mais on retrouve les mêmes problèmes que pour l'analyse ROC : il faut choisir un paramètre.

7.3.2 Courbes Alternative Free-ROC

Les courbes A-FROC [Chakraborty and Winter, 1990] sont des extensions des courbes Free-ROC présentées précédemment mais qui ne prennent en compte que le faux positif de plus haut score par image, ce qui ne pénalise pas le cas où un observateur indique un grand nombre de localisations sans lésions.

7.3.3 Comparaison des courbes

Plusieurs techniques ont été développées pour permettre de réaliser des comparaisons. De la même manière que pour les courbe ROC, il est possible de comparer les courbes F-ROC en fonction de la FLL pour un nombre de faux positifs donnés. Cependant, étant donné que les courbes F-ROC n'ont pas de fin déterminée, il n'est pas possible d'utiliser l'aire sous la courbe. JAFROC[Chakraborty and Berbaum, 2004] (JAcknife Free Receiver Operating Curve) est un algorithme et un logiciel développé par Chakraborty et se base sur une FDM non liée directement à la courbe.

Cette mesure de performance utilise un algorithme dérivé des études A-FROC, ce qui signifie qu'il n'utilise pas l'ensemble des informations disponibles dans les courbes Free-ROC. Il va comparer les scores des faux positifs de plus forte note pour chaque image avec les notes des vrais positifs. La FDM mesure la probabilité d'avoir un score de vrai positif supérieur à celui d'un faux positif (de n'importe quelle image).

Soit θ la valeur de la FDM, N_T le nombre total d'images, indexés par i , N_A le nombre total de cas pathologiques, indexés par j . n_j est le nombre total de lésions dans le cas anormal j .

$$\theta = \frac{1}{N_T N_A} \sum_{i=1}^{N_T} \sum_{j=1}^{N_A} \sum_{k=1}^{n_j} W_{jk} \psi(X_i, Y_{jk})$$

$$\psi(X, Y) = \begin{cases} 1.0 & \text{si } Y > X \\ 0.5 & \text{si } Y = X \\ 0.0 & \text{si } Y < X \end{cases} \quad (7.3)$$

$$\text{avec } \sum_{k=1}^{n_j} W_{jk} = 1$$

X_i le score du plus haut Faux positif de l'image i , Y_{jk} est la note de la lésion k de l'image j . Si une lésion n'a pas été détectée, alors sa note sera par défaut de "0".

Les poids W_{jk} correspondent à l'importance relative de détecter la lésion k dans l'image j pour le diagnostique. Pour chaque image, la somme des poids doit être égale à 1.

Une seconde version de JAfroc existe avec un pouvoir statistique plus important, mais elle nécessite de disposer d'un grand nombre de cas non pathologiques. La formule est la même que celle précédente (eq. 7.3). La seule différence est que la première sommation se fait sur l'ensemble des cas non pathologiques N_N (eq. 7.4).

$$\theta = \frac{1}{N_T N_A} \sum_{i=1}^{N_N} \sum_{j=1}^{N_A} \sum_{k=1}^{n_j} W_{jk} \psi(X_i, Y_{jk}) \quad (7.4)$$

Chapitre 8

Systèmes de détection

8.1 Les CAD en TEP

Les systèmes CAD (Computer-Aided-Detection) sont des algorithmes permettant d'assister le praticien dans la détection des lésions ou le classement des images médicales. Dans le cadre de l'imagerie TEP oncologique, le besoin principal est celui du suivi thérapeutique. Pour cela, il est important de détecter d'éventuelles lésions résiduelles. Pour cela, il faut que le système CAD soit particulièrement adapté à la recherche de petites lésions de faible contraste qui pourraient échapper au médecin. Cependant, le diagnostique, qui consiste à évaluer la dangerosité des lésions, et leur caractère pathologique est une tâche plus complexe qui relève plus des système d'aide au diagnostique, qui ne seront pas traités ici.

Le développement des systèmes CAD a débuté dans les années 1980 [Chan et al., 1987], notamment pour détecter les micro calcifications en mammographie. Bien qu'il existe plusieurs systèmes CAD commerciaux pour l'imagerie TDM (xLNA pour philips par exemple), aucun CAD commercial pour la TEP n'existe à ma connaissance.

8.2 Types de classification

8.2.1 supervisée - méthodologie

Les classifieurs supervisés nécessitent une connaissance a priori des classes. On entraîne le classifieur en lui fournissant des *exemples* de cas avec l'étiquette associée. A partir de cette base de données d'entraînement, le classifieur va générer un *modèle* predictif permettant de classer de futurs exemples non encore connus.

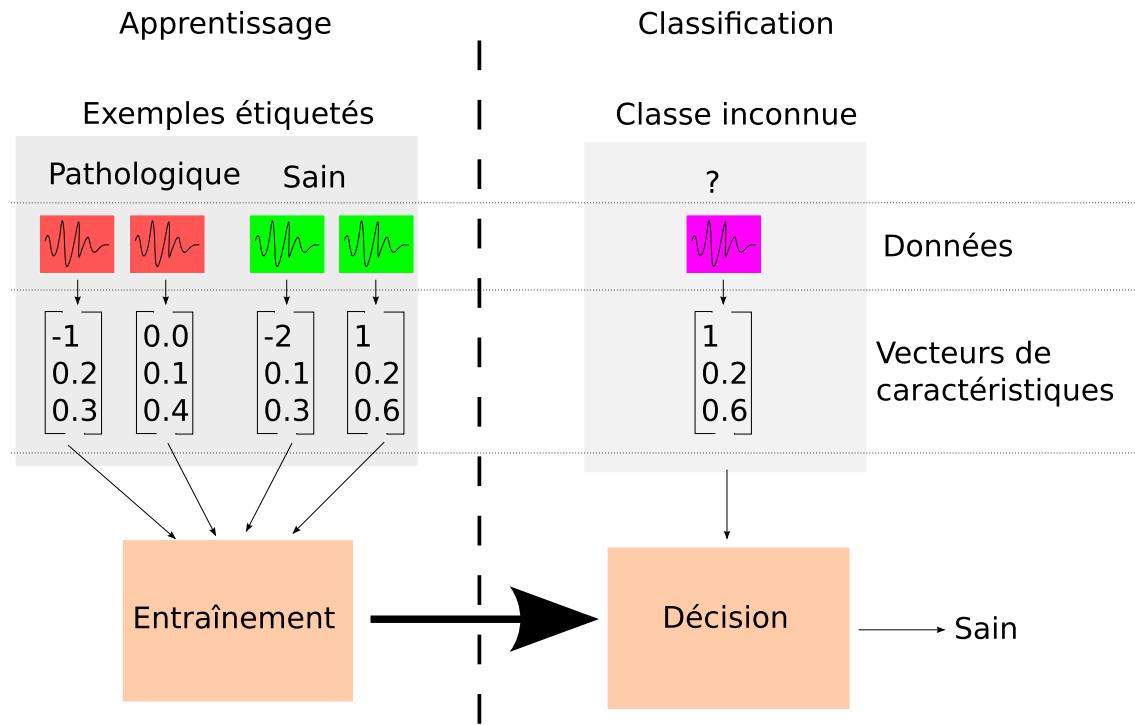


FIG. 8.1: Fonctionnement d'un classifieur supervisé : Les données d'apprentissage servent à entraîner le classifieur pour générer un modèle. Ce modèle permettra de rattacher des observations aux classes apprises.

Machines à vecteur de support (SVM)

La “Machine à Vecteur de Support”, aussi appelée “Séparateur à Vaste Marge”, ou “Support Vector Machines” en Anglais, est un classifieur qui comme son nom l’indique vise à maximiser la marge[Boser et al., 1992], qui est la distance entre les points des données et la surface séparatrice (voir figure 8.2.1).

8.2.2 non supervisée - méthodologie

Dans le système de classification non supervisé, on fournit directement au classifieur l’ensemble des données à traiter. Il devra de lui-même les classer par similitude en groupes. On utilise ce type de classifieur si on ne connaît pas a priori les classes 8.2.2.

La classification non supervisée repose sur une méthode statistique utilisant une fonction de proximité.

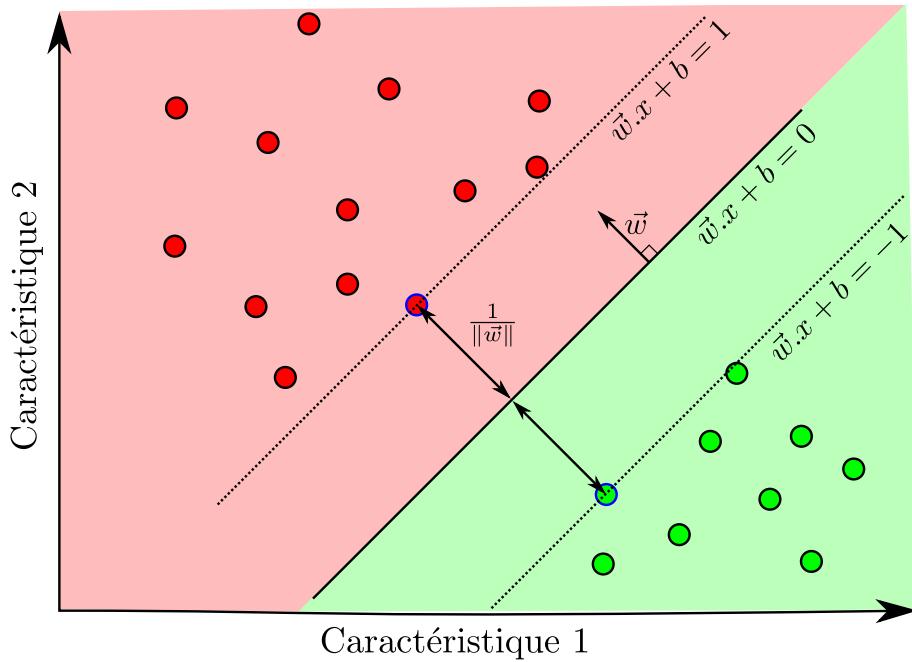


FIG. 8.2: Machine à Vecteur de Support : Les points vecteur de support (entourés de bleu) sont les seuls utilisés pour calculer la surface de séparation d'équation $\vec{w} \cdot x + b = 0$. Le vecteur \vec{w} est normal à la surface de séparation et permet de calculer la marge $\frac{1}{\|\vec{w}\|}$.

8.3 classifiers

CROSS-VALIDATION

8.3.1 SVM (Separateur à Vaste Marge)

Ce type de classifieur se base sur

8.3.2 LDA

8.4 Systèmes humain

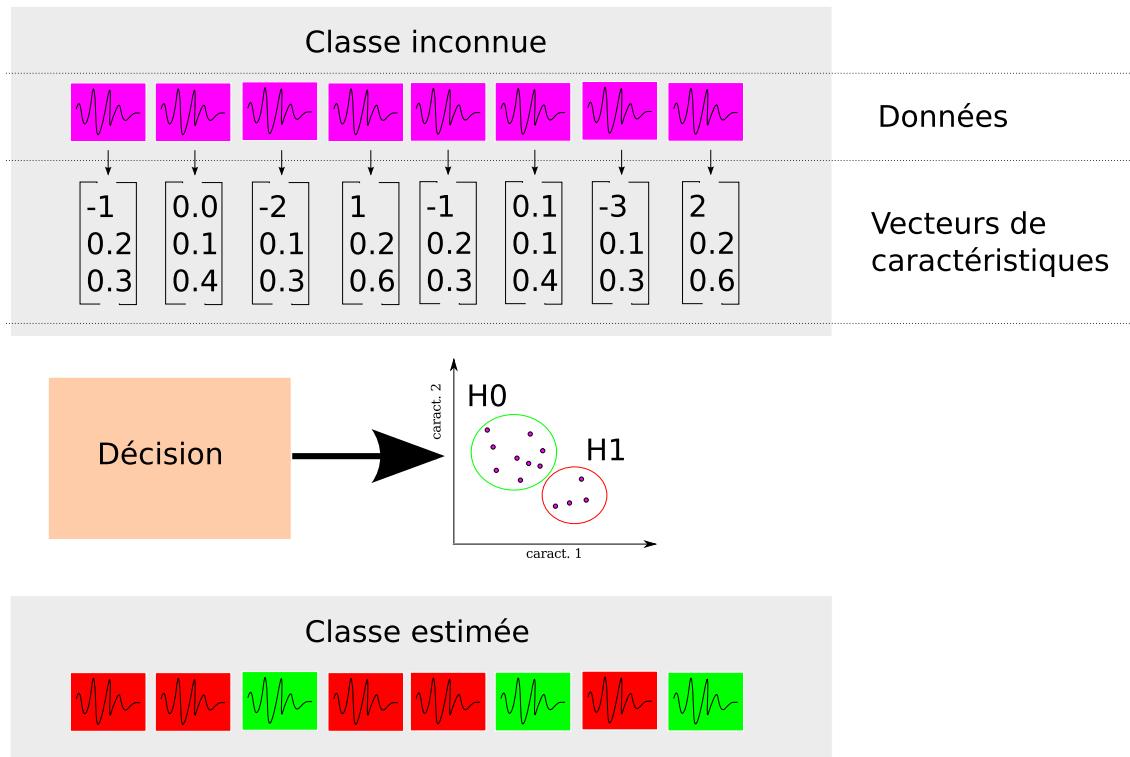


FIG. 8.3: Fonctionnement d'un classifieur non supervisé : Les données brutes sont envoyées au classifieur qui va les regrouper en classes en fonction de leur répartition dans l'espace des caractéristiques.

Cinquième partie

Simulation et base de données (MIC 2011)

Chapitre 9

Simulations

9.1 principe des simulations

9.1.1 monte carlo

9.1.2 analytiques

9.1.3 MC accélérés

9.2 simulateurs disponibles

9.3 processus de simulation avec SORTEO

9.4 Contribution à SORTEO

Chapitre 10

Base de donnée

10.1 Présentation

10.2 Modèles

10.3 données clef

Sixième partie

Résultats

Chapitre 11

Méthodes

J'ai utilisé la base de donnée présentée dans la partie V pour évaluer les performances des techniques de correction du mouvement respiratoire présentés dans le chapitre 6.

Les techniques de correction du mouvement implémentées sont les suivantes :

1. Correction pendant la reconstruction par modification de la matrice système (voir section 6.4)
2. Correction post-reconstruction par recalage des images prises à différents instants du cycle (voir section 6.2)

Elles sont comparées avec les images non corrigées et des images statiques (qui représentent une correction parfaite).

Dans le cas présent, l'objectif est d'évaluer les performances des techniques de correction du mouvement sur la détection des lésions de faible contraste et faible diamètre. Pour cela, les performances d'un système de détection automatique seront mesurées à l'aide des courbes F-ROC.

11.1 Système CAD

Le système CAD que nous utilisons a été développé à l'origine pendant les travaux de thèse de Sandrine Tomei ainsi que mes travaux de master [?, Lartizien et al., 2010]. Nous l'avons amélioré et adapté aux besoins de cette étude, notamment en développant les mesures de performances.

Le CAD utilise des informations fréquentielles obtenue par décomposition non décimée des images en ondelettes Biorthogonale 4.4. Ces données sont utilisées par le système de classification basé sur un SVM travaillant voxel par voxel. Une étape de réduction des faux positifs est ajoutée par la suite. Étant donné le faible nombre de tumeurs présentes

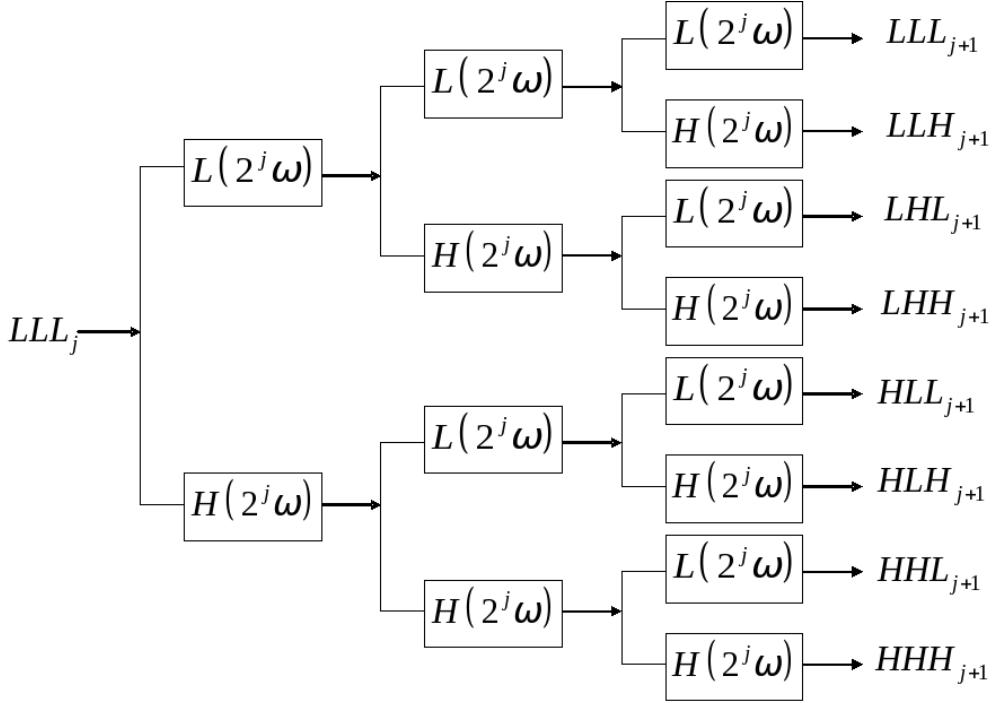


FIG. 11.1: *Décomposition en ondelettes par banc de filtres : Chaque image est filtrée selon les 3 dimensions pour obtenir les coefficients d'ondelettes et d'échelle de chaque voxel de l'image.*

dans les organes considérés, la base d'apprentissage est générée à partir des images utilisées pour la base de test.

11.1.1 Vecteur de Caractéristiques

Nous avons choisi d'utiliser une décomposition en ondelettes 3D non décimée par banc de filtres. Dans le cas tridimensionnel, la décomposition par banc de filtre est résumé par la Figure 11.1.1. L'image de départ est filtrée séparément dans les trois directions de l'espace par un filtre fréquentiel passe-haut correspondant à la fonction d'ondelettes (noté H) et passe-bas correspondant à la fonction d'échelle (noté L).

Ainsi sept images de détails (HHH, LLH, ...) et une image d'approximation (LLL) sont produites pour chaque niveau j de décomposition. Les huit images du niveau suivant $j + 1$ sont générées de la même manière, mais en considérant l'image d'approximation LLL_j du niveau précédent comme image de départ. Les caractéristiques des images, rassemblées dans un vecteur descripteur de taille $8 \times j$, correspondent ici à l'ensemble de ces coefficients pour chaque voxel de l'image.

11.1.2 Génération de la base d'apprentissage

Les données générées par la décomposition en ondelette de chaque image se présentent sous la forme de volumes 4D, indiquant pour chaque voxels de l'image 3D l'ensemble des coefficients associés.

La base d'apprentissage sert à entraîner le classifieur dans un processus de classification supervisée, en lui fournissant un ensemble d'exemples avec leur classes associées (voir figure 8.2.1) page 52).

La génération de cette base demande l'extraction de points de la classe "pathologique", et de la classe "sain". Les points correspondants à la classe pathologique sont extraits des centres de chaque tumeur de chaque image de la base pour l'organe considéré. Les points de la classe "sain" sont extraits de manière aléatoire dans les volumes de toutes les images (hors tumeurs). Pour des raisons de simplicité d'implémentation, un nombre fixe de points sont extraits de chaque image.

11.1.3 Apprentissage de la Machine à Vecteur de Support (SVM)

Une fois la base de données générée, elle sera fournie en entrée au classifieur SVM (Machine à Vecteur de Support) pour l'apprentissage.

Le principe des SVM est, de trouver l'hyperplan optimal de séparation, qui va maximiser la marge de séparation entre les deux classes "pathologique" et "sain" (voir 8.2.1). Cette marge correspond à la distance à l'hyperplan entre les plus proches vecteurs de caractéristiques appartenant à chacune des classes. La définition de cette marge et donc de l'hyperplan, se fait uniquement à partir de vecteurs de support, qui correspondent à l'enveloppe du groupement de point de l'une des deux classes.

Le SVM va fournir en sortie un modèle décrivant l'hyperplan permettant de séparer les données.

11.1.4 Génération des sites présumés

Une fois le classifieur entraîné, on lui soumettra les voxels des images pour qu'il les classe. On aura donc en sortie un ensemble de carte de score (une par image), dans lesquelles chaque voxel correspond au score indiqué par le SVM. Ce score est une mesure de distance par rapport à l'hyperplan de séparation dans l'espace des points d'apprentissage, et donne donc une indication de la "certitude" du SVM vis-à-vis de sa classification.

Cette carte de score est ensuite seuillée pour fournir une carte binaire, indiquant pour chaque voxel la classe fournie par le SVM. Le mécanisme de sélection du seuil est défini plus tard.

A cette étape, le résultat est sous forme "voxels", et non "région". Nous avons choisi de travailler sur des agrégats de points plutôt que directement sur les voxels car les cartes binarisées sont très bruitées, et donc non directement exploitables. Les voxels "pathologiques" sont donc agrégés selon une connexité 26 (3x3x3). Un score est associé à chaque agrégat, correspondant au score le plus important observé dans l'aggrégat.

11.1.5 Évaluation du résultat

Pour évaluer le résultat, nous avons mis au point des règles permettant de classer les agrégats si nous avons accès à la vérité terrain. Ce sont ces algorithmes qui sont utilisés dans les évaluateurs de performance présentés ci-après.

Les agrégats sont classés en LL (Lésion localisée) et NL (Lésion Non Localisée) selon qu'ils peuvent être considérés comme des vrai positifs ou des faux positifs :

Soit L l'ensemble des points de la lésion, A les points correspondant à l'amas candidat.

Les agrégats seront considérés comme des vrai positifs si ils intersectent une tumeur selon les règles suivantes, ou comme de faux positifs sinon. Cependant, si leur taille est inférieure à la taille minimale définie par première règle ($\alpha \times \text{card}(L)$), l'amas n'est pas considéré.

$\text{card}(L \cap A) > \alpha \times \text{card}(L)$ avec α fixé de manière empirique à 0.05 qui fixe la proportion minimale de la tumeur qui doit être présente dans l'amas. Elle permet d'éviter les amas qui intersecteraient la tumeur par accident.

$\text{card}(L \cap A) > \beta \times \text{card}(A)$ avec β fixé de manière empirique à 0.20 limite l'étendue de l'amas en dehors de la tumeur.

11.1.6 Résumé

Ce pseudo-code décrit le CAD, de l'importation des images à l'extraction des lésions potentielles :

1. Décomposition des images en ondelettes : Pour chaque voxel de l'image d'origine, on obtient entre 8 et 32 coefficients suivant le niveau de décomposition de l'image, qui correspondent au vecteur de caractéristiques utilisés par le classifieur
2. Extraction de la base d'apprentissage : Les coefficients des centres de toutes les tumeurs sont extraits des volumes décomposés, et vont former la base d'apprentissage pathologique. Un certain nombre de voxels sont tirés aléatoirement dans les zones normales de chaque images et leur coefficients sont ajoutés à la base saine.
3. Apprentissage : Le classifieur SVM est entraîné sur cette base d'apprentissage pour générer le modèle qui sera utilisé pour le test.

4. Tests : Le SVM entraîné est utilisé pour classer chaque voxel contenu dans les organes à évaluer (poumon et foie).
5. Réduction des Faux-positifs : Les points sont agrégés en composantes connexes (connexité 26 en 3 dimensions).
6. Évaluation : Si nous avons accès à la vérité terrain, il est possible d'évaluer les performances du CAD en classant les agrégats en LL (Lésion Localisée) et LN (Lésion Non Localisée).

11.2 Optimisation des paramètres du système CAD

Les différentes étapes de l'évaluation des performances nécessitent la fixation d'un grand nombre de paramètres.

11.2.1 Paramètres à optimiser

Paramètres de la base d'apprentissage

Nous avons sélectionné plusieurs paramètres qui selon nous peuvent modifier les performances des bases d'apprentissage. Ces paramètres sont assez généraux et seront donc sélectionnés une fois et appliqués à toutes les modalités de la même manière :

Normalisation : Les données sont normalisées de manière à ce que la moyenne et l'écart-type de chaque caractéristique soit de 1 ($(\mu, \sigma) = (1, 1)$) (*moyenne*), ou alors pour que l'ensemble des valeurs soit comprises entre -1 et +1 (*écart*).

Nombre de points de la base d'apprentissage : Le nombre de points extraits de chaque image pour alimenter la base d'exemples normaux peut avoir une influence sur les résultats. Trois valeurs sont testées. 100 pts/im. (soit 1500 pts. négatifs), 200 pts/im. (soit 3000 pts. négatifs) et 1000 pts/im. (soit 15000 pts. négatifs).

positions des points extraits : Les points normaux extraits de la base d'image peuvent être extraits de tout le volume de l'organe hors tumeurs, ou bien extraits sur une partie seulement des volumes hors tumeur (érosion morphologique de 2 voxels) pour simplifier .

Le but de la normalisation est d'homogénéiser les plages de valeurs des différentes caractéristiques pour faciliter le travail du classifieur, dont les paramètres C et gamma dépendent de la distance entre les points et ne permettent pas de gérer des différences trop importantes d'étendues dans les caractéristiques. La première méthode de normalisation (*moyenne*) a l'avantage d'être relativement peu sensible aux valeurs extrêmes, contrairement à la seconde (*écart*).

Le nombre de point de la base d'apprentissage détermine directement la qualité de l'apprentissage. Le nombre de caractéristique est de $8 \times j$, avec j le niveau de décomposition des images. Il n'existe pas de règle définitive pour déterminer le nombre d'exemples nécessaire en fonction du nombre de caractéristiques, mais les SVM sont relativement bon pour éviter le sur-apprentissage. Dans notre cas, il vaudrait donc mieux avoir plus de données. Cependant, le nombre de points notés "tumeurs" est limité par le nombre de tumeurs présentes dans la base d'apprentissage (173 tumeurs pour le poumon, 106 pour le foie). Il y a donc un risque de déséquilibre de la base d'apprentissage, qui devrait idéalement avoir le même nombre d'exemples "tumeur" que "normales". Des techniques de correction existent pour les SVM pour contrebalancer ces déséquilibres en indiquant un paramètre C différent pour chaque classe, mais les tests que nous avons réalisés ne montrent pas d'améliorations significatives des performances.

Les points normaux extraits des images pour alimenter la base vont avoir une influence directe sur la qualité des résultats. Idéalement ils devraient être représentatif de l'ensemble des cas rencontrés dans la base de tests, mais les bordures de certains organes peuvent ressembler à des tumeurs, et rendre l'estimation de la surface de séparation plus difficile. J'ai donc voulu évaluer la performance du CAD sur une base dépourvue de ces données ambiguës. Pour cela nous avons réalisé une Érosion de 2 voxels sur les masques des volumes à extraire.

Paramètres du classifieur

Le classifieur (SVM) va récupérer en entrée les données d'apprentissage formatées selon les choix opérés précédemment et va générer un modèle. Cependant cet algorithme dispose de ses propres paramètres, qui sont beaucoup plus dépendant des données. Un nouveau jeu de paramètres sera donc calculé pour chaque modalité.

niveau de décomposition j : Il correspond au nombre de niveaux de décomposition pris en compte par le SVM. Les niveaux de décomposition élevés correspondant à des informations de très basse fréquence, les informations qu'ils apportent ne sont pas forcément pertinentes pour la détection des lésions. Cependant, les caractéristiques fréquentielles des images générées par les différentes modalités étant différentes, il est nécessaire d'adapter ce paramètre pour chacune.

Coefficient de pénalisation C : Lors du calcul de l'hyperplan de séparation des données, chaque point mal classé va pénaliser la surface d'un facteur proportionnel à C .

Largeur de bande γ : Cette valeur influe directement sur la largeur de bande du noyau utilisé par le classifieur (RBF).

11.2.2 Méthodes d'optimisation

La sélection des deux types de paramètres vu précédemment est obtenue à l'aide des algorithmes présentés ci-après.

Choix des paramètres du classifieur

Pour obtenir les meilleurs paramètres du classifieur, je réalise une recherche exhaustive par grille. Cela correspond à mesurer un ensemble de critères de performance pour chaque jeu de paramètre possible.

Par exemple, si j'ai deux paramètres A et B, et que je recherche le meilleur jeu de paramètres, je vais tout d'abord sélectionner pour chaque paramètre la taille de la zone de recherche. Pour A, ce sera les valeurs -1, 0, 1, 2 et pour B les valeurs 100, 10000, 50000. Les critères de performance seront évalués pour toutes les combinaisons des valeurs de A et B, tels que (-1, 100), (-1, 10000), (-1, 50000), (0, 100), ..., (2, 50000). Le jeu de paramètres ayant obtenu les meilleures performances sera donc sélectionné.

Les critères de performance que nous avons sélectionné sont la sensibilité et la spécificité obtenue pour une validation croisée à 5 éléments réalisée sur la base d'apprentissage comme présenté sur la figure 11.2.2. La validation croisée est une technique permettant de minimiser le biais sur les performances.

Pour choisir les meilleurs paramètres de classifieur, j'ai effectué une recherche exhaustive par grille avec les paramètres suivants :

C : de 1 à 10000 en 15 pas logarithmique

gamma : de 0.0001 à 1 en 15 pas logarithmique

j : de 1 à 4, soit de 8 à 32 caractéristiques

L'optimisation a été réalisée à l'aide du logiciel rapid-i [Mierswa et al., 2006] pour chaque modalité. Les indicateurs de performance calculés par validation croisée sont les suivants : Sensibilité, Spécificité. Le triplet de paramètre retenu est celui qui maximise la sensibilité.

Nous avons représenté pour chaque modalité le nuage de point positionnant chaque triplet dans un espace à deux dimensions ("Sensibilité", "Spécificité"). Cela permet de vérifier que le critère choisi (maximisation de la sensibilité) ne se fait pas trop au détriment de la spécificité.

De ce nuage de point nous pouvons voir le front de pareto. Ce type de diagramme permet de rechercher un optimum selon plusieurs critères incompatibles. Dans notre cas, nous voulons à la fois une sensibilité et une spécificité importante, sachant qu'il n'existe pas de jeu de paramètres "parfaits" qui permettent d'avoir 100% aux deux. Dans notre

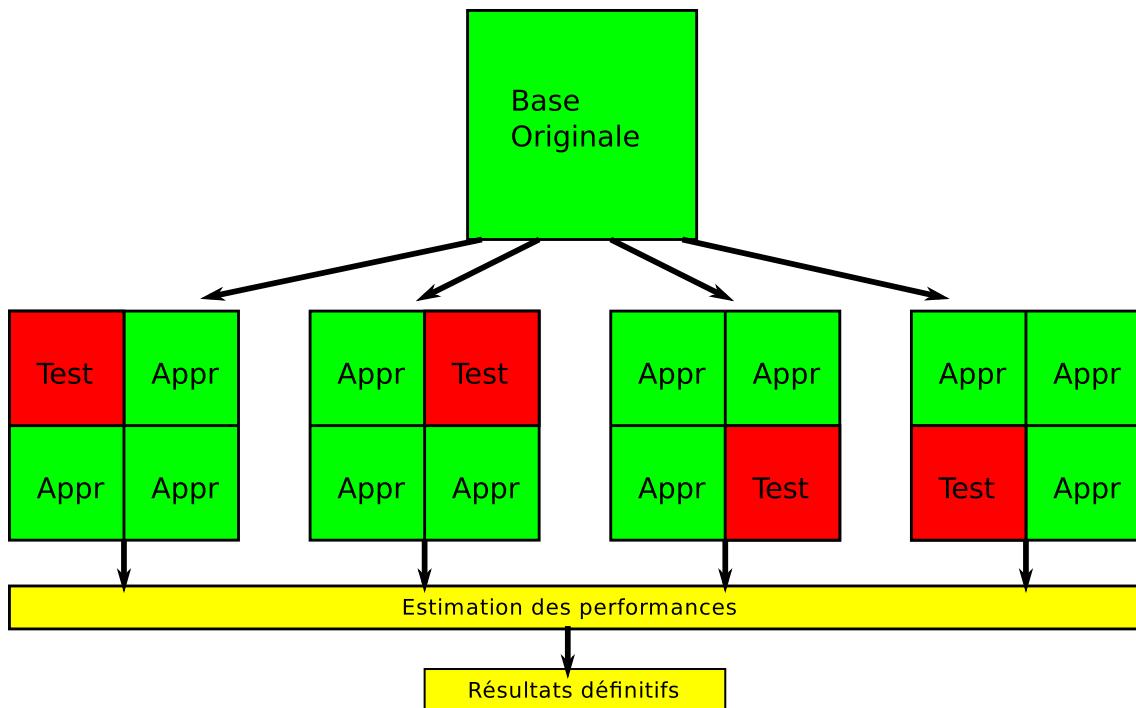


FIG. 11.2: Réalisation d'une validation croisée à n éléments (ici 4) : La base d'exemples d'origine est décomposée en n parties égales. La mesure de performance se fait n fois, avec à chaque fois un apprentissage sur $n - 1$ éléments et un test sur l'élément restant. La mesure de performance est réalisée sur les résultats des n tests.

cas, le front de pareto va permettre de vérifier que le choix par maximisation de la sensibilité ne se fait pas au détriment de la spécificité.

Mesure de performance de la base d'apprentissage

La mesure de performance décrite ici permet de sélectionner les meilleurs paramètres de génération de la base d'apprentissage, mais aussi une fois ces paramètres sélectionnés, de comparer les différentes techniques de correction du mouvement respiratoire.

Pour un seuil donné, Le CAD va générer un ensemble d'aggrégats associés à un score 11.1.4. Il est ensuite possible d'évaluer la performance de ce CAD à l'aide de courbes F-ROC décrites en 7.3.1 en connaissant la vérité terrain.

Le seuil sélectionné pour générer les aggrégats sera celui qui va maximiser la LLF (Fraction de Localisation de Lésion), c'est à dire celui qui va permettre de détecter un maximum de lésions.

Cette sélection est réalisée selon l'algorithme suivant :

- *Sensibilité := []*

- Pour chaque seuil s entre -2 et +2 (40 pas), faire :
- *CarteSeuillée* := Seuiller *carte de Score* au seuil s
- *Agrégats* := Calcul des agrégats de *CarteSeuillée* (*Score*, classe, étendue).
- *Sensibilité[s]* := Calcul *Sensibilité*(*Agrégats*)
- fin faire
- *Seuil Optimal* := Seuil qui maximise la *Sensibilité*

Ce critère de sélection va naturellement engendrer un grand nombre de faux positifs, mais il faut garder à l'esprit qu'il sera utilisé pour réaliser des courbes F-ROC, qui indiquent une spécificité pour chaque nombre de faux positif en jouant sur un second seuil, toujours supérieur à celui retenu pour extraire les agrégats.

Le second critère utilisé pour comparer les bases d'apprentissage est la figure de mérite décrite en 7.3.3. Elle va comparer les scores des faux positifs de plus fort score pour chaque image avec les score des vrais positifs de l'image.

Chapitre 12

Analyse des résultats

Dans ce chapitre je vais détailler les résultats obtenus par les méthodes présentées précédemment. Je commencerais par présenter les courbes obtenues lors de l'étape d'optimisation et d'adaptation du CAD aux données, puis je parlerais des résultats obtenus par ce CAD pour les différentes modalités.

12.1 Optimisation des paramètres

Nous avons réalisé des mesures de performances pour 5 jeux de paramètres différents présentés ci-après. Les paramètres en *gras* correspondent à ceux utilisés pour la base témoin.

- Nombre de points de la base d'apprentissage (100, 200, 1000)
- Normalisation des données (*moyenne*, +1/-1)
- Position des points de la base d'apprentissage (*organe*, Érosion)

Témoin : contient des données normalisées par la méthode moyenne avec 200 points extraits de chaque organe.

Érosion : contient des données normalisées par la méthode moyenne avec 200 points extraits de chaque organe érodée. C'est la seule base pour laquelle nous n'avons pas utilisé l'ensemble de l'organe hors tumeurs pour l'extraction des points.

Appauvrie : contient des données normalisées par la méthode moyenne avec 100 points extraits de chaque organe.

Enrichie : contient des données normalisées par la méthode moyenne avec 1000 points extraits de chaque organe.

Normalisation : contient des données normalisées par la méthode +1/1 avec 200 points extraits de chaque organe.

Pour vérifier que les paramètres du classifieur sont stables, une nouvelle base Témoin (Témoin 2) a été générée avec les données de seulement 14 images sur les 15 que nous avons simulé. Les points "normaux" ne sont pas non plus extraits aux mêmes endroits que pour la base Témoin. Nous allons donc vérifier que les paramètres optimaux du CAD ne sont pas dépendant de la base mais de ses paramètres.

Sélection des meilleurs paramètres du classifieur

Les paramètres du classifieur sont déterminés par une recherche par grille. Elle consiste à rechercher l'optimum en évaluant la performance ce chaque jeu de paramètre dans un ensemble déterminé à l'avance pour représenter . La performance de chaque triplet (C , γ , j) est estimée en réalisant une cross-validation à 5 validations sur l'ensemble de la base d'apprentissage.

Les paramètres sont choisis à partir du front de pareto figure 12.1 en maximisant la sensibilité.

Dans l'ensemble, on peut voir clairement que les points correspondant au premier niveau de décomposition (bleu foncé) ont une performance systématiquement inférieure aux autres. Pour la base témoin, la performance maximal est atteinte pour environ 15% de sensibilité et une spécificité de 93.5%, ce qui correspond à la valeur de spécificité la plus faible de tous les points de la base témoin. On observe cependant un front de pareto marqué, bien qu'en fort retrait par rapport aux autres niveaux de décomposition. Ce même constat se retrouve pour les bases appauvries et enrichies.

Les deux autres bases ont des comportements différents. Pour la base normalisée, les performances s'effondrent suffisamment pour que le classifieur soit quasiment incapable de discerner les classes, avec une sensibilité maximale de 1% environ. Cela indique qu'il ne parvient pas à trouver une surface de séparation des données avec les paramètres indiqués. Puisque la normalisation moyenne parvient à mieux séparer les données pour ce niveau de décomposition, il est probable que la normalisation ne parvienne pas à homogénéiser les différentes dimensions de manière satisfaisante. Dans le cas de la base érodée, la simplification du problème de classification fait que les performances sont nettement en hausse avec des performances de l'ordre de 55% de sensibilité pour 99.2% de spécificité, ce qui est le score le plus élevé de toutes les bases.

Les performances du second niveau de décomposition (bleu ciel) sont toujours situées environ au barycentre entre les performances du premier niveau et celles des niveaux 3 et 4.

Les performances des décomposition de niveaux 3 (vert) et 4 (rouge) sont systématiquement meilleures que les autres mais sont entremêlées. La base Témoin et la base Appauvrie montrent une claire avance tant en terme de sensibilité que de spécificité pour 3 niveaux de décomposition. La base érodée ne montre quand à elle aucune différence de spécificité entre ces deux niveaux de décomposition, mais montre cependant une faible amélioration

de la sensibilité pour les 4 niveaux (de 0.5%). Quand à la base normalisée, on n'observe pas de réelles différences entre les deux niveaux de décomposition.

La figure 12.1 référence tous les paramètres sélectionnés à partir des courbes de pareto. On peut observer que la sensibilité la plus importante est atteinte pour la base appauvrie, avec 82% de bonne détection. Ce taux est sensiblement le même que celui de la base Érosion (80%). Ces valeurs importantes par rapport aux autres peuvent s'expliquer par le fait que ces deux bases proposent un problème simplifié. Dans le cas de la base Érosion, les cas litigieux (discontinuités proches des bords des organes) ont été retirés de la base, ce qui simplifie le problème, tandis que pour la base appauvrie, c'est le nombre de points "sain" plus faible de la base d'apprentissage (1500 contre 3000) qui permet de rendre le problème plus simple à traiter au détriment de la généralisation du résultat trouvé aux images complètes.

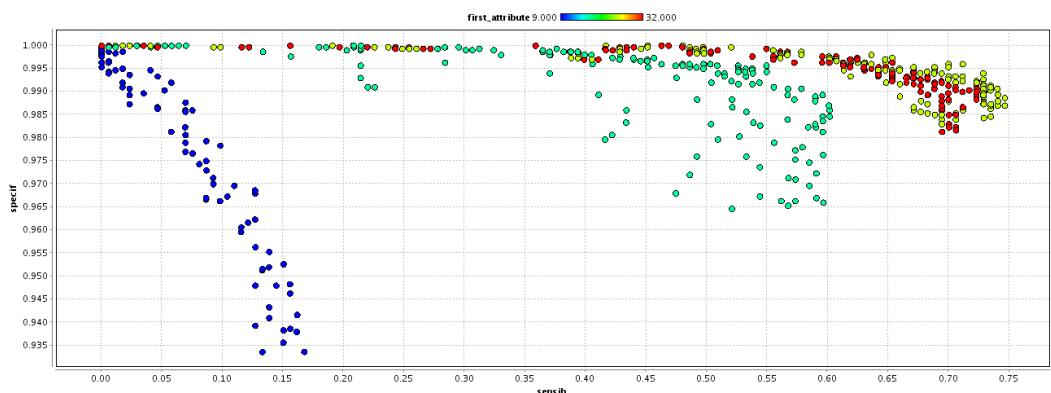
Il est intéressant de constater que les performances en sensibilité pour les bases Appauvrie, Témoin et Enrichie sont inversement proportionnels à la taille de la base. En effet, plus la complexité de la base est importante, plus il devient difficile de trouver une surface de séparation efficace. Cependant, une base trop simpliste va engendrer une solution qui sera sans rapport avec la réalité, comme nous le verrons plus tard avec les courbes F-ROC.

Les valeurs de spécificité sont toutes très supérieures à 99%, ce qui montre que le classifieur n'a pas de problèmes pour classer correctement les points sains. En effet, la base étant très déséquilibrée, avec un rapport de 17 points sains pour 1 point "lésion" dans la base témoin, il est normal il est normal que le classifieur favorise la classification des points "sain". Il existe des techniques permettant de compenser ce déséquilibre lors de l'apprentissage, mais tous les tests que nous avons réalisés n'ont pas montré d'amélioration du résultat. En effet, l'étape de sélection du seuil (voir section suivante) permet de compenser ces différences.

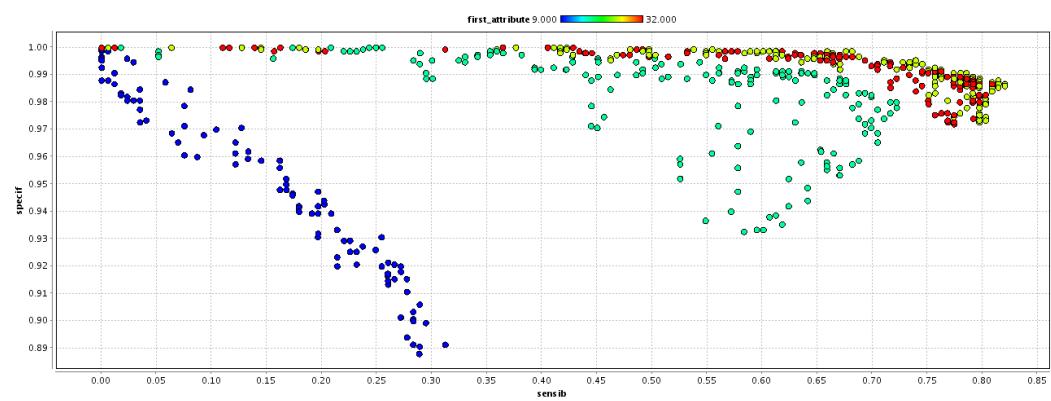
Il est intéressant d'observer que les fronts de pareto de la base Témoin 2 sont très semblables à ceux de la base Témoin pour une décomposition au troisième niveau. Des disparités sont apparues Cela semble indiquer que la base est stable avec un nombre de points suffisant.

a	Base Témoin	Base Érosion	Base appauvrie	Base enrichie	Base normalisée
C	464	74	5412	5412	10000
γ	0.0053	0.0094	0.00031	0.0017	0.052
j	3	3	3	4	3
Sensibilité	0.75	0.80	0.82	0.60	0.76
Spécificité	0.99	0.99	0.99	0.99	0.99
Précision	0.98	0.98	0.97	0.99	0.98

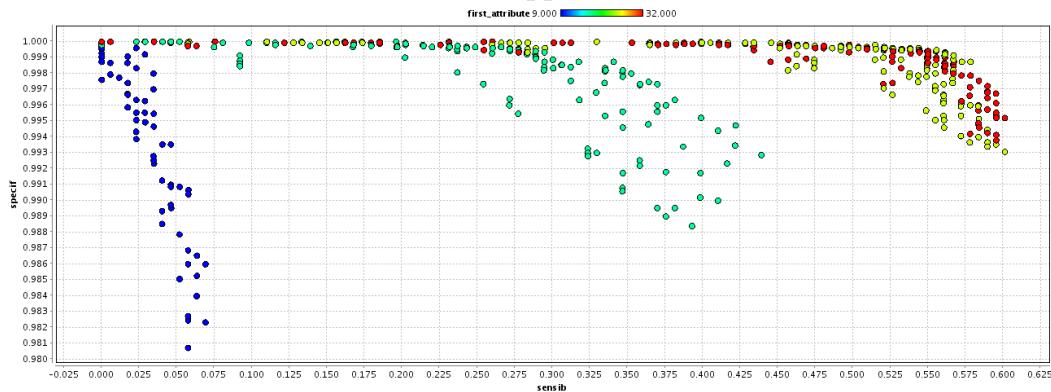
TAB. 12.1: Paramètres sélectionnés pour l'optimisation des performances. Sont indiqués pour chaque base le triplet de paramètres sélectionné ainsi que sa position sur le front de pareto.



a) Base Témoin

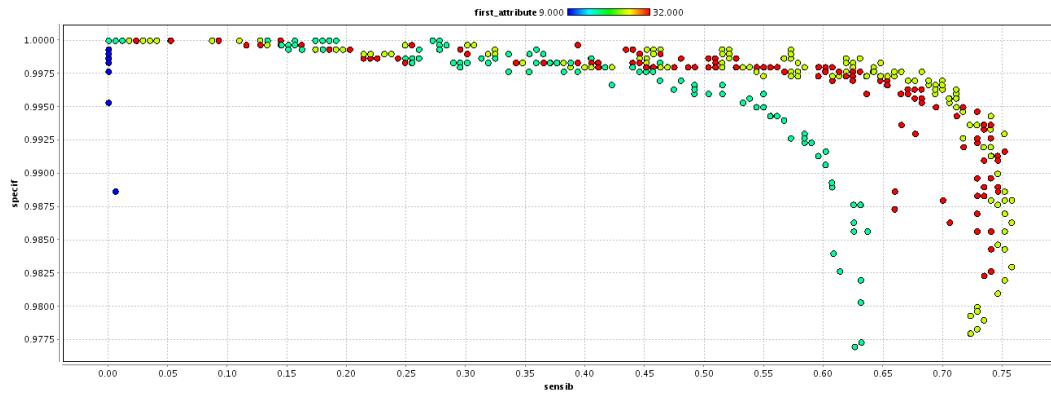


b) Base appauvrie

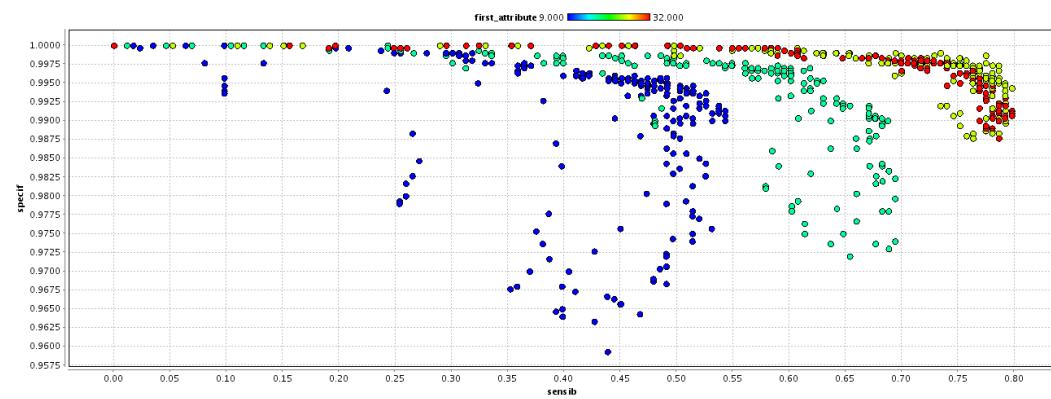


c) Base enrichie

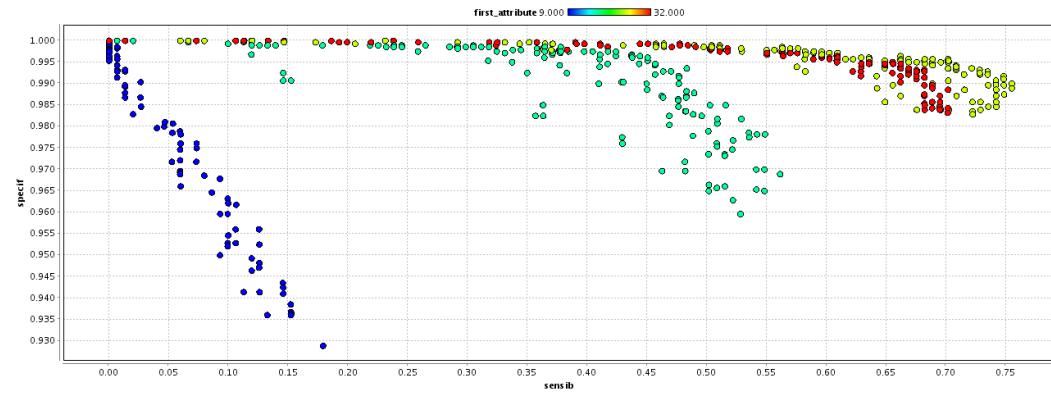
FIG. 12.1: Fronts de pareto des résultats de la recherche des meilleurs paramètres du classifieur (1/2). Pour chaque triplet de paramètres (C, γ, j), la sensibilité et la spécificité sont reportées sur le graphique. Le code couleur correspond à la valeur de j . En a), la base témoin, avec 200 points négatifs par image et une normalisation moyenne, en b) la base appauvrie avec 100 points négatifs par image et une normalisation moyenne, et en c) la base enrichie avec 1000 points négatifs par image et une normalisation moyenne.



a) Base Normalisée -1/+1



b) Base Érodée



c) Base Témoin 2

FIG. 12.2: Fronts de pareto des résultats de la recherche des meilleurs paramètres du classifieur (2/2). Pour chaque triplet de paramètres (C, γ, j) , la sensibilité et la spécificité sont reportées sur le graphique. Le code couleur correspond à la valeur de j . En a) la base normalisation avec 200 points négatifs par image et une normalisation +1/-1. En b), la base érodée, avec 200 points négatifs par image et une normalisation moyenne, en c) la base Témoin 2, réalisée de la même manière que la base témoin mais en retirant une image..

12.1.1 Courbe Free-ROC

Les courbes Free-ROC de la figure 12.1.1 permettent de comparer les performances du CAD sur les différentes bases d'apprentissage. Les courbes ont volontairement été tronquées à 40 faux positifs par image, car ce nombre est déjà trop important pour un système CAD.

On peut observer que les courbes correspondants aux bases Témoin et Enrichie atteignent leur maximum de performances pour un nombre de faux positifs relativement faible par rapport aux autres bases : entre 17 et 20 faux positifs pour ces bases, contre plus de 40 pour les bases Appauvrie et Érosion. Cela tend à montrer que le système CAD est plus performant pour ces bases car il crée moins d'agrégats là où il n'y a pas de lésions.

En ce qui concerne la sensibilité maximale obtenue sur les courbes, elle est atteinte pour la base témoin avec environ 62% de sensibilité, suivie par la base appauvrie avec 60%, mais pour un nombre de faux positifs beaucoup plus important (38 contre 18 pour la base Témoin). La troisième courbe est la courbe Normalisation, suivie par la base enrichie puis la base Érosion. Il est important de noter que les sensibilités maximales observées sont très proches, entre 55% et 62%, ce qui indique que la qualité de la base d'apprentissage n'a pas d'impact réel sur la sensibilité maximale atteinte par le CAD, mais qu'il pourra être plus ou moins difficile pour ce dernier de différencier les lésions du bruit de fond.

En pratique, on choisira le seuil pour avoir la certitude d'avoir un nombre de faux positifs "raisonnable" par image. Dans notre cas, les images contiennent environ 10 lésions par image. Il peut être intéressant de comparer les performances des bases pour ratio de 1 faux positif par lésion, soit pour 10 faux positifs. Dans ce cas, la base Témoin est au coude à coude avec la base Enrichie, à environ 55%, ce qui est déjà très proche de leur performances maximales. La base Normalisation et la base Appauvrie sont quand à elles à 40% de sensibilité, tandis que la base Érosion atteint 35%.

Dans tous les cas, nous avons donc vu que le maximum de performances est apporté par la base Témoin.

12.1.2 Comparaison des performances JAFROC

La comparaison des performances obtenues par l'algorithme JAFROC [Chakraborty and Winter, 1999] de la figure 12.4 nous montre les FDM (Figure de Mérite) obtenues pour les différentes bases. Les FDM des bases Témoin, Érosion et Enrichie sont quasiment au même niveau à 0.18, mais les barres d'erreurs semblent montrer un léger avantage pour la base Témoin.

Les bases Normalisation et Appauvrie quand à elles ont une FDM de 0.1 environ, ce qui indique une performance plus faible que les autres.

D'un point de vue statistique, La p-value fournie par le logiciel est de 0.049, ce qui ne

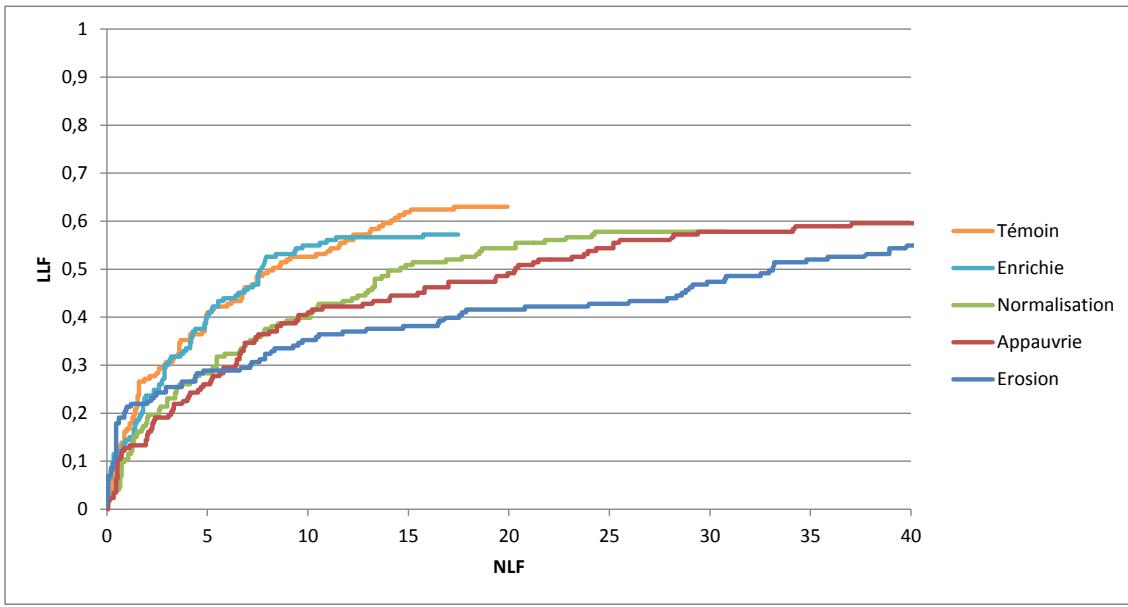


FIG. 12.3: Courbe Free-ROC comparant les performances du CAD sur une base témoin (normalisation moyenne et 200 points négatifs par image), sur une base enrichie (1000 points négatifs par image), sur une base appauvrie (100 points négatifs par image), sur une base normalisée différemment (normalisation entre -1 et +1 et 200 points négatifs par image) et enfin sur une base de 100 points négatifs par image mais dont les volumes ont été érodés de 2 voxels.

permet pas de pouvoir annoncer avec une fiabilité de 95% que le test est significatif, c-à-d que les FDM sont effectivement toutes différentes. Cela se vérifie aisément en regardant l'étendue des barres d'erreur. Mais il faut noter que cette FDM est basée sur une méthode avec une puissance statistique faible, ce qui signifie qu'elle sous-estime la p-value.

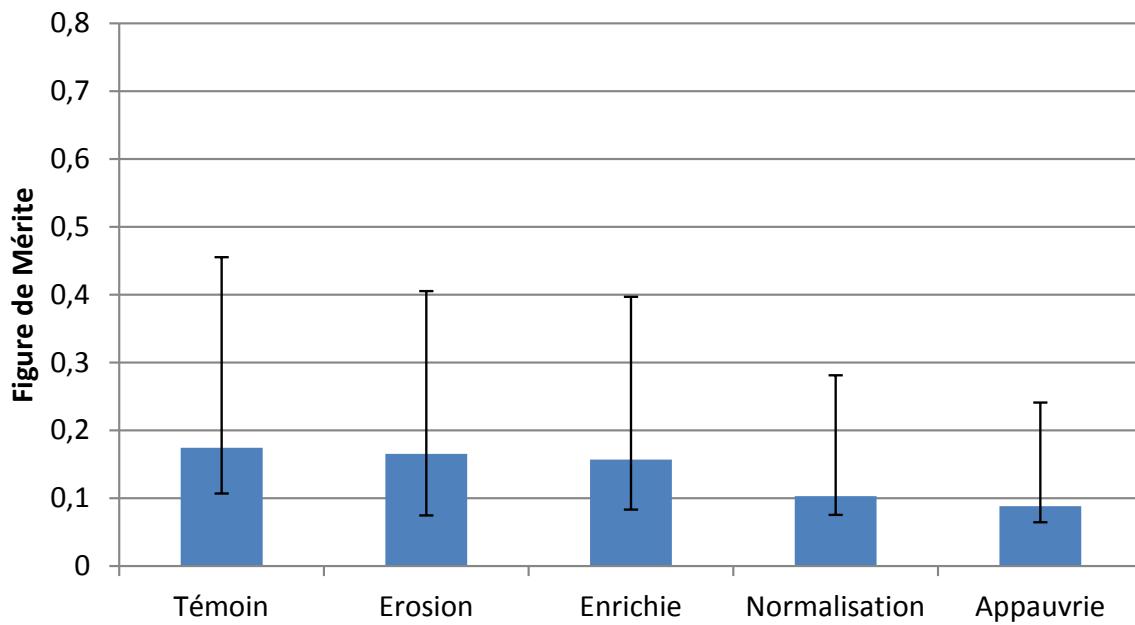


FIG. 12.4: Les FOM (Figure de Mérite) obtenues pour les différents paramètres

12.2 Comparaison des performances des différentes méthodes sur le Poumon

Les paramètres de génération de la base d'apprentissage retenus sont ceux de la base Témoin. Ils correspondent aux choix suivants :

- 200 points tirés aléatoirement dans le volume complet du poumon de chaque image (hors tumeurs)
- normalisation par neutralisation de la moyenne et de la variance , tels que $\mu, \sigma) = (1, 1)$

Quatre modalités seront comparées. ET-Static correspond aux images de la “vérité terrain”, à savoir des images sans mouvement respiratoire. Elle doit donner la performance haute. ET-NoCorr à l'inverse correspond aux images simulées avec mouvement respiratoire mais reconstruites sans aucune correction de mouvement. Elle représente la marge basse. ET-IM correspond aux images reconstruites avec la correction de mouvement post-reconstruction, et ET-LOR aux images reconstruites avec correction de mouvement pendant la reconstruction.

Sélection des meilleurs paramètres du classifieur

De la même manière que pour la sélection de la méthode de génération de la base d'apprentissage, il faut adapter les paramètres du classifieur aux différentes bases d'apprentissage.

La figure 12.5 montre les nuages de points associés aux différents triplets de paramètres (C, γ, j) pour chaque modalité. La référence sera toujours la base Témoin de la figure 12.1.a, car elle correspond à la vérité terrain.

Il est étonnant de constater que la base 12.1.a, qui correspond à la modalité statique, ou vérité terrain, offre les pires performances pour la décomposition de niveau 1 (17% de sensibilité) de toutes les modalités, avec la base non corrigée. La mauvaise performance de cette base d'apprentissage se retrouve pour tous les niveaux de décomposition. A l'inverse, tous les nuages de points ont la même répartition que ceux de la base Statique, par opposition aux distributions des bases érodées ou normalisées vues dans la partie précédente.

La modalité ayant les meilleures performances pour la décomposition de niveau 2 est ET-IM, avec plus de 70% de sensibilité, comparés aux 60% de la base Statique.

Les meilleures performances du CAD sont atteintes pour le troisième niveau de décomposition pour toutes les modalités sauf ET-LOR, avec une sensibilité d'environ 81 à 84% excepté pour ET-Static avec 75%, confirmant la tendance observée au premier niveau de décomposition.

La spécificité observée pour les meilleurs jeux de paramètres est sensiblement la même quelquesoient la modalité, entre 99% et 99.5%.

12.2.1 Courbes Free-ROC

Les courbes F-ROC obtenues sur les différentes modalités sont présentées dans la figure 12.7.

Toutes les courbes ont une NLF (Nombre de faux positifs moyen par image) à peu près équivalente entre 20 et 24, excepté ET-LOR avec une NLF de 34. Cependant, les performances en LLF (proportion des lésions localisées) de ET-LOR sont constantes à partir d'une LNF de 20 environ.

L'ordre des courbes est constant pour tous les NLF à partir de 5 faux positifs par image. Au delà de cette limite, les performances des images statiques sont supérieures à celles de ET-IM de 1 à 5%, tandis que ET-IM a des performances supérieures de 3 à 10% à celles des images non Corrigées et de ET-LOR. Ces deux dernières modalités ont des performances quasiment identiques.

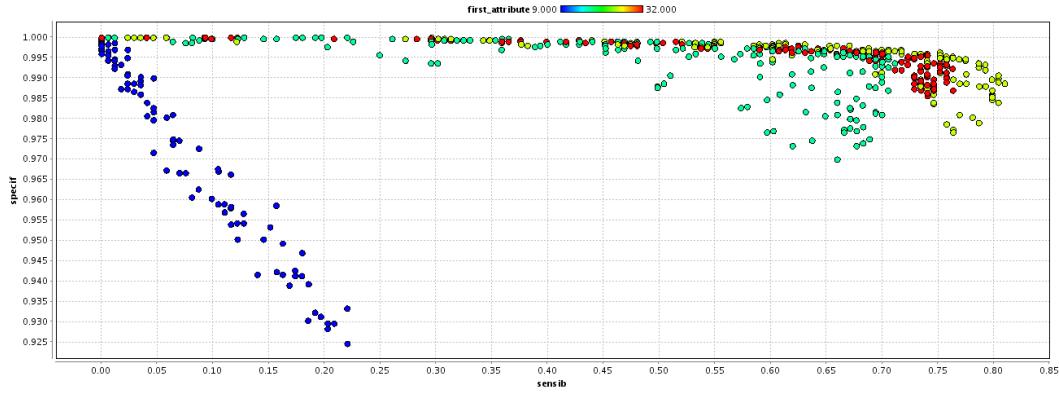
Cependant, en dessous de 5 faux positifs par image, les courbes sont trop proches pour pouvoir en déduire une tendance.

Dans sa globalité, le graphique montre que le maximum de performances est apporté par la base Statique statiques, suivi par la base ET-IM. Les bases ET-LOR et Non corrigées sont très proches l'une de l'autre mais nettement en dessous des deux premières en terme de LLF à NLF égale.

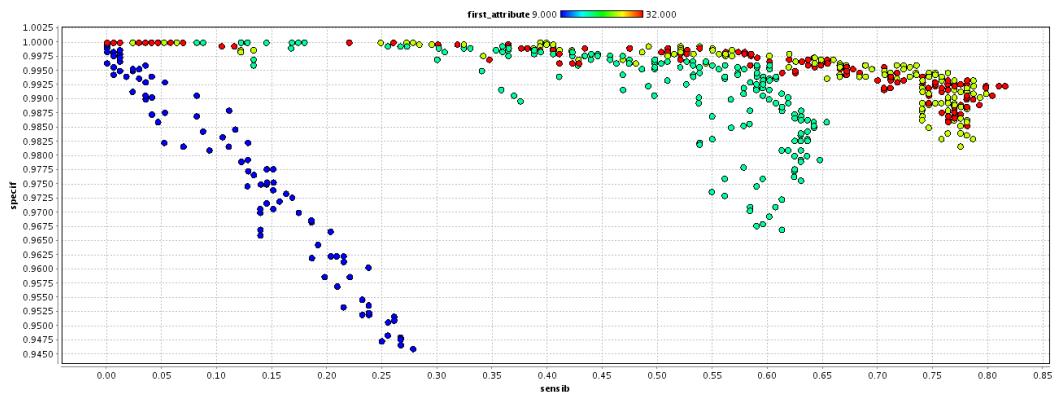
12.2.2 Comparaison des performances JAFROC

Je vais comparer les différentes modalités à partir du résultat de l'analyse JAFROC.

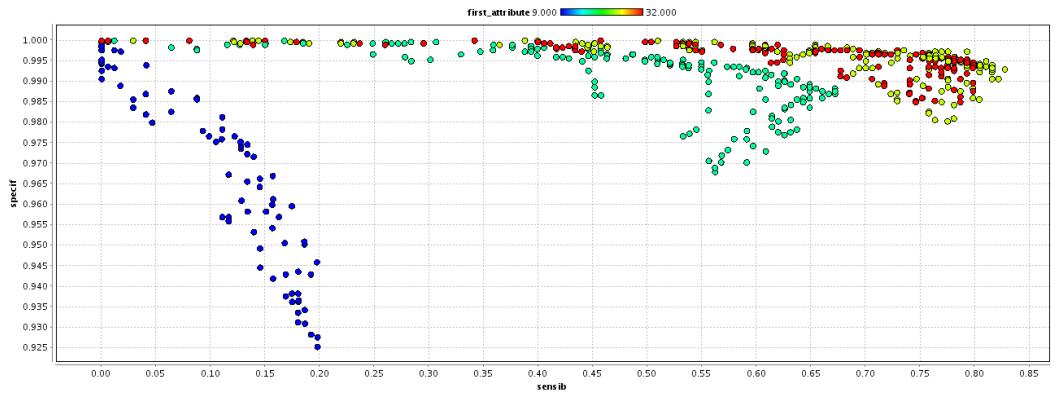
La p-value est de 0.10, ce qui ne permet pas de déclarer que statistiquement les données sont différentes : 12.8.



a) ET-IM



b) ET-LOR



c) ET-NoCorr

FIG. 12.5: Fronts de pareto des résultats de la recherche des meilleurs paramètres du classifieur pour les différentes modalités, avec 200 points négatifs par image. Pour chaque triplet de paramètres (C, γ, j) , la sensibilité et la spécificité sont reportées sur le graphique. Le code couleur correspond à la valeur de j . a) représente la correction d'image ET-IM, b) les images non corrigées du mouvement, et c) les images corrigées par la méthode LOR.

a	Base Statique	Base IM	Base LOR	Base NoCorr
C	464	10000	10000	10000
γ	0.0053	0.00097	0.00031	0.00055
j	3	3	4	3
Sensibilité	0.75	0.81	0.82	0.83
Spécificité	0.99	0.99	0.99	0.99
Précision	0.98	0.98	0.98	0.98

FIG. 12.6: Paramètres sélectionnés pour l'optimisation des performances du Poumon. Sont indiqués pour chaque base le triplet de paramètres sélectionné ainsi que sa position sur le front de pareto.

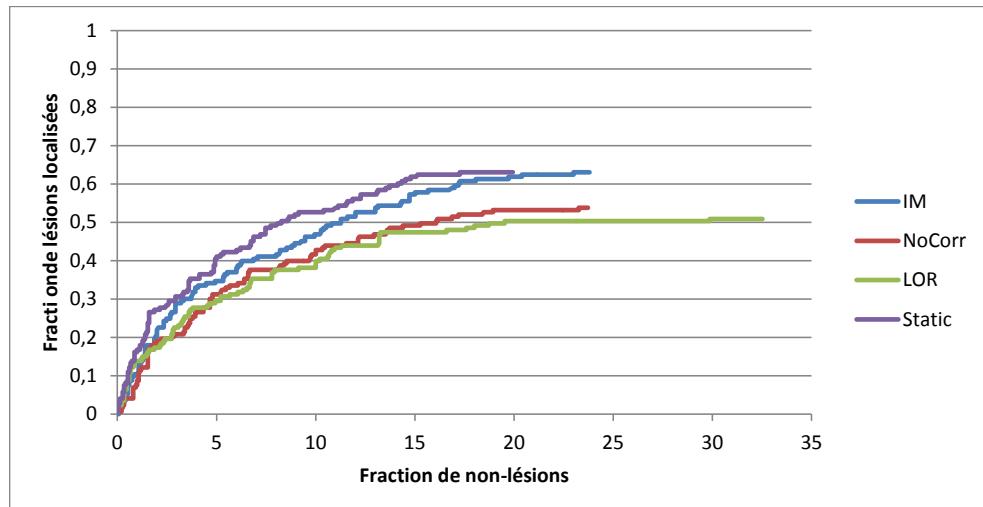


FIG. 12.7: Courbe Free-ROC comparant les performances du CAD selon les modalités de correction du mouvement respiratoire.

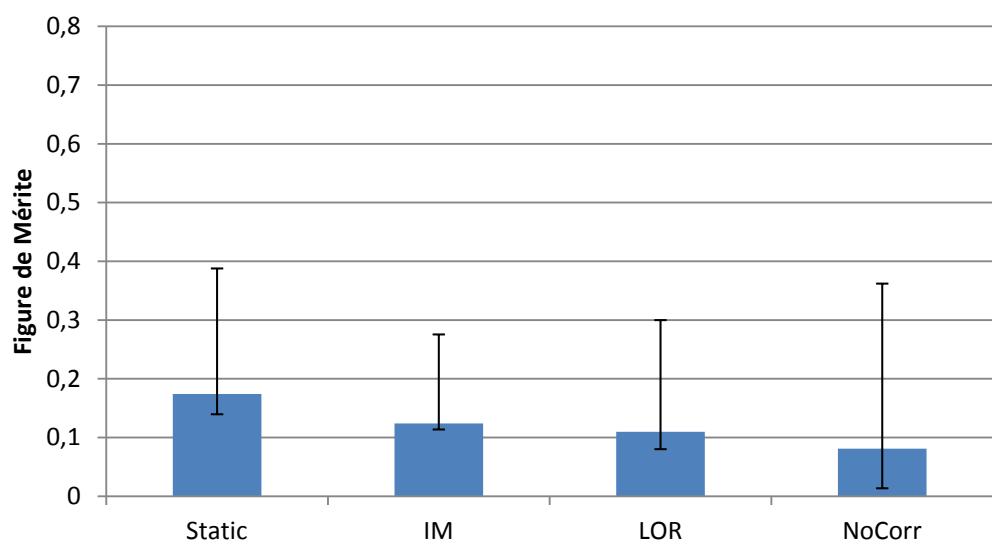


FIG. 12.8: Les FDM (Figure de Mérite) obtenues pour les différentes modalités.

12.3 Comparaison des performances des différentes méthodes Foie

Les caractéristiques utilisées pour obtenir ces résultats sont les suivants :

- 200 points tirés aléatoirement dans le volume de chaque image (hors tumeurs)
- normalisation par moyennage et neutralisation de la variance

12.3.1 Comparaison des performances JAFROC

La p-value est de 0.1, ce qui ne permet pas de déclarer que statistiquement les données sont différentes : 12.9

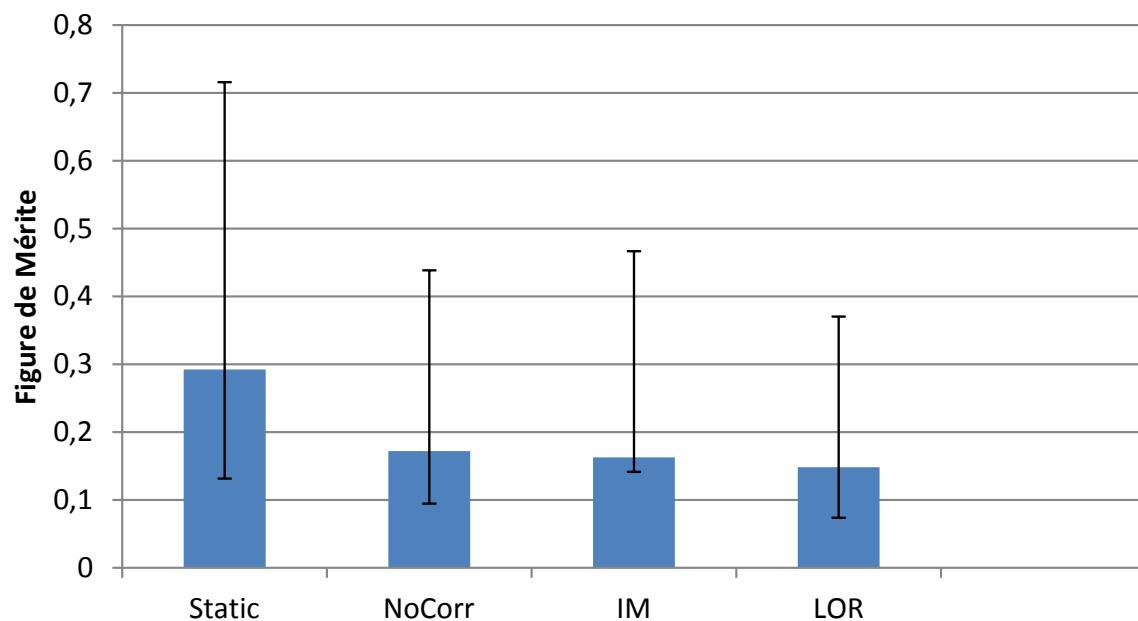


FIG. 12.9: Les FDM (Figure de Mérite) obtenues pour les différentes modalités.

12.3.2 Courbes Free-ROC

Voir figure 12.10. Le maximum de performances est apporté par les images statiques, suivi par les images ET-IM.

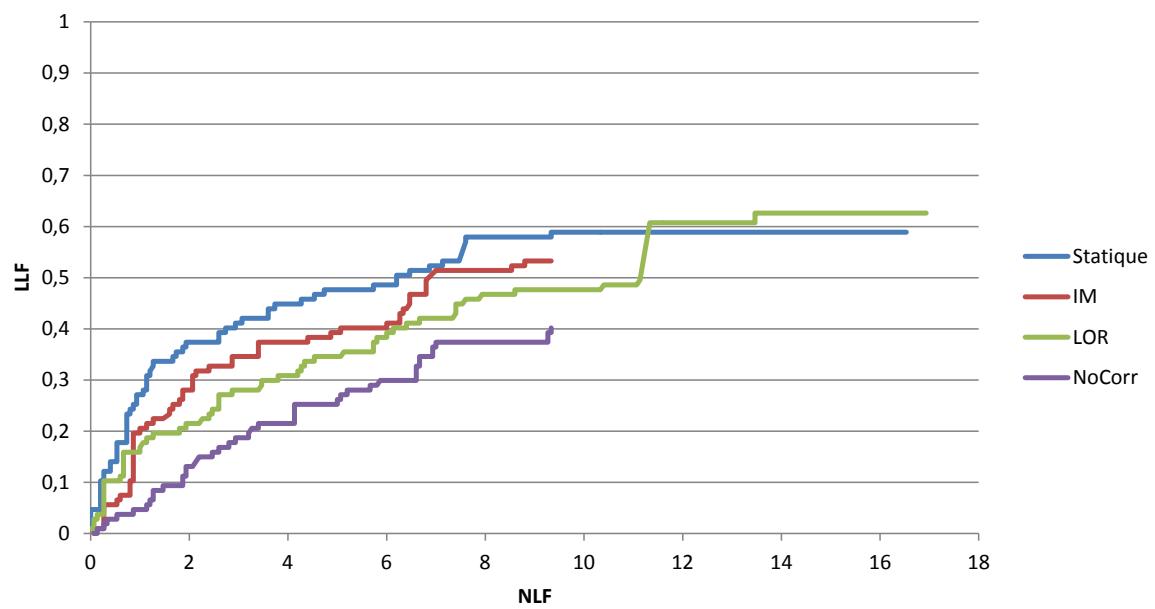


FIG. 12.10: Courbe Free-ROC comparant les performances du CAD selon les modalités de correction du mouvement respiratoire.

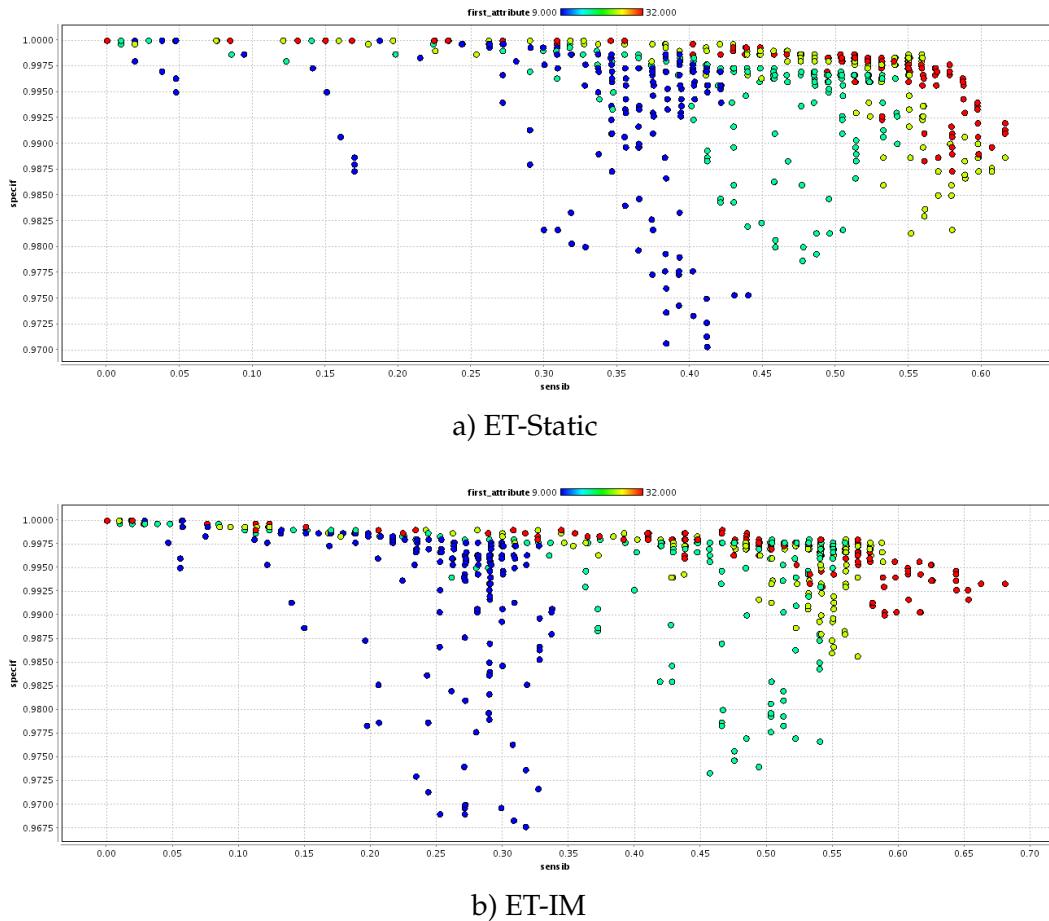
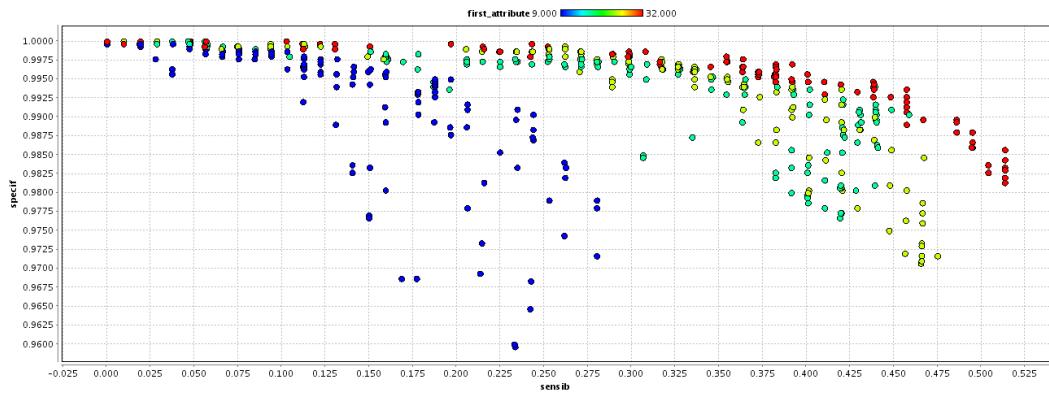
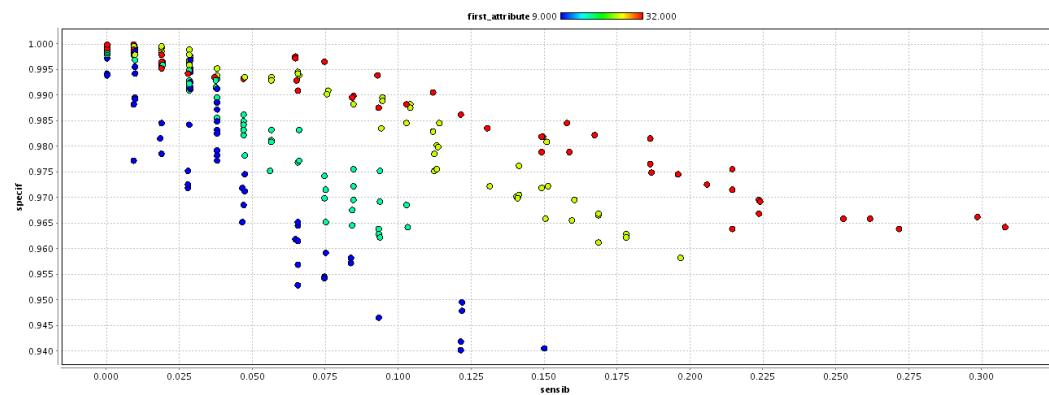


FIG. 12.11: Fronts de pareto des résultats de la recherche des meilleurs paramètres du classifieur pour les différentes modalités, avec 200 points négatifs par image. Pour chaque triplet de paramètres (C, γ, j) , la sensibilité et la spécificité sont reportées sur le graphique. Le code couleur correspond à la valeur de j . a) représente la correction d'image ET-Static, b) les images corrigées du mouvement post-reconstruction.



c) ET-LOR



d) ET-NoCorr

FIG. 12.12: Fronts de pareto des résultats de la recherche des meilleurs paramètres du classifieur pour les différentes modalités, avec 200 points négatifs par image. Pour chaque triplet de paramètres (C, γ, j), la sensibilité et la spécificité sont reportées sur le graphique. Le code couleur correspond à la valeur de j . c) représente la correction d'image pendant la reconstruction et d) les images non corrigées.

a	Base Statique	Base IM	Base LOR	Base NoCorr
C	858	5412	251	5412
γ	0.002	0.00055	0.0053	0.0017
j	4	4	4	4
Sensibilité	0.62	0.68	0.51	0.31
Spécificité	0.99	0.99	0.99	0.96
Précision	0.98	0.98	0.97	0.94

FIG. 12.13: Paramètres sélectionnés pour l'optimisation des performances du Foie. Sont indiqués pour chaque base le triplet de paramètres sélectionné ainsi que sa position sur le front de pareto.

Septième partie

Conclusion

Huitième partie

Bibliographie

Bibliographie

- [Bai and Brady, 2009] Bai, W. and Brady, M. (2009). Regularized B-spline deformable registration for respiratory motion correction in PET images. *Physics in Medicine and Biology*, 54 :2719.
- [Barrett et al., 1993] Barrett, H., Yao, J., Rolland, J., and Myers, K. (1993). Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21) :9758.
- [Bloomfield et al., 2003] Bloomfield, P., Spinks, T., Reed, J., Schnorr, L., Westrip, A., Livieratos, L., Fulton, R., and Jones, T. (2003). The design and implementation of a motion correction scheme for neurological PET. *Physics in Medicine and Biology*, 48 :959.
- [Bordessoule et al., 2006] Bordessoule, D., Jaccard, A., Chaury, M., Desfougères, M., and Monteil, J. (2006). Impact de la tep dans la prise en charge des lymphomes malins hodgkiniens et non hodgkiniens. *Oncologie*, 8(4) :322–330.
- [Boser et al., 1992] Boser, B., Guyon, I., and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM.
- [Boucher et al., 2004] Boucher, L., Rodrigue, S., Lecomte, R., and Benard, F. (2004). Respiratory gating for 3-dimensional PET of the thorax : feasibility and initial results. *Journal of Nuclear Medicine*, 45(2) :214.
- [Bunch et al., 1978] Bunch, P., Hamilton, J., Sanderson, G., and Simmons, A. (1978). A free-response approach to the measurement and characterization of radiographic-observer performance. *J. Appl. Photogr. Eng*, 4(4) :166–171.
- [Bundschuh et al., 2007] Bundschuh, R., Martinez-Moeller, A., Essler, M., Martinez, M., Nekolla, S., Ziegler, S., and Schwaiger, M. (2007). Postacquisition detection of tumor motion in the lung and upper abdomen using list-mode PET data : a feasibility study. *Journal of Nuclear Medicine*, 48(5) :758.
- [Chakraborty and Berbaum, 2004] Chakraborty, D. and Berbaum, K. (2004). Observer studies involving detection and localization : modeling, analysis, and validation. *Medical physics*, 31 :2313.
- [Chakraborty and Winter, 1990] Chakraborty, D. and Winter, L. (1990). Free-response methodology : alternate analysis and a new observer-performance experiment. *Radiology*, 174(3) :873.

- [Chan et al., 1987] Chan, H., Doi, K., Galhotra, S., Vyborny, C., MacMahon, H., and Jokich, P. (1987). Image feature analysis and computer-aided diagnosis in digital radiography. i. automated detection of microcalcifications in mammography. *Medical Physics*, 14 :538.
- [Chang et al., 2010] Chang, G., Chang, T., Pan, T., Clark, John W., J., and Mawlawi, O. R. (2010). Implementation of an Automated Respiratory Amplitude Gating Technique for PET/CT : Clinical Evaluation. *J Nucl Med*, 51(1) :16–24.
- [Dawood et al., 2008] Dawood, M., Buther, F., Jiang, X., and Schafers, K. (2008). Respiratory motion correction in 3-d pet data with advanced optical flow algorithms. *Medical Imaging, IEEE Transactions on*, 27(8) :1164 –1175.
- [Dawood et al., 2006] Dawood, M., Lang, N., Jiang, X., and Schafers, K. (2006). Lung motion correction on respiratory gated 3-D PET/CT images. *IEEE transactions on medical imaging*, 25(4) :476.
- [Detorie and Dahlbom, 2008] Detorie, N. C. and Dahlbom, M. (2008). Quantitative evaluation of reconstruction algorithms for motion compensated pet. In *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pages 5366 –5372.
- [Dimitrakopoulou-Strauss et al., 2002] Dimitrakopoulou-Strauss, A., Strauss, L., Heichel, T., Wu, H., Burger, C., Bernd, L., and Ewerbeck, V. (2002). The role of quantitative 18F-FDG PET studies for the differentiation of malignant and benign bone lesions. *Journal of Nuclear Medicine*, 43(4) :510.
- [El Naqa et al., 2006] El Naqa, I., Low, D., Bradley, J., Vicic, M., and Deasy, J. (2006). Deblurring of breathing motion artifacts in thoracic PET images by deconvolution methods. *Medical physics*, 33 :3587.
- [Erdi et al., 2004] Erdi, Y., Nehmeh, S., Pan, T., Pevsner, A., Rosenzweig, K., Mageras, G., Yorke, E., Schoder, H., Hsiao, W., Squire, O., et al. (2004). The CT motion quantitation of lung lesions and its impact on PET-measured SUVs. *Journal of Nuclear Medicine*, 45(8) :1287.
- [Farquhar et al., 1999] Farquhar, T., Llacer, J., Hoh, C., Czernin, J., Gambhir, S., Seltzer, M., Silverman, D., Qi, J., Hsu, C., and Hoffman, E. (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.
- [Fayad et al., 2010] Fayad, H., Pan, T., Roux, C., and Visvikis, D. (2010). Application of a generic motion model for pet respiratory motion correction. *Journal of Nuclear Medicine*, 51(2) :524.
- [Genestie et al., 1998] Genestie, C., Zafrani, B., Asselain, B., Fourquet, A., Rozan, S., Validire, P., Vincent-Salomon, A., and Sastre-Garau, X. (1998). Comparison of the prognostic value of scarff-bloom-richardson and nottingham histological grades in a series of 825 cases of breast cancer : major importance of the mitotic count as a component of both grading systems. *Anticancer research*, 18(1B) :571–576.

- [Guivarc'h et al., 2004] Guivarc'h, O., Turzo, A., Visvikis, D., and Bizais, Y. (2004). Synchronization of pulmonary scintigraphy by respiratory flow and by impedance plethysmography. In *Proceedings of SPIE*, volume 5370, page 1166.
- [Hudson, 1994] Hudson, H. (1994). Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 13(4).
- [Jan et al., 2004] Jan, S., Santin, G., Strul, D., Staelens, S., Assie, K., Autret, D., Avner, S., Barbier, R., Bardies, M., Bloomfield, P., et al. (2004). GATE : a simulation toolkit for PET and SPECT. *Physics in medicine and biology*, 49 :4543.
- [Klein and Huesman, 1997] Klein, G. and Huesman, R. (1997). A 3D optical flow approach to addition of deformable PET volumes. *nam*, page 0136.
- [Klein et al., 2001] Klein, G., Reutter, R., and Huesman, R. (2001). Four-dimensional affine registration models for respiratory-gated pet. *Nuclear Science, IEEE Transactions on*, 48(3) :756–760.
- [Krak et al., 2005] Krak, N., Boellaard, R., Hoekstra, O., Twisk, J., Hoekstra, C., and Lammeretsma, A. (2005). Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *European journal of nuclear medicine and molecular imaging*, 32(3) :294–301.
- [Lamare et al., 2007a] Lamare, F., Cresson, T., Savean, J., Cheze Le Rest, C., Reader, A., and Visvikis, D. (2007a). Respiratory motion correction using affine transformation of list mode data in PET. *Physics in medicine and biology*, 52 :121–140.
- [Lamare et al., 2007b] Lamare, F., Ledesma Carbayo, M., Cresson, T., Kontaxakis, G., Santos, A., Le Rest, C., Reader, A., and Visvikis, D. (2007b). List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations. *Physics in medicine and biology*, 52(17) :5187–5204.
- [Lartizien et al., 2010] Lartizien, C., Marache-Francisco, S., Le Pogam, A., Tomei, S., and Visvikis, D. (2010). Impact of partial volume correction in whole-body PET imaging : A computer-aided detection study. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE*, pages 3214–3217. IEEE.
- [Livieratos et al., 2005] Livieratos, L., Stegger, L., Bloomfield, P., Schafers, K., Bailey, D., and Camici, P. (2005). Rigid-body transformation of list-mode projection data for respiratory motion correction in cardiac PET. *Physics in Medicine and Biology*, 50 :3313.
- [Metz, 1986] Metz, C. (1986). Roc methodology in radiologic imaging. *Investigative Radiology*, 21(9) :720.
- [Mierswa et al., 2006] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T. (2006). Yale : Rapid prototyping for complex data mining tasks. In Ungar, L., Craven, M., Gunopulos, D., and Eliassi-Rad, T., editors, *KDD '06 : Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940, New York, NY, USA. ACM.

- [Nehmeh et al., 2002] Nehmeh, S., Erdi, Y., Ling, C., Rosenzweig, K., Schoder, H., Larson, S., Macapinlac, H., Squire, O., and Humm, J. (2002). Effect of respiratory gating on quantifying PET images of lung cancer. *Journal of Nuclear Medicine*, 43(7) :876.
- [Nehmeh et al., 2003] Nehmeh, S., Erdi, Y., Rosenzweig, K., Schoder, H., Larson, S., Squire, O., and Humm, J. (2003). Reduction of respiratory motion artifacts in PET imaging of lung cancer by respiratory correlated dynamic PET : methodology and comparison with respiratory gated PET. *Journal of Nuclear Medicine*, 44(10) :1644.
- [Nie et al., 2006] Nie, Y., Li, Q., Li, F., Pu, Y., Appelbaum, D., and Doi, K. (July 2006). Integrating pet and ct information to improve diagnostic accuracy for lung nodules : A semiautomatic computer-aided method. *Journal of Nuclear Medicine*, 47(7) :1075–1080.
- [Qiao et al., 2006] Qiao, F., Pan, T., John Jr, W., and Mawlawi, O. (2006). A motion-incorporated reconstruction method for gated PET studies. *Physics in Medicine and Biology*, 51 :3769.
- [Rahmim et al.,] Rahmim, A., Tang, J., Ay, M., and Bengel, F. 4D Respiratory Motion-Corrected Rb-82 Myocardial Perfusion PET Image Reconstruction.
- [Reader et al., 2002] Reader, A., Ally, S., Bakatselos, F., Manavaki, R., Walledge, R., Jeavons, A., Julyan, P., Zhao, S., Hastings, D., and Zweit, J. (2002). One-pass list-mode em algorithm for high-resolution 3-d pet image reconstruction into large arrays. *Nuclear Science, IEEE Transactions on*, 49(3) :693 – 699.
- [Segars, 2001] Segars, W. (2001). *Development and Application of the new dynamic nurbs-based cardiac-torso (NCAT) phantom*. PhD thesis.
- [servant laval, 2007] servant laval, A. (2007). Cours anatome fonctionnelle. Disponible sur Internet à l'adresse <http://www.chups.jussieu.fr/polysPSM/anatfonctPSM2/poly/POLY.Chp.8.html>.
- [Sherwood, 2006] Sherwood, L. (2006). *Fundamentals of physiology : a human perspective*. Thomson Brooks/Cole.
- [Stephan et al., 2003] Stephan, C., Wesseling, S., Schink, T., and Jung, K. (2003). Comparison of eight computer programs for receiver-operating characteristic analysis. *Clin Chem*, 49(3) :433–439.
- [Swets and Pickett, 1982] Swets, J. and Pickett, R. (1982). *Evaluation of diagnostic systems : methods from signal detection theory*. Academic Press New York.
- [Thielemans et al., 2006] Thielemans, K., Manjeshwar, R., Tao, X., and Asma, E. (2006). Lesion detectability in motion compensated image reconstruction of respiratory gated pet/ct. In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 6, pages 3278 –3282.
- [Thorndyke et al., 2006] Thorndyke, B., Schreibmann, E., Koong, A., and Xing, L. (2006). Reducing respiratory motion artifacts in positron emission tomography through retrospective stacking. *Medical physics*, 33 :2632.

[Zou et al., 2007] Zou, K., O'Malley, A., and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5) :654.