

N° d'ordre : 2009-ISAL-0119

Année 2009

Thèse

Présentée devant
L'Institut National des Sciences Appliquées de Lyon

Pour obtenir
Le grade de docteur

Ecole doctorale :
Electronique, Electrotechnique et Automatique

Formation doctorale :
Sciences de l'Information, des Dispositifs et des Systèmes

Par

Sandrine TOMEI LE DIGARCHER
Ingénieur CPE Lyon 2006

Détection automatique de fixations radioactives dans les images TEP oncologiques

Soutenue le 10 décembre 2009 devant la Commission d'examen

Jury

Directrice de thèse	LARTIZIEN Carole	CR CNRS
Directeur de thèse	ODET Christophe	Professeur (INSA Lyon)
Rapporteur	VISVIKIS Dimitris	DR INSERM (LATIM, Brest)
Rapporteur	CANU Stéphane	Professeur (LITIS, Rouen)
Examineur	GIAMMARILE Francesco	Professeur (UCB Lyon 1)
Examineur	PROST Remy	Professeur (INSA Lyon)

Résumé

Le cancer fait partie des causes majeures de mortalité en France et touche une population de tout âge et de tout sexe. Les progrès actuels de la médecine, de la biologie et des techniques d'imagerie médicale permettent d'améliorer la prise en charge du patient et d'optimiser les traitements administrés. L'imagerie de Tomographie par Emission de Positons (TEP) permet de visualiser l'information fonctionnelle des cellules et notamment leur consommation en FDG, un analogue du glucose, marqué au fluor 18 (18F). Les cellules tumorales sont de grandes consommatrices de cette substance radioactive. Il n'est pas rare que la TEP soit plus sensible qu'une imagerie anatomique à la détection de tumeurs de petite taille. Ainsi, la TEP s'apprête à devenir un examen de référence dans le suivi pathologique de patients atteints de certains cancers. C'est notamment le cas du lymphome qui est une pathologie caractérisée par une dissémination de petites tumeurs dans tout le corps du patient.

Un ensemble de développements méthodologiques vise à améliorer l'exploitation de ce type d'images par le clinicien. L'optimisation de l'utilisation de l'imagerie TEP en clinique passe par une amélioration de la quantification des images observées, de l'instrumentation des caméras et par la mise en place d'un ensemble de méthodes optimisant l'interprétation des images provenant des examens cliniques. Malgré tous les efforts actuels pour faciliter l'utilisation de l'imagerie TEP d'un point de vue diagnostic, les performances de détection humaine propres à cette imagerie pourraient encore être améliorées dans le cas de certains cancers. De ces performances dépendent l'optimisation et la réadaptation des traitements conventionnels pour chaque cas et chaque patient, afin de limiter les risques de rechute ou de complications à long terme. Cette thèse est menée dans le cadre de cette optimisation et repose sur le développement d'un outil d'aide à la détection pour les images TEP 3D corps entier se focalisant plus spécifiquement sur les tumeurs difficilement visibles (faible diamètre et/ou faible contraste). L'objectif est de compléter le clinicien en lui fournissant une carte volumique de présence des tumeurs pour chaque image TEP observée.

Pour ce faire, nous avons suivi la méthodologie standardisée des systèmes de classification supervisée. Nous avons proposé une approche basée sur la combinaison de deux classifieurs, l'analyse discriminante linéaire (linear discriminant analysis ou LDA en anglais) et les séparateurs à vaste marge (SVM). Tous deux sont associés à des caractéristiques descriptives de l'image de type ondelettes. Chaque classifieur, préalablement entraîné, fournit une carte tridimensionnelle de score propre à chaque image. Ces scores, obtenus pour chaque voxel, quantifient leur probabilité de correspondre à une tumeur. Ces cartes sont ensuite seuillées à l'aide d'une méthode originale basée sur une analyse 3D par courbe FROC (Free Response Operating Characteristics). Elles produisent ainsi des cartes binaires de détection. Une étape de réduction de faux positifs (RFP) permet de nettoyer les cartes binaires provenant de chaque classifieur séparément. Celles-ci sont ensuite combinées et fournissent la sortie du système complet. Nous avons proposé une stratégie d'évaluation 3D du système encore peu utilisée dans une littérature surtout orientée vers l'étude de cas ou le 2D.

Afin de mesurer les performances de notre système, nous avons utilisé des images simulées. Celles-ci garantissent la connaissance indispensable des caractéristiques des tumeurs cibles. Deux jeux de données de réalisme croissant ont servi à l'évaluation quantitative et qualitative du système. La première base d'images simulées avait été définie avant la thèse à partir du fantôme mathématique MCAT et calibrée dans le cadre de l'optimisation des protocoles d'imagerie TEP corps entier pour la détection des hyperfixations. La seconde base d'images a été développée dans le cadre de cette thèse et présente cent volumes TEP corps entier obtenus par simulation Monte Carlo à partir du fantôme anthropomorphique de Zubal. Elle contient 50 cas sains et 50 cas pathologiques présentant un ensemble de 375 tumeurs sphériques de diamètre et contraste calibrés. Le réalisme de ces images a été évalué à l'aide d'index statistiques et cliniques calculées sur les images simulées et cliniques.

Les résultats obtenus à partir des volumes binaires provenant du système de détection développé se montrent prometteurs. Ils ouvrent la perspective d'enrichir l'information généralement fournie au clinicien par ce genre de système par l'intermédiaire d'une carte paramétrique quantifiant la pertinence de chaque tumeur détectée.

Table des matières

Table des figures	6
Liste des tableaux	10
Liste des abréviations	12
Introduction générale.....	13
Chapitre 1 Contexte	15
1.1 Introduction	15
1.2 Principe d'utilisation de la tomographie par émission de positons (TEP) en oncologie	15
1.2.1 Introduction	15
1.2.2 Principes et spécificités de l'imagerie TEP	15
1.2.3 Intérêt et spécificité du FDG en oncologie.....	17
1.2.4 Quantification d'un examen TEP au 18F-FDG.....	19
1.2.5 Acquisition corps entier	20
1.3 Performances de la TEP en oncologie et comparaison aux autres modalités	21
1.3.1 Introduction	21
1.3.2 Le diagnostic initial et la stadification	21
1.3.3 Le suivi au cours de la thérapie et la réévaluation du stade	21
1.3.4 Le suivi après traitement	22
1.3.5 Conclusions	22
1.4 Le cas particulier du lymphome	23
1.4.1 Description médicale de la pathologie	23
1.4.2 Statistiques et différents stades de la maladie	25
1.4.3 La TEP pour le lymphome : avantages	26
1.4.4 La TEP pour le lymphome : limitations et besoins	28
1.5 Détection en imagerie TEP oncologique.....	30
1.5.1 Les verrous de la TEP au FDG en oncologie	30
Chapitre 2 Les systèmes d'aide à la détection en imagerie médicale	32
2.1 Introduction	32
2.2 Les CAD en imagerie TEP : une application en développement.....	32
2.2.1 Introduction	32
2.2.2 Définitions préliminaires.....	33
2.2.3 Les systèmes CAD utilisant uniquement l'information TEP.....	33
2.2.4 Les CAD combinant l'information anatomique et fonctionnelle.....	35
2.2.5 Conclusions	36
2.3 Classification supervisée ou non supervisée ?	36
2.3.1 Introduction	36
2.3.2 La classification non supervisée.....	36
2.3.3 La classification supervisée.....	37
2.3.4 Conclusions	38
2.4 Les CAD supervisés : un schéma méthodologique standardisé.....	38
2.4.1 Introduction	38
2.4.2 Les différentes étapes d'un système CAD supervisé	38
2.4.3 Extraction initiale des tumeurs potentielles.....	39
2.4.4 Réduction de faux positifs (RFP).....	50
2.5 Evaluation des performances des systèmes CAD	54
2.5.1 Introduction	54

2.5.2	Comptage des vrais positifs (VP) et faux positifs (FP)	54
2.5.3	Indices de performances	55
2.5.4	Sélection des bases d'images d'apprentissage et de test	57
2.5.5	Conclusions	58
2.6	Conclusions de l'état de l'art.....	59
Chapitre 3	Choix méthodologiques pour le système CAD proposé	60
3.1	Introduction	60
3.2	Motivations.....	60
3.3	Choix méthodologiques.....	61
3.3.1	Les images TEP utilisées.....	61
3.3.2	Choix des caractéristiques	61
3.3.3	Choix du classifieur.....	62
Chapitre 4	Simulation d'une base d'images TEP pour l'oncologie : oncoPET_DB	63
4.1	Introduction	63
4.2	La simulation TEP Monte Carlo	63
4.3	Les bases d'images simulées TEP existantes	64
4.4	Conclusions	66
4.5	Génération de la base d'images simulées : description du modèle de distribution de 18F-FDG	66
4.5.1	Méthode générale de simulation d'images TEP	66
4.5.2	Choix du fantôme numérique	67
4.5.3	Génération d'un modèle d'activité de 18F-FDG.....	68
4.5.4	Génération du modèle de tumeurs : extension, taille et contraste	69
4.5.5	De la simulation à la reconstruction d'oncoPET_DB	75
4.6	La base d'images simulées résultante	76
4.6.1	Les images 'saines' d'oncoPET_DB.....	77
4.6.2	Les images 'pathologiques' d'oncoPET_DB.....	77
4.7	Validation de la base d'images simulées.....	78
4.7.1	Introduction	78
4.7.2	Mesures statistiques sur les activités de 18F-FDG.....	78
4.7.3	Mesures cliniques	80
4.8	Mise en ligne d'oncoPET_DB	82
4.9	Discussions.....	83
4.10	Conclusions sur le développement d'oncoPET_DB	84
Chapitre 5	Proposition d'un système CAD pour l'imagerie TEP corps entier	85
5.1	Introduction	85
5.2	Description du système CAD	86
5.2.1	Extraction initiale des tumeurs candidates	86
5.2.2	Réduction de faux positifs (RFP)	102
5.2.3	Combinaison de classifieurs	103
5.3	Optimisation des paramètres du système CAD	103
5.3.1	Introduction	103
5.3.2	Optimisation des paramètres intrinsèques du SVM	104
5.3.3	Choix des exemples utilisées pour l'apprentissage	105
5.3.4	Optimisation du vecteur de caractéristiques V	106
5.3.5	Optimisation de l'étape de RFP	107
5.4	Performances du système CAD.....	107
5.4.1	Introduction	107
5.4.2	Evaluation du système CAD complet	107
5.4.3	Evaluation du système CAD complet au regard des performances humaines	108

5.5	Performances du système CAD sur une base d'images simulées simplifiée	108
5.5.1	Introduction	108
5.5.2	Caractéristiques de la base d'images simulées utilisées.....	108
5.5.3	Optimisation des paramètres du CAD.....	109
5.5.4	Evaluation du système CAD final	119
5.5.5	Conclusions sur l'évaluation du système CAD par les données simulées simplifiées	124
Chapitre 6	Evaluation du système CAD sur la base d'images simulées oncoPET_DB ..	127
6.1	Introduction	127
6.2	Caractéristiques des images d'oncoPET_DB.....	127
6.3	Rappels sur le système CAD et méthodes d'évaluation.....	129
6.3.1	Rappels sur le système CAD	129
6.3.2	Méthodes d'évaluation du CAD sur les images d'oncoPET_DB	129
6.4	Evaluation du CAD sur les images d'oncoPET_DB.....	131
6.4.1	Exploitation et discussions des performances du CAD complet.....	131
6.4.2	Utilisation d'une information additionnelle de score pour la présentation des résultats	137
6.5	Conclusions sur l'évaluation par les données simulées réalistes	144
6.5.1	Conclusions et discussions	144
6.5.2	Perspectives	146
	Conclusions, discussions et perspectives	147
	ANNEXES	150
	Annexe A : L'imagerie TEP : principes physiques et limitations.....	151
	Annexe B : Rappels sur la théorie des ondelettes	160
	Annexe C : Description de la base d'images simulées simplifiée.....	162
	Bibliographie.....	164
	Bibliographie personnelle	183

Table des figures

Figure 1-1 : Les différentes étapes de la création d'une image TEP.....	16
Figure 1-2 : Les trois types d'acquisition, transmission, blanc et émission.....	16
Figure 1-3 : Les différentes étapes de la glycolyse	17
Figure 1-4 : Projection en maximum d'intensité (MIP) représentant la distribution normale du 18F-FDG dans l'organisme.....	18
Figure 1-5 : Protocole d'acquisition multi-lit corps entier.....	20
Figure 1-6 : Exemples d'images TEP au 18F-FDG provenant d'examens de patients atteints d'un cancer du poumon, a), b), d'un lymphome c), d), de cancer de la tête et du cou e) et d'un cancer de l'œsophage f).....	23
Figure 1-7 : Les principales composantes du système lymphatique humain	24
Figure 1-8 : Rôle des lymphocytes.....	25
Figure 1-9 : Les différents sites d'affections initiaux du lymphome illustrés sur une image TEP au 18F-FDG d'un patient atteint de ce cancer	26
Figure 1-10 : Suivi médical et utilisation de la TEP pour les patients atteints de lymphome..	27
Figure 1-11 : Exemples d'image TEP au 18F-FDG : patiente présentant un lymphome malin non hodgkinien de haut grade, a) image TEP avant traitement montrant des cibles hypermétaboliques médiastinales et sus-claviculaires gauches, b) après chimiothérapie montrant un aspect de réponse partielle. Autre exemple : patiente présentant une maladie de Hodgkin, c) image TEP avant traitement avec présence d'une masse ganglionnaire médiastinale fixant très intensément le 18F-FDG, d) après chimiothérapie et montrant un aspect de réponse métabolique complète. [Cachin et al., 2006]	30
Figure 2-1 : Résultats qualitatifs du CAD détaillé par [Guan et al., 2006], avec l'image TEP de départ (a), l'image obtenue après l'algorithme de compétition-diffusion (b), après croissance de régions (c).	34
Figure 2-2 : Résultats qualitatifs obtenus par la méthode développée par [Tozaki et al., 2003; Tozaki et al., 2004] pour deux images cliniques différentes (a, b) et (c, d). Les candidats retenus par le CAD apparaissent en rouge.	34
Figure 2-3 : Coupes d'une image TDM (a), TEP (b) contenant une tumeur dans le poumon. Cette tumeur est correctement détectée par le CAD (c).....	35
Figure 2-4 : Coupes représentatives des résultats obtenus par la méthode de [Kanakatte et al., 2008] : image TDM originale (a), image TEP non corrigée de l'atténuation (b), image TEP corrigée (c) et image représentant les candidats potentiels.	36
Figure 2-5 : Les différentes étapes d'un système CAD.....	39
Figure 2-6 : Coupe de poumons segmentée d'une image TEP corrigée de l'atténuation. La segmentation est obtenue à partir d'une méthode par croissance de régions sur l'image acquise en transmission correspondante.	40
Figure 2-7 : Ensemble de ROIs de taille 9 x 9 pixels extraites de mammographies (El-Naqa, Yongyi et al. 2002) correspondant à des microcalcifications (classe tumeur présente) a) et des structures saines (classe tumeur absente) b).....	41
Figure 2-8 : Illustration de la méthodologie de discrimination linéaire dans l'hypothèse de multinormalité des classes H0 et H1 pour un vecteur de caractéristiques à deux composantes (V0 et V1).....	46

Figure 2-9 : Structures d'un réseau de neurones : un ensemble de couches dont chacune d'entre elle suit le même modèle.....	47
Figure 2-10 : Exemple simpliste d'un arbre de décision à deux nœuds feuilles (malade, bien portant).	48
Figure 2-11 : Principales étapes du processus de sélection par algorithme génétique.....	51
Figure 2-12 : Distributions de probabilité de la variable de décision dans le cas d'évaluation d'une tâche de détection entre deux classes : H0 pour une observation négative (saine) et H1 pour une observation positive (pathologique).	55
Figure 2-13 : Exemple de courbe ROC, chaque croix représente une valeur du critère de décision qui permet d'extraire un couple sensibilité/1-spécificité.	56
Figure 2-14 : Présentation schématique des différentes méthodes d'évaluation de l'erreur de mauvaises classifications concernant l'utilisation et le partitionnement des images disponibles.....	58
Figure 4-1 : De gauche à droite : coupes d'un modèles anatomique, du fantôme labellisé, d'images simulées au 18F-FDG, au 18F-Dopa, au 11C-Raclopride, et image recalée IRM clinique/TEP simulée.	65
Figure 4-2 : Etapes du processus de simulation d'une acquisition TEP réaliste.....	66
Figure 4-3 : Rendu volumique du fantôme NCAT a). Coupe coronale du fantôme Zubal b) .	67
Figure 4-4 : Exemples de coupes coronales d'images cliniques provenant de trois patients différents utilisés pour l'extraction des activités moyennes à appliquer au fantôme numérique Zubal, coupes transverses (ligne du haut), et coronales (ligne du bas)	69
Figure 4-5 : Coupes coronales du Zubal avec les zones d'affection potentielle par le lymphome pointées par les flèches, dans le cas 'médiastinale' a) et 'lominaire' b)	70
Figure 4-6 : Coupes d'images TEP simulées au 18F-FDG générées pour l'étude de calibration. Ces images illustrent les cas extrêmes de détectabilité des tumeurs, avec le plus haut contraste (20 :1) avec un diamètre de 14 mm et le plus petit contraste (2 :1) avec un diamètre de 7 mm, pour les poumons (a, b), le foie (c, d), la rate (e, f) et les artères (g, h)	72
Figure 4-7 : Evolution du pourcentage de tumeurs réellement trouvées par les observateurs en fonction de la distance d'acceptation en mm	74
Figure 4-8 : Variation du niveau moyen de certitude assigné à chaque tumeur de l'étude de calibration. Les résultats sont exprimés en fonction du contraste vrai et du diamètre des tumeurs localisées dans le poumon a), le foie b), la rate c) et les artères d).	74
Figure 4-9 : Le choix de cinq contrastes par organe et par diamètre de tumeurs : 7 mm a), 10,5 mm b), 14 mm c)	75
Figure 4-10 : Les sept lits permettant de couvrir le Zubal par simulation TEP	76
Figure 4-11 : Coupes coronales d'une image TEP simulée a) et d'une image clinique b)	77
Figure 4-12 : Coupes coronales de deux images TEP simulées contenant chacune trois tumeurs dans les artères.....	78
Figure 4-13 : Comparaison des activités simulées et réelles de 18F-FDG dans les principaux organes d'intérêt.....	79
Figure 5-1 : Schéma bloc reprenant les différentes étapes du système CAD développé	86
Figure 5-2 : Schéma descriptif de l'extraction initiale des tumeurs candidates.....	86
Figure 5-3 : Implémentation par banc de filtres d'une décomposition en ondelettes non décimée.....	87
Figure 5-4 : Fonction d'ondelettes et d'échelle pour la Bior4.4 (a, b) et la Coif2 (c, d).....	89
Figure 5-5 : Coupes coronales d'images de décomposition en ondelettes pour le niveau $j=2$ (b-i) obtenue à partir d'une image originale (a) provenant de la base d'images simulées oncoPET_DB.	90
Figure 5-6 : Illustration de la méthode de classification par LDA reposant sur le calcul d'un projecteur optimal des classes	93

Figure 5-7 : Histogramme de répartitions des projetés λ des vecteur V initiaux pour la classe H_0 (rose) et H_1 (vert) pour les poumons (a), le foie (b) et les tissus mous (c) avec un modèle prédictif de LDA calculé pour un niveau de décomposition $j=2$. Les courbes (trait plein) affichées correspondent aux distributions normales adaptées à chaque histogramme.....	97
Figure 5-8 : Histogramme de répartitions des pour la classe H_0 (rose) et H_1 (vert) pour les poumons (a), le foie (b) et les tissus mous (c) avec un modèle prédictif de LDA calculé pour un niveau de décomposition $j=5$. Les courbes (trait plein) affichées correspondent aux distributions normales adaptées à chaque histogramme.	98
Figure 5-9 : Schéma explicatif de la méthode de classification par SVM basée sur la maximisation de la marge entre les deux classes	101
Figure 5-10 : Schéma bloc des étapes de réduction de faux positifs du système CAD	102
Figure 5-11 : Schéma récapitulatif de la sélection des tumeurs candidates les plus suspectes	103
Figure 5-12 : Grilles de recherche permettant d'extraire la meilleure valeur des paramètres C et de σ , pour le RBF (a), le noyau polynome de degré 3 (b) et de degré 4 (c).....	105
Figure 5-15 : Coupe coronale (a) et transverse (b) d'une image de la base simplifiée	109
Figure 5-16 : Courbes FROC obtenues pour deux niveaux de décomposition extrêmes, $j=2$ (ligne pleine) et $j=5$ (ligne pointillée) pour un apprentissage organe spécifique (aucun symbole) ou sur le corps entier (symbole). Les tumeurs du poumon (a), du foie (b) et des tissus mous (c) sont considérées.....	111
Figure 5-17 : Courbes FROC pour deux niveaux de décomposition $j=2$ et $j=5$ sans aucune sélection de caractéristiques (symboles vides) et avec sélection de caractéristiques (ligne pointillée et symbole plein).	112
Figure 5-18 : Vue MIP 3D d'une image TEP simulée corps entier superposée à la carte binaire correspondant à $j=2$ (a) et $j=5$ (b) sans sélection de caractéristiques pour le LDA. Les sept tumeurs disponibles sont pointées par les flèches rouges.	113
Figure 5-19 : Courbes FROC obtenues avec le SVM pour $j=2$ et $j=5$ sans sélection de caractéristiques (symboles vides) et $j=5$ avec sélection de caractéristiques (symboles pleins) pour un apprentissage sur tous les organes et un test sur les poumons (a), le foie (b) et les tissus mous (c).....	116
Figure 5-20 : Courbes FROC obtenues avec le SVM pour $j=5$ sans (symboles) et avec enrichissement de la base d'apprentissage H_0 , notée H_0+ (sans symboles) pour les poumons (a), le foie (b) et les tissus mous (c).	117
Figure 5-21 : Classement de chaque cluster VP détecté par rapport à tous les autres candidats détectés pour le LDA en considérant a) le maximum du cluster comme score de groupe et en comptant les monovoxels, b) en enlevant les monovoxels, c) en considérant la moyenne du cluster comme score et sans les monovoxels, d) pour le SVM en considérant le maximum du cluster et sans les monovoxels.	119
Figure 5-22 : Coupes axiales des cartes binaires superposées aux images originales pour une tumeur dans le poumon (a) et le foie (b), provenant du LDA (au milieu) et du SVM (en bas). L'image originale est présentée (en haut) avec les tumeurs fléchées.....	121
Figure 5-23 : Vue 3D d'une image TEP simulée corps entier contenant 7 tumeurs fléchées superposée à sa carte binaire sortie du LDA (a), du SVM (b) et après les l'élimination des monovoxels et la combinaison par vote à la majorité (c). Les flèches blanches indiquent les tumeurs manquées tandis que les rouges indiquent les tumeurs détectées.	122
Figure 5-24 : Graphes présentant les performances en sensibilité du système CAD complet pour chaque contraste et organe comparé aux performances humaines correspondantes.	123
Figure 5-25 : Graphes montrant la sensibilité moyenne du système CAD pour une tâche de sélection des sept tumeurs les plus suspectes par image. Chaque contraste, définie par la	

sensibilité humaine obtenue pour la même tâche et chaque organe, les poumons (a), le foie (b) et les tissus mous (c) sont traités.	126
Figure 6-1 : Schéma bloc rappelant les étapes du système CAD développé	129
Figure 6-2 : Vue 3D de deux images de test superposées à leurs cartes binaires correspondantes, provenant du LDA suivi de l'élimination des monovoxels (a, d), du SVM sans monovoxels (b, e) et du CAD final (c, f).....	132
Figure 6-3 : Zoom d'une coupe coronale provenant de la superposition d'une image de test avec une carte seuillée provenant du LDA (a), du SVM (b) et de la combinaison de deux (c). La tumeur ciblée est localisée dans le foie.	135
Figure 6-4 : Zoom d'une coupe coronale provenant de la superposition d'une image de test avec une carte seuillée provenant du LDA (a), du SVM (b) et de la combinaison de deux (c). La tumeur ciblée, de contraste minimum, est localisée dans les poumons.	135
Figure 6-5 : Coupe contenant une tumeur cible dans les <i>poumons</i> de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.	139
Figure 6-6 : Coupe contenant une tumeur cible dans le <i>foie</i> de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.....	140
Figure 6-7 : Coupe contenant une tumeur cible dans les <i>ganglions</i> de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.	141
Figure 6-8 : Amélioration de la visibilité des VP au milieu de l'ensemble des candidats par l'insertion du degré de suspicion provenant (c) du LDA, (d) du SVM. Les cartes de candidats ordonnés sont superposées aux images correspondantes d'origine présentées en (a). (b) propose des cartes de détection binaire superposées aux images d'origine. Les tumeurs sont numérotées de haut en bas sur chaque image.....	142
Figure 6-9 : Rang obtenu par chaque candidat VP en considérant tous les organes pour les combinaisons CAD/LDA (triangles verts) et CAD/SVM (carrés rouges). L'idéal est de présenter un VP dans les premières positions dans le classement (0).....	143
Figure 6-10 : Rang obtenu par chaque candidat VP en considérant tous les organes pour les combinaisons CAD/LDA (triangles verts) et CAD/SVM (carrés rouges). Les tumeurs cibles appartiennent, (a) aux contrastes faibles correspondant à 10% et 30% de sensibilité humaine, (b) au contraste intermédiaire (50%) et aux (c) contrastes forts (70%-90%). L'idéal est de présenter un VP dans les premières positions dans le classement (0).....	144

Liste des tableaux

Tableau 2-1 : Exemple de matrice confusion permettant de définir les notions de VP, FP, VN et FN.....	33
Tableau 2-2 : Avantages et inconvénients des classifieurs existants	50
Tableau 2-3 : Les différents stades du processus de classification et leur influence sur la combinaison de classifieur.	53
Tableau 4-1 : Activités de 18F-FDG et coefficients d'atténuation appliqués aux organes principaux du Zubal	69
Tableau 4-2 : DSF calculées à partir de cinq ROI tracées dans les organes principaux des images simulées et cliniques	80
Tableau 4-3 : Paramètres cliniques des études provenant des différents articles de référence et du CERMEP pour la comparaison des SUV	81
Tableau 4-4: Comparaison statistiques des SUVs homme et femme obtenu pour un nombre identique de patients.....	81
Tableau 4-5 : Comparaison des SUV publiés et expérimentaux.....	82
Tableau 5-1 : Principales caractéristiques des ondelettes Bior4.4, Coif2, Daubechies.....	88
Tableau 5-2 : Variances correspondant aux distributions de λ extraites du LDA sur la base de données simulées simplifiée pour $j=2$ et $j=5$ et chacune des classes.	96
Tableau 5-3 : Meilleures performances associées aux valeurs des paramètres correspondants pour chaque fonction noyau	105
Tableau 5-4 : Valeurs de FPI extraits des courbes FROC à une sensibilité fixée de 90% obtenues pour chaque combinaison de paramètres et chaque organe	111
Tableau 5-5 : Valeurs de FPI extraits des courbes FROC à une sensibilité fixée de 90% obtenues pour $j=2$ et chaque organe	113
Tableau 5-6 : Comparaison des caractéristiques sélectionnées par l'algorithme génétique ou le test de Student en considérant chaque organe cible et un niveau de décomposition $j=2$	114
Tableau 5-7 : Evolution des performances après chaque étape de la RFP en termes de sensibilité/FPI.....	120
Tableau 5-8 : Contrastes et organes cibles pour chaque tumeur présentée sur l'image TEP corps entier de la Figure 5-21. Les tumeurs sont présentées de haut en bas sur l'image cible et les contrastes se rapportent aux performances humaines leur correspondant.	122
Tableau 6-1 : Gamme de contrastes définie pour des tumeurs situées dans chaque organe de la base d'images simulées réalistes à partir des performances de détection humaine.	128
Tableau 6-2 : Répartition du nombre de tumeurs par organe pour chaque diamètre et contraste dans les 12 images d'apprentissage et de validation. Le contraste est défini par rapport à la sensibilité humaine correspondante. Seuls les poumons, le foie et les ganglions sont ciblés.128	
Tableau 6-3 : Répartition du nombre de tumeurs par organe pour chaque diamètre et contraste dans les 13 images de test. Le contraste est défini par rapport à la sensibilité humaine correspondante. Seuls les poumons, le foie et les ganglions sont ciblés.....	128
Tableau 6-4 : Contrastes et organes cibles pour chaque tumeur présentée sur les images TEP corps entier de la Figure 6-2. Les tumeurs sont présentées de haut en bas et les contrastes se rapportent aux performances humaines leur correspondant.....	131

Tableau 6-5 : Evolution exprimée en termes de sensibilité/FPI des performances après l'élimination des monovoxels et la combinaison. Les performances sont moyennées sur les 13 images de test dont le nombre de tumeurs cibles est précisé sur la première ligne du tableau.	133
Tableau 6-6 : Sensibilités en pourcentage par contraste et diamètre pour le foie	134
Tableau 6-7 : Sensibilités en pourcentage par contraste et diamètre pour les ganglions	134
Tableau 6-8 : Sensibilités en pourcentage par contraste et diamètre pour les poumons	136
Tableau 6-9 : Taille moyenne en voxels des candidats VP détectés par le LDA, les SVM et le CAD final pour les poumons, le foie et les ganglions.....	136
Tableau 6-10 : Exploitation des rangs des candidats VP dont les valeurs sont extraites de la Figure 6-9 pour chaque combinaison, CAD/LDA et le CAD/SVM. Les meilleures performances sont mises en gras.	144

Liste des abréviations

2D: Bidimensionnel
3D : Tridimensionnel
AG : Algorithme Génétique
CAD : Computer Aided Detection
CHO : Channelized Hotelling Observer
DSF : Déviation Standard Fractionnelle
FDG : 2-fluoro-2-deoxy-D-glucose
FN : Faux Négatifs
FP : Faux Positifs
FPI : Faux Positifs par Image
FROC : Free Receiver Operating Characteristics
IRM: Imagerie par Résonance Magnétique
kPPV : K plus proches voisins
LDA : Linear Discriminant Analysis
LROC : Localization Receiver Operating Characteristics
MCAT: Mathematical CARDiac Torso
NCAT: Nurbs CARDiac Torso
ROC: Receiver Operating Characteristics
SUV: Standardized Uptake Value
SVM : Séparateurs à Vaste Marge
TDM: Tomodensitométrie
TEP : Tomographie par Emission de Positons
VN : Vrais Négatifs
VP: Vrais Positifs

Introduction générale

Contexte de la thèse

Avec environ 320 000 nouveaux cas, 150 000 décès et 800 000 personnes malades en 2005 (source : iNCA), le cancer fait partie des causes majeures de mortalité en France et touche une population de tout âge et de tout sexe. Le nombre de décès par cancer est en diminution lorsqu'on considère l'augmentation de la population. La part globale des cas de cancer dépistés dans la population augmente cependant régulièrement. Cette augmentation est d'une part due à l'hygiène de vie actuelle (tabac, alcool, toxiques professionnels, obésité, pollution...), d'autre part à l'amélioration des techniques diagnostiques et de dépistage. Ainsi, des cancers caractérisés par des tumeurs de petite taille, au niveau du sein ou des ganglions par exemple, sont décelés précocement et sont donc plus facilement soignés.

Les progrès actuels de la médecine, de la biologie et des techniques d'imagerie médicale permettent ainsi de sauver la vie de nombreux malades, d'améliorer la prise en charge du patient et d'optimiser les traitements administrés. L'imagerie TEP appartient à cette nouvelle génération d'imagerie médicale et est utilisée depuis les années 90 en clinique. Cette imagerie permet de visualiser l'information fonctionnelle des cellules et notamment leur consommation en FDG, un analogue du glucose, marqué au fluor 18 (18F). Les cellules tumorales, nécessitant une grande énergie pour se multiplier rapidement, sont de grandes consommatrices de cette substance radioactive, appelée radiotracer. Il n'est pas rare que la TEP soit plus sensible qu'une imagerie anatomique à la détection de tumeurs de petite taille (phase précoce d'évolution ou après traitement). En effet, l'imagerie anatomique ne permet que de caractériser les modifications physiques générées par la présence de la tumeur, alors que les modifications précoces sont d'abord fonctionnelles.

L'optimisation de l'utilisation de l'imagerie TEP en clinique passe par une amélioration de l'instrumentation des caméras et la mise en place d'un ensemble de méthodes optimisant l'interprétation des images provenant des examens cliniques. La quantification des examens TEP est particulièrement étudiée avec pour objectif principal de fournir la meilleure information au clinicien sur la captation du traceur par les cellules humaines. Les développements méthodologiques de traitement des images pour la TEP oncologique se sont jusqu'à présent surtout concentrés sur l'optimisation des méthodes de formation de l'image ou de corrections d'effets nuisant à leur interprétation correcte sur le plan diagnostique (mouvements respiratoires, effet de volume partiel...). Le récent développement de scanners bimodaux, TEP/TDM notamment, a également favorisé le développement de méthodes de recalage ou de segmentation applicables aux images générées. Malgré tous les efforts actuels pour faciliter l'utilisation de l'imagerie TEP d'un point de vue diagnostique, la sensibilité et la spécificité de cette imagerie pourraient encore être améliorées dans le cas de certains cancers.

Certaines pathologies présentent des tumeurs résiduelles fréquentes et difficilement visibles. Les lymphomes, comme le cancer du poumon à petites cellules ou le mélanome, se caractérisent par la dissémination de tumeurs de petite taille dans tout le corps, notamment à des stades d'évolution avancés. Contrairement au cancer du poumon à petites cellules qui se soigne difficilement, les lymphomes répondent généralement bien aux traitements conventionnels. Ils touchent en majorité les enfants et les jeunes adultes. Il est donc primordial d'optimiser les traitements conventionnels pour chaque cas et chaque patient afin de limiter les risques de rechute ou de complications à long terme dus à une trop forte toxicité de traitements inadaptés. L'imagerie TEP est considérée comme un examen nécessaire et crucial pour cette pathologie.

Problématique de la thèse

Motivés par ces développements récents de l'imagerie TEP dans la prise en charge du cancer et l'émergence d'une problématique d'optimisation du diagnostic, nous nous sommes penchés sur la difficile problématique de l'aide à la détection. Cette problématique trouve sa réponse au travers des systèmes d'aide à la détection, dits CAD pour '*Computer Aided Detection*'. Ils proposent une décision quant à la nature d'une image, d'un organe ou d'un ensemble de voxels de l'image, pouvant être

confrontée au diagnostic initial du clinicien. Ils sont applicables dans les cas pathologiques complexes et problématiques, comme lors du suivi thérapeutique du lymphome.

Largement exploités en mammographie et en imagerie TDM pour le cancer du foie ou des poumons, ces systèmes commencent à peine à être développés en imagerie TEP oncologique. Au commencement de la thèse, seules deux publications étaient disponibles sur ce type de développement et à ce jour aucun système commercial n'est disponible. L'intérêt des équipes de recherche augmente cependant avec une émergence de nouvelles études, dont certaines coécrites par des constructeurs comme Siemens. Conformément aux performances obtenues pour les autres modalités d'imagerie, de tels systèmes dédiés promettent une prise en charge plus précise des différentes pathologies. Cette amélioration passe notamment par une augmentation des performances du clinicien en termes de détection et de stadification initiale et après traitement, et une diminution des variations inter- et intra-cliniciens. Dans le cas de tumeurs de taille réduite présentes dans les phases précoces de la maladie ou à la fin des premiers cycles de traitement, les CAD peuvent apporter un second avis au clinicien sous la forme d'images paramétriques ou binaires.

La mise en place de ces méthodes repose généralement sur une optimisation précise et multiparamétrique de toutes les étapes les constituants. Cette optimisation nécessite un nombre non négligeable d'images dont la vérité terrain des tumeurs doit être connue. Celles-ci sont idéalement cliniques et annotées par le médecin. La collecte de telles images apparaît cependant souvent complexe et problématique. Ainsi, il est fréquent d'avoir en premier lieu recours à des images simulées. Ces images simulées permettent d'apporter un réalisme suffisant et une connaissance de la vérité terrain en termes de tumeurs. Elles doivent être suffisamment nombreuses pour l'élaboration et la validation de méthodes par la suite dédiées à des images cliniques.

En retenant toutes ces considérations, les travaux de cette thèse reposent sur ces deux principaux objectifs :

- La proposition d'un système CAD adapté aux images TEP oncologiques 3D dédié aux tumeurs problématiques,
- L'évaluation de ce système CAD à l'aide de deux bases d'images TEP oncologiques simulées dont l'une d'elles a été développée au cours de cette thèse.

Plan de la thèse

Le contexte de cette thèse est présenté en premier lieu. Une brève description de l'imagerie TEP est donnée, suivie d'un développement sur son application en oncologie. L'intérêt du radiotracer FDG et les différentes caractéristiques d'un examen TEP corps entier sont ainsi précisés. Un état de l'art de l'utilisation et des performances de la TEP au FDG pour différentes pathologies cancéreuses d'abord, ciblées sur les lymphomes ensuite, est réalisé et comparé aux autres modalités d'imagerie. Le Chapitre 1 se termine sur la description précise de la problématique abordée lors de cette thèse.

Le Chapitre 2 présente un état de l'art des quelques systèmes CAD existants en imagerie TEP. Un élargissement aux systèmes développés pour les autres modalités est aussi réalisé. L'organisation de ces méthodes en étapes standardisées est ensuite précisée. Enfin les méthodologies d'évaluation de tels systèmes sont présentées.

Les choix méthodologiques retenus au cours de cette thèse pour la proposition du système CAD dédié aux images TEP corps entier sont présentés au Chapitre 3, duquel ressort également l'intérêt de générer une base d'images TEP simulées réalistes dédiée à l'oncologie.

La méthodologie suivie pour le développement de la base d'images TEP simulées est présentée au Chapitre 4. Un modèle de patient virtuel sain en termes de répartition de FDG est associé à un ensemble de tumeurs dont les caractéristiques suivent la problématique clinique du lymphome. La validation de cette base d'images simulées contenant des cas sains et pathologiques est ensuite détaillée. Celle-ci repose à la fois sur des mesures statistiques et cliniques.

Le Chapitre 5 propose un système CAD dédié à l'imagerie TEP oncologique 3D à partir des choix méthodologiques définis au Chapitre 3. Une étude de faisabilité est ensuite réalisée à partir d'une base d'images simulées simplifiée. Différents critères d'évaluation qualitatifs et quantitatifs sont abordés et les performances numériques comparées aux performances humaines.

Le Chapitre 6 présente l'évaluation du système CAD à partir des images plus réalistes proposées au Chapitre 4. L'ensemble des résultats est ensuite discuté et ouvre plusieurs perspectives.

Chapitre 1 Contexte

1.1 Introduction

La Tomographie par Emission de Positons (TEP) est une modalité d'imagerie médicale de type fonctionnelle qui existe depuis le début des années 1970, mais dont l'utilisation en recherche et clinique ne date que des années 1990. Utilisée en recherche pour l'exploration cérébrale et la cardiologie, la TEP est de plus en plus préconisée en clinique pour la prise en charge des cancers. Un nombre croissant de centres hospitaliers possèdent leur propre caméra TEP combinée ou pas. Ce chapitre résume brièvement les principes physiques de la formation de l'image TEP et développe le cas particulier de son application en oncologie.

1.2 Principe d'utilisation de la tomographie par émission de positons (TEP) en oncologie

1.2.1 Introduction

La TEP est un examen d'imagerie fonctionnelle basée sur l'injection de molécules marquées par un atome radioactif émetteur de positons. Ces molécules, propres à la fonctionnalité que l'on souhaite observer, sont transportées par des processus physiologiques jusqu'à la cellule. Les atomes radioactifs se désintègrent en émettant des photons gamma qui sont détectés à l'extérieur du corps humain par une caméra TEP. Le paragraphe suivant reprend brièvement les fondements de l'imagerie TEP avant de préciser son utilisation dans le domaine de l'oncologie.

1.2.2 Principes et spécificités de l'imagerie TEP

L'obtention de la cartographie 3D de distribution d'un traceur radioactif repose sur un processus complexe résumé sur la Figure 1-1. Ce processus se base sur trois étapes principales qui sont l'acquisition des données brutes, leurs corrections et leurs reconstructions. L'acquisition TEP est basée sur la détection de deux photons d'énergie identique en quasi opposition de phase. Ces événements sont appelés coïncidences par la nature simultanée en temps et énergie des deux photons détectés. Un système de détection est placé autour du patient et enregistre les photons arrivant sur deux détecteurs opposés et couplés par un dispositif électronique, dans un intervalle de temps inférieur à une fenêtre temporelle variant de 4 à 20 ns, dite fenêtre de coïncidences. Cette mesure indique la présence d'une annihilation à l'intérieur du volume reliant les deux détecteurs, aussi appelé ligne de réponse (LOR). Les coïncidences vraies correspondent à la détection en coïncidence de deux photons de 511 keV dans la même fenêtre temporelle. Ces événements constituent l'information utile pour la formation de l'image TEP. L'ensemble des coïncidences vraies enregistrées dans une LOR est proportionnel à l'intégrale de la distribution radioactive le long de cette ligne.

Bien que l'information utile de l'examen TEP repose sur la détection de coïncidences vraies, d'autres coïncidences (aléatoires, diffusées) provenant d'autres types d'évènements peuvent aussi être détectées. Celles-ci brulent la formation de l'image TEP. Différentes corrections permettent de réduire les limites de l'imagerie qu'elles soient dues à ces phénomènes physiques (atténuation, diffusion...) ou à des processus intrinsèques à la caméra (normalisation, temps mort...).

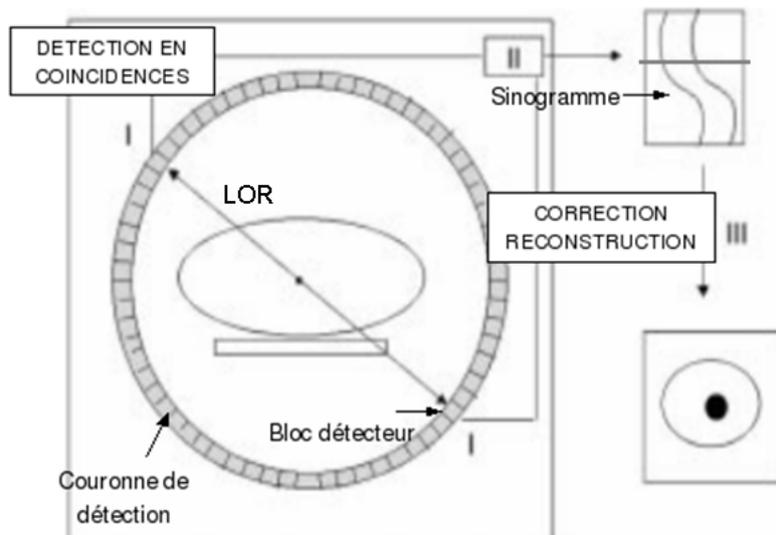


Figure 1-1 : Les différentes étapes de la création d'une image TEP

Il existe trois types d'acquisition TEP. La première méthode d'acquisition est dite en émission. Elle correspond à celle précédemment détaillée, pour laquelle le système de détection tourne autour du patient dont on veut extraire la distribution du radiotracer injecté. La seconde méthode correspond à une acquisition en transmission comme l'illustre la Figure 1-2. Dans ce cas, l'acquisition est réalisée à partir d'une source externe, tournant autour du patient. Cette source peut être du germanium 68 (^{68}Ge), du césium 137 (^{137}Cs) ou encore des rayons X. Le temps de mesure en transmission est souvent inférieur au temps d'acquisition en émission. Cette acquisition permet de mesurer pour chaque LOR l'atténuation des photons dans le corps du patient afin de corriger ultérieurement les données de l'affaiblissement due aux événements atténués. La dernière acquisition est appelée 'blanc' (Figure 1-2), elle est réalisée à vide, c'est-à-dire sans patient et suivant le même protocole que l'acquisition en transmission. Ces deux dernières acquisitions permettent de corriger les données de l'atténuation comme expliqué dans la suite.

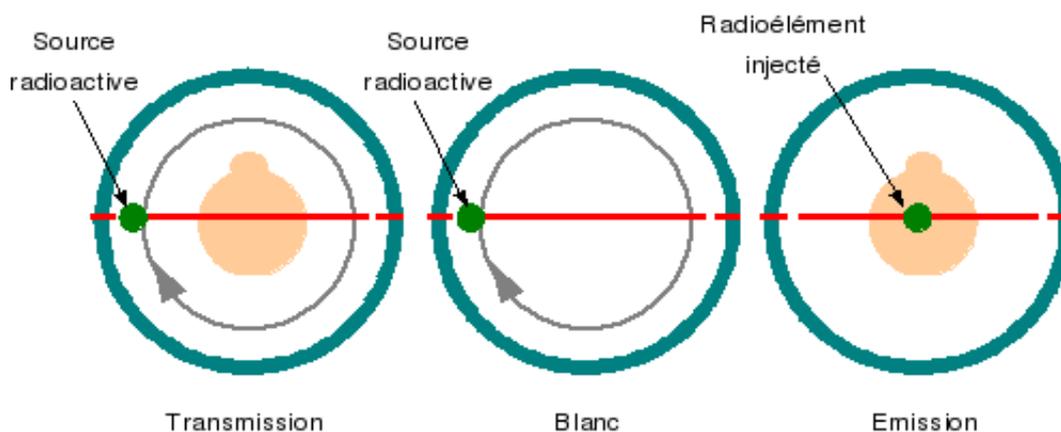


Figure 1-2 : Les trois types d'acquisition, transmission, blanc et émission

La reconstruction permet, à partir de ces données optimisées, l'obtention d'une image 3D représentant la distribution en concentration du traceur radioactif dans l'organisme. Les principales méthodes de reconstruction sont séparées en deux catégories : analytiques et itératives. Les méthodes itératives sont favorisées en imagerie TEP oncologique car elles permettent de mieux prendre en compte les problèmes spécifiques liés à cette application, notamment la faible statistique en termes d'événements (coïncidences) détectés lors de l'acquisition. Une explication détaillée de l'ensemble de ces phénomènes peut être trouvée en Annexe.

En recherche, l'imagerie TEP permet l'exploration fonctionnelle du cerveau afin de favoriser une meilleure connaissance des maladies dégénératives (Parkinson, Alzheimer...). Elle est aussi appliquée dans les centres cliniques en cancérologie depuis quelques années. Elle permet d'apporter une information fonctionnelle utile dans des cas où l'imagerie anatomique atteint ses limites dans les cas de la détection de tumeurs de petite taille notamment. Les paragraphes suivants résument la méthodologie d'acquisition TEP en oncologie et les différentes applications ciblées dans la prise en charge du patient cancéreux.

1.2.3 Intérêt et spécificité du FDG en oncologie

Comme expliqué dans le chapitre précédent, l'acquisition TEP est rendue possible grâce à l'injection de molécules propres à une fonctionnalité de l'organisme, marquée par un traceur radioactif. Dans le cas du cancer, la fonctionnalité observée est l'absorption de glucose par l'organisme et plus particulièrement par les cellules cancéreuses.

Le glucose est considéré comme un carburant essentiel à l'organisme qui peut l'assimiler directement. Les cellules cancéreuses, qui prolifèrent rapidement, sont particulièrement friandes de glucose. Celui-ci est transporté à l'intérieur des cellules par des transporteurs spécifiques de la membrane afin d'être dégradé par un ensemble de réactions appelé glycolyse dont les deux principales étapes sont présentées sur la Figure 1-3. Le glucose est d'abord phosphorylé en glucose-6-phosphate, catalysé par l'enzyme hexokinase. Le glucose-6-phosphate est ensuite transformé en fructose-6-phosphate qui subit à son tour plusieurs transformations successives pour être finalement excrété par la cellule sous forme de lactate selon le cycle de Krebs. Le cycle de Krebs assure la plus grande part des besoins énergétiques de la cellule grâce à la formation de coenzymes réduits qui seront réoxydés dans la chaîne respiratoire. C'est au niveau du cycle de Krebs que se déroule la majeure partie des oxydations et de la production d'énergie d'une cellule.

Le fluorodeoxyglucose ou FDG est un analogue du glucose avec un aspect structural très proche. Il est donc aussi capté par les cellules mais inutilisable dans leur métabolisme. En effet, comme l'indique la Figure 1-3, le FDG suit le même processus métabolique que le glucose jusqu'à l'étape de phosphorylation en FDG-6-phosphate. Il n'est cependant pas métabolisé par l'enzyme glucose-6-phosphatase et s'accumule donc dans la cellule sous la forme de FDG-6-phosphate. Cette accumulation permet de distinguer la présence de tumeurs, généralement beaucoup plus gourmandes en énergie que les tissus sains et donc marquées par une présence élevée de FDG-6-phosphate. Le taux de métabolisme du glucose permet de définir le FDG comme marqueur de l'agressivité de la tumeur [Warburg et al., 1930]. Par son métabolisme et son importance pour les cellules cancéreuses, le FDG est donc actuellement préconisé en clinique pour les examens concernant des patients en cancérologie. Il est généralement marqué au fluor 18 donnant le ^{18}F -FDG.

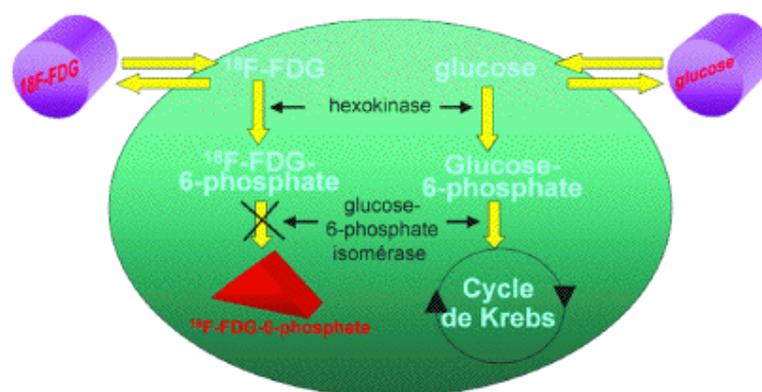


Figure 1-3 : Les différentes étapes de la glycolyse

Bien que les cellules cancéreuses soient considérées comme plus gourmandes en énergie et donc fixant plus le ^{18}F -FDG, celui-ci n'est pas spécifique aux tumeurs dans la limite où tous les organes ont besoin d'énergie et présentent une fixation physiologique de cette molécule [Zincirkeser et al., 2007]. Le cerveau, et notamment la matière grise, capte intensément le ^{18}F -FDG. Le cœur présente une fixation variable d'un patient à l'autre ou chez un même patient. La variabilité de

captation dépend du niveau d'insuline et de la consommation en acide gras. Elle diminue par exemple après une période de jeûne, le cœur utilisant alors la voie de dégradation des acides gras plutôt que celle du glucose. Il est donc fréquent de favoriser une période de jeûne de 4 à 6 h avant l'examen TEP afin de limiter les artefacts dus à la forte captation du cœur. La captation d'autres organes comme le foie, la rate, la moelle osseuse est modérée et varie également selon le patient. La rate et la moelle osseuse présentent d'ailleurs une fixation plus basse que celle du foie. Le thymus est souvent fixant chez les adultes jeunes ou les enfants avec un aspect en V inversé. La graisse brune peut également fixer chez les jeunes adultes dans les régions cervicales en particulier l'hiver pour lutter contre le froid et chez les patients maigres. Le poumon fixe faiblement le 18F-FDG en particulier sur les données non corrigées de l'atténuation. Les voies urinaires et les reins présentent une fixation assez importante du 18F-FDG notamment due à l'élimination du traceur par la voie urinaire, contrairement au glucose qui est réabsorbé par les cellules rénales. Le 18F-FDG est ainsi souvent présent tout au long des voies urinaires. Il est donc parfois difficile de les distinguer d'une fixation ganglionnaire pelvienne. Les muscles peuvent fixer intensément le 18F-FDG, notamment après une intervention chirurgicale et si le patient n'est pas correctement immobilisé (sans parler, sans mâcher). La connaissance par les cliniciens de toutes ces fixations physiologiques normales, illustrées sur la Figure 1-4, permettent de faciliter leur prise en compte dans l'évaluation et le diagnostic des images TEP.



Figure 1-4 : Projection en maximum d'intensité (MIP) représentant la distribution normale du 18F-FDG dans l'organisme

La principale limite de la TEP au 18F-FDG repose donc sur sa non-spécificité tumorale et les processus inflammatoires fixant également le 18F-FDG. L'intérêt de la recherche se porte récemment sur le développement de nouveaux radiotraceurs se liant spécifiquement aux cellules tumorales. [Kumar et al., 2008] présente l'ensemble des nouveaux radiotraceurs marqués au fluor 18 ou autre actuellement en phase de développement. Ces radiotraceurs sont généralement spécifiques à un type de cancer bien ciblé. La choline au carbone 11 (11C-CH) présente par exemple une sensibilité de 81 à 97% et une spécificité de 62 à 87% pour le cancer de la prostate. Ce radiotraceur reflète l'augmentation du métabolisme lipidique et fait face à une utilisation du 18F-FDG souffrant d'un faible métabolisme du glucose et d'une forte élimination urinaire. La fluorothymidine (18F-FLT) et la méthionine au carbone 11 se sont montrées efficaces pour les tumeurs cérébrales face à un cerveau hypercaptant au FDG. D'autres radiotraceurs présentent des performances prometteuses pour différents types de cancer. Plusieurs causes sont à l'origine de la faible utilisation de ces nouvelles molécules pour réaliser les examens TEP. D'une part, ces radiotraceurs n'ont pas encore l'autorisation d'être utilisés en clinique. Leur supériorité au FDG n'a pas forcément encore été démontrée. Aussi, certains d'entre eux sont non plus associés au 18F, mais au 11C. Cet isotope est assez peu disponible et implique un coût de synthèse trop élevé.

1.2.4 Quantification d'un examen TEP au 18F-FDG

Les images TEP au 18F-FDG permettent donc de quantifier l'accumulation de ce traceur dans chaque structure de l'organisme du patient. L'image observée par le clinicien est donc intrinsèquement quantitative et reflète la concentration d'activité de 18F-FDG en kBq/mL pour chaque voxel de l'image. Le Becquerel (Bq) est une unité qui mesure le nombre de désintégrations de noyaux radioactifs par seconde. La concentration tissulaire observée est la somme de la concentration intravasculaire du FDG non phosphorylé et des concentrations intracellulaires du FDG phosphorylé et non phosphorylé. Seule la composante phosphorylée intracellulaire est directement liée à l'activité métabolique de la cellule tumorale. Les médecins préfèrent généralement observer une unité biologique sur les images. Il est donc courant d'utiliser une analyse à compartiments modélisant les échanges du FDG dans les cellules. Il existe de nombreuses approches théoriques de quantification du métabolisme tumoral du glucose, mais la plus connue est l'analyse bicompartimentale simplifiée de Patlak-Gjedde [Patlak et al., 1983]. Après une phase d'équilibre, la concentration $C(t)$ tissulaire au temps t peut être donnée par la formule (1-1).

$$C(t) = \lambda \cdot C_p(t) + K_i \cdot \int C_p(t) dt \quad (1-1)$$

Où $C_p(t)$ la concentration plasmatique (sanguine) de 18F-FDG, λ est la fraction de 18F-FDG non métabolisé. K_i représente le flux net de FDG entrant dans les tissus et est donc directement lié à l'activité métabolique. Des prélèvements sanguins fournissent la concentration de FDG dans le plasma en fonction du temps ($C_p(t)$) et des mesures de la glycémie au cours de l'examen. Réaliser un examen TEP dynamique en oncologie semble cependant assez difficile à mettre en place. Si on effectue la mesure suffisamment longtemps après l'injection intraveineuse du FDG, le premier terme de l'équation (1-1) peut être négligé. Le coefficient K_i étant alors décrit par le rapport $K_i = \frac{C(t)}{\int C_p(t) dt}$. Ainsi, si on

admet que l'intégrale est proportionnelle à l'activité injectée divisée par le poids corporel, on obtient une valeur normalisée du K_i , appelée SUV (standard uptake value). Cette mesure semi-quantitative est donnée par la formule (1-2) :

$$SUV = \frac{C(t)}{\frac{\text{dose injectée}}{\text{poids corporel}}} \quad (1-2)$$

Cette simplification suppose que le volume de distribution et la clairance plasmatique (volume apparent de plasma épuré par unité de temps) du FDG sont identiques chez tous les patients, ce qui est loin d'être le cas. Chez le patient obèse par exemple, les SUV sont habituellement augmentés, la concentration dans les tissus adipeux, impliquée dans le stockage des graisses, étant significativement abaissée. On peut cependant considérer ces paramètres comme stables chez un même patient, le SUV devenant fiable pour un suivi quantitatif. Des études ont d'ailleurs montré une bonne corrélation entre la valeur de K_i obtenue par la méthode de Patlak et le SUV, et une même valeur clinique dans le suivi de patients traités par chimiothérapie [Weber et al., 2003]. Le SUV est en fait corrélé à certaines caractéristiques morphologiques et physiologiques telles que le poids et la glycémie. Le SUV dépend peu de la durée du scanner d'émission mais peut énormément varier avec la carte d'atténuation utilisée pour la correction d'atténuation. En effet, celle-ci est à l'origine de la correction correcte ou pas des images dont est extrait le SUV maximum ou moyenne de la tumeur. Certaines études ont même montré des différences de SUV allant jusqu'à un facteur 2 dues à des différences d'acquisition et de traitement des données [Feuardent et al., 2003]. D'autres types de SUV, non plus normalisés par le poids, mais par d'autres mesures morphologiques ont donc été mis en place. Elles se basent sur une normalisation par la surface corporelle (en m^2) [Yeung et al., 2002] ou la masse maigre (g) [Zasadny and Whal, 1993; Sugawara et al., 1999]. Tout comme le SUV normalisé par le poids, une consommation de FDG proportionnelle au critère corporel est supposée. Un rapport $\frac{SUV_{tumeur}}{SUV_{fond}}$ supérieur à 2 indique généralement une masse considérée comme tumorale en clinique. Une analyse plus précise du clinicien (cytologie, biopsie...) est dans ce cas mise en place.

1.2.5 Acquisition corps entier

Les acquisitions TEP pour l'oncologie concernent en général le corps entier du patient, de manière à réellement définir l'affection sur l'ensemble de l'organisme. Etant donné le champ axial réduit des imageurs TEP (en général une quinzaine de cm), il est nécessaire de déplacer le patient dans le champ de vue de la caméra par incréments successifs. Ce type d'acquisition est ainsi qualifié de multi-lit comme le montre la Figure 1-5. Le nombre de lits est défini par la taille du patient et du champ de vue du tomographe. Les acquisitions correspondant aux différentes positions de lit sont reconstruites séparément et recollées afin de former le volume entier. Afin de limiter la perte d'information en bordure des images due à la plus faible sensibilité aux extrémités du scanner, des chevauchements de lits sont induits. Plusieurs plans peuvent ainsi se superposer aux extrémités des données.

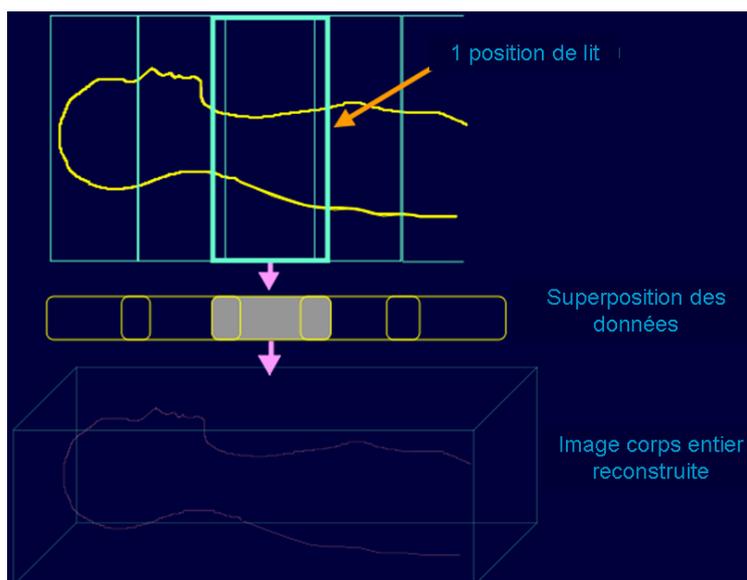


Figure 1-5 : Protocole d'acquisition multi-lit corps entier

Les durées des acquisitions TEP corps entier sont fixées de manière à obtenir une qualité d'image satisfaisante et un confort du patient acceptable. Les protocoles standard établis pour des acquisitions statiques et d'une seule position de lit comme en neurologie et cardiologie prévoient des durées d'examen d'une vingtaine de minutes. En imagerie TEP corps entier, le temps total d'acquisition doit tenir compte des déplacements du lit et assurer un nombre suffisant d'événements par position de lit. Le nombre de lits par acquisition corps entier varie avec la caméra utilisée. Une acquisition corps entier compte en général 6 à 7 lits pour une caméra Siemens, plus pour une Philips. Malheureusement, les temps précisés précédemment sont beaucoup trop courts pour couvrir tous les lits. Un examen TEP doit être réalisé en moins d'une heure pour être admis en routine clinique. Ainsi la durée d'acquisition d'une seule position de lit est réduite d'un facteur 1 à 4 par rapport aux mêmes acquisitions en neurologie et cardiologie. Il faut savoir qu'en imagerie TEP, la durée d'acquisition contrôle la qualité statistique des données. En effet, l'erreur relative sur la mesure de l'activité est proportionnelle à l'inverse de la racine carrée du temps d'acquisition. Diminuer le temps d'acquisition en oncologie d'un facteur 1 à 4 par rapport aux acquisitions en neurologie et cardiologie implique une réduction du rapport signal sur bruit de l'image d'un facteur 1 à 2. Les acquisitions en transmission sont également réduites ce qui augmente le bruit dans l'estimation des coefficients d'atténuation. Cette correction, primordiale en imagerie corps entier où l'atténuation est beaucoup plus élevée qu'en TEP cérébrale, en est donc affectée.

1.3 Performances de la TEP en oncologie et comparaison aux autres modalités

1.3.1 Introduction

Le diagnostic et le suivi du cancer reposent sur un certain nombre d'examen allant de la palpation directe à des méthodes d'imagerie conventionnelles comme la tomodensitométrie (TDM) et l'imagerie par résonance magnétique (IRM) et à des procédures invasives comme la biopsie. Cependant, depuis quelques années l'utilisation de données non plus anatomiques mais fonctionnelles est également recommandée. L'imagerie TEP permet de détecter des structures cancéreuses sur la base de processus biochimiques et moléculaires et non pas seulement morphologiques. Il n'est pas rare que les modifications biochimiques des zones tumorales précèdent les variations anatomiques observables au TDM ou à l'IRM. La TEP peut donc s'avérer plus sensible que les autres modalités pour la détection de tumeurs de petite taille. De nombreuses études cliniques ont été entreprises afin d'évaluer les indications cliniques de la TEP en fonction du type de cancer [Eschmann et al., 2002; Jerusalem et al., 2003; Brink et al., 2004; Eubank and Mankoff, 2005] et certains articles balayent l'ensemble des applications de la TEP en oncologie [Rohren et al., 2004]. Un résumé des contributions de la TEP pour les différentes étapes de la prise en charge des patients cancéreux est donc donné dans les paragraphes suivants.

1.3.2 Le diagnostic initial et la stadification

Le diagnostic initial et la stadification consistent à déterminer le degré de malignité de la pathologie détectée de manière à appliquer au patient un traitement qui corresponde à son degré d'affection. L'évaluation des cancers pulmonaires est une des principales indications de la TEP et représente la moitié des examens effectués.

L'étude de [Gould et al., 2001], effectuée sur 1474 nodules pulmonaires solitaires, a montré une sensibilité de 97 % et une spécificité de 78 % pour l'examen TEP. Cette performance est supérieure à celle des examens cliniques combinée à l'imagerie anatomique. Elle ne considère cependant que les tumeurs de taille supérieure à 8 mm. La sensibilité de la TEP pour la stadification ganglionnaire est très élevée et peut modifier le stade jusqu'à 40 % des patients. La TEP est plus efficace que la TDM pour la détection d'extension extra-ganglionnaire de la maladie [Gould et al., 2001].

La sensibilité de la TEP pour les métastases hépatiques est de 95 % comparée à 76 % et 65 % respectivement pour l'IRM et la TDM [Cohade et al., 2003]. Cependant, la TEP seule ne peut pas être prise en compte. En général, ce type de cancer est traité par chirurgie, d'où un besoin en information anatomique important, qui ne peut être fourni que par un autre type d'imagerie.

La sensibilité et la spécificité de la TEP pour la stadification initiale du lymphome sont respectivement de 86 % et 96 % contre 81 % et 41 % pour la TDM [Stumple et al., 1998]. La TEP présente 89 % de sensibilité et 100 % de spécificité en ce qui concerne la détection simple de la maladie.

La TEP n'est cependant pas conseillée pour le diagnostic initial de cancer du sein. La mammographie reste dans ces cas là l'examen le plus performant.

Peu d'études concluantes ont à ce jour été réalisées sur l'utilisation de la TEP dans le diagnostic initial des cancers œsophagiens. Les résultats existants sont plus controversés et ne permettent pas encore d'affirmer l'intérêt de l'utilisation de la TEP pour l'évaluation du stade de ce type de maladie [Yeung et al., 1999; Meltzer et al., 2000].

1.3.3 Le suivi au cours de la thérapie et la réévaluation du stade

Une fois le diagnostic initial établi, un traitement est mis en place (cycles de radiothérapie, chimiothérapie, actes chirurgicaux). Il est important de pouvoir ensuite suivre l'évolution de la maladie au cours ou à la fin du traitement. Une réadaptation du traitement peut ainsi être appliquée le plus rapidement possible afin de favoriser la survie à long terme du patient et lui éviter des éventuels désagréments supplémentaires.

La TEP s'avère mieux adaptée que la TDM à la différenciation de changements post-chirurgicaux, pouvant être confondus avec des tumeurs résiduelles [Duhaylongsod et al., 1995]. Cependant, les traitements par radiothérapie rendent parfois difficiles l'interprétation des examens TEP à cause des captations de sites inflammatoires persistant parfois jusqu'à six mois après la fin de la thérapie [Frank et al., 1995].

Comme précédemment, la TEP n'est pas particulièrement conseillée pour le cancer du sein. Elle peut cependant être utilisée pour l'évaluation précoce de la réponse au traitement dans le cas de forme fortement métastatiques au niveau ganglionnaire.

En ce qui concerne le lymphome, la TEP offre une réponse rapide de l'évolution de la maladie au cours du traitement. Elle permet aussi de dériver l'état de rémission totale ou complète des patients comme le montre l'étude de [Jerusalem et al., 1999]. Après un premier cycle de traitements sur 54 patients lymphomes, la TEP montre un diagnostic pertinent avec des valeurs prédictives positives de 100 % contre 42 % pour la TDM.

Les cancers œsophagiens doivent être suivis de manière prudente par la TEP. Plusieurs études, dont [Kole et al., 1998], ont montré une sensibilité de 100 % identique à celle de modalités conventionnelles d'imagerie médicale. Cependant la variabilité de la captation de l'œsophage sain est à l'origine de nombreuses fausses détections qui génèrent une spécificité plus mauvaise de 57 %.

L'information TEP associée aux données anatomiques de la TDM après une intervention chirurgicale permet d'atteindre une sensibilité et une spécificité de 81 % et 100 % pour les cancers de la tête et du cou [Wong et al., 1997]. Elle apparaît plus précise que les imageries anatomiques pour la détection de tumeurs récurrentes. Cependant la proximité d'organes hypercaptants rend difficile l'interprétation de localisation. De nos jours, l'utilisation de scanners combinés TEP/TDM est donc favorisée.

1.3.4 Le suivi après traitement

Une fois les cycles de traitement achevés et la rémission établie, les patients sont régulièrement surveillés et peuvent subir des examens TDM, IRM. La TEP apparaît parfois pour pallier la faible efficacité des modalités d'imagerie anatomique pour la détection de récidives de petite taille ou situées en dehors du champ d'exploration. Dans le diagnostic de récurrence des cancers du cou et la tête, la TEP a toute sa place. Cependant, sa réalisation systématique pour la surveillance du patient n'est pas encore validée [Colavolpe et al., 2008]. La principale utilisation de la TEP en gastroentérologie est la mise en évidence de récidives précoces des cancers colorectaux ou la suspicion d'une récurrence locale qui n'a pas fait sa preuve avec les examens traditionnels [Akhurst and Larson, 1999].

1.3.5 Conclusions

Pour conclure, la TEP présente de nombreuses applications et recommandations pour différents types de cancer. La Figure 1-6 présente des images TEP au 18F-FDG provenant de patients atteints de différentes pathologies (cancer du poumon, a), b), lymphome c), d), cancer de la tête et du cou e) et cancer de l'œsophage f)). Toutes ces images sont extraites de l'article de [Rohren et al., 2004]. Les flèches noires indiquent la présence de foyers de forte captation de 18F-FDG généralement synonymes de tumeurs. Les performances de la TEP varient selon les cancers le stade de la prise en charge. Les études montrent notamment que la TEP semble adaptée à la détection de petites tumeurs potentiellement disséminées dans tout le corps du patient. C'est en général dans ce type de cas que des progrès sont encore à faire au niveau de l'utilisation de cette imagerie fonctionnelle. Le cas particulier du lymphome est donc plus précisément abordé dans le paragraphe suivant. Cette pathologie cancéreuse affecte le système lymphatique du patient et se caractérise par la possible dissémination de tumeurs de petites tailles le long des chaînes ganglionnaires.

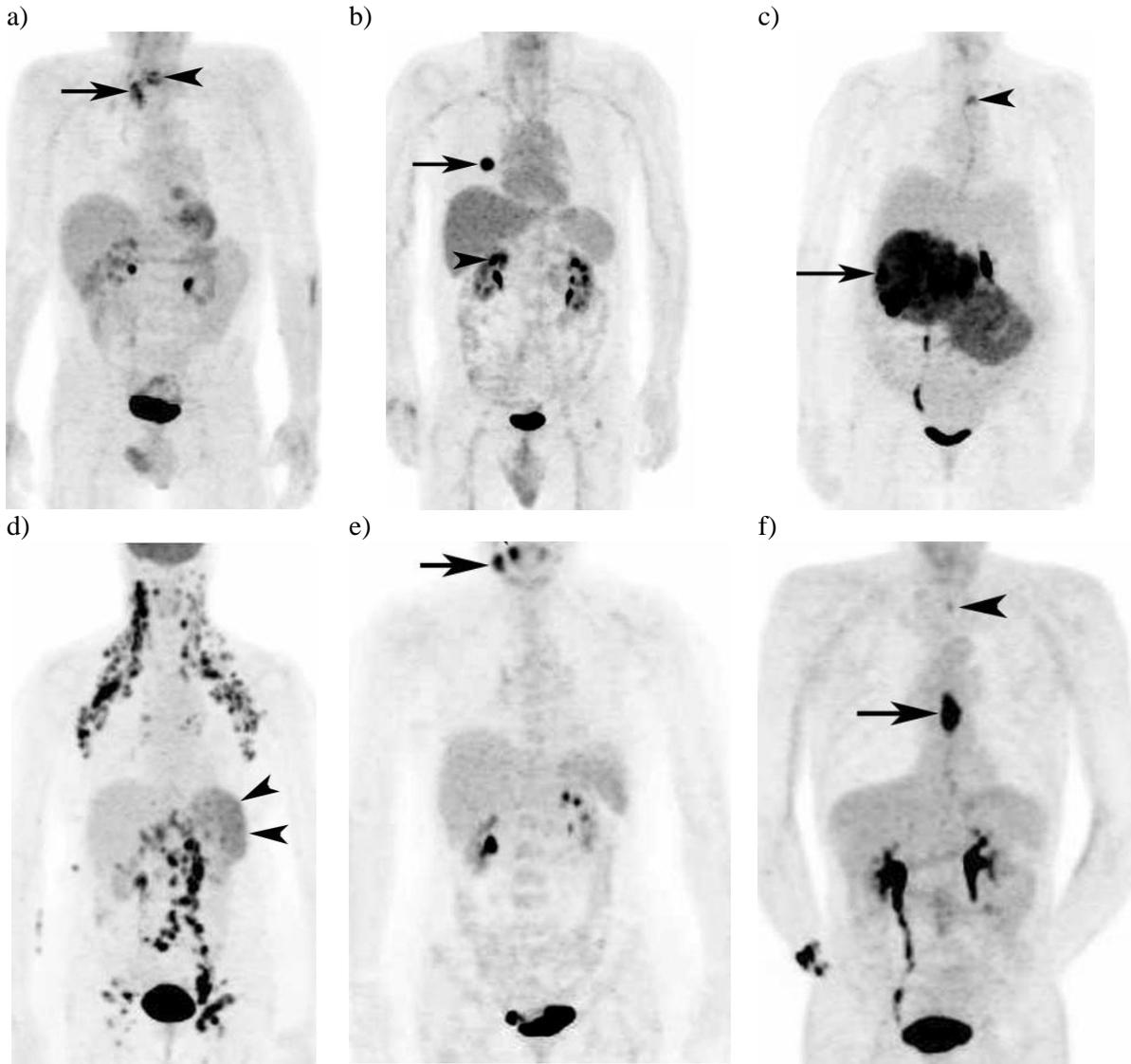


Figure 1-6 : Exemples d'images TEP au 18F-FDG provenant d'examens de patients atteints d'un cancer du poumon, a), b), d'un lymphome c), d), de cancer de la tête et du cou e) et d'un cancer de l'œsophage f)

1.4 Le cas particulier du lymphome

1.4.1 Description médicale de la pathologie

Les lymphomes sont des maladies du sang, caractérisées par la prolifération maligne de cellules lymphoïdiques et réticulaires qui tendent à infiltrer tout l'organisme (source : www.gela.org). Ce cancer du système lymphatique se développe lorsqu'une erreur survient au niveau de la fabrication des lymphocytes, conduisant à la production de cellules anormales. Celles-ci peuvent proliférer de deux manières : en se divisant plus vite que les lymphocytes normaux ou en vivant plus longtemps que ces derniers. Les lymphomes peuvent toucher isolément les ganglions lymphatiques ou d'autres organes lymphoïdiques, mais ils peuvent également atteindre le cerveau, l'estomac, le foie, les glandes salivaires, la peau ou les poumons. Avant de se lancer dans des explications plus précises sur les deux types de maladies caractérisant le lymphome, ses affections et ses différents stades, revenons sur l'origine même de ce cancer, c'est-à-dire une déficience du système immunitaire.

Le système immunitaire correspond au système de défense d'un humain face aux agressions extérieures et aux mutations internes. Celui-ci est constitué d'un ensemble de cellules et organes dont la fonction est d'identifier, de maîtriser et de détruire les particules étrangères ainsi que les cellules anormales, avant qu'elles n'affectent l'organisme. Le système lymphatique fait partie intégrante du système de défense de l'organisme et est illustré sur la Figure 1-7. Il est principalement constitué des

chaines ganglionnaires et des principaux organes lymphatiques (rate, thymus...). Les vaisseaux lymphatiques forment un réseau à l'intérieur du corps humain au sein duquel est transportée la lymphe, un liquide contenant des globules blancs ou lymphocytes. Au sein de ce réseau de vaisseaux lymphatiques se trouvent des milliers de ganglions lymphatiques qui filtrent la lymphe, tout comme le font certains organes spécifiques comme la rate, les amygdales, la moelle osseuse et le thymus. Ce filtrage permet d'extraire des bactéries, virus et autres substances étrangères. Quand un nombre important de bactéries est filtré, un ou plusieurs ganglions lymphatiques peuvent gonfler. Ainsi, la plupart du temps, des ganglions sont les signes que le corps réagit à l'infection.

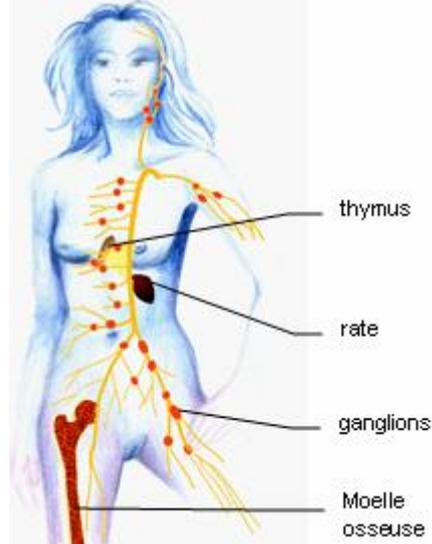


Figure 1-7 : Les principales composantes du système lymphatique humain

Comme expliqué précédemment, les lymphocytes sont des globules blancs fabriqués par les organes lymphoïdiques. Les lymphocytes sont séparés en deux familles : les lymphocytes B et les lymphocytes T. La Figure 1-8 résume le rôle des lymphocytes B jusqu'à leur différenciation en sécrétants d'anticorps et en lymphocytes B mémoire. Les lymphocytes B matures fabriquent des anticorps membranaires pour lutter contre les infections. Les anticorps sont des récepteurs pour un antigène étranger précis. Les lymphocytes B porteurs de ces anticorps, spécifiques des antigènes des substances étrangères, se multiplient en de nombreux clones de ce même lymphocyte B. Lorsque les anticorps ne détectent ou n'éradiquent pas les substances étrangères, les lymphocytes dits de type T interviennent. Ces lymphocytes T reconnaissent quand les cellules ont été infectées et les détruisent directement. Après une attaque, les lymphocytes B se transforment en groupe de cellules mémoire pouvant être réactivée dès la réapparition de l'antigène donné. Le lymphome provient donc d'un mauvais fonctionnement au niveau du système lymphatique et peut avoir pour origine un problème au niveau des lymphocytes T ou B.

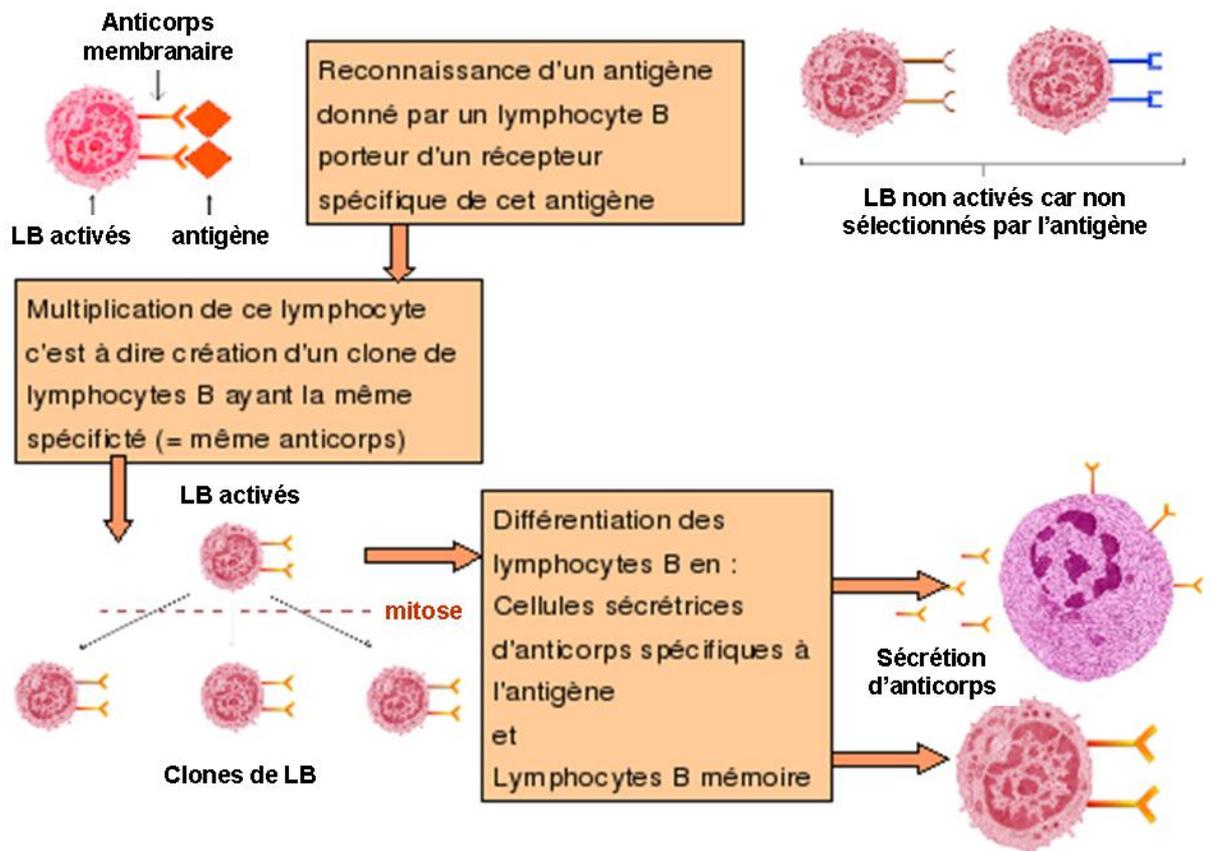


Figure 1-8 : Rôle des lymphocytes

1.4.2 Statistiques et différents stades de la maladie

Comme pour la plupart des tumeurs, la cause exacte des lymphomes n'est pas connue. On sait cependant que les personnes dont les défenses immunitaires sont affaiblies (après un traitement ou lors de différentes maladies) ont un plus fort risque de développer un lymphome. La majorité des chiffres et définitions présentés ci-après proviennent de la Ligue Contre le Cancer (www.ligue-cancer.net) ou du GELA (www.gela.org). Il y a environ 11 000 nouveaux cas de lymphomes chaque année en France et plus de 287 000 nouveaux cas chaque année dans le monde, principalement dans les pays développés. C'est une des tumeurs dont la fréquence augmente le plus, son incidence ayant pratiquement doublée en vingt ans. Le lymphome est la plus fréquente des hémopathies (tumeurs malignes des cellules du sang) et le troisième cancer le plus répandu chez les enfants. Elle touche en majorité des hommes et le taux de survie à 5 ans est de 55 %. Il existe deux types principaux de cancers du système lymphatique : le lymphome hodgkinien (maladie de Hodgkin ou MDH) et le lymphome non hodgkinien (LNH). Le lymphome hodgkinien a été décrit par T. Hodgkin pour la première fois en 1832 [Hodgkin, 1832]. Un demi-siècle plus tard, Wilms [Wilms, 1865] propose l'appellation de maladie de Hodgkin. Il représente 25 % des lymphomes et survient avant 60 ans dans trois quarts des cas, le plus souvent entre 15 et 30 ans. Sa particularité est qu'il reste généralement localisé dans les ganglions lymphatiques et qu'il touche presque toujours les lymphocytes B ganglionnaires. Ce type de tumeur répond bien au traitement avec un taux de guérison élevé. Le lymphome non hodgkinien représente 75 % des cas. Comme pour la maladie de Hodgkin, deux tiers des cas surviennent avant 60 ans et le dernier tiers au-delà de 75 ans. Les lymphomes non hodgkiniens, sont répartis en plus de trente sous-groupes. Ils peuvent se localiser dans tout l'organisme. Chaque groupe se distingue par son évolution dans le corps du patient et sa réponse aux traitements.

La classification internationale d'Ann Arbor [Carbone et al., 1971] permet de répartir les lymphomes en quatre stades cliniques en fonction des zones d'atteinte.

- Le stade I correspond à l'implication d'un seul territoire ganglionnaire (I) et est caractérisée par une survie à dix ans d'environ 90 %.

- Le stade II implique plusieurs chaînes ganglionnaires homolatérales par rapport au diaphragme ou d'une autre région extralymphatique contiguë. La survie à dix ans est de 80 %.
- Le stade III correspond à une atteinte non plus uniquement sus-diaphragmatique mais aussi sous-diaphragmatique incluant la rate. La survie à dix ans est de 70 %.
- Le stade IV est l'affection de plusieurs organes extralymphatiques. La survie à dix ans est de 50 %.

L'extension du lymphome dépend de son lieu d'affection initial (www.leucemie-espoir.org). Les affections initiales médiastinales (zone rouge de la Figure 1-9), correspondent à environ 65 % des cas, pour lesquelles l'extension se fait du médiastin vers les creux sus-claviculaires. Les formes initiales cervicales hautes (zone verte de la Figure 1-9), touchant 25 % des cas, voient l'affection s'étendre vers la base du cou. Les formes initiales ilio-lombaires (zone orange de la Figure 1-9) et axillaires (zone bleue de la Figure 1-9) représentent chacune 5 % des cas et correspondent également à une extension sus-claviculaire. En général, les foyers tumoraux se situent à 80 % au niveau des ganglions lymphatiques et à 20 % dans le reste des organes cibles. La Figure 1-9 présente grossièrement les différents foyers d'affections initiaux sur une image TEP d'un patient atteint d'un lymphome.

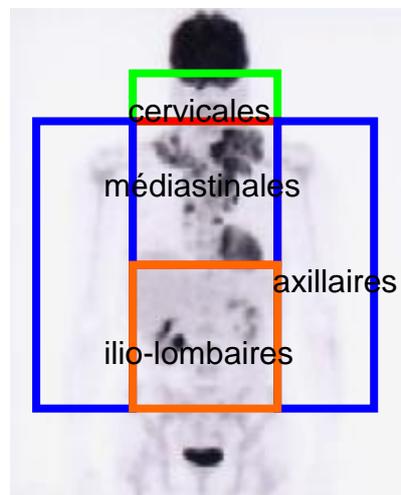


Figure 1-9 : Les différents sites d'affections initiaux du lymphome illustrés sur une image TEP au 18F-FDG d'un patient atteint de ce cancer

1.4.3 La TEP pour le lymphome : avantages

1.4.3.1 Introduction

En cas de suspicion de lymphome, le médecin fait passer un certain nombre d'examen au patient permettant d'exclure ou de confirmer le diagnostic. Parmi ces examens, la TEP au 18F-FDG a toute son importance. Plusieurs études [Bar-Shalom et al., 2001; Segall, 2001; Juweid and Cheson, 2005] ont montré son importance croissante dans la stadification et la réadaptation de la réponse au traitement par rapport à la scintigraphie au Gallium 67 qui était depuis trente ans l'imagerie de référence pour l'exploration des lymphomes. La TEP au 18F-FDG apparaît plus sensible pour la stadification initiale du cancer avec des sensibilités comprises entre 72 % et 100 % contre 63 % à 83 % pour la scintigraphie au Gallium 67 [Workman Jr and Coleman, 2006]. La TEP devient donc l'examen de type fonctionnel de référence pour le lymphome. La Figure 1-10 présente les différentes étapes du traitement des lymphomes et met en évidence les indications des examens TEP correspondantes [Juweid et al., 2007]. Les performances de la TEP pour l'ensemble de ces indications sont résumées dans les paragraphes suivants.

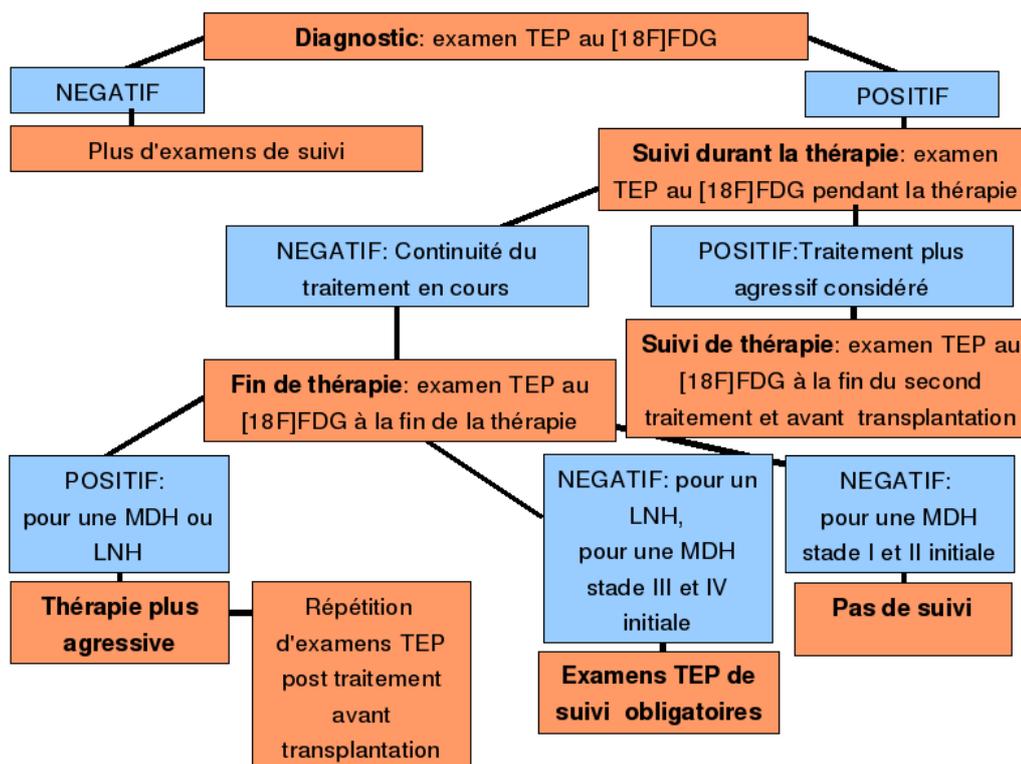


Figure 1-10 : Suivi médical et utilisation de la TEP pour les patients atteints de lymphome

1.4.3.2 Le diagnostic initial et la stadification du lymphome

Parmi les examens standards comprenant l'interrogatoire et l'examen clinique complet, les scanners ont un rôle central pour la détermination du stade initial de la maladie. Pour le cas particulier de la maladie de Hodgkin, la TEP affiche de meilleures performances que les examens anatomiques pour la détection d'affections inguinales, axillaires et claviculaires (zones décrites sur la Figure 1-9) [Bangerter et al., 1998]. Toujours selon cette étude, la TEP ne semble pas montrer de supériorité par rapport aux méthodes d'imagerie anatomique conventionnelles pour l'appréciation de l'extension sous-diaphragmatique. Elle apporte cependant de meilleurs résultats pour l'appréciation de l'extension viscérale et médullaire. Une autre étude [Partridge et al., 2000] a montré l'influence de la TEP sur la détermination du stade par rapport aux autres imageries. 18 patients sur 44 ont vu leur stade augmenter et 3 patients sur 44 ont vu leur stade diminuer. Ces conclusions ont permis de modifier le traitement initial de 10 patients parmi les 18 dont la gravité de l'état était surestimée et d'1 patient sur les 3 dont l'état est finalement plus pessimiste. Une étude similaire [Depas et al., 2004] portant cette fois sur les lymphomes pédiatriques a montré des résultats comparables en termes de stadification. Bien que la TEP et les techniques d'imagerie conventionnelles soient concordantes pour 16 patients sur 19, un d'eux a vu son évaluation initiale passée d'un stade IV à un stade II. Un autre a vu son évaluation passée du stade III au stade IV, la TEP étant l'unique modalité ayant détecté une tumeur dans les os. La TEP a donné une évaluation du stade incorrecte pour un patient sur les trois où l'IRM s'est avérée plus efficace. En termes de sensibilité/spécificité, la TEP affiche des performances de 86 % et 96 % respectivement, contre 86 % et 41 % pour les imageries anatomiques pour la maladie de Hodgkin [Stumple et al., 1998]. Pour les lymphomes non hodgkiniens, le couple sensibilité/spécificité atteint 89 % et 100 % pour la TEP contre 81 % et 41 % pour les imageries anatomiques [Stumple et al., 1998]. Il semble donc plutôt évident que la réalisation d'un examen TEP au 18F-FDG lors de la stadification initiale de la maladie ait une importance et permette d'améliorer le choix initial du traitement. La détection précoce d'un stade élevé permet de réaliser directement des traitements lourds, sans passer par de futures thérapies de 'rattrapage'. De même pour un stade plus optimiste, un allègement des traitements permet d'assurer une meilleure sécurité du patient en évitant les éventuelles complications tardives dues à de trop fortes toxicités.

1.4.3.3 Le suivi au cours de la thérapie et la réévaluation du stade

L'évaluation précoce de la réponse au traitement le garantit d'une survie plus longue du patient. Elle permet de réadapter ou d'arrêter le traitement en cours de manière à affranchir le patient de futures complications. Les lymphomes sont des pathologies répondant généralement bien aux traitements. Les tumeurs ciblées subissent ainsi souvent une importante réduction de taille, qui se visualise sur les différents types d'imagerie et particulièrement en TEP. Cependant, l'imagerie anatomique ne peut pas différencier la fibrose résiduelle de cellules tumorales viables. L'imagerie fonctionnelle comme la TEP au 18F-FDG est parfois capable de faire cette différenciation de part son apport d'information fonctionnelle. L'étude de [Dittmann et al., 2001] a montré des résultats prometteurs sur la différenciation des masses résiduelles et tumeurs viables post-traitement. L'étude porte sur 40 patients atteints de la maladie de Hodgkin. La sensibilité et la spécificité obtenues sont respectivement de 87,5 % et 94,7 % pour la TEP contre 25 % et 56 % pour la TDM. Plusieurs études ont d'ailleurs été menées pour valider l'intérêt de la TEP pour la prédiction de la réponse thérapeutique. [Jerusalem et al., 1999] compare la TEP et la TDM après un premier cycle de traitements sur 54 patients lymphomes. La TEP montre un diagnostic pertinent avec des valeurs prédictives positives de 100 % contre 42 % pour la TDM. La survie avec arrêt immédiat du traitement et la survie générale de tous les patients est fixée à 86 % et 92 % pour les patients détectés comme guéris après ce premier cycle de traitement par la TEP. Les mêmes chiffres pour les patients détectés comme non totalement guéris sont 0 % et 50 %. Des études similaires [Maisey et al., 2000; Mikhaeel et al., 2000; Spaepen et al., 2000] ont également affiché des résultats prometteurs sur la valeur pronostique de la TEP au 18F-FDG après un ou plusieurs cycles de traitement.

1.4.3.4 Le suivi après traitement

Les lymphomes Hodgkiniens ont un taux de rechute de 30 % environ. Ces rechutes se produisent en général dans les 5 ans, ce qui impose une surveillance étroite des patients (tous les trois mois pendant deux ans puis avec un allègement progressif). Cette surveillance repose essentiellement sur des interrogatoires du patient et d'examen cliniques mais les scanners sont parfois utilisés. Actuellement, peu d'études sont consacrées à l'utilisation de la TEP sur le suivi long terme.

1.4.4 La TEP pour le lymphome : limitations et besoins

Plusieurs études ont été menées de manière à définir de façon précise à la fois les limites de la TEP au 18F-FDG [Barrington and O'Doherty, 2003] et les besoins des cliniciens [Hoskin, 2003] concernant le traitement des lymphomes. La première limite de la TEP au 18F-FDG ne concerne pas seulement son application au lymphome, mais repose simplement sur le métabolisme même du 18F-FDG qui n'est malheureusement pas spécifique aux cellules cancéreuses mais à toutes les cellules demandant de l'énergie. Même si le niveau de fixation dans les tumeurs est généralement plus élevé que dans les organes sains, il peut y avoir des effets de recouvrement. Ce recouvrement peut entraîner des cas de détection difficiles notamment pour des tumeurs captant faiblement le 18F-FDG. Ainsi, la Figure 1-11 présente des images TEP au 18F-FDG de patientes atteintes de lymphome avant et après traitement. La première patiente présente une réponse partielle alors que la seconde une réponse totale. D'une manière plus générale, la TEP au 18F-FDG peut être à l'origine de quelques erreurs typiques indépendamment du stade de la prise en charge du lymphome. Elles sont classées en trois catégories, les tumeurs manquées, les tumeurs masquées et les tumeurs confondues.

Considérons d'abord le cas des tumeurs manquées. Il paraît évident que les tumeurs les plus souvent ratées par l'œil humain sont de petites tailles et de faible captation [O'Doherty et al., 2002]. Les patients scannés pendant ou juste après une chimiothérapie présentent aussi des tumeurs difficilement détectables, dues au chamboulement généré par ce traitement au niveau morphologique et physiologique. Un autre problème plus difficile à résoudre concerne des types de lymphomes qui ne captent pas le 18F-FDG. Certains lymphomes de bas grade présentent une fixation variable de radiotraceur [Stumple et al., 1998; Jerusalem et al., 1999], tandis que certains lymphomes dit de MALT (mucosa-associated lymphoid tissue) ne captent pas du tout le 18F-FDG [Hoffman et al., 1999]. Une étude a également montré que les tumeurs de moins d'un centimètre pouvait être manquées sur une image TEP au 18F-FDG [Jerusalem et al., 2001], ce qui est gênant pour cette pathologie.

Passons maintenant au cas des tumeurs masquées. Une forte captation de radiotracteur dans les organes sains peut résulter en une difficulté à détecter les tumeurs réelles. Le cerveau et plus spécifiquement la matière grise est particulièrement gourmande en 18F-FDG ce qui implique que les lymphomes cérébraux ne peuvent pas toujours être suivis par des examens TEP au 18F-FDG. De même, La captation physiologique variable de l'appareil digestif peut aussi être à l'origine de l'omission de petites tumeurs. Le cœur peut également être très captant et masquer des tumeurs situés dans le médiastin le long des principales artères proches des chaînes ganglionnaires (Figure 1-7) [Barrington and O'Doherty, 2003]. Les deux types d'erreur précédemment décrites (tumeurs manquées et tumeurs masquées) génèrent des faux-négatifs lors de l'évaluation des images TEP par le clinicien.

La dernière limitation de la TEP au 18F-FDG est la présence de foyers de fixation imitant le comportement d'une tumeur réelle comme les inflammations des tissus [Zhuang and Alavi, 2002]. Celles-ci sont fréquentes plusieurs jours après une chimiothérapie. D'autres exemples précis de cas d'imitation de tumeurs peuvent être cités. Une captation focalisée dans le cœur peut être interprété à tort comme une captation des ganglions lymphatiques paracardiaque. Une fixation élevée de radiotracteur focalisée dans les urètres peut être confondue avec une captation dans les ganglions para-aortiques et iliaques. Une fixation élevée de la nuque et de la partie haute de la poitrine peut être confondue quant à elle avec une captation des ganglions lymphatiques cervicaux. Ce même phénomène peut être observé avec la fixation de la graisse brune par les enfants ou jeunes adultes. L'ensemble de ces cas peuvent être à l'origine de faux positifs dont le nombre est généralement réduit par la combinaison de deux modalités d'imagerie (exemple : TEP/TDM).

Le diabète et l'hyperglycémie sont deux exemples de facteurs supplémentaires pouvant compliquer l'interprétation des images TEP par les cliniciens [Barrington and O'Doherty, 2003]. Cette remarque est particulièrement vraie notamment dans le cas de tumeurs faiblement captantes dont la taille atteint parfois la résolution de l'image.

Les examens TEP au 18F-FDG sont fortement recommandés pour la prise en charge des lymphomes. Cependant, l'interprétation de ces images est encore limitée par les problèmes précédemment décrit et pourrait être facilitée et améliorée. Selon [Hoskin, 2003], un des besoins primordial des cliniciens apparaît au niveau de la réévaluation du stade après un ou plusieurs cycles de traitement. En effet, la réévaluation correcte de la réponse thérapeutique est le garant d'une survie plus longue du patient. On distingue deux cas de réponses thérapeutiques

- La réponse thérapeutique est déclarée totale. Dans ce cas là, les traitements sont interrompus et le patient passe au stade de suivi trimestriel, voir annuel, selon le stade d'affection initial d'Ann Arbor (cf. paragraphe 1.4.2). En cas de diagnostic erroné, l'état du patient continuera à se détériorer et il sera soumis à une thérapie plus agressive, donc nocive pour son futur à court et long terme.
- La réponse thérapeutique est déclarée partielle. Une batterie de traitements plus agressifs (autogreffe, chimio- radiothérapie plus intenses) est appliquée au patient afin d'éliminer définitivement la maladie. Une erreur de diagnostic augmentent les facteurs de risques à long terme dus à la toxicité des traitements (insuffisances cardiaques, secondes affections induites...) pour le patient complètement remis.

Les lymphomes touchent en majorité les personnes jeunes et répondent correctement aux thérapies ciblées. Ainsi, l'assurance d'un diagnostic sans faille par le clinicien est parfois difficile notamment dans les cas précédemment cités. Faciliter la mise en place de ce diagnostic à l'aide d'un outil numérique, non soumis à l'objectivité de l'œil humain, serait ainsi un atout pour rendre cette modalité d'imagerie médicale encore plus utilisée.

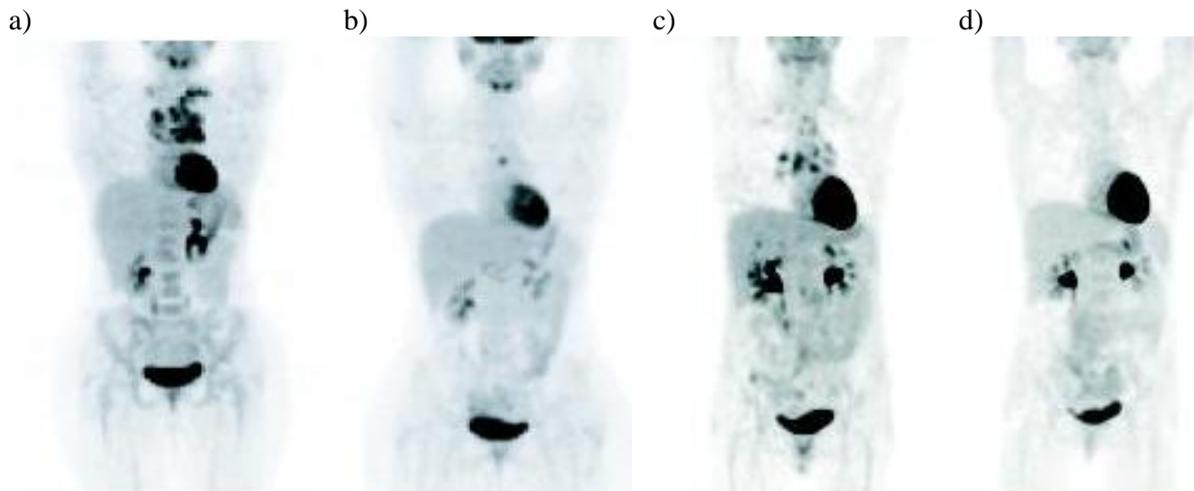


Figure 1-11 : Exemples d'image TEP au 18F-FDG : patiente présentant un lymphome malin non hodgkinien de haut grade, a) image TEP avant traitement montrant des cibles hypermétaboliques médiastinales et sus-claviculaires gauches, b) après chimiothérapie montrant un aspect de réponse partielle. Autre exemple : patiente présentant une maladie de Hodgkin, c) image TEP avant traitement avec présence d'une masse ganglionnaire médiastinale fixant très intensément le 18F-FDG, d) après chimiothérapie et montrant un aspect de réponse métabolique complète. [Cachin et al., 2006]

1.5 Détection en imagerie TEP oncologique

1.5.1 Les verrous de la TEP au FDG en oncologie

1.5.1.1 La qualité des images TEP

Les paragraphes précédents montrent clairement l'utilité de l'imagerie TEP dans les différentes étapes de la prise en charge des cancers. On peut en effet déjà observer des performances en sensibilité et spécificité supérieures aux modalités d'imagerie conventionnelle pour certains types de cancer comme le lymphome.

Cependant, l'imagerie TEP présente plusieurs inconvénients qui limitent leur utilité diagnostique. Outre le fait que le FDG ne soit pas spécifique qu'aux tumeurs, l'imagerie TEP n'apporte aucune information anatomique. Les images TEP sont aussi généralement qualifiées d'images de 'faible qualité visuelle' par rapport aux autres modalités d'imagerie médicale, et ce pour plusieurs raisons. Elles sont tout d'abord caractérisées par une résolution plutôt mauvaise comparée aux autres modalités d'imagerie médicale. Cette résolution est comprise entre 3 et 5 mm pour la TEP humaine contre une résolution d'un millimètre pour les moins bonnes performances pour les images TDM et IRM. Les gammes d'intensité de l'image TEP sont en général très larges et peuvent dépendre de facteurs physiologiques et cliniques (dose injectée, temps d'acquisition, poids du patient...). Les tumeurs ne sont pas forcément les structures les plus fixantes en 18F-FDG. Enfin et non des moindres, les images TEP corps entier sont fortement bruitées. Déjà lors de l'acquisition des sinogrammes, un bruit de type Poissonien, dû aux différents processus physiques et à la décroissance du radiotracer utilisé, entâche les données brutes. Ce bruit est d'autant plus important en imagerie corps entier pour lesquelles les durées d'acquisition et donc la statistique sont réduites (paragraphe 1.2.5). De même, la réduction de la durée des mesures en transmission augmente le bruit dans l'estimation des coefficients d'atténuation. Cet effet est particulièrement pénalisant pour les acquisitions du thorax et de l'abdomen, caractérisées par une atténuation très élevée. Le traitement numérique des images TEP oncologiques corps entier repose donc sur une correcte prise en compte de chacune de ces limitations, de manière à réduire les potentielles erreurs.

1.5.1.2 Les développements méthodologiques pour une utilisation facilitée de la TEP en oncologie

Différentes méthodes sont actuellement en développement avec pour objectif principal de faciliter l'intégration de la TEP comme examen clinique de référence dans la prise en charge du

cancer. Ces développements portent principalement sur l'amélioration de la qualité des images TEP afin d'augmenter son utilité diagnostic en clinique.

Ils reposent d'une part sur l'amélioration des méthodes de correction et reconstruction déjà existantes et implantées sur les caméras ou la prise en compte de nouveaux effets limitants. Les méthodes de correction d'atténuation [Pan et al., 2005; Qiao et al., 2006] et de reconstruction de type statistiques [Qiao et al., 2006] peuvent être citées en exemple dans la première catégorie tandis que les corrections d'effet de volume partiel [Soret et al., 2007; Le Pogam et al., 2008; Rousset et al., 2008] et de mouvement respiratoire [Nehmeh et al., 2003; Visvikis et al., 2006; Lamare et al., 2007] illustrent la seconde catégorie.

D'autre part des méthodes permettant d'améliorer la quantification des images TEP sont évaluées. Celles-ci reposent sur une exploitation plus fiable des SUV (SUV moyen, SUV maximum...) afin d'aider le médecin à prendre une décision sur une zone suspecte [Buvat et al., 2007; Tylski et al., 2008].

Le développement exponentiel de caméras couplées (TEP/TDM, TEP/IRM...) implique l'émergence de méthodes de traitement d'image de type recalage [Makela et al., 2003; Rizzo et al., 2005; Ireland et al., 2007] et segmentation de tumeurs pré-détectées [Aristophanous et al., 2007; Li et al., 2008; Hatt et al., 2009]. Elles sont particulièrement nécessaires pour les traitements de type radiothérapie où l'anatomie et le volume des tumeurs doivent être connus de manière précise.

Comme nous l'avons indiqué en introduction, un besoin non négligeable en TEP oncologique concerne l'aide à la décision dans le cadre du suivi thérapeutique. Cette aide passe principalement par le développement de méthodes d'aide à la détection. Ce domaine méthodologique, parcouru depuis de nombreuses années en imagerie TDM et mammographie, est encore très peu développé en TEP de part la difficulté d'exploitation de ce type d'imagerie fonctionnelle. Le cas des tumeurs de petite taille et de faible activité difficilement visibles ou masquées par les fixations physiologiques des autres organes est fréquent. Les médecins affirment aussi qu'après une journée de visualisation d'images TEP, leur œil ne peut pas être aussi efficace et sensibles aux petits détails. Leur proposer un second avis 'numérique' sur ces images problématiques aiguillerait ou confirmerait leur diagnostic. Ainsi, cette thèse aborde le développement d'un tel système CAD pour l'imagerie TEP corps entier oncologique.

Le chapitre suivant constitue un état de l'art des méthodes de détection automatique existantes et appliquées en imagerie médicale.

Chapitre 2 Les systèmes d'aide à la détection en imagerie médicale

2.1 Introduction

Un système d'aide à la prise de décision dédiée à l'imagerie médicale est un programme numérique qui consiste à assister le clinicien dans la tâche parfois délicate d'interprétation et de classement des images médicales. Ce domaine est représenté par deux applications principales : les systèmes d'aide à la détection (*Computer Aided Detection* ou *CAD*) et les systèmes d'aide au diagnostic (*Computer Aided Diagnosis* ou *CADi*). Les CAD ont pour rôle de détecter et d'attribuer un score à une ou plusieurs tumeurs dans une image. Les CADi différencient les tumeurs malignes et bénignes parmi celles détectées par les CAD ou fournies par un clinicien. Comme précisé lors de l'introduction, le besoin principal des cliniciens en imagerie TEP oncologique se focalise sur le suivi thérapeutique. Pour cela, il est primordial de correctement détecter les éventuelles tumeurs résiduelles. Ainsi, nous nous focalisons davantage sur l'aspect CAD, correspondant à la détection de tumeurs problématiques dans les images TEP corps entier, qu'au diagnostic de ces tumeurs par leur caractérisation de tumeurs bénignes ou malignes.

En termes vulgarisés, un CAD peut être assimilé à un observateur numérique analysant l'image et y posant le même type de décision qu'un clinicien. Cette décision numérique n'a pas pour ambition de remplacer l'avis humain, mais de le compléter. Ainsi, le clinicien peut utiliser ce second avis pour revenir sur son appréciation initiale, en la confirmant, en l'infirant, ou en attirant son attention sur d'autres localisations potentiellement pathologiques. La technologie CAD est interdisciplinaire et combine des éléments d'intelligence artificielle et de traitement d'images.

Le développement de systèmes CAD en imagerie médicale a débuté dans les années 80 [Chan et al., 1987; Giger et al., 1987] et ciblait des applications à forts enjeux médicaux comme la mammographie. Les études de [Freer and Ulissey, 2001; Gilbert et al., 2006; Doi, 2007; Marten and Engelke, 2007] présentent un état de l'art non exhaustif d'une grande partie des méthodes CAD existantes. Elles montrent que ces systèmes permettent une meilleure prise en charge des pathologies cancéreuses, notamment en augmentant les performances en terme de détection et de stadification et en diminuant les variations inter- et intra-cliniciens. Certaines méthodes sont d'ailleurs déjà implantées sur les consoles cliniques. On peut citer en exemple Image Checker de R2 technologies (<http://www.r2tech.com>), iCAD (<http://www.icadmed.com>), Kodak mammography CAD engine (<http://www.kodak.com>) et syngo mammoCAD de Siemens qui assure l'aide au diagnostic en mammographie. syngo CT de Siemens comporte un ensemble de modules dédiés au cancer du poumon ou du colon (<http://www.medical.siemens.com>) pour l'imagerie TDM, tandis que Philips utilise les programmes xLNA (www.medical.philips.com) pour l'imagerie TDM. Les rares méthodes de type CAD récemment proposées en imagerie TEP 3D n'ont pas encore fait l'objet d'un transfert industriel. Pourtant, les principaux constructeurs (Siemens, Philips...) sont associés à ces travaux par le biais de collaboration avec des équipes académiques.

Les paragraphes suivants présentent tout d'abord les quelques systèmes CAD récemment développés en imagerie TEP puis confrontent deux familles de classifieurs. Un schéma méthodologique standardisé est ensuite présenté dont chacune des étapes est commentée au gré d'exemples de CAD en imagerie TDM et en mammographie. Les règles permettant l'évaluation et la comparaison de tels systèmes sont enfin données.

2.2 Les CAD en imagerie TEP : une application en développement

2.2.1 Introduction

Ce paragraphe détaille les systèmes CAD publiés dans la littérature. Les méthodes présentées ci-après sont réparties en deux catégories : les CAD n'utilisant que l'information fonctionnelle et les CAD combinant les informations fonctionnelles et anatomiques. On peut distinguer deux types de méthodes, celles reposant sur la classification supervisée et celles s'inspirant de méthodes de

traitement d'images. Un paragraphe préliminaire précise les quelques définitions utiles à la compréhension des performances de ce type de système.

2.2.2 Définitions préliminaires

La nature des performances quantitatives d'un système de détection repose sur quatre notions essentielles que sont les vrais positifs (VP), faux positifs (FP), vrais négatifs (VN) et faux négatifs (FN). Un VP quantifie la détection d'une vraie tumeur tandis qu'un FP correspond à la détection d'une tumeur qui n'en est pas réellement une. De même, les FN qualifient les vraies tumeurs non détectées par le système. Une matrice de confusion, présentée dans le Tableau 2-1, propose une définition simple des mesures de la qualité d'un système CAD qui sont comprises entre 0 et 1 (ou 0 et 100 %).

Tableau 2-1 : Exemple de matrice confusion permettant de définir les notions de VP, FP, VN et FN

		Classe estimée	
		Sain	Pathologique
Classe réelle	Sain	Vrai Négatif (VN)	Faux Positif (FP)
	Pathologique	Faux Négatif (FN)	Vrai Positif (VP)

Plusieurs valeurs peuvent être extraites de cette matrice de confusion de manière à quantifier les différentes capacités du CAD. La sensibilité, définie en (2-1), correspond à la capacité du CAD à détecter les vraies tumeurs, tandis que la spécificité, décrite en (2-2) correspond à sa capacité à ne pas détecter de tumeurs là où il n'y en a pas. La précision (2-3), quant à elle, précise l'exactitude de la détection. Ces valeurs sont utilisées par la suite pour quantifier les performances des CAD présentés.

$$\text{Sensibilité} = \frac{VP}{VP + FN} \quad (2-1)$$

$$\text{Spécificité} = \frac{VN}{VN + FP} \quad (2-2)$$

$$\text{Précision} = \frac{VP}{VP + FP} \quad (2-3)$$

2.2.3 Les systèmes CAD utilisant uniquement l'information TEP

Les premiers systèmes CAD présentés dans ce paragraphe se basent sur les deux types de méthodes précédemment citées, du traitement d'image et des classifieurs supervisés. Ils utilisent uniquement l'information fonctionnelle.

La méthode de [Guan et al., 2006], qui a aussi fait l'objet d'un brevet [Huang et al., 2007], débute par une labellisation des parties du corps humain à l'aide de modèles de Markov caché [Rabiner, 1989] permettant d'extraire le cou, les poumons et l'abdomen. Les tumeur candidates sont ensuite extraites grâce à un algorithme de compétition-diffusion [Kubota and Espinal, 2000]. Celui-ci fait évoluer deux régions de l'espace vers deux états majoritaires (tumeurs/non tumeurs) par minimisation d'une fonctionnelle dépendant de paramètres discriminants. Un SUV de 2 pour les poumons, proche de celui significatif en clinique (2,5), est utilisé pour l'initialisation de l'algorithme de compétition-diffusion, c'est-à-dire le choix des régions initiales. Les états majoritaires sont définis à partir de valeurs de SUV propres à chaque partie du corps. La segmentation finale des volumes tumoraux se fait quant à elle par croissance de régions 3D dont la condition d'arrêt repose sur un pourcentage du SUV maximal initial. La Figure 2-1 présente trois coupes correspondant aux différentes étapes du CAD : l'image TEP originale (a), les régions d'intérêt (b) et la segmentation finale par croissance de régions (c). Bien que les résultats qualitatifs présentés prouvent l'efficacité de la méthode sur des images cliniques contenant des tumeurs de taille moyenne, aucun résultat quantitatif n'appuie ces performances. De plus, on constate sur la Figure 2-1c que la méthode détecte également les fixations physiologiques des

reins. L'inconvénient principal de cette méthode repose aussi sur l'utilisation du SUV, qui comme expliqué au paragraphe 1.2.4, peut être assez instable selon de nombreux critères instrumentaux, méthodologiques et cliniques. La segmentation initiale des parties du corps humain apparaît cependant très intéressante pour l'imagerie TEP souffrant du manque d'information anatomique.



Figure 2-1 : Résultats qualitatifs du CAD détaillé par [Guan et al., 2006], avec l'image TEP de départ (a), l'image obtenue après l'algorithme de compétition-diffusion (b), après croissance de régions (c).

Une autre méthode de détection basée sur la sélection de pixels de courbures de Gauss 2D maximales permet à [Tozaki et al., 2003; Tozaki et al., 2004] de réaliser une première extraction des tumeurs candidates. Ces courbures correspondent au produit des courbures principales. Celles-ci sont calculées par combinaisons linéaires des gradients des intensités dans les deux coordonnées de l'espace. Une recherche de maxima sur des images de différence inter coupe permet de réduire les fausses détections parmi tous ces candidats. L'inconvénient de la méthode est son aspect bidimensionnel alors que les tumeurs portent des informations tridimensionnelles. La Figure 2-2 présente des images TEP cliniques après détection originale des candidats (a, c) et après diminution des fausses détections (b, d). Cette figure montre également la taille moyenne des tumeurs ciblées qui est élevée et commence à sortir de la problématique de détection difficile. De nombreuses fixations physiologiques persistent (dans le cerveau par exemple) et les seuls résultats quantitatifs présentés reposent sur une diminution des candidats détectés au cours des étapes du système. La méthode n'est cependant suffisamment détaillée dans l'article de proceeding. Elle n'a d'ailleurs fait que l'objet d'une conférence en 2004, qui n'a pas été complétée par un article publié.

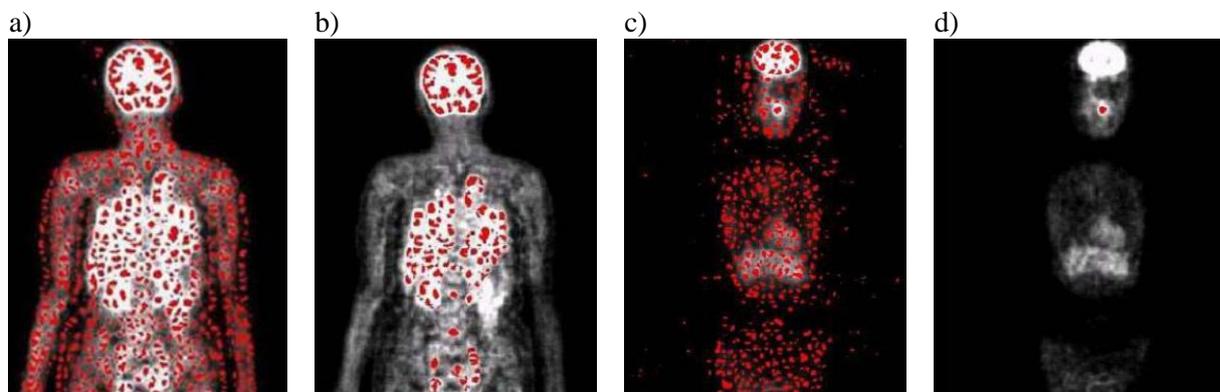


Figure 2-2 : Résultats qualitatifs obtenus par la méthode développée par [Tozaki et al., 2003; Tozaki et al., 2004] pour deux images cliniques différentes (a, b) et (c, d). Les candidats retenus par le CAD apparaissent en rouge.

Le système CAD développé par [Kanakatte et al., 2008] se base sur l'utilisation de classifieurs supervisés que sont les k plus proches voisins et les séparateurs à vaste marge (SVM) dont nous détaillons le principe au paragraphe 2.4.3.3. Une première sélection de voxels potentiellement pathologiques est extraite par seuillage de l'image TEP originale. Le seuil est reporté à une valeur de SUV de 2 proche de celui utilisé en clinique (2,5) pour extraire les tumeurs potentielles. Un ensemble de caractéristiques statistiques de type moment du premier et second ordre est extrait des ROIs de taille 32x32 voxels contenant les candidats initiaux. Des caractéristiques spatiales (transformée en cosinus discrète) et fréquentielles (moments statistiques de coefficients d'ondelette) sont aussi ajoutées pour chaque candidat. Les ondelettes sélectionnées sont la Daubechies d'ordre 2 et 3 (db2 et db3), la Coiflet (coif2, coif3) et la Biorthogonales (bior2.8, bio3.9). Ces caractéristiques sont ensuite envoyées

aux deux classifieurs supervisés (k plus proches voisins et SVM) qui sont évalués séparément. L'étude montre que les performances optimales sont atteintes pour une combinaison des k plus proches voisins et de la transformée en cosinus discrète. Une précision élevée de 98% est atteinte correspondant à une sensibilité de 92% et une spécificité de 99%. Ces performances sont obtenues par validation croisée sur un nombre réduit de huit images TEP cliniques, chacune constituée de 47 coupes. Il faut cependant remarquer que le calcul de fausses détections (faux positifs) est réalisé par coupe. En d'autres termes, la présence d'un ou plusieurs candidats FP sur une coupe ne contenant pas de tumeurs réelles n'incrémente le compteur des FP que d'une unité. Cette méthodologie ne permet donc pas de réellement quantifier l'affection de l'image volumique par les fausses détections. L'image TDM correspondant à l'image TEP étudiée permet enfin le contourage des tumeurs. La Figure 2-3 présente une coupe d'image TDM originale (a), la coupe de l'image TEP correspondante (b), dont est extraite la tumeur cible apparaissant en rouge sur l'image recalée (c). On peut aussi constater que la tumeur cible est de taille importante.

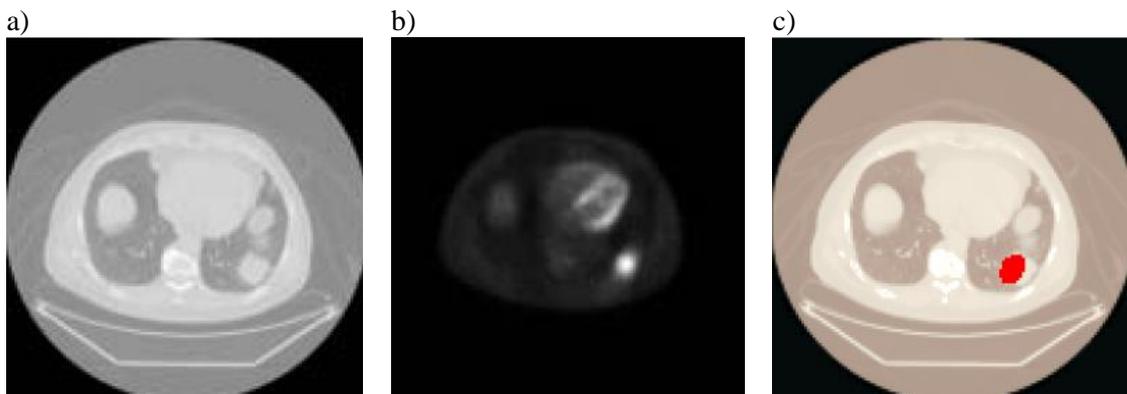


Figure 2-3 : Coupes d'une image TDM (a), TEP (b) contenant une tumeur dans le poumon. Cette tumeur est correctement détectée par le CAD (c).

2.2.4 Les CAD combinant l'information anatomique et fonctionnelle

Plusieurs études récentes présentent des systèmes CAD adaptés aux images TEP/TDM de manière à bénéficier d'une meilleure localisation des tumeurs.

[Cui et al., 2008] utilise l'information TDM pour segmenter les poumons des images bimodales, qui sont ensuite traités par seuillage de mesures de SUV (valeur SUV seuil de 1,5). Les candidats restants sont agrégés par groupe de voxels de SUV proches. L'image TDM est ensuite utilisée afin de ne conserver que les vraies tumeurs, devant répondre à des critères de distributions d'intensité, de forme et de localisation vérifiés dans les deux types d'imagerie. La méthode est testée sur 9 images cliniques contenant 31 nodules pulmonaires, dont 30 sont correctement détectés. Ces performances sont plus que satisfaisantes et prouvent l'efficacité de l'information anatomique additionnelle. Aucune mesure de spécificité n'est fournie par l'auteur étant donné que les images cliniques n'ont pas toutes été précisément annotées (positions de toutes les tumeurs notamment).

[Jafar et al., 2006] utilise un multi-seuillage d'intensité et une sélection par le volume des groupes de voxels conservés, avec comme critère un diamètre compris entre 5 mm et 3 cm. La méthode se base sur des images TDM, TEP corrigées et TEP non corrigées de l'atténuation. Ces dernières ont pour rôle d'éviter la potentielle perte d'information due à une correction d'atténuation imparfaite. Un ensemble de règles de logique floue basé sur des mesures de SUV, de contraste et d'intensité des structures permet ensuite de classer les différents candidats. La validation du système CAD utilise également des performances 'par coupe' comme [Kanakatte et al., 2008] et ne repose que sur trois images TEP/TDM cliniques. La sensibilité atteint 75% pour 100% de spécificité. Un exemple de coupe d'images TDM (a), TEP non corrigée de l'atténuation (b), TEP corrigée (c) et de candidats (d) est présenté sur la Figure 2-4. La tumeur, correctement détectée, est pointée par la flèche. L'utilisation de la logique floue semble adaptée et présente plusieurs niveaux de degré de suspicion et la méthode profite de l'information mutuelle de différents types d'images.

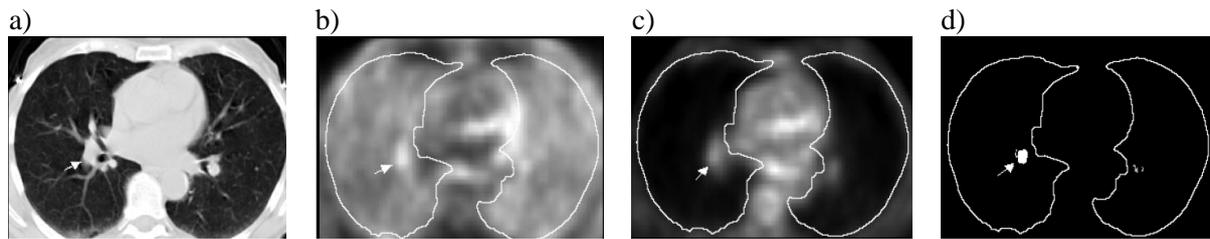


Figure 2-4 : Coupes représentatives des résultats obtenus par la méthode de [Kanakatte et al., 2008] : image TDM originale (a), image TEP non corrigée de l'atténuation (b), image TEP corrigée (c) et image représentant les candidats potentiels.

[Nie et al., 2006] introduit quant à lui les réseaux de neurones associés à des informations sémantiques, anatomiques et de type SUV sur des tumeurs pulmonaires préalablement segmentées manuellement. Le système développé est plutôt référencé comme un CADi que comme un CAD. Les tumeurs candidates sont en effet sélectionnées manuellement. Les performances de cette méthode, exprimées en terme d'aire sous la courbe ROC qui sera présentée ultérieurement au paragraphe 2.5.3 ($Az = 0,95$), ne peuvent ainsi pas être comparées avec celles des études précédentes. Cette étude porte sur la différenciation de tumeurs pré-détectés et non pas sur la détection propre de ces tumeurs. Il est cependant intéressant de voir l'intégration des informations fonctionnelles extraites des images TEP à un CADi dédié à l'imagerie TDM. Ainsi, l'inconvénient majeur de ce CADi repose sur son aspect 'semi-automatique', nécessitant la délimitation préliminaire des nodules cibles par le clinicien.

2.2.5 Conclusions

Actuellement peu de systèmes CAD dédiés à l'imagerie TEP existent dans la littérature et aucun n'est disponible sur les consoles des caméras cliniques. Les développements, assez récents, tendent cependant à s'amplifier. On peut remarquer qu'il est souvent difficile de définir les méthodes les plus pertinentes et efficaces.

D'une part, les méthodes d'évaluation ne sont pas standardisées et les rares performances quantitatives fournies sont d'origine variable (par coupe, par cas, par volume). D'autre part, les données utilisées pour l'évaluation des systèmes CAD sont parfois difficiles à collecter, notamment car la vérité terrain est indispensable. Il est ainsi fréquent de voir les méthodes de détection évaluées quantitativement sur moins de dix images cliniques. Enfin, une spécificité de l'imagerie TEP est son aspect 3D corps entier. Les autres domaines d'application, notamment la mammographie et la TDM, reposent sur des analyses 2D ou focalisées sur un seul organe (les poumons ou le foie...). Aucune stratégie d'évaluation des CAD pour l'imagerie 3D, c'est-à-dire tenant compte de l'ensemble des voxels de l'image, n'est actuellement précisément définie.

Cette analyse renforce d'une part **l'originalité que représente un système CAD pour la détection de tumeurs problématiques de faible taille et contraste en imagerie TEP 3D corps entier**. D'autre part, elle souligne **la nécessité corollaire de développer une stratégie d'évaluation des performances de CAD qui intègre la spécificité 3D corps entier de la TEP**.

2.3 Classification supervisée ou non supervisée ?

2.3.1 Introduction

La classification, à l'origine même des systèmes CAD, permet de regrouper des données en sous-ensembles partageant des caractéristiques communes. La suite du paragraphe décrit brièvement les deux familles de classification présentes dans la littérature : la classification supervisée et la classification non supervisée.

2.3.2 La classification non supervisée

L'idée générale de la classification non supervisée est de découper l'espace des observations en un certain nombre de régions disjointes, définies par des frontières, et de décréter que toutes les

observations situées dans une même région de l'espace appartiennent à une même classe. L'espace des observations a la dimension du nombre de caractéristiques descriptives de chaque observation. L'intérêt de transcrire les informations disponibles de l'observation dans un espace de caractéristiques de dimension élevée pour la classification non supervisée a été décrit dans la thèse de Thomas Grenier [Grenier, 2005]. Ainsi ce partitionnement peut être vu comme de la segmentation d'images lorsqu'on considère celle-ci comme un traitement dans l'espace des caractéristiques. Intuitivement, le partitionnement est facilement imaginable en deux ou trois dimensions, il l'est un peu moins en N dimensions.

Le partitionnement repose en général sur une méthode statistique d'analyse de données grâce à laquelle toutes les observations présentes dans chaque région sont proches au sens des critères de proximité définis à l'aide de mesures de distances. Le partitionnement idéal est obtenu lorsque 1) la covariance intraclasse est minimisée impliquant des régions les plus homogènes possibles, 2) l'inertie interclasse est maximisée impliquant des régions bien différenciées.

Il existe plusieurs méthodes de classification non supervisée, ou segmentation. On peut citer les K-moyennes aussi connues sous le nom de nuées dynamiques, les cartes de Kohonen et le regroupement ascendant hiérarchique.

La méthode des K-moyennes [Herwig et al., 1999] ou nuées dynamiques permet d'obtenir les meilleures régions de l'espace regroupant les observations dont la distance de chacune d'elles au représentant de cette région est minimisée (minimisation de la variance). L'écart entre les classes est également maximisé. La valeur de K, correspondant au nombre de classes finales, est fixée par l'utilisateur. Chaque région de l'espace est représentée par son centre ou par un ensemble de points représentatifs de la région.

Les cartes de Kohonen [Kohonen, 1982] reposent sur une projection (combinaison linéaire) optimale de l'espace des observations dans un espace de dimension inférieure tout en conservant leur positionnement relatif dans l'espace des données (topologie).

Le regroupement ascendant hiérarchique [Ward, 1963] permet de répartir un ensemble d'observations dans un certain nombre de classes selon une mesure de dissimilarité entre ces mêmes observations. Le point de départ repose sur un nombre de classes égales au nombre d'individus ou observations (un individu = une classe). A chaque étape, les deux classes de similarité maximale sont fusionnées jusqu'à atteindre le nombre de classes défini par l'utilisateur. L'objectif de ces méthodes est de définir le nombre de classes optimales pour le problème soulevé. Un dendrogramme permet de représenter les itérations du regroupement hiérarchique. Il s'agit d'une représentation graphique où les abscisses correspondent aux observations. Le regroupement de deux individus en une même classe est matérialisé par un trait horizontal.

Dans le cas de la classification non supervisée, les observations que l'on souhaite classifier sont les seules disponibles pour réaliser le partitionnement. La classification non supervisée n'a pas pour objectif final de caractériser les régions de l'espace obtenues par partitionnement, mais plutôt d'extraire de manière optimale les frontières de ces régions sans en connaître leur étiquette (tumeur/non tumeur).

2.3.3 La classification supervisée

L'objectif de la classification supervisée apparaît nuancé par rapport à celui de la classification non supervisée. Plus qu'un partitionnement de l'espace en régions disjointes, l'objectif principal est d'étiqueter chaque nouvelle observation selon son appartenance à une des classes définies par apprentissage. Le classifieur dispose d'exemples d'apprentissage étiquetés, c'est-à-dire pour lesquels la classe d'appartenance est connue a priori. Le classifieur apprend donc à reconnaître les différentes caractéristiques discriminant chaque classe et utilise ce modèle pour classifier toute nouvelle observation. Les classes d'appartenance (saine/pathologique) sont définies au préalable par l'utilisateur. L'inconvénient principal de la classification supervisée repose sur la nécessité d'avoir une base d'observations suffisamment large et représentative pour l'apprentissage et le test du classifieur.

2.3.4 Conclusions

Il apparaît assez clairement que les objectifs des deux familles de classification décrites précédemment sont légèrement différents. La classification non supervisée, souvent référée à de la segmentation, apparaît en amont ou en aval de la détection en tant que telle des tumeurs. Elle permet par exemple de découper une image en organes ou structures différents. Un algorithme de classification supervisée peut ensuite être appliqué à ces régions pré-segmentées. La classification supervisée peut ainsi caractériser les régions découpées de l'espace par une étiquette de classe connue a priori.

Les CAD dédiés à l'imagerie TEP reposent, soit sur du traitement d'image, seuillage notamment, soit sur des méthodes de classification supervisée. On verra par la suite que la majorité des CAD dédiés aux autres modalités (mammographie, TDM) privilégient l'utilisation des classificateurs supervisés.

Notre choix principal s'est donc porté sur l'utilisation d'algorithmes de classification supervisée, jouissant de la connaissance a priori des étiquettes de classe permettant la génération d'un modèle prédictif caractéristique du problème traité, prêt à classer toute nouvelle observation.

Les CAD supervisés reposent généralement sur un schéma méthodologique standardisé. L'ensemble de ces étapes associé à des exemples de la littérature est détaillé dans le paragraphe suivant.

2.4 Les CAD supervisés : un schéma méthodologique standardisé

2.4.1 Introduction

Les méthodes d'aide à la détection dédiées à l'imagerie TEP présentées au paragraphe 2.2 restent limitées de part leur nombre et leurs critères de validation. Le travail de cette thèse s'est donc largement inspiré de la méthodologie suivie lors du développement de systèmes CAD dans d'autres modalités d'imagerie médicale et particulièrement la mammographie et l'imagerie TDM. La classification supervisée permet de profiter de la disponibilité a priori d'images préalablement classées pour fournir un 'superviseur' efficace et paramétrable. Le paragraphe suivant définit les différentes étapes standardisées d'un système CAD supervisé au gré d'exemples de la littérature.

2.4.2 Les différentes étapes d'un système CAD supervisé

Comme rappelé par [Li, 2007], un système CAD est constitué de deux étapes majeures : l'identification initiale des tumeurs et la réduction de faux positifs (notion définie au paragraphe 2.2.2).

L'objectif de l'identification initiale des tumeurs (partie haute de la Figure 2-5) est de localiser les structures potentiellement suspectes de l'image avec la plus grande sensibilité possible au coût d'un grand nombre de fausses détections. Différents choix méthodologiques doivent être précisés pour cette étape :

- Les régions ou organes d'intérêt du système, lorsque celui-ci n'est pas dédié au corps entier, peuvent être séparés. L'image ou les régions d'intérêt de l'image peuvent ensuite être améliorées lors d'une étape de pré-traitement, reposant par exemple sur du rehaussement du contraste ou du débruitage (point 1 de la Figure 2-5).
- Le processus d'extraction des tumeurs suspectes passe ensuite par l'utilisation d'un système de classification (point 3 de la Figure 2-5), optimisé par un apprentissage, s'appliquant sur un ensemble de caractéristiques de l'image ou de la structure cible (point 2 de la Figure 2-5). Ce dernier a pour but de déterminer la classe, « tumeur présente » ou « tumeur absente », de l'image ou de la structure cible.

L'objectif de l'étape de réduction de faux positifs (partie basse de la Figure 2-5) est de fortement réduire les fausses détections extraites de l'étape précédente, tout en conservant une sensibilité élevée. Cette étape rassemble un ensemble de méthodes permettant de trier les tumeurs candidates précédemment extraites afin de ne conserver que les candidats les plus 'suspects' (point 4 de la Figure 2-5).

L'ensemble de ces points est repris au paragraphe suivant.

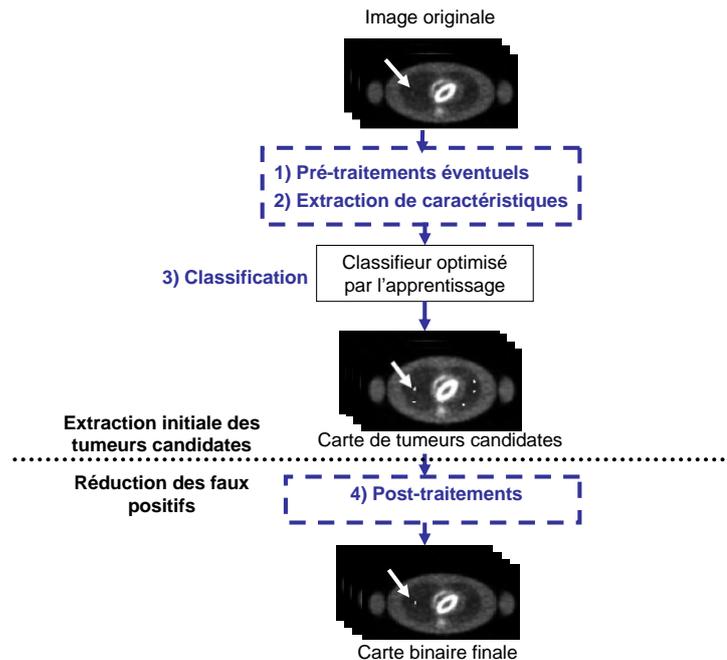


Figure 2-5 : Les différentes étapes d'un système CAD

2.4.3 Extraction initiale des tumeurs potentielles

2.4.3.1 Prétraitements de l'image

Cette étape (point 1 de la Figure 2-5) a pour objectif de préparer et améliorer l'image initiale à l'extraction des tumeurs candidates.

Il est fréquent qu'un système CAD soit évalué et optimisé pour un organe particulier, même si la logique veut que celui-ci puisse être adaptable à n'importe quel organe potentiellement affecté par des tumeurs. Cette focalisation est particulièrement adaptée pour les images TDM et mammographiques qui ne ciblent intrinsèquement qu'un seul organe (les poumons, le foie ou le sein...). L'image TEP a, quant à elle, la spécificité d'être corps entier et tridimensionnelle, ce qui peut impliquer la localisation des différents organes. L'intérêt d'initialement focaliser le développement des méthodes de classification sur un seul organe cible permet de mieux tenir compte de ses caractéristiques particulières en terme de bruit et de texture par exemple. Son inconvénient est la génération de systèmes à adapter à chaque nouvelle région cible et de ce fait ne pouvant pas réaliser de traitement global sur les images corps entier.

Des méthodes généralement simples sont utilisées pour segmenter l'organe cible avant tout traitement. Des seuillages ou des méthodes de partitionnement par k-moyennes (classification non supervisée) sont généralement utilisés sur des images anatomiques [Brown et al., 2003; Bae et al., 2005; Kim et al., 2005; Jafar et al., 2006].

Dans le cas de l'imagerie TEP, fournissant une information fonctionnelle et non pas anatomique, la segmentation des organes est assez problématique. [Ying et al., 2004] utilisent l'image de transmission acquise pendant l'examen TEP classique. Les poumons à segmenter apparaissent sur l'image en transmission par un niveau de gris modéré (rempli d'air), séparé du reste du corps d'atténuation plus élevée. Un seuillage par histogramme permet d'extraire les points de départ d'une croissance de régions dont le critère d'arrêt est choisi empiriquement afin d'optimiser les résultats visuels. La Figure 2-6 illustre une coupe de poumons segmentée par cette méthode. Une autre méthode repose sur l'utilisation de l'information anatomique apportée par l'image TDM recalée sur le volume TEP. Cette méthode est rendue possible par la disponibilité des caméras combinées TEP/TDM permettant d'extraire des volumes bimodaux rigoureusement recalés. Cette approche nécessite cependant un plus grand espace de stockage (données TDM et TEP) et requiert un recalage précis des images TDM et TEP, même lorsque celles-ci ne sont pas issues de caméra multimodale. [Guan et al., 2006] utilise une autre technique basée un modèle de Markov caché basé sur deux caractéristiques que

sont la captation du FDG dans les poumons et la projection sagittale de la courbe de la nuque. D'autres études dédiées à la TEP portent directement sur l'image entière sans cibler d'organes particuliers. C'est notamment le cas de [Tozaki et al., 2003; Tozaki et al., 2004] qui utilise l'image TEP corps entier.

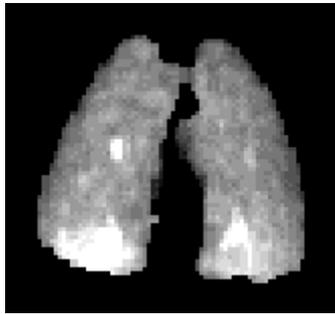


Figure 2-6 : Coupe de poumons segmentée d'une image TEP corrigée de l'atténuation. La segmentation est obtenue à partir d'une méthode par croissance de régions sur l'image acquise en transmission correspondante.

En ce qui concerne l'amélioration de l'image à traiter, cette étape n'est pas obligatoire et certaines études utilisent simplement l'image originale [Mougiakakou et al., 2007]. D'autres préfèrent améliorer l'image initiale, de manière à influencer les performances finales de classification. En imagerie TDM, des méthodes de rehaussement de contraste sont mises en place à partir de filtres sélectifs [Li et al., 2003]. Elles permettent aussi parfois d'éliminer les structures gênantes (vaisseaux sanguins dans les poumons...). En mammographie, il est également courant de débruiter les images par des techniques de filtrage [Bazzani et al., 2001; Boroczky et al., 2006; Papadopoulos et al., 2008]. Quelque soit l'imagerie, le rehaussement des images permet dans la majeure partie du temps d'améliorer les résultats de classification [Paik et al., 2004; Tourassi et al., 2008]. Cependant, l'étude de [Ge et al., 2005] a montré que le débruitage peut aussi être à l'origine d'une baisse de la sensibilité lorsque le filtre n'est pas adapté aux images ciblées.

2.4.3.2 Les caractéristiques descriptives

Un système de classification ne traite que très rarement l'image brute telle quelle. Différentes caractéristiques descriptives de l'image ou du cas clinique (patient) (point 2 de la Figure 2-5) sont généralement préférées. Celles-ci sont en effet porteuses d'un nombre plus important d'informations facilitant leur interprétation. Ces caractéristiques sont assimilées à ce que l'œil humain analyse avant de poser son diagnostic sur la présence ou pas de tumeurs. Elles sont l'interprétation de l'image d'un point de vue numérique et leur choix est primordial. Elles doivent en effet permettre au classifieur de discriminer les éléments pathologiques des éléments sains et être caractéristiques de chacune des classes.

[Cheng et al., 2006] présente un état de l'art des différents types de caractéristiques utilisées classiquement, classées selon leurs catégories. On peut différencier plusieurs familles de caractéristiques selon leur type :

- les régions d'intérêt de l'image
- les caractéristiques statistiques
- les caractéristiques cliniques
- les caractéristiques fréquentielles
- les caractéristiques sémantiques

Les paragraphes suivants reprennent chacune de ces catégories au travers d'exemples d'utilisation de systèmes CAD dédiés à différentes modalités d'imagerie médicale.

Régions d'intérêt de l'image (ROIs) et intensités

Que ce soit en imagerie anatomique ou fonctionnelle, les tumeurs apparaissent visuellement avec un degré de distinction variable par rapport au fond de l'organe d'appartenance. Dans le cas de la mammographie et de la TEP au 18F-FDG, les tumeurs sont de captation généralement plus élevée que

le tissu environnant. L'imagerie TEP associée à d'autres radiotraceurs que le 18F-FDG peut aussi être caractérisée par un hyposignal dans les tumeurs. Dans le cas de la TDM, les tumeurs correspondent à des nodules de forme spécifique (ellipsoïde) et d'atténuation plus élevée. Etant donné que les tumeurs peuvent, dans certains cas, être caractérisées par la valeur même des pixels ou voxels les constituant, certaines études les utilisent telle quelle comme caractéristiques d'image.

En mammographie, [El-Naqa et al., 2002] sélectionne par exemple des ROIs de taille légèrement supérieure aux microcalcifications ciblées (9 x 9 pixels) comme caractéristiques suffisamment discriminantes à envoyer au classifieur supervisé. [Boroczky et al., 2006] décrit également les nodules pulmonaires cibles par leur distribution d'intensité tridimensionnelle bien que cette caractéristique soit accompagnée d'autres types de composantes discriminantes. Dans certains cas, il n'est pas suffisant de caractériser une tumeur par son intensité, notamment lorsque le contraste est trop faible, que l'image est fortement bruitée ou que des structures anatomiques ou physiologiques saines se rapprochent de la forme observée. La Figure 2-7 met en évidence une telle limite à partir d'un ensemble de ROIs mammographiques extraites de la base d'images sélectionnée par [El-Naqa et al., 2002] pour évaluer le CAD développé dans son étude. La différence entre les ROIs représentant les microcalcifications présentes (Figure 2-7a) et absentes (Figure 2-7b) est en effet parfois très confuse.

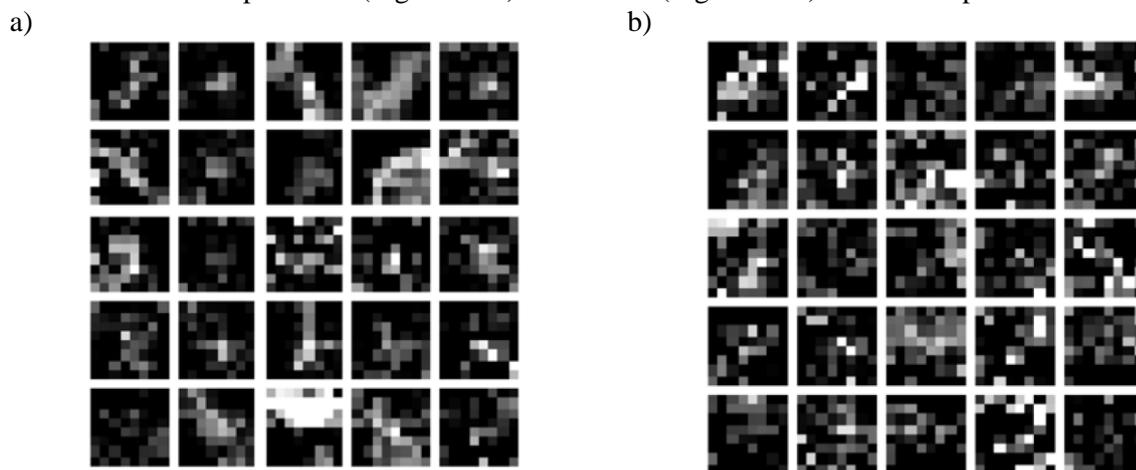


Figure 2-7 : Ensemble de ROIs de taille 9 x 9 pixels extraites de mammographies (El-Naqa, Yongyi et al. 2002) correspondant à des microcalcifications (classe tumeur présente) a) et des structures saines (classe tumeur absente) b).

Caractéristiques statistiques

Les caractéristiques statistiques reflètent généralement les propriétés de bruit et de texture des zones étudiées. Les statistiques du premier ordre caractérisent les niveaux de gris de l'image sans prendre en compte leur distribution relative. Des mesures de moyennes ou d'histogrammes peuvent être extraites. Les statistiques d'ordre deux représentent généralement des propriétés de texture. Celle-ci est assimilée à l'apparence locale de l'image, c'est à dire l'organisation des détails d'une petite partie de l'image telle que le système visuel humain l'aperçoit.

[Haralick et al., 1973] a proposé un ensemble de 14 caractéristiques qui peuvent être utilisées afin d'homogénéiser les mesures extraites pour les différentes modalités d'imagerie. Ces mesures proviennent de la matrice de cooccurrence de l'image. Celle-ci représente l'estimation de la probabilité de transition d'un niveau de gris à un autre. En d'autres termes, elle détermine la fréquence d'apparition d'un motif formé par deux pixels de niveau de gris différents. Chaque élément $P(\theta, d)$ de la matrice de cooccurrence P représente le nombre de pixels de niveau de gris a ayant pour voisin un pixel de niveau de gris b à la distance $d = (dx, dy) = (d \cos \theta, d \sin \theta)$. θ représente l'angle de la droite reliant les deux pixels par rapport à l'horizontale. $P(\theta, d)$ est de taille $N \times N$ avec N le nombre de niveau de gris a et b de l'image. L'exploitation de cette matrice de cooccurrence se fait au travers d'indices de texture tel que l'entropie, le moment angulaire du deuxième ordre (énergie), le contraste, la corrélation et l'homogénéité...

Ainsi, les études de [Gletsos et al., 2003; Lin et al., 2005] décrivent leurs ROIs en mammographie par un ensemble de 48 caractéristiques statistiques toutes dérivées de huit composantes extraites des matrices de cooccurrence pour chaque θ et d comme l'entropie et la corrélation, dont les formules sont données aux équations (2-4) et (2-5).

$$\text{entropie} = -\sum_a \sum_b P_{ab} \ln(P_{ab}) \quad (2-4)$$

$$\text{corr\u00e9lation} = -\frac{\sum_a \sum_b (ab) P_{ab} - \mu^2}{\sigma^2} \quad (2-5)$$

μ et σ sont respectivement la moyenne et l'\u00e9cart type de la matrice de co-occurrence et P_{ab} est l'\u00e9l\u00e9ment (a,b) de la matrice de co-occurrence, calcul\u00e9s pour chaque θ et d .

Dans l'article [Mougiakakou et al., 2007], qui est une am\u00e9lioration de [Gletsos et al., 2003], les moments d'ordre un apparaissent d'ailleurs comme les plus performants parmi toutes les caract\u00e9ristiques.

En imagerie TEP, la taille r\u00e9duite des zones d'int\u00e9r\u00eat (tumeurs) implique une caract\u00e9risation de leur texture limit\u00e9e. Cette famille de caract\u00e9ristiques peut donc parfois \u00eatre rendue insuffisante ou inappropri\u00e9e pour une t\u00e2che de discrimination.

Caract\u00e9ristiques g\u00e9om\u00e9triques

Les caract\u00e9ristiques g\u00e9om\u00e9triques des tumeurs, exprim\u00e9es par exemple par des crit\u00e8res de convexit\u00e9, peuvent aussi \u00eatre discriminantes. Elles sont g\u00e9n\u00e9ralement simples \u00e0 interpr\u00e9ter car elles correspondent aux d\u00e9tails visuels sur lesquels se focalisent consciemment les cliniciens. Cette famille de caract\u00e9ristiques est g\u00e9n\u00e9ralement d\u00e9di\u00e9e aux imageries de type anatomique \u00e9tant donn\u00e9 qu'elle repose sur une d\u00e9lin\u00e9ation pr\u00e9cise des structures \u00e9tudi\u00e9es.

Plusieurs CAD adoptent cette technique notamment pour l'imagerie TDM. Dans ce cas, les tumeurs initiales sont g\u00e9n\u00e9ralement pr\u00e9-segment\u00e9es manuellement ou semi-manuellement et leur forme peut facilement \u00eatre extraite. Le syst\u00e8me CAD cibl\u00e9 sur le cancer du poumon d\u00e9velopp\u00e9 par [Li et al., 2008] se base sur l'extraction d'un ensemble de caract\u00e9ristiques de forme discriminantes. On peut citer l'irr\u00e9gularit\u00e9 de la tumeur pr\u00e9 segment\u00e9e, son degr\u00e9 de compacit\u00e9 et son diam\u00e8tre effectif. [Lin et al., 2005; Boroczky et al., 2006] ajoutent aussi la notion de sph\u00e9ricit\u00e9 pour des nodules pulmonaires. Il est \u00e9galement possible d'utiliser un crit\u00e8re de localisation, et notamment de sym\u00e9trie afin de caract\u00e9riser la pr\u00e9sence de tumeurs thoraciques comme le fait [Yoshida, 2004] en imagerie TDM.

Le principal inconv\u00e9nient des caract\u00e9ristiques de forme est d'\u00eatre fortement corr\u00e9l\u00e9 \u00e0 la segmentation initiale des tumeurs \u00e9tudi\u00e9es. Celle-ci peut parfois s'av\u00e9rer approximative, notamment dans le cas de tumeurs de petite taille. Cette famille de caract\u00e9ristiques est aussi difficilement exploitable dans le cas de l'imagerie fonctionnelle.

Caract\u00e9ristiques fr\u00e9quentielles

Selon [Sachs et al., 1971], le syst\u00e8me visuel humain traite l'information \u00e0 travers des canaux s\u00e9lectifs en fr\u00e9quence. Il est ainsi possible de se rapprocher d'une mod\u00e9lisation assez fiable du syst\u00e8me visuel humain en extrayant des caract\u00e9ristiques fr\u00e9quentielles des images cibles.

La d\u00e9composition en ondelettes a une place pr\u00e9pond\u00e9rante dans la famille des caract\u00e9ristiques d'images de type fr\u00e9quentielles. Elle repose sur une base de fonctions form\u00e9e d'une ondelette «m\u00e8re», qui est une fonction de moyenne nulle et de support compact, dilat\u00e9e ou compress\u00e9e (ce qui introduit un facteur d'\u00e9chelle) et translat\u00e9e afin de couvrir tout le domaine de la fonction d\u00e9compos\u00e9e. Le facteur de translation d\u00e9termine sur quel intervalle du domaine de d\u00e9finition de la fonction on se place, alors que le facteur d'\u00e9chelle calibre le domaine fr\u00e9quentiel. Le positionnement conjoint dans les domaines spatial et fr\u00e9quentiel et l'aspect multir\u00e9solution de cette transformation est d'un attrait particulier dans la d\u00e9tection d'objets Gaussiens de taille r\u00e9duite noy\u00e9s dans du bruit haute fr\u00e9quence.

D\u00e8s le milieu des ann\u00e9es 90, les \u00e9tudes de [Strickland and Hahn, 1996; Strickland and Hahn, 1997] ont mis en \u00e9vidence l'int\u00e9r\u00eat des ondelettes d\u00e9finies par [Rioul and Vetterli, 1991; Daubechies, 1992] pour la d\u00e9tection automatique de microcalcifications en mammographie. La version discr\u00e8te non

décimée de la décomposition en ondelettes [Shensa, 1992] est privilégiée dans ces deux études de manière à conserver la même taille des images de détails et d'approximation le long des niveaux d'échelle. [Strickland and Hahn, 1996; Strickland and Hahn, 1997] montrent également que l'utilisation d'ondelettes biorthogonales apparait comme une alternative à l'utilisation de filtres adaptés dans le cas particulier de la détection d'objets d'allure Gaussienne dans un bruit de statistique Gaussienne. L'étude de [Riyahi-Alam et al., 2004] utilise également les coefficients d'une décomposition en ondelettes non décimée dont il extrait les maxima le long des échelles pour un CAD en mammographie. Toujours pour la mammographie, la méthode de [Regentova et al., 2007] extrait également les maxima de coefficients d'une décomposition en ondelettes décimée et les associe à un arbre de Markov caché.

La difficulté essentielle de cette famille de caractéristiques repose sur le choix de la fonction ondelettes et de son niveau de décomposition qui peut influencer sur les images de détail et d'approximation obtenues.

Caractéristiques textuelles

La dernière grande famille de caractéristiques rassemble des informations qui ne sont plus directement extraites des images, mais du dossier médical du patient. Ce type de caractéristiques est communément appelé sémantique. L'ensemble des informations personnelles du patient se retrouve dans cette catégorie, allant de son âge à ses habitudes alimentaires. Cette famille de caractéristiques permet de tenir compte de l'influence de paramètres extérieurs à l'image, mais qui influencent néanmoins le diagnostic du clinicien au cours de l'interprétation des images.

Ces caractéristiques ont été jusqu'à présent principalement incorporées dans des CAD dédiés à l'imagerie TDM. Elles caractérisent les différentes familles de malades et les critères personnels complexes influençant l'évolution des pathologies cancéreuses. [Nie et al., 2006] inclue par exemple l'âge, le sexe, l'historique des maladies et le statut de fumeur ou non fumeur dans son système CAD dédié aux images couplées TEP/TDM pulmonaires. [Tang et al., 2008] profite également d'informations sémantiques pour enrichir son système CAD initialement basé sur des informations de type clinique (forme et localisation des microcalcifications). Cette méthode permet de mieux interpréter le comportement des cliniciens face aux différentes images et patients. [Raicu et al., 2007] a d'ailleurs démontré de fortes corrélations entre certaines caractéristiques de forme et un ensemble de caractéristiques sémantiques associées par les cliniciens sur ces mêmes images pour la détection de tumeurs pulmonaires en imagerie TDM.

Les caractéristiques sémantiques sont intéressantes dans le sens où elles reproduisent le comportement du clinicien face à des patients dont l'état de santé dépend généralement d'un contexte extérieur à l'imagerie. Cependant, l'utilisation d'informations sémantiques uniquement peut parfois paraître restrictive dans les cas où les patients ne répondent pas directement aux critères de santé attendus. Ce type de caractéristiques demande également un nombre non négligeable de données supplémentaires à l'imagerie et devant être fourni par le clinicien. Enfin, il apparait parfois difficile de traduire les informations sémantiques de manière numérique comme nécessaire pour certains classifieurs supervisés.

2.4.3.3 Le classifieur supervisé

Introduction

Un problème de classification consiste à définir l'étiquette d'une donnée (image, pixels, groupe de pixels) parmi les différentes classes possibles (point 3 de la Figure 2-5). Dans le cas de classification supervisée, un ensemble de données dont l'étiquette de classes est connue, permet d'apprendre au classifieur à reconnaître les caractéristiques propres de chaque classe. En d'autres termes, un classifieur est un modèle prédictif dont le rôle est d'accomplir la prédiction de la classe d'une observation en ayant en mémoire les classes disponibles. Pour cela, il apprend à reconnaître les données des différentes classes en extrayant de l'ensemble des informations (ensemble de caractéristiques défini au paragraphe 2.4.3.2) relatives aux données, celles qui sont les plus pertinentes et utiles. Chaque nouvelle observation est ensuite comparée avec celles déjà classées de manière à trouver la classe avec laquelle elle a la plus forte ressemblance. Un classifieur peut également associer

une valeur de score à chaque nouvelle observation. Ce score est représentative la probabilité d'appartenance à une classe.

Cette thèse se place dans une problématique de classification bi-classe avec H0 la classe tumeur absente et H1 la classe tumeur présente.

La mise en place d'un modèle prédictif se heurte généralement à divers problèmes dont deux prépondérants que sont 1) le chevauchement possible des classes et 2) la forme de ces classes qui peut être complexe dans l'espace de départ. La modélisation de tout classifieur est généralement soumise au compromis problématique entre la minimisation de l'erreur de classification et la complexité de la règle de décision. Ce compromis est exprimé au travers de la théorie de l'apprentissage de Vapnik et Cortes qui est détaillé au paragraphe suivant. Les classifieurs supervisés les plus fréquemment utilisés dans les systèmes CAD en imagerie médicale sont ensuite définis et illustrés au travers d'exemples de la littérature.

Apprentissage et théorie de Vapnik et Cortes

Vapnik et Cortes [Vapnik, 1995] sont à l'origine du développement de la théorie statistique de l'apprentissage. Elle est à l'origine de tout problème de classification supervisée et définit l'estimation statistique du modèle prédictif à partir d'échantillons d'apprentissage de taille finie. La théorie de Vapnik et Cortes (VC) prend en compte la taille, finie, de l'échantillon disponible pour l'apprentissage. Elle fournit une description quantitative du compromis entre complexité du modèle de classification et information d'apprentissage disponible.

Considérons une paire entrée-sortie $z = (V, y)$ de densité de probabilité $P(z)$. Le vecteur v correspond à l'ensemble des caractéristiques de l'échantillon et y à sa classe. Un ensemble d'échantillons disponibles pour l'apprentissage est défini par l couples z_1, \dots, z_l . Ces exemples d'apprentissage sont associés à une densité de probabilité $P(z)$ et une fonction coût $Q(z, \omega)$, pour laquelle ω représente la séparation optimale des deux classes. L'objectif de l'apprentissage est de définir la fonction de classification minimisant le risque théorique R . Ce risque correspond à l'erreur que commettrait le classifieur sur un échantillon d'apprentissage de taille infini et est défini par l'équation (2-6).

$$R(z, \omega) = \int Q(z, \omega) dP(z) \quad (2-6)$$

La fonction $Q(z, \omega)$ peut aussi être notée $L(y, f(V, \omega))$, où f est la fonction qui permet d'obtenir la classe y à partir du vecteur v selon le plan de séparation ω . Cette fonction « coût » est spécifique à chaque problème d'apprentissage. La fonction $L(y, f(V, \omega))$ peut par exemple correspondre à la distance entre la classe donnée en sortie du classifieur et la classe réelle.

Etant donné que la distribution de probabilité P de toute nouvelle donnée est inconnue, il est impossible de directement minimiser le coût L et donc le risque théorique. Deux méthodes permettant de remédier à ce problème sont décrites.

1) La première repose sur la substitution du risque théorique par une autre fonction qui lui est proche. Cette fonction de décision, notée risque empirique R_{emp} , repose sur la disponibilité des l exemples d'apprentissage et est définie par l'équation (2-7).

$$R_{emp}(f) = \frac{1}{l} \sum_{i=1}^l Q[z_i, \omega] = \frac{1}{l} \sum_{i=1}^l L[f(V_i, \omega), y_i] \quad (2-7)$$

Le problème majeur de cette méthode repose sur le fait que le risque empirique R_{emp} est toujours biaisé à cause du choix fini des exemples d'apprentissage. Ces échantillons ne peuvent en effet pas

représenter la totalité des exemples possibles (infini). Le risque théorique, quant à lui, ne dépendant pas d'un échantillon particulier, est bien plus consistant, mais inaccessible.

2) La seconde méthode consiste à estimer la densité de probabilité $P(z)$ à partir des données d'apprentissage disponibles et de trouver l'estimée optimale de la fonction $f(V_i, \omega)$.

Nous verrons dans le paragraphe suivant que l'ensemble des classifieurs supervisés repose sur la théorie statistique de l'apprentissage de VC. L'obtention du modèle de classification optimal se base un compromis entre une minimisation des risques de classification et le choix d'un apprentissage de richesse adéquate avec le problème.

Le classifieur naïf de Bayes

La théorie de la décision Bayésienne repart de la minimisation du risque théorique au travers de l'expression des densités de probabilité du problème. La décision Bayésienne [Bayes, 1763] établit que la stratégie qui consiste à affecter une nouvelle observation à la classe ayant la plus grande probabilité a posteriori est optimale, c'est à dire génère un nombre d'erreurs plus petit que toute autre stratégie. La probabilité a posteriori correspond à la probabilité pour qu'une observation appartienne à une certaine classe connaissant son vecteur de caractéristiques. Le classifieur de Bayes est dit naïf dans le sens où il suppose que les vecteurs de caractéristiques sont indépendants.

La difficulté d'utilisation du classifieur de Bayes repose sur l'estimation des probabilités a posteriori par sa formulation donnée à l'équation (2-8).

$$P(H_i / V) = \frac{P(V / H_i) \cdot P(H_i)}{P(V)} \quad (2-8)$$

$P(V / H_i)$ est la probabilité conditionnelle de la classe i pour le vecteur V (c'est-à-dire la proportion des observations dans la classe H_i correspondant à V), $P(H_i)$ est la probabilité a priori de la classe i et $P(V)$ est la probabilité inconditionnelle de v . Une méthode permettant d'estimer la probabilité a posteriori du système repose sur un postulat sur les distributions H_i , comme pour l'analyse discriminante linéaire.

L'analyse discriminante de Fisher-Hotelling

L'analyse discriminante linéaire de Hotelling [Hotelling, 1931] ou de Fisher [Fisher, 1936], aussi appelée LDA pour *linear discriminant analysis* en anglais, se base sur la méthodologie générale de la discrimination linéaire. Celle-ci a pour objectif de dériver une combinaison linéaire des caractéristiques des images de telle sorte que les caractéristiques projetées dans le nouvel espace soient séparables de manière optimale, c'est-à-dire en minimisant le risque empirique défini précédemment. Cette dérivation aboutit à un plan de projection ω et ne se base que sur le calcul de moments statistiques d'ordre un et deux de ces caractéristiques. Le plan de séparation optimale entre les classes correspond à la normale au plan ω . La Figure 2-8 illustre la méthodologie LDA en considérant deux classes H_1 et H_0 multinomiales dont les caractéristiques sont projetées de manière optimale sur le plan ω . ω est calculé selon l'équation (2-9) nécessitant la connaissance de la moyenne des vecteurs de caractéristiques de chaque classe et de la covariance intra-classe pondérée Σ . L'ensemble de ces mesures sont extraites par apprentissage.

$$\omega = (\langle V_1 \rangle - \langle V_0 \rangle)' \Sigma^{-1} \text{ avec } \Sigma = \frac{1}{2} \left((V - \langle V_0 \rangle)(V - \langle V_0 \rangle)^T + (V - \langle V_1 \rangle)(V - \langle V_1 \rangle)^T \right) \quad (2-9)$$

Le LDA classique dérive lui-même du classifieur idéal de Bayes pour lequel sont inférées des hypothèses de distributions multinormales et de matrice de variances covariances identiques d'une classe à l'autre. Géométriquement, cela veut dire que les nuages de points représentant les observations disponibles pour les deux classes ont la même forme (et volume) dans l'espace de

représentation. Cette démonstration apparaît dans le Chapitre 5. Cependant les adaptations de Hotelling et Fisher s'avèrent également efficaces lorsque l'hypothèse de multinormalité n'est pas satisfaite.

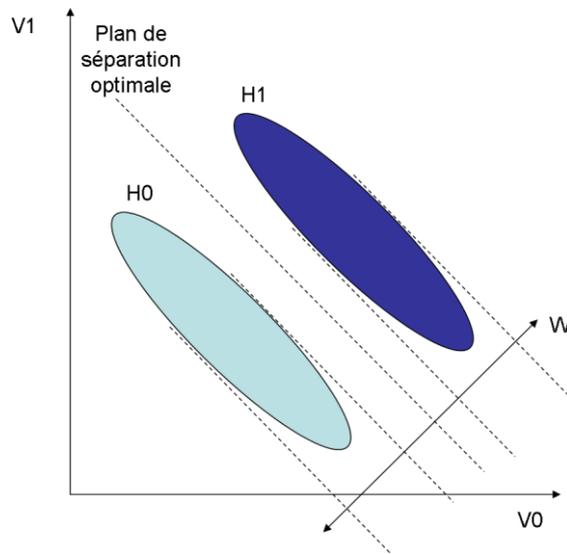


Figure 2-8 : Illustration de la méthodologie de discrimination linéaire dans l'hypothèse de multinormalité des classes H0 et H1 pour un vecteur de caractéristiques à deux composantes (V0 et V1).

Les réseaux de neurones

Le réseau de neurones [Haykin, 1999] est un classifieur non linéaire se basant sur le fonctionnement des neurones humains de manière à imiter le fonctionnement de notre cerveau. Un réseau de neurones est constitué de neurones artificiels interconnectés dont le rôle est de mimer les propriétés des neurones biologiques. Un neurone permet de définir une fonction discriminante dans l'espace d'entrée en réalisant une combinaison linéaire de son vecteur de caractéristiques. Les neurones sont arrangés par couches et comprennent une couche d'entrée, une couche de sortie et des couches cachées. Pour les couches cachées, aucune connexion n'existe entre les couches d'entrée et de sortie. Leur objectif est d'apprendre des caractéristiques d'entrée en utilisant leurs a priori cachés (poids ω de la Figure 2-9). La règle d'arrêt de l'apprentissage repose sur la minimisation du risque empirique (sans postulat sur les densités de probabilité donc). La Figure 2-9 détaille la structure type d'un réseau de neurones.

Malgré des avantages incontestables au vue de performances généralement bonnes, de l'absence d'hypothèse de linéarité et d'un pouvoir de généralisation important, les réseaux de neurones sont assez difficiles à mettre en place et souffrent de leur approche boîte noire. Il est en effet difficile de définir le nombre de couches cachées du classifieur et le nombre de neurones par couche optimal. Dans la littérature actuelle des CAD appliqués en imagerie médicale, les réseaux de neurones sont généralement associés et comparés à une autre famille de classifieurs non linéaires : les séparateurs à vaste marge (ou *support vector machines* en anglais, SVM).

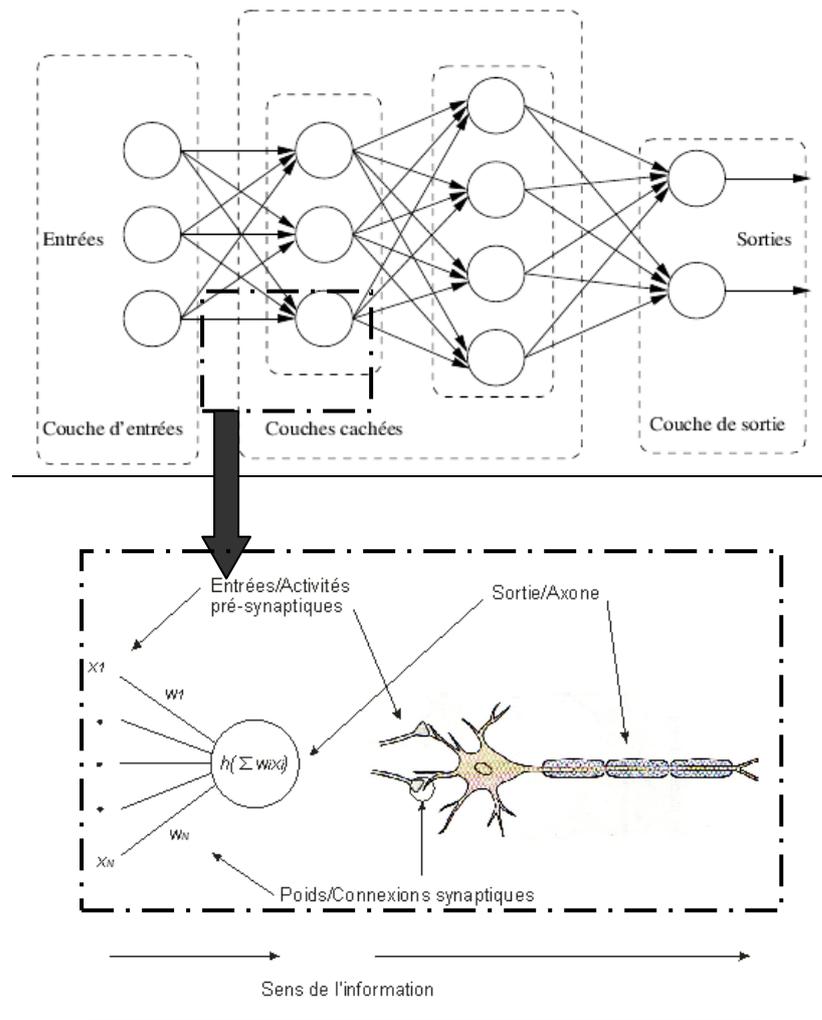


Figure 2-9 : Structures d'un réseau de neurones : un ensemble de couches dont chacune d'entre elle suit le même modèle

Les séparateurs à vaste marge (SVM)

La méthodologie des SVM [Boser et al., 1992; Vapnik, 1995] repose directement sur la minimisation du risque théorique de mauvaise classification détaillé précédemment. L'idée basique des SVM est, comme leur nom l'indique, de trouver l'hyperplan optimal de séparation qui maximise la marge de séparation entre les deux classes H_1 et H_0 . Cette marge correspond en d'autres termes 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. Ces vecteurs de support correspondent aux cas extrêmes des deux classes qui pourraient poser potentiellement des problèmes de classification. Dans le cas où la frontière de décision ne peut pas être définie de manière linéaire, des fonctions à noyaux sont utilisées pour projeter le problème de classification dans un espace de dimension supérieure. Ce nouvel espace a ainsi l'avantage de présenter un hyperplan de séparation optimal linéaire. La théorie des SVM est plus précisément présentée au Chapitre 5.

Les SVM présentent généralement des performances proches de celles des réseaux de neurones. Ils sont connus pour être adaptés aux bases d'apprentissage de petite taille et des ensembles d'observations de grande dimension (nombre de caractéristiques élevé). Leur inconvénient majeur repose sur leur fort aspect paramétrique puisque la modélisation des SVM dépend d'un certain nombre de critères : choix du noyau et de ses paramètres et de différents paramètres intrinsèques.

Les K plus proches voisins (k-PPV)

Cette méthode [Dasarathy, 1991] assez simpliste repose sur la comparaison de toute nouvelle observation aux observations d'apprentissage. La méthode consiste à sélectionner les k plus proches voisins de la base d'apprentissage, au sens de la métrique considérée et à associer à l'observation la classe majoritaire de ces k voisins. En d'autres termes, la méthode des k-PPV n'effectuent aucune hypothèse sur les distributions de probabilité du problème. Elle en propose une procédure d'estimation locale au voisinage de l'observation traitée, permettant de construire la règle de décision f optimale définie par Vapnik et Cortes. La notion de proximité entre voisins est définie par une métrique comme la distance Euclidienne. Il est montré que l'erreur de décision des k-PPV est proche de l'erreur de décision bayésienne dans le cas où k augmente [Fukanaga and Hostetler, 1975]. Le cas le plus simple est pour k=1 pour lequel on cherche le voisin le plus proche de la nouvelle observation et on lui affecte sa classe. L'extension à k>1 impose ainsi d'appliquer une règle de décision à la majorité.

Le principal inconvénient de ce classifieur est sa complexité linéaire au nombre d'exemples d'apprentissage pour le calcul de k plus proches voisins. Les temps de calcul peuvent rapidement devenir prohibitifs. Diverses méthodes permettent cependant de réduire ces temps de calcul.

Les arbres de décision

D'un tout autre style, les arbres de décision [Breiman et al., 1984] sont des classifieurs assez populaires. Ils permettent de séparer l'espace d'entrée en régions (donc classes) disjointes de manière adaptative afin de construire une frontière finale de décision optimale. Les régions sont choisies à partir d'une procédure d'optimisation où chaque étape correspond au choix de la meilleure séparation des classes en accord avec une fonction coût. Cette fonction coût reflète le risque empirique de mauvaise classification précédemment décrit. La séparation en régions est modélisée par une arborescence où chaque feuille contient une fraction de l'échantillon d'apprentissage appartenant presque tous à une seule et même classe. La classification se fait par parcours d'un nouveau vecteur de caractéristiques de la racine de l'arbre à une unique feuille terminale. Ce parcours est contrôlé par les valeurs de ses attributs. Les attributs correspondent à un ensemble de règles, fixé par l'utilisateur, qui peuvent être de type différents (numérique, logique, sémantique...) comme illustré sur la Figure 2-10.

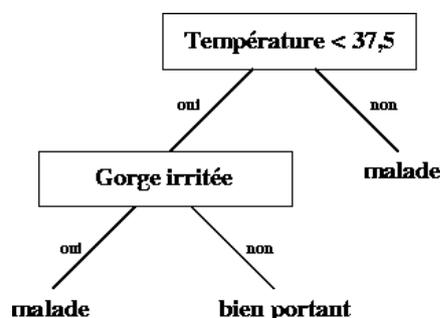


Figure 2-10 : Exemple simpliste d'un arbre de décision à deux nœuds feuilles (malade, bien portant).

Les arbres de décision sont généralement rapides et adaptés au traitement d'un grand nombre d'observations. La décision finale de la classe d'appartenance repose cependant sur un ensemble de règles logiques facilement interprétables, notamment d'un point de vue médical.

Exemples d'utilisation en imagerie médicale

De nombreux articles de revue résument l'utilisation des différents classifieurs dans le domaine de la mammographie et de l'imagerie TDM. Le choix du classifieur et de ces caractéristiques associées varie d'un article à l'autre et les performances de chaque classifieur ne sont pas toujours comparables. Il est donc quasiment impossible d'ordonner les classifieurs selon leur capacité à être ou pas les plus adaptés aux problématiques ciblées. [Cheng et al., 2006] décrit un ensemble de méthodes de classification utilisées pour le développement de systèmes CAD en mammographie, tandis que [Li, 2007] fait de même pour l'imagerie TDM. Au vue d'une large littérature disponible, la suite du paragraphe ne détaille que quelques exemples de classifieurs supervisés utilisés en mammographie et TDM.

[Strickland and Hahn, 1997; Baydush et al., 2007] obtiennent des performances prometteuses avec le LDA combiné à des caractéristiques fréquentielles pour la classification de microcalcifications en mammographie. Pour la même application, [Sahiner et al., 2001] optimise les performances du LDA sur des caractéristiques morphologiques et de texture. Pour la même application, [Samulski et al., 2007] obtient quant à lui des performances similaires par l'utilisation d'un classifieur naïf de Bayes d'une part, c'est-à-dire considérant les distributions des deux classes totalement indépendantes et de SVM d'autre part.

[Campadelli et al., 2005] analyse la robustesse des SVM à un apprentissage à partir de classes de taille très inégale pour la détection de tumeurs pulmonaires en imagerie TDM. Des coefficients de régularisation sont ainsi ajoutés de manière à obtenir des performances sensiblement identiques à des cas de classes correctement réparties. [Boroczky et al., 2006; Lu et al., 2009] utilisent aussi les SVM sur des images TDM pulmonaires.

En ce qui concerne les réseaux de neurones, [Gletsos et al., 2003] base son système CADi dédié aux tumeurs du foie en TDM sur une combinaison séquentielle de trois réseaux de neurones permettant de classer les nodules pré segmentés manuellement successivement selon leur malignité et le type de maladies qu'ils représentent. [Cheng and Cui, 2004] adapte un réseau de neurones à quatre couches à l'aide de règles de logique floue permettant de détecter toutes les microcalcifications présentes dans 100 mammographies au prix d'un faible nombre de faux positifs par image. En mammographie, [Sajda et al., 2002] génère un réseau de neurones hiérarchique pyramidale où chaque entrée reçoit des informations qui lui sont propres ou qui proviennent de la sortie d'un réseau de neurones de hiérarchie inférieure. De nombreuses études similaires portent sur l'adaptation des classifieurs de base vers des algorithmes plus rapides et efficaces [Wu et al., 2002; Zhang and Jiang, 2007; Gil-Pita and Yao, 2008; Leung et al., 2008]

Les différents classifieurs supervisés ont également été comparés à travers différents articles. L'étude de [Corrêa Silva et al., 2005] compare le LDA, les SVM et un réseau de neurones pour la classification de nodules pulmonaires pré-segmentés dont sont extraits des caractéristiques tridimensionnelles. Sur 36 nodules dont 7 malins, le SVM s'est avéré le moins performant au niveau de la sensibilité à spécificité équivalente. [Bazzani et al., 2001; El-Naqa et al., 2002], dans leurs études portant sur un CAD dédié mammographie, présentent des performances légèrement supérieures en faveur des SVM comparées aux réseaux de neurones et au LDA. Cette conclusion est notamment vraie dans le cas d'une base d'apprentissage de taille réduite. Toujours concernant la détection de microcalcifications en mammographie, l'étude de [Garcia-Orellana et al., 2007] reste plus modérée et conclue sur des performances quasiment identiques pour les SVM et les réseaux de neurones. L'étude proposée par [Mougiakakou et al., 2007] en imagerie TDM, évalue cinq systèmes de détection basés sur la combinaison de réseaux de neurones probabilistes et d'un type de réseaux de neurones particuliers, les perceptrons multicouches, auxquels s'ajoute un classifieur par k-plus proches voisins. Les meilleures performances sont obtenues par le système combinant différents classifieurs afin de traiter les caractéristiques de manière variée.

D'autres systèmes reposent sur la combinaison de plusieurs classifieurs supervisés. C'est le cas du CAD développé par [Park et al., 2009] pour la mammographie. Celui-ci se base sur la comparaison et la combinaison des scores obtenus d'un classifieur par k-PPV et de réseaux de neurones. Les auteurs mettent en évidence une faible corrélation entre les résultats des deux classifieurs, justifiant d'une combinaison optimale des deux types de score. L'intérêt de la combinaison implique une amélioration significative des performances initiales. [Zhen and Chan, 2001] génère quant à lui un CAD pour la mammographie basé sur les arbres de décision recevant un ensemble de caractéristiques d'origine fréquentielles extraites de régions suspectes pré-segmentées numériquement.

Conclusions

La classification supervisée répond à la problématique d'étiquetage de toute nouvelle observation à partir d'observations d'apprentissage déjà classées. L'ensemble des méthodes présenté dans les paragraphes précédents répond à ce problème de manière plus ou moins adaptée et optimale. Le Tableau 2-2 résume les principaux avantages et inconvénients de chacune de ces méthodes.

Tableau 2-2 : Avantages et inconvénients des classifieurs existants

Classifieur	Avantages	Inconvénients
LDA	Bonnes performances dans les cas linéairement séparables et respectant l'hypothèse de multinormalité, Bonne robustesse dans le cas où l'hypothèse de multinormalité est approximée	Difficultés à gérer les cas non séparables
Réseaux de neurones	Robuste, aucune règle à préciser, Applicable sur un large spectre de caractéristiques	Mise en œuvre difficile, Temps d'apprentissage parfois long Aspect « Boîte noire »
SVM	Flexible et rapide Formalisme théorique robuste	Mise en œuvre parfois difficile due à un grand nombre de paramètres à fixer
K-PPV	Mise en place simpliste	Complexité linéaire avec la taille de la base d'apprentissage
Arbres de décision	Faible complexité, Applicable sur un large spectre de caractéristiques	Précision dépendant uniquement des paramètres de chaque attribut Moins performants que d'autres classifieurs

2.4.4 Réduction de faux positifs (RFP)

2.4.4.1 Introduction

Après l'identification initiale des tumeurs candidates, l'étape de réduction de faux positifs a pour objectif de trier l'ensemble de ces tumeurs et d'extraire les plus probables tout en éliminant un maximum de tumeurs faussement détectées. Diverses méthodologies peuvent être utilisées pour cette optimisation, cependant les trois principales catégories sont 1) la sélection de caractéristiques en amont du classifieur, 2) l'ajout de nouvelles caractéristiques en aval du classifieur, 3) la combinaison de deux classifieurs différents. Les paragraphes suivants reprennent ces différentes méthodes utilisées dans la littérature.

2.4.4.2 La sélection de caractéristiques et la réduction de dimension

Introduction

Le choix des caractéristiques optimales à envoyer à un classifieur repose, soit sur des méthodes de sélection de certaines d'entre elles, soit sur des méthodes de réduction de dimension de leur espace représentatif.

La première approche permet de choisir les composantes les plus pertinentes et discriminantes du vecteur de caractéristiques par rapport à une fonction coût à optimiser. L'objectif final est d'affiner les performances du classifieur recevant ces caractéristiques et de limiter les temps de calcul et l'espace mémoire nécessaire. La seconde approche consiste à transcrire les caractéristiques initiales dans un espace de dimension inférieure.

La sélection des caractéristiques suit un problème d'optimisation dont la fonction coût est définie par l'utilisateur. Le choix de cette fonction repose sur deux logiques différentes :

- la première évalue les composantes du vecteur de caractéristiques par rapport à leur indépendance les unes aux autres (distance interclasse, corrélation) sans tenir compte des performances correspondantes obtenues par le classifieur. Cette méthode, notée *filter* en anglais, propose une exécution rapide sans apprentissage, mais a tendance à sélectionner de larges groupes. Le groupe optimal s'adapte ainsi à un grand nombre de classifieur. On peut citer l'utilisation de tests statistiques, scores de Fisher ou de Pearson.
- La seconde, appelée *wrapper* en anglais, évalue les caractéristiques par la précision de prédiction par rapport au classifieur ciblé. Celle-ci peut être extraite à partir de la probabilité d'erreur obtenue par validation croisée, décrite au paragraphe 2.5.4.2, par exemple. Cette méthode implique une exécution lente nécessitant un apprentissage par le classifieur ciblé. Elle peut souffrir d'un manque de généralité car le groupe optimal est

spécifique au classifieur testé, mais en est justement plus précise, car réglée par rapport aux interactions spécifiques entre les données et le classifieur.

La technique la plus utilisée de réduction de dimensionnalité est l'analyse en composantes principales (ACP). Elle permet de projeter l'ensemble des composantes du vecteur de caractéristiques dans un espace de dimension inférieure. Cette projection rend les nouvelles composantes, combinaisons linéaires des composantes originales, optimales mais plus forcément interprétables.

Nous allons, dans le paragraphe suivant, décrire deux méthodes de sélection de caractéristiques, les algorithmes génétiques (AG) et les algorithmes de recherche séquentielle, et la méthode la plus répandue de réduction de dimensionnalité qu'est l'ACP.

Les algorithmes génétiques (AG)

Les algorithmes génétiques (AG) proposent une recherche aléatoire des composantes les plus pertinentes du vecteur de caractéristiques. L'aspect aléatoire implique une recherche qui peut être lente mais particulièrement efficace, notamment lorsque les paramètres intrinsèques de l'algorithme sont correctement choisis.

Les AG [Sankar and Wang, 1996] se basent sur la théorie de l'évolution. Ils permettent de faire évoluer la solution du problème d'optimisation, c'est-à-dire le choix d'un vecteur de caractéristique optimal, par croisement d'individus (qui sont les composantes du vecteur de caractéristiques), mutation et sélection. Les deux premières opérations répondent à des critères probabilistes. L'ensemble de ces évolutions se fait de manière à optimiser une fonction de performances (précision, sensibilité, spécificité...). La Figure 2-11 illustre les principales étapes du fonctionnement d'un algorithme génétique.

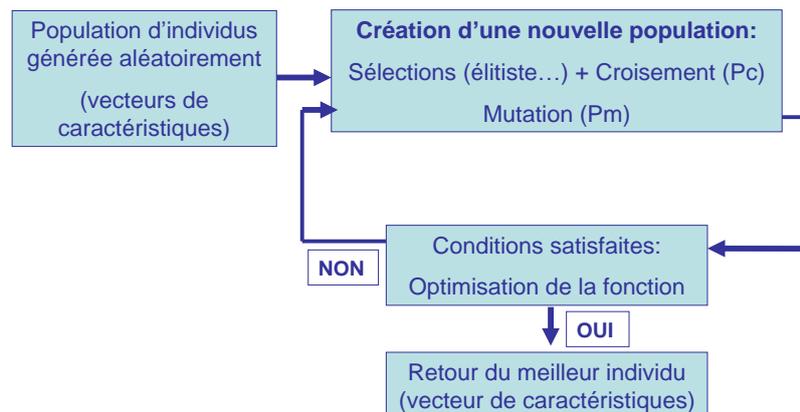


Figure 2-11 : Principales étapes du processus de sélection par algorithme génétique

Plusieurs exemples de la littérature valorisent l'utilisation des AG. [Boroczky et al., 2006], par exemple, sélectionne 10 caractéristiques sur les 23 initialement calculées pour réduire le nombre de FP de 56.4 % tout en gardant la même sensibilité pour la détection de tumeurs pulmonaires en imagerie TDM. Sa fonction de performances à optimiser se base sur une hiérarchie comprenant la sensibilité, la spécificité et le nombre de caractéristiques. [Mougiakakou et al., 2007] améliore la précision de son système CAD de 1 à 5 % grâce à la sélection de caractéristiques pour la détection de tumeurs du foie en imagerie TDM. [Zhao et al., 2007; Lu et al., 2009] montrent également un nombre de FP réduits au prix d'une sensibilité identique ou faiblement diminuée grâce à un AG en imagerie TDM pulmonaire.

Les algorithmes de recherche séquentiels

Ce paragraphe détaille brièvement les algorithmes dont la stratégie de recherche est séquentielle.

Ces algorithmes de recherche séquentiels [Devijver and Kittler, 1982] sont simples mais parfois très lents. La méthode générale consiste à ajouter ou enlever des caractéristiques au vecteur

descripteur jusqu'à obtention d'une fonction d'objectif optimale. La méthode *sequential forward selection*, en anglais, ou SFS, rajoute itérativement des composantes au vecteur tandis que la méthode *sequential backward selection* ou SBS les enlève. Des adaptations de ces méthodes les rendant plus robustes, autorisent les retours en arrière [Pudil et al., 1994] ou des recherches bidirectionnelles.

Quelques exemples de la littérature prouvent l'efficacité de ces méthodes. [Yu, 2000] réduit le nombre de caractéristiques de 31 à 15 par SFS et SBS en mammographie. Les caractéristiques statistiques et de forme les plus discriminantes concordent entre les deux algorithmes, même si leur ordre d'importance diffère. Les meilleures performances semblent cependant obtenues avec le SBS. L'étude de [Nandi et al., 2006], portant sur le développement d'un système CAD pour la mammographie, compare également la SBS, la SFS et des tests statistiques pour la sélection des caractéristiques les plus discriminantes. Sur les 22 caractéristiques initiales (contour, forme, texture...), les neuf premières composantes résultantes sont identiques et correspondent essentiellement aux caractéristiques de forme.

L'analyse en composantes principales (ACP)

L'ACP, introduite en 1901 par K. Pearson [Pearson, 1901] et développée par H. Hotelling en 1933, est aussi connue sous le nom de transformée de Karhunen-Loève ou de transformée de Hotelling. Comme pour le classifieur de Hotelling, le résultat d'une ACP permet de déterminer les axes qui expliquent de manière optimale la dispersion des caractéristiques disponibles.

Chaque donnée étant représentée dans un espace à p dimensions (caractéristiques), l'ensemble des données forme un nuage de l points dans \mathbb{R}^p . Le principe de l'ACP est d'obtenir une représentation approchée du nuage dans un sous-espace de dimension plus faible q par projection sur des axes bien choisis. Cette projection s'apparente à une combinaison linéaire optimale des caractéristiques. Une métrique dans \mathbb{R}^p est choisie de manière à calculer les distances des points projetés à leur centre de gravité. Les q axes principaux sont ceux qui maximisent l'inertie du nuage projeté, c'est-à-dire la moyenne pondérée des carrés des distances des points projetés à leur centre de gravité. Les composantes principales sont les l vecteurs ayant pour coordonnées celles des projections orthogonales des l éléments du nuage sur les q axes principaux.

L'ACP construit ainsi de nouvelles caractéristiques de dimension inférieure, artificielles, et des représentations graphiques permettant de visualiser les relations entre elles. Les proximités entre les composantes principales et les caractéristiques initiales sont mesurées par leurs corrélations.

Cette méthode, utilisée pour le développement de CAD en imagerie médicale [Papadopoulos et al., 2008], a l'avantage de ne pas être spécifique au classifieur utilisé. Elle implique l'extraction de l'ensemble des caractéristiques initiales à chaque étape du CAD (apprentissage/test). Le temps d'extraction initial des caractéristiques reste long étant donné que l'ensemble des composantes doit être calculé.

2.4.4.3 Ajout de nouvelles caractéristiques post classification

Cette méthode repose sur une étape de post-traitement utilisant les tumeurs candidates extraites de la première étape de classification. Il s'agit d'exploiter de nouvelles caractéristiques pour ces candidats afin d'extraire de nouvelles règles de discrimination pertinente entre les vraies tumeurs (VP) et les mauvaises détections éventuelles (FP). Ces caractéristiques sont généralement différentes de celles initialement envoyées au classifieur. Elles peuvent ensuite être envoyées à un autre classifieur ou analysées telles quelles.

[Masotti et al., 2009] extrait des caractéristiques de type texture à partir des images de ranklet pour les régions d'intérêt survivant son système CAD développé pour la mammographie. Un SVM traite ensuite ces caractéristiques et permet de réduire les fausses détections de 91% à 71% par mammographie en considérant une sensibilité fixée à 80%. Dans son CAD dédié aux tumeurs pulmonaires pour les radiographies thoraciques, l'étude de [Yoshida, 2004] base son RFP sur des mesures de corrélation entre une tumeur initialement détectée dans le poumon et son symétrique dans l'autre poumon défini par recalage. [Wu et al., 2006; Ikedo et al., 2007] appliquent la même méthode en extrayant des caractéristiques bilatérales de mammographies respectivement pour une réduction de FP d'un facteur 3 et 2.

2.4.4.4 La combinaison de classifieurs

Une autre méthode de RFP repose sur la combinaison de classifieurs pour laquelle [Kittler et al., 1998], bien qu'il n'en soit pas le concepteur, expose une architecture théorique commune en 1998. Cette méthode est fréquemment utilisée dans les domaines de l'aide à la détection (CAD) et au diagnostic (CADi) quelque soit l'application. Elle repose d'une part sur le fait qu'il n'existe pas de meilleur classifieur capable de traiter et donc d'apprendre n'importe quelle distribution des données d'apprentissage. D'autre part, le réglage d'un classifieur est un problème extrêmement difficile avec des choix intrinsèques initiaux importants qui peuvent totalement modifier les performances finales.

Ainsi, l'exploitation de leur complémentarité est une manière de créer un classifieur plus fort à partir de ceux existants. La combinaison de classifieurs est possible du moment qu'au moins un des paramètres suivants : l'espace des caractéristiques, la base d'apprentissage, la règle de décision ou le type des sorties, diffère entre les deux classifieurs. Lorsqu'aucun de ces paramètres n'est différent, il est aussi possible de les adapter à l'aide des méthodes détaillées dans la première colonne du Tableau 2-3. Ce tableau présente aussi les différentes catégories de méthodes de combinaison de classifieurs qui sont généralement utilisées dans la littérature.

Tableau 2-3 : Les différents stades du processus de classification et leur influence sur la combinaison de classifieur.

Données d'apprentissage	Caractéristiques	Classifieurs	Combinaison des types de sortie
<p>Bagging [Breiman, 1996] : Réplication des données d'apprentissage de manière à ce que chaque classifieur n'apprend pas exactement sur les mêmes données (bootstrap)</p> <p>Boosting [Freund and Schapire, 1996; Freund and Schapire, 1997] : Manipulation d'ensemble des données de manière à créer itérativement un ensemble de classifieur. La probabilité des données d'être utilisées par un nouveau classifieur pour son apprentissage dépend de leur classement à l'itération précédente.</p>	<p>Random Subset Method [Ho, 1998] : Sélection aléatoire d'un certain nombre de sous-espaces de l'espace de caractéristiques original et apprentissage d'un classifieur différent sur chaque sous-espace</p>	<p>Association séquentielle : Les classifieurs sont mis en série, l'un derrière l'autre et la sortie d'un classifieur peut servir d'entrée à un autre classifieur</p> <p>Association parallèle [Duin and Tax, 2000] : Les classifieurs sont utilisés de manière séparés et leurs sorties sont combinées</p> <p>Association hybride : Mêle les deux méthodologies de combinaison</p>	<p>Classe [Kimura and Shridhar, 1991] : Vote à la pluralité, majorité, unanime, notoire...</p> <p>Rang (pour le multiclasse) : Meilleur rang...</p> <p>Mesure [Tax et al., 2000] : Médiane, maximum, produit, minimum...</p>

L'étude de [Li et al., 2001] en mammographie prouve que la combinaison d'un classifieur complexe en cascade d'un classifieur simple, réalisant la sélection initiale des candidats, diminue le taux de FP de 56 % au prix d'une perte en sensibilité de 1 %. Cette méthode apparait d'ailleurs nettement plus efficace que la simple augmentation du seuil appliqué aux cartes de score de présence extraites du classifieur initial (plus le score est élevé, plus la probabilité de présence de tumeurs est grande). [Gurcan et al., 2001] utilise une combinaison séquentielle de classifieur afin de réduire de moitié le nombre de FP obtenu à partir d'images TDM thoraciques. L'étude de [Chan et al., 2007] combine quant à elle les scores obtenus de deux LDA, recevant respectivement des caractéristiques provenant de volumes 3D d'images de tomosynthèse et de projections 2D correspondantes. La diminution des FP est d'un facteur 1.5. Ainsi, la combinaison de classifieurs permet de notablement diminuer le taux des FP tout en conservant une bonne sensibilité.

Le travail de mise en œuvre et de paramétrisation d'un classifieur se retrouve cependant répliqué avec toutes les possibles optimisations et choix intrinsèques que cela suppose. Une mise en corrélation des scores obtenus par deux classifieurs de nature différente (un réseau de neurones et un

k-PPV) permet à [Park et al., 2009] de vérifier l'intérêt de leur combinaison parallèle de type score. La sensibilité du système dédié à la mammographie passe de 75,8 % (réseaux de neurone) et 65,9 % (k-PPV) à 80,3 % pour la combinaison.

2.4.4.5 Autres méthodes possibles

Un large panel de méthodes, autre que celles détaillées précédemment, permet de réduire le nombre de fausses détections. Certaines, originales, ne sont utilisées que dans quelques études, mais restent très intéressantes. On peut citer deux exemples types que sont l'élimination des candidats monovoxels [Yu and Guan, 2000] et le classement des tumeurs par leur score, reflétant leur probabilité d'appartenance à la classe tumeur [Jafar et al., 2006]. Ces méthodes tiennent compte des observations réalisées post-classification et peuvent dépendre de paramètres définis empiriquement. C'est par exemple le cas de la sélection d'un certain nombre de candidats potentiels par classement de leur score.

2.5 Evaluation des performances des systèmes CAD

2.5.1 Introduction

On a constaté au cours des paragraphes précédent qu'il est très difficile de comparer les performances des systèmes CAD pour une même modalité. Les études ne précisent en effet pas toujours toutes les règles d'évaluation utilisées. Trois types règles sont nécessaires à la correcte interprétation des performances des systèmes CAD.

La première consiste à définir le comptage des vrais positifs (VP) et faux positifs (FP) décrits au paragraphe 2.2.2. Une image entière ou une coupe peuvent être classées comme VP (étude de cas), mais ce comptage se fait parfois par nombre de candidats réels par image ou coupe.

Plus largement, il faut également fixer une stratégie d'évaluation des performances en sensibilité et spécificité, termes définis au paragraphe 2.2.2. Ainsi, la troisième règle consiste à définir les critères de performances. Un certain nombre d'outils basés sur la méthodologie ROC, initialement prédéfinis pour l'évaluation de la détectabilité dans des études d'observateurs humains, est généralement utilisé et adapté pour la problématique de la classification.

Enfin, la troisième règle est propre aux systèmes CAD supervisés et consiste à définir le partitionnement des images en base d'apprentissage et base de test. Les images d'apprentissage servent à estimer les paramètres optimaux du classifieur et les images de test permettant de calculer la proportion finale des erreurs de classification. Celui peut en effet influencer les performances de manière pessimiste ou optimiste.

Les paragraphes suivants présentent un état de l'art de l'ensemble de ces règles afin d'introduire la méthodologie choisie pour l'évaluation du système CAD 3D développé dans cette thèse.

2.5.2 Comptage des vrais positifs (VP) et faux positifs (FP)

Ce paragraphe développe les exemples de la littérature concernant le mode de comptage des VP et FP.

La sortie d'un système CAD est généralement une carte paramétrique (binaire ou pas) pour laquelle un voxel ou groupe de voxels reçoit une valeur propre caractéristique de sa classe d'appartenance (tumeur/non tumeur). Afin de déterminer si un voxel étiqueté 'tumeur' est bien un VP, on définit une distance d'acceptation entre la position trouvée et celle de la vérité terrain, en dessous de laquelle une tumeur suspecte est considérée comme correctement détectée. Dans la majorité des cas, cette distance est calculée entre les centres des deux positions. Par exemple, Paik [Paik et al., 2004] la définit comme la moitié du diamètre d'une vraie tumeur. En d'autres termes, pour être correctement détectée, le centre d'une tumeur suspecte doit être situé à l'intérieur d'une tumeur réelle.

La méthode de comptage des tumeurs suspectes étant réellement des tumeurs (VP) ou des faux positifs (FP) peut être appliquée par image (3D), par coupe (2D) ou par cas. Ce dernier type de comptage relève généralement plus d'une caractérisation (CADi) des tumeurs pré-détectées que d'une détection initiale de tumeurs (CAD).

C'est dans le cas du comptage des FP que les performances peuvent fortement différer selon un calcul par image (3D) ou par coupe (2D). Lorsque le comptage des FP est réalisé pour une image 3D, l'ensemble des voxels ou groupes de voxels candidats est considéré, ce qui produit généralement un taux de FP conséquent. C'est le cas des études de [Paik et al., 2004], avec une sensibilité de 100% et 165 faux positifs par image TDM et [Zhao et al., 2003], avec une sensibilité de 94,4% et 906 faux positifs par image TDM. Ces performances peuvent sembler assez mauvaises comparées aux autres méthodes d'évaluation (de l'ordre de quelques FP), mais elles ne peuvent pas vraiment leur être comparées. Lorsque le comptage est réalisé par coupe, le nombre réel de candidats FP visibles sur l'image n'est pas reflété. En effet, une coupe contenant un nombre N de fausses détections ($N \geq 1$) n'incrmente le compteur total des FP que d'une unité au lieu de N , comme visible sur l'image. Cette méthode, utilisée par [Ge et al., 2005; Jafar et al., 2006; Kanakatte et al., 2008], est généralement favorisée par les auteurs et donnent un nombre de fausses détections plus réduit. Elle ne peut pas être comparée avec la précédente.

2.5.3 Indices de performances

La sensibilité et la spécificité d'un système CAD mesurent son efficacité quantitative à partir des volumes binaires obtenus. Cependant, d'autres méthodes d'évaluation, basées sur les courbes de type ROC (*Receiver Operating Characteristics*) [Swets and Pickett, 1982; Metz, 1986], sont souvent utilisées pour enrichir les comparaisons inter-CAD ou fixer les différents paramètres des algorithmes.

La méthodologie ROC tient son origine de l'évaluation de la détectabilité humaine pour les études d'optimisation de la qualité des images. Elle fournit une mesure objective des performances d'un observateur humain dans une tâche de détection entre les deux classes saine et pathologique. Elle se base sur un modèle de décision représentée sur la Figure 2-12 par deux distributions de probabilité évoluant en fonction du critère de décision ou score, généralement dénotée λ . Ce modèle suppose que l'ensemble des valeurs de λ évalué sur des cas négatifs (sains) suit une distribution de probabilité $P(\lambda_0, \sigma_0)$ de valeur moyenne λ_0 et d'écart type σ_0 . De même, l'ensemble des valeurs λ évalué sur des cas positifs (pathologiques) suit une distribution de probabilité $P(\lambda_1, \sigma_1)$ de valeur moyenne λ_1 et d'écart type σ_1 . Généralement les valeurs de λ sont plus élevées en moyenne pour les cas positifs. Le mécanisme de décision humain ou numérique repose sur le choix d'un critère de décision seuil λ_s au-delà duquel toutes les observations sont considérées comme pathologiques.

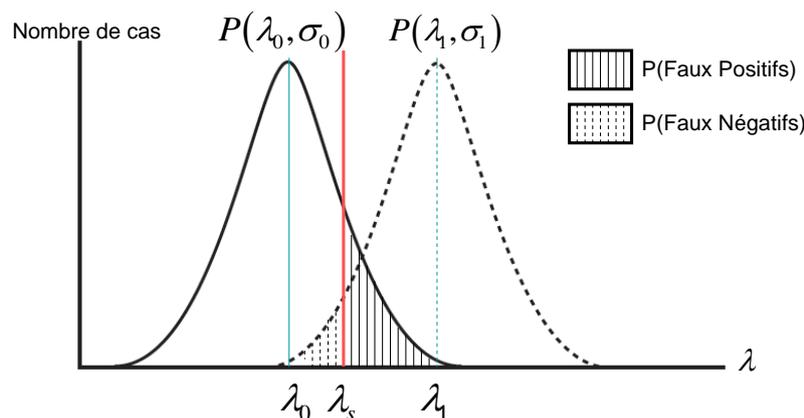


Figure 2-12 : Distributions de probabilité de la variable de décision dans le cas d'évaluation d'une tâche de détection entre deux classes : H_0 pour une observation négative (saine) et H_1 pour une observation positive (pathologique).

Le choix du seuil permet de définir les vrais et faux positifs selon la matrice de confusion (cf. Tableau 2-1). Une courbe ROC permet de caractériser un système humain ou numérique en évaluant le couple de paramètres (VP/VN) pour différentes valeurs du seuil de décision. En d'autres termes, si on considère un système de détection numérique, sa sortie de type score est seuillée pour différentes valeurs de seuil de décision. De la carte binaire obtenue est extrait un couple VP/VN (correspondant à la matrice de confusion définie au paragraphe 2.2.2) qui est reporté comme un point de la courbe

ROC. Une courbe ROC, dont un exemple est donné sur la Figure 2-13, permet d'observer les performances d'un système de détection humaine ou numérique en éliminant l'influence du critère de décision (seuillage) qui est utilisé pour obtenir les performances finales de type VP et FP.

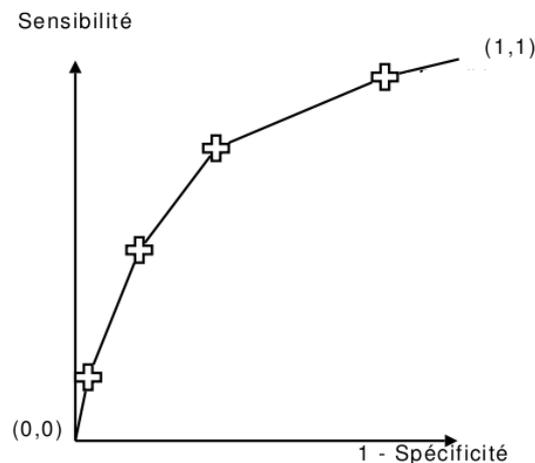


Figure 2-13 : Exemple de courbe ROC, chaque croix représente une valeur du critère de décision qui permet d'extraire un couple sensibilité/1-spécificité.

La courbe ROC optimale est proche de la marche d'escalier avec une sensibilité de 1 pour une spécificité de 0. Il faut noter que le tracé des courbes ROC s'applique à la détection d'une seule anomalie dans l'image sans aucune indication sur sa localisation. En d'autres termes, l'image visualisée ne peut être que soit négative soit positive dans son ensemble et la localisation de l'éventuelle tumeur n'est pas prise en compte. Ainsi une image peut être classée positivement à juste titre même si la localisation suspectée par l'observateur est fautive. En pratique, le clinicien recherche la présence d'une ou plusieurs tumeurs de taille et de contraste variable situées dans différentes régions de l'image. Cette tâche, beaucoup plus complexe, a impliqué l'émergence de nouvelles méthodologies d'évaluation plus proche de la réalité, toujours basées sur les études ROC. Le formalisme statistique de l'ensemble de ces méthodes est résumé dans l'article de Swensson [Swensson, 1996]. [Chakraborty and Yoon, 2008] reprend l'ensemble de ce formalisme de manière à adapter les performances obtenues par les systèmes CAD aux performances humaines. Les principales adaptations se basent sur l'incorporation de la notion de localisation des tumeurs dans l'image.

L'analyse LROC (*localization response operating characteristics*) mesure les performances du système diagnostique pour la détection et la localisation des tumeurs réelles de l'image. Dans ce cas là, l'image peut contenir plusieurs tumeurs et l'observateur doit indiquer la position de la tumeur la plus suspecte (1 par image). Celle-ci est comptée comme un vrai positif si la distance entre la position de la tumeur la plus suspecte et la position réelle de la tumeur ne dépasse pas une limite de tolérance fixée. La courbe LROC représente donc les variations des vrais positifs correctement localisés en fonction des faux positifs mesurés sur les images saines pour différentes valeurs du critère de décision λ_s .

L'analyse AFROC (*alternative free response operating characteristics*) est une extension de l'analyse LROC proposée par [Swensson, 1996] et [Chakraborty and Winter, 1990]. Dans ce cas, l'observateur ne reporte pas uniquement la position la plus suspecte, mais toutes celles qui dépassent la valeur du seuil de décision qu'il a fixé. Les courbes AFROC représentent les variations des vrais positifs en fonction de la fraction d'images contenant au moins un faux positif (comme les courbes ROC). L'hypothèse supplémentaire de ce modèle par rapport à l'analyse LROC concerne la stationnarité du critère de décision pendant l'interprétation d'une même image. Il faut en effet vérifier que l'observateur (humain) n'augmente pas son seuil de décision une fois qu'il a détecté une tumeur.

L'analyse FROC (*free response operating characteristics*) se rapproche de l'analyse AFROC. Le paramètre qui change est le comptage des faux positifs en abscisse, qui est calculé par image (2D ou 3D), générant généralement des valeurs assez élevées.

Cette dernière analyse semble la plus adaptée pour notre problématique d'évaluation d'un système CAD pour des images 3D corps entier. Elle permet en effet de tenir compte d'un nombre de tumeurs par image supérieur à un, tout en considérant l'ensemble des faux positifs pour une image 3D. Elle correspond ainsi à l'analyse qui se rapproche le plus de l'analyse visuelle qui peut être donnée de la sortie du classifieur.

2.5.4 Sélection des bases d'images d'apprentissage et de test

2.5.4.1 Problématique de la définition du « gold standard » ou vérité terrain

Les étapes de mise en place d'un CAD supervisé nécessitent généralement un grand nombre d'images annotées par le clinicien. Cette annotation correspond à la connaissance de la vérité terrain en termes de diagnostic et de caractéristiques de la ou des tumeurs réellement présentes dans l'image. Ces informations sont non seulement utiles pour la détermination des modèles de classification supervisée, mais également pour le calcul des performances quantitatives détaillées précédemment. La collection d'images cliniques annotées est un travail de longue haleine nécessitant le concours de différentes personnes, et principalement des cliniciens. Deux choix s'offrent aux utilisateurs :

- des images d'études rétrospectives à récupérer pour d'anciens malades, ce qui peut être très coûteux en temps pour le clinicien dont l'expertise est nécessaire notamment pour le ciblage des tumeurs,
- des images d'études actuelles sur lesquelles le clinicien fait son diagnostic en direct, sachant que celui-ci ne peut être réellement effectif qu'après une histologie ou postérieurement au traitement.

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

- la Digital Database For Screening DDSM [Heath et al., 2001],
- la Mammographic Image Analysis Society MIAS [Suckling et al., 1994]
- la Lung Image Database Consortium LIDC [Armato et al., 2004] pour la TDM.

Ces bases contiennent chacune un grand nombre d'images cliniques annotées et associées à des cas pathologiques de difficulté diagnostique croissante. Elles sont généralement le résultat d'un travail groupé de centres de recherche et cliniques.

Quelques bases de cas cliniques sont disponibles en imagerie TEP oncologiques. Celles-ci ne sont cependant pas distribuées à la communauté scientifique, sont rarement annotées et ne présentent pas forcément assez d'images de chaque classe. La littérature des CAD en imagerie TEP (paragraphe 2.2) confirme ce manque de bases d'images cliniques distribuées car les auteurs utilisent généralement quelques cas fournis par les hôpitaux partenaires de leur étude.

Une alternative à l'utilisation d'images cliniques repose sur la simulation d'examen cliniques. Les images simulées sont obtenues à partir d'un programme numérique ou simulateur, reproduisant virtuellement et le plus fidèlement possible la réalité clinique en termes d'acquisition, de bruit et de modélisation des tumeurs. La simulation permet de remédier aux limitations des images cliniques, notamment l'absence de vérité terrain sur la tumeur (captation, délimitation, contraste...) et de contrôler des paramètres bruitant l'image TEP. Il apparaît ainsi intéressant de profiter des avantages de la simulation numérique pour disposer d'un nombre suffisant d'images pour l'évaluation initiale des systèmes de classification. Lorsque la faisabilité du système est prouvée sur des images simulées, les performances du CAD peuvent ensuite être validées sur des images cliniques. Nous détaillons plus précisément les techniques de simulation dans le Chapitre 4.

2.5.4.2 Partitionnement des images disponibles

Une fois le type d'images défini, il convient de les séparer en différents groupes selon leur utilisation pour la mise en place et l'évaluation du CAD supervisé. L'ensemble des images doit contenir des cas pathologiques et des cas sains. Le nombre d'images disponibles et les caractéristiques des tumeurs (diamètres, positions, organes cibles...) doivent être précisés pour de futures

comparaisons de performances. Tous ces critères peuvent en effet parfois expliquer des différences de performances entre plusieurs systèmes.

Plusieurs méthodes de partitionnement des bases d'images pour l'estimation de l'erreur de classification sont données dans la littérature. On peut citer les méthodes de *validation croisée* et de *leave-one-out*, de *resubstitution*, de *bootstrap* [Efron and Tibshirani, 1997] et de *hold-out*. Des détails théoriques sur ces méthodes et leur comparaison peuvent être trouvés dans les articles de [Toussaint, 1974; Fukunaga and Hayes, 1989]. Dans le cas idéal, les images d'apprentissage et de test sont totalement différents comme pour la méthode par hold-out. Lorsque la base d'images disponible est trop petite, des méthodologies adaptées permettent de répliquer les images disponibles (bootstrap). Il est aussi possible de séparer les images en plusieurs échantillons de petite taille et de moyenner les erreurs obtenues sur toutes les combinaisons d'échantillons d'apprentissage et de test. Cette méthode correspond à la validation croisée. Lorsqu'un seul échantillon sert au test du classifieur et tous les autres à l'apprentissage, la méthode est appelée leave-one-out. La méthode de resubstitution, quant à elle, utilise toutes les images disponibles pour l'apprentissage et teste le modèle du classifieur sur chacune de ces mêmes images. La Figure 2-14 résume l'ensemble de ces méthodes.

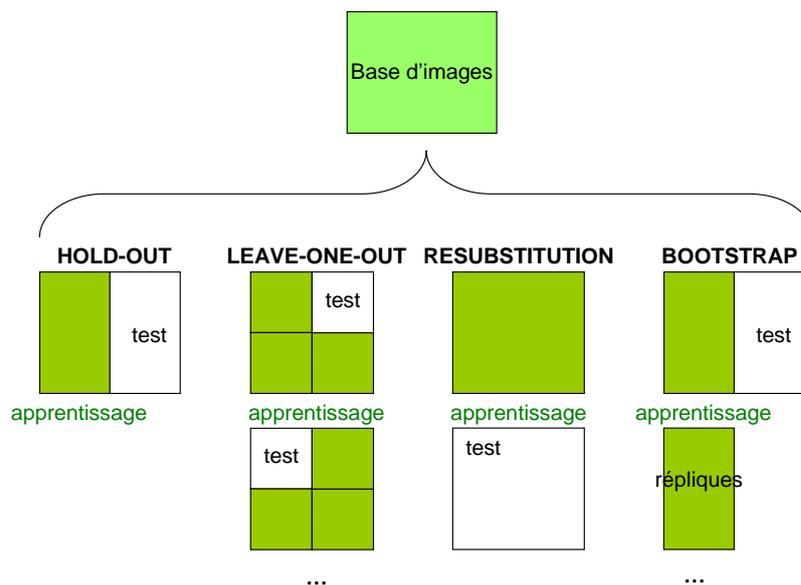


Figure 2-14 : Présentation schématique des différentes méthodes d'évaluation de l'erreur de mauvaises classifications concernant l'utilisation et le partitionnement des images disponibles.

Les méthodes de hold-out et de leave-one-out sont considérées comme non-biaisées, même si la méthode de hold-out aurait plutôt tendance à surestimer l'erreur de classification. Dans le cas de la méthode de leave-one-out ou validation croisée, la répétition de la procédure pour les différents échantillons amène à une erreur moyenne estimée proche de l'erreur théorique du classifieur pour l'ensemble des échantillons disponibles. La méthode de resubstitution a, quant à elle, tendance à sous-estimer parfois fortement l'erreur de classification et a fournir une variance sur la mesure de l'erreur plus grande [Sahiner et al., 2000].

2.5.5 Conclusions

En conclusion, la méthodologie d'évaluation des systèmes CAD comporte un grand nombre de paramètres à préciser. La comparaison directe des systèmes CAD dédiés à la même application reste cependant fragile tant la moindre différence concernant les aspects précédemment cités, la taille de base d'images, son partitionnement et le comptage des vrais et faux positifs, peut influencer les performances finales.

En ce qui concerne la méthodologie d'évaluation 3D et plus précisément corps entier, elle doit encore être précisée et affinée. Elle ne peut ainsi pas facilement être comparée aux performances d'un même système sur des images 2D, notamment pour le nombre de fausses détections.

2.6 Conclusions de l'état de l'art

Nous avons proposé dans ce chapitre un état de l'art des systèmes CAD existants en imagerie médicale et plus précisément en imagerie TEP. Les différentes étapes standardisées du CAD supervisé qui est développé dans ce manuscrit font l'objet d'un cahier des charges. Celui-ci est détaillé au chapitre suivant.

Chapitre 3 Choix méthodologiques pour le système CAD proposé

3.1 Introduction

Le Chapitre 1 a mis en évidence le souci pour les cliniciens d'optimiser la prise en charge du patient cancéreux. Pour cela, la détection précoce initiale ou au cours des traitements doit être efficace et précise. Les différents types d'imagerie permettent de faciliter la prise de décision du médecin et en particulier l'imagerie TEP au 18F-FDG. Cependant la détection de tumeurs peut parfois paraître difficile, notamment dans le cas de tumeurs faiblement visibles comme observées après une chimiothérapie ou une radiothérapie. Ce chapitre précise les motivations de ce travail de thèse et les choix méthodologiques réalisés à partir des constatations de l'état de l'art.

3.2 Motivations

L'intérêt des systèmes CAD a été prouvé dans d'autres modalités d'imagerie médicale (mammographie, TDM) [Freer and Ulissey, 2001; Gilbert et al., 2006; Doi, 2007; Marten and Engelke, 2007]. Ils permettent en effet d'augmenter les performances en termes de détection et de stadification en diminuant les variations inter- et intra-cliniciens.

L'état de l'art, présenté au Chapitre 2, a montré un développement encore très réduit des systèmes CAD en imagerie TEP oncologique 3D. L'ensemble de ces études sont cependant très récentes. Elles suscitent d'ailleurs l'intérêt des principaux constructeurs comme Siemens qui est associé à des publications académiques [Guan et al., 2006] et a déposé un brevet sur sa méthode de détection automatique en imagerie TEP [Huang et al., 2007]. Un autre brevet a aussi été déposé sur la méthodologie standardisée du développement d'un tel système dédié à des tumeurs 3D [Yoshida et al., 2002]. Il n'existe cependant à ce jour aucun CAD dédié à l'imagerie TEP oncologique implémenté sur les consoles des caméras cliniques. Les développements futurs pourraient néanmoins remédier à ce manque et améliorer l'utilisation de l'imagerie TEP 3D corps entier dans la prise en charge du cancer.

Des discussions fructueuses avec deux médecins nucléaires du Centre Léon Bérard et du Centre Hospitalier Lyon Sud, le Dr Thomas Mognetti et le Pr. Francesco Giammarile, ont d'ailleurs confirmé leur intérêt au niveau clinique d'un tel système. Un système CAD peut s'avérer intéressant dans le cas de cancers disséminés et pouvant présenter des fixations résiduelles critiques au cours du traitement. Ceci est notamment vrai si on considère le nombre conséquent d'images observées chaque jour et certains types de pathologies problématiques. L'application clinique principale a d'ailleurs été mise en place grâce à cette collaboration. Le lymphome est ainsi apparu comme la pathologie la plus susceptible de bénéficier d'un tel système. En effet, ce type de cancer est caractérisé par des tumeurs de taille réduite ou diffuse et de contraste parfois réduit. Il présente une bonne réponse aux traitements conventionnels (chimiothérapie, radiothérapie). Le lymphome touche en majorité des patients très jeunes pour lesquels l'espérance de vie après rémission est par conséquent très élevée. Le Dr Thomas Mognetti et le Pr. Francesco Giammarile ont ainsi souligné l'intérêt de déterminer le plus tôt possible une réponse totale ou partielle afin d'éviter des complications éventuelles à long terme pour ces patients (cf. paragraphe 1.4).

Comme détaillé au chapitre précédent, un système d'aide à la détection, afin d'orienter le clinicien, doit lui fournir un score ou une étiquette (binaire) de présence de tumeurs en chaque endroit, ou voxel de l'image étudiée. Un CAD repose sur un ensemble de choix méthodologiques standardisés que nous définissons et justifions ci-après. Ces choix sont ensuite précisés et détaillés au Chapitre 5. Remarquons que nous ne considérons que l'information fonctionnelle fournie par la TEP corps entier dans le cadre de cette thèse.

3.3 Choix méthodologiques

3.3.1 Les images TEP utilisées

3.3.1.1 Pré-traitements des données

Les images TEP 3D corps entier ciblées dans cette thèse sont utilisées sans pré-traitement préalable. Les seules corrections exploitées sont celles généralement appliquées dans les protocoles cliniques (atténuation, diffusion, temps mort...).

Il aurait pu être intéressant de tester l'influence de certaines méthodes de traitement d'images TEP qui font l'objet de développements importants dans la communauté scientifiques et suscitent des débats. On peut citer par exemple les méthodes de débruitage [Turkheimer et al., 2008] ou de correction de volume partiel [Soret et al., 2007; Le Pogam et al., 2008; Rousset et al., 2008].

Les méthodes de débruitage, en lissant l'image, peuvent faire disparaître les petites tumeurs de faible contraste et donc diminuer la sensibilité. Cependant, la diminution de la texture granuleuse due au bruit Gaussien, permettrait aussi de réduire le nombre de fausses détections. La correction d'effet de volume partiel a, quant à elle l'avantage de rehausser les vraies tumeurs, mais également le bruit. Un réel compromis est alors à trouver entre une bonne sensibilité et un nombre de fausses détections réduit. Il faut également noter que les corrections de volume partiel n'ont jusqu'à présent jamais été appliquées sur une image entière 3D, mais de manière localisée sur une tumeur ou un organe. Aucune étude n'a d'ailleurs déjà prouvé l'intérêt d'utiliser des telles méthodes de correction d'effet de volume partiel sur une imagerie TEP corps entier en termes de détectabilité humaine.

Cependant ces méthodes ne sont pas appliquées en routine clinique, ni implantées sur les consoles cliniques, ou encore en phase de validation. Notre objectif est de démontrer l'intérêt du CAD sur des images les plus représentatives de la réalité clinique actuelle. L'outil développé dans cette thèse pourra ensuite être utilisé pour évaluer d'autres images (débruitées et corrigées). C'est d'ailleurs l'objectif d'une étude préliminaire initiée récemment en collaboration avec le LATIM, Brest, afin d'évaluer l'impact des méthodes de correction d'effet de volume partiel qu'ils développent sur les performances de détectabilité mesurées par le système CAD [Lartzien et al., 2009].

Les premiers choix méthodologiques de cette thèse se sont donc focalisés sur une image TEP 3D corps entier telle qu'elle est visualisée après un examen clinique actuelle sans aucun traitement supplémentaire.

3.3.1.2 Données cliniques ou simulées

Le paragraphe 2.5.4.1 a introduit l'intérêt des images simulées pour l'étude de faisabilité d'un système CAD dédié à l'imagerie TEP 3D. Celles-ci ont ainsi l'avantage de fournir la vérité terrain sur les caractéristiques (captation, forme...) et les localisations des tumeurs et de contrôler l'ensemble des effets dégradant l'image TEP. L'obtention d'un large nombre d'images simulées TEP oncologiques est généralement moins complexe que la collection d'images cliniques annotées, même si quelques cas sont déjà distribués en ligne par les centres hospitaliers. Nous avons donc profité de la disposition d'outils de simulation performants au laboratoire pour faciliter le développement et l'évaluation du système présenté dans ce manuscrit. Un nombre suffisant d'images TEP réalistes saines et pathologiques a pour ce faire été simulé et utilisé. Les tumeurs introduites dans cette base d'images simulées miment les cas problématiques du lymphome et leurs propriétés de répartition dans le corps humain et la population. Le développement de ces images simulées est présenté au Chapitre 4.

3.3.2 Choix des caractéristiques

Le choix des caractéristiques décrivant les images ou les groupes de voxels à classer est d'une importance capitale. Un ensemble de familles de caractéristiques descriptives a été donné au paragraphe 2.4.3.2. Les caractéristiques doivent être les plus représentatives et discriminantes possible.

Nous avons choisi dans un premier temps de ne pas conserver de caractéristiques descriptives de forme. D'une part, l'imagerie TEP est, comme nous l'avons déjà soulignée, une imagerie fonctionnelle peu riche en information anatomique. D'autre part, nous plaçons notre problématique de détection dans le cas de tumeurs de faible diamètre assimilables à des petites sphères. Leur description

anatomique apparaît donc très limitée. Aussi, le principal inconvénient des caractéristiques de forme est d'être fortement corrélé à la segmentation initiale des tumeurs étudiées qui peut parfois être approximative dans le cas de tumeurs de petite taille. Cette famille de caractéristiques est ainsi difficilement exploitable dans le cas de l'utilisation exclusive de l'imagerie fonctionnelle.

Nous n'avons pas non plus tenu compte des informations sémantiques des patients. D'une part, celles-ci ne sont pas disponibles lors d'une étude de faisabilité sur des images simulées. D'autre part, ces caractéristiques peuvent apparaître difficilement traduisibles pour des classifieurs nécessitant une information de type numérique en entrée (SVM et LDA par exemple).

Les caractéristiques statistiques des différents ordres des tumeurs cibles (faible diamètre et contraste) risquent de ne pas être suffisamment discriminantes. Les images TEP présentent en effet une texture parfois irrégulière ou une différenciation entre tumeurs et fond incertaine en considérant le fort bruit Gaussien les affectant.

Nous avons donc considéré que les caractéristiques fréquentielles et plus précisément celles extraites d'une décomposition en ondelettes, apparaissent adaptées à notre imagerie 3D. Leur positionnement conjoint dans les domaines spatial et fréquentiel et l'aspect multirésolution de cette transformation est d'un attrait particulier dans la détection d'objets Gaussiens de taille réduite (tumeurs) noyés dans du bruit.

3.3.3 Choix du classifieur

Comme précisé précédemment, notre choix s'est porté sur la classification supervisée.

Le paragraphe 2.4.3.3 a détaillé les différents classifieurs supervisés utilisés dans la littérature. Il est assez difficile de sélectionner le classifieur idéal tant leurs performances varient d'une problématique à l'autre et selon les caractéristiques descriptives reçues. Dans le cadre de cette thèse, nous nous sommes focalisés sur l'utilisation et la comparaison de deux classifieurs de fondements différents, l'analyse discriminante linéaire ou *linear discriminant analysis* en anglais (LDA) et les séparateurs à vaste marge (SVM).

Le LDA, malgré son aspect linéaire et sa simplicité, n'a jamais été utilisé pour des tâches de détection automatique en imagerie TEP 3D corps entier. Il est cependant souvent approché par un observateur linéaire connu pour l'évaluation de la détectabilité pour l'évaluation de la qualité des images. Le Channelized Hotelling Observer (CHO) [Barrett et al., 1993] s'est d'ailleurs montré fortement corrélé aux performances de détection humaine de signaux connus dans un fond connu en imagerie TEP [Lartizien et al., 2004]. Il semblait donc intéressant de poursuivre les travaux de C. Lartizien en utilisant le CHO comme méthode de classification des voxels des images TEP 3D corps entier.

Les résultats obtenus avec ce système CAD (Chapitre 5) ont montré que les performances de détection du LDA seul sont assez limitées, notamment à cause d'un nombre trop élevée de fausses détections (FP). Nous nous sommes donc penchés sur l'utilisation d'un autre classifieur supervisé en vue d'une combinaison finale de leurs performances comme suggéré par [Park et al., 2009] et détaillé au paragraphe 2.4.4.4. Les SVM reposent sur une théorie robuste et sont ajustés au travers différents paramètres totalement contrôlables. Leurs adaptations aux bases d'apprentissage de petite taille en font un classifieur attractif pour notre problématique. Les bonnes performances de détection obtenues lors des différentes études présentées au paragraphe 2.4.3.3 nous ont orienté à évaluer les SVM comme deuxième classifieur.

Un grand nombre de classifieurs supervisés a été présenté dans le paragraphe 2.4.3.3. Notre choix s'est focalisé sur les deux d'entre eux qui semblaient être les plus adaptés à notre problématique par rapport à l'état de l'art. Il apparaît cependant difficile de savoir si l'optimisation des deux classifieurs séparément aboutira à des performances optimales.

Chapitre 4 Simulation d'une base d'images TEP pour l'oncologie : oncoPET_DB

4.1 Introduction

La simulation d'acquisitions TEP réalistes apporte une alternative à l'utilisation d'images cliniques. La collection d'images cliniques étiquetées est en effet un travail de longue haleine nécessitant le concours des cliniciens. En classification supervisée, la vérité terrain doit être précisément définie, notamment en termes de localisation et de forme des tumeurs. Des informations portant sur la captation de la tumeur et donc son contraste avec le fond apparaissent également utiles pour l'évaluation des systèmes potentiellement développés. L'ensemble de ces informations est intrinsèquement connue par simulation. Comme leur nom l'indique, les images simulées correspondent à des examens virtuels d'objets ou de patients eux aussi virtuels (fantômes). Les simulateurs modélisent la géométrie d'une caméra donnée favorisant ainsi l'accès aux facteurs de dégradation de l'image TEP de manière à permettre des études de performances quantitatives de chaque algorithme avant un éventuel transfert sur des machines et consoles cliniques. La simulation permet également de mettre à disposition un nombre important d'images nécessaires à l'apprentissage et à la validation d'un système CAD.

L'utilisation de la simulation dans les différents domaines d'imagerie médicale au cours des dernières années prouvent son intérêt. Pour citer quelques exemples en instrumentation, les simulations ont été utilisées en Imagerie par Résonance Magnétique (IRM) afin d'évaluer les distributions de taux d'absorptions spécifiques [Wang et al., 2007] propres à un système d'acquisition ; en imagerie ultrasonore (US) pour optimiser la conception de sondes [Mckeighen, 1996] et en imagerie nucléaire pour la conception de nouveaux scanners [Lartizien et al., 2003] ou l'évaluation de la dose radioactive délivrée aux patients durant la radiothérapie [Badel et al., 2007]. En ce qui concerne les méthodes de traitement d'images, les simulations ont été utilisées pour tester des méthodes de segmentation de microcalcifications en mammographie [Morrison and Linnett, 1999] ; pour l'évaluation de méthodes de segmentation de contours en US [Brusseau et al., 2004] ; pour valider des méthodes de recalage d'images TEMP et IRM cérébrales [Aubert-Broche et al., 2006]. En imagerie nucléaire, la simulation a notamment servi à l'évaluation de méthodes de correction d'effet de volume partiel (EVP) [Reilhac et al., 2004; Boussion et al., 2006] ou de mouvement du patient [Perruchot et al., 2004].

Il apparaît donc évident qu'une base d'images TEP simulées, contenant des cas sains et pathologiques étiquetés, serait d'une grande utilité pour réaliser l'étude de faisabilité du système CAD proposé dans ce manuscrit. Au commencement de la thèse, aucune base d'images ce type pour l'oncologie n'était disponible. Les paragraphes suivants décrivent brièvement le principe de la simulation Monte Carlo et son utilisation pour la réalisation d'examens TEP virtuels réalistes. Nous revenons ensuite sur les bases d'images TEP non oncologiques disponibles dans la littérature. Enfin, nous détaillons la méthodologie suivie pour la génération de la base d'images TEP simulées dédiée à l'oncologie.

Afin de simplifier son appellation et sa distribution, cette base d'images TEP simulées oncologiques a été dénommée 'oncoPET_DB'. Certes l'originalité du mot est limitée, mais elle permet une compréhension directe par l'utilisateur des images qu'il serait amené à utiliser: **oncologiePET_DataBase**.

4.2 La simulation TEP Monte Carlo

On appelle méthode de Monte-Carlo (MC) [Metropolis, 1947; Metropolis and Stanislas Ulam, 1949] toute méthode visant à calculer une valeur numérique, et utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. Le nom de ces méthodes fait allusion aux jeux de hasard pratiqués à Monte-Carlo. L'imagerie TEP est particulièrement propice à l'utilisation de la simulation MC par la nature probabiliste des processus physiques qui la compose. En effet, la simulation MC permet de

modéliser l'émission et le parcours du positon, l'annihilation et les interactions photon-matière jusqu'à la probabilité de détection des photons dans le détecteur. De plus amples détails sur les processus physiques permettant la formation de l'image TEP peuvent être trouvés en Annexe.

Plusieurs simulateurs TEP MC rigoureusement validés sont actuellement disponibles. On peut par exemple citer SimSET [Harrison et al., 1993], EIDOLON [Zaidi and Morel, 1999], PET-EGS [Castiglioni et al., 1999], GATE [Jan et al., 2004] et PET-SORTEO [Reilhac et al., 2005].

Le simulateur GATE a été développé dans le cadre de la collaboration OpenGate et se base sur la librairie générique pour la simulation des particules, GEANT4 [Agostinelli et al., 2003]. Il est puissant et flexible et le processus de simulations des données inclut la modélisation de tous les phénomènes majeurs et permet de modéliser fidèlement la plupart des géométries existantes. Son principal inconvénient reste le temps de calcul nécessaire pour faire une simulation, bien qu'à la date de rédaction de la thèse, de nombreux efforts aient été faits pour réduire ces temps exhaustifs (notamment dans le cadre de l'ANR fGATE).

Le développement de PET-SORTEO a été initié avant celui de GATE dans le cadre de la thèse de [Reilhac, 2007] et bénéficie de temps de simulation plus réduits. PET-SORTEO a été validé pour différents micro-scanners [Lartizien et al., 2005] et pour la caméra TEP Siemens ECAT/EXACT HR+ (CTI Siemens) [Reilhac et al., 2005].

Une comparaison des codes de simulation en termes de réalisme, de temps de calcul et de flexibilité a été réalisée par [Buvat et al., 2005] et dans la thèse de [Reilhac, 2007]. Notre choix s'est porté sur l'utilisation du simulateur PET-SORTEO. La raison principale de ce choix repose sur les temps de simulation affichés par PET-SORTEO et GATE au commencement de la thèse. Fin 2006, seul PET-SORTEO pouvait nous permettre de simuler des images TEP corps entier comme celles ciblées par notre problématique en un temps raisonnable.

4.3 Les bases d'images simulées TEP existantes

Au démarrage de la thèse, il n'y avait que deux bases d'images simulées TEP accessibles à la communauté scientifique.

La première développée par [Reilhac et al., 2005] contient des images TEP cérébrales. Les fantômes utilisés proviennent de l'utilisation et de l'enrichissement de deux modèles cérébraux labellisés existants, le Zubal [Zubal et al., 1994] et le Jacob [Kabani et al., 1998] par des images cliniques IRM de patients différents incluant ainsi une variabilité anatomique inter-modèle. Des images TEP réelles ont été utilisées pour définir les distributions dynamiques de radiotracer dans le cerveau pour du 18F-FDG, du 18F-Dopa, du 11C-Raclopride et de 18F-MPPF. Les images disponibles, présentées sur la Figure 4-1, ont été validées de manière statistique en comparaison à des images cliniques réelles. Cette base d'images simulées est notamment utilisée pour la validation de méthodes de segmentation [Saad et al., 2008] ou de correction d'effet de volume partiel [Tokha and Reilhac, 2007].

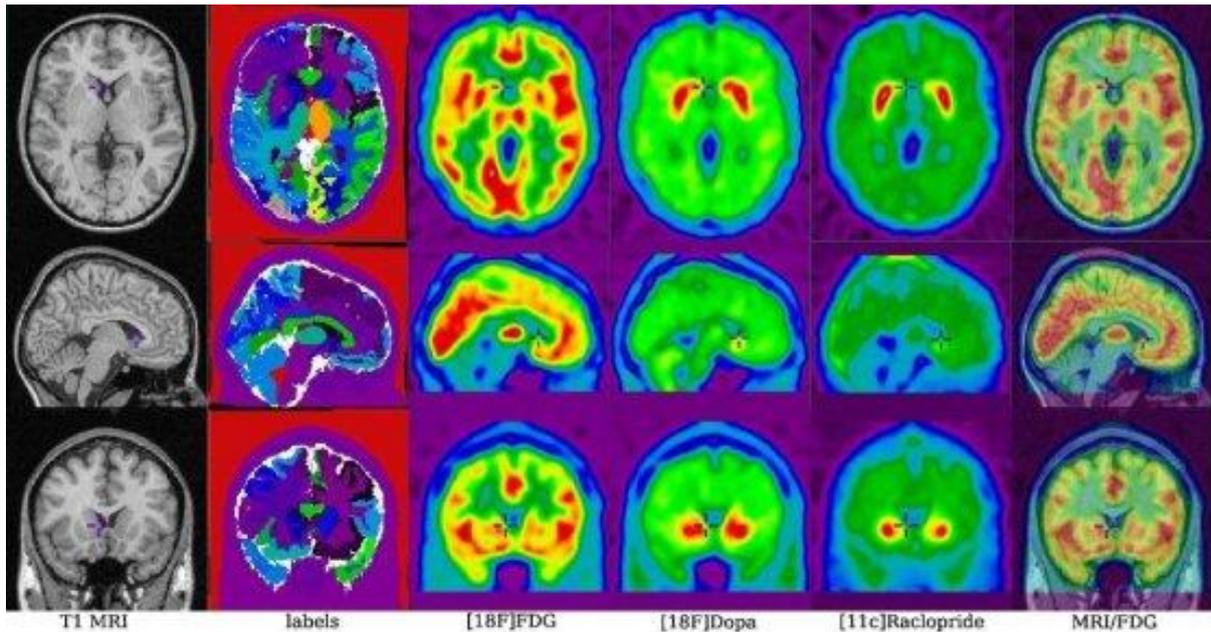


Figure 4-1 : De gauche à droite : coupes d'un modèles anatomique, du fantôme labellisé, d'images simulées au 18F-FDG, au 18F-Dopa, au 11C-Raclopride, et image recalée IRM clinique/TEP simulée.

La seconde base d'images développée par [Castiglioni et al., 2005] fournit des images TEP simulées à partir de fantômes, de simulateurs et de radiotraceurs différents. Les images simulées correspondent à des acquisitions de fantômes géométriques, corps entier et cerveaux. Les géométries de scanner modélisées sont également variées (GE Advance, Philips Prism 3000...). Cette base d'images contient un large nombre d'images permettant de tester et d'évaluer un grand nombre d'algorithmes de traitement d'images.

Ces deux bases d'images TEP simulées ne sont malheureusement pas compatibles avec la validation d'un algorithme d'aide à la détection ciblée sur l'oncologie. En effet, la validation d'un tel algorithme nécessite un nombre assez conséquent d'images TEP corps entier saines et pathologiques dont les caractéristiques des tumeurs sont connues et exploitables. Ces images doivent correspondre à celles exploitées par les cliniciens dans la réalité et donc être essentiellement obtenue à partir du traceur 18F-FDG. Afin de modéliser la grande variété de cas de détection de tumeurs résiduelles possibles, les caractéristiques des tumeurs cibles doivent être adaptées.

La frilosité des laboratoires au développement de base d'images TEP simulées s'explique en partie par le temps nécessaire à sa génération et à la prise main de simulateurs parfois complexes. L'opération de simulation d'une acquisition TEP repose en effet sur des logiciels de complexité avérée et nécessite une grande puissance de calcul. Le principe d'un simulateur Monte Carlo implique la définition de nombreux paramètres comme l'illustre la Figure 4-2:

- un protocole d'acquisition (durée d'acquisition réaliste)
- le choix ou la création d'un fantôme réaliste du point de vue anatomique
- la définition de cartes d'activités en radiotraceur reflétant la réalité clinique

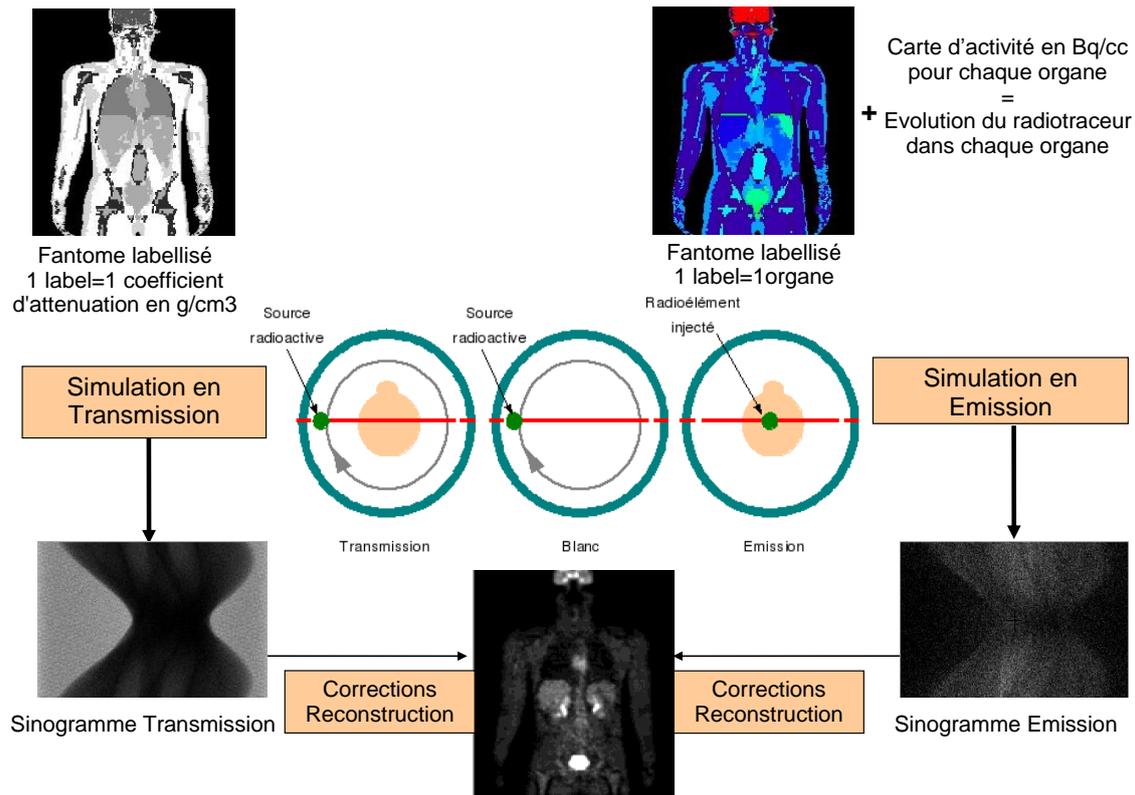


Figure 4-2 : Etapes du processus de simulation d'une acquisition TEP réaliste

4.4 Conclusions

La simulation MC est un outil adapté à la génération d'exams TEP virtuels et plusieurs simulateurs MC sont actuellement validés et mis à la disposition de la communauté scientifique. Cependant aucune base d'images TEP simulées dédiée à l'oncologie n'était disponible au commencement de la thèse fin 2006. Or l'utilisation d'une telle base d'images, garantit à la fois la vérité terrain et un nombre d'exemples sains et pathologiques suffisant pour l'apprentissage des classifieurs supervisés. Elle permet ainsi une mise en place et une validation initiale d'un système CAD simple et robuste. Nous avons donc choisi de développer une base d'images TEP simulées dédiée à l'oncologie, utilisable non seulement pour la validation du système CAD proposé dans cette thèse, mais aussi par toute la communauté scientifique pour l'évaluation de différentes méthodes de traitement destinées à améliorer les performances de détection.

4.5 Génération de la base d'images simulées : description du modèle de distribution de ¹⁸F-FDG

4.5.1 Méthode générale de simulation d'images TEP

Comme décrit sur la Figure 4-2, la simulation d'images TEP repose sur plusieurs étapes permettant de reproduire le plus fidèlement possible les différents phénomènes physiques impliqués dans l'acquisition réelle d'images TEP. L'étape principale est la génération d'images provenant d'acquisitions en émission simulées à partir d'un modèle anatomique virtuel. Ce modèle anatomique représente le patient virtuel subissant l'examen TEP virtuel. Le modèle anatomique choisi dans cette étude est le fantôme de Zubal [Zubal et al., 1994] qui est présenté dans la partie 4.5.2. Il correspond à la description d'un corps humain voxelisé où chaque voxel est associé au label de son organe d'appartenance. A chacun de ces labels doit être assigné une valeur de concentration de radiotraceur exprimé en kBq/mL et une valeur d'atténuation du tissu lui correspondant. A l'aide de ces trois informations (le fantôme, l'activité du radiotraceur et la valeur de l'atténuation), le simulateur Monte Carlo peut reproduire les différents phénomènes d'une acquisition TEP réelle, comme les processus statistiques de désintégrations radioactives, d'interactions entre particules et de détection des photons.

Le paragraphe 4.5.3 présente les valeurs de concentration en 18F-FDG dérivées pour les principaux organes d'intérêt à partir de l'analyse de 70 images TEP cliniques au 18F-FDG. Des tumeurs de caractéristiques définies de manière à reproduire l'extension du lymphome sont ensuite ajoutées au modèle de patient virtuel initial pour produire une base de modèles pathologiques. Cette description est présentée au paragraphe 4.5.4.

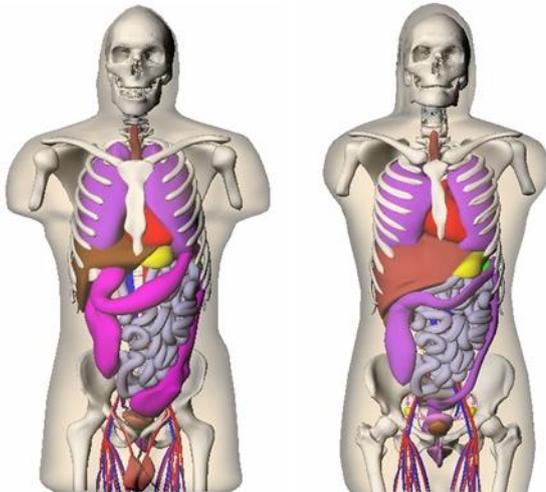
4.5.2 Choix du fantôme numérique

Un fantôme de corps humain est une description géométrique virtuelle de la morphologie de l'organisme qui peut avoir plusieurs niveaux de détails. Deux types de représentations existent.

La première présente les différentes structures de l'organisme comme un ensemble d'objets de forme géométrique de complexité variable, des ellipses ou des B-splines généralisées (NURBS). Le fantôme NCAT [Segars, 2001] et son extension XCAT [Segars et al., 2009] peuvent être cités en exemple. Les organes de ce fantôme ont été construits à partir de surfaces B-splines associés aux données TDM 3D du *Visible Human Project* [Banvard, 2002]. Ainsi ces surfaces peuvent être facilement modifiées de manière à modéliser les variations anatomiques et les mouvements cardiaques et respiratoires du patient. L'avantage de ces fantômes repose essentiellement sur une résolution plus fine et la facilité de décrire une grande collection de variabilités physiologiques (organe de taille et poids différents) en ajustant les paramètres de ces formes géométriques. Cependant ces fantômes sont en général de ressemblance anatomique plus approximative.

La seconde catégorie de représentations concerne les fantômes numériques provenant de segmentation d'examen radiologiques réels (TDM ou IRM) comme le Zubal [Zubal et al., 1994]. Un label unique est assigné à chaque organe. L'avantage principal de cette représentation repose sur une vraisemblance anatomique plus élevée, à défaut d'une résolution spatiale fixée par le pas d'échantillonnage du volume numérisé. Il est également impossible de décrire une variabilité physiologique du modèle ainsi figé. La Figure 4-3 illustre un rendu du fantôme NCAT a) et du fantôme Zubal b)

a)



b)

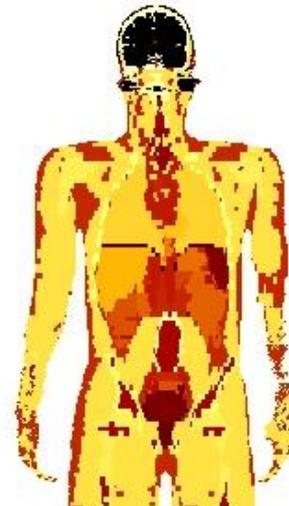


Figure 4-3 : Rendu volumique du fantôme NCAT a). Coupe coronale du fantôme Zubal b)

Notre choix s'est porté sur un fantôme de ce dernier type, le fantôme de Zubal.

Le fantôme de Zubal [Zubal et al., 1994] provient de la segmentation d'un examen TDM torse et tête et de coupes IRM cérébrales de deux patients masculins et de morphologie moyenne (175 cm et 70,3 kg). Il comporte 63 structures cérébrales et extra-cérébrales et 62 organes pour le corps entier. Comme expliqué au Chapitre 3, le modèle de cancer suivi pour fixer les caractéristiques des tumeurs est le lymphome. Cette maladie touche à 80 % les ganglions lymphatiques qui suivent les principales artères du corps humain. Or aucun fantôme de la littérature ne présente de description détaillée des chaînes ganglionnaires. En effet, le nombre des ganglions lymphatiques est d'environ 700 avec un diamètre approximatif de 6 mm, ce qui rend leur description anatomique difficilement réalisable. Cependant, le fantôme Zubal présente une description assez précise des principales artères qui servent

généralement de localisation de référence pour les ganglions lymphatiques. Notre attention s'est portée sur un total de 30 structures parmi toutes celles disponibles dans le Zubal. Le fantôme utilisé correspond à une matrice 3D constituée de 192 x 96 x 243 voxels de dimension 3,57 mm dans les trois directions. Ainsi le champ de vue axial couvert par le fantôme est d'environ 86 cm de la tête aux mi-cuisses. Ce champ de vue correspond aux examens TEP les plus représentatifs réalisés sur des patients atteints de lymphome.

4.5.3 Génération d'un modèle d'activité de 18F-FDG

Comme expliqué au paragraphe 4.1, la simulation Monte Carlo d'images TEP repose sur la création d'un modèle d'activité du radiotracer. En d'autres termes, une valeur d'activité de 18F-FDG en kBq/mL doit être assignée à chaque organe, c'est-à-dire chaque label du fantôme numérique. Cette valeur correspond à la concentration moyenne de radiotracer captée par l'organe dans des conditions cliniques. Une valeur d'atténuation du tissu à énergie fixée de 511 keV est également affectée à chaque organe ou label. Six valeurs de référence sont utilisées pour modéliser les facteurs d'atténuation de chaque structure : celle de l'eau ($0,096 \text{ cm}^{-1}$), des os ($0,151 \text{ cm}^{-1}$), du cerveau ($0,098 \text{ cm}^{-1}$), des muscles ($0,088 \text{ cm}^{-1}$), de l'air ($0,025 \text{ cm}^{-1}$) et de la graisse ($0,088 \text{ cm}^{-1}$).

Les valeurs d'activité affectées à chaque organe ont été dérivées d'images cliniques fournies par le CERMEP, Lyon. L'ensemble des patients se répartit en 22 hommes et 13 femmes de poids moyen 68 ± 13 kg. Un total de 70 images est disponible avec 43 acquisitions pour les hommes et 27 acquisitions pour les femmes, chaque patient ayant subi entre un et trois examens. Les images TEP au 18F-FDG proviennent d'acquisition sur une caméra de type Siemens, ECAT EXACT HR+ (CTI/Siemens, Knoxville). Toutes les images ont été corrigées des effets dégradants (atténuation, normalisation, diffusés...) et ont été reconstruites avec l'algorithme AW-OSEM [Comtat et al., 1998] utilisant 6 itérations, 16 subsets suivi d'un postfiltrage Gaussien de largeur à mi-hauteur de 8 mm dans les trois directions. Ces méthodes respectent celles habituellement réalisées en clinique sur ce type de scanner. Des régions d'intérêt (ROI) dessinées à l'aide du logiciel AMIDE [Loening and Gambhir, 2003] sur les principaux organes d'intérêt des images cliniques ont permis d'extraire des valeurs moyennes d'activité de 18F-FDG. Les ROI ont été placées sur les coupes transverses au centre de chaque structure d'intérêt. Leur taille varie d'un organe d'intérêt à un autre, allant de 30 voxels pour les glandes salivaires à 245 voxels pour le foie. L'ensemble de ces résultats a été validé par deux médecins nucléaires, Dr. T. Mognetti du Centre Leon Berard, Lyon et Pr. F. Giammarile du centre hospitalier Lyon Sud. L'activité des structures cérébrales a été définie à partir de la base d'images simulées cérébrale existantes développée par [Reilhac et al., 2005]. L'ensemble des structures cérébrales de cette base est classée en deux catégories : la matière grise et la matière blanche/peau/muscle/graisse. Chacun de ces deux groupes a reçu une activité de 18F-FDG respectivement de 22,99 kBq/mL et 8,45 kBq/mL. La Figure 4-4 illustre quelques exemples d'images TEP cliniques utilisées pour le contourage des ROI et l'extraction des activités moyennes typiques de 18F-FDG.

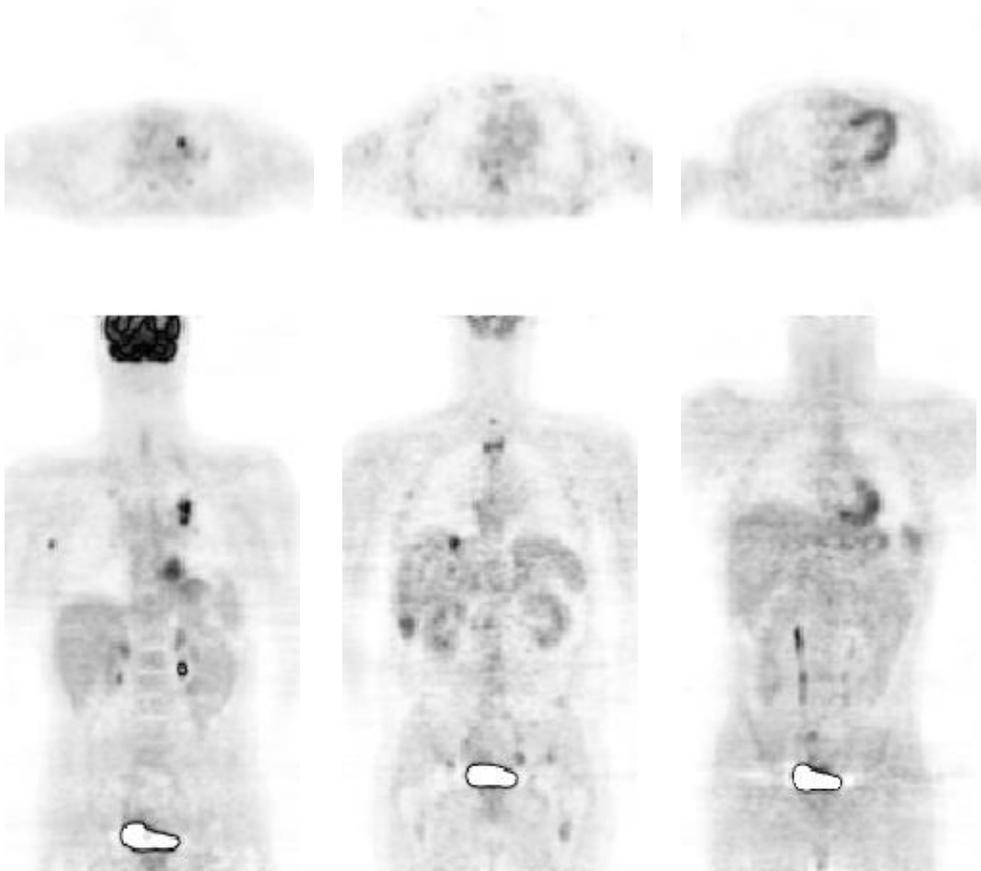


Figure 4-4 : Exemples de coupes coronales d'images cliniques provenant de trois patients différents utilisés pour l'extraction des activités moyennes à appliquer au fantôme numérique Zubal, coupes transverses (ligne du haut), et coronales (ligne du bas)

Le Tableau 4-1 résume les différentes activités extraites des images cliniques et assignées aux labels du fantôme numérique ainsi que les valeurs d'atténuation correspondantes, nécessaires à la simulation.

Tableau 4-1 : Activités de 18F-FDG et coefficients d'atténuation appliqués aux organes principaux du Zubal

Organe	Activité de 18F-FDG (kBq/mL)	Coefficient d'atténuation (cm ⁻¹)
Foie	5,95 +/- 1,31	0,096
Poumons	1,34 +/- 0,48	0,025
Rate	4,39 +/- 0,13	0,096
Cœur	16,85 +/- 10,12	0,096
Vessie	44,09 +/- 19,72	0,096
Moelle osseuse	6,41 +/- 0,16	0,096
Reins	8,22 +/- 3,03	0,096
Muscle	2,58 +/- 0,68	0,088
Matière blanche	8,45	0,098
Matière grise	22,99	0,098

4.5.4 Génération du modèle de tumeurs : extension, taille et contraste

4.5.4.1 Adaptation de notre modèle numérique au lymphome

Comme définie dans la problématique, l'application principale ciblée lors de cette thèse est le lymphome. Cette pathologie est caractérisée par la dissémination de tumeurs de petites tailles parfois difficilement détectables notamment lors du suivi au cours du traitement et de la réadaptation du stade. Comme nous l'avons souligné, les données simulées doivent être le plus réalistes possible pour

permettre une évaluation fiable de notre système de détection. Ce réalisme repose sur deux points : 1) les activités de radiotracer affectées à chaque organe du fantôme, 2) le modèle de tumeurs proposées.

Les tailles des tumeurs sont définies de manière à correspondre à des cas difficiles de détection. Le diamètre des ganglions lymphatiques est approximativement de 6 à 9 mm dans les trois directions. L'étude rétrospective de Adams et al [Adams et al., 1998], a d'ailleurs montré que la TEP permet de détecter des tumeurs de taille minimale 6 mm et que 40% des tumeurs métastasées ont un diamètre inférieur à 1 cm. Pour oncoPET_DB, les tumeurs sont donc approximées par des sphères de 7, 10,5 et 14 mm de diamètre et sont considérées comme non-nécrotiques, c'est-à-dire présentant une fixation homogène de radiotracer. Une distance minimale de quatre voxels, environ 14 mm, entre le centre des tumeurs et le bord des organes est imposée, de manière à limiter les effets de bord. Cette condition n'est cependant pas appliquée à la structure 'artère' modélisant les ganglions lymphatiques, qui contiendra la majorité des tumeurs, de part son volume trop réduit (958 cm³). De même, le recouvrement des tumeurs est interdit afin d'éviter le phénomène de création d'amas de tumeurs. Le coefficient d'atténuation de chaque tumeur est celui de l'eau.

Un détail des différentes affections du lymphome en fonction de la localisation initiale des tumeurs par rapport au diaphragme peut être trouvé au paragraphe 1.4.1. Afin de suivre les cas généralement rencontrés en clinique, nous considérons les affections médiastinales et lombaires. Elles représentent respectivement 50 % et 10 % de tous les cas de lymphome. Dans la génération de notre base d'images simulées, les derniers 40 % correspondent à une propagation diffuse des tumeurs dans les organes cibles du torse du patient. Pour chacune de ces disséminations, les tumeurs sont situées pour 80 % des cas dans les ganglions lymphatiques et 20 % dans les autres organes, tels que les poumons, la rate et le foie. La Figure 4-5 présente les zones d'insertion possibles de tumeurs sur une coupe coronale du Zubal, dans le cas médiastinale a) et lombaire b). Le cas de propagation diffuse des tumeurs correspond à la somme des zones a) et b).

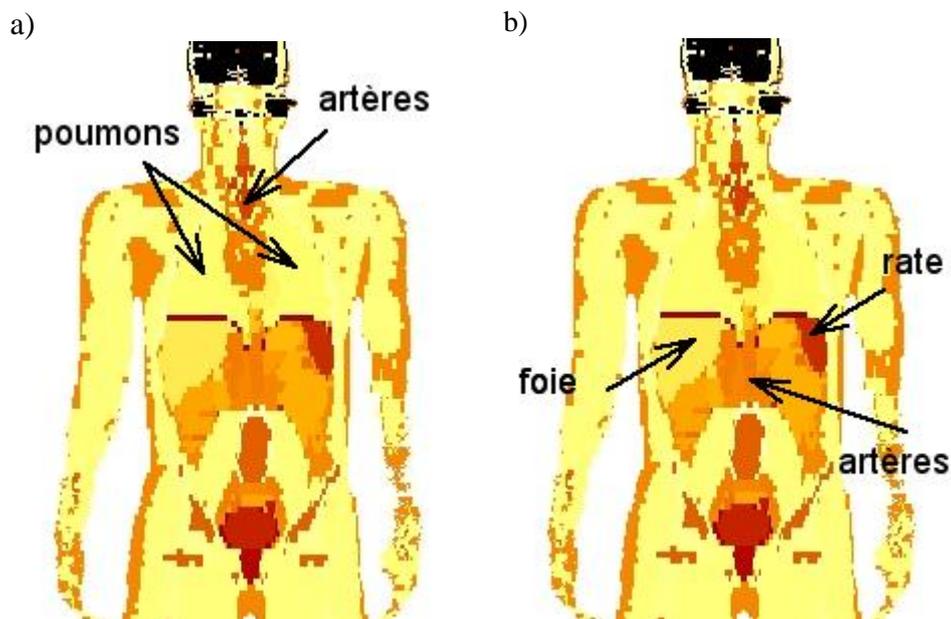


Figure 4-5 : Coupes coronales du Zubal avec les zones d'affection potentielle par le lymphome pointées par les flèches, dans le cas 'médiastinale' a) et 'lombaire' b)

La base d'images TEP simulées finale peut être séparée en deux parties. La première partie des cas pathologiques suit la distribution des différentes affections du lymphome en pourcentage des cas totaux et la localisation des tumeurs dans les organes cibles. Le nombre total de tumeurs par image est fixé à cinq, de manière à éviter le recouvrement de plusieurs tumeurs dans les artères de taille réduite. Le système vasculaire qui sert de référence pour le positionnement des tumeurs des ganglions lymphatiques représente en effet un volume réduit et contient 80 % du nombre total de tumeurs. Les tumeurs sont donc réparties aléatoirement dans chaque organe en fonction de l'affection considérée (lombaire, médiastinale, propagation diffuse).

La deuxième partie de la base d'images simulées contient des cas pathologiques qui ne suivent pas la distribution clinique des affections du lymphome. Dix tumeurs de même caractéristique que précédemment sont ajoutées dans chaque image et réparties aléatoirement dans tous les organes cibles du torse humain (poumons, foie, rate, artères). La distribution des tumeurs dans les organes cibles est approximativement de 30 % dans le foie, les poumons et les artères et de 10 % dans la rate. Cette seconde série d'images permet un compromis entre le temps de simulation Monte Carlo prohibitif et le besoin d'un nombre élevé de tumeurs par image pour la mise en place de notre système d'aide à la détection.

4.5.4.2 Etude de calibration des tumeurs

Après avoir défini la taille, la forme et les possibles localisations des tumeurs, le choix du contraste doit être rigoureusement étudié. Celui-ci est défini comme le rapport entre l'activité de la tumeur et celle du fond. La gamme de contraste à affecter aux tumeurs doit être choisie de telle sorte que celles-ci soient suffisamment visibles et donc détectables, sans être trop évidentes. Une étude de calibration se basant sur des performances de détectabilité humaine permet d'obtenir un ensemble de cinq contrastes par organe et diamètre de tumeur. Une méthodologie similaire à celle développée par [Lartzien et al., 2004] a été utilisée. Celle-ci comporte quatre étapes :

- la génération d'images simulées se basant sur le modèle d'activité défini au paragraphe 4.5.3 contenant N tumeurs de contrastes compris dans un large intervalle pour chaque diamètre et organe cible,
- le report par des observateurs humains des coordonnées des N tumeurs les plus probables de chaque image associées à un degré de certitude pour la position pointée,
- la définition d'une distance d'acceptation entre la position pointée par l'observateur humain et la position réelle et connue du centre de la tumeur qui permet de définir un vrai positif (VP),
- l'échantillonnage régulier des performances humaines en considérant cette distance d'acceptance afin d'extraire l'ensemble des cinq contrastes compris entre les valeurs extrêmes des degrés de certitude.

Les images simulées utilisées pour l'étude de calibration

Douze images TEP au 18F-FDG basées sur le modèle d'activité décrit au paragraphe 4.5.3 contenant chacune N=5 tumeurs sont simulées. Chacune de ces images correspond à un champ axial couvrant le torse humain du haut des poumons au bas du foie. Les tumeurs sont distribuées aléatoirement dans les poumons, le foie, la rate et les artères à raison d'une seule tumeur pour chaque couple diamètre/contraste. Le diamètre des tumeurs dans le fantôme est de 7, 10,5, 14 mm (respectivement 2, 3 et 4 voxels) et les contrastes appartiennent à l'intervalle [2 :20] pour chaque organe. L'étude originelle utilisait également des tumeurs de diamètre 18 mm (5 voxels). Celles-ci ont rapidement été considérées comme trop grandes pour répondre à des cas difficiles de détectabilité et ne figurent pas dans les résultats qui seront présentés dans les parties suivantes. L'ensemble des images utilisées représente donc un total de 60 tumeurs défini par : $3 \text{ diamètres} \times 5 \text{ contrastes} \times 4 \text{ organes}$. La Figure 4-6 illustre des exemples d'images simulées générées pour cette étude de calibration. Chaque image contient un cas extrême de tumeurs en termes de couple diamètre/contraste.

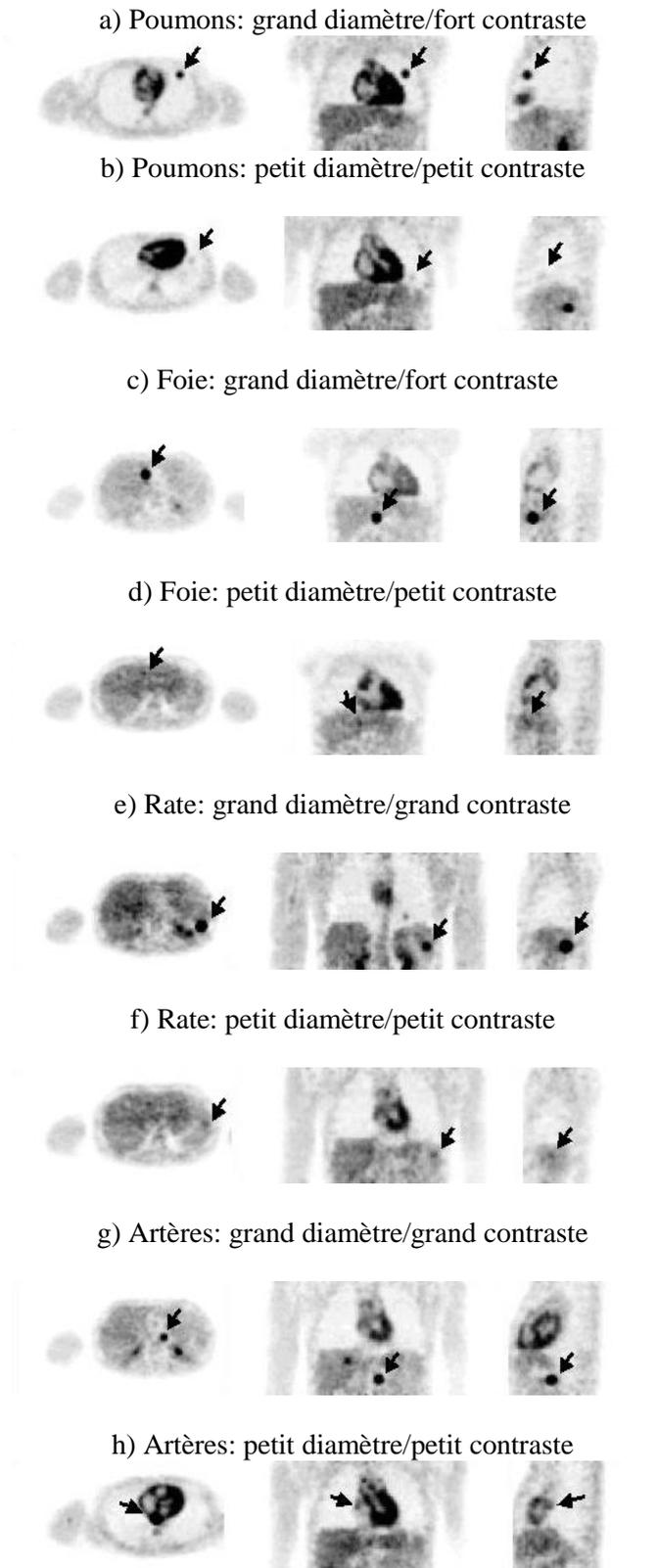


Figure 4-6 : Coupes d'images TEP simulées au ^{18}F -FDG générées pour l'étude de calibration. Ces images illustrent les cas extrêmes de détectabilité des tumeurs, avec le plus haut contraste (20 :1) avec un diamètre de 14 mm et le plus petit contraste (2 :1) avec un diamètre de 7 mm, pour les poumons (a, b), le foie (c, d), la rate (e, f) et les artères (g, h)

La tâche de détection des observateurs humains

Deux observateurs humains ont visualisé ces douze images. Ils ont choisi les cinq positions ayant les plus fortes probabilités de correspondre à des centres des tumeurs en utilisant un degré de certitude choisi sur une échelle de 1 à 5. La valeur 1 correspond à une tumeur très peu visible et 5 une tumeur évidente. Chaque observateur connaissait les organes cibles, le nombre de tumeurs total par image (N=5) et pouvait modifier les paramètres d'affichage des images.

Le choix d'une distance d'acceptation

Dans les cas difficiles, il est possible que l'observateur humain ne pointe pas exactement le centre réel de la tumeur même s'il a correctement identifié un vrai positif. Il est donc nécessaire de fixer une distance d'acceptation séparant la position pointée par l'observateur et la position réelle du centre de la tumeur. Cette distance d'acceptation est simplement une distance euclidienne 3D afin de prendre en compte le fait que l'observateur peut visualiser l'image simulée selon les trois vues transverse, coronale et sagittale. Au-delà de cette distance, la position pointée par l'observateur est considérée comme trop éloignée pour correspondre à la tumeur réelle. Afin de déterminer cette distance, le nombre total de tumeurs effectivement trouvés par les observateurs pour différentes distances d'acceptation est reporté en fonction de la distance d'acceptation variant entre 5 et 35 mm. La Figure 4-7 illustre la courbe obtenue en moyennant les performances des deux observateurs et permet de fixer la distance d'acceptation à 14 mm, valeur au-delà de laquelle la courbe des performances totales se stabilise.

Calibration de la gamme de contraste

Cette étape permet de fixer des cas représentatifs des différents niveaux de difficulté de détection. Les performances des observateurs sont extraites à partir de la distance d'acceptation de 14 mm définie précédemment. Chaque organe est étudié séparément de manière à obtenir une gamme de contraste qui lui est propre. Chacune des tumeurs observées est donc associée à une position et à un degré de certitude précisé par l'observateur humain. La position pointée est validée par sa distance euclidienne à la position vraie par rapport à la distance d'acceptation. Ainsi dans cette partie, seules les positions validées comme bien trouvées entrent en compte pour la détermination des contrastes. Les tumeurs non trouvées par les observateurs sont assignées d'un degré de certitude de 0. Les degrés de certitude moyennés sur les deux observateurs sont affichés en fonction de la catégorie de la tumeur, définie par son organe d'appartenance et son contraste réel. La Figure 4-8 montre les courbes correspondantes pour les poumons a), le foie b), la rate c) et les artères d). Les barres d'erreur correspondent à l'écart type des indices fixés par les deux observateurs. On constate que les courbes sont croissantes avec le contraste jusqu'à saturation. Cette saturation est due au fait qu'une tumeur évidente pour un diamètre et un contraste donné est déjà à 5 et l'augmentation d'un des deux paramètres n'implique pas un meilleur score que celui maximal. De même, les écart-types représentant la variabilité inter-observateur devraient être de plus en plus petits lorsque le contraste augmente. Cependant le nombre peu élevé d'observateurs humains implique une disparité parfois grande des résultats pour certains couples contraste/diamètre dépendant par exemple de la localisation des tumeurs dans l'organe cible. Cette disparité a été prise en compte pour la définition des contrastes.

On remarque que les scores associés au diamètre le plus élevé (14 mm) sont parfois inférieurs aux scores associés à un diamètre plus petit. C'est par exemple le cas des poumons (Figure 4-8a) pour un contraste de 5 :1. Ces résultats inattendus peuvent s'expliquer par la présence d'une et une seule tumeur par contraste et diamètre pour chaque organe sur l'ensemble des images de calibration. Il est donc possible que cette tumeur spécifique de diamètre 14 mm et contraste 5 :1 soit localisée dans une région difficile, c'est-à-dire au bord de l'organe cible ou près du cœur. Cette localisation pourrait impacter le score donné par les deux observateurs humains. Une telle remarque peut aussi être faite certains couples de caractéristiques pour le foie (Figure 4-8b).

Une gamme de cinq valeurs de contraste est définie par un échantillonnage des courbes de la Figure 4-8 pour chacun des organes. Les remarques faites précédemment sur l'allure générale des courbes ont été prises en compte. Ces contrastes couvrent l'ensemble des performances de détectabilité humaine, allant du plus faible niveau de confiance au plus élevé (0,5, 1,5, 2,5, 3,5, 4,5). Les contrastes ainsi choisis sont rassemblés sur la Figure 4-9 pour chaque organe et diamètre de

tumeurs. Chaque barre correspond à un contraste pour un couple diamètre/organe. Ces contrastes sont ensuite appliqués aux tumeurs de la base d'images simulées oncoPET_DB. Pour la première partie de la base simulée, les probabilités de distribution dans les organes cibles sont celles définies par les cas cliniques du lymphome en 4.5.4.1. Pour la seconde partie, aucune distinction d'affection n'est réalisée. Le choix de la localisation de chaque tumeur dans cet organe, de son diamètre et du contraste correspondants, répond à une distribution normale de probabilité décroissante de relation inverse avec le diamètre des tumeurs (10% pour chaque couple contraste/organe pour les tumeurs de 7 mm, 6% pour les tumeurs de 10,5 mm et 4% pour les tumeurs de 14 mm).

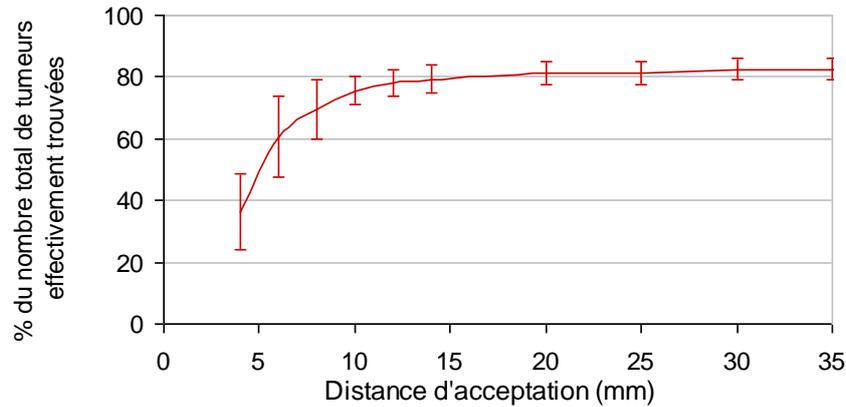


Figure 4-7 : Evolution du pourcentage de tumeurs réellement trouvées par les observateurs en fonction de la distance d'acceptation en mm

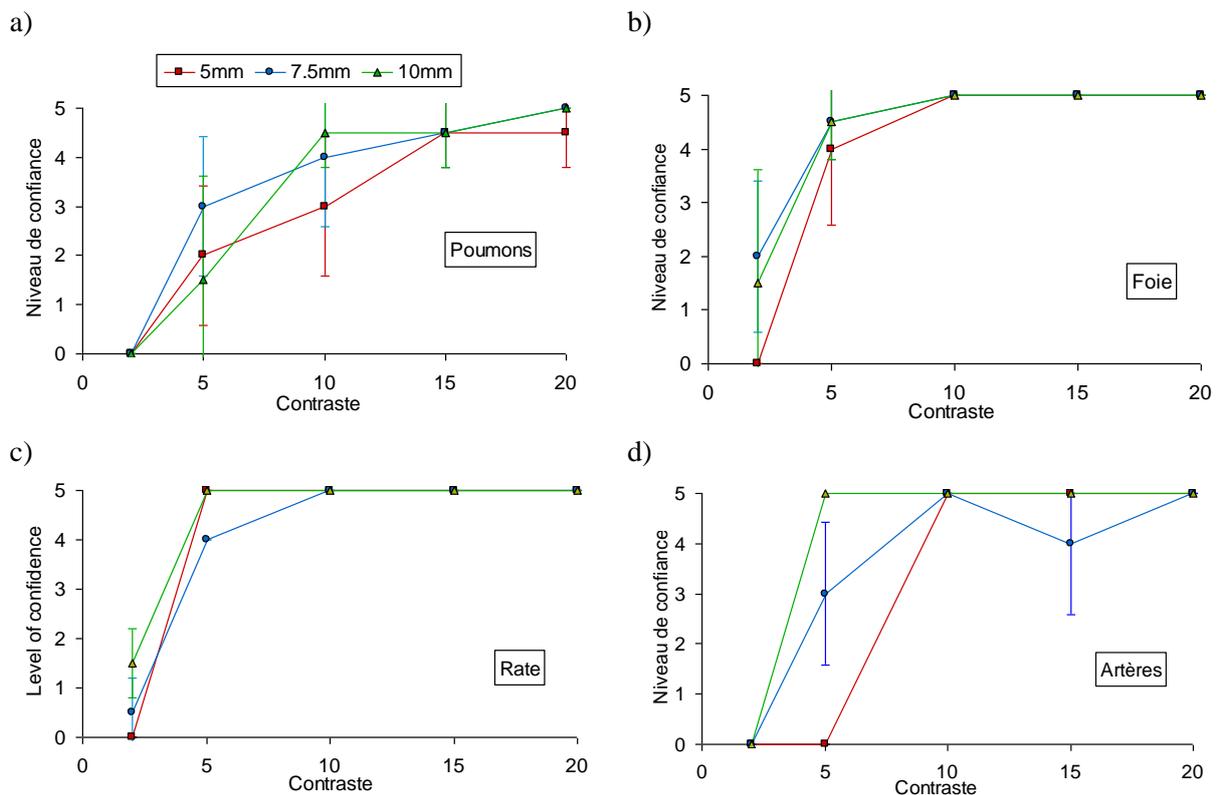
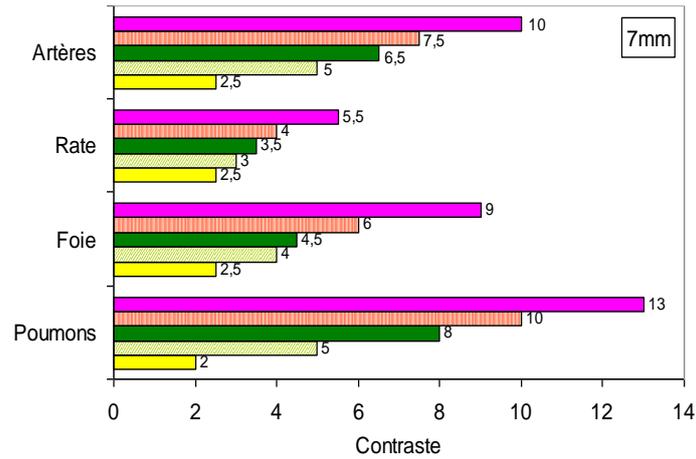
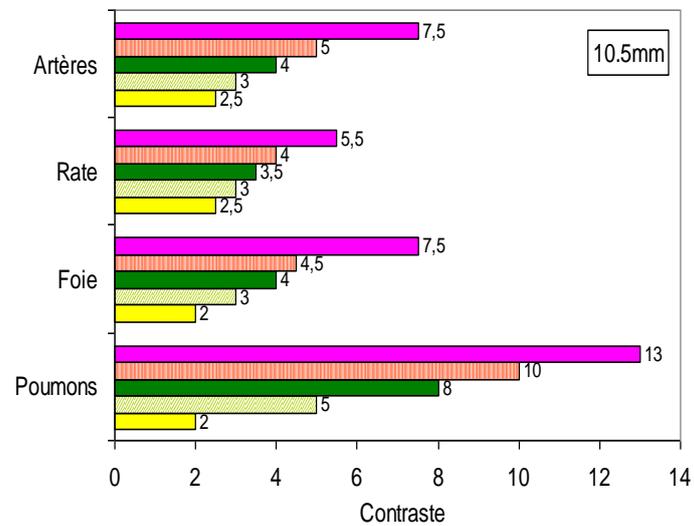


Figure 4-8 : Variation du niveau moyen de certitude assigné à chaque tumeur de l'étude de calibration. Les résultats sont exprimés en fonction du contraste vrai et du diamètre des tumeurs localisées dans le poumon a), le foie b), la rate c) et les artères d).

a)



b)



c)

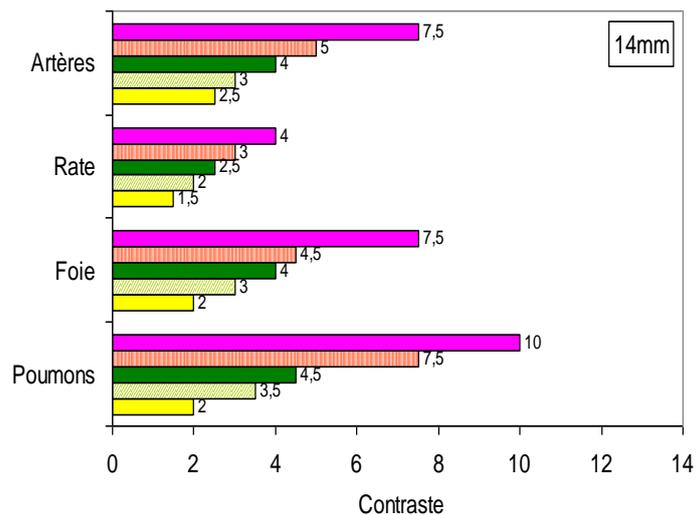


Figure 4-9 : Le choix de cinq contrastes par organe et par diamètre de tumeurs : 7 mm a), 10,5 mm b), 14 mm c)

4.5.5 De la simulation à la reconstruction d'oncoPET_DB

Les sinogrammes d'émission et de transmission sont générés à l'aide du simulateur Monte Carlo PET SORTEO (<http://sorteo.cermep.fr>) développé par [Reilhac et al., 2005]. Afin de reproduire

le plus fidèlement possible les acquisitions TEP cliniques fournies par le CERMEP, nous avons choisi de reproduire la géométrie de la caméra Siemens ECAT EXACT HR+ (CTI/Siemens, Knoxville), pour laquelle le simulateur PET-SORTEO a été validé [Reilhac et al., 2005]. Les données sont simulées en mode d'acquisition 3D et corrigées de tous les effets traités en routine clinique. Ces corrections incluent toutes celles définies au paragraphe 1.2.2, auxquelles s'ajoute la correction de décroissance du radiotracer. Chaque sinogramme d'émission est corrigé du phénomène d'atténuation qui entache les données à l'aide de son propre sinogramme de transmission simulé. Actuellement, cette correction est le plus souvent réalisée en clinique grâce à une image TDM provenant d'une acquisition combinée TEP/TDM [Zaidi et al., 2007]. Notre approche également utilisée dans certains centres de recherche est basée sur le sinogramme de transmission et permet de nous affranchir de l'intégration du bruit généré lors de la simulation d'acquisitions TDM réalistes. Nous revenons sur ce choix dans la discussion du paragraphe 4.10.

Comme expliqué au paragraphe 1.2.5, les acquisitions cliniques corps entier sont réalisées par juxtapositions de plusieurs acquisitions permettant de couvrir l'ensemble du patient. Ceci est obtenu par un déplacement pas à pas du lit dans le champ de vue axial. Cette méthode permet de couvrir l'ensemble du patient. Le simulateur PET SORTEO permet de modéliser cette acquisition multi-lit. Sept lits sont nécessaires pour couvrir le fantôme Zubal de la tête aux mi-cuisses, comme illustré sur la Figure 4-10. La simulation correspond à 7 min d'acquisition en émission et 3 min d'acquisition en transmission par lit, débutant 60 min post-injection.

Les images simulées sont reconstruites à l'aide de l'algorithme itératif AW-OSEM utilisant 6 itérations et 16 subsets, suivi d'un postfiltrage Gaussien de largeur à mi-hauteur de 8 mm dans les trois directions. Les images résultantes sont des volumes de 128 x 128 x 375 voxels de dimension 5,0625 x 5,0625 x 2,4250 mm.

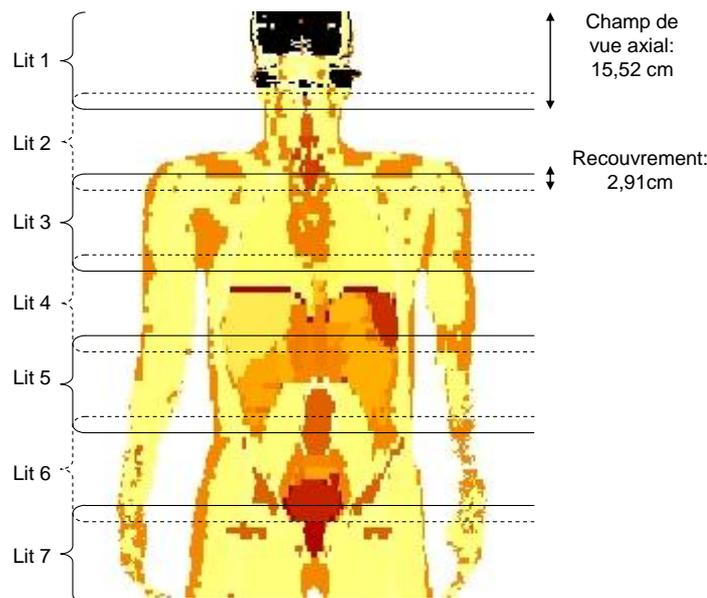


Figure 4-10 : Les sept lits permettant de couvrir le Zubal par simulation TEP

La simulation d'une image TEP tridimensionnelle corps entier est un procédé très lent. Afin de limiter les temps de calcul, les ressources du centre de calcul de l'IN2P3 (<http://cc.in2p3.fr>) ont été mises à contribution. La simulation de chaque lit est donc séparée en plusieurs tâches soumises en parallèle et aboutissant à un total de 90 heures CPU sur vingt processeurs 32 bits sous licence Linux. Il faut cependant noter que les temps de simulation incluent les délais de soumission des tâches qui dépendent des utilisations extérieures.

4.6 La base d'images simulées résultante

OncoPET_DB contient cinquante images TEP simulées type 'saines' sans tumeur et cinquante images type 'pathologiques' avec tumeurs.

4.6.1 Les images 'saines' d'oncoPET_DB

La série des cas 'sains' consiste en 50 réalisations de bruit différentes de la distribution de ^{18}F -FDG du Zubal décrite en 4.5.3. En d'autres termes, le même patient virtuel a subi 50 fois le même examen en considérant la même captation initiale des organes en radiotracteur. La seule variation provient de l'initialisation des tirages aléatoires qui régissent les effets physiques modélisés lors de la simulation. La Figure 4-11 présente une coupe coronale d'une image 'saine' a) comparée à une image clinique b) provenant du CERMEP, Lyon. Visuellement, les deux images présentent de fortes similitudes, même si les os du bassin de l'image simulée semblent plus captants et la séparation entre les poumons et le foie un peu trop horizontale et marquée par le diaphragme.

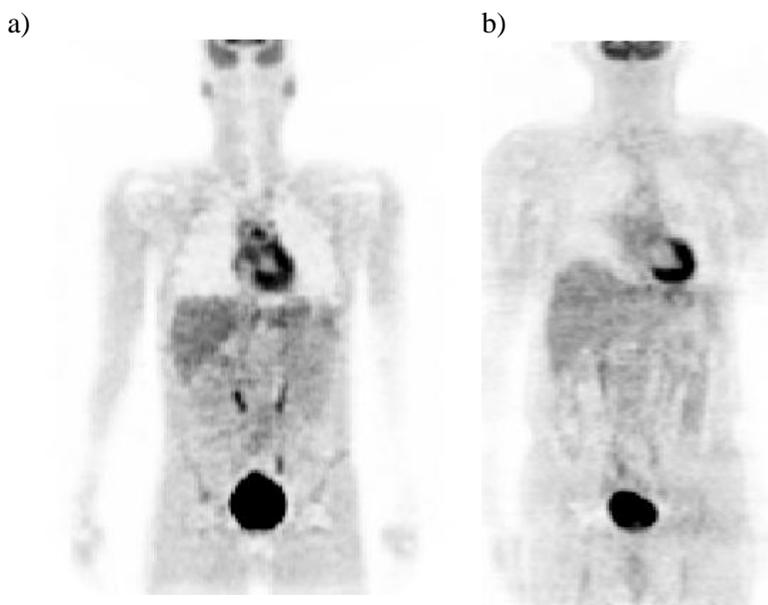


Figure 4-11 : Coupes coronales d'une image TEP simulée a) et d'une image clinique b)

4.6.2 Les images 'pathologiques' d'oncoPET_DB

La série des cas 'pathologiques' est elle-même séparée en deux catégories. La première catégorie de 25 images reproduit fidèlement la distribution statistique des cas de lymphome comme expliqué en 4.5.4.1. Chaque image contient cinq tumeurs de contraste et de diamètre différents. Conformément aux cas cliniques de lymphome, 12 images pathologiques correspondent à une affection médiastinale (50 %), 2 à une affection lombaire (10 %) et 11 à une propagation diffuse dans le torse du patient (40 %). Un total de 125 tumeurs sphériques dont 105 dans les artères, 4 dans le foie, 15 dans les poumons et 1 dans la rate a été distribué aléatoirement dans chacune de ces trois catégories.

La seconde catégorie de 25 images inclut dix tumeurs chacune, réparties dans les poumons, le foie, les artères et la rate avec une probabilité de 30 % pour les trois premiers organes et de 10 % pour le dernier en raison de son volume restreint. La rate présente en effet une texture et une captation proche du foie. La taille de celui-ci permet une plus forte distribution de tumeurs pour une texture très proche. Les mêmes couples de contraste et diamètre sont utilisés, même si la distribution statistique des cas de lymphome n'est pas respectée. Un total de 250 tumeurs sphériques, 75 dans les artères, les poumons et le foie et 15 dans la rate est distribué aléatoirement. La Figure 4-12 présente des coupes coronales de deux images simulées pathologiques différentes provenant d'oncoPET_DB. Chaque coupe contient trois tumeurs localisées dans les artères. Pour la coupe a), les couples diamètre/contraste de haut en bas sont : 14 mm/4 :1, 10,5 mm/7,5 :1, 10,5 mm/7,5 :1. Pour la coupe b), les couples diamètre/contraste de haut en bas sont : 10,5 mm/5 :1, 14 mm/5 :1, 10 mm/4 :1. Notons qu'aucune de ces tumeurs n'a son centre dans la même coupe coronale contrairement à ce que l'on pourrait croire au vue des images.

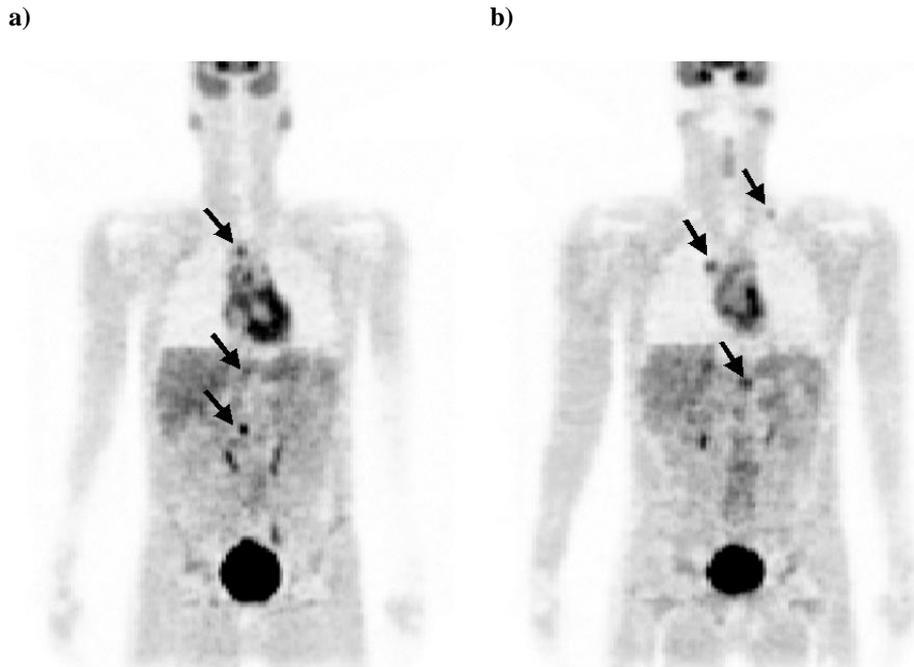


Figure 4-12 : Coupes coronales de deux images TEP simulées contenant chacune trois tumeurs dans les artères

4.7 Validation de la base d'images simulées

4.7.1 Introduction

Cette partie expose les différentes mesures quantitatives utilisées pour valider la pertinence d'oncoPET_DB par rapport :

- aux images cliniques du CERMEP qui ont permis la mise en place du modèle d'activité,
- à des études cliniques provenant de trois articles de référence traitant de patients 'sains' et 'pathologiques'.

4.7.2 Mesures statistiques sur les activités de 18F-FDG

4.7.2.1 Introduction

Bien que l'analyse visuelle des images TEP simulées semble concluante, il est important de valider leur pertinence de manière quantitative. Plusieurs mesures quantitatives, incluant des moments statistiques d'ordre un et deux des activités simulées, sont donc utilisées afin d'évaluer la similarité des images simulées et cliniques. Une mesure de déviation standard fractionnelle (DSF) est aussi calculée sur une série de cinq images simulées et cinq images patients suivant la méthodologie de [Aristophanous et al., 2008]. Cette mesure est définie comme la déviation standard d'une ROI divisée par sa moyenne.

4.7.2.2 Moyenne et déviation standard des activités de 18F-FDG

Des ROI sont dessinées sur $N = 25$ images simulées et les $N = 70$ images cliniques de manière à extraire les moyennes des activités sur ces régions. La déviation standard (std) intra-ROI de ces activités moyennes, elle-même moyennée sur ces N images est calculée selon la formule (4-1).

$$std = \frac{\sqrt{\sum_{i=1}^N \sigma_i^2}}{N} \quad \text{ou } N \text{ est le nombre d'images et } \sigma_i \text{ est la déviation standard de la région } i \quad (4-1)$$

La Figure 4-13 résume les valeurs statistiques obtenues pour les images simulées et cliniques. A l'exception de la matière grise, du foie et du cœur, les activités simulées et cliniques sont bien corrélées, notamment en considérant leurs déviations standards respectives. Le foie est certainement contaminé (effet de volume partiel) par l'activité du rein droit hypercaptant situé proche de lui. Cet effet est aussi observé pour la rate qui, contaminée par le rein gauche, présente une activité moyenne simulée légèrement trop élevée. En ce qui concerne la matière grise et le cœur, les différences peuvent être expliquées par un tracé de ROI parfois difficile du à la complexité des organes considérés (cavités pour le cœur, matière grise et matière blanche enchevêtrées pour le cerveau). De plus, les activités des structures cérébrales ont été fixées à l'aide de valeurs mesurées dans [Reilhac et al., 2005] pour lesquelles n'apparaissent pas les déviations standards. Il est donc difficile d'évaluer si les activités simulées appartiennent à l'intervalle possible des activités mesurées. Néanmoins, dans le cas particulier du lymphome, le cerveau est plus considéré comme une source de contamination d'activité en dehors du champ de vue qu'une structure précisément ciblée. Cette différence d'activité peut donc être considérée comme anodine. Pour résumer, les moyennes des activités sur les différents organes simulés sont globalement bien corrélées aux mesures cliniques.

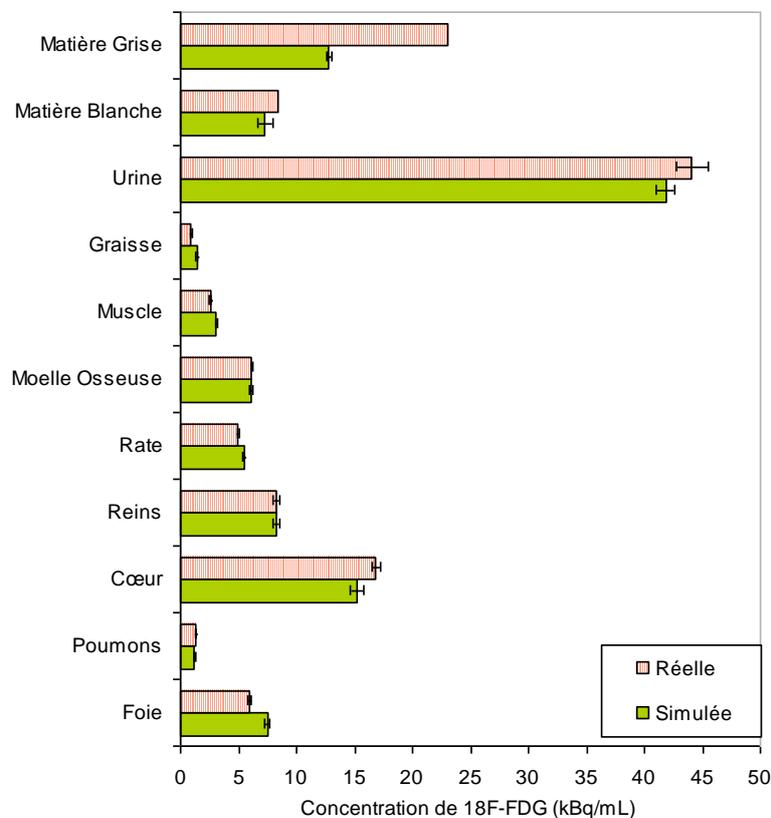


Figure 4-13 : Comparaison des activités simulées et réelles de 18F-FDG dans les principaux organes d'intérêt

4.7.2.3 Déviations standard fractionnelles (DSF)

Cette mesure, définie en (4-2), permet de comparer la déviation standard de groupes de pixels dans des situations variées.

$$DSF = \frac{\sigma_i}{\mu_i} \text{ où } \sigma_i \text{ et } \mu_i \text{ sont les déviations standards et la moyenne de la région } i \quad (4-2)$$

En effet, la normalisation par la moyenne des pixels rend la DSF indépendante de la valeur d'activité absolue. Le Tableau 4-2 résume les DSF extraites et moyennées sur cinq ROI définies sur les principaux organes des images simulées et cliniques. Les tailles des ROI utilisées pour les données simulées et cliniques, également présentées dans le Tableau 4-2, ont été choisies de même ordre de grandeur de manière à ne pas influencer la comparaison de ces mesures. Les DSF des images simulées et cliniques sont généralement proches, même si certains organes simulés comme le foie, présentent

des valeurs légèrement plus élevées. Les valeurs cliniques pour les reins et la vessie affichent également une plus grande hétérogénéité que les organes simulés. La vessie simulée, apparaissant comme une structure uniformément remplie, ne peut refléter la grande variabilité inter-patient due à des paramètres physiologiques tels que l'élimination du FDG par les voies urinaires avant et après injection. Les variations autour des valeurs de DSF cliniques englobent cependant les valeurs simulées. En conclusion, la proximité des DSF cliniques et simulées reflète une proche correspondance entre les deux catégories d'images.

Tableau 4-2 : DSF calculées à partir de cinq ROI tracées dans les organes principaux des images simulées et cliniques

	Taille des ROIs simulées (voxels)	DSF simulées	Taille des ROIs cliniques (voxels)	DSF cliniques
Foie	370 +/- 10	0,15 +/- 0,01	378 +/- 7	0,10 +/- 0,02
Poumons	346 +/- 6	0,23 +/- 0,04	354 +/- 6	0,20 +/- 0,03
Rate	125 +/- 4	0,11 +/- 0,02	123 +/- 1	0,11 +/- 0,03
Cœur	162 +/- 7	0,14 +/- 0,03	158 +/- 5	0,12 +/- 0,03
Vessie	111 +/- 7	0,06 +/- 0,01	110 +/- 5	0,13 +/- 0,12
Reins	200 +/- 6	0,23 +/- 0,03	194 +/- 4	0,30 +/- 0,09
Muscle	172 +/- 2	0,11 +/- 0,02	177 +/- 2	0,12 +/- 0,01

4.7.3 Mesures cliniques

4.7.3.1 Introduction

Le paragraphe précédent a permis de montrer la bonne correspondance entre les images simulées et cliniques provenant du CERMEP, en termes de mesures statistiques et de bruit. Quelques différences pour certains organes cibles sont présentées mais explicables, notamment par l'effet de volume partiel. Il est également important de vérifier que les données simulées correspondent aux données réelles d'un point de vue clinique. Dans cet objectif, nous nous sommes intéressés à deux paramètres, la dose injectée de radiotracer et la valeur semi-quantitative standard qu'est le SUV (cf. paragraphe 1.2.4).

4.7.3.2 Dose mesurée vs. Dose injectée

La dose de radiotracer injectée au patient se mesure en mCi ou en MBq sachant qu'un mCi équivaut à 37 MBq. Dans le cas où le corps entier du patient, de la tête au pied, est visible sur l'image TEP, la somme des activités de tous les voxels, est égale à cette dose injectée. Ceci est valable uniquement si les images reconstruites ont été corrigées du phénomène de décroissance radioactive et reflètent l'activité supposée présente au moment de l'injection. Cependant, le champ de vue axial de l'acquisition est généralement réduit à quelques lits qui couvrent, dans notre cas, le patient de la tête aux mi-cuisses. La somme des activités mesurée est donc inférieure à la dose injectée. Il semble donc ainsi adapté de comparer, non pas les doses injectées aux patients mais les sommes des activités dans le champ de vue restreint. Cette somme d'activité est mesurée et moyennée sur cinq patients CERMEP de morphologie similaire au Zubal. La Figure 4-11b montre une coupe coronale d'un des patients utilisés. L'activité moyenne mesurée sur cinq images simulées atteint 3,92 +/- 0,015 mCi. Elle est comparée à une valeur clinique de 3,38 +/- 1,35 mCi pour les cinq patients sélectionnés. Les valeurs cliniques et simulées apparaissent ainsi bien corrélées.

Il semble évident qu'aucune dose injectée cohérente ne peut être extraite des images simulées. En effet, bien que la valeur mesurée de 3,92 +/- 0,015 mCi corresponde à un certain pourcentage de la dose injectée, il nous est impossible de définir précisément ce pourcentage. La dose injectée est cependant une mesure essentielle à la détermination de paramètres semi-quantitatifs utiles à la validation des images simulées tel que le SUV. Etant donné la bonne concordance des images simulées et cliniques sur les différents points ci-détaillés, le paragraphe suivant utilise seulement les images cliniques et leurs SUV en référence pour la dernière étape de validation d'oncoPET_DB.

4.7.3.3 Valeurs de SUV

Le SUV, défini au paragraphe 1.2.4, est une mesure semi-quantitative très fréquemment utilisée pour l'évaluation des images TEP cliniques. Les SUV calculés pour différents organes des images cliniques du CERMEP ont ainsi été comparés avec ceux extraits d'articles de référence [Tan and Ong, 2004; Souvatzoglou et al., 2007; Zincirkeser et al., 2007]. Seuls les SUV normalisés par le poids du patient sont présentés dans ce paragraphe. Etant donné que cette mesure peut varier avec un certain nombre de critères physiologiques et morphologiques [Feuardent et al., 2003], une liste de ces paramètres pour chaque étude est présentée dans le Tableau 4-3.

Tableau 4-3 : Paramètres cliniques des études provenant des différents articles de référence et du CERMEP pour la comparaison des SUV

	Zincirkeser, Sahn et al. 2007	Souvatzoglou, Ziegler et al. 2007	Tan and Ong 2004	CERMEP
Dose (mCi)	11,7 +/- 1	10,7 +/- 1,4	13,4	8,6 +/- 1,5
Nb Homme	38	27	52	22
Nb Femme	32	8	48	13
Poids (kg)	73 +/- 17 a	–	59 [38-72]	68 +/- 12 [50-102]
Diète	4h	4h	4h	4h
Pathologie	Sain	Différents cancers*	Sain	Lymphome
Scanner	Biograph	HR+/Biograph	Biograph	HR+

*Lymphome/ Cancer de l'œsophage/ Carcinome de la thyroïde

Etant donné que [Tan and Ong, 2004; Souvatzoglou et al., 2007] considèrent les SUV provenant de patients masculins et féminins en un seul groupe, nous avons vérifié cette équivalence entre les deux catégories d'individus sur les données cliniques du CERMEP. La Tableau 4-4 expose les résultats de tests statistiques de Student (t-test) et de Fisher (F-test) réalisés sur les SUV mesurés à partir d'une sélection d'organes des images cliniques du CERMEP. Le seuil $p = 0,05$ est considéré comme significatif. Les degrés de liberté (ddl) affichés dans le Tableau 4-4 correspondent au nombre de ROI utilisées pour chaque organe et sont identiques pour chaque catégorie de patients (homme/ femme).

Tableau 4-4: Comparaison statistiques des SUVs homme et femme obtenu pour un nombre identique de patients

	ddl	F-test	p	t-test	p
Foie	27	1,18	0,51	0,92	0,37
Poumon gauche	27	1,81	0,06	1,23	0,22
Rate	24	1,89	0,11	1,35	0,19
Cœur	25	1,04	0,9	0,69	0,49
Rate	27	1,77	0,07	1,94	0,06
Moelle Osseuse	27	1,32	0,58	1,01	0,32
Reins	20	1,85	0,17	2,21	0,04
Muscle	26	1,57	0,7	0,85	0,4

Les différences significatives ($p < 0,05$) sont affichées en gras

Les p-valeurs sont généralement supérieures au seuil significatif de 0,05, excepté pour les reins, dont la p-valeur est de 0,04. Le Tableau 4-4 confirme donc l'hypothèse des articles de référence étant donné qu'aucune différence significative n'a été trouvée entre les patients hommes et femmes en termes de SUV.

Suivant cette hypothèse de groupe unique de patients, le Tableau 4-5 compare les SUV de référence [Tan and Ong, 2004; Souvatzoglou et al., 2007; Zincirkeser et al., 2007] avec ceux extraits des images du CERMEP pour des organes observés conjointement dans la majorité des études.

Tableau 4-5 : Comparaison des SUV publiés et expérimentaux

	Zincirkeser, Sahn et al. 2007	Souvatzoglou, Ziegler et al. 2007	Tan and Ong 2004	CERMEP
Foie	3,2 +/- 0,8	2,2 +/- 0,5	2,8	1,3+/-0,3
Poumons	0,6 +/- 0,2	0,5 +/- 0,1	0,6	0,3 +/- 0,1
Coeur	5,2 +/- 3,8	–	–	4,7 +/- 2,1
Moelle Osseuse	–	2,1 +/- 0,5	–	1,3 +/- 0,3
Rate	2,4 +/- 0,6	1,8 +/- 0,4	2,3	0,9 +/- 0,3
Cerveau	6,7 +/- 1,3	5,1 +/- 1,2	6	–

Pour les poumons, les valeurs de SUV CERMEP sont proches de celles extraites de l'étude de [Souvatzoglou et al., 2007]. Les deux études traitent de patients atteints de la même pathologie (lymphome) dont les acquisitions sont réalisées sur la même plateforme (Siemens ECAT EXACT HR+). Les SUV correspondants aux autres organes apparaissent cependant plus faibles que les SUV de référence. Cette sous-estimation peut éventuellement être expliquée par le fait que la base d'images CERMEP contient à la fois des images TEP pré- et post-traitement. Des phénomènes inflammatoires peuvent apparaître à l'issue des traitements et influence le calcul des SUV [Cheson, 2007]. L'étude de [Souvatzoglou et al., 2007] ne concerne en effet que des acquisitions pré-traitement.

Les SUV calculés dans l'étude de [Tan and Ong, 2004] et [Zincirkeser et al., 2007] apparaissent également plus élevés que ceux du CERMEP et de [Souvatzoglou et al., 2007]. L'intervalle de valeurs pour le cœur et les poumons inclue cependant les valeurs du CERMEP. Une explication possible d'une telle différence repose sur l'utilisation de cas sains dans [Tan and Ong, 2004] et [Zincirkeser et al., 2007]. Les SUV reflètent donc des fixations physiologiques différentes de celles des cas pathologiques. Il est de toute manière évident que les valeurs de SUV varient d'une étude à l'autre et ne peuvent pas totalement se recouvrir.

Cette comparaison entre les SUV du CERMEP et les SUV publiés dans des articles de référence n'est cependant pas totalement concluante et ne permet pas de valider entièrement oncoPET_DB sur cette mesure. En effet, même si certains effets physiologiques peuvent expliquer une faible différence de SUV, ils ne peuvent justifier la sous-estimation importante d'un organe tel que le foie.

4.8 Mise en ligne d'oncoPET_DB

L'objectif principal de la génération d'oncoPET_DB est de permettre l'évaluation quantitative précise des performances du système d'aide à la détection développé dans cette thèse. Cependant, oncoPET_DB peut également s'avérer intéressante pour d'autres applications et il a semblé pertinent de pouvoir la mettre à disposition de la communauté scientifique intéressée. L'ensemble des images peut être présenté comme un outil facile et rapide d'utilisation pour l'évaluation d'algorithmes améliorant la qualité des images TEP. Des algorithmes impactant l'amélioration du contraste de petites structures (reconstruction, correction d'effet de volume partiel) ou des systèmes d'aide à la détection pourraient en effet être validés à l'aide d'oncoPET_DB. Quelques équipes nous ont d'ailleurs déjà

contactées afin d'utiliser les images simulées pour la validation d'algorithmes de segmentation de tumeurs (PINlab en Suisse, Université Polytechnique de Hong Kong).

Un portail web, http://www.creatis.insa-lyon.fr/oncoPET_DB/, développé avec le concours de l'équipe informatique et plus spécifiquement Sorina Pop, permettra l'échange des données avec les utilisateurs extérieurs. Les cent images simulées saines et pathologiques (12 Mo chacune), associées à leur sept sinogrammes d'émission (19 Mo chacun) seront disponibles. De même, les modèles Zubal modifiés contenant les différentes tumeurs seront mis à disposition permettant par exemple de corriger les données de l'atténuation, non plus en se basant sur les acquisitions en transmission, mais sur des scans TDM simulés bruités. Afin de faciliter la requête d'utilisateurs qui souhaiteraient reconstruire les données à l'aide d'algorithmes ou paramètres différents de ceux actuels, les acquisitions « blanc », les images de normalisation et un lien vers des codes libres de reconstruction, seront également accessibles.

4.9 Discussions

OncoPET_DB est une base d'images simulées TEP au 18F-FDG basée sur le fantôme Zubal associé à une carte d'activité extraite d'images cliniques et dont le modèle d'extension de tumeurs suit la réalité des cas de lymphome. Elle a l'avantage d'éviter une perte de temps considérable due à la prise en main d'un simulateur Monte Carlo et la simulation des images. Initialement développée pour permettre l'évaluation quantitative des performances du système CAD développé dans ce manuscrit, elle apparaît intéressante pour d'autres applications. Parmi ces applications peuvent être cités les algorithmes impactant le contraste des petites structures (effet de volume partiel, corrections), de quantification et de traitement d'image (segmentation).

OncoPET_DB présente 50 images pathologiques contenant des tumeurs et 50 images saines. La méthodologie de la génération d'oncoPET_DB repose sur plusieurs choix stratégiques détaillés au paragraphe 4.5. La base d'images simulées a été validée sur plusieurs points statistiques et cliniques, résumés dans le paragraphe 4.7, qui la rendent pertinente par rapport à des acquisitions cliniques réelles. Elle présente néanmoins plusieurs limites.

Dans un premier temps, la valeur de SUV correspondant au foie est beaucoup trop faible par rapport aux valeurs fournies par les articles de référence présentés au paragraphe 4.7.3.3. Il serait donc intéressant de la réévaluer afin d'assurer un plus grand réalisme de la carte d'activités des images d'oncoPET_DB.

Concernant le choix du modèle humain d'activité en 18F-FDG, nous avons choisi de combiner le fantôme Zubal à une sélection d'activités réalistes extraites d'images TEP réelles pour les principaux organes d'intérêt. Une alternative aurait été d'initier notre modèle à partir des images TEP réelles en y insérant des tumeurs simulées de contraste calibré. Ce choix aurait induit une distribution réaliste du traceur et une variabilité anatomique inter-patient, qui correspond à deux points faibles de la base d'images simulées actuelles. Cette solution présente cependant des inconvénients. Le premier repose sur l'accessibilité à un large nombre d'images saines, incluant les sinogrammes et l'image de normalisation, indispensables à la reconstruction des données. Le second inconvénient se base sur une insertion correcte des tumeurs simulées aux sinogrammes réels. Il semble en effet difficile de reproduire parfaitement les effets physiques impactant l'acquisition lorsque les tumeurs uniques sont simulées, impliquant des structures de bruit différentes de la réalité. De même, il semble difficile de calibrer les contrastes des tumeurs simulées lorsque le fond de celles-ci n'est pas parfaitement connu. Nous avons donc préféré contrôler l'ensemble du processus de simulation en fixant l'anatomie du patient tout en extrayant les activités des principaux organes à partir d'images cliniques. La comparaison réalisée entre la dose injectée calculée dans le champ de vue sur les images simulées (3,92 mCi) et sur les images cliniques fournies par le CERMEP (3,38 mCi) prouve d'ailleurs une bonne corrélation entre les modèles d'activité.

Une autre limitation de la simulation concerne l'utilisation de la géométrie de la caméra HR+. Cette caméra peut actuellement apparaître dépassée, notamment lorsqu'on considère les nouvelles caméras TEP/TDM disponibles en application clinique. Les performances en termes de taux de comptage peuvent en effet apparaître supérieures pour les nouvelles caméras du fait d'une électronique plus rapide et de l'utilisation de détecteurs LSO. La résolution intrinsèque de la caméra HR+ reste quant à elle très proche de celles des caméras actuelles [Adam et al., 1997; Jakoby et al., 2006; Surti et

al., 2007; Teras and Tolvanen, 2007]. Les améliorations principales des nouvelles caméras concernent également la partie post-traitement et proposent notamment des algorithmes de reconstruction itératifs plus performants prenant en compte la PSF (fonction d'étalement du point ou *Point Spread Function* en anglais). Une résolution de 2 à 3 mm peut ainsi être obtenue grâce à leur utilisation.

Comme mentionné au paragraphe 4.5.5, les sinogrammes d'émission n'ont pas été corrigés de l'atténuation en utilisant l'image TDM correspondante comme actuellement réalisée en routine clinique. Nous avons utilisé une méthode similaire basée sur l'image en transmission. Nous avons ainsi voulu éviter l'inclusion de bruit additionnel due à la simulation TDM comme paramètre de l'étude. Dans le cadre de la distribution d'oncoPET_DB, nous mettrons à disposition les volumes d'atténuation utilisés pour la simulation. Les utilisateurs intéressés pourront ainsi générer les projections bruitées en combinant les modèles labellisés (fantômes) à un simulateur CT analytique [Comtat et al., 1999] ou Monte Carlo [Tabary et al., 2007].

Enfin le choix du fantôme Zubal permet une grande précision au niveau de la description artérielle, mais ne permet pas de modéliser les mouvements respiratoires et cardiaques qui pourrait impacter la détection des tumeurs dans les poumons. L'utilisation d'un fantôme tel que le XCAT [Segars et al., 2009] décrit au paragraphe 4.5.2, permettrait d'introduire une variabilité intra-individuelle absente de la base actuelle. L'utilisation du XCAT permettrait aussi de tenir compte de l'aspect mouvement respiratoire qui joue un rôle important dans les acquisitions TEP. Cette perspective est par ailleurs en cours de réalisation dans le cadre d'une collaboration entre le LATIM, Brest et le CREATIS-LRMN [Le Maitre et al., 2009].

4.10 Conclusions sur le développement d'oncoPET_DB

A ce jour, il n'existe aucune base d'images TEP simulées spécifiques à l'oncologie correspondant à la structure d'une base d'images cliniques réelles. Les deux seules bases d'images existantes comportent pour l'une, des acquisitions cérébrales uniquement [Reilhac et al., 2005] et pour l'autre un ensemble d'acquisitions de fantômes et de radiotraceurs divers utilisant des simulateurs variés [Castiglioni et al., 2005]. Cependant le besoin en volume important d'images TEP simulées propres à l'oncologie augmente face au développement de nombreux algorithmes dédiés à l'amélioration de la qualité des images et nécessitant une interprétation en termes de détectabilité.

Dans ce chapitre, nous avons proposé une méthodologie permettant de générer une base d'images TEP simulées au 18F-FDG et bien représentative des acquisitions cliniques. L'ajout de tumeurs de caractéristiques calibrées répond aux cas cliniques et réalistes du lymphome.

La méthodologie basée sur différents choix pour la mise en place de modèles 'patients' réalistes semble efficace. Cependant, quelques améliorations peuvent encore être apportées, notamment du point de vue réalisme du mouvement respiratoire et variabilité inter-patient.

Le chapitre suivant se consacre au développement d'un système CAD dont l'évaluation a initialement mis en évidence le besoin du développement d'une telle base d'images TEP oncologiques simulées.

L'ensemble de ces travaux a été présenté à la conférence *IEEE Medical Imaging Conference 2008* [Tomei et al., 2008] et accepté pour publication dans le journal *IEEE Transactions on Nuclear Science* [Tomei et al., 2009].

Chapitre 5 Proposition d'un système CAD pour l'imagerie TEP corps entier

5.1 Introduction

Ce chapitre propose la méthodologie d'un système CAD pour les images TEP oncologiques 3D corps entier. Ce système doit être capable de détecter des tumeurs de petite taille et de faible contraste de manière à compléter le clinicien dans la difficile tâche d'évaluation d'images TEP pathologiques.

Le Chapitre 2 a présenté les étapes standardisées d'un CAD supervisé et les principales méthodes proposées dans la littérature pour chacune d'entre elles. Un système CAD vise à extraire des informations pertinentes des images, ou cas traités, de manière à les envoyer à un classifieur supervisé. Celui-ci permet d'extraire une cartographie des tumeurs candidates contenant aussi un nombre négligeable de faux positifs. Une étape de réduction de ces mauvaises détections permet de nettoyer ces cartes tout en conservant une bonne sensibilité.

Dans le Chapitre 3, nous avons décrit et justifié les choix méthodologiques propres au système CAD proposé dans ce manuscrit. Nous avons justifié le choix des méthodes que nous avons décidé de retenir pour chacune des étapes et qui sont résumées sur le schéma bloc de la Figure 5-1. Dans un premier temps, un ensemble de caractéristiques de type ondelettes est extrait des images TEP originales. Celles-ci sont envoyées à deux classifieurs en parallèle, le LDA et les SVM. L'extraction initiale des candidats résulte en deux cartes binaires de détection qui sont ensuite traitées de manière à éliminer les potentielles fausses détections. Les deux nouvelles cartes ainsi nettoyées sont ensuite combinées. L'utilisation de deux classifieurs et l'approche par combinaison de classifieurs sont proposées comme une alternative au système de classification unique (LDA ou SVM) dont les performances individuelles sont trop faibles. L'utilisation de la complémentarité du LDA et du SVM de théorie différente nous a donc paru intéressante pour remédier à ces fausses détections trop importantes.

La suite du chapitre est présentée comme suit. Le paragraphe 5.2 présente la mise en place du système CAD constituée pour chaque classifieur, 1) d'une étape d'extraction des tumeurs initialement candidates, 2) d'un nettoyage des faux positifs et 3) d'une combinaison des cartes de détection des LDA et SVM. L'étape d'optimisation des paramètres intrinsèques et extrinsèques des classifieurs supervisés est ensuite définie au paragraphe 5.3. Dans ce chapitre, la stratégie d'évaluation du système CAD final basée sur une méthode de partitionnement des images par hold-out (cf. 2.5.4.2) est enfin exposée au paragraphe 5.4, tandis que le paragraphe 5.5 illustre les résultats obtenus à partir de la base d'images simulées simplifiée.

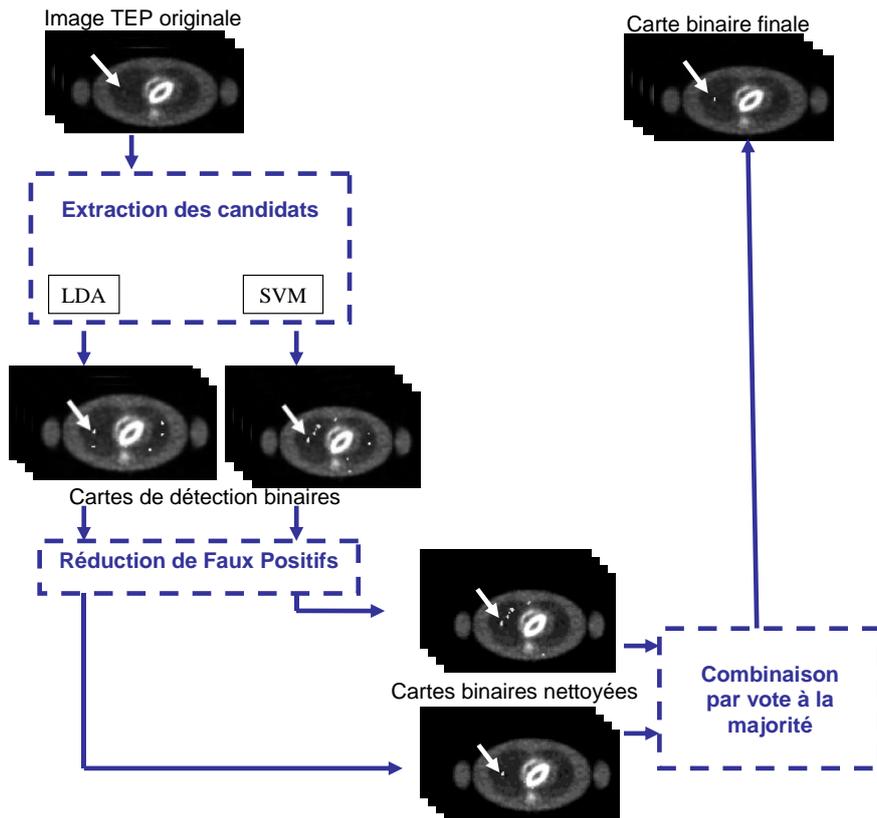


Figure 5-1 : Schéma bloc reprenant les différentes étapes du système CAD développé

5.2 Description du système CAD

Cette section expose les méthodologies sélectionnées pour la mise en place du système d'aide à la détection. Deux étapes sont précisées : 1) l'extraction initiale des tumeurs candidates pour une image au travers d'une carte binaire de détection et 2) le nettoyage de cette carte.

5.2.1 Extraction initiale des tumeurs candidates

La méthodologie suivie pour l'extraction initiale des tumeurs candidates est décrite sur la Figure 5-2. Les caractéristiques des images sont dans un premier temps extraites sous la forme d'un vecteur descriptif V . Ce vecteur descriptif est ensuite envoyé au modèle prédictif du classifieur supervisé, obtenu lors d'un apprentissage sur le même type de caractéristiques de classe connue. La carte de score obtenue en sortie du classifieur est ensuite seuillée afin d'obtenir la carte binaire contenant les localisations des tumeurs initialement candidates.

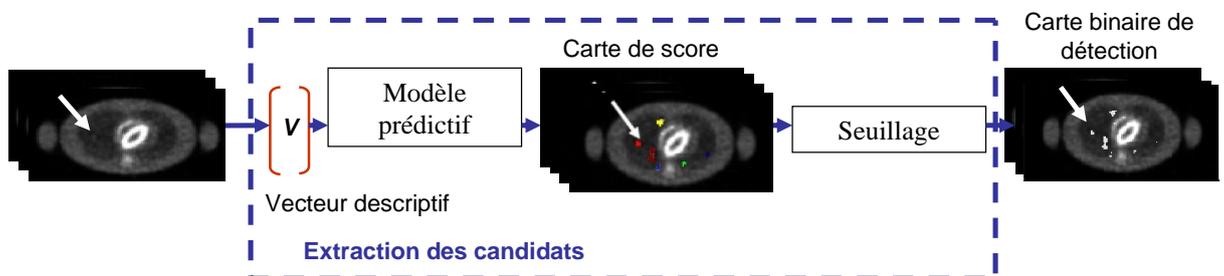


Figure 5-2 : Schéma descriptif de l'extraction initiale des tumeurs candidates

5.2.1.1 Choix des caractéristiques des images

Les caractéristiques descriptives des images d'intérêt sont assemblées sous la forme d'un vecteur V . Ce vecteur V est calculé en chaque voxel d'intérêt de l'image. Comme nous l'avons vu au

paragraphe 2.4.3.2, les caractéristiques descriptives sont l'interprétation de l'image d'un point de vue numérique et leur choix est primordial.

Comme nous l'avons expliqué dans le Chapitre 3, nous avons choisi de nous intéresser aux caractéristiques fréquentielles qui permettent de synthétiser le contenu fréquentiel de l'image. De plus, l'extraction d'informations fréquentielles et plus précisément d'ondelettes permet de reproduire le système de perception humain traitant l'information à travers des canaux sélectifs en fréquences spatiales [Sachs et al., 1971]. La décomposition d'une image en ondelettes permet d'obtenir des composantes de l'image correctement localisées dans l'espace pour lesquelles les structures ou formes des signaux sont facilement accessibles (localisation dans le domaine fréquentiel). Cette transformation a l'avantage d'être flexible par rapport à la résolution de l'image et robuste à sa qualité variable. Elle permet d'extraire les informations fréquentielles, non seulement de la tumeur, mais aussi du bruit environnant. Celui-ci est ainsi plus facilement décorrélié de la tumeur cible. Ce paragraphe n'a pas pour ambition de détailler de manière exhaustive la théorie des ondelettes. Le lecteur intéressé par la théorie des ondelettes pourra se référer aux articles de [Rioul and Vetterli, 1991; Daubechies, 1992]. Un bref aperçu théorique des ondelettes est aussi donné en Annexe B.

Nous nous focalisons sur l'utilisation d'une décomposition en ondelettes discrètes adaptée à nos images discrètes.

Nous avons choisi d'utiliser une décomposition en ondelettes 3D par banc de filtres comme généralement en traitement d'images discrètes. Dans le cas tridimensionnel, la décomposition par banc de filtre se résume par la Figure 5-3. L'image de départ est filtrée séparément dans les trois directions de l'espace x , y et z 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). Le lien de ces filtres avec les fonctions d'ondelettes est précisé en Annexe B. Ainsi sept images de détails et une image d'approximation 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 le vecteur descripteur V de taille $8 \times j$, correspondent ici à l'ensemble de ces coefficients pour chaque voxel (x, y, z) de l'image ((5-1)).

$$V = [LLL_j, LLH_j, LHL_j, LHH_j, HLL_j, HLH_j, HHL_j, HHH_j], \quad j = \{2, 3, 4, 5\} \quad (5-1)$$

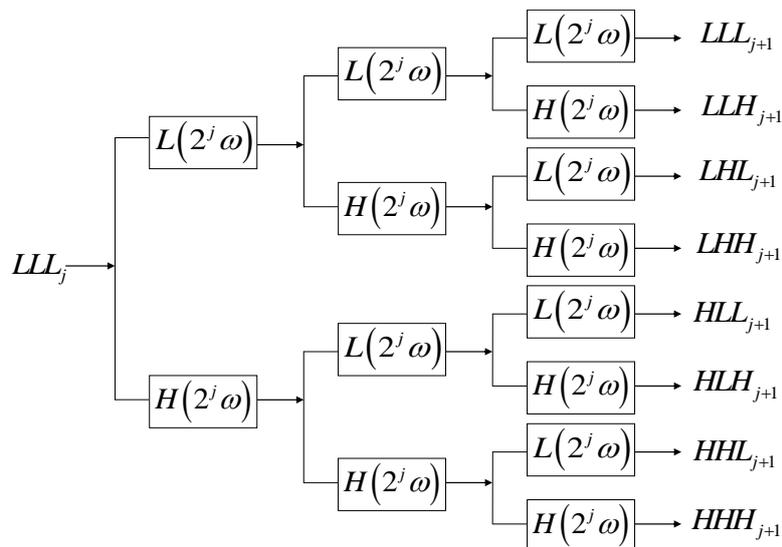


Figure 5-3 : Implémentation par banc de filtres d'une décomposition en ondelettes non décimée

L'utilisation d'une décomposition en ondelettes est optimale lorsque la fonction ondelette est judicieusement choisie. L'idéal est d'utiliser une ondelette dont la forme est proche de celle d'une tumeur à détecter. Cela permet d'obtenir des images de coefficients d'ondelettes ayant leurs maximums localisés sur les tumeurs. Les tumeurs ciblées dans notre problématique de détection difficile correspondent à des signaux approximativement Gaussiens de largeur à mi-hauteur 1 cm

environ dans l'image reconstruite. Elles sont noyées dans un bruit quasi Gaussien affectant l'image. Les caractéristiques de la fonction d'ondelettes conditionnent la qualité de la décomposition en ondelettes :

- l'ondelette doit être symétrique afin de ne pas introduire de distorsion de phase dans la décomposition et par conséquent permettre une meilleure localisation des tumeurs cibles (sans délocalisation),
- l'ondelette doit avoir un petit support, c'est-à-dire un nombre de coefficients du filtre limité afin d'obtenir une meilleure localisation des tumeurs dans le domaine spatial,
- le nombre de moments nuls de l'ondelette doit être plutôt élevé afin d'obtenir une meilleure décorrélation du bruit des tumeurs cibles dans le domaine des ondelettes.

Les deux dernières caractéristiques ne peuvent malheureusement pas être simultanément optimisées et répondent à un compromis.

La plupart des études de CAD en imagerie médicale basées sur des caractéristiques discriminantes fréquentielles [Strickland and Hahn, 1996; Strickland and Hahn, 1997; Yu, 2000; Riyahi-Alam et al., 2004; Regentova et al., 2007] ont exploité les ondelettes biorthogonale 4.4 (Bior4.4) [Cohen et al., 1992] et la Coiflet2 (Coif2) [Daubechies, 1992]. Le Tableau 5-1 reprend les caractéristiques précédemment citées pour ces deux ondelettes tandis que la Figure 5-4 illustre l'allure de leur filtre passe haut et passe bas correspondants.

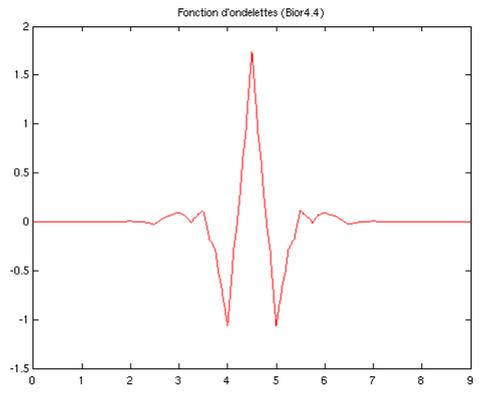
Tableau 5-1 : Principales caractéristiques des ondelettes Bior4.4, Coif2, Daubechies

	Orthogonale	Biorthogonale	Symétrique	Taille du filtre Passe-Bas	Taille du filtre Passe-haut	Nombre de moments
Bior4.4	non	oui	oui	9	7	4
Coif2	oui	oui	presque	12	12	3

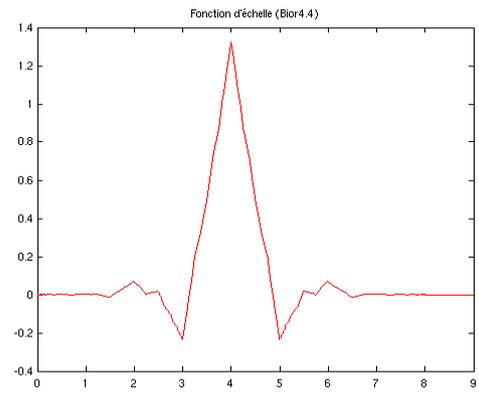
L'ondelette Bior4.4 présente des caractéristiques permettant un bon compromis entre facilité de localisation et décorrélation des signaux. La forme de l'ondelette semble de plus adaptée à la problématique de la thèse d'autant plus que sa forme rappelle celle des tumeurs cibles en imagerie TEP oncologique. Remarquons que la Coif2 a aussi été testée lors d'une étude préliminaire, mais les résultats ne sont pas apparus concluants en termes de faux positifs. En considérant l'utilisation d'une ondelette adaptée, les images de détails doivent présenter leurs extrema au niveau des frontières des tumeurs (maxima) et au centre des tumeurs (minima) et les images d'approximation doivent les présenter au centre des tumeurs.

Cette tendance est nettement observée sur la Figure 5-5 qui présente des coupes d'images de détail (b-i) et d'approximations (j) de la décomposition d'une image TEP simulée d'oncoPET_DB présentée au Chapitre 3. Celle-ci contient deux tumeurs dans les poumons, une dans le foie et une dans la rate.

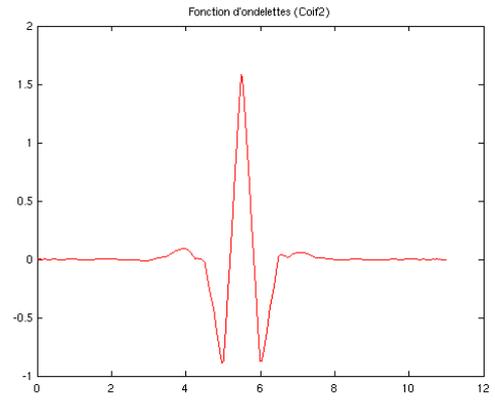
a)



b)



c)



d)

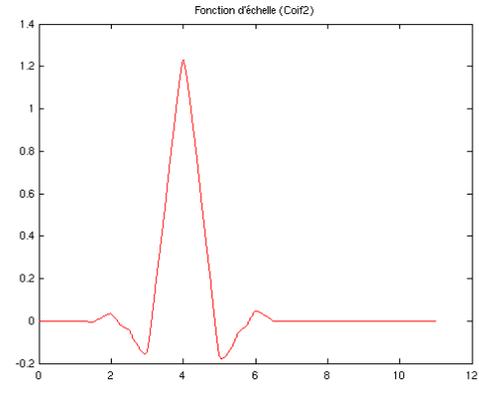


Figure 5-4 : Fonction d'ondelettes et d'échelle pour la Bior4.4 (a, b) et la Coif2 (c, d)

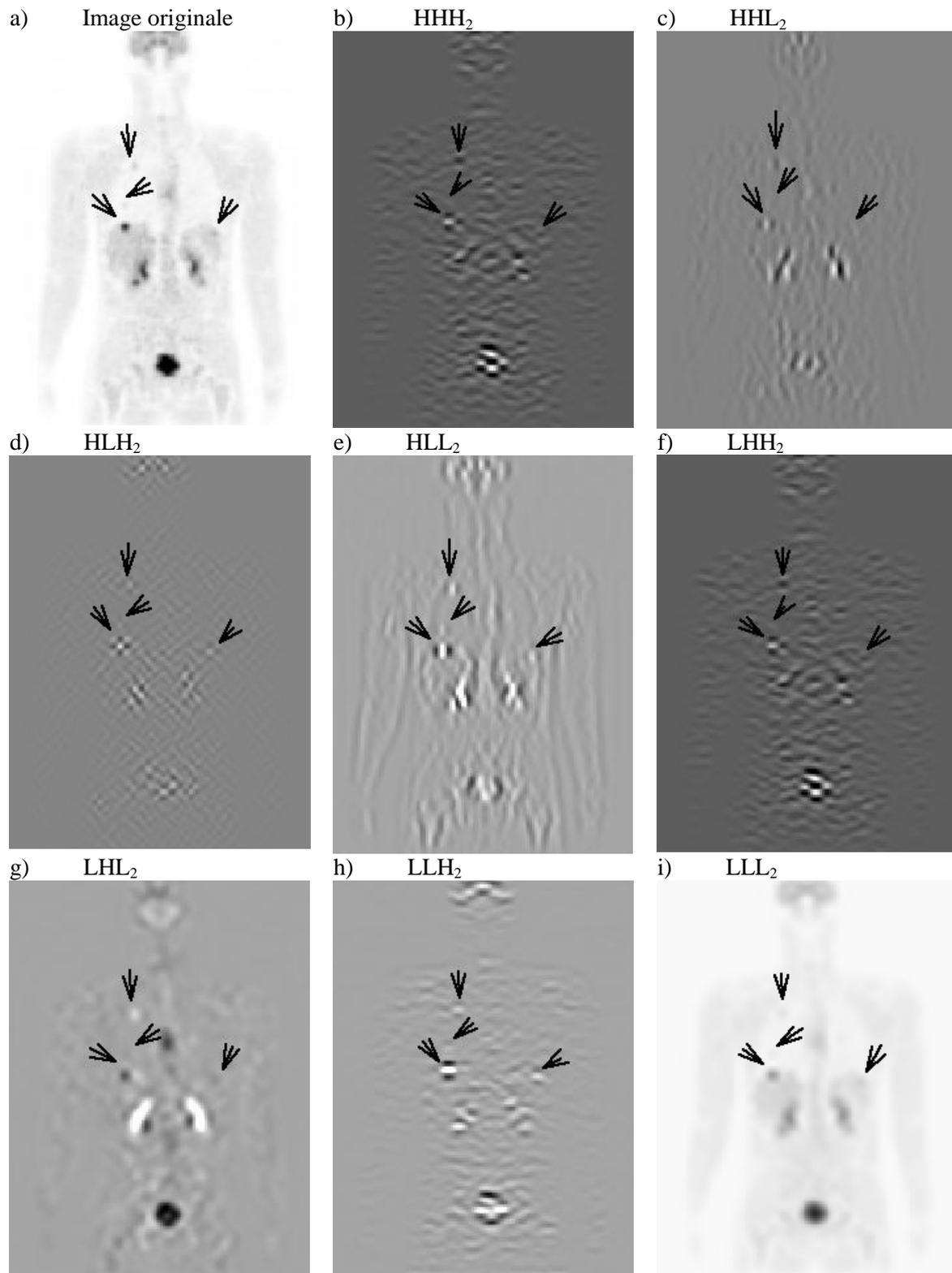


Figure 5-5 : Coupes coronales d'images de décomposition en ondelettes pour le niveau $j=2$ (b-i) obtenue à partir d'une image originale (a) provenant de la base d'images simulées oncoPET_DB.

5.2.1.2 Choix des classifieurs

Introduction

Cette thèse se positionne dans une problématique de classification bi-classe, où H_0 est la classe « tumeur absente » et H_1 la classe « tumeur présente ». Un classifieur supervisé a pour rôle de classer des observations inconnues dans une de ces deux classes après un apprentissage sur des observations d'étiquette connue. Une liste des principaux classifieurs supervisés utilisés dans la littérature a été donnée dans l'état de l'art (Chapitre 2). Comme expliqué en introduction et au Chapitre 3, nous avons choisi de tester et d'évaluer deux classifieurs supervisés, le LDA et le SVM. Ces deux classifieurs, se basant sur des approches différentes, sont ainsi exploités de manière complémentaire afin d'améliorer leurs résultats individuels

Les paragraphes suivants décrivent de manière plus détaillée ces deux classifieurs en présentant le calcul des scores qu'ils attribuent à chaque observation (comme rappelé sur la Figure 5-2) et leurs paramètres intrinsèques.

L'analyse discriminante linéaire (LDA) et ses dérivées

La première méthode de classification utilisée pour le développement du système CAD repose sur la théorie de l'analyse discriminante linéaire (ou linear discriminant analysis, *LDA* en anglais). Nous nous plaçons dans un problème de classification bi-classe avec H_0 , la classe dite 'tumeur absente' et H_1 , la classe dite 'tumeur présente'. Le vecteur descriptif V contient les caractéristiques représentatives de l'observation traitée.

1) Le LDA classique

Le processus de classification ne peut être parfait et implique forcément des erreurs comme le décrit la théorie de Vapnik et Cortes définie au Chapitre 2.4.3.3. L'incertitude de la classification, aussi appelée risque empirique R_{emp} de mauvaises classifications, est modélisée par des probabilités d'appartenance à une des deux classes du problème. Comme souligné, tous les classifieurs supervisés sont liés à la théorie de Vapnik et Cortes. Nous montrons dans ce paragraphe le lien entre le LDA et la théorie de VC. Plus la probabilité d'une classe est grande, plus il y a de chance que le classement de l'observation dans cette classe soit correct au vu de son vecteur descripteur V . La recherche de la probabilité a posteriori $P(H_i/V)$, avec $i = \{0,1\}$, maximale définit l'étiquetage de l'observation décrite par V dans une des deux classes i . Ainsi, l'expression du problème de classification passe par la formule de Bayes qui s'énonce en (5-2).

$$P(H_i/V) = \frac{P(V/H_i) \cdot P(H_i)}{P(V)} \text{ avec } i = \{0,1\} \quad (5-2)$$

$P(V/H_i)$ est la probabilité conditionnelle ou vraisemblance de la classe i pour le vecteur V , c'est-à-dire la proportion des observations dans la classe H_i correspondant à V , $P(H_i)$ est la probabilité a priori de la classe i et $P(V)$ est la probabilité inconditionnelle de V .

On constate que quelque soit la classe d'appartenance ($i = \{0,1\}$), le dénominateur du théorème de Bayes reste identique et égal à $P(V)$. Il n'est donc pas le facteur intervenant dans la classification du vecteur V .

La probabilité a priori $P(H_i)$ et la probabilité de vraisemblance $P(V/H_i)$ pour la classe i doivent être définies.

Lorsqu'aucune connaissance a priori sur les classes n'est disponible, on choisit les probabilités $P(H_i)$ égales pour $i = \{0,1\}$, ce qui correspond à un nombre identique d'observations disponibles dans chaque classe pour la définition de la règle de classification, c'est-à-dire l'apprentissage du LDA.

Il reste donc à définir $P(V/H_i)$ afin de pouvoir calculer la probabilité a posteriori $P(H_i/V)$ et trouver pour quelle valeur de $i = \{0,1\}$, elle est maximale.

L'intérêt du LDA est de permettre le calcul de cette probabilité $P(V/H_i)$ en postulant des classes d'appartenance de distributions multinormales de moyenne μ_i et d'écart type σ_i . Les caractéristiques μ_i et σ_i des distributions de chaque classe sont estimées par la formule (5-3) à partir d'un ensemble de n observations d'apprentissage dont la classe i est connue.

$$\mu_i = \frac{1}{n} \sum_{j=1}^n V_{j/i} \text{ et } \sigma_i = \frac{1}{n-1} \sum_{j=1}^n (V_{j/i} - \mu_{j/i})(V_{j/i} - \mu_{j/i})^T, \quad (5-3)$$

Les moyennes μ_i et variances σ_i ainsi estimées, l'équation (5-2) est utilisée pour connaître les probabilités d'affectation de l'observation décrite par V à chacune des deux classes. La classe optimale est celle de probabilité a posteriori maximale soit celle correspondant à $\arg \max(P(H_i/V))$.

2) Interprétation géométrique

L'analyse discriminante linéaire (LDA) peut aussi être interprétée de manière géométrique. On considère dans un premier temps que le vecteur descripteur V est bidimensionnel (c'est-à-dire ne possédant que deux caractéristiques descriptives). Le vecteur V est donc un point que l'on place dans un plan et on suppose que tous les points d'une même classe sont situés dans une même région de l'espace. Les connaissances a priori sur les classes sont considérées comme nulles, de telle sorte qu'elles sont considérées comme équiprobables ($P(H_i) = \frac{1}{2}$). On cherche la surface séparatrice entre ces deux classes. Celle-ci est définie par les points dont la probabilité d'appartenance à une classe est identique à celle de l'autre classe. Cela se traduit en équation par l'égalité des probabilités (5-4).

$$P(V/H_1) = P(V/H_0) \quad (5-4)$$

Sachant qu'on postule des distributions normales pour les deux classes, on peut alors remplacer les termes de l'égalité (5-4) par la formule d'une loi Gaussienne (5-5).

$$P(V/H_i) = \frac{1}{2\pi\sigma_i} \exp^{-\frac{1}{2}(V-\mu_i)^T \sigma_i^{-1} (V-\mu_i)} \text{ avec } i = \{0,1\} \quad (5-5)$$

Ce qui donne l'égalité (5-6).

$$\frac{1}{2\pi\sigma_0} \exp^{-\frac{1}{2}(V-\mu_0)^T \sigma_0^{-1} (V-\mu_0)} = \frac{1}{2\pi\sigma_1} \exp^{-\frac{1}{2}(V-\mu_1)^T \sigma_1^{-1} (V-\mu_1)} \quad (5-6)$$

En passant au log, on obtient l'égalité (5-7).

$$\log\left(\frac{\sigma_1}{\sigma_0}\right) - \frac{1}{2} V^T (\sigma_0^{-1} - \sigma_1^{-1}) V + V^T (\sigma_0^{-1} \mu_0 - \sigma_1^{-1} \mu_1) - \frac{1}{2} (\mu_0^T \sigma_0^{-1} \mu_0 - \mu_1^T \sigma_1^{-1} \mu_1) = 0 \quad (5-7)$$

Dans le cas où les distributions de deux classes sont de même variance $\sigma = \sigma_0 = \sigma_1$, les simplifications de l'égalité (5-7) donnent (5-8).

$$V^T \sigma^{-1} (\mu_0 - \mu_1) - \frac{1}{2} \sigma^{-1} (\mu_0^T \mu_0 - \mu_1^T \mu_1) = 0 \quad (5-8)$$

Cette égalité peut aussi s'exprimer par la formule (5-9) en développant l'égalité (5-8)

$$V^T \sigma^{-1} (\mu_0 - \mu_1) - \frac{1}{2} (\mu_0 + \mu_1)^T \sigma^{-1} (\mu_0 - \mu_1) = 0 \quad (5-9)$$

Cette expression se présente sous la forme d'une équation de droite ou d'un hyperplan lorsque le vecteur V est multidimensionnel. La position du plan entre les deux distributions des observations des deux classes est définie par la constante $c = \frac{1}{2} (\mu_0 + \mu_1)^T \sigma^{-1} (\mu_0 - \mu_1)$.

On peut transformer l'équation (5-9) en l'inégalité (5-10) qui peut également être traduite de manière géométrique.

$$V^T \sigma^{-1} (\mu_0 - \mu_1) > c \quad (5-10)$$

La classe d'appartenance d'une nouvelle observation est fonction de la position de la projection de son vecteur de caractéristique V dans la direction du vecteur $\omega = \sigma^{-1} (\mu_0 - \mu_1)$, défini par apprentissage, par rapport à l'hyperplan de séparation. ω est la normale à cet hyperplan. La condition d'appartenance de l'observation à une des deux classes dépend de la position de son vecteur descripteur V d'un côté ou de l'autre de cet hyperplan perpendiculaire à ω .

On se place maintenant dans un espace de dimension supérieure à 2. On note par la suite λ le résultat numérique de la projection du vecteur descriptif sur l'hyperplan ω qui quantifie également la distance du vecteur descripteur projeté à l'hyperplan optimal.

La Figure 5-6 présente la méthode de classification par LDA. Le projecteur optimal ω définit l'hyperplan séparateur optimal entre les deux classes H0 (cercles rouges) et H1 (cercles bleus).

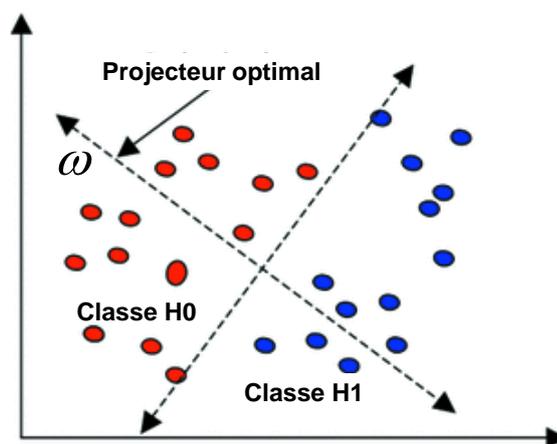


Figure 5-6 : Illustration de la méthode de classification par LDA reposant sur le calcul d'un projecteur optimal des classes

On peut remarquer que cette approche n'est pas sans rappeler celle de l'analyse en composantes principales (ACP) décrite au paragraphe 2.4.4.2. La différence ici repose sur la prise en compte de l'aspect classification dans la projection des données.

2) L'analyse discriminante de Fisher et Hotelling

Dans l'article de Barrett [Barrett et al., 1993], deux théories de classification très proches du LDA classique sont résumées et apparaissent très liées : l'analyse de Hotelling et l'analyse de Fisher. L'analyse discriminante de Hotelling (ou Hotelling-LDA) [Hotelling, 1931] est généralement plutôt décrite comme un observateur numérique mimant le comportement humain face à un problème de classification de signal connu dans un fond connu. C'est pourquoi il est plus fréquent d'utiliser l'adaptation de l'analyse discriminante de Fisher aussi notée Fisher-LDA [Fisher, 1936]. Les termes Fisher-LDA ou Hotelling-LDA et LDA sont généralement interchangeables.

Ces deux adaptations du LDA classique ont également pour objectif de trouver la classe optimale pour chaque nouvelle observation de manière à minimiser le risque empirique défini par Vapnik et Cortes. Pourtant le classifieur de Fisher et de Hotelling repose sur une logique quelque peu différente. D'une part, aucune hypothèse n'est faite sur les densités de probabilité des deux classes H_0 et H_1 (multinormales et égalité des variances). D'autre part, le tracé de l'hyperplan séparateur optimal ne repose plus sur l'égalité des probabilités a posteriori d'appartenance aux deux classes définie en (5-4), mais plutôt sur l'éloignement des moyennes de ces densités de probabilité. Ainsi l'optimisation de la séparation des classes repose sur la maximisation du critère de Fisher donné en (5-11).

$$J(\omega) = \frac{(m_1 - m_0)^2}{s_1^2 + s_0^2} \quad (5-11)$$

Avec $m_i = \omega^T \mu_i$ la moyenne projetée sur la normale à la surface séparatrice optimale et s_i la variance projetée, de chaque classe $i = \{0,1\}$. Le critère $J(\omega)$ doit être maximisé selon le projecteur ω afin d'optimiser la séparation des deux classes. Minimiser le coût $J(\omega)$ revient à minimiser le coût $Q(\omega)$ défini par Vapnik et Cortes au paragraphe 2.4.3.3, et qui permet d'obtenir une classe estimée de toute nouvelle observation correcte. L'expression de $J(\omega)$ peut être simplifiée si on reprend chacun des termes du rapport.

En reprenant la formulation de la variance non projetée σ_i pour la classe i (5-3), sa projection s_i sur ω est définie par la formule (5-12).

$$s_i^2 = \sum_V (\omega^T V - \omega^T \mu_i)^2 = \sum_V \omega^T (V - \mu_i)(V - \mu_i)^T \omega \quad (5-12)$$

Considérons alors la matrice de dispersion intra-classe S_w définie en (5-13) et précisant la variabilité dans chacune des deux classes,

$$S_w = \sum_V (V - \mu_0)(V - \mu_0)^T + \sum_V (V - \mu_1)(V - \mu_1)^T \quad (5-13)$$

On peut intégrer cette matrice S_w dans l'équation (5-12) ce qui nous donne l'égalité (5-14) correspondant au dénominateur de l'équation (5-11).

$$s_0^2 + s_1^2 = \omega^T S_w \omega \quad (5-14)$$

De la même manière, on obtient (5-16) en exprimant le numérateur de $J(\omega)$ avec la matrice de dispersion inter-classe S_B définie par l'équation (5-15).

$$S_B = (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T \quad (5-15)$$

$$(\tilde{\mu}_1 - \tilde{\mu}_0)^2 = (\omega^T \mu_1 - \omega^T \mu_0)^2 = \omega^T (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T \omega = \omega^T S_B \omega \quad (5-16)$$

En remplaçant chacun des termes de l'expression du rapport $J(\omega)$, l'égalité (5-11) est maintenant donnée par (5-17).

$$J(\omega) = \frac{\omega^T S_B \omega}{\omega^T S_w \omega} \quad (5-17)$$

L'optimisation de la séparation des deux classes passe par la maximisation de $J(\omega)$ par rapport à ω . Cela revient à chercher le maximum de la forme quadratique $\omega^T S_B \omega$ sous la contrainte $\omega^T S_w \omega = 1$. Ce problème est résolu par maximisation du Lagrangien \mathcal{L} (5-18), c'est-à-dire annulation de ses dérivées.

$$\mathcal{L}(\omega) = \omega^T S_B \omega - \alpha (\omega^T S_w \omega - 1) \quad (5-18)$$

Soit,

$$2S_B \omega - \alpha 2S_w \omega = 0 \quad (5-19)$$

Ce qui donne

$$S_B \omega = \alpha S_w \omega \quad (5-20)$$

Dans le cas où S_w n'est pas singulière (donc inversible), le problème s'écrit comme un problème conventionnel de valeur propre.

$$S_w^{-1} S_B \omega = \alpha \omega \quad (5-21)$$

On sait que $S_B \omega$ est toujours dans la direction de $\mu_1 - \mu_0$ car $S_B = (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T$. Il n'est donc pas nécessaire de résoudre le problème pour les valeurs et vecteurs propres de $S_w^{-1} S_B$. On obtient directement la solution donnée par la formule (5-22).

$$\omega = S_w^{-1} (\mu_1 - \mu_0) \quad (5-22)$$

Rappelons que la matrice de dispersion intra-classe S_w est calculée comme la moyenne pondérée des matrices de variance de chaque classe (5-13). On retombe donc sur la même expression du projecteur ω que précédemment à une constante près pour le LDA (5-10) permettant le calcul de la variable λ .

3) Appellations dans le cadre de la thèse

Le LDA classique est considéré comme simple et efficace. Il impose néanmoins des conditions de distributions de probabilité assez strictes qui sont rarement respectées en imagerie médicale.

Les Figure 5-7 et Figure 5-8 présentent les distributions de répartition des λ pour chacune des classes H0 et H1 et les densités de probabilités estimées correspondantes. Les valeurs des scores λ , montrées sur ces figures, ont été obtenues par application du LDA sur des vecteurs descripteurs de localisations saines et pathologiques extraites de la base d'images simulées simplifiée. En d'autres

termes, nous présentons la répartition condensée de la distribution des vecteurs V dans l'espace d'origine sur un axe unique de projection. Ces projections sont notées λ et sont définies par l'équation (5-10). Ces répartitions permettent ainsi d'extraire des conclusions quant à l'allure des distributions de l'espace d'origine.

Le vecteur de caractéristiques correspond aux niveaux de décomposition en ondelettes extrêmes $j=2$ (Figure 5-7) et $j=5$ (Figure 5-8) pour trois organes, les poumons, le foie et les tissus mous.

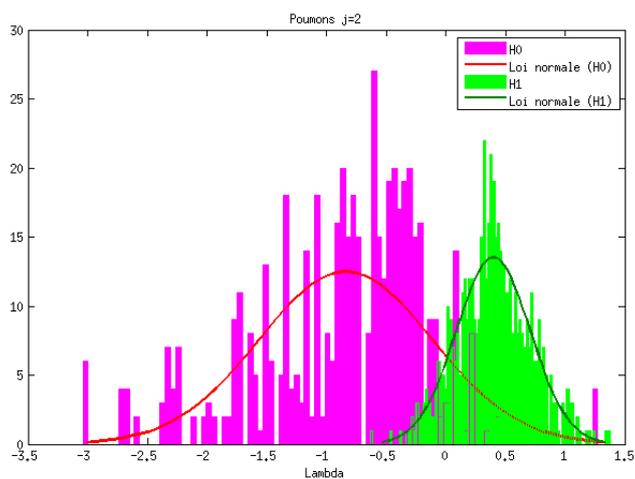
Les distributions de probabilité du foie s'éloignent d'une loi normale notamment pour la classe H_0 . Le Tableau 5-2 présente les variances calculées sur les scores λ des deux classes illustrés sur les Figure 5-7 et Figure 5-8. Dans le cas idéal, celles-ci doivent être identiques pour les deux classes. Les variances pour les deux classes sont assez proches pour $j=2$ et $j=5$ à l'exception des tissus mous qui présentent des valeurs très éloignées. Que ce soit par la forme des distributions de probabilité ou par la valeur des covariances de chaque classe, les hypothèses de base du LDA semblent parfois difficilement respectées pour notre problématique et nos images cibles.

Tableau 5-2 : Variances correspondant aux distributions de λ extraites du LDA sur la base de données simulées simplifiée pour $j=2$ et $j=5$ et chacune des classes.

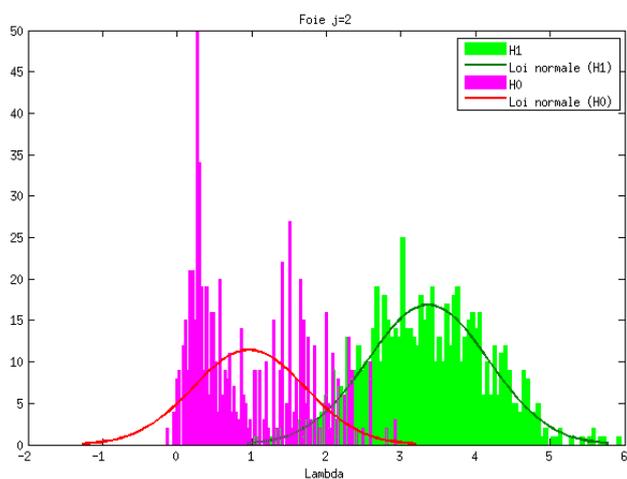
Classe		H0	H1
j=2	Poumons	0,5222	0,7862
	Foie	0,5546	0,6472
	Tissus mous	0,4106	0,0963
j=5	Poumons	0,7289	0,9891
	Foie	0,6195	0,9111
	Tissus mous	0,7386	0,1975

Le LDA, et notamment la version de Fisher ou Hotelling, est cependant connu pour être assez robuste dans le cas où ses hypothèses de base ne sont pas respectées. Nous utilisons par la suite l'appellation LDA pour référer au Fisher-LDA comme généralement dans la littérature dans des problèmes de classification [Jesneck et al., 2007; Hamdi et al., 2008].

a)



b)



c)

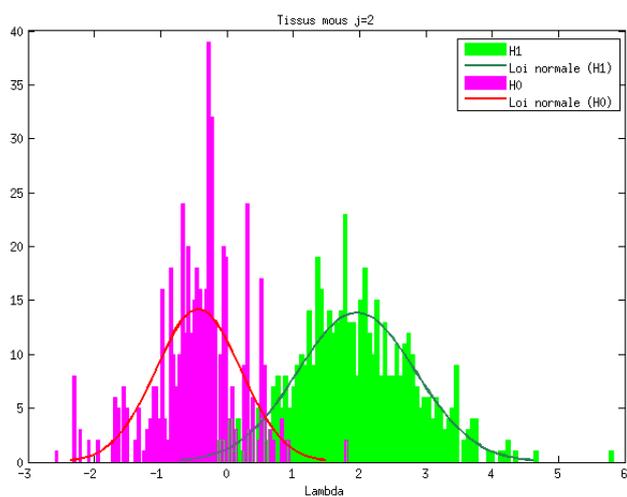
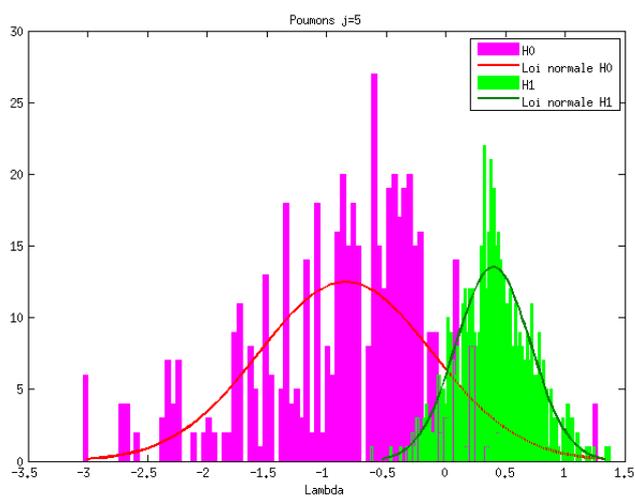
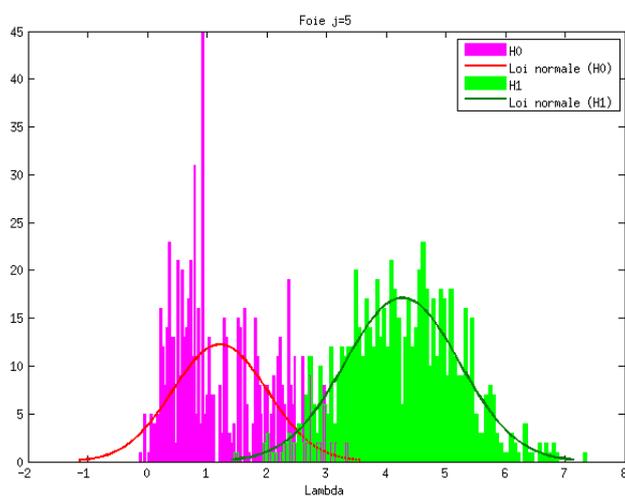


Figure 5-7 : Histogramme de répartition des projetés λ des vecteur V initiaux pour la classe H0 (rose) et H1 (vert) pour les poumons (a), le foie (b) et les tissus mous (c) avec un modèle prédictif de LDA calculé pour un niveau de décomposition $j=2$. Les courbes (trait plein) affichées correspondent aux distributions normales adaptées à chaque histogramme.

a)



b)



c)

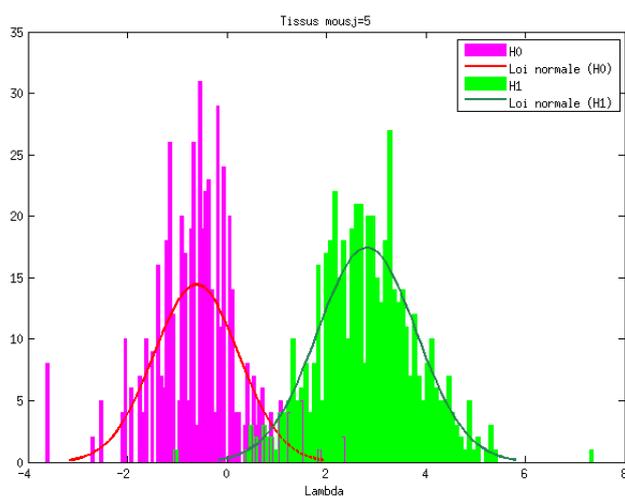


Figure 5-8 : Histogramme de répartition des pour la classe H0 (rose) et H1 (vert) pour les poumons (a), le foie (b) et les tissus mous (c) avec un modèle prédictif de LDA calculé pour un niveau de décomposition $j=5$. Les courbes (trait plein) affichées correspondent aux distributions normales adaptées à chaque histogramme.

Les Séparateurs à vaste marge (SVM)

Les SVM ont été développés dans les années 90 et reposent sur la robuste théorie de Vapnik et Cortes [Vapnik, 1995] et la minimisation du risque structurel. Ils ont rapidement été adoptés pour leur capacité à s'adapter à des données de grandes dimensions (nombre de caractéristiques important), une base d'apprentissage réduite, le fait qu'ils soient bien fondés théoriquement, et leurs bons résultats en pratique. Les SVM reposent sur deux idées clés existantes depuis plusieurs années : la notion de marge maximale [Vapnik and Lerner, 1963; Duda and Hart, 1973] et la notion de fonctions à noyaux [Mercer, 1909].

Considérons le même ensemble d'exemples d'apprentissage que précédemment $(V_i, y_i), V_i \in \mathcal{R}^n, y_i \in (-1, +1)$ sont les labels des deux classes et $i=(1, 2, ..l)$. Si les deux classes sont linéairement séparables alors il existe un hyperplan défini par $\omega \cdot V + b = 0$ séparant chaque V_i de telle sorte que tous les points d'une même classe se retrouvent du même côté de l'hyperplan. La fonction de décision λ est donnée par l'équation (5-23).

$$\lambda_{\omega, b} = \text{sign}((\omega \cdot V) + b) \quad (5-23)$$

Dans le cas linéairement séparable, l'hyperplan de séparation optimal est trouvé par minimisation du risque théorique régularisé de Vapnik et Cortes (VC) qui quantifie la complexité de l'équation de la surface séparatrice définie par $\omega \cdot V + b = 0$.

Parmi l'ensemble des surfaces satisfaisants ces conditions, les SVM cherchent l'hyperplan qui maximise sa distance aux points les plus proches de chaque classe.

La distance de l'hyperplan à un point A de la classe H1 est définie par $\frac{\omega \cdot V_A + b}{\|\omega\|}$, et pour un point B

de la classe H0 par $\frac{\omega \cdot V_B + b}{\|\omega\|}$. La marge correspond à la distance minimale m entre les deux classes selon l'équation (5-24).

$$m(\omega, b) = \min_{x_i/y_i=-1} \frac{\omega \cdot V_i + b}{\|\omega\|} + \min_{x_i/y_i=1} \frac{\omega \cdot V_i + b}{\|\omega\|} = \frac{1}{\|\omega\|} \left(\min_{x_i/y_i=-1} (\omega \cdot V_i + b) + \min_{x_i/y_i=1} (\omega \cdot V_i + b) \right) \quad (5-24)$$

On considère la surface séparatrice comme un hyperplan canonique de manière à normaliser le problème d'optimiser. On impose ainsi $\min_{x_i/y_i=-1} (\omega \cdot V_i + b) = 1$ et la distance m entre les deux classes est

définie par $\frac{2}{\|\omega\|}$. Cette marge m doit être maximisée afin d'avoir les deux classes les plus séparées possible.

Ainsi le problème global équivaut à un problème de minimisation sous contrainte défini par (5-25).

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 \\ y_i ((\omega \cdot V_i) + b) \geq 1, i = (1, 2, ..l) \end{cases} \quad (5-25)$$

Ce problème peut être résolu par la minimisation du Lagrangien (5-26) par rapport à ω et b et sa maximisation par rapport aux coefficients de Lagrange α_i .

$$La(\omega, b, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^l \alpha_i (y_i ((\omega \cdot V_i) + b) - 1) \quad (5-26)$$

La solution est donnée par l'équation (5-27). Seuls les α_i non nuls sont ceux correspondants aux vecteurs V critiques, situés sur la marge. Ils sont les plus proches de l'hyperplan optimal de séparation des deux classes.

$$\begin{cases} \omega = \sum_{i=1}^l (\alpha_i (y_i \cdot V_i)) \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases} \quad \text{avec } \alpha_i \geq 0 \quad (5-27)$$

La classification de toute nouvelle observation V_i en une des deux classes H1 et H0 est donnée par sa position par rapport à l'hyperplan optimal défini par ω et b . La solution binaire $\lambda_{\omega,b}$ du problème de classification s'écrit donc (5-28).

$$\lambda_{\omega,b} = \text{sign} \left(\sum_{i=1}^l \alpha_i y_i (V \cdot V_i) + b \right) \quad (5-28)$$

Les seuls α_i qui peuvent être non nuls dans la solution donnée par l'équation (5-28) sont ceux pour lesquels la contrainte (5-25) correspond à une égalité. Le terme *vecteurs de support* leur est appliqué car ils sont les seuls nécessaires pour la détermination de la marge. La classification d'un nouveau vecteur de caractéristiques V dépend donc maintenant simplement du signe de $\omega \cdot V + b$ considérant les valeurs de ω et b . La sortie des SVM peut être binaire (classe H0 ou H1) ou peut contenir la valeur du score λ correspondant au degré de ressemblance du vecteur descriptif V avec celui caractéristique d'une tumeur cible. Ce score λ , non borné, est défini comme la distance entre le vecteur de caractéristiques V et l'hyperplan optimal de séparation.

Dans le cas où les données ne sont pas linéairement séparables à cause d'un recouvrement partiel des deux classes, c'est à dire dans beaucoup de cas réels, l variables positives de relaxation ξ_i , $i = (1, 2, \dots, l)$, sont introduites. Elles réduisent les contraintes sur la marge optimale en autorisant mais pénalisant les mauvaises classifications. Le problème d'optimisation quadratique (5-25) est donc formulé par l'équation (5-29).

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ y_i ((\omega_i \cdot V_i) + b) \geq 1 - \xi_i \end{cases} \quad (5-29)$$

La variable C quantifie le compromis entre la minimisation des erreurs d'apprentissage et la minimisation de la complexité du modèle. La solution du nouveau problème d'optimisation est similaire au cas séparable excepté pour la contrainte sur α_i (5-27) qui devient $C \geq \alpha_i \geq 0$. Cette variable doit généralement être optimisée lors de la phase d'apprentissage du classifieur.

Dans le cas totalement non linéairement séparable, les vecteurs de caractéristiques d'entrée sont projetés dans un espace de dimension supérieure par l'intermédiaire d'une fonction noyau. Cette opération se traduit par la substitution du produit scalaire dans les formules précédentes par $\Phi(V, V_i)$. Les conditions de Mercer [Mercer, 1909] doivent être respectées pour correctement définir une fonction noyau. C'est le cas pour les fonctions à base radiale Gaussiennes, ou *Gaussian Radial Basis Function* (RBF) en anglais, définie par l'équation (5-30).

$$\Phi(V, V_i) = \exp \left(\frac{-\|V - V_i\|^2}{2\sigma^2} \right) \quad (5-30)$$

L'écart type σ de la fonction noyau est fixé lors de la phase d'apprentissage, conjointement au

paramètre de pénalisation C . Les performances des SVM sont particulièrement dépendantes de ces choix qui régissent l'efficacité du modèle de classification.

La Figure 5-9 présente un schéma récapitulatif de la méthode de classification par SVM dans le cas d'un problème bi-classe : H0 (carré) et H1 (cercle). Les vecteurs de support du système apparaissent par des symboles pleins. Le mauvais classement d'un cercle dans la classe des carrés est autorisé par la variable de relaxation ξ .

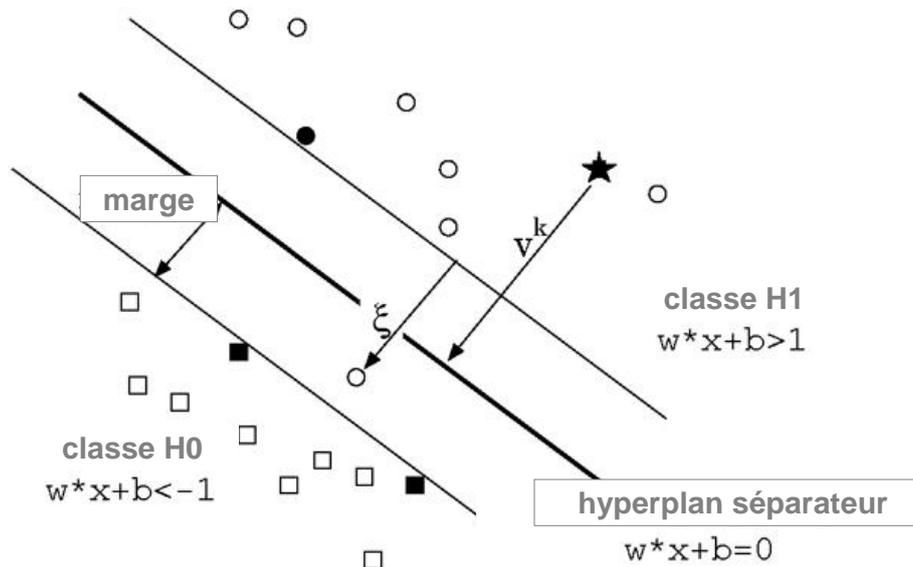


Figure 5-9 : Schéma explicatif de la méthode de classification par SVM basée sur la maximisation de la marge entre les deux classes

5.2.1.3 Seuillage et agrégation des tumeurs candidates

Les sorties des classifieurs LDA et SVM correspondent aux scores, notés λ , obtenus par chaque voxel de l'image 3D. Plus ce score est élevé, plus la probabilité que ce voxel corresponde à une tumeur est grande. Un seuillage de ces cartes 3D de score permet d'extraire des cartes binaires, où 1 représente une tumeur potentiellement présente, et 0 une tumeur potentiellement absente.

Courbes FROC pour le seuillage des cartes de score

Dans cette étude, le seuil des cartes de score est dérivé d'une analyse de courbes FROC [Swenson, 1996] et [Chakraborty and Winter, 1990]. Celles-ci ont été décrites au Chapitre 2 (2.5.3). Les analyses FROC ont été mises en place pour permettre l'évaluation des performances de détection d'un observateur humain ou numérique dans une tâche plus réaliste que celle des analyses ROC. Il est en effet possible d'évaluer la précision de la localisation de multiples tumeurs par image. Les courbes FROC permettent ainsi de proposer une évaluation tridimensionnelle des performances de classification en considérant les faux positifs par rapport à l'ensemble de l'image. Elles tracent en effet la sensibilité moyenne pour chaque seuil λ_s de la carte de score λ en fonction du nombre de faux positifs par image (dénoté FPI dans la suite). Une série de 100 valeurs de seuil λ_s est extraite des cartes de score λ provenant de chaque classifieur par un échantillonnage régulier entre les valeurs minimales et maximales. Les cartes seuillées ainsi obtenues sont utilisées pour calculer un couple (sensibilité/FPI). Chaque couple est moyenné pour toutes les cartes de score extraites du classifieur considéré.

Le seuil λ_s est ensuite extrait des courbes FROC de manière à assurer une sensibilité de 90 % quelque soit le taux de FPI. Ce choix est souvent considéré comme suffisant pour assurer la légitimité des systèmes CAD [Bornefalk and Hermansson, 2005]. [Stumple et al., 1998] présente d'ailleurs des performances humaines avoisinant les 90 % de sensibilité pour le lymphome, tandis que [Gould et al., 2001] met en évidence une sensibilité humaine de 97 % pour le cancer du poumon.

Agrégation des voxels candidats

Le comptage des vrais et faux positifs définissant les valeurs nécessaires au tracé des courbes FROC doit être précisé.

Un candidat retenu par classification est défini comme un agrégat en connexité 26 de taille maximum de 125 voxels obtenu par un parcours lexicographique de la carte binaire de détection.

Un VP est un candidat pour lequel au moins un de ces voxels appartient à une tumeur réelle de centre et de diamètre connu. La distance d'acceptation d'un VP, calculée entre le centre d'une tumeur réelle et un voxel d'un cluster candidat (cf. paragraphe 2.5.2), correspond au rayon d'une tumeur cible. Cette définition suit celle proposée par [Paik et al., 2004]. Aussi, si deux voxels disjoints appartiennent à une même tumeur réelle, le nombre de VP n'est augmenté que d'une unité.

Le nombre de tumeurs candidates à tort (FP) est défini comme le nombre réel de mauvaises détections dans toute l'image ou tout le volume d'intérêt (si on ne considère qu'un organe) afin de donner la meilleure représentation possible de l'aspect visuel réel du volume.

5.2.2 Réduction de faux positifs (RFP)

5.2.2.1 Introduction

L'étape précédente permet de détecter les candidats les plus suspects à partir des cartes binaires 3D, mais ne permet pas de discriminer tous les faux positifs des vraies tumeurs. Comme précisé sur la Figure 5-1, l'étape finale permettant de réduire les faux positifs observés sur les cartes binaires provenant des deux classificateurs repose sur la combinaison de ces cartes binaires.

Nous avons cependant évalué l'intérêt d'ajouter une étape de nettoyage préliminaire afin d'obtenir les cartes binaires les plus propres possibles avant cette combinaison. Ce nettoyage, représenté sur la Figure 5-10, est détaillé dans la suite du paragraphe. Il comprend deux étapes principales : l'élimination des candidats monovoxels et la sélection des candidats par leur degré de suspicion.

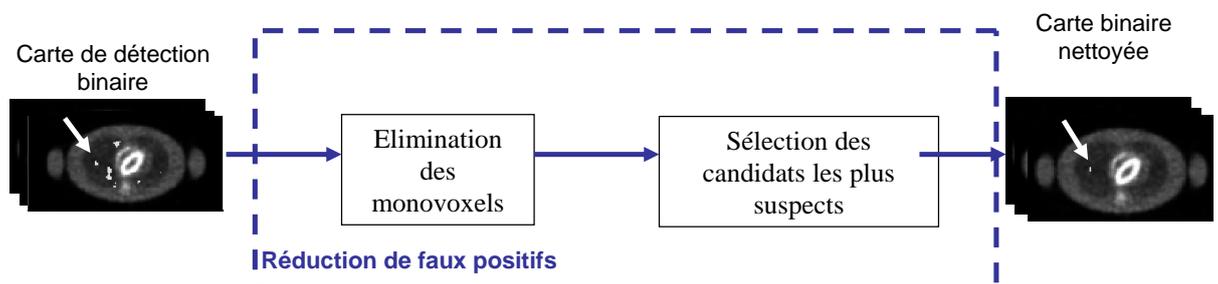


Figure 5-10 : Schéma bloc des étapes de réduction de faux positifs du système CAD

5.2.2.2 Elimination des monovoxels

Une analyse rapide des tumeurs (clusters) extraites lors de l'identification initiale par les classificateurs permet de mettre en évidence une forte quantité de faux positifs monovoxels, c'est-à-dire de clusters de taille un voxel. Ils sont généralement dus au bruit quasi Gaussien affectant les images TEP et qui rend les organes cibles assez granuleux. Ils sont ainsi détectés par les classificateurs comme des tumeurs potentielles qu'ils miment. Cette tendance avait aussi été observée par [Yu and Guan, 2000] lors de l'étude de faisabilité de son système CAD pour la mammographie. L'élimination de ces monovoxels sur les cartes binaires extraites des classificateurs est ainsi évaluée comme première étape de réduction de FPI.

5.2.2.3 Sélection des candidats les plus suspects

Un second nettoyage des cartes binaires est proposé et repose sur la sélection des candidats potentiels les plus suspects parmi tous. Comme expliqué précédemment, l'étape de classification résulte en une carte de score (Figure 5-11a), qui par seuillage, donne une carte binaire des candidats potentiels (Figure 5-11b). La carte de score assigne à chaque voxel une valeur représentative de sa probabilité d'appartenir à une tumeur.

Les deux types de cartes, de score et seuillée, sont combinés afin d'assigner une force (extraite de la carte de score) à chacun des clusters détectés par la carte seuillée. Deux forces sont considérées,

la première correspond à la valeur maximale des scores des voxels constituant le candidat détecté, et la seconde correspond à sa valeur moyenne. Ces forces permettent de quantifier le degré de suspicion de chaque candidat vu comme un amas de voxel, quant au classifieur considéré. Les candidats potentiels sont triés par ordre décroissant de force (Figure 5-11c) et leur rang est exprimé en fonction des forces obtenues par les autres candidats. Le rapport entre le rang du candidat et le nombre de candidats total est calculé. Un rang faible (0 %) correspond à un candidat très suspect (score élevé) et un rang élevé (100 %) à un candidat peu suspect. Une sélection des candidats les plus suspects est réalisée à partir de cette liste ordonnée. Les candidats de rang inférieur à un seuil défini empiriquement à partir des images de validation sont conservés, les autres sont éliminés. Ce rang seuil est défini de manière à maintenir une sensibilité de détection élevée tout en éliminant une large portion des faux positifs initialement détectés (Figure 5-11d).

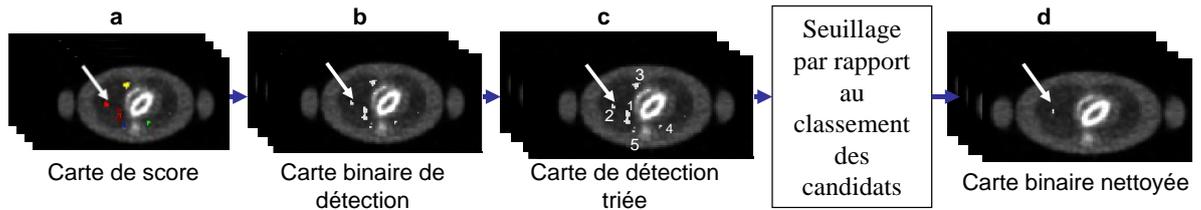


Figure 5-11 : Schéma récapitulatif de la sélection des tumeurs candidates les plus suspectes

5.2.3 Combinaison de classifieurs

[Park et al., 2009] et [Li et al., 2001], dans deux systèmes CAD respectivement dédiés à la mammographie et à l'imagerie TDM, ont montré l'intérêt de combiner deux classifieurs de nature différentes comme méthode efficace de réduction de faux positifs.

Inspirée par ces articles, nous avons évalué l'utilité de la combinaison de type classe par vote à la majorité des deux classifieurs choisis, le LDA et le SVM. Cette combinaison correspond, dans le cas où seulement deux classifieurs sont utilisés, à la multiplication terme à terme des cartes binaires exprimant les classes de chaque voxel. Pour qu'une tumeur candidate soit conservée, elle doit être définie comme telle par les deux classifieurs. Cette combinaison des cartes binaires est optimale lorsque la distribution des VP des deux classifieurs est corrélée et celle des FP décorrélée.

Les deux classifieurs se basent sur une logique différente reposant sur l'utilisation des moments statistiques pour le LDA et sur la maximisation de la marge entre les cas extrêmes des deux classes pour le SVM. Pour une sensibilité fixée de 90 %, il semble logique que la majorité des tumeurs soit correctement détectée par les deux classifieurs. Il semble intéressant de vérifier que la distribution des faux positifs due au bruit et au fond réponde à une répartition spatiale différente selon le modèle du classifieur généré. Ainsi ces FP peuvent disparaître par combinaison.

5.3 Optimisation des paramètres du système CAD

5.3.1 Introduction

L'optimisation du CAD s'applique aux deux étapes principales du système détaillées au paragraphe 5.2 : 1) l'extraction des candidats initiaux, 2) la maximisation des performances de la réduction de faux positifs.

1) L'extraction des candidats à partir des classifieurs

Les paramètres intrinsèques du SVM sont optimisés lors de l'étape d'apprentissage. Ils permettent le calcul du modèle prédictif des SVM. Le calcul du modèle prédictif du LDA ne nécessite pas d'optimisation étant donné qu'il repose sur l'estimation de moments statistiques à partir des données d'apprentissage.

Ensuite, l'extraction des candidats initiaux est optimisée à travers l'étude de deux paramètres :

- les caractéristiques de la base d'apprentissage
- les composantes du vecteur de caractéristiques V

Cette évaluation se base sur les courbes FROC extraites des cartes de score dérivées des images de validation. Elle ne tient donc pas compte des étapes de RFP. Notre optimisation du CAD

repose sur la sélection des jeux de paramètres correspondant au plus faible FPI pour une sensibilité fixée de 90%.

2) La réduction de faux positifs

Les deux méthodes de la réduction de faux positifs décrites au paragraphe 5.2.2 sont évaluées séquentiellement. Elles doivent permettre de conserver une sensibilité assez haute tout en éliminant les mauvaises détections par nettoyage des cartes binaires de détection.

5.3.2 Optimisation des paramètres intrinsèques du SVM

La première étape précédant l'utilisation des SVM consiste à fixer ses paramètres intrinsèques qui sont la déviation standard σ de la fonction noyau et le paramètre de pénalisation C (cf. paragraphe 5.2.1.2). Nous avons examiné deux fonctions noyaux, la fonction polynomiale de degré 2 à 4 et les fonctions à base radiale Gaussienne. Leur formule est donnée par les équations (5-31) et (5-32) où p correspond au degré de la fonction polynomiale.

$$\Phi(V, V_i) = \exp\left(\frac{-\|V - V_i\|^2}{2\sigma^2}\right) \quad (5-31)$$

$$\Phi(V, V_i) = (\sigma \times V^T V_i + 1)^p \quad (5-32)$$

La suite du paragraphe donne un exemple d'optimisation de ces deux paramètres dans le cas de la base d'images simulées simplifiée.

Une grille de recherche exhaustive a été utilisée pour couvrir simultanément l'espace des σ entre 2^{-12} et 2^2 et l'espace des C entre 2^0 et 2^{20} . Une échelle logarithmique base 2 en fonction des valeurs de C et σ est utilisée. Pour chaque valeur de σ et C , une validation croisée à sept ensembles a été appliquée aux vecteurs de caractéristiques de la base d'apprentissage pour extraire les performances du SVM correspondant à chacun de ces couples. En d'autres termes, la base d'apprentissage est séparée en sept sous-ensembles, six d'entre eux servent à extraire le modèle prédictif du SVM pour chaque couple de valeurs et l'appliquent sur le septième groupe. Chaque itération modifie l'ordre de sélection des sous-ensembles et les sept résultats sont moyennés pour chaque C et σ afin d'extraire les valeurs qui optimisent la précision du système. Les performances moyennes pour la base d'images simulées simplifiée détaillée au paragraphe 5.5.2 sont présentées sur la Figure 5-12, pour le RBF a), le noyau polynôme de degré 3 b) et de degré 4 c). Les performances du noyau polynôme de degré 2 ne sont pas montrées étant donné que celui-ci converge rarement vers un hyperplan optimal. L'échelle de couleur correspond à la précision de classification obtenue avec le SVM considérant chaque noyau et valeurs de C et σ . Cette figure prouve clairement que le RBF est le noyau le plus adapté à notre problème de classification avec les plus hautes performances (rose sur l'échelle de couleurs). Le Tableau 5-3 donne les paramètres optimaux obtenus pour chaque noyau, associés au couple de paramètres correspondant. Le RBF augmente la précision de 2 % par rapport aux autres fonctions. Dans la suite, le paramètre de pénalisation C et la déviation standard σ sont fixés à 57 et 0.044 (cercle rouge de la Figure 5-12).

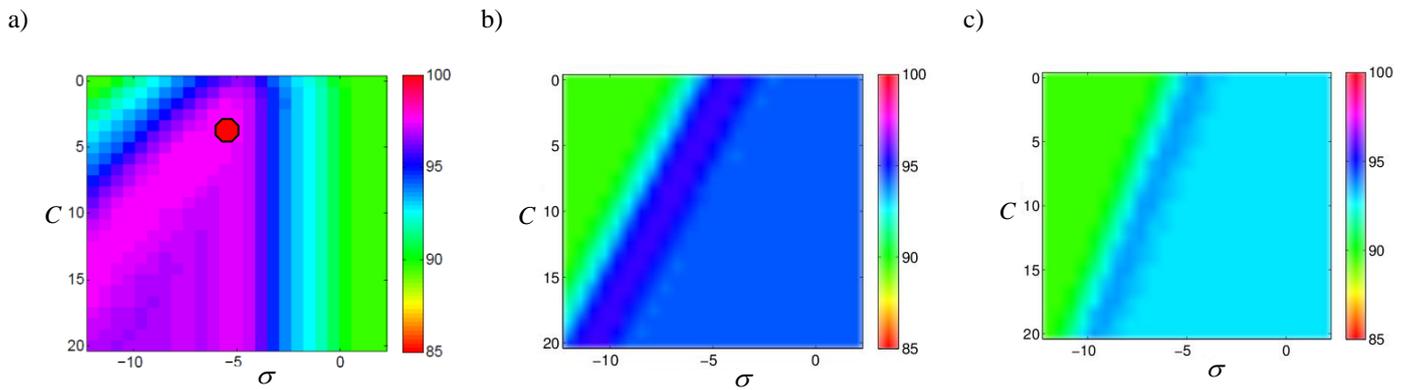


Figure 5-12 : Grilles de recherche permettant d'extraire la meilleure valeur des paramètres C et de σ , pour le RBF (a), le noyau polynôme de degré 3 (b) et de degré 4 (c)

Tableau 5-3 : Meilleures performances associées aux valeurs des paramètres correspondants pour chaque fonction noyau

	C	σ	Précision
RBF	57	0.0044	97.65
Polynomial degré 3	322	0.0093	95.78
Polynomial degré 4	1029	0.0042	95.47

5.3.3 Choix des exemples utilisées pour l'apprentissage

Afin de favoriser le pouvoir de généralisation des classifieurs linéaires, il est nécessaire que l'apprentissage soit réalisé sur un large éventail d'exemples de tumeurs et de non tumeurs différents. Pour ce faire, nous avons choisi une base d'apprentissage contenant au minimum autant d'exemples de la classe H1 que de la classe H0. Ce choix permet de se placer dans les conditions énoncées au paragraphe 5.2.1.2 pour le LDA qui considère autant d'exemples d'apprentissage dans les deux classes pour la simplification du calcul de la formule de Bayes. En d'autres termes, les probabilités $P(H_i)$ sont égales pour $i = \{0,1\}$, simplifiant l'équation (5-2).

Les vecteurs de caractéristiques V des exemples d'apprentissage sont uniquement calculés au centre de chaque tumeur des images dites d'apprentissage pour la classe H1 et aux mêmes coordonnées des images de la base d'apprentissage H0.

En ce qui concerne la taille de la base d'apprentissage, l'influence du nombre de tumeurs sélectionnées parmi toutes celles disponibles n'est pas directement détaillée dans ce manuscrit. Ce questionnement a cependant donné lieu à une communication parallèle [Lartizien et al., 2009] évaluant l'intérêt de répliquer les exemples de la base d'apprentissage par bootstrap. Cette étude préliminaire met en exergue le difficile processus d'optimisation de l'apprentissage également décrit par [Sahiner et al., 2000].

5.3.3.1 L'apprentissage organe spécifique

Une considération concernant la paramétrisation du CAD est de tenir compte de l'hétérogénéité du bruit de fond des organes cibles affectés par les tumeurs. Il est en effet probable qu'un modèle prédictif adapté à un organe spécifique donne de meilleures performances qu'un modèle entraîné sur des tumeurs réparties dans tout le corps. Différents apprentissages et tests focalisés sur chaque organe cible ont été réalisés et leurs performances comparées à celles d'un apprentissage sur l'ensemble du corps humain.

5.3.3.2 L'apprentissage avec une base de tumeurs enrichie en cas sains

Pour le LDA, nous avons choisi un échantillon d'apprentissage contenant l'ensemble des vecteurs V de caractéristiques des tumeurs et non tumeurs disponibles dans la base d'apprentissage. L'apprentissage du LDA est ainsi réalisé sur un nombre de vecteurs descripteurs équilibré entre les deux classes.

Un enrichissement de la base d'apprentissage, noté $H0+$, a été évalué afin d'affiner le modèle prédictif des SVM. La base enrichie contient ainsi plus d'exemples négatifs, c'est-à-dire de voxels non-tumeurs choisis de manière aléatoire parmi l'ensemble des voxels disponibles. Une comparaison des performances du SVM sans ou avec enrichissement des exemples d'apprentissage évaluera sa pertinence par rapport à nos images cibles contenant très peu de vrais positifs (tumeurs) et un grand nombre de vrais négatifs (tous les autres voxels). 5000 voxels supplémentaires, aléatoirement choisis parmi l'ensemble des volumes de la base d'apprentissage $H0$, ont ainsi été ajoutés pour chacun des trois organes d'intérêt.

5.3.4 Optimisation du vecteur de caractéristiques V

Comme décrit précédemment, les éléments des vecteurs de caractéristiques de l'image correspondent aux coefficients d'une décomposition en ondelettes 3D non décimée de niveau j utilisant l'ondelette Bior4.4. Les composantes de ce vecteur descripteur doivent être optimisées afin de limiter le nombre de FPI.

5.3.4.1 Le niveau de décomposition en ondelettes

La valeur du niveau de décomposition en ondelettes est d'abord évaluée avec une échelle de plus en plus dilatée allant de 2 à 5. Comme mentionné précédemment, la problématique de cette thèse est focalisée sur la détection de tumeurs quasi Gaussiennes dans l'image reconstruite et de faible largeur à mi hauteur. Il peut donc apparaître excessif de tester des niveaux d'échelle si dilatés ($j = 5$ par exemple). On peut en effet s'attendre à ce que le signal cible n'apparaisse qu'au niveau des basses échelles (plus détaillées). Rappelons cependant que les ondelettes ne permettent pas seulement d'extraire la tumeur type, mais également d'extraire les caractéristiques du bruit environnant. Des caractéristiques d'échelle élevée (dilatée), associées à celles d'échelle réduite, pourraient ainsi être utiles à la discrimination des tumeurs du fond.

Il peut parfois apparaître que la décomposition en ondelettes de niveau entier rate certaines tumeurs de trop petit contraste pour cause d'échantillonnage initial trop large de l'espace des ondelettes [Strickland and Hahn, 1996; Strickland and Hahn, 1997]. Par conséquent, nous avons testé l'intérêt d'ajouter de sous-niveaux de décomposition ou *voices*, comme proposé par Rioul [Rioul and Vetterli, 1991] et utilisé par Strickland [Strickland and Hahn, 1996; Strickland and Hahn, 1997]. Ces *voices* sont des demi-niveaux intégrées au niveau entier de manière à affiner l'échantillonnage initiale de l'ondelette et la rapprocher d'une transformée continue. L'extraction de ces *voices* pour chaque niveau j de décomposition consiste à calculer une autre transformée en ondelettes avec les mêmes filtres passe bas L et passe haut H . Des images interpolées d'un facteur $\sqrt{2}$, au lieu des images originales sont cependant utilisées.

5.3.4.2 La sélection de caractéristiques

Afin de limiter les calculs et le surapprentissage des classifieurs, un algorithme de sélection de caractéristiques, couplé au LDA et au SVM a été utilisé. L'algorithme génétique (AG) [Sankar and Wang, 1996] a été utilisé pour évaluer le pouvoir discriminant de chaque composante des vecteurs de caractéristiques de niveau j . La sélection de caractéristiques, si nécessaire, doit permettre une erreur minimale de classification avec un jeu minimal de caractéristiques des images. L'AG, décrit au paragraphe 2.4.4.2, est une technique d'optimisation où chaque solution d'un problème, ici le choix d'un ensemble de composantes, est traitée comme un individu compétitif. La meilleure solution évolue par croisements, mutations et sélections des individus. Chaque individu est évalué conformément à une fonction objectif qui quantifie le(s) critère(s) à optimiser. Nous avons fixé cette fonction à la précision de classification, définie par l'équation (5-33).

$$précision = \frac{VP}{VP + FP} \quad (5-33)$$

Le choix du vecteur V optimal est donc dépendant et spécifique du classifieur utilisé pour le calcul de cette fonction objectif. La probabilité de mutation est fixée à 10 % avec une population initiale de quinze individus soit quinze vecteurs de caractéristiques différents. Leurs composantes sont choisies aléatoirement à l'initialisation du processus. Le processus de sélection de composantes les plus discriminantes est itératif. Notre critère d'arrêt est fixé à sept nouvelles générations de vecteurs de caractéristiques sans aucune amélioration des performances. L'AG associé à chaque classifieur pour le calcul de la précision n'est appliqué que sur les tumeurs de la base d'apprentissage.

Les composantes du vecteur V , sélectionnées par l'AG, sont comparées à celles sélectionnées par les tests statistiques de Student et de Fisher. Ces deux tests ont aussi été utilisés comme méthode de sélection de caractéristiques dans la littérature par [Nandi et al., 2006]. Les deux tests statistiques permettent de mesurer la significativité des différences entre les caractéristiques des deux bases d'apprentissage H_0 et H_1 . Le test de Student [Press et al., 1992] permet de vérifier que les deux jeux de caractéristiques (sains et pathologiques) se discriminent par leurs moyennes, tandis que le test de Fisher [Fisher, 1925], généralement calculé en amont, étudie les différences de leurs variances. Contrairement à l'AG qui fixe les composantes du vecteur V par optimisation des performances des classifieurs, les tests statistiques ne reposent que sur une observation de chaque composante provenant des deux classes sans intervention des classifieurs.

5.3.5 Optimisation de l'étape de RFP

L'étape de RFP se base dans un premier temps sur l'élimination des monovoxels à partir des cartes binaires obtenues à partir du seuillage des cartes de score provenant des classifieurs pour les images de validation. La sensibilité et le taux de FPI sont comparés avant et après ce nettoyage afin d'en montrer l'intérêt.

Il en est de même avec la méthode de classement des scores. Le rang seuil optimal à appliquer à la liste ordonnée des tumeurs candidates est calibré de manière à conserver au moins 90 % des candidats VP. Par ailleurs, cette valeur seuil considère la potentielle baisse de sensibilité due à l'élimination des monovoxels.

5.4 Performances du système CAD

5.4.1 Introduction

Cette section expose les critères d'évaluation des performances du système CAD final, sachant que l'ensemble des résultats de cette partie est obtenu à partir des images de test.

Comme introduit dans l'état de l'art (Chapitre 2) et lors de la mise en place du cahier des charges (Chapitre 3), aucun formalisme pour l'évaluation 3D n'est réellement proposé actuellement dans la littérature. Les quelques études qui présentent des performances 3D [Zhao et al., 2003; Paik et al., 2004] se heurtent généralement à des valeurs de faux positifs très élevées par rapport à celles obtenues à partir d'évaluations 2D (cf. paragraphe 2.2).

5.4.2 Evaluation du système CAD complet

Les performances finales sont évaluées à partir de la base d'images de test. Dans un premier temps, les performances quantitatives en termes de sensibilité et FPI sont considérées. L'aspect visuel des cartes binaires 3D superposées aux images originales est aussi observé. Les cartes doivent être les plus propres possible (minimum de faux positifs) et présenter les vraies tumeurs comme correctement détectées, même si ce n'est qu'au travers d'un voxel.

Cette évaluation quantitative et visuelle est dans un premier temps réalisée à partir des images de détection nettoyées obtenues pour chaque classifieur. Une observation visuelle de la distribution des vrais positifs et des faux positifs sur chacune des cartes binaires provenant de chaque image de test est réalisée de manière à valider l'intérêt de leur combinaison. Cette combinaison, dite par vote à la

majorité, aboutit à une carte binaire représentative du système CAD final pour chaque image de test. Ces cartes binaires permettent aussi l'évaluation du CAD complet et sont donc observées.

5.4.3 Evaluation du système CAD complet au regard des performances humaines

Un système CAD se doit d'être au moins aussi efficace qu'un observateur humain pour pouvoir être perçu comme un second avis pertinent. La dernière étape de la validation du CAD repose sur la comparaison des performances quantitatives du système développé aux performances humaines obtenues sur les mêmes images.

Les contrastes des tumeurs des images simulées proviennent d'études de calibration à l'aide d'observateurs humains. Cette étude de calibration est détaillée en [Lartizien et al., 2004] et en Annexe C pour la base d'images simulées à partir d'un modèle simplifié et au paragraphe 4.5.4.2 pour la base d'images oncoPET_DB. Elle permet d'échantillonner les contrastes des tumeurs de chaque organe de manière à recouvrir l'ensemble des performances humaines possibles (de 10% à 90% de sensibilité). Ainsi, le contraste de chaque tumeur de couple (organe/diamètre) est associé à une probabilité de détection par l'observateur humain d'autant plus élevée que le contraste est fort.

Les performances de détection du CAD ont été dans un premier temps extraites pour chaque contraste et chaque organe de manière à les comparer aux sensibilités humaines de référence.

Dans un second temps, le CAD a été placé dans le cas d'une tâche de détection plus restrictive reproduisant l'étude réalisée par les observateurs humains lors de la calibration des contrastes. Non seulement les performances par contraste sont calculées, mais le CAD doit choisir, tout comme l'a fait l'observateur humain, les N tumeurs candidates les plus évidentes parmi toutes, N correspondant au nombre de tumeurs réellement présentes dans les images. L'ensemble des tumeurs candidates est extrait des cartes binaires provenant du CAD final. Les degrés de suspicion correspondant à ces tumeurs candidates sont extraits des cartes de score provenant des deux classifieurs séparément, comme au paragraphe 5.2.2.3 pour l'ensemble des voxels du cluster. Les N tumeurs de plus haut score sont conservées comme les plus suspectes parmi toutes. Ainsi, la sensibilité du système CAD est calculée en ne considérant que ces N tumeurs candidates. Elle est ensuite comparée aux sensibilités humaines correspondantes pour chaque couple (organe/diamètre). Les deux combinaisons carte binaire (permettant d'extraire les agrégats candidats) et carte de score (permettant d'extraire le score des candidats) sont les suivantes : CAD/LDA et CAD/SVM.

Pour une assurer l'intérêt du CAD, ces deux évaluations doivent fournir des performances au moins aussi élevées que celles des observateurs humains.

5.5 Performances du système CAD sur une base d'images simulées simplifiée

5.5.1 Introduction

Le CAD a dans un premier temps été évalué à l'aide de la base d'images TEP simulées à partir d'un modèle de corps humain simplifié. Cette première étape permet de vérifier la pertinence de chacune des étapes et la faisabilité du système à partir d'images simples.

Le paragraphe 5.5.3 présente les résultats d'optimisation des paramètres du CAD tant pour l'extraction initiale des tumeurs candidates que pour la réduction de faux positifs. Les performances finales du CAD après combinaison par vote à la majorité sont ensuite évaluées dans le paragraphe 5.5.3.4.

5.5.2 Caractéristiques de la base d'images simulées utilisées

Dans ce chapitre, l'étude de faisabilité du système CAD est réalisée à partir d'images simulées présentant un modèle simplifié du corps humain adapté du MCAT [Tsui et al., 1993], dans lequel des tumeurs sphériques ont été virtuellement insérées dans les tissus mous, le foie et les poumons. La Figure 5-14 présente une image de cette base simulée.

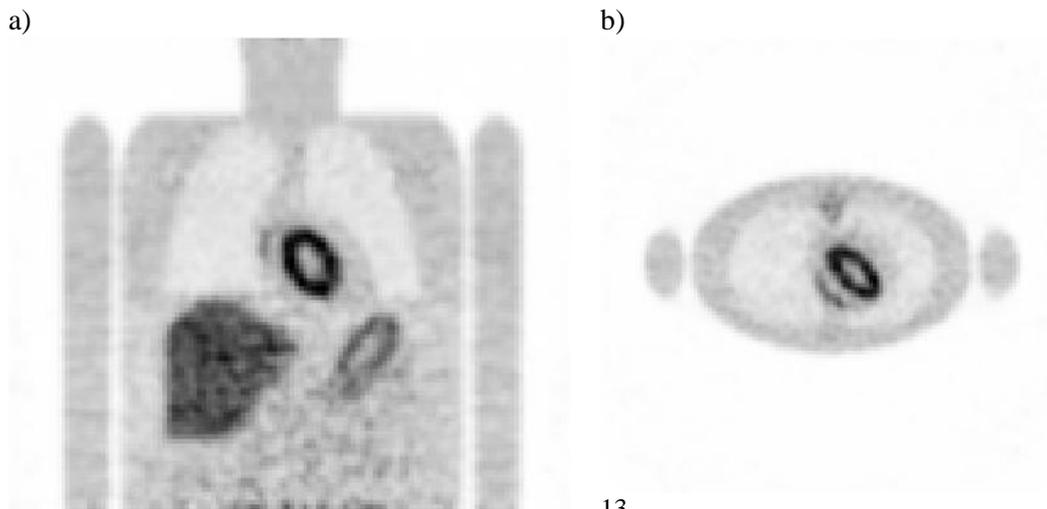


Figure 5-14 : Coupe coronale (a) et transverse (b) d'une image de la base simplifiée

Cette base d'images TEP simulées simplifiée a été développée lors de la thèse de C. Lartizien [Lartizien, 2001]. Un ensemble de volume de $128 \times 128 \times 108$ voxels de dimension 5,0625 mm dans les trois directions est disponible. Une description des caractéristiques de cette base d'images peut être trouvée en Annexe C et dans l'article [Lartizien et al., 2004].

L'étude des performances du système CAD développé s'appuie sur une méthode d'évaluation par hold-out (cf. paragraphe 2.5.4). La base d'images simulées est ainsi partagée en deux groupes distincts, le premier étant utilisé pour l'apprentissage intrinsèque du classifieur, le second pour le test du modèle dérivé. Une troisième base d'images, dite de validation, peut aussi être exploitée pour le choix des paramètres du système CAD extrinsèques au classifieur. Le partitionnement des images disponibles de la base simulée est plus précisément détaillé ci-après.

- La base d'apprentissage permet le calcul des modèles prédictifs des deux classifieurs, c'est-à-dire l'hyperplan optimal pour le SVM et le poids ω pour le LDA. Dans cette thèse, les vecteurs de caractéristiques pour l'apprentissage ne sont calculés qu'aux coordonnées des centres des tumeurs pour les images pathologiques et les coordonnées correspondantes dans les images saines. Ces images consistent en 250 volumes pathologiques contenant un total de 1750 tumeurs localisées dans les trois organes cibles et 250 volumes sains sans tumeur. Les 250 volumes pathologiques correspondent à 25 modèles différents de distribution de tumeurs associés à 10 réalisations de bruit différentes. Ainsi seules 175 tumeurs ont des caractéristiques différentes parmi les 1750 disponibles, dont seul le bruit environnant varie. Les 250 volumes sains correspondent à autant de réalisations de bruit différentes du modèle sain. En d'autres termes, le même patient virtuel subit 250 fois l'examen TEP, dont l'image résultante ne varie que par la distribution de bruit obtenue.
- La base d'images de validation sert à l'ajustement des autres paramètres du CAD comme le vecteur descripteur de caractéristiques. Elle contient 25 images simulées de modèles différents contenant 175 tumeurs dans les trois organes cibles.

La base d'images de test permet le calcul des performances du système CAD final. Elle comprend 25 images différentes, soit 175 tumeurs, exclusivement utilisées pour le test.

5.5.3 Optimisation des paramètres du CAD

5.5.3.1 Introduction

Les paramètres du CAD conditionnant les performances des deux étapes principales du CAD que sont l'extraction des tumeurs candidates et l'élimination de faux positifs sont optimisées dans ce paragraphe.

Comme nous l'avons défini au paragraphe 5.3, l'optimisation multiparamétrique de l'extraction des tumeurs candidates porte simultanément sur le choix des exemples de tumeurs pris en compte pour l'apprentissage et sur les composantes du vecteur de caractéristiques le représentant. La meilleure combinaison de paramètres est celle qui limite le nombre de FPI pour une sensibilité donnée de 90%. Ces performances sont mesurées à partir des courbes FROC extraites des images de validation. Une combinaison optimale de paramètres pour chacun des classifieurs est définie. Rappelons que le taux de FPI est calculé sur l'ensemble de l'image TEP 3D corps entier selon la définition donnée au paragraphe 5.2.1.3.

Les deux méthodes de réduction de faux positifs basées sur un nettoyage des cartes binaires de détection décrites au paragraphe 5.2.2 sont ensuite évaluées.

5.5.3.2 Optimisation des paramètres du LDA pour l'extraction initiale des tumeurs

Caractéristiques de la base d'apprentissage

L'optimisation multiparamétrique débute pour le LDA par l'influence d'un apprentissage organe spécifique. Un modèle prédictif pour chacun des trois organes d'intérêt est généré à partir des tumeurs de la base d'apprentissage localisées dans cet organe. La Figure 5-15 présente l'influence d'un apprentissage ciblé sur un organe spécifique (sans symbole) comparé à un apprentissage sur toutes les tumeurs disponibles (symbole). Les courbes FROC correspondent à deux niveaux extrêmes de décomposition $j=2$ (trait plein) et $j=5$ (trait pointillé) sur le poumon (a), le foie (b) et les tissus mous (c).

La Figure 5-15b prouve que le classifieur dérivé d'un apprentissage propre à chaque organe est particulièrement adapté pour le foie. En effet, le nombre de FPI diminue de 140 à 90 pour $j=2$ et de 60 à 40 pour $j=5$ pour une sensibilité de 90%. Comme on le voit sur l'image de la Figure 5-14, la texture du foie lui est spécifique et repose sur une granulosité importante. Cette granulosité est très différenciée des autres organes et particulièrement de la texture des poumons. Ainsi le LDA, entraîné sur les caractéristiques de bruit particulières du foie, permet de plus facilement discriminer les VP du bruit environnant.

L'amélioration des performances par un apprentissage organe spécifique sur les poumons est plus modérée. Comme le montre la Figure 5-15a, des résultats comparables sont obtenus pour $j=5$, même si le taux de FPI est réduit de 125 à 60 pour $j=2$, en considérant une sensibilité de 90%.

Concernant les tissus mous, les performances à sensibilité fixée de 90% sont similaires avec ou sans apprentissage spécifique quelque soit le niveau j de décomposition. Ce résultat peut certainement s'expliquer par l'hétérogénéité de cette structure répartie dans tout le corps. Les tumeurs à détecter sont certainement affectées par un bruit de nature très variable en fonction de leur localisation et de la contamination des organes voisins. On peut donc supposer que la notion d'organe spécifique a dans le cas des tissus mous peut-être moins de sens.

Les courbes FROC correspondant à un niveau de décomposition en ondelettes $j=3$ et $j=4$ ne sont pas représentées sur la Figure 5-15, mais suivent la tendance des courbes de leur plus proche j voisin. Ces conclusions nous confirment l'intérêt d'utiliser un apprentissage différent pour chaque organe d'intérêt, appelé apprentissage organe spécifique.

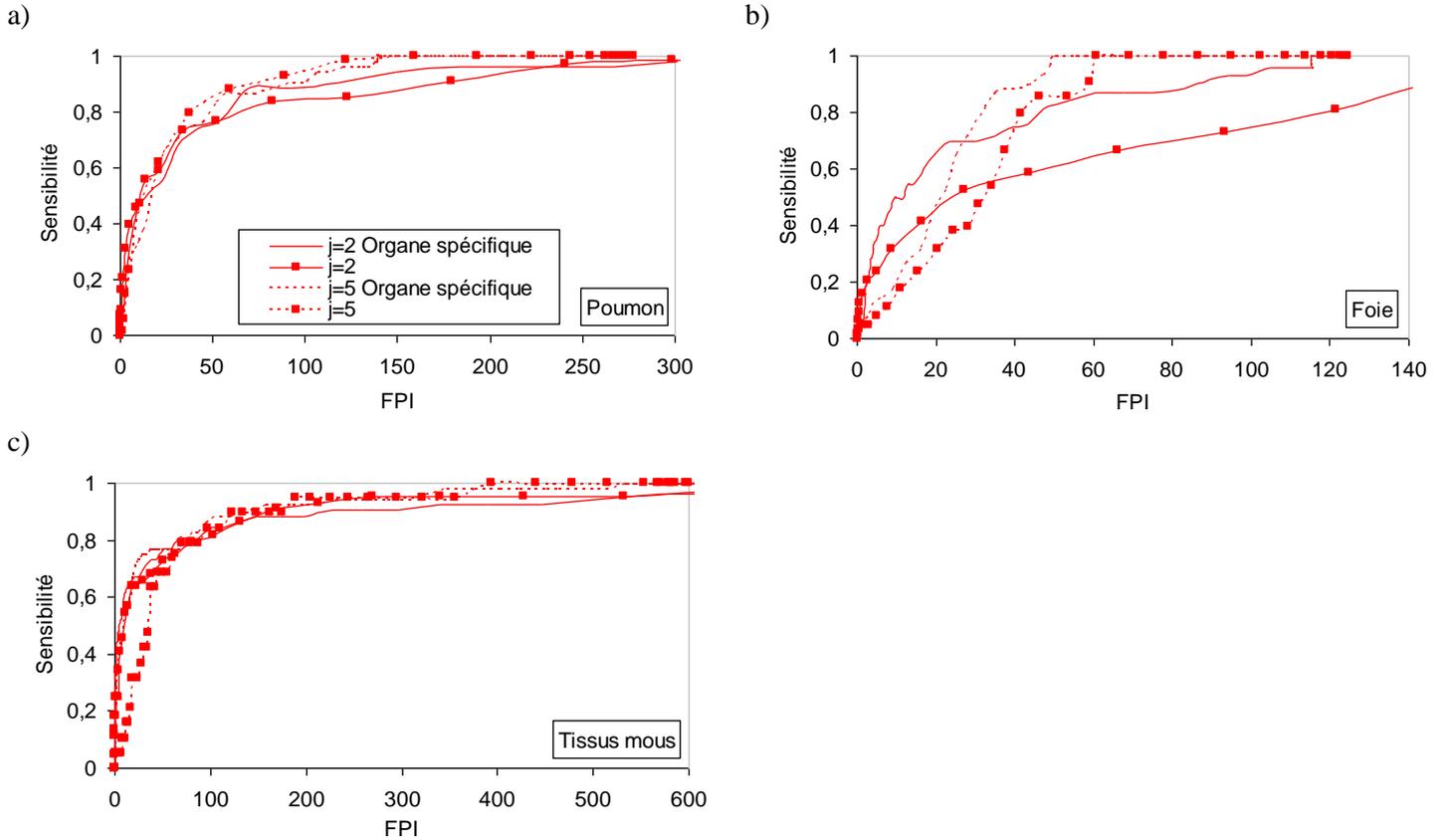


Figure 5-15 : Courbes FROC obtenues pour deux niveaux de décomposition extrêmes, $j=2$ (ligne pleine) et $j=5$ (ligne pointillée) pour un apprentissage organe spécifique (aucun symbole) ou sur le corps entier (symbole). Les tumeurs du poumon (a), du foie (b) et des tissus mous (c) sont considérées.

Optimisation des composantes du vecteur V de caractéristiques

Le second paramètre à optimiser pour l'extraction initiale des tumeurs candidates concerne les composantes du vecteur de caractéristiques V et plus précisément le niveau j de décomposition en ondelettes. Les niveaux entiers de $j = 2$ à $j = 5$ ont été observés, tandis que l'intérêt de sous niveaux de décomposition appelés voix décrits au paragraphe 5.3.4 a été analysé. Un algorithme de sélection de caractéristiques, l'algorithme génétique, a été appliqué de manière à évaluer le pouvoir discriminant des caractéristiques fréquentielles extraites de la décomposition en ondelettes 3D.

La Figure 5-16 montre les courbes FROC dérivées d'un apprentissage organe spécifique sans (symboles vides) et avec sélection de caractéristiques (lignes pointillées et symboles pleins) pour $j=2$ et $j=5$ pour les poumons (a), le foie (b) et les tissus mous (c). L'utilisation des voix est aussi illustrée sur les courbes FROC pour $j=2$ (cercle vide sur trait plein).

Le Tableau 5-4 résume les FPI extraits de la Figure 5-16 pour chaque combinaison de paramètres pour une sensibilité de 90%. Pour information, la taille du vecteur de caractéristiques V pour $j=2$ est réduite de 16 à 7 composantes pour les poumons et de 16 à 11 composantes pour le foie et les tissus mous avec l'algorithme génétique. Pour $j=5$, le vecteur de caractéristiques V passe de 40 à 30 composantes pour les poumons, à 26 pour le foie et à 25 pour les tissus mous. L'utilisation des voix double la taille originale du vecteur de caractéristiques.

Tableau 5-4 : Valeurs de FPI extraits des courbes FROC à une sensibilité fixée de 90% obtenues pour chaque combinaison de paramètres et chaque organe

	$j=2$	$j=2$ avec AG	$j=2$ avec Voices	$j=5$	$j=5$ avec AG
Poumons	60	200	60	65	65
Foie	90	55	85	40	55
Tissus mous	200	120	250	120	120

Concernant le foie et les tissus mous, le nombre de FPI sans sélection de caractéristiques est plus bas avec $j=5$ qu'avec $j=2$. L'AG ne permet pas de réduire le nombre de FPI pour $j=5$, mais le fait de manière significative pour $j=2$ l'amenant au même niveau que $j=5$ (55 FPI pour le foie et 120 pour les tissus mous).

Quelque soit l'organe ciblé, l'ajout de sous-niveaux de décomposition (voices) ne permet pas réellement d'améliorer les performances, mais tout au plus d'égaliser les performances correspondantes sans voix. Il se pourrait que la subdivision par deux des niveaux entiers n'approxime encore pas suffisamment la transformée continue pour permettre de faire ressortir les petites tumeurs problématiques.

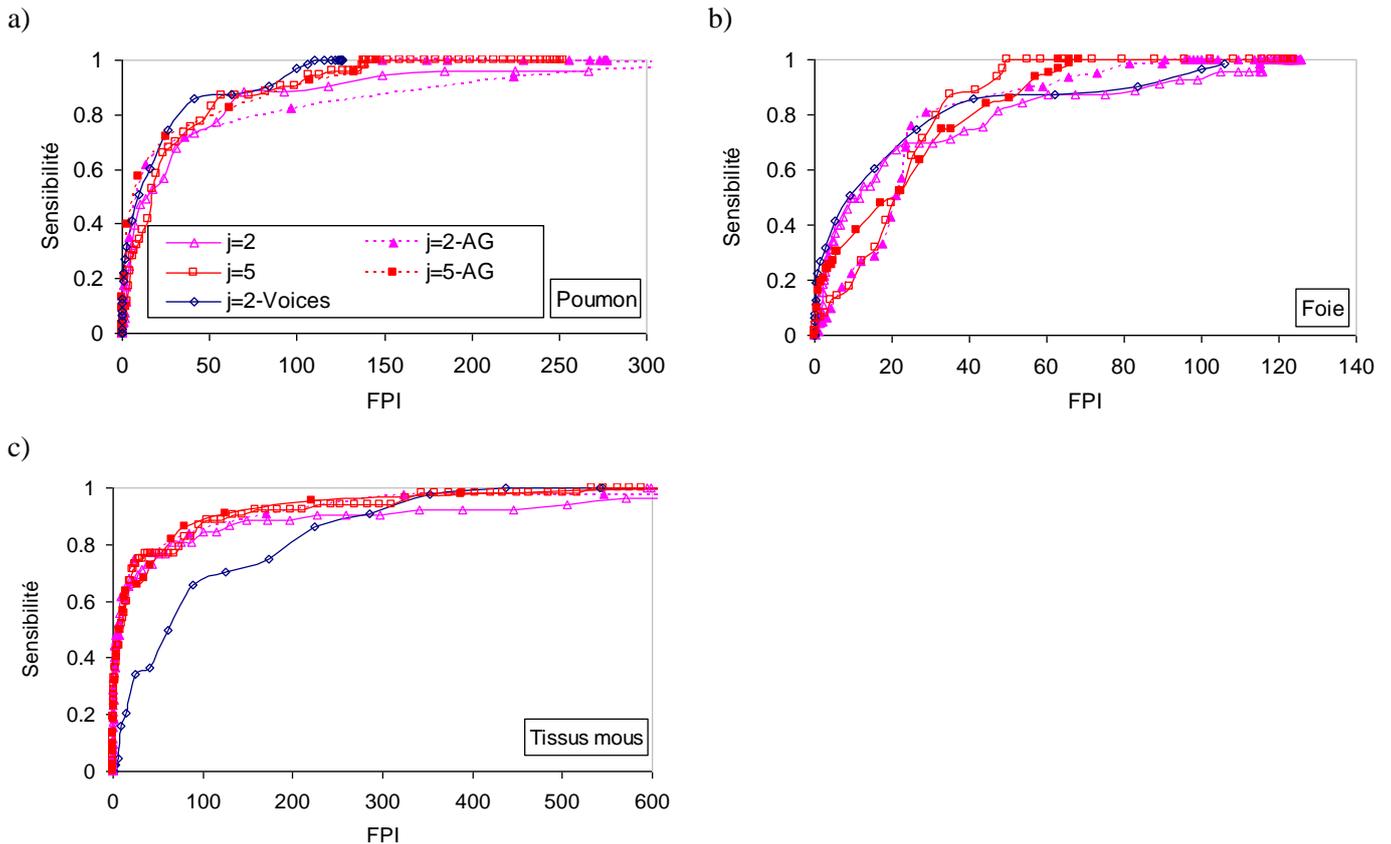


Figure 5-16 : Courbes FROC pour deux niveaux de décomposition $j=2$ et $j=5$ sans aucune sélection de caractéristiques (symboles vides) et avec sélection de caractéristiques (ligne pointillée et symbole plein).

La grande différence observée du nombre de FPI entre $j=2$ et $j=5$ sans sélection de caractéristiques pour le foie et les tissus mous est expliquée par l'observation de la Figure 5-17. Celle-ci présente une vue en MIP d'une image simulée superposée à sa carte binaire pour $j=2$ (a) et $j=5$ (b). Les sept tumeurs disponibles pour cette image, toutes détectées par le LDA, sont fléchées. La carte binaire pour $j=5$ (b) présente un plus petit nombre de clusters détectés que pour $j=2$ (a) ce qui correspond bien aux résultats du Tableau 5-4. Ces clusters sont cependant beaucoup plus gros et apparaissent dilatés par le niveau élevé de décomposition en ondelettes $j=5$. On peut supposer qu'ils correspondent à l'agglomération de plus petits clusters observés pour $j=2$. De tels gros clusters risquent d'être plus difficiles à éliminer pendant la phase de réduction de faux positifs. Ils contiennent en effet à la fois des VP et des FP. Nous avons donc choisi de tenir compte de ce critère visuel qui n'est pas reflété par l'analyse FROC et pas pris en compte dans le processus d'agrégation des voxels et du besoin d'entrer la meilleure carte binaire au processus de RFP.

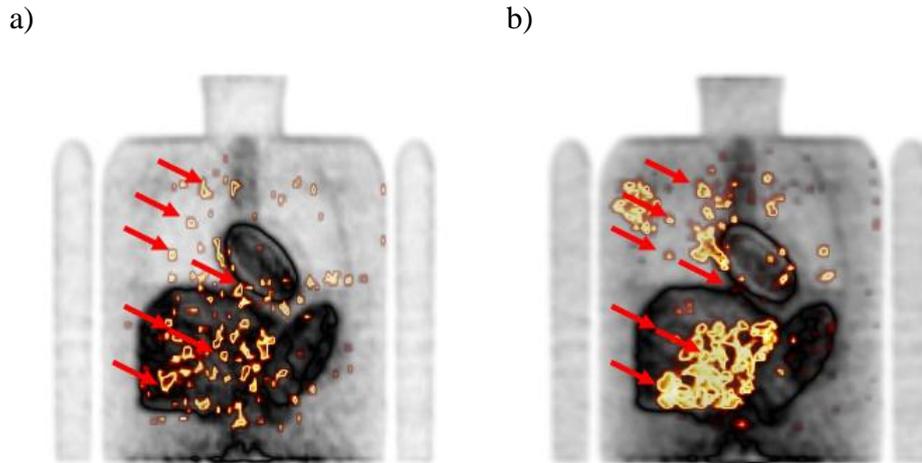


Figure 5-17 : Vue MIP 3D d'une image TEP simulée corps entier superposée à la carte binaire correspondant à $j=2$ (a) et $j=5$ (b) sans sélection de caractéristiques pour le LDA. Les sept tumeurs disponibles sont pointées par les flèches rouges.

L'utilisation de l'AG n'influence pas les performances des poumons pour $j=5$, mais augmente le taux de FPI pour $j=2$. Cette tendance peut paraître conflictuelle avec les bonnes performances de l'AG pour le foie et les tissus mous. Cependant, les performances initiales obtenues pendant la phase d'apprentissage de l'AG se montraient déjà inférieures pour les poumons comme l'explique le Tableau 5-5 pour $j=2$. Le Tableau 5-5 présente la précision associée à la sensibilité correspondante obtenue sur les tumeurs de la base d'apprentissage lors de la sélection des composantes du vecteur de caractéristiques. Rappelons que le paragraphe 2.4.4.2 précise le fonctionnement de l'AG tandis que les différents paramètres de l'algorithme sont donnés au paragraphe 5.3.4. L'algorithme utilise les tumeurs de la base d'apprentissage pour évaluer les différentes combinaisons de caractéristiques. L'optimisation du vecteur de caractéristiques est réalisée par l'intermédiaire des performances qu'il obtient avec le LDA.

Selon le Tableau 5-5, les meilleurs résultats sont obtenus pour les tissus mous, puis le foie et enfin les poumons avec une sensibilité assez faible. De tels résultats pourraient être expliqués par un arrêt trop précoce du processus itératif de sélection de caractéristiques (sept itérations). La meilleure combinaison de caractéristiques n'est donc peut-être pas atteinte par mutations, croisements et sélections pour cet organe et ce niveau de décomposition.

Tableau 5-5 : Valeurs de FPI extraits des courbes FROC à une sensibilité fixée de 90% obtenues pour $j=2$ et chaque organe

	Précision	Sensibilité
Poumons	95	81
Foie	96	92
Tissus mous	96	96

Aussi, nous avons voulu vérifier la concordance des caractéristiques sélectionnées par l'AG avec celles sélectionnées par le test statistique de Student. Celui-ci permet de ne conserver que les composantes vérifiant l'hypothèse de moyennes différentes, c'est-à-dire dont la p-valeur est inférieure au test de significativité 0,05. Un test de Fisher, vérifiant l'hypothèse d'égalité de variance entre les composantes du vecteur V pour les deux classes a conclu sur des p-valeurs toutes supérieures au seuil de significativité (0,05). Ce résultat a ainsi mis en évidence la légitimité d'utiliser un test de Student. Les échantillons utilisés pour le calcul des p-valeurs du test de Student correspondent aux 1750 vecteurs de caractéristiques de la base d'apprentissage pour chaque classe. Contrairement à l'AG, aucune notion d'optimisation des performances du classifieur n'est dans ce cas prise en compte.

Seuls les résultats correspondant aux caractéristiques pour $j=2$ soit $j \times 8 = 16$ composantes sont présentés dans le Tableau 5-6. Les caractéristiques initialement disponibles sont numérotées de 1 à 16 selon la formule (5-34) pour $j=2$.

$$V = [LLL_j, LLH_j, LHL_j, LHH_j, HLL_j, HLH_j, HHH_j], j = \{2, 3, 4, 5\} \quad (5-34)$$

Tableau 5-6 : Comparaison des caractéristiques sélectionnées par l'algorithme génétique ou le test de Student en considérant chaque organe cible et un niveau de décomposition $j=2$.

	Méthodes de sélection	Composantes sélectionnées	Taille du vecteur V
Poumons	Algorithme génétique	1, 8, 9, 10, 11, 14, 16	7
	Test de Student	2, 3, 5, 6, 8, 10, 11, 12, 15, 16	10
Foie	Algorithme génétique	1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 16	11
	Test de Student	5, 6, 7, 8, 11, 12, 13, 14, 15, 16	10
Tissus mous	Algorithme génétique	1, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16	11
	Test de Student	7, 9, 10, 11, 12, 13, 14, 15, 16	9

Le Tableau 5-6 montre en caractère gras les composantes conjointement sélectionnées par les deux méthodes de sélection. Aucune tendance particulière de sélection le long des niveaux de décomposition en ondelettes n'apparaît clairement, ni aucune direction de l'espace particulière. Il aurait été intéressant de pouvoir ordonner les caractéristiques par ordre de sélection afin de vérifier si le rang de chacune concordait entre les deux méthodes. Ce classement n'est cependant pas possible avec l'algorithme génétique qui traite chaque vecteur de caractéristiques comme un unique individu sans réellement distinguer les éléments le constituant.

On constate que la taille du vecteur descriptif V optimal obtenu pour les poumons avec l'AG est inférieure à celle du vecteur obtenu par le test statistique de Student. Ce résultat pourrait confirmer l'hypothèse d'un critère d'arrêt trop rapide de l'AG pour les poumons.

Conclusions

En se basant sur les résultats et discussions précédents dédiés au LDA, **la meilleure combinaison de paramètres pour le vecteur de caractéristiques V est fixée à $j=2$ pour tous les organes, sans sélection de caractéristiques pour les poumons et avec sélection de caractéristiques pour le foie et les tissus mous** (cellules mises en gras du Tableau 5-4). Le choix de $j=2$ avec AG est favorisé par rapport à $j=5$ pour le foie et les tissus mous car il permet d'obtenir une taille réduite de V pour un nombre de FPI identique mais de taille raisonnable avec l'AG.

Les résultats préliminaires de cette évaluation ont été présentés au colloque GRETSI 2007 [Tomei et al., 2007].

5.5.3.3 Optimisation des paramètres du SVM pour l'extraction initiale des tumeurs

L'optimisation de l'étape d'extraction initiale des tumeurs candidates par le SVM repose, non plus sur deux, mais trois paramètres. Deux d'entre eux portent sur le choix des exemples d'apprentissage (organe spécifique et H0+). Afin de limiter les combinaisons paramétriques multiples, nous avons, contrairement au LDA, d'abord optimisé les composantes du vecteur de caractéristiques V . Une fois cette optimisation effectuée, l'intérêt d'utiliser des échantillons d'apprentissage propre à chaque organe ou enrichi en cas sains (H0+) est évalué. Comme précédemment, l'optimisation des paramètres extrinsèques au classifieur repose sur le tracé de courbes FROC à partir des images de validation. La priorité est donnée à un nombre de FPI le plus bas possible pour une même sensibilité de 90%.

J'ai eu l'occasion d'encadrer Simon Marache-Francisco pour son stage de Master Recherche dans le cadre des développements portant sur cette partie de la thèse dédiée aux SVM.

Optimisation des composantes du vecteur V de caractéristiques

La Figure 5-18 présente les courbes FROC dérivées du SVM pour deux niveaux de décomposition ($j=2$ et 5), un apprentissage sur l'ensemble des organes et un test sur les poumons (a), le foie (b) et les tissus mous (c) avec (symboles pleins) et sans sélection de caractéristiques (symboles vides). Afin d'alléger les figures, nous ne montrons que la meilleure courbe obtenue pour l'AG c'est-à-dire celle correspondant au plus haut niveau de décomposition en ondelettes $j=5$ (carré plein).

Les performances de détection sont améliorées avec un niveau de décomposition élevé, particulièrement pour une sensibilité fixe de 90%. Pour donner un exemple, le nombre de FPI diminue de 260 pour $j=2$ sans AG à 180 pour $j=5$ sans AG pour les poumons (a). Contrairement au LDA, les amas de voxels classifiés positivement par le SVM ne sont pas de taille excessive.

Cette tendance, contradictoire avec celle du LDA, pourrait s'expliquer par une utilisation différente des vecteurs de caractéristiques par les deux classifieurs. Pour le LDA, le filtre adaptatif ω ((5-22)) contient des valeurs dont l'ordre de grandeur dépend du niveau d'échelle j , c'est-à-dire élevé pour un fort niveau de dilatation et faible pour un faible niveau d'échelle. La carte de score λ calculée pour l'image entière affiche donc de larges zones de valeurs élevées dues au produit scalaire des vecteurs ω et V de composantes élevées (relatifs à un j élevé et donc à une dilatation trop grande). Le seuillage de ces cartes de haut score donne donc de gros amas de voxels candidats. Le SVM quant à lui n'utilise pas les caractéristiques pour $j=2$ à $j=5$ par le calcul de moments d'ordre 1 et 2. Chaque caractéristique du vecteur V est traitée comme une composante indépendante apportant une information supplémentaire à la séparation des espaces. La notion d'échelle n'est donc pas prise en compte de la même manière et ne pénalise pas les résultats.

La Figure 5-18 indique également que les performances incluant une sélection de caractéristiques par l'AG avec $j=5$ restent inférieures à celles de même niveau de décomposition sans AG pour le foie (de 70 à 100 FPI) et les tissus mous (de 275 à 425 FPI). Les poumons présentent des résultats légèrement inférieurs avec sélection de caractéristiques (de 90 à 95 FPI). L'efficacité restreinte de l'AG sur nos jeux de données pourrait comme précédemment être expliqué par un arrêt trop précoce du processus de sélection de caractéristiques. L'AG cherche en effet à rassembler de manière optimale les composantes d'ondelettes par croisements, mutations et sélections en maximisant la fonction objectif. Il peut donc très bien ne pas atteindre la combinaison optimale s'il est stoppé trop tôt.

Compte tenu des observations faites, une combinaison optimale d'un niveau de décomposition en ondelettes $j=5$ sans AG est favorisée.

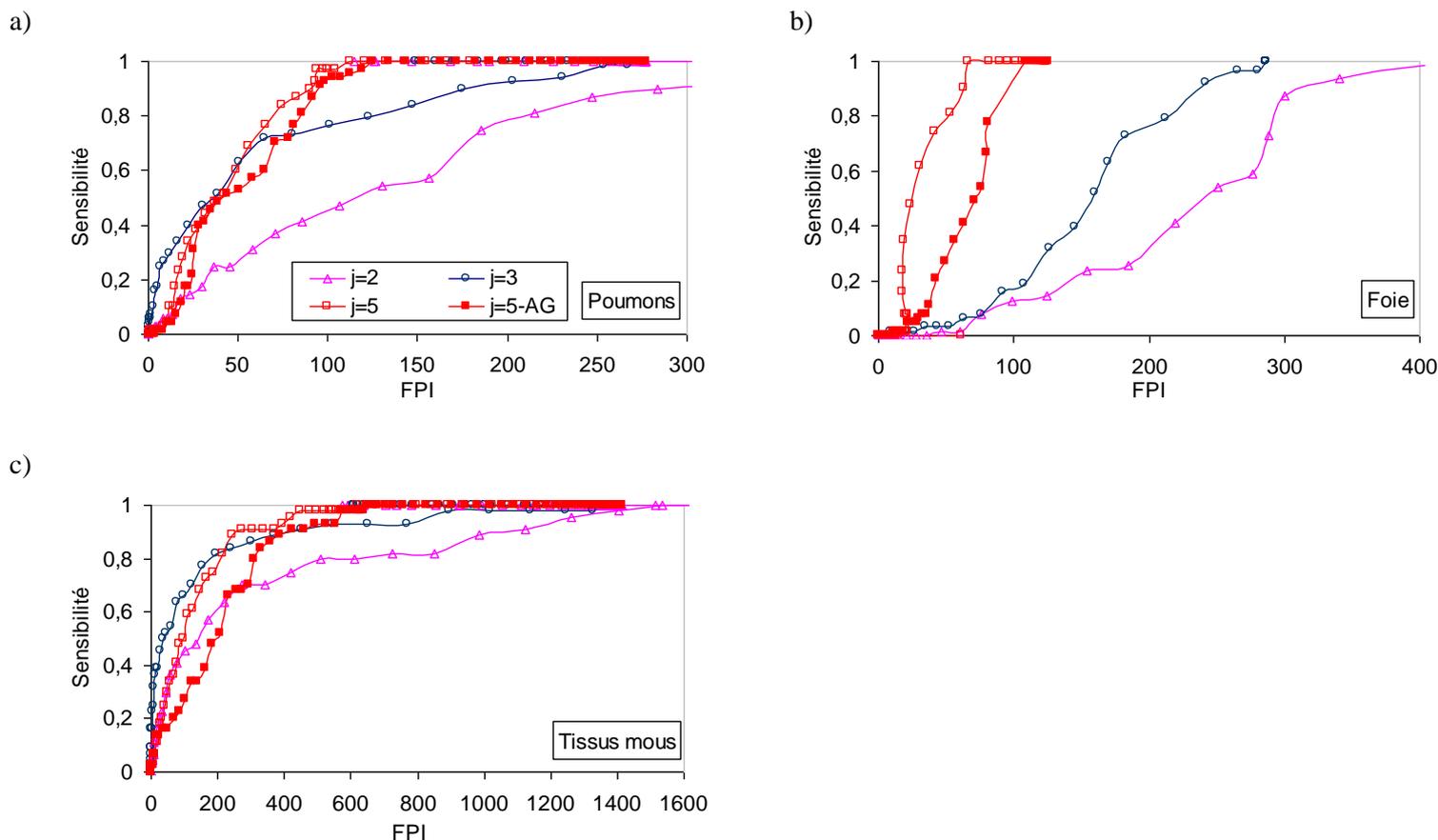


Figure 5-18 : Courbes FROC obtenues avec le SVM pour $j=2$ et $j=5$ sans sélection de caractéristiques (symboles vides) et $j=5$ avec sélection de caractéristiques (symboles pleins) pour un apprentissage sur tous les organes et un test sur les poumons (a), le foie (b) et les tissus mous (c).

Enrichissement des cas de la base d'apprentissage H_0+ et apprentissage organe spécifique

La Figure 5-19 illustre simultanément l'influence d'une base enrichie en cas négatifs (H_0+) (carré vide) et d'un apprentissage organe spécifique (pas de symbole) pour $j=5$. Pour obtenir la base d'apprentissage H_0+ , 5000 nouveaux points caractérisant les fausses détections causées par le bruit ou la texture des organes en dehors des tumeurs cibles, sont ajoutés. Ces courbes FROC sont à comparer à celle obtenue avec une base d'apprentissage non enrichie et basé sur des tumeurs de tous les organes confondus.

Considérons d'abord l'intérêt de l'enrichissement de la base d'images H_0 , noté H_0+ (triangles vides de la Figure 5-19) par rapport à un apprentissage classique (carrés vides). Le modèle prédictif semble amélioré par la base enrichie avec une diminution d'un facteur 1,3 (a) des FPI pour les poumons et avec un facteur 1,7 (b) pour le foie. Les performances pour les tissus mous apparaissent plus équilibrées avec un nombre de FPI équivalent pour une même sensibilité de 90%. Un tel résultat pourrait s'expliquer par un nombre de points additionnels trop réduits par rapport à la taille de cette structure (d'environ 208000 voxels), comparée aux 16000 et 35000 voxels du foie et des poumons. Il semblerait que la base d'apprentissage H_0 ne soit pas assez enrichie pour influencer le modèle prédictif et le rendre plus sensible aux possibles fausses détections.

En ce qui concerne l'intérêt de l'apprentissage dédié à un seul organe, la Figure 5-19 montre une nette diminution des FPI pour une sensibilité de 90% d'un facteur 1,7 pour le foie (b) et d'un facteur 3 pour les tissus mous (c) comparé à un apprentissage sur tous les organes utilisant cependant la base H_0 enrichie. Les poumons (a) présentent des performances équivalentes avec ou sans apprentissage organe spécifique.

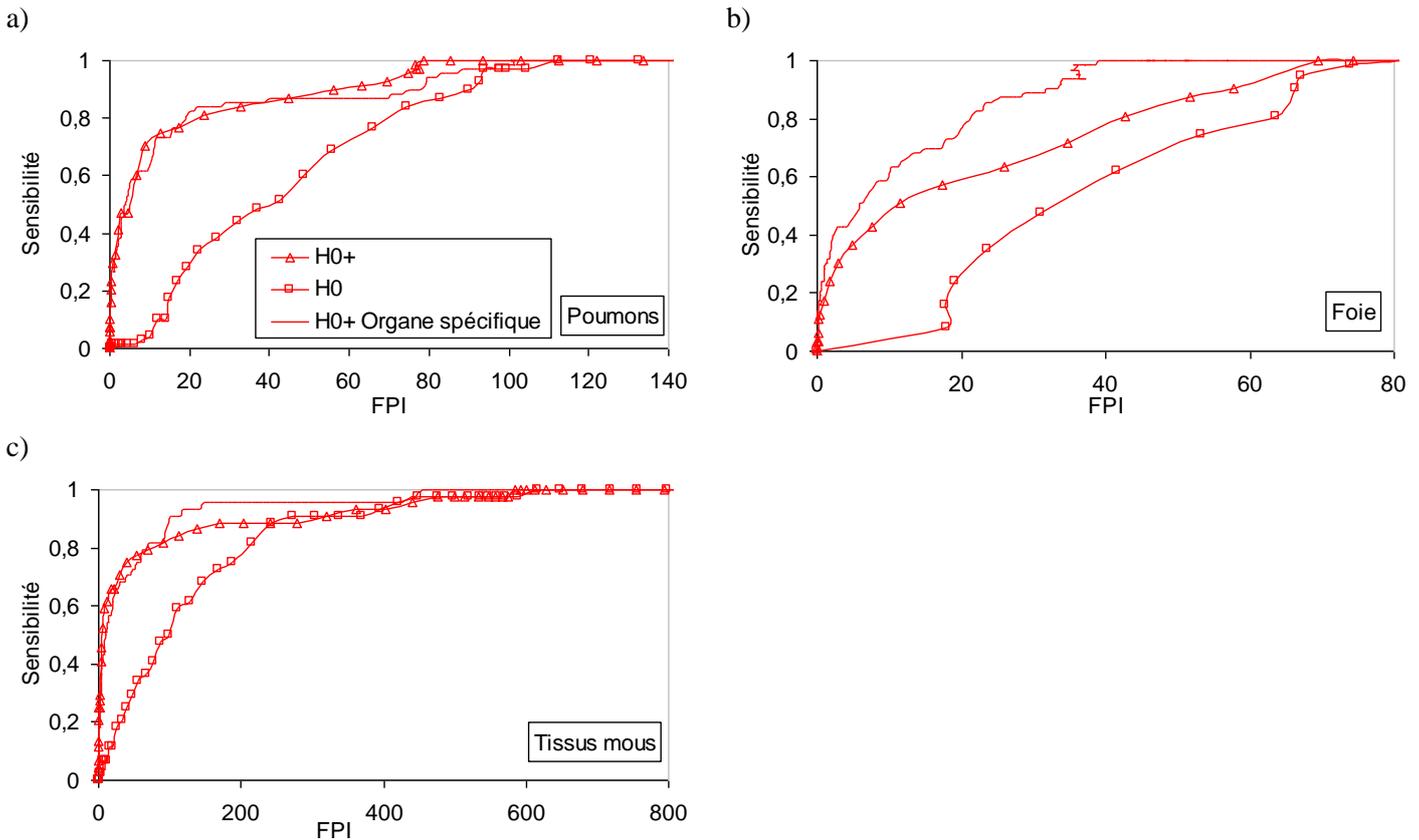


Figure 5-19 : Courbes FROC obtenues avec le SVM pour $j=5$ sans (symboles) et avec enrichissement de la base d'apprentissage H0, notée H0+ (sans symboles) pour les poumons (a), le foie (b) et les tissus mous (c).

Conclusions

Compte tenu de ces résultats, **la meilleure combinaison de paramètres pour le SVM est fixée à un niveau de décomposition $j=5$ associé à un apprentissage organe spécifique et une base H0 enrichie (H0+).**

L'ensemble de ces résultats portant sur l'optimisation de l'extraction des candidats pour le LDA et le SVM a été présenté au colloque MIC 2008 [Tomei et al., 2008].

5.5.3.4 Evaluation de la réduction des faux positifs pour le LDA et les SVM

La partie précédente a permis d'optimiser l'apprentissage des deux modèles prédictifs. Ces deux modèles de classification permettent d'extraire des cartes de score pour chaque image à classifier. Ces cartes de score sont ensuite seuillées à l'aide d'une valeur déterminée par une analyse FROC. Comme souligné sur la Figure 5-2, ces courbes FROC permettent d'extraire un seuil correspondant à une sensibilité moyenne de 90 % sur les images de validation. Cette sensibilité reflète la pertinence de l'extraction initiale des candidats potentiels. Le nombre de FPI reste cependant trop élevé pour de futures applications cliniques et nécessite d'être réduit.

L'étape suivante consiste donc à réduire le nombre de fausses détections. Comme expliqué au paragraphe 5.2.2, deux méthodes de RFP permettant de nettoyer les cartes binaires sont proposées et évaluées. Les cartes binaires ainsi optimisées sont ensuite comparées puis combinées pour aboutir au système CAD complet.

Optimisation du nettoyage initial des cartes binaires

Les deux méthodes de réduction de faux positifs proposées sont 1) l'élimination des monovoxels considérés en majorité comme du bruit et 2) la conservation des tumeurs candidates les plus suspectes. La pertinence de ces deux étapes repose sur leur capacité à effectivement nettoyer les

cartes binaires par élimination des fausses détections sans pour autant trop pénaliser la sensibilité. L'ensemble des résultats présentés dans cette partie provient des 25 images de validation.

La Figure 5-20 présente le rang en pourcentage de chaque cluster VP détecté en fonction de l'ensemble des clusters détectés avec monovoxels (a) et sans monovoxels (b-d). Rappelons que ce rang correspond au classement de la force affectée au candidat par rapport à l'ensemble des scores de tous les candidats obtenus par classification comme expliqué au paragraphe 5.2.2.3. Les Figure 5-20a, b et d correspondent aux rangs obtenus lorsque le maximum des scores des voxels d'un candidat est considéré et la Figure 5-20c correspond aux rangs obtenus lorsque la moyenne des scores des voxels constituant un candidat est affectée.

Les résultats pour le LDA (a, b, c) confirment l'intérêt d'éliminer les monovoxels comme nettoyage des cartes binaires. En effet, la Figure 5-20a présente de nombreux candidats VP mal classés (rang > 40%) comparé à la Figure 5-20b. Dans le premier cas, les FP monovoxels conservés affectent le rang des candidats VP en les décalant à la fin du classement. Ces FP monovoxels, présentant un rang élevé, sont liés à de petits blobs de forte captation en 18F-FDG dus au bruit Gaussien affectant les images TEP. La Figure 5-20b semble donc plus adaptée à la sélection de candidats VP par leur classement. Une tendance apparaît clairement et place une majorité des tumeurs VP candidates en premières positions (rang < 20%). Pour citer un exemple, 98% des candidats VP des poumons se situent dans les 40% plus forts candidats. De la même manière, 94% des candidats VP du foie sont dans les 40% des plus hauts scores, 97% des candidats des tissus mous sont dans les 35% des plus hauts scores.

Toujours pour le LDA, la Figure 5-20c présente le classement en pourcentage de chaque candidat VP sans monovoxel et en utilisant comme force la moyenne des scores des voxels le constituant. Une comparaison avec la Figure 5-20b montre une allure quasiment identique des deux diagrammes avec une majorité des VP bien classée. La seule différence repose sur le rang des candidats les plus mal classés (rang > 40%) pour lesquelles la méthode utilisant le maximum semble plus efficace. Nous avons donc décidé d'assigner le maximum des voxels du cluster comme score du candidat.

La Figure 5-20d présente des résultats pour les SVM, avec une force calculée comme le maximum des voxels du cluster, plus balancés avec un nombre de VP mal classés plus important.

Les Figure 5-20b et d permettent de fixer empiriquement le rang seuil afin de sélectionner une majorité des clusters VP (rang > seuil) et d'enlever les FP les plus faibles (rang < seuil). Ces pourcentages seuils sont fixés à **40% pour les poumons et le foie et 35% pour les tissus mous avec le LDA et à 40% pour tous les organes avec le SVM.**

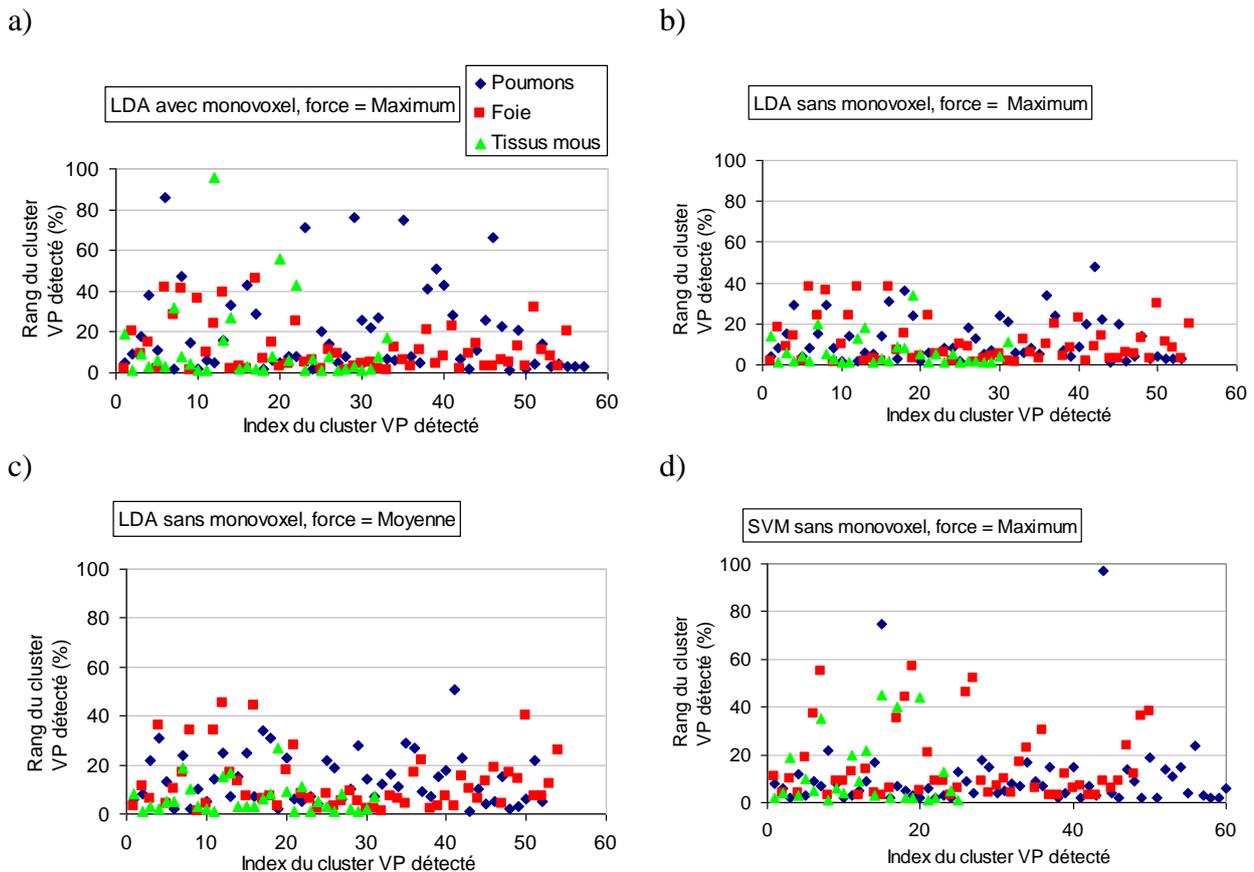


Figure 5-20 : Classement de chaque cluster VP détecté par rapport à tous les autres candidats détectés pour le LDA en considérant a) le maximum du cluster comme score de groupe et en comptant les monovoxels, b) en enlevant les monovoxels, c) en considérant la moyenne du cluster comme score et sans les monovoxels, d) pour le SVM en considérant le maximum du cluster et sans les monovoxels.

5.5.4 Evaluation du système CAD final

5.5.4.1 Introduction

Cette section présente une évaluation quantitative et qualitative du CAD final. Les performances des deux classifieurs avant et après les étapes de nettoyage sont comparées avant d'être combinées. L'évaluation des cartes binaires finales obtenue repose sur des valeurs de sensibilité et nombre de FPI. Elles sont ensuite comparées aux performances humaines par l'intermédiaire de deux mises en situation du CAD. L'ensemble des valeurs données dans cette section ont été extraites des images de test.

5.5.4.2 Comparaison des performances des classifieurs

Résultats quantitatifs

Le Tableau 5-7 présente l'évolution des couples sensibilité/FPI avant et après chaque étape de nettoyage des cartes binaires pour chacun des classifieurs et chacun des organes en considérant les images de test.

Les résultats présentés dans la première ligne du Tableau 5-7 montrent une différence supérieure à 10% en termes de FPI entre les sorties brutes des deux classifieurs pour les poumons. Les tissus mous et le foie sont quant à eux particulièrement favorisés par les SVM avec des taux de FPI très inférieures à ceux du LDA (165 vs. 89 FPI et 72 vs. 32 FPI).

Tableau 5-7 : Evolution des performances après chaque étape de la RFP en termes de sensibilité/FPI

	LDA			SVM		
	Poumons	Foie	Tissus mous	Poumons	Foie	Tissus mous
Sortie du classifieur	82%/88	92%/72	84%/165	90%/78	89%/32	81%/96
Après l'élimination des monovoxels	81%/42	89%/54	81%/67	87%/63	89%/31	81%/94
Après la sélection par le score	70%/34	85%/33	74%/46	83%/30	80%/12	69%/51

Selon la deuxième ligne du Tableau 5-7, l'élimination des monovoxels est plus efficace pour le LDA que pour le SVM. La diminution du nombre de FPI correspond à un facteur de 1,5 (pour le foie) à 2,5 (pour les tissus mous) pour le LDA, tandis que ce nombre ne varie quasiment pas pour le SVM. Cette différence peut s'expliquer par le niveau de décomposition en ondelettes j utilisé pour l'extraction du vecteur V de caractéristiques. Tandis que le modèle prédictif du SVM est optimisé par l'utilisation d'un large intervalle de niveaux d'échelle, jusqu'à une échelle très dilatée $j=5$, celui du LDA est favorisé par une échelle compactée $j=2$ (avec ou sans AG). Les agrégats de bruit détectés par les classifieurs apparaissent de taille réduite et souvent monovoxels pour $j=2$ et un peu plus dilatés pour le SVM et $j=5$. En ce qui concerne le LDA, une moyenne de 65% de toutes les tumeurs candidates initiales pour les tissus mous, 43% pour le foie et 42% pour les poumons sont des monovoxels. Parmi tous ces monovoxels, une majorité provient du bruit Gaussien entachant les images. Seule une moyenne de 0,08% d'entre eux pour les tissus mous, 0,1% pour le foie et 0,4% pour les poumons correspondent à des VP.

L'analyse de la troisième ligne du Tableau 5-7 montre que la sélection des candidats par leur rang diminue le nombre de FPI initial par 2 tout en diminuant la sensibilité d'environ 10% pour chacun des organes et classifieurs. Cette diminution apparaît cependant trop importante en vue d'une future combinaison des cartes binaires, même si le nombre de FPI est nettement amélioré. Ces performances sont peut être dues au choix empirique du rang seuil qui est le compromis entre un nettoyage efficace et la conservation d'une sensibilité élevée.

En considérant que la combinaison des classifieurs diminuera certainement aussi la sensibilité, ce deuxième nettoyage des cartes binaires provenant des deux classifieurs n'est pas pris en compte pour la suite.

Résultats qualitatifs

La Figure 5-21 présente des coupes axiales de cartes binaires superposées aux images originales correspondantes et considérant une tumeur de faible contraste, respectivement 30% et 10% de sensibilité humaine, dans les poumons (a) et dans le foie (b). La première ligne montre l'image d'origine, tandis que les lignes suivantes portent sur les sorties brutes du LDA et du SVM sans nettoyage. L'analyse visuelle de ces deux dernières lignes indique que les deux classifieurs détectent correctement les mêmes tumeurs (sensibilité requise de 90%), mais génère une distribution spatiale de FP différente. Il semble donc intéressant de combiner les deux cartes binaires obtenues en parallèle par les deux classifieurs afin d'éliminer ces fausses détections non spatialement décorréelées.

Justifications numériques de la combinaison par vote à la majorité

Afin d'appuyer cette remarque, nous avons observé la densité des voxels faux positifs détectés par chaque classifieur en considérant le volume total des trois organes cibles que sont les poumons, le foie et les tissus mous. Ce volume d'intérêt contient un total de 258416 voxels.

Nous sommes revenus dans ce paragraphe à la notion de voxels, plutôt que celles d'amas ou clusters précédemment utilisés pour le calcul des performances quantitatives (courbes FROC, sensibilité et FPI), afin de nous projeter dans la méthode de combinaison par vote à la majorité des cartes binaires. Cette combinaison correspond en effet, dans notre cas où seulement deux classifieurs sont utilisés, à réaliser un 'et logique' entre chaque voxel des deux cartes binaires. Cela revient en quelque sorte à calculer une carte d'intersection des deux cartes binaires originalement obtenues des deux classifieurs.

On considère pour ces calculs les 25 images simulées de test. Une moyenne de 690 voxels pour les cartes binaires sorties du LDA et de 1262 voxels pour celles des SVM, toutes deux après élimination des monovoxels, sont classées faux positifs (donc mal détectée). La combinaison des cartes binaires provenant du LDA et du SVM (après élimination des monovoxels) permet de diminuer le nombre de voxels détecté à une moyenne de 113 voxels faux positifs. Le nombre de voxels détectés décroît donc de 83% par rapport aux cartes de LDA sans monovoxels et de 91% par rapport aux cartes de SVM sans monovoxels. Cette diminution significative du nombre de voxels faux positifs valide la présence d'une distribution spatiale des fausses détections décorrélée (l'intersection des cartes élimine les FP). La sensibilité est quant à elle supérieure à 80 % (deuxième ligne du le Tableau 5-7) pour les deux classifieurs et a donc de forte chance de ne pas être trop diminuée par la combinaison de classifieurs. Ces deux raisons confirment l'intérêt de la combinaison par vote à la majorité des cartes binaires.

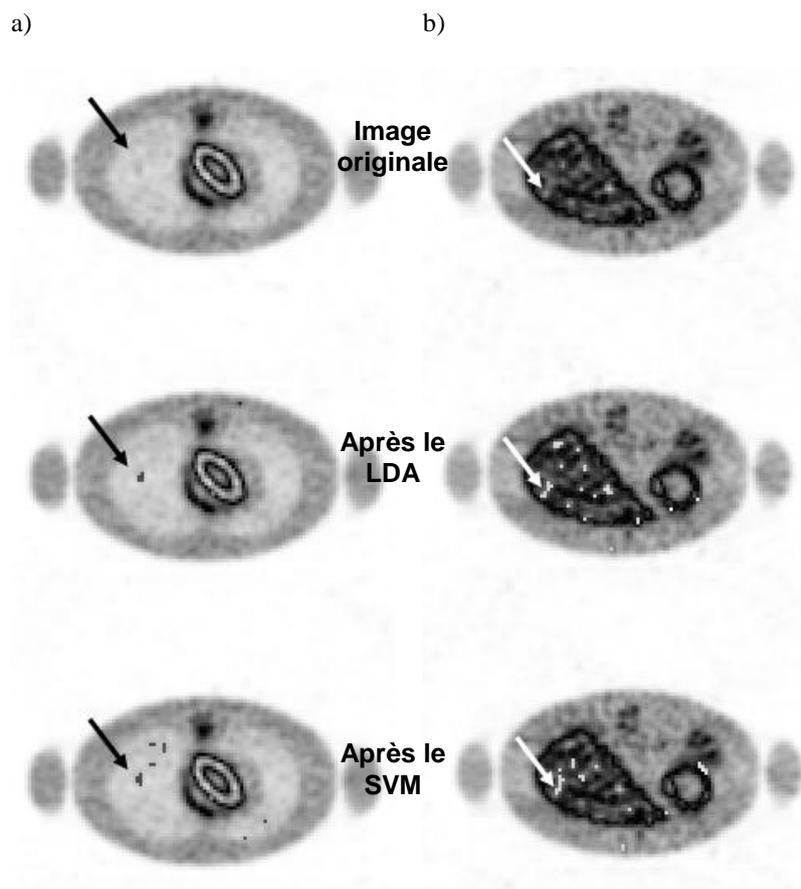


Figure 5-21 : Coupes axiales des cartes binaires superposées aux images originales pour une tumeur dans le poumon (a) et le foie (b), provenant du LDA (au milieu) et du SVM (en bas). L'image originale est présentée (en haut) avec les tumeurs fléchées.

5.5.4.3 Performances finales du système CAD développé

Performances quantitatives et qualitatives du système final

Après nettoyage et combinaison des cartes binaires simultanément extraites des classifieurs, les performances (sensibilité/FPI) obtenues apparaissent prometteuses avec 79%/17 FPI pour les poumons, 81%/17 FPI pour le foie et 76%/25 FPI pour les tissus mous. La sensibilité finale obtenue est généralement autour de 80% tandis que le nombre de FPI est diminué d'un facteur 5 à 8.

La Figure 5-22 permet de visualiser l'amélioration impliquée par l'élimination des monovoxels et la combinaison par vote à la majorité par rapport aux sorties brutes des classifieurs. Un ensemble de vues 3D d'une image contenant sept tumeurs superposée aux cartes binaires correspondantes pour le LDA (a), le SVM (b) et le CAD final (c) est présenté. Les flèches rouges

correspondent à des tumeurs correctement détectées tandis que les blanches correspondent aux tumeurs manquées. Les tumeurs ont des contrastes variés et leur centre n'est pas localisé sur la même coupe. Le Tableau 5-8 résume les contrastes et localisations des tumeurs présentées sur la Figure 5-22. Les tumeurs sont numérotées de haut en bas sur la coupe axiale. Les contrastes sont exprimés par rapport aux performances de détection humaines qui sont à l'origine de leur calibration comme le précise l'Annexe C. L'analyse visuelle des cartes binaires prouve la nette amélioration des performances liées au nettoyage des monovoxels et à la combinaison des cartes. On observe en effet une diminution des fausses détections tandis que six tumeurs sur les sept ciblées sont conservées.

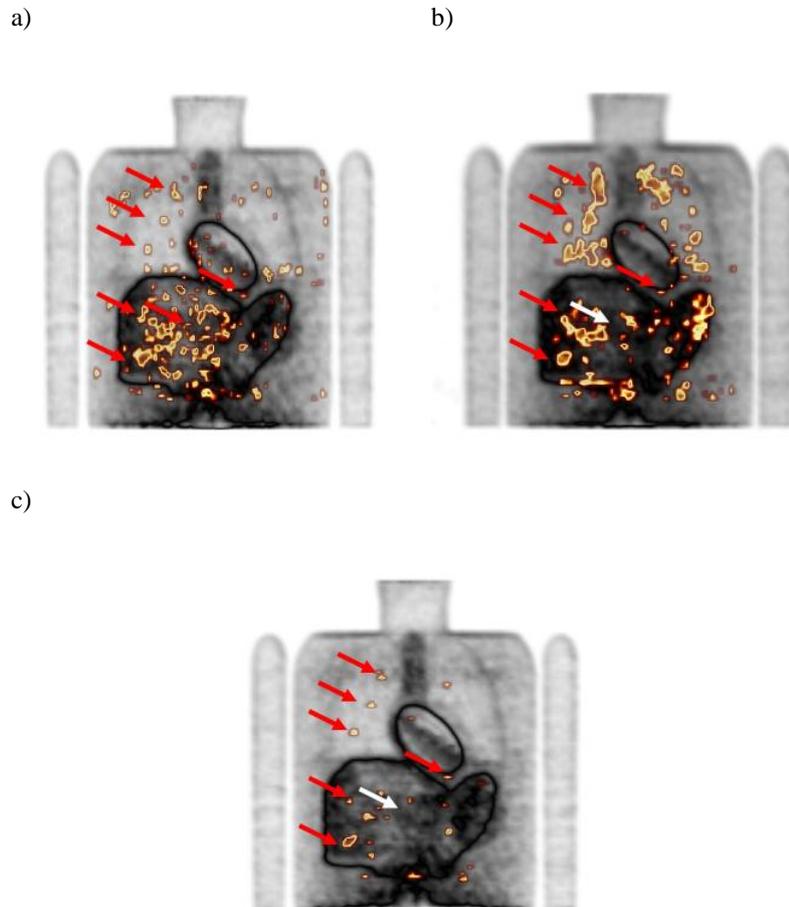


Figure 5-22 : Vue 3D d'une image TEP simulée corps entier contenant 7 tumeurs fléchées superposée à sa carte binaire sortie du LDA (a), du SVM (b) et après les l'élimination des monovoxels et la combinaison par vote à la majorité (c). Les flèches blanches indiquent les tumeurs manquées tandis que les rouges indiquent les tumeurs détectées.

Tableau 5-8 : Contrastes et organes cibles pour chaque tumeur présentée sur l'image TEP corps entier de la Figure 5-22. Les tumeurs sont présentées de haut en bas sur l'image cible et les contrastes se rapportent aux performances humaines leur correspondant.

Tumeur	1	2	3	4	5	6	7
Organe	Poumons	Poumons	Poumons	Tissus mous	Foie	Foie	Foie
Contraste	30%	70%	90%	50%	10%	50%	90%

Comparaison du CAD final avec les performances humaines

Le paragraphe précédent a prouvé l'efficacité globale du système CAD sur des images simulées simplifiées comme en témoignent les cartes volumiques binaires nettoyées. Les performances du CAD sont maintenant comparées à celles obtenues lors de l'étude de détection humaine réalisée sur les mêmes images.

A) Sensibilité du système CAD par contraste

La Figure 5-23 présente la sensibilité obtenue par le CAD final moyennée sur les 25 images de test en considérant les poumons (hachures diagonales), le foie (hachures horizontales) et les tissus mous (hachures verticales) et chaque contraste séparément. Les sensibilités humaines de référence (10%, 30%, 50%, 70% et 90%) ayant permis de fixer chaque contraste sont données en abscisse tandis que celles du CAD sont présentées en ordonnée. La Figure 5-23 montre que le CAD surpasse les sensibilités humaines utilisées pour fixer les contrastes des tumeurs, à l'exception des tissus mous pour une sensibilité humaine de référence de 50%. Il faut également noter que les performances du CAD ne sont pas croissantes avec les sensibilités humaines de référence. On s'attend à ce qu'une tumeur soit plus facilement détectée à contraste supérieur. Or ce n'est pas toujours le cas. Les tissus mous et le foie présentent par exemple une sensibilité plus basse pour une référence de 50% que pour 30%. Ces différences peuvent s'expliquer par le bruit environnant et la localisation de la tumeur. Une tumeur de plus haut contraste peut donc être plus difficilement détectée qu'une tumeur de plus faible contraste. De plus, les performances obtenues pour les différentes tumeurs sont également liées à sa position dans l'organe et au possible effet de volume partiel affectant cet objet de taille réduite et dévalant son activité initiale. Les poumons, qui possèdent un fond beaucoup plus lisse, présentent par ailleurs une des performances croissantes avec le contraste des tumeurs.

On constate que la sensibilité du CAD pour les tumeurs de faible contraste, correspondant à des performances de détection humaine de 10% et 30%, est nettement supérieure aux performances humaines. Une sensibilité supérieure à 57% est observée pour des tumeurs de plus petit contraste localisées dans les poumons, le foie et les tissus mous et atteint 88% pour des tumeurs des tissus mous correspondant à des performances humaines de détection de 30%. Ce point précise l'intérêt du système CAD qui permet d'aider le clinicien pour la détection des cas les plus problématiques.

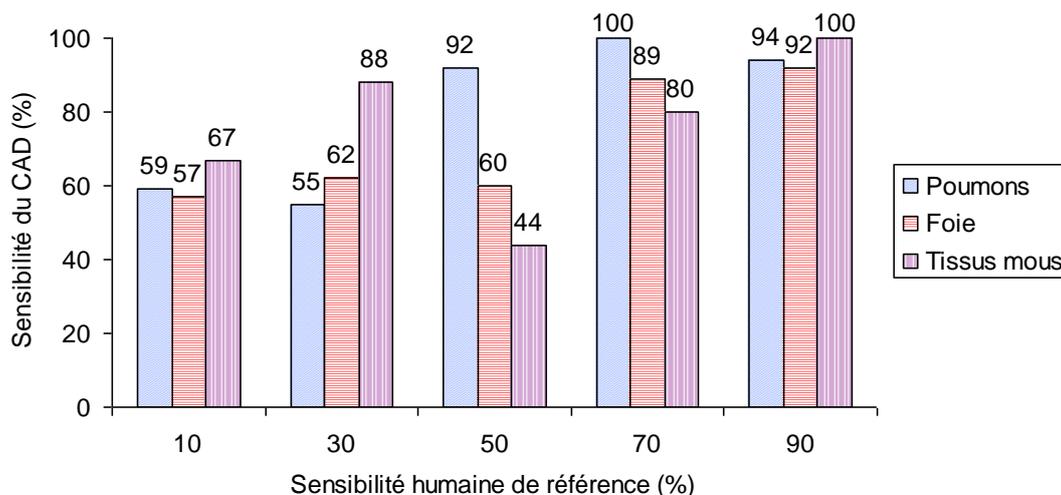


Figure 5-23 : Graphes présentant les performances en sensibilité du système CAD complet pour chaque contraste et organe comparé aux performances humaines correspondantes.

B) Mise en situation du système CAD

Comme expliqué au paragraphe 5.4.3, le système CAD est placé dans les mêmes conditions que les observateurs humains lors de la calibration des contrastes des tumeurs simulées. Le système CAD doit choisir les N=7 tumeurs candidates les plus suspectes parmi toutes celles détectées afin de reproduire le plus fidèlement la tâche de détection de l'observateur humain.

La Figure 5-24 illustre les sensibilités obtenues par les différentes combinaisons carte binaire/carte de score pour les poumons (a), le foie (b) et les tissus mous (c). Les combinaisons (CAD/LDA) et (CAD/SVM) sont d'une part à comparer aux performances humaines, d'autre part à comparer entre elles. Elles proposent en effet une comparaison de la capacité de chacun des classifieurs à donner les scores les plus élevés aux vraies tumeurs.

Les performances pour le foie et les poumons sont généralement meilleures que pour les tissus mous. La configuration (CAD/SVM) paraît plus adaptée que la configuration (CAD/LDA) pour le

foie, excepté pour de faibles contrastes. La configuration (CAD/LDA) conduit à de meilleurs résultats pour les poumons, tandis que les tissus mous sont favorisés par la configuration (CAD/SVM). Cependant, ces observations varient notablement avec le contraste des tumeurs. Il est donc assez difficile d'extraire une logique au niveau des performances de classification.

Généralement le LDA (configuration CAD/LDA) est plus performant que le SVM pour les tumeurs de petit contraste. Remarquons que toutes les configurations sont limitées à une sensibilité de 80%. Elles présentent des performances inférieures aux observateurs humains pour les tumeurs de fort contraste mais équivalentes ou supérieures pour les tumeurs de petit contraste. Comme précédemment, l'utilité du CAD apparaît au travers de ses performances pour les tumeurs apparaissant problématiques pour l'œil humain (10% et 30% de détectabilité humaine).

5.5.5 Conclusions sur l'évaluation du système CAD par les données simulées simplifiées

Le système CAD développé dans ce manuscrit pour les images TEP 3D oncologiques se base sur une combinaison de deux classifieurs : le LDA et le SVM. Les paramètres du CAD, intrinsèques et extrinsèques aux deux classifieurs, ont été optimisés. La carte volumique binaire de détection fournie par ce système affiche les foyers potentiellement pathologiques qui doivent attirer l'œil du clinicien. Nous avons détaillé les étapes d'optimisation et l'étude de faisabilité de ce système à partir d'une base d'images TEP simulées simplifiée. L'ensemble de cette méthode et son étude de faisabilité à l'aide des images simulées simplifiées ont été soumis sous la forme de deux articles au journal IEEE Transactions on Medical Imaging [Tomei et al., 2009; Tomei et al., 2009].

On peut noter que pour une utilisation potentielle du CAD dans les conditions cliniques, les méthodes développées doivent être les plus rapides possibles. Les développements concernant l'extraction de caractéristiques et le LDA ont été réalisés à l'aide de Python. Concernant les SVM, l'apprentissage et le test ont été réalisés à l'aide de l'implémentation python de la librairie libre *libSVM* [Chang and Lin, 2001]. La sélection de caractéristiques se base quant elle sur l'outil Java *RapidMiner* [Mierswa et al., 2006], initialement dédié aux algorithmes de datamining. Actuellement, le temps nécessaire pour obtenir une carte binaire à partir d'une image 128x128x108 voxels est de 2 min pour le LDA et 6 min pour le SVM. Ces temps pourraient certainement être améliorés par une implémentation en C++.

Les performances du CAD final apparaissent prometteuses avec des couples sensibilité/FPI de 79%/17 pour les poumons, 81%/17 pour le foie et 76%/25 pour les tissus mous. Cependant, le taux de FPI est encore trop élevé et la sensibilité un peu trop faible pour de futures applications cliniques. Remarquons cependant que le partitionnement de la base d'images suit la méthode de hold out impliquant une surestimation des erreurs de classification [Sahiner et al., 2000]. Cette raison unique ne peut cependant pas totalement expliquer le nombre de FPI encore trop élevé pour des applications cliniques. Les sensibilités obtenues apparaissent aussi un peu trop basses par rapport à celles actuellement atteintes en imagerie TEP clinique au 18F-FDG pour de nombreux cancers. Pour ne citer qu'un exemple, l'utilisation des images TEP assure une sensibilité de 89% pour le lymphome [Stumple et al., 1998]. Cette valeur est obtenue à partir d'une étude rétrospective de 71 patients imagés par TEP et présentant des lymphomes de bas et hauts grades.

Comme nous l'avons souligné dans ce manuscrit, le système CAD développé n'aspire pas à remplacer le clinicien, mais plutôt à le compléter. Ainsi, seule une étude ROC humaine utilisant des images TEP au 18F-FDG avec ou sans l'aide du système CAD pourrait évaluer son influence réelle au niveau clinique. En effet, les résultats présentés sur les Figure 5-23 et Figure 5-24 semblent indiquer que le système CAD donne des performances tout à fait satisfaisante pour les tumeurs problématiques de faible contraste.

En attendant cette potentielle étude ROC humaine, l'insertion d'informations additionnelles quantifiant le degré de suspicion de chaque tumeur candidate pourrait enrichir l'exploitation des cartes binaires provenant du système CAD. Cette éventualité est présentée dans le chapitre suivant.

Dans ce chapitre, les performances du CAD ont également été comparées aux performances humaines sur ces mêmes images. Le CAD apparaît plus performant que les observateurs humains pour les tumeurs de faible contraste, c'est-à-dire correspondant à 10 ou 30% de détectabilité humaine.

L'ensemble de ces performances n'ont cependant pas pu être comparées par rapport à d'autres CAD dédiés à l'imagerie TEP. Ceci peut notamment être expliqué par le nombre limité de systèmes existants pour ce type d'imagerie et au manque de telles évaluations quantitatives dans la littérature (cf. paragraphe 2.2). Une comparaison des performances apparaît d'autant plus difficile que notre évaluation repose sur une approche 3D de manière à représenter concrètement l'aspect visuel des cartes binaires, ce qui n'est pas le cas dans la littérature actuelle.

Le chapitre suivant évalue également la pertinence du système CAD sur les images d'oncoPET_DB, plus réalistes d'un point de vue anatomique et variabilité des caractéristiques tumeurs.

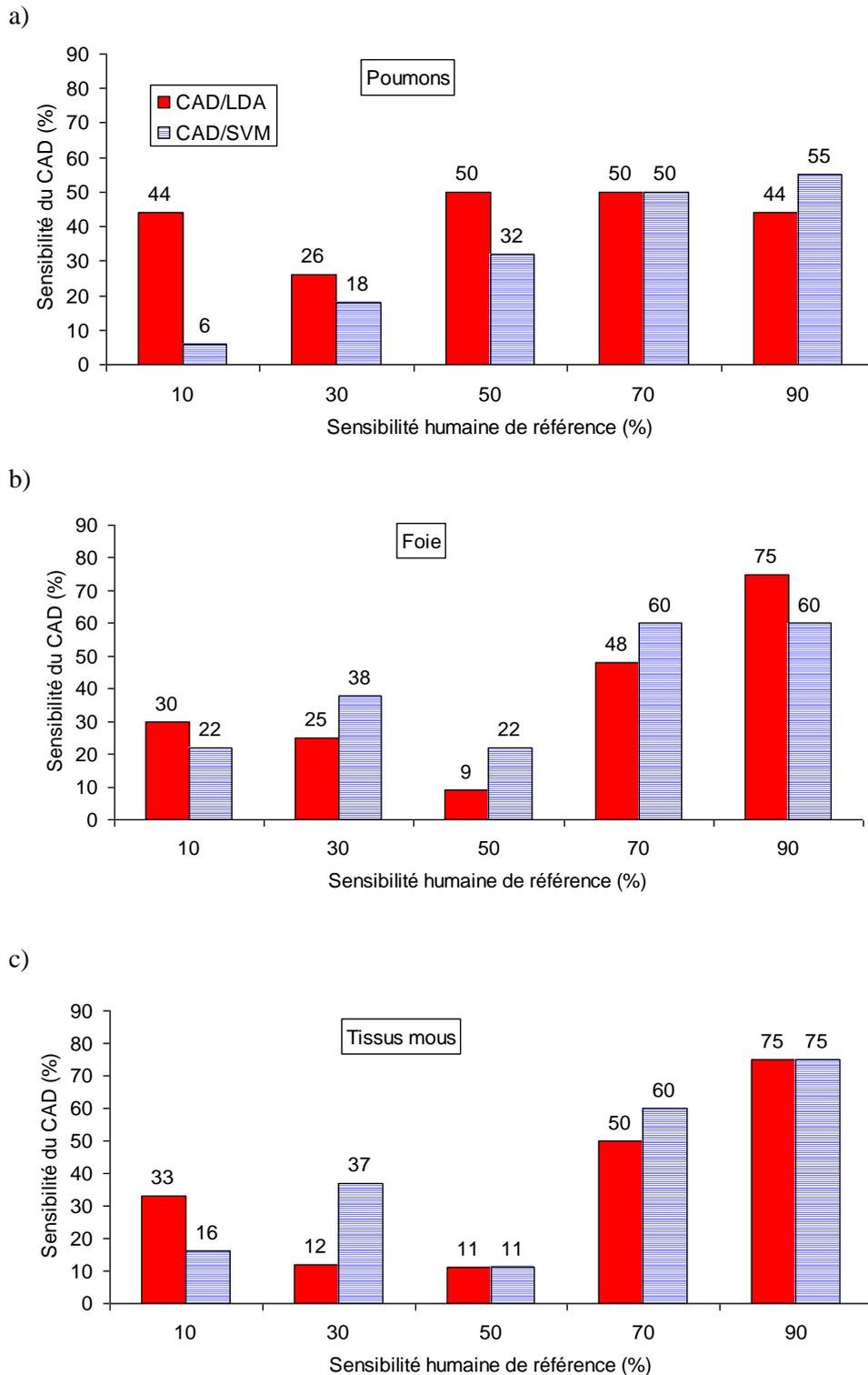


Figure 5-24 : Graphes montrant la sensibilité moyenne du système CAD pour une tâche de sélection des sept tumeurs les plus suspectes par image. Chaque contraste, définie par la sensibilité humaine obtenue pour la même tâche et chaque organe, les poumons (a), le foie (b) et les tissus mous (c) sont traités.

Chapitre 6 Evaluation du système CAD sur la base d'images simulées oncoPET_DB

6.1 Introduction

Le chapitre précédent a proposé un système CAD pour l'imagerie TEP 3D et a décrit les résultats d'une première étude de faisabilité sur des images simulées provenant d'un modèle humain simplifié. Cette étude a mis en évidence la limite du système actuel notamment à cause d'un nombre de FPI trop important pour de futures applications cliniques. Nous avons avancé l'intérêt d'ajouter une information de score modélisant la probabilité d'appartenance à la classe 'tumeur' de chaque candidat.

Ce chapitre présente la synthèse des contributions de ce travail de thèse en combinant le système CAD aux images simulées d'oncoPET_DB. L'exploitation d'images plus réalistes et qui contiennent des tumeurs de contraste mais aussi de diamètre variés permet d'évaluer les performances du CAD dans des conditions proches de la réalité clinique et d'envisager les adaptations à y apporter. Cette évaluation repose sur une méthodologie similaire à celle détaillée dans le Chapitre 5. Cependant, nous consacrons une part plus importante de notre étude à l'analyse des résultats qualitatifs. Les tendances générales des performances obtenues sur chaque type d'images sont comparées afin de tenter d'extraire un comportement commun. L'utilisation de cartes paramétriques de score au lieu de cartes binaires de détection à la sortie du système est enfin présentée.

6.2 Caractéristiques des images d'oncoPET_DB

Les caractéristiques d'oncoPET_DB ont été présentées de manière détaillée au Chapitre 4. Parmi les 100 images disponibles, nous n'en utilisons que 25 pathologiques et 25 saines. Nous nous focalisons seulement sur les images pathologiques contenant 10 tumeurs chacune, même si celles-ci ne respectent pas la distribution statistique des cas cliniques du lymphome. Elles présentent en effet une répartition plus adaptée à l'apprentissage des tumeurs dans les différents organes, présentant une distribution équivalente des tumeurs dans ces organes (cf. paragraphe 4.6). La distribution des caractéristiques des tumeurs cibles en termes de diamètre, d'organe et de contraste est donc mieux représentée, ce qui rend l'analyse quantitative des performances par groupe de caractéristiques plus robuste statistiquement. Les résultats quantitatifs portant sur la rate ne sont cependant pas présentés étant donné que cet organe est faiblement représenté en nombre de tumeurs dans les 25 images pathologiques (cf. paragraphe 4.6).

Un total de 25 images pathologiques contenant 10 tumeurs chacune et 25 images saines est donc disponible. Ces tumeurs cibles se répartissent en trois organes, trois diamètres et cinq contrastes. Comme au Chapitre 5, notre stratégie d'évaluation repose sur une méthode par *hold-out*. Ainsi la répartition de l'ensemble des images pathologiques est réalisée comme suit :

- Les tumeurs d'apprentissage, utiles au calcul des modèles prédictifs, sont extraites au centre des tumeurs de 12 images pathologiques et aux coordonnées correspondantes dans les 12 images saines. Chaque image pathologique contient 10 tumeurs sphériques, soit un total de 120 positions, de contraste et de diamètres variés réparties aléatoirement dans les poumons, le foie, la rate et les ganglions.
- La base de validation nécessaire à l'optimisation des paramètres extrinsèques au classifieur du système CAD est constituée des mêmes 12 images pathologiques précédemment décrites.
- La base de test permettant le calcul des performances finales du système correspond à 13 images pathologiques contenant chacune 10 tumeurs réparties dans les mêmes organes cibles que précédemment. Ces images sont différentes de celles précédemment utilisées pour l'apprentissage et la validation.

Tableau 6-1 rappelle les différents contrastes calibrés pour chaque organe et chaque diamètre tandis que les Tableau 6-2 et Tableau 6-3 résument les caractéristiques des tumeurs disponibles pour l'apprentissage et celles disponibles pour le test.

Tableau 6-1 : Gamme de contrastes définie pour des tumeurs situées dans chaque organe de la base d'images simulées réalistes à partir des performances de détection humaine.

Performances humaines (%)		10	30	50	70	90
Poumons	7 mm	2	5	8	10	13
	10,5 mm	2	5	8	10	13
	14 mm	2	3.5	4.5	7.5	10
Foie	7 mm	2.5	4	4.5	6	9
	10,5 mm	2	3	4	4.5	7.5
	14 mm	2	3	4	4.5	7.5
Ganglions	7 mm	2.5	3	3.5	4	5.5
	10,5 mm	2.5	3	3.5	4	5.5
	14 mm	1.5	2	2.5	3	4

Tableau 6-2 : Répartition du nombre de tumeurs par organe pour chaque diamètre et contraste dans les 12 images d'apprentissage et de validation. Le contraste est défini par rapport à la sensibilité humaine correspondante. Seuls les poumons, le foie et les ganglions sont ciblés.

Performances humaines (%)		10	30	50	70	90
Poumons	7 mm	1	3	3	6	2
	10,5 mm	2	1	8	1	2
	14 mm	0	1	1	7	1
Foie	7 mm	3	3	4	2	2
	10,5 mm	6	1	1	2	1
	14 mm	2	1	2	1	3
Ganglions	7 mm	4	4	5	3	6
	10,5 mm	0	1	0	3	1
	14 mm	1	3	1	1	2

Tableau 6-3 : Répartition du nombre de tumeurs par organe pour chaque diamètre et contraste dans les 13 images de test. Le contraste est défini par rapport à la sensibilité humaine correspondante. Seuls les poumons, le foie et les ganglions sont ciblés.

Performances humaines (%)		10	30	50	70	90
Poumons	7 mm	5	3	5	0	2
	10,5 mm	0	4	2	3	3
	14 mm	3	2	2	1	1
Foie	7 mm	2	1	6	8	5
	10,5 mm	0	1	6	0	2
	14 mm	3	1	3	2	1
Ganglions	7 mm	4	5	3	2	3
	10,5 mm	2	5	3	0	3
	14 mm	2	5	0	2	1

Les Tableau 6-2 et le Tableau 6-3 nous permettent de constater que le nombre de tumeurs disponibles dans la base de test pour chaque groupe de caractéristiques (organe/diamètre/contraste) est assez limité. Certaines tumeurs ne sont pas représentées comme le groupe (foie/10,5mm/70%). Cette répartition aléatoire traduit néanmoins davantage la réalité clinique. Il aurait ainsi peut être été plus pertinent de sélectionner chaque volume de la base d'apprentissage et de la base de test de manière à couvrir l'ensemble des groupes de caractéristiques de tumeurs.

Plusieurs choix s'imposent au vue de ce tableau. D'une part, nous ne présentons pas dans la suite les performances quantitatives globales de la rate à cause d'une statistique trop réduite. D'autre part, l'analyse des performances globales du CAD selon les contrastes des tumeurs sont exploités à l'aide d'une méthodologie un peu différente de celle du Chapitre 5. La méthodologie utilisée dans ce chapitre est présentée dans le paragraphe suivant.

6.3 Rappels sur le système CAD et méthodes d'évaluation

6.3.1 Rappels sur le système CAD

Le CAD utilisé dans ce chapitre suit la méthodologie présentée au Chapitre 5. Le schéma de la Figure 6-1 rappelle les principales étapes du système.

L'extraction initiale des tumeurs candidates utilise la combinaison de coefficients d'ondelettes (point 1 de la Figure 6-1) et de deux classifieurs (le LDA et le SVM) en parallèle (point 2 de la Figure 6-1). Comme précédemment, le seuillage des cartes de score simultanément fournies par les deux classifieurs repose sur des valeurs de seuil extraites de courbes FROC tracées sur les images de validation. Ces seuils correspondent à une sensibilité de 90 % et sont générés pour chaque classifieur et chaque organe.

La réduction des faux positifs visibles sur les cartes de détection obtenues repose ensuite sur l'élimination des candidats monovoxels (point 3 de la Figure 6-1). Les cartes binaires provenant du LDA et des SVM ainsi nettoyées sont ensuite combinées de manière à profiter de la décorrélation spatiale des fausses détections pour les réduire. La sélection des candidats par leur force, présentée au paragraphe 5.2.2.3 comme méthode de nettoyage des cartes binaires avant leur combinaison, n'est pas exploitée dans ce chapitre. Le Chapitre 5 a en effet montré qu'elle diminuait trop fortement la sensibilité.

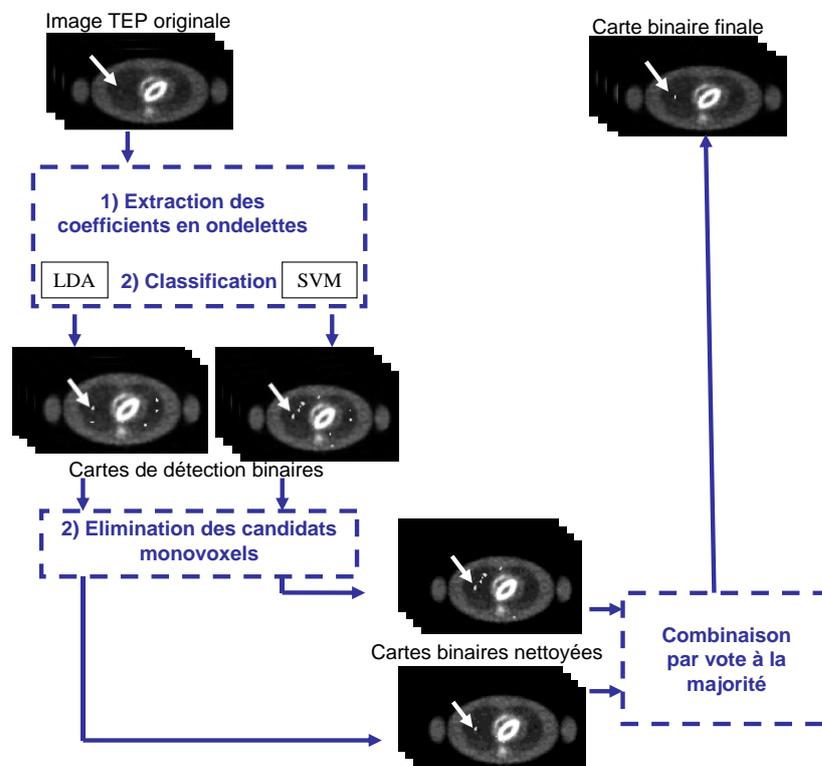


Figure 6-1 : Schéma bloc rappelant les étapes du système CAD développé

6.3.2 Méthodes d'évaluation du CAD sur les images d'oncoPET_DB

Les étapes d'optimisation largement présentées au Chapitre 5 ne sont pas répétées dans ce chapitre. Les mêmes paramètres que précédemment en ce qui concerne l'optimisation de l'apprentissage par le choix des tumeurs cibles et du vecteur descripteur de caractéristiques V sont conservés.

6.3.2.1 Performances générales du CAD

Nous nous attachons dans ce chapitre à évaluer les performances quantitatives et qualitatives du système CAD complet. L'utilisation des images réalistes d'oncoPET_DB permet en effet d'analyser les performances du système dans des conditions proches de la réalité clinique. Pour ce faire, nous analysons les valeurs de FPI et de sensibilité provenant des images binaires extraites de

chaque classifieur séparément et du CAD complet. D'autre part, l'aspect visuel des cartes binaires provenant du CAD est présenté. Elles correspondent en effet à l'information que le clinicien exploite au cours de son diagnostic. Les cartes binaires doivent être le plus propres possible (minimisation des FPI) et flécher une majorité des vraies tumeurs. Les cartes binaires sont superposées à leurs images d'origine afin de faciliter leur exploitation.

6.3.2.2 Utilisation d'une information additionnelle de score pour la présentation des résultats

Comme expliqué au paragraphe 6.2, le nombre insuffisant de tumeurs de test par groupe de caractéristiques, présentée dans le Tableau 6-3, implique une exploitation moins précise des performances par organe/diamètre/contraste que celle illustrée par les Figure 5-23 et Figure 5-24 du Chapitre 5. OncoPET_DB a l'avantage de proposer une grande variabilité de tumeurs de contraste et diamètre variés, mais ne permet pas, dans son état actuel, d'offrir un nombre statistiquement significatif de chaque tumeur. Nous discutons de ce point dans le paragraphe 6.5 de ce chapitre.

Le chapitre précédent a postulé la possibilité d'intégrer des informations additionnelles aux tumeurs candidates afin de faciliter la visualisation des images sorties du CAD. Ainsi, la notion de 'score' ou 'force' d'appartenance des candidats dans la classe 'tumeur présente' est étudiée dans ce chapitre. Cette notion est présentée au paragraphe 5.2.2.3 du Chapitre 5 dans le cadre de la sélection des tumeurs les plus suspectes. Comme précédemment, les forces ou scores obtenus par chaque candidat sont classés par ordre décroissant. Ainsi un candidat de force élevée (haut degré de suspicion) apparaît dans les premières places du classement. Remarquons que ce score appartient à un intervalle de valeurs dépendant de l'organe et du classifieur considéré. Le rang obtenu par le classement pour chaque candidat à partir de son score d'appartenance est donc normalisé par le nombre total de candidats de manière à obtenir un classement relatif entre 0 et 100%. Chaque organe est traité séparément. Le score initial d'un candidat est défini comme la valeur maximale des scores des voxels le constituant. Comme au paragraphe 5.2.2.3, les combinaisons possibles de cartes binaires pour extraire les candidats et de cartes de score pour leur affecter une force sont disponibles : CAD/LDA, et CAD/SVM.

Notre évaluation, permettant d'analyser l'insertion des rangs aux cartes binaires de détection, en considérant deux cas, 1) l'étude qualitative de tumeurs VP ciblées par rapport à leur contraste, 2) l'étude quantitative globale des tumeurs correspondant à chacune des gammes de contraste. Dans le premier cas, une tumeur 'vraie positive' par organe cible est sélectionnée pour le diamètre minimum (7mm) et chaque contraste. Le classement de la force de chaque VP par rapport à tous les autres candidats FP de l'organe ainsi que sa représentation en niveau de couleur sur l'image de détection sont présentés. Idéalement, les VP doivent présenter les meilleurs classements de manière à ce que la contamination des cartes binaires de détection, fournies par le système CAD final, par les fausses détections apparaisse moins gênante visuellement. Dans le second cas, le classement de l'ensemble des candidats VP pour chaque gamme de contraste, sans distinction de diamètre, est affiché. L'analyse globale de ces rangs a pour objectif d'extraire une tendance générale en vue de l'insertion de cette information aux cartes de détection fournies par le CAD.

6.4 Evaluation du CAD sur les images d'oncoPET_DB

6.4.1 Exploitation et discussions des performances du CAD complet

6.4.1.1 Performances visuelles et quantitatives des cartes binaires fournies par le CAD

Les performances qualitatives et quantitatives générales du système CAD sont présentées dans ce paragraphe. Elles se basent, comme dans le chapitre précédent, sur une analyse 3D des cartes de détection fournies par le CAD finale. Cette approche originale permet de mieux corrélérer l'aspect visuel des cartes binaires et les taux de fausses détections et de sensibilité réellement obtenus. Les organes cibles sont les poumons, le foie et les ganglions lymphatiques disséminés dans le corps du patient le long des principales artères.

La Figure 6-2 présente des cartes binaires superposées à leur image d'origine en projection MIP pour deux cas de la base de test différents. Les cartes binaires proviennent du LDA seul (a, d), du SVM seul (b, e), toutes deux après élimination des monovoxels, et du CAD final (c, f). Dix tumeurs par image sont ciblées par les flèches, avec une flèche rouge pour une tumeur correctement trouvée et une flèche blanche pour une tumeur manquée. Les caractéristiques des tumeurs cibles, numérotées de haut en bas sur les Figure 6-2a et Figure 6-2d, sont données dans le Tableau 6-4.

Tableau 6-4 : Contrastes et organes cibles pour chaque tumeur présentée sur les images TEP corps entier de la Figure 6-2. Les tumeurs sont présentées de haut en bas et les contrastes se rapportent aux performances humaines leur correspondant.

Tumeur	Image de test 1			Image de test 2		
	Organe	Diamètre	Contraste	Organe	Diamètre	Contraste
1	Ganglions	7mm	70%	Ganglions	7mm	10%
2	Poumons	7mm	50%	Ganglions	14mm	30%
3	Ganglions	7mm	90%	Poumons	14mm	50%
4	Poumons	7mm	50%	Poumons	7mm	90%
5	Poumons	7mm	50%	Poumons	7mm	10%
6	Foie	14mm	50%	Poumons	10,5mm	70%
7	Foie	7mm	70%	Foie	7mm	10%
8	Foie	7mm	70%	Ganglions	10,5mm	30%
9	Rate	10,5mm	30%	Ganglions	10,5mm	90%
10	Ganglions	7mm	30%	Ganglions	10,5mm	30%

Quelque soit le classifieur (LDA ou SVM), l'aspect visuel des cartes binaires apparaît d'assez mauvaise qualité pour les poumons, notamment à cause d'un nombre élevé de fausses détections. Le foie et les ganglions, quant à eux, semblent propres et faiblement affectés par les fausses détections. On constate l'élimination de la tumeur 8 de l'image 1, localisée dans le foie, de petit diamètre et de petit contraste. Celle-ci était pourtant visible sur les cartes de détection sorties de chaque classifieur indépendamment. Un inconvénient intrinsèque de la combinaison par vote à la majorité associée à la définition des VP choisie dans ce manuscrit, apparaît ainsi clairement. Ce type de combinaison implique en effet une sensibilité au mieux équivalente à la plus mauvaise des deux classifieurs puisque dans le cas où deux classifieurs sont utilisés, la notion de vote à la majorité s'apparente à un 'ET' logique entre chaque voxel.

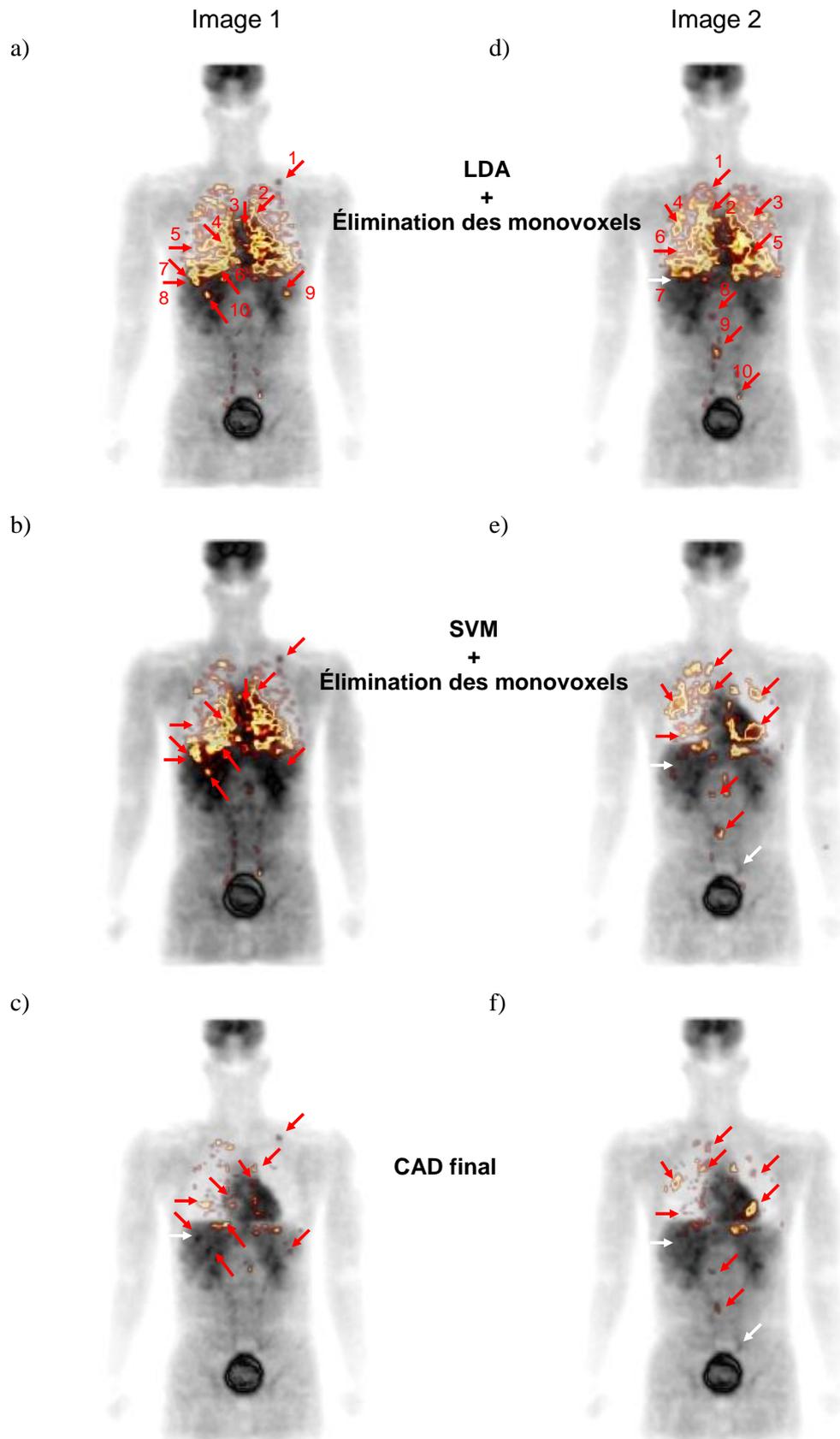


Figure 6-2 : Vue 3D de deux images de test superposées à leurs cartes binaires correspondantes, provenant du LDA suivi de l'élimination des monovoxels (a, d), du SVM sans monovoxels (b, e) et du CAD final (c, f).

Le Tableau 6-5 présente les résultats quantitatifs en termes de sensibilité et FPI obtenus à la sortie de chaque classifieur (deuxième ligne), après l'élimination des monovoxels (troisième ligne) et après la combinaison (quatrième ligne). Le nombre de tumeurs cibles pour chaque organe est définie sur la première ligne du Tableau 6-5.

L'analyse du Tableau 6-5 indique que, comme précédemment, l'élimination des candidats monovoxels permet de notablement diminuer le nombre de FPI (d'un facteur 1,5 à 2,5) pour le LDA. Contrairement aux performances sur la base simulée simplifiée, le SVM est plus influencé par l'élimination des monovoxels. Une diminution du FPI d'un facteur 2 est généralement obtenue pour une sensibilité faiblement affectée.

Le LDA présente de meilleures performances que le SVM avant et après l'élimination des monovoxels pour les poumons et le foie, tandis que les ganglions sont favorisés par le SVM.

Les couples sensibilité/FPI calculés à partir des 13 images de test pour le CAD final apparaissent très mitigés avec une bonne sensibilité pour les poumons associés à un FPI conséquent (81%/57) et une sensibilité beaucoup trop basse pour le foie (54%) et les ganglions (62%), même si le nombre de FPI est minime. Au regard des bonnes performances de chacun des classifieurs pris séparément, on constate que la combinaison résulte en une diminution trop forte de la sensibilité dans le foie (48%) et les ganglions (30%), alors qu'elle est limitée à 10% pour les poumons. On peut aussi remarquer que les performances pour le foie apparaissent meilleures avant la combinaison avec une bonne sensibilité pour un nombre de FPI très restreint.

Si on compare les performances quantitatives du Tableau 6-5 de ce chapitre à celles du Tableau 5-7 du chapitre précédent, on constate que les performances du CAD complet obtenues sur les images simplifiées sont meilleures (sensibilité plus élevée et FPI plus faible). La présence de tumeurs de taille variée dans les images d'oncoPET_DB semble donc faire apparaître de nouvelles limitations de la méthode.

Tableau 6-5 : Evolution exprimée en termes de sensibilité/FPI des performances après l'élimination des monovoxels et la combinaison. Les performances sont moyennées sur les 13 images de test dont le nombre de tumeurs cibles est précisé sur la première ligne du tableau.

	LDA			SVM		
	Poumons	Foie	Ganglions	Poumons	Foie	Ganglions
Nombre de tumeurs	36	41	40	36	41	40
Sortie du classifieur	97%/203	95%/23	95%/53	92%/96	89%/11	100%/31
Après l'élimination des monovoxels	94%/130	92%/8	90%/35	92%/55	84%/4	100%/16
Après l'élimination des monovoxels et la combinaison de classifieurs	CAD final					
	Poumons	Foie	Ganglions			
	36	41	40			
	81%/57	54%/1	62%/2			

6.4.1.2 Discussions des performances fournies par le CAD final

La diminution des sensibilités obtenues par le CAD final extraites du Tableau 6-5 peuvent être expliquées par différents phénomènes que nous reprenons dans ce paragraphe.

Les Tableaux 6-6 et Tableau 6-7 présentent les sensibilités de chacun des classifieurs après élimination des candidats monovoxels (a, b), et avec le CAD final (c) pour les différents contrastes des tumeurs du foie et les ganglions. Le nombre de tumeurs disponibles pour chaque groupe de caractéristiques, extrait du Tableau 6-3, est précisé entre parenthèses dans les Tableaux 6-6 et Tableau 6-7. Les cellules noircies correspondent à une absence de tumeurs de la catégorie correspondante parmi toutes les cibles des images de test. Les cellules surlignées correspondent à des tumeurs généralement bien détectées par le LDA et le SVM, et dont la sensibilité est fortement diminuée par la combinaison des cartes. Pour citer quelques exemples du foie, les tumeurs de diamètre minimum (7mm) et de contraste élevé, correspondant à des performances humaines de 50% et 70%, voient respectivement leurs sensibilités diminuer de à 67% et à 25%. De même, les tumeurs de plus fort diamètre (14mm) ne sont plus du tout détectées après combinaison des classifieurs alors qu'elles l'étaient à 67% par le LDA et le SVM seul. Pour les ganglions, seules deux catégories sont affectées

par une chute de la sensibilité post combinaison. Il s'agit d'une part des tumeurs de plus grand diamètre 14mm et de contraste minimum (10%) qui étaient détectées à 100% pour le LDA et 50% pour les SVM contre 0% pour le CAD final. D'autre part, la sensibilité des tumeurs de diamètre intermédiaire 10,5mm et de contraste 30% est diminuée de 80% pour le LDA et 60% pour le SVM à 40% pour le CAD. Il faut cependant remarquer que ces résultats doivent être observés en considération d'une répartition statistique limite des tumeurs de test dans les différentes catégories (diamètre/contraste).

Tableau 6-6 : Sensibilités en pourcentage par contraste et diamètre pour le foie

(a) extraits des cartes binaires sans monovoxels fournies par le LDA

LDA	10	30	50	70	100
7mm	50 (2)	0 (1)	100 (6)	88 (8)	100 (5)
10,5mm		100 (1)	100 (6)		100 (2)
14mm	67 (3)	100 (1)	100 (3)	100 (2)	100 (1)

(b) extraits des cartes binaires sans monovoxels fournies par le SVM

SVM	10	30	50	70	100
7mm	0 (2)	0 (1)	100 (6)	62 (8)	100 (5)
10,5mm		100 (1)	100 (6)		100 (2)
14mm	67 (3)	100 (1)	100 (3)	100 (2)	100 (1)

(c) extraits des cartes binaires fournies par le CAD complet

CAD	10	30	50	70	90
7mm	0 (2)	0 (1)	67 (6)	25 (8)	100 (5)
10,5mm		0 (1)	100 (6)		50 (2)
14mm	0 (3)	100 (1)	100 (3)	0 (2)	100 (1)

Tableau 6-7 : Sensibilités en pourcentage par contraste et diamètre pour les ganglions

(a) extraits des cartes binaires sans monovoxels fournies par le LDA

LDA	10	30	50	70	100
7mm	25 (4)	60 (5)	100 (3)	100 (2)	100 (3)
10,5mm	50 (2)	80 (5)	100 (3)		100 (3)
14mm	100 (2)	100 (5)		100 (2)	100 (1)

(b) extraits des cartes binaires sans monovoxels fournies par le SVM

SVM	10	30	50	70	100
7mm	75 (4)	80 (5)	100 (3)	100 (2)	100 (3)
10,5mm	100 (2)	60 (5)	100 (3)		100 (3)
14mm	50 (2)	40 (5)		100 (2)	100 (1)

(c) extraits des cartes binaires fournies par le CAD complet

CAD	10	30	50	70	100
7mm	25 (4)	60 (5)	100 (3)	100 (2)	67 (3)
10,5mm	50 (2)	40 (5)	100 (3)		100 (3)
14mm	0 (2)	40 (5)		100 (2)	100 (1)

La combinaison par vote à la majorité associée à la définition des VP faite dans ce manuscrit implique deux inconvénients :

1) L'inconvénient principal intrinsèque à la méthode et dans le cas où seulement deux classifieurs sont utilisés, repose sur l'indexation des performances du CAD final sur celles du plus mauvais classifieur. C'est ce qu'on peut observer dans le Tableau 6-5. Il apparaît même que les sensibilités finales sont souvent inférieures aux plus mauvaises sensibilités de chaque classifieur obtenues sur les cartes binaires. Une raison simple à cela repose sur le fait qu'à sensibilité égale, les

deux classifieurs ne détectent pas forcément les mêmes tumeurs. Les tumeurs discordantes sont donc éliminées par combinaison par vote à la majorité.

2) Le second inconvénient de la méthode est mis en évidence par les Tableau 6-6 et Tableau 6-7. Ces deux tableaux présentent des tumeurs détectées avec une sensibilité de 100% par les deux classifieurs qui sont ensuite éliminées lors de la combinaison. L'hypothèse (1) présentée précédemment n'entre pas en application pour les contrastes élevés, car toutes les tumeurs cibles sont conjointement détectées par les deux classifieurs (sensibilité de 100%). Une analyse visuelle des cartes binaires permet de vérifier l'origine de cette élimination des candidats VP, alors qu'ils sont détectés par les deux classifieurs. La Figure 6-3 cible une tumeur du foie de diamètre 14 mm et de contraste correspondant à une sensibilité humaine de 70% (en gras dans le Tableau 6-6). Les cartes binaires provenant (a) du LDA, (b) du SVM et (c) du CAD final, sont superposées à l'image d'origine. Nous constatons que les deux classifieurs détectent correctement cette tumeur mais pas à travers les mêmes voxels. Le LDA favorise la classification des voxels du bord de la tumeur, tandis que les SVM favorisent plutôt ceux du centre. Ainsi la combinaison des deux cartes binaires par vote à la majorité, correspondant à un 'ET logique' sur chacun de leurs voxels, implique une élimination de cette tumeur pour laquelle aucun voxel n'est conjoint spatialement entre les deux classifieurs. Le même effet est observé sur les tumeurs des ganglions.

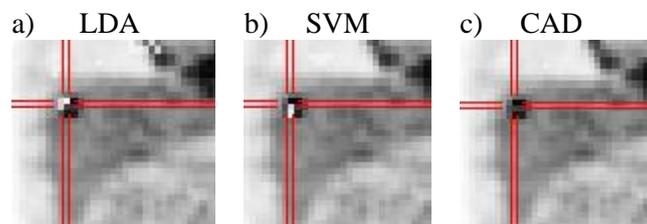


Figure 6-3 : Zoom d'une coupe coronale provenant de la superposition d'une image de test avec une carte seuillée provenant du LDA (a), du SVM (b) et de la combinaison de deux (c). La tumeur ciblée est localisée dans le foie.

De même, la Figure 6-4 présente un zoom d'une coupe coronale provenant de la superposition d'une image de test avec une carte seuillée provenant du LDA (a), du SVM (b) et du CAD final. La tumeur ciblée est localisée dans les poumons, a un diamètre de 14mm et un contraste correspondant à 10% de sensibilité humaine. On constate que la détection de voxels disjoints par chaque classifieur pour un même candidat VP élimine ce même candidat pour le CAD final. Le Tableau 6-8 illustre les sensibilités obtenues sur les tumeurs des poumons pour chaque combinaison de caractéristiques, le LDA, le SVM et le CAD final. Les tumeurs des poumons de diamètre maximal 14mm et de contraste minimal voient ainsi leur sensibilité passer de 80% pour le LDA et 100% pour le SVM à 40% pour le CAD. De même, les tumeurs de diamètre intermédiaire (10,5mm) et de contraste correspondant à 30% de détection humaine passent de 100% de sensibilité pour chaque classifieur à 50% pour le CAD complet.

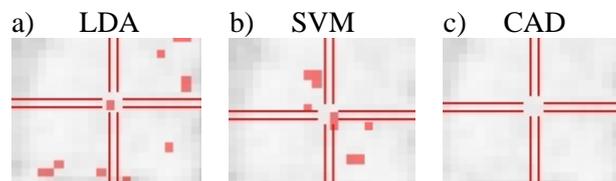


Figure 6-4 : Zoom d'une coupe coronale provenant de la superposition d'une image de test avec une carte seuillée provenant du LDA (a), du SVM (b) et de la combinaison de deux (c). La tumeur ciblée, de contraste minimum, est localisée dans les poumons.

Tableau 6-8 : Sensibilités en pourcentage par contraste et diamètre pour les poumons

(a) extraits des cartes binaires sans monovoxels fournies par le LDA

LDA	10	30	50	70	100
7mm	25 (4)	60 (5)	100 (3)	100 (2)	100 (3)
10,5mm	50 (2)	80 (5)	100 (3)		100 (3)
14mm	100 (2)	100 (5)		100 (2)	100 (1)

(b) extraits des cartes binaires sans monovoxels fournies par le SVM

SVM	10	30	50	70	100
7mm	75 (4)	80 (5)	100 (3)	100 (2)	100 (3)
10,5mm	100 (2)	60 (5)	100 (3)		100 (3)
14mm	50 (2)	40 (5)		100 (2)	100 (1)

(c) extraits des cartes binaires fournies par le CAD complet

CAD	10	30	50	70	100
7mm	25 (4)	60 (5)	100 (3)	100 (2)	67 (3)
10,5mm	50 (2)	40 (5)	100 (3)		100 (3)
14mm	0 (2)	40 (5)		100 (2)	100 (1)

Cette diminution de la sensibilité de quelques groupes de caractéristiques de tumeurs pour les poumons a cependant moins d'impact sur les performances générales, exprimées dans le Tableau 6-3, que pour le foie et les ganglions. Bien que la sensibilité du CAD final décroisse de 10% à 20% par rapport aux sensibilités de chaque classifieur séparément, celle-ci reste tout à fait acceptable pour cet organe (supérieure à 80%).

La raison principale pouvant expliquer la conservation d'une sensibilité correcte pour les poumons repose sur la taille en voxels des candidats VP de cet organe. Ces tailles sont présentées dans le Tableau 6-9 pour chaque classifieur sans monovoxels, puis pour le CAD complet en considérant les trois organes cibles. Ces valeurs sont extraites et moyennées à partir des 13 images de test. Les candidats VP localisés dans les poumons apparaissent plus gros en moyenne que ceux localisés dans le foie et les ganglions. Comme expliqué au paragraphe 5.2.1.3 du chapitre précédent, l'agrégation des voxels candidats est définie par parcours lexicographique de la carte binaire. La taille conséquente des candidats VP dans les poumons pour le LDA et les SVM permet de plus facilement conserver une sensibilité élevée par combinaison. La chance de recouvrement des voxels appartenant à un plus gros cluster candidat est en effet supérieure à celle de voxels appartenant à un amas de taille réduite.

Tableau 6-9 : Taille moyenne en voxels des candidats VP détectés par le LDA, les SVM et le CAD final pour les poumons, le foie et les ganglions.

Taille des candidats VP en voxels	Poumons	Foie	Ganglions
LDA	33	14	10
SVM	38	8	19
CAD	22	3	6

6.4.1.3 Conclusions sur les performances fournies par le CAD final

Un problème non négligeable a été soulevé dans cette discussion. La méthode de combinaison par vote à la majorité est certes efficace, mais doit être appliquée avec discernement. Les adaptations potentielles pourraient porter sur deux points, 1) la redéfinition d'un VP et 2) l'utilisation d'une combinaison de score et non plus de classe. Ces adaptations sont brièvement détaillées ci-après :

1) La combinaison par vote à la majorité apparaît efficace pour les poumons, mais élimine trop de vraies tumeurs par une combinaison de leurs voxels disjoints spatialement. Il serait donc intéressant de redéfinir la méthode de comptage des VP, afin qu'un candidat détecté par les deux classifieurs restent classifié comme tel par le CAD final, même si aucun de ces voxels ne se recouvre. Actuellement, une tumeur est considérée comme correctement détectée si au moins un de ces voxels est détecté par le classifieur. Ce ou ces voxels ainsi classifié(s) constitue(nt) le candidat VP localisé sur

cette tumeur. Lorsqu'un VP est détecté par les deux classifieurs, il doit pouvoir être conservé lors de leur combinaison, même si aucun des voxels classés appartenant à la tumeur réelle n'est concordant spatialement. Une solution pourrait être d'utiliser la morphologie mathématique ou la segmentation par croissance de régions pour redéfinir le candidat obtenu sur la carte binaire à partir de sa représentation dans l'image de départ. Cette méthode est utilisée par [Guan et al., 2006] afin de définir de manière précise les volumes tumoraux dont les graines (initialisation de l'algorithme) ont été obtenues par une classification par compétition-diffusion. Cependant, nous nous heurterions à la difficile paramétrisation des méthodes de segmentation et le fait que les clusters dus au bruit seraient aussi redéfinis par segmentation, et donc plus difficilement éliminés.

2) La seconde solution repose sur l'utilisation d'une combinaison de classifieurs, non plus de type *classe* mais de type *score*. Utilisée par [Park et al., 2009] lors du développement d'un CAD pour la mammographie, cette combinaison se base sur un calcul de moyenne, de maximum ou de minimum des scores directement obtenus à partir des classifieurs. Cette méthode permet de considérer les scores d'appartenance à la classe H1, plus riches d'informations que la simple classe de chaque voxel. Les problèmes d'élimination des candidats détectés de manière spatialement décorrélés par les deux classifieurs pourraient ainsi se réduire étant donné que le score final d'un candidat serait une combinaison linéaire des scores pour chaque classifieur et non plus une loi du 'tout ou rien'.

Au-delà de ces adaptations méthodologiques, nous exploitons dans le prochain paragraphe la possibilité d'intégrer une information de suspicion aux candidats fournis par le CAD pour chaque image. Cette astuce vise à limiter la nuisance visuelle due aux forts taux de FPI observés pour les poumons notamment.

6.4.2 Utilisation d'une information additionnelle de score pour la présentation des résultats

On a constaté sur la Figure 6-2 que les cartes de détection binaires sorties de chaque classifieur séparément et du CAD final sont toujours entachées par un nombre trop important de FPI. La visualisation des cartes binaires est rendue particulièrement délicate dans le cadre d'un futur passage en routine clinique. Le clinicien peut en effet avoir du mal à distinguer les fausses détections des vraies tumeurs et ne pas voir l'utilité d'un tel système d'aide à la détection. Nous exploitons ici la possibilité d'introduire une indication relative au degré de suspicion de chaque candidat afin d'améliorer la qualité parfois limitée des cartes binaires.

Nous avons vu au paragraphe 5.2.2.3 du chapitre précédent puis au paragraphe 6.3.2 du présent chapitre, qu'il est possible de classer les tumeurs candidates par rapport à leur score d'appartenance à la classe H1 'tumeur présente'. Afin de vérifier la pertinence de l'addition de ce type d'informations sur les cartes sorties du système CAD, nous avons analysé les candidats extraits de ces cartes en considérant les 13 images de test. Comme expliqué dans la méthodologie détaillée au paragraphe 6.3.2, la force d'un candidat est exprimée comme le maximum des scores des voxels le constituant. Ces valeurs sont extraites des cartes de score obtenues à partir des deux classifieurs exploités en parallèle. On recense donc deux types de combinaisons carte seuillée/carte de score qui sont 1) CAD/LDA et 2) CAD/SVM. Le rang d'un candidat est calculé en pourcentage comme le rapport entre le classement de sa force et le nombre total de candidats. Comme dans le chapitre précédent, le candidat le plus suspect est classé dans le haut du tableau (premières places) et les moins suspects dans le bas du tableau (dernières places).

6.4.2.1 Analyse visuelle de quelques cas

Les Figure 6-5, Figure 6-6 et Figure 6-7 se focalisent sur quelques tumeurs localisées dans les poumons et les ganglions. Les tumeurs ciblées sont de diamètre 7mm dans le fantôme d'origine pour un contraste lié à une sensibilité humaine échantillonnée de 10% à 90%. L'image d'origine contenant la tumeur cible du contraste étudié est présentée en (a) tandis que la carte de détection avec les candidats rangés par rapport à leur degré de suspicion est affichée pour le LDA en (b) et pour le SVM en (c). Aucune tumeur des poumons de diamètre 7mm et de contraste correspondant à 70% n'est illustrée sur la Figure 6-5 car ce groupe de caractéristiques n'est pas représenté parmi les images de test (cf. Tableau 6-3). De même pour la Figure 6-6, aucune tumeur de diamètre 7mm et de contraste

correspondant à 10% et 30% de sensibilité humaine n'est présentée, car elles ne sont pas détectées par le CAD complet. Le nombre moyen de candidats fournis par le CAD est de 53 pour les poumons, 2 pour le foie et 4 pour les ganglions.

Les cartes obtenues à partir des scores du LDA associés aux candidats du CAD final, présentées sur les Figure 6-5b et Figure 6-7b, montrent que les vraies tumeurs occupent presque toutes les premières places du classement (rouge, orange de l'échelle de couleurs). Les vrais positifs sont en effet toujours placés parmi les premiers 40% des candidats les plus suspects, même pour des petits contrastes (30%). Deux exceptions concernent la tumeur des ganglions de plus haut contraste et la tumeur des poumons de plus faible contraste (avec des rangs supérieurs à 50%). Les rangs correspondant aux tumeurs du foie, présentés sur la Figure 6-6b, sont quant à eux, un peu moins bons, allant jusqu'à une valeur élevée de 80%.

Les résultats obtenus à partir des scores du SVM associés aux candidats du CAD final sont plus mitigés (Figure 6-5c et Figure 6-7c) pour les poumons et les ganglions. On observe en effet que seuls les candidats de contraste élevé (70% et 90% de sensibilité de référence) bénéficient d'un rang inférieur à 25% pour les poumons et les ganglions. Les candidats VP représentés par quelques voxels ressortent donc moins de l'ensemble des candidats par un niveau intermédiaire sur l'échelle de couleurs. En revanche, les scores fournis par les SVM pour les tumeurs du foie sont meilleurs que ceux du LDA.

Les conclusions de cette analyse visuelle semblent rejoindre celles du Chapitre 5 sur la comparaison des deux classifieurs. Elles indiquaient en effet un meilleur comportement du LDA avec des tumeurs cibles de faible contraste et du SVM avec des tumeurs cibles de fort contraste (cf. Figure 5-23 et Figure 5-24).

La Figure 6-8 illustre l'amélioration de la visibilité des VP au milieu de l'ensemble des candidats par l'insertion du degré de suspicion provenant (c) du LDA, (d) du SVM. Les cartes de candidats ordonnés (b) sont superposées aux images d'origine correspondantes présentées en (a). Les caractéristiques des tumeurs numérotées de haut en bas visibles sur les coupes coronales sont également données sur la Figure 6-8. Les coupes observées apparaissent visiblement plus 'nettoyées' par l'utilisation de l'échelle de couleur pour représenter les degrés de suspicion des candidats. Les candidats VP attirent ainsi plus facilement l'œil que sur les cartes binaires uniquement (Figure 6-8b).

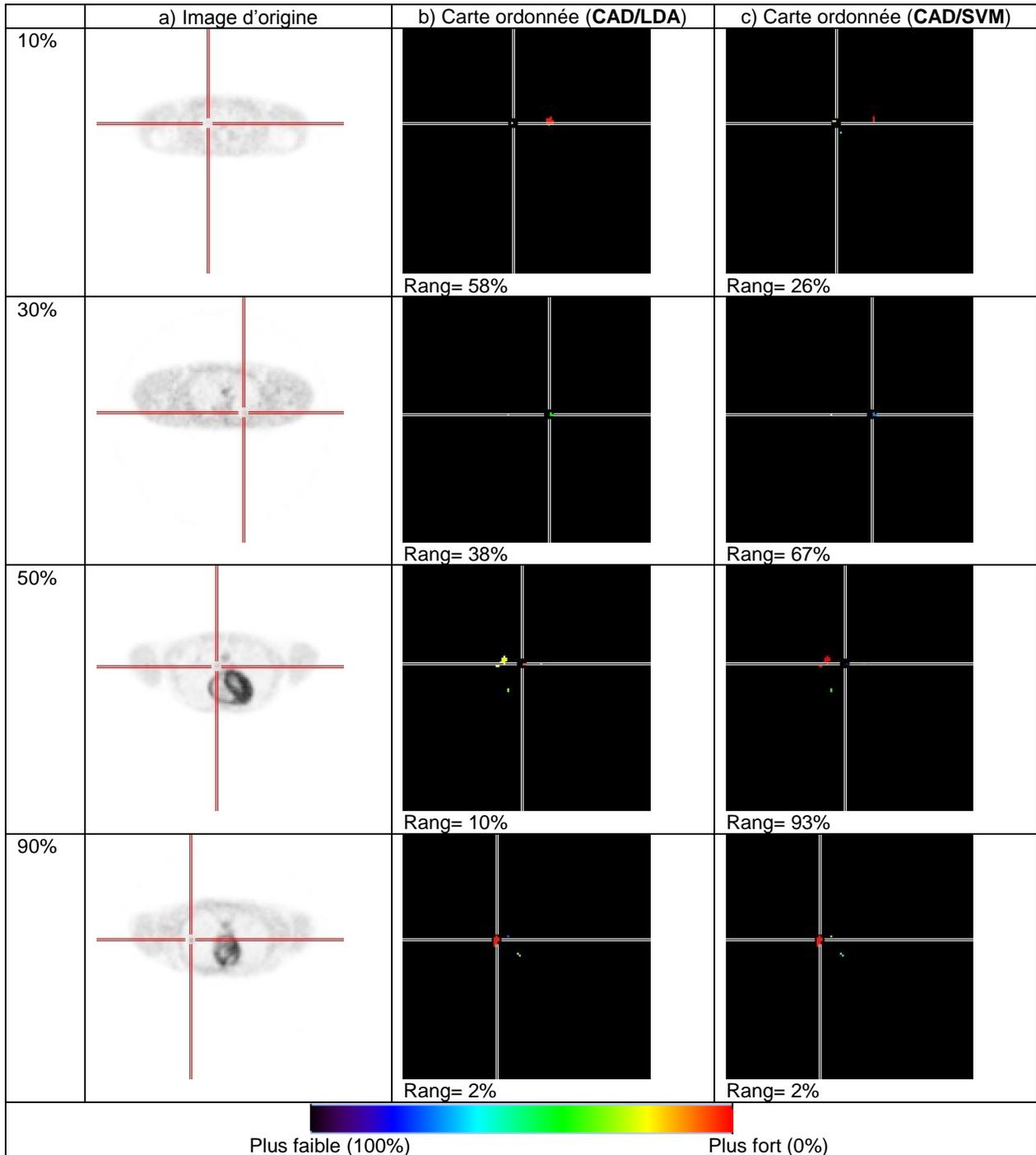


Figure 6-5 : Coupe contenant une tumeur cible dans les *poumons* de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.

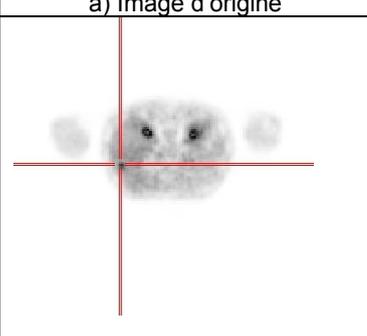
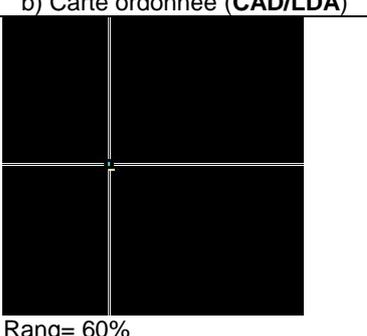
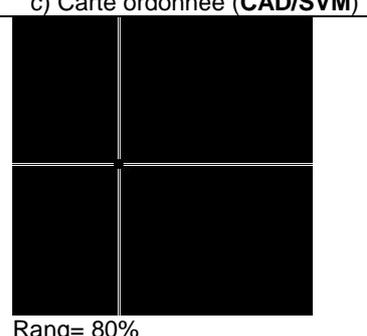
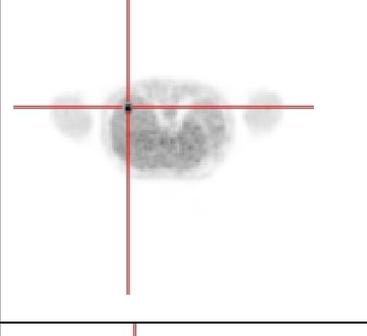
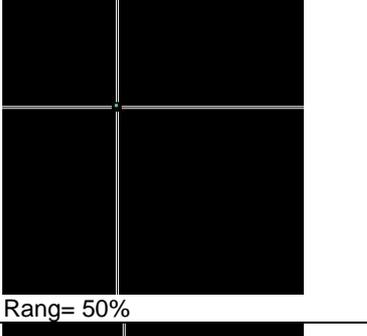
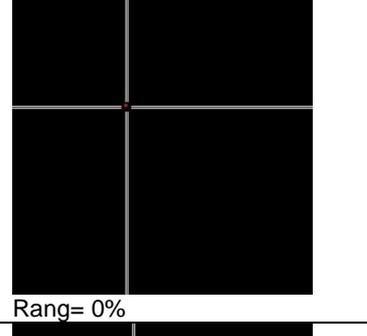
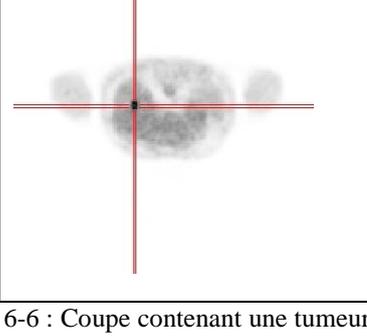
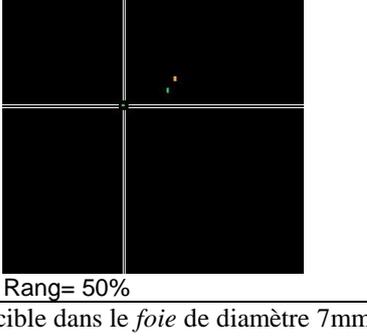
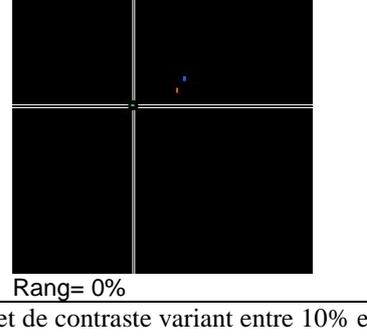
	a) Image d'origine	b) Carte ordonnée (CAD/LDA)	c) Carte ordonnée (CAD/SVM)
50%		 Rang= 60%	 Rang= 80%
70%		 Rang= 50%	 Rang= 0%
90%		 Rang= 50%	 Rang= 0%

Figure 6-6 : Coupe contenant une tumeur cible dans le *foie* de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.

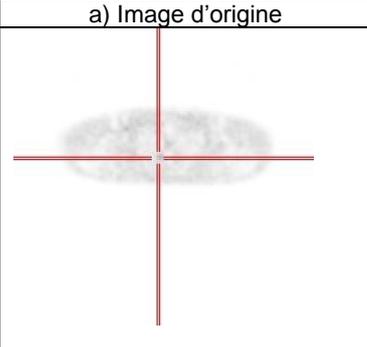
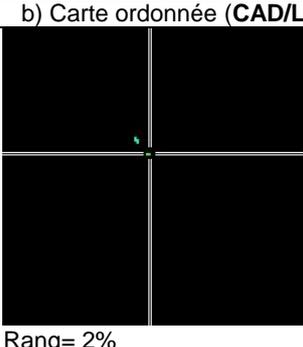
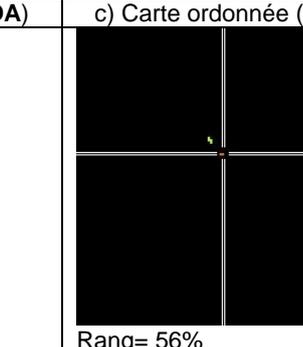
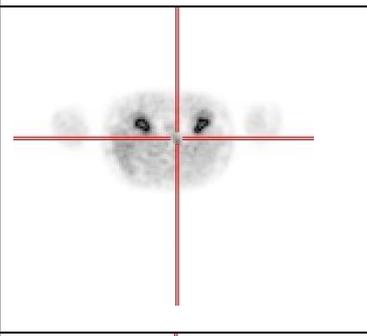
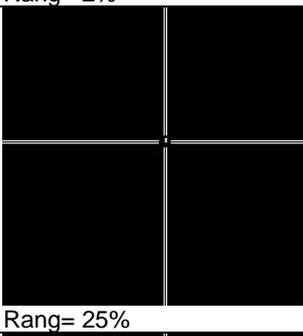
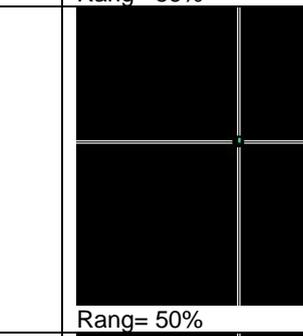
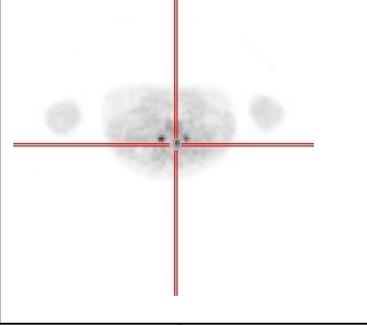
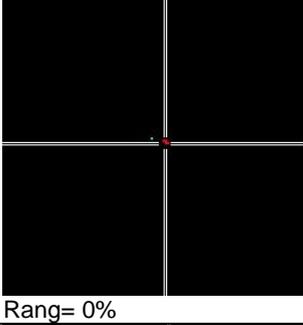
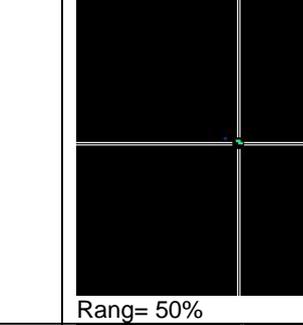
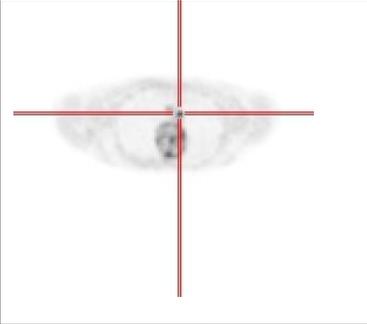
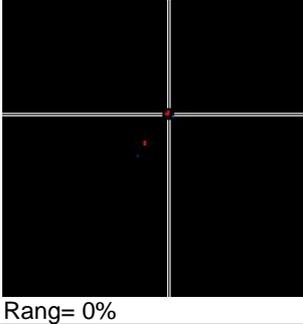
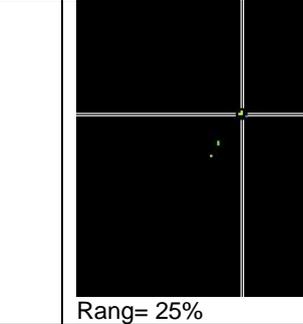
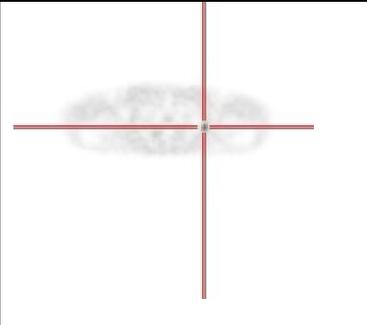
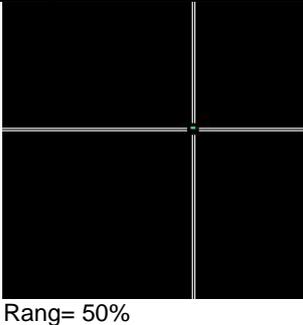
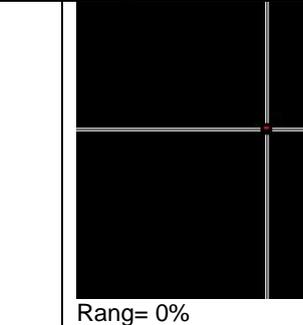
	a) Image d'origine	b) Carte ordonnée (CAD/LDA)	c) Carte ordonnée (CAD/SVM)
10%		 Rang= 2%	 Rang= 56%
30%		 Rang= 25%	 Rang= 50%
50%		 Rang= 0%	 Rang= 50%
70%		 Rang= 0%	 Rang= 25%
90%		 Rang= 50%	 Rang= 0%

Figure 6-7 : Coupe contenant une tumeur cible dans les *ganglions* de diamètre 7mm et de contraste variant entre 10% et 90% de sensibilité humaine de référence. (a) : Image originale, (b) Carte de force rangée pour la combinaison CAD/LDA, (c) pour la combinaison CAD/SVM. Le rang correspond au classement de la force de la tumeur par rapport à tous les candidats de l'organe considéré.

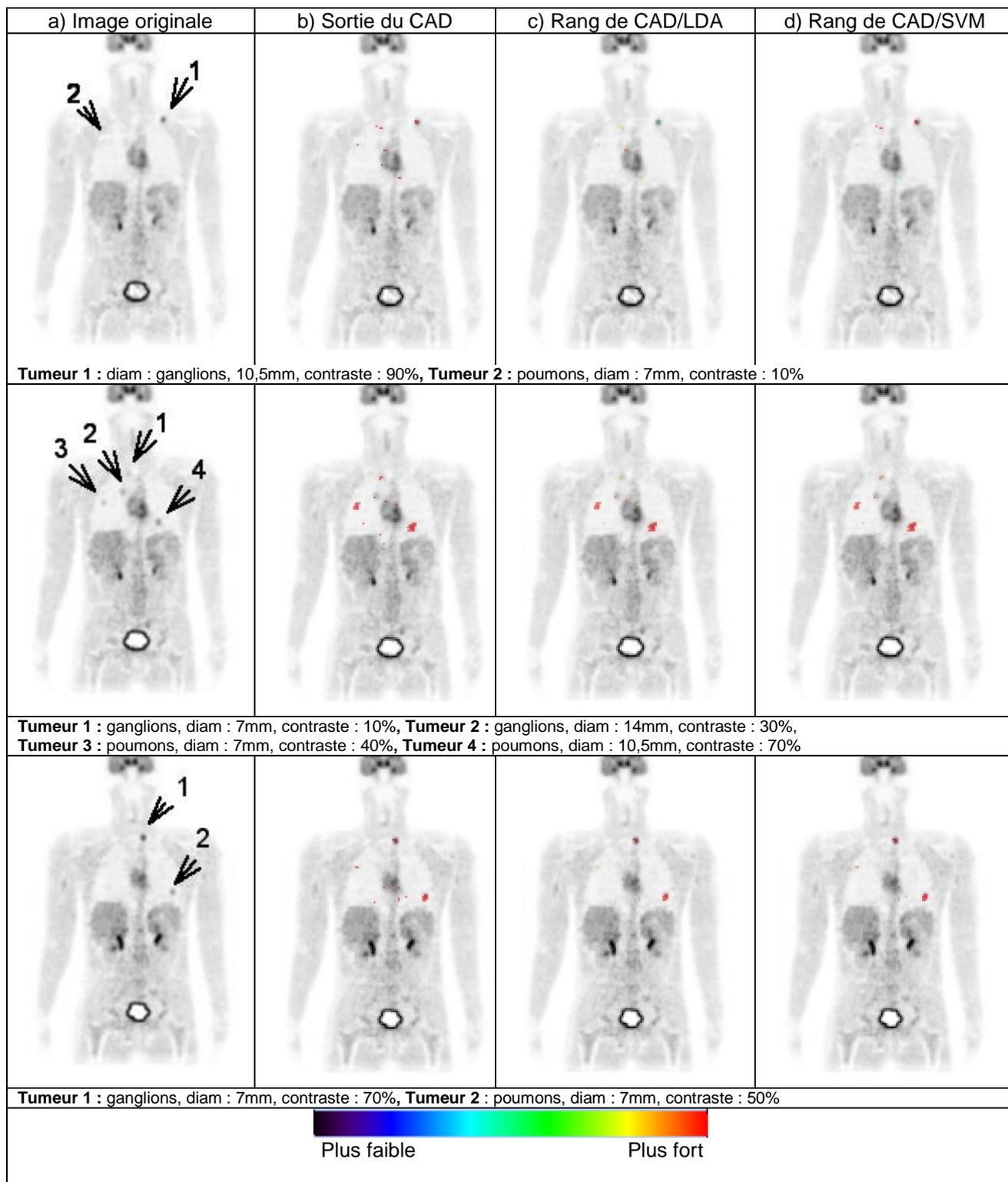


Figure 6-8 : Amélioration de la visibilité des VP au milieu de l'ensemble des candidats par l'insertion du degré de suspicion provenant (c) du LDA, (d) du SVM. Les cartes de candidats ordonnés sont superposées aux images correspondantes d'origine présentées en (a). (b) propose des cartes de détection binaire superposées aux images d'origine. Les tumeurs sont numérotées de haut en bas sur chaque image.

6.4.2.2 Etude de la globalité des candidats vrais positifs

Afin de vérifier la significativité globale de ces résultats visuels, nous avons considéré l'ensemble des candidats VP des poumons et des ganglions (sans distinction de contraste) obtenus par le CAD sur les 13 images de test. La Figure 6-9 illustre le classement des forces de ces candidats VP pour les combinaisons CAD/LDA (triangle vert) et CAD/SVM (triangle rouge) présentées précédemment. Ce diagramme présente une majorité des candidats VP bien classés avec un rang inférieur au 40% des candidats les plus suspects pour la combinaison CAD/LDA. La combinaison CAD/SVM exhibe un plus grand nombre de candidats VP très mal classés (rang>60%). On remarque que de nombreux VP sont au premier rang du classement des candidats les plus suspects, confirmant l'intérêt de l'utilisation de cette information pour mieux valoriser les candidats VP parmi tous.

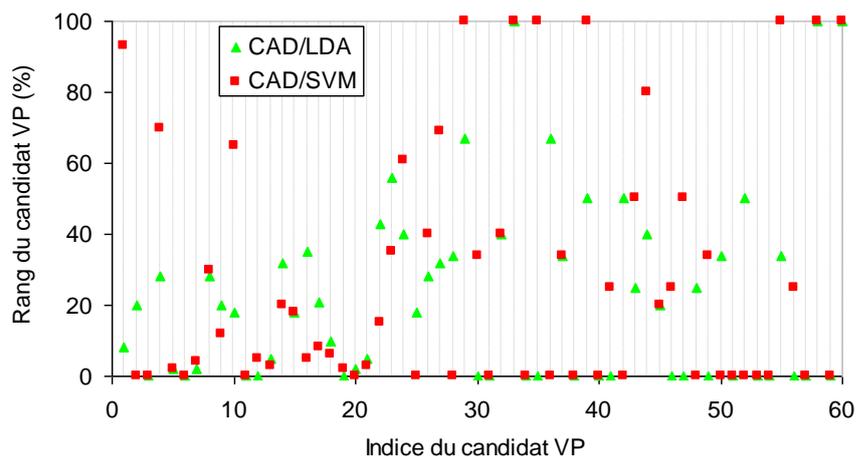


Figure 6-9 : Rang obtenu par chaque candidat VP en considérant tous les organes pour les combinaisons CAD/LDA (triangles verts) et CAD/SVM (carrés rouges). L'idéal est de présenter un VP dans les premières positions dans le classement (0).

La Figure 6-10 illustre le même classement que la Figure 6-9, mais en séparant les tumeurs selon leur contraste relatif à la sensibilité humaine correspondante. On considère (a) des contrastes faibles (10% et 30% de sensibilité humaine), (b) un contraste intermédiaire (50%) et (c) des contrastes élevés (70% et 90%).

Le Tableau 6-10 résume les rangs obtenus sur la Figure 6-10 pour chaque gamme de contraste en considérant chaque configuration carte binaire/carte de score. On observe une diminution des rangs moyens lorsque le contraste augmente. Les candidats VP, plus facilement détectables apparaissent donc plus suspects. On observe un nombre de candidats de plus mauvais rang (100%) deux à trois fois moins important que le nombre de candidats de meilleur rang (0%). Une comparaison des valeurs extraites par chaque combinaison CAD/LDA et CAD/SVM indique un meilleur comportement de la première configuration (CAD/LDA). La combinaison CAD/LDA présente en moyenne des candidats VP de meilleur rang que la combinaison CAD/SVM. Cette différence semble s'amplifier lorsque le contraste augmente. L'observation de la Figure 6-9c montre cependant que cette différence significative est surtout due à la présence des candidats VP de très mauvais rang (100%) pour la combinaison CAD/SVM. Ceux-ci augmentent la moyenne générale des rangs donnée dans le Tableau 6-10. Même si le nombre de candidat VP de rang 0% est quasiment identique pour les deux configurations, l'utilisation du LDA pour l'extraction des rangs résulte en deux fois moins de candidat VP de plus mauvais rang.

Ces résultats semblent indiquer que le LDA est en moyenne plus adapté à l'inclusion d'un degré de suspicion comme indication liée aux candidats extraits du CAD.

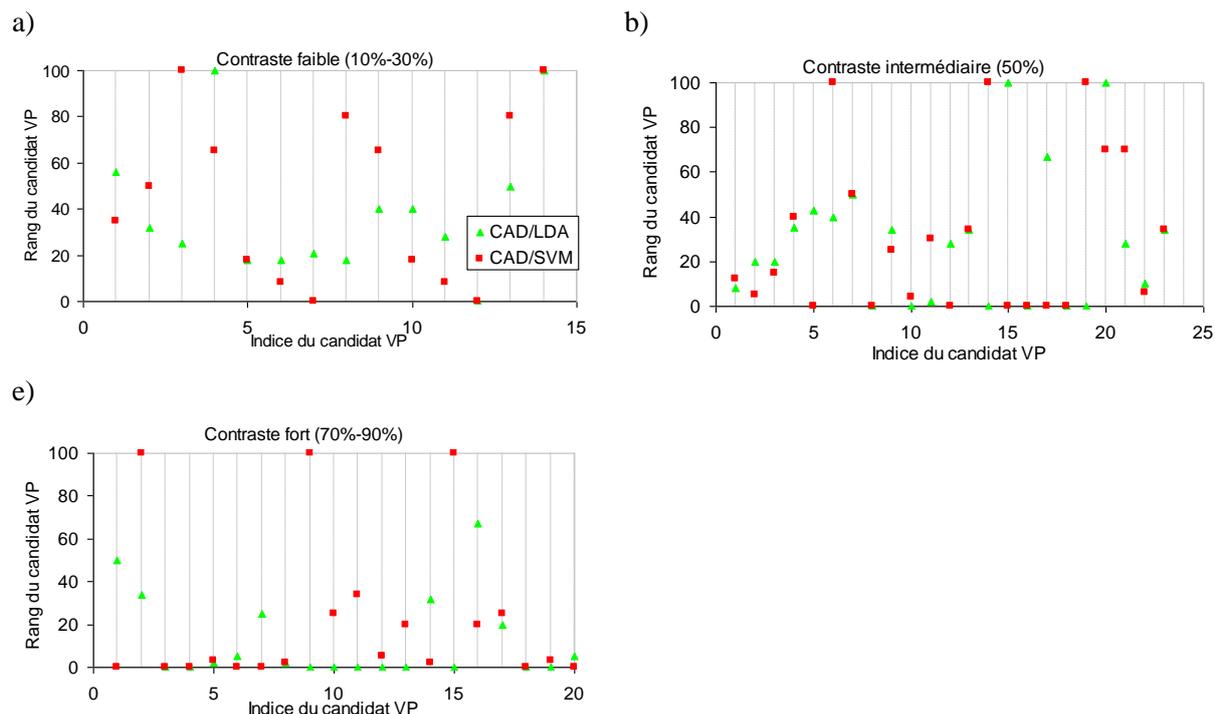


Figure 6-10 : Rang obtenu par chaque candidat VP en considérant tous les organes pour les combinaisons CAD/LDA (triangles verts) et CAD/SVM (carrés rouges). Les tumeurs cibles appartiennent, (a) aux contrastes faibles correspondant à 10% et 30% de sensibilité humaine, (b) au contraste intermédiaire (50%) et aux (c) contrastes forts (70%-90%). L'idéal est de présenter un VP dans les premières positions dans le classement (0).

Tableau 6-10 : Exploitation des rangs des candidats VP dont les valeurs sont extraites de la Figure 6-9 pour chaque combinaison, CAD/LDA et le CAD/SVM. Les meilleures performances sont mises en gras.

		CAD/LDA	CAD/SVM
Nombre de candidat VP de rang 100%		4	8
Nombre de candidat VP de rang 0%		17	16
Moyenne des rangs	Gamme de contraste 10%-30%	39	45
	Gamme de contraste 50%	28	30
	Gamme de contraste 70%-90%	12	22

Ces résultats soulignent l'intérêt d'introduire une information dépendante du degré de suspicion de chaque candidat fourni par le CAD complet, à partir des scores extraits du LDA ou des SVM. Cette information supplémentaire a pour objectif d'attirer l'œil du clinicien sur les amas de voxels les plus suspects. Une perspective pourrait être d'indiquer le meilleur rang obtenu parmi les deux classifieurs.

6.5 Conclusions sur l'évaluation par les données simulées réalistes

6.5.1 Conclusions et discussions

Ce chapitre a abordé l'évaluation du système CAD sur un ensemble d'images TEP 3D simulées réalistes. L'utilisation de telles images permet une étude de faisabilité intermédiaire avant les images cliniques qui sont généralement plus complexes, présentent des organes et des tumeurs plus hétérogènes. L'utilisation de tumeurs de diamètre variable localisées dans des organes virtuels et du bruit physiologique plus complexes que ceux de la base d'images simplifiées utilisées au Chapitre 5 nous a ainsi rapprochés de cas réels.

6.5.1.1 Performances globales

Les performances globales du système CAD complet, exprimées sous la forme de couples sensibilité/FPI, sont de 81%/57 pour les poumons, 54%/1 pour le foie et 86%/3 pour les ganglions (cf.

Tableau 6-5). Le nombre de faux positifs par image est calculé dans tout l'organe ciblé et dépend donc fortement de son volume. Ces performances sont assez décevantes. Le foie et les ganglions présentent une sensibilité beaucoup trop basse et les poumons sont entachés par un nombre de faux positifs trop important. Les performances obtenues à partir des images d'oncoPET_DB ne sont d'ailleurs pas vraiment corrélées à celles extraites des images simplifiées présentées au Chapitre 5.

D'une part, les poumons des images réalistes apparaissent beaucoup moins 'propres' que ceux du Chapitre 5, avec trois fois plus de FPI pour une sensibilité équivalente (cf. paragraphe 5.5.4.3 du Chapitre 5). Une distribution de bruit différente dans les deux types d'images (simplifiées et réalistes) pourrait être à l'origine d'un tel résultat. On constate aussi sur les cartes binaires que les bords des poumons, notamment du côté du diaphragme et du cœur sont particulièrement détectés par le LDA et les SVM et donc le CAD final. Une contamination de l'activité et de la texture des poumons par les organes environnants plus réalistes pourraient expliquer l'obtention de cartes binaires si bruitées.

D'autre part, la sensibilité obtenue sur le foie des images d'oncoPET_DB (54%) est nettement inférieure à celle des images simplifiées (81%). Cette différence pourrait s'expliquer par une plus grande variabilité des caractéristiques des tumeurs en termes de diamètre et de contraste. Les mauvaises sensibilités pour le foie et les ganglions localisées dans les images plus réalistes peut être expliquée par 1) la limite intrinsèque de la combinaison de classe par vote à la majorité, 2) une définition des candidats VP peu adaptée à cette combinaison. Ainsi, des candidats VP conjointement détectés par les deux classifieurs, mais en des voxels disjoints spatialement, peuvent être éliminés par le CAD complet. Le choix de la meilleure combinaison des classifieurs (type score ou classe), mais surtout de la définition spatiale des candidats extraits du LDA et des SVM, peuvent donc être remis en cause.

6.5.1.2 L'utilisation d'oncoPET_DB pour l'évaluation du CAD

Ce chapitre a permis d'évaluer l'intérêt d'utiliser oncoPET_DB dans le cadre de la validation du système CAD développé dans cette thèse.

Une première remarque porte sur le nombre de tumeurs disponibles dans les images d'oncoPET_DB et plus précisément la répartition statistique des différents groupes de caractéristiques dans les organes. Certaines tumeurs sont en effet trop peu représentées pour certains organes par l'ensemble des 25 images avec un nombre inférieur à trois tumeurs (cf. Tableau 6-2 et Tableau 6-3). L'analyse quantitative des performances par diamètre et contraste reste donc assez limitée.

Aussi, les Tableau 6-6, Tableau 6-7 et Tableau 6-8 ont montré que le LDA et le SVM présentent une sensibilité rapidement saturée pour les trois organes cibles. Ainsi, la majorité, voire toutes les tumeurs de contraste intermédiaire à fort sont détectées, alors qu'elles correspondent à une sensibilité humaine de 50% à 90%. Deux raisons pourraient expliquer la détection particulièrement facile de ces tumeurs par les classifieurs, 1) la calibration du contraste des tumeurs, effectuée au Chapitre 4, a induit des tumeurs trop facilement visibles, 2) les classifieurs sont plus performants que l'œil humain pour ces gammes de tumeurs. La Figure 4-8 du Chapitre 4 justifie le choix des gammes de contraste pour chacun des organes à partir de deux observateurs humains ayant visualisés les images simulées d'oncoPET_DB. Cette saturation de la sensibilité pourrait donc indiquer une bonne adaptation du LDA et des SVM à ce type de tumeurs. Nos observations doivent cependant rester prudentes de part la répartition des tumeurs dans certaines catégories parfois trop limitée (cf. Tableau 6-5).

6.5.1.3 Avantages de l'insertion du degré de suspicion des candidats

Le paragraphe 6.4.2 a mis en évidence l'intérêt de présenter les candidats extraits du système CAD complet associés à un rang de suspicion. Ce rang de suspicion correspond au classement obtenu par chaque candidat selon son degré d'appartenance à la classe 'tumeur'. Cette valeur est extraite d'une des deux cartes de score fournie par le LDA ou les SVM en parallèle.

L'étude visuelle des Figure 6-7 à Figure 6-8 a montré des vrais positifs correspondant aux meilleurs rangs, même dans le cas de tumeurs de faible contraste. L'œil du clinicien aurait ainsi plus de facilité à se focaliser sur les vraies tumeurs, même si elles sont noyées au milieu des fausses détections dans les cartes binaires de détection. De même, le Tableau 6-10 indique des rangs moyens toujours inférieurs à 40% quelque soit la gamme de contraste, avec des valeurs généralement

meilleures lorsque les scores sont fournis par le LDA. Ces valeurs de rang sont d'un intérêt tout particulier pour les tumeurs de faible contraste, difficilement visibles à l'œil nu.

6.5.2 Perspectives

La piste suivie pour le développement du CAD semble prometteuse, bien que les performances quantitatives et qualitatives actuelles sur les cartes 3D binaires soient encore insuffisantes pour une application en clinique et remettent en cause certains choix. Les performances finales du système CAD doivent également être estimées sur des images plus représentatives des différentes tumeurs disponibles. Une méthode de partitionnement des images disponibles par 'leave one out' (définie au paragraphe 2.5.4.2 du Chapitre 2) ou 'resubstitution' pourrait être envisagée bien que celles-ci ne garantissent toujours pas une statistique de tumeurs très élevée. L'utilisation de données répliquées (plusieurs simulations du même modèle) ou bootstrappées permettrait de contourner ce problème et de faire mieux apprendre les classifieurs.

Le système semble présenter des problèmes de détection de tumeurs trop visibles, ce qui peut ne pas apparaître en soi comme un inconvénient, étant donné que ce type de tumeurs est facilement visible à l'œil nu et n'est donc pas problématique pour le clinicien. Un comptage plus adapté des vrais positifs d'une image pourrait être mis en place, notamment par l'intégration d'une méthode d'agrégation des voxels plus efficace qu'un simple parcours lexicographique. Une méthode de classification non supervisée ou segmentation, par croissance de régions par exemple, permettrait de mieux définir les clusters candidats.

La confrontation des performances du CAD à celles des observateurs humains sur les mêmes images laissent penser qu'au-delà de l'aspect binaire d'une classification par classe H1/H0, les scores des candidats pourraient être exploités. Un affichage par classement de suspicion des tumeurs candidates pourrait ainsi apporter une information supplémentaire pertinente au clinicien. En utilisant la même catégorie d'information, une combinaison de classifieurs plus adaptée à l'utilisation de seulement deux classifieurs pourrait être évaluée. Utilisé par [Park et al., 2009], la conservation d'informations de type score avant ou après le seuillage initial des cartes pourrait respectivement éviter l'élimination de tumeurs détectées par les deux en des voxels différents, ou lorsqu'un des classifieurs présente un score faible et l'autre un score élevé.

Conclusions, discussions et perspectives

Bilan des contributions et discussions

Les protocoles actuels de recherche pour lutter contre le cancer favorisent la prise en charge optimale des patients, la détection précoce des tumeurs et l'adaptation rapide des traitements. Depuis une dizaine d'années, l'imagerie TEP au 18F-FDG voit son importance s'accroître dans les différentes étapes du suivi thérapeutique du patient. Ses performances, bien que déjà élevées voire parfois supérieures à d'autres imageries anatomiques, peuvent cependant encore être améliorées, notamment en termes de sensibilité et spécificité sur des cas particuliers et problématiques.

Ce manuscrit a abordé un des aspects de l'optimisation de l'utilisation de cette imagerie dans l'objectif de la placer comme examen de référence en cancérologie. L'aide à la détection permet d'orienter le clinicien lors du diagnostic d'une image TEP à partir de la détection et de la localisation des différentes tumeurs qu'elle présente. Certains cas de tumeurs sont en effet difficilement détectables à l'œil humain de par leur contraste faible ou/et leur petit diamètre et/ou leur localisation problématique dans le corps humain.

Les systèmes d'aide à la détection ou CAD en anglais, ont largement été développés en mammographie pour le cancer du sein et en TDM pour le cancer du poumon et du foie. Certains algorithmes sont même implantés sur les consoles cliniques de constructeurs comme Siemens ou Philips. L'imagerie TEP reste malheureusement encore novice dans le domaine des systèmes CAD. Si l'on en croit les études pour les autres modalités (mammographie, TDM), de tels systèmes permettraient d'améliorer les performances générales du clinicien et de diminuer les variations inter et intra-clinicien.

Nous avons dans un premier chapitre rappelé les principes physiques de la TEP au 18F-FDG et son application en oncologie. Ce chapitre nous a conduits à nous orienter vers une problématique de cancer caractérisé par des tumeurs résiduelles faiblement visibles et disséminées dans tout le corps du patient. Afin de répondre au manque de système d'aide à la détection dans cette modalité d'imagerie, nous nous sommes consacrés au développement d'un tel système. Les images TEP oncologiques sont des cibles particulièrement difficiles à exploiter de part leurs caractéristiques de bruit et leur aspect corps entier. Un état de l'art des CAD développés pour cette imagerie nous en a montré leurs limites, tant au niveau de leur quantité que de leur validation. Nous avons donc élargi nos recherches aux CAD dédiés à l'imagerie anatomique, TDM ou mammographie. Notre choix s'est focalisé sur une approche supervisée de la classification de manière à assurer une correcte représentation des différentes régions de l'image obtenues. Un cahier des charges a ensuite été dressé de manière à formaliser les choix effectués pour les différentes étapes standardisées de mise en place du système.

L'approche supervisée de la classification requiert un large jeu d'images annotées, saines et pathologiques, pour l'apprentissage et l'évaluation du système. Ce type de données est difficilement accessible dans le cadre d'études cliniques. Nous avons alors décrit un procédé de développement de bases d'images TEP simulées corps entier dédiées à l'oncologie. La simulation nous a semblé représenter un moyen efficace de contourner les problèmes d'obtention d'images cliniques pour une étude de faisabilité du système. Un ensemble de 100 images saines et pathologiques a été généré par simulation Monte Carlo. Les tumeurs cibles ont été définies de manière à échantillonner les performances humaines de détection, tout en suivant les modes de dissémination du lymphome. Une collaboration étroite avec deux médecins nucléaires, Francesco Giammarile et Thomas Mognetti, respectivement du Centre Hospitalier Lyon Sud et du Centre Léon Bérard, a permis de focaliser notre attention sur ce cancer comme problématique principale. La pertinence des activités de FDG simulées dans l'ensemble des organes et tumeurs et du bruit entachant les images a été vérifiée à travers diverses mesures statistiques et cliniques. Une comparaison des images simulées avec les images cliniques ayant servi à l'élaboration des modèles numériques a permis de mettre en évidence une bonne concordance tant au niveau visuel que quantitatif. Le principal problème de validation de la

base d'images simulées actuelle repose certainement sur le SUV du foie, calculé sur les images cliniques de référence, qui apparaît trop bas par rapport aux valeurs de la littérature.

Nous avons ensuite décrit le développement du système CAD dédié aux images TEP oncologiques et la méthode d'évaluation des cartes binaires 3D obtenues. Les principales étapes d'extraction de caractéristiques, de classification et de nettoyage des cartes binaires ont été évaluées. Deux classifieurs différents ont été utilisés et combinés afin d'obtenir une carte 3D pour chaque nouvelle image où chaque voxel est classé en tumeur ou non tumeur. Afin de considérer l'aspect tridimensionnel de la carte finale obtenue, une définition des performances par agrégats de voxels, correctement détectés ou pas, a été favorisée. Ainsi le nombre réel de FP pour le volume d'intérêt est comptabilisé, représenté la qualité visuelle réelle de l'image. Cette méthodologie d'évaluation quantitative n'a cependant jamais été utilisée en imagerie TEP 3D, et rarement employée en imagerie anatomique 3D. Une évaluation par cas ou par coupe lui est généralement préférée. Aucune comparaison de performances avec d'autres méthodes actuellement utilisées n'a donc été réalisée.

Deux jeux d'images simulées ont été utilisés pour évaluer les performances du CAD. Les premières images, modélisant un modèle du corps humain simplifié, ont permis de mettre en évidence des résultats prometteurs tant visuels que quantitatifs. Une étude multiparamétrique a notamment permis de fixer l'échantillon d'apprentissage et les caractéristiques optimales. Le second jeu d'images, développé lors de cette thèse a permis de placer le CAD dans une configuration beaucoup plus réaliste avec des tumeurs de caractéristiques plus variées et un modèle de corps humain plus complexe et complet. Les performances finales du CAD apparaissent cependant assez mitigées, notamment pour les poumons qui souffrent d'un nombre conséquent de FP et pour le foie qui exhibe une sensibilité limitée. Cependant, les tumeurs des ganglions, cibles principales du lymphome, sont correctement classifiées avec un nombre de FPI limité.

La méthodologie générale s'avère prometteuse, notamment si on considère les performances du système associées à celles des observateurs humains ayant permis la calibration des contrastes des tumeurs cibles. Le CAD souffre cependant d'une limitation non négligeable reposant sur la combinaison même des cartes binaires. Cette limitation, avérée lors de l'utilisation des images simulées réalistes, implique une nette diminution de la sensibilité du système pour des tumeurs cibles de diamètre important. De même, le nombre de FPI actuellement obtenu reste encore trop élevé pour correctement répondre aux besoins des médecins pour une application clinique. Une étude préliminaire a cependant évalué l'intérêt de classer les candidats retenus sur les cartes binaires par leur degré de suspicion. L'affichage d'une carte paramétrique de détection et non plus binaire permettrait d'orienter le médecin et de limiter l'impact visuel de la présence de ces FPI.

Perspectives

Le travail développé dans le cadre de cette thèse ouvre ainsi de nombreuses perspectives.

Le premier axe de poursuite du travail concerne **l'évaluation d'une combinaison des cartes de score (et non plus binaires) et une redéfinition des agglomérats de voxels candidats**. Comme expliqué précédemment, la perte d'information impliquée par l'utilisation exclusive des informations de type classe pourrait être limitée par la prise en compte des scores de chaque candidat. Une redéfinition spatiale des candidats extraits de chaque classifieur permettrait aussi une meilleure correspondance entre les bonnes détections de chaque carte. Cette nouvelle définition autoriserait une analyse FROC plus judicieuse. En effet, la limitation de la taille d'un amas candidat à 125 voxels, défini par un parcours lexicographique de l'image, ne permet pas de correctement refléter l'aspect visuel des cartes binaires obtenues par seuillage. Elle est également à l'origine de petites irrégularités du nombre de fausses détections dans le cas de gros amas détectés qui peuvent être ainsi coupés en plus petits amas de manière non optimale. Il apparaît ainsi parfois difficile de comparer des courbes FROC très proches.

Ce raisonnement nous mène à une autre perspective spécifique à l'utilisation des courbes FROC pour l'extraction d'un seuil optimal pour chaque classifieur, organe et combinaison de paramètres. Bien que la supériorité de certaines combinaisons de paramètres semble évidente (le niveau de décomposition pour les SVM par exemple), certains cas peuvent paraître équivoques

(certaines courbes FROC obtenues pour le LDA par exemple). Outre leur aspect visuel parfois décorrélé du à une définition limitée des agrégats candidats ($j=2$ vs. $j=5$), il est parfois difficile d'établir une différence réellement significative statistiquement entre les deux courbes. Cette remarque est notamment vraie lorsqu'on considère la forte corrélation des courbes aux échantillons pris en compte pour les tracer. Il serait ainsi judicieux **d'analyser statistiquement la différence entre les courbes FROC** représentatives des combinaisons de paramètres à optimiser. La méthode de Chakraborty [Chakraborty and Winter, 1990; Chakraborty and Yoon, 2008] permet par exemple d'extraire différents indices d'évaluation de la significativité des différences entre courbes FROC. Il est également possible, lorsque tous les critères d'étude sont réunis, de faire correspondre les courbes FROC à des courbes ROC plus traditionnelles. Parmi ces critères, il est impératif de réaliser un test sur un nombre équivalent d'images saines et pathologiques. Ce n'est actuellement pas le cas étant donné que seules des images pathologiques sont testées pour la mise en place du CAD.

La troisième piste de travail concerne conjointement **l'enrichissement de la base d'images TEP simulées et la validation finale du CAD sur des images quasi-cliniques**. OncoPET_DB souffre en effet d'une homogénéité de l'anatomie et des distributions d'activité des patients. Un ensemble de modèles de simulation basé sur le fantôme XCAT [Segars, 2001] permettrait d'enrichir la base d'images simulées à l'aide d'images de patients d'anatomie et de sexe variable, pour lesquels les mouvements respiratoires et cardiaques seraient pris en compte. Ce fantôme peut en effet s'adapter à différentes images cliniques réelles par la modification des structures mathématiques (NURBS) modélisant les organes. Cette perspective, accompagnant ce travail de thèse et plus largement le projet de recherche associé, a déjà été amorcée par une collaboration avec le LATIM, Brest [Le Maitre et al., 2009] et plus précisément les travaux de thèse de Simon Marache-Francisco. L'apport indubitable de réalisme aux images TEP simulées permettrait de fixer définitivement les différents paramètres du CAD afin de les appliquer pour la classification d'images cliniques réelles.

Enfin un élargissement du projet concernerait **l'évaluation et la combinaison d'autres classifieurs que ceux initialement sélectionnés et l'inclusion d'information de type anatomique au système complet**. En effet, la littérature prouve qu'un large panel de méthodologies peut encore être testé et optimisé. La mise en place d'un CAD relève un peu de la 'recette de cuisine' où chaque élément peut être remplacé jusqu'à obtenir la combinaison idéale des ingrédients. Les informations additionnelles permettraient notamment de rapprocher la méthode numérique de détection à celle utilisée par le clinicien, qui, de nos jours, n'interprète plus que très rarement les images TEP seules.

Quelques unes de ces perspectives seront notamment traitées par Simon Marache-Francisco, prenant la suite de ma thèse.

Le travail accompli au cours de cette thèse a fait émerger le besoin de méthodologies d'aide à la détection en imagerie TEP 3D corps entier associées à une approche d'évaluation 3D. Ce travail a permis d'évaluer la pertinence d'une méthode basée sur la combinaison de deux classifieurs supervisés. L'approche 3D de l'évaluation reste encore assez précoce en imagerie médicale et mérite d'être encore améliorée afin que les résultats quantitatifs, notamment en termes de faux positifs, apparaissent meilleurs.

Les récents travaux de recherche ont montré l'engouement des laboratoires quant au développement et à la distribution de jeux d'images annotées TEP simulées complexes et réalistes [Stute et al., 2008; Le Maitre et al., 2009; Stute et al., 2009], mais aussi d'acquisitions réelles à partir de fantômes physiques dont la géométrie et les caractéristiques des tumeurs sont maîtrisées [Kadmas, 2009]. Il apparaît donc que la simulation et son utilisation optimale, permettant de fournir des images dont le réalisme frôle la réalité, présentent un grand potentiel dans l'évaluation des méthodologies d'amélioration des images TEP. Les deux contributions présentées tout au long de ce manuscrit pourraient ainsi servir de base au développement et à l'amélioration des méthodologies d'aide à la décision en imagerie TEP corps entier oncologique et de leur évaluation.

ANNEXES

Annexe A : L'imagerie TEP : principes physiques et limitations

1) Les émetteurs de positons et la détection de photons

L'obtention de la cartographie 3D de la distribution du traceur radioactif dans le corps d'un patient est le résultat d'un ensemble de mécanismes allant de phénomènes physiques d'émission de positons, d'interactions photon-matière, à la détection de ces photons et à l'utilisation de ces données détectées pour recréer un volume tridimensionnel de captation par l'organisme. L'utilisation d'émetteurs de positons dont l'évolution rend possible l'acquisition TEP. Les émetteurs de positons sont des isotopes radioactifs présentant un excès de charge positive, ou proton, dans leurs noyaux. Un processus de désintégration, correspondant à la transformation d'un proton p en un neutron n , leur permet de migrer vers un état stable. Cette désintégration résulte en l'émission d'un neutrino ν et d'un positon e^+ selon l'équation (A-1). Le positon est une particule de même masse que l'électron mais de charge opposée.



Une fois émis le positon e^+ parcourt quelques millimètres dans les tissus au cours desquels il dissipe presque totalement son énergie cinétique par collision avec les électrons du milieu. La distance moyenne parcourue par le positon dans le milieu dépend de son énergie cinétique initiale et de la densité électronique de la matière traversée (eau, os...). Cette distance ou libre parcours moyen est généralement de 1 à 2 mm dans l'eau qui constitue l'essentiel des tissus biologiques. La perte énergétique continue ainsi jusqu'à ce que le positon atteigne un état d'équilibre thermique avec le milieu. Le positon interagit alors une dernière fois avec un électron du milieu par une réaction appelée annihilation, illustrée sur la Figure A-1, qui génère deux photons gamma de 511keV en quasi-opposition de phase. En réalité, les photons ne sont pas émis de manière totalement colinéaire, mais selon un angle dépendant du principe de conservation de la quantité de mouvement.

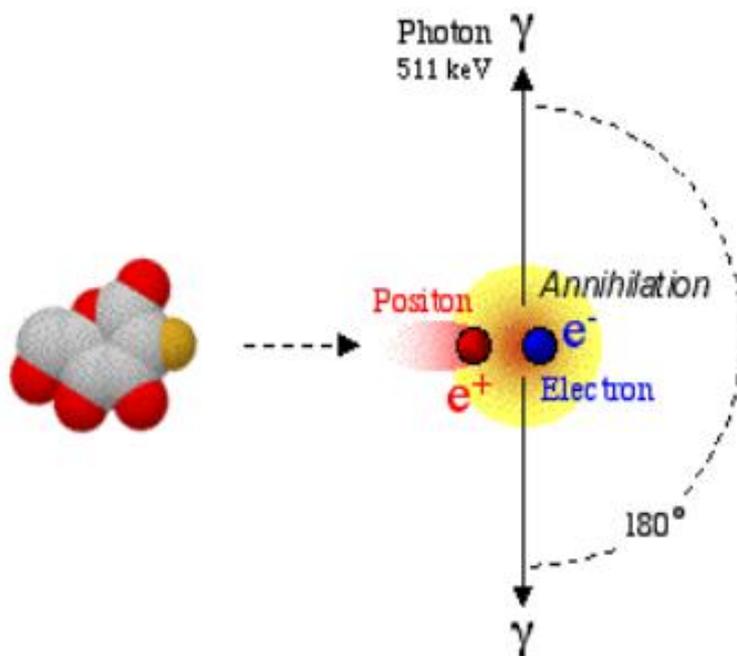


Figure A-1 : De l'émission du positon à l'annihilation

Le principe de base de la TEP est la détection de ces deux photons de 511keV en quasi opposition de phase, afin de déterminer le lieu exact de l'émission du positon. En réalité, l'information collectée lors d'un examen TEP correspond au lieu de l'annihilation et non pas à celui de l'émission de l'émission du positon e^+ . Le libre parcours moyen du positon est donc une limite intrinsèque de la technique TEP qui ne mesure pas directement la distribution spatiale d'activité dans le corps. Le Tableau A-2 donne les principales caractéristiques des émetteurs de positons utilisés en TEP.

Tableau A-2 : Propriétés des traceurs radioactifs les plus couramment utilisés en TEP

Isotopes	Période (minutes)	Coefficient de branchement (%)	Energie cinétique maximum (keV)	Parcours moyen du positon dans l'eau (mm)
15O	2	100	1723	2,22
13N	10	99	1190	1,44
11C	20	99	981	1,12
18F	109,8	97	635	0,6
76Br	960	54	3440	5,0
68Ga	68,2	89	1899	2,4
82Rb	1,3	95	3350	4,7

La période de décroissance des isotopes est relativement courte pour les émetteurs de positons les plus utilisés en TEP (18F, 11C, 15O, 11N) dont le maximum est 109 min pour le ^{18}F . Il est donc nécessaire que les isotopes soient synthétisés à l'aide d'un cyclotron sur le lieu d'utilisation du traceur. Cette propriété physique permet aussi de ne pas trop pénaliser le patient par une dose radioactive trop longue à disparaître du corps.

2) Les différentes coïncidences et la formation de l'image TEP

Les coïncidences vraies

La Figure A-2 décrit le cheminement de la formation de l'image TEP séparé en trois étapes : la détection de l'information utile de l'examen TEP (I), permettant la génération des données TEP résumées en matrices spécifiques appelées sinogrammes (II) qui sont ensuite corrigés et reconstruits pour extraire l'image TEP (III).

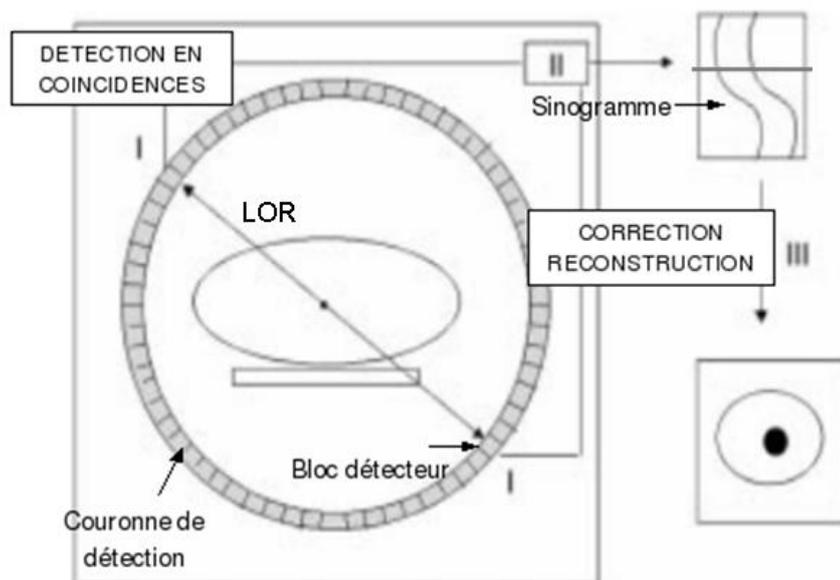
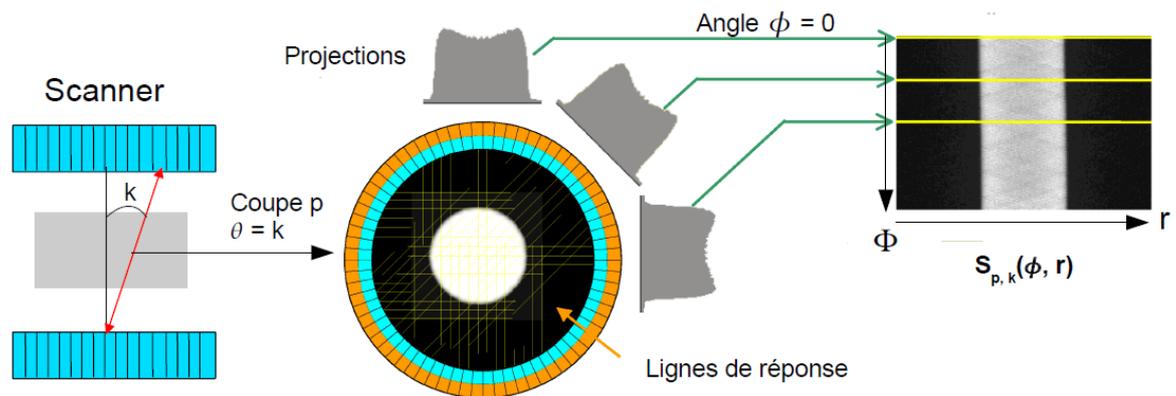


Figure A-2 : Processus de détection et reconstruction d'une image TEP

L'acquisition TEP est basée sur la détection de deux photons d'énergie identique en quasi-opposition de phase. Ces événements sont appelés coïncidences par la nature simultanée en temps et énergie des deux photons détectés. Un système de détection est placé autour du patient et enregistre les photons arrivant sur deux détecteurs opposés et couplés par un dispositif électronique, dans un

intervalle de temps inférieur à une fenêtre temporelle variant de 4 à 20 ns, dite fenêtre de coïncidences. Cette mesure indique la présence d'une annihilation à l'intérieur du volume reliant les deux détecteurs, aussi appelé ligne de réponse (LOR). Les coïncidences vraies correspondent à la détection en coïncidence de deux photons de 511keV dans la même fenêtre temporelle. Ces événements constituent l'information utile pour la formation de l'image TEP. L'ensemble des coïncidences vraies enregistrées dans une LOR est proportionnel à l'intégrale de la distribution radioactive le long de cette ligne. La formation de l'image TEP passe par la mesure d'un ensemble de projections de la distribution radioactive pour différentes valeurs d'angle de projection comme le montre la Figure A-3. Cette figure distingue différents angles et positions par rapport au repère cartésien (x,y,z) du tomographe :

- l'angle polaire θ avec l'axe z du repère cartésien du tomographe,
- l'angle azimutal Φ entre la projetée de la LOR dans le plan (x,y) et l'axe x
- la position radiale r qui représente la distance séparant la ligne de projection à l'axe centre du tomographe



$S_{p,k}(\phi, r)$: sinogramme de la coupe p, $\theta = k$

Figure A-3 : Processus de détection en coïncidences et formation du sinogramme

L'ensemble des projections est stocké dans une matrice appelée sinogramme $S_{p,\theta}(\Phi,r)$. Chaque ligne de cette matrice correspond à la projection monodimensionnelle de la coupe p pour un angle azimutal Φ particulier. Cette matrice possède autant de lignes que d'angles de mesure, et autant de colonnes que d'éléments de projection pour une position angulaire θ .

Les coïncidences aléatoires

Bien que l'information utile de l'examen TEP repose sur la détection de coïncidences vraies, d'autres coïncidences provenant d'autres types d'évènements peuvent aussi être détectées. Celles-ci brulent la formation de l'image TEP. Le premier exemple est les coïncidences aléatoires qui correspondent à la détection de deux photons issus de deux annihilations différentes mais détectées, par hasard, dans la même fenêtre temporelle (Figure A-4).

La distribution des coïncidences aléatoires est plus ou moins uniforme dans le champ de vue. Le taux d'évènements aléatoires mesuré sur une LOR est fonction de la largeur de la fenêtre de coïncidences temporelles (2τ) et des taux de photons simples (aussi appelé singles) détectés par les deux modules considérés. Ces deux grandeurs sont notées S_1 et S_2 . On a donc un taux de coïncidences aléatoires correspondant à l'équation (A-2).

$$R = 2\tau S_1 S_2 \quad (\text{A-2})$$

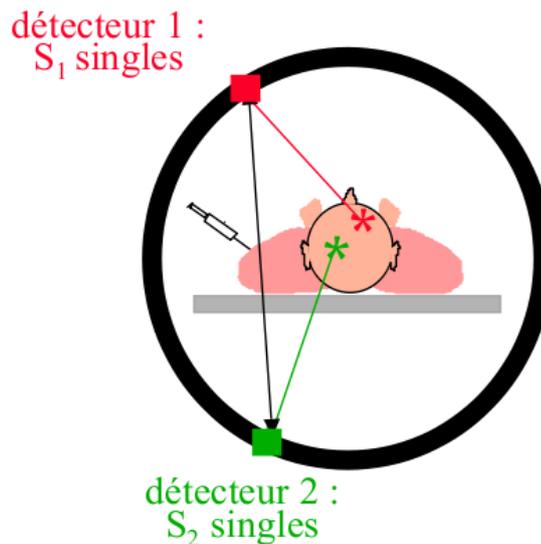


Figure A-4 : Processus de détection des coïncidences aléatoires

Les coïncidences diffusées

Les coïncidences diffusées sont aussi à l'origine d'ajout de bruit dans les données extraites de l'examen TEP. Elles correspondent à la détection en coïncidence de deux photons issus d'une même annihilation mais dont l'un ou l'autre a subi une ou plusieurs diffusions dans les tissus avant de parvenir au système de détection (Figure A-5). Une partie de ces coïncidences est éliminée grâce à une fenêtre en énergie permettant de ne garder que les photons d'énergie proche de 511 keV. Cependant, cette sélection n'est pas suffisante à l'élimination de toutes les coïncidences diffusées.

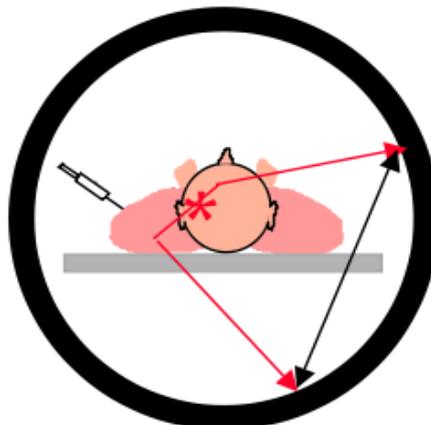


Figure A-5 : Processus de détection de coïncidences diffusées

3) Les modes d'acquisition pour la TEP

Mode 2D et 3D

Comme expliqué dans la section précédente, l'acquisition TEP passe par la détection de deux photons d'énergie 511keV émis en quasi-opposition de phase. Il existe deux modes d'acquisition à proprement parlé en TEP correspondant à la manière dont les LOR sont prises en compte pour le calcul des événements : le mode 2D et le mode 3D.

- Le mode 2D était l'unique mode d'acquisition possible avec les premiers imageurs TEP. Des anneaux de plomb ou tungstènes, appelés septas, placés entre les couronnes de détection de manière à limiter la contamination du signal enregistré sur une coupe par celui enregistré sur les coupes axiales voisines. Seules les coïncidences correspondantes à ces plans directs (une seule coupe) sont censées être détectées. Cependant, les événements détectés par les couronnes adjacentes sont également permis afin d'augmenter l'échantillonnage spatial en créant des plans virtuels intercalés entre les plans directs

comme le montre la Figure A-5a. Ainsi une caméra contenant N couronnes permet l'acquisition de $2 \times N - 1$ plans correspondants à N plans directs et N-1 plans intercalés. Une manière de rendre le mode 2D moins restrictifs en termes d'acceptation des événements repose sur l'enregistrement d'événements correspondants à des angles d'incidence plus importants. Dans ce cas, les événements sont aussi comptés comme provenant des plans directs ou intercalés le plus proche. Le paramètre fixant l'angle d'acceptation est noté angle maximum d'ouverture (ou maximum ring difference, mrd) et correspond en général au nombre d'anneaux séparant les extrémités de la LOR inclinée de l'angle maximum d'ouverture. La Figure A-5 présente un mrd de 1, n'autorisant que les événements provenant des plans directement adjacents à celui étudié.

- Le mode 3D repose sur la rétractation des septas, permettant la comptabilisation des événements reliant deux détecteurs de n'importe quelle couronne comme l'illustre la Figure A-5b. La sensibilité du système de détection est de ce fait fortement améliorée, ce qui permet de diminuer les doses injectées au patient lors d'examen cliniques. Dans ce cas le mrd est égal au nombre de couronnes. L'absence de septa se traduit cependant par une augmentation du taux d'événements diffusés et aléatoires. Il faut compter sur l'enregistrement des photons simples qui augmente également et contribue à l'augmentation du bruit statistique en saturant le système d'acquisition à cause de l'effet de temps mort qui sera détaillé plus loin. De nos jours, seules les caméras mode 3D sans septa sont commercialisées pour les applications cliniques.

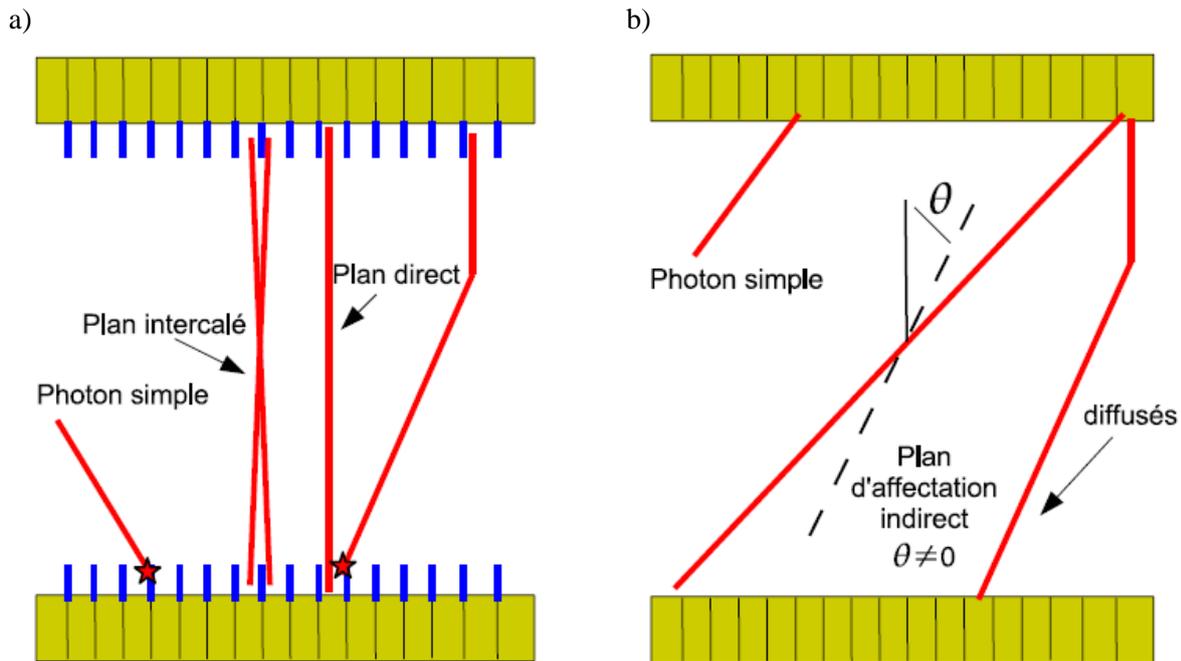


Figure A-5 : Principe du mode d'acquisition 2D a) et 3D b)

Les différentes acquisitions TEP

Il existe trois types d'acquisition TEP. La première méthode d'acquisition est l'acquisition en émission, qui correspond à celle précédemment détaillée, où le système de détection tourne autour du patient dont on veut extraire la distribution du radiotracer injectée. La seconde méthode correspond à une acquisition en transmission comme l'illustre la Figure A-6a. Dans ce cas, l'acquisition est réalisée à partir d'une source externe, tournant autour du patient. Cette source peut être du germanium 68 (^{68}Ge), du césium 137 (^{137}Cs) ou encore des rayons X. Le temps de mesure en transmission est souvent inférieur au temps d'acquisition en émission. Cette acquisition permet de mesurer pour chaque LOR l'atténuation des photons dans le corps du patient pour pouvoir ultérieurement corriger les données de l'affaiblissement due aux événements atténués. La dernière acquisition est appelée blanc (Figure A-6), elle est réalisée à vide, c'est-à-dire sans patient et suivant le même protocole que l'acquisition en transmission. Ces deux dernières acquisitions permettent de corriger les données de l'atténuation comme expliqué dans la suite.

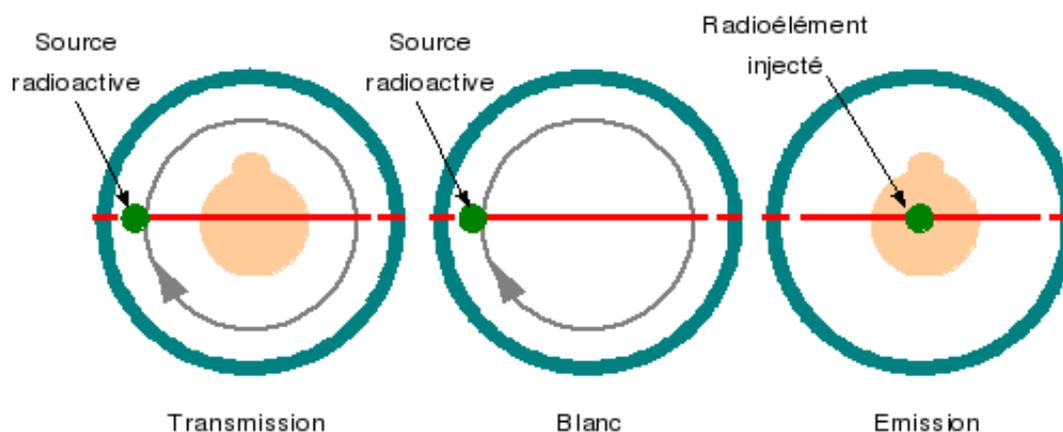


Figure A-6 : Les trois types d'acquisition, transmission, blanc et émission

4) Les différentes corrections

Les données acquises lors d'un examen TEP ne peuvent pas être utilisées directement à cause de leur contamination par différentes sources de bruit ayant pour origine des phénomènes physiques ou des caractéristiques du système de détection utilisé. Il est donc nécessaire de corriger les données acquises avant la reconstruction de l'image TEP à partir des sinogrammes.

Correction d'atténuation

Une proportion importante des photons de 511 keV est atténuée de manière non isotrope par les tissus du patient suite à diverses interactions avec la matière. Les images TEP non corrigée de l'atténuation, sous-évaluent les fixations profondes et amplifient l'activité du bord comme le montre la Figure A-7. La correction d'atténuation permet de restituer une activité correcte quelque soit la position dans le corps du patient. L'ensemble des approches existantes [Zaidi et al., 2007] se base sur les deux acquisitions (en transmission et blanc) présentées précédemment. La combinaison des données provenant des deux scans permet de déduire l'atténuation subie par les photons sur chaque LOR du sinogramme et donc les coefficients de correction à utiliser.

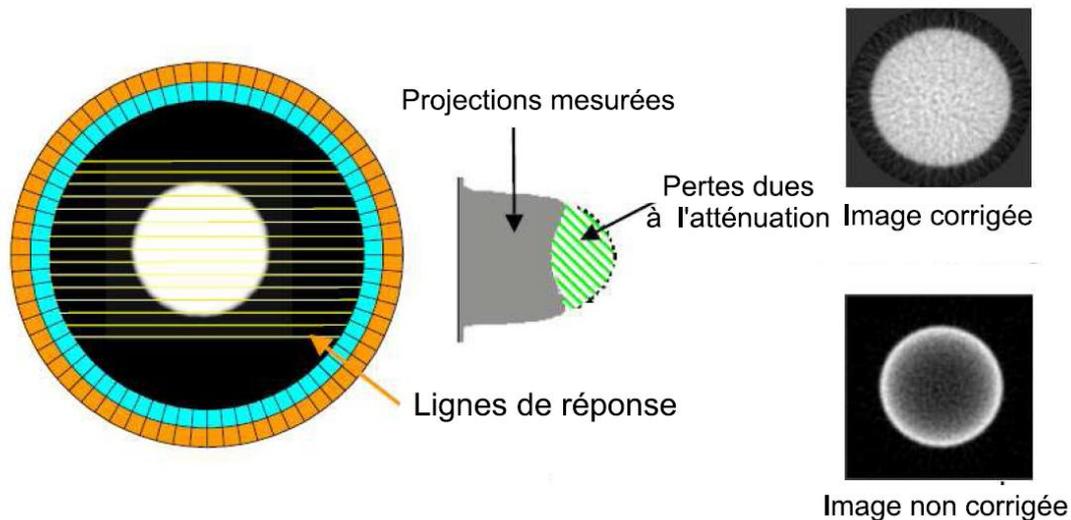


Figure A-7 : Effet de l'atténuation sur les images TEP

Correction de temps mort

Le temps mort est un phénomène limitant trouvant son origine dans le système de détection lui-même. Il s'agit de l'intervalle de temps durant lequel le dispositif est insensible après le passage d'une impulsion électrique. Lorsque deux impulsions surviennent avec un décalage temporel inférieur au temps d'intégration, seule la première est traitée et la deuxième est perdue. La correction de ce temps mort consiste à modéliser les pertes de comptage en événements pour chaque module de détection. En effet, le taux de photons simples s et le facteur de temps mort, noté d_{tk} sont liés par la relation (A-3) où les constantes A et B apparaissant sont déterminées expérimentalement.

$$d_{tk} = 1 + As + Bs^2 \quad (A-3)$$

Correction de coïncidences diffusées

Les coïncidences diffusées, bien que déjà triées par les fenêtres énergétiques peuvent excéder 50 % des coïncidences détectées. Les images non corrigées présentent une diminution du contraste, du rapport signal sur bruit et une perte en résolution spatiale. Trois types méthodes existent afin de corriger les données de ces coïncidences diffusées [Zidi 2004, Watson 1996].

- La première famille de méthode combine les données acquises dans au moins deux fenêtres en énergie.
- La seconde exploite l'information spatiale de localisation erronée des coïncidences diffusées
- La dernière se base sur un calcul direct de la distribution des diffusés de manière analytique pour un patient donné ou en utilisant des méthodes de simulation de type Monte Carlo.

Correction des coïncidences aléatoires

Les coïncidences aléatoires provenant de la détection de deux photons issus de deux annihilations différentes mais détectés dans la même fenêtre temporelle bruite les données mesurées. La correction de ce phénomène consiste à soustraire au niveau du sinogramme les coïncidences aléatoires estimées des coïncidences vraies. Ces coïncidences aléatoires peuvent être estimées de trois façons [Brasse et al., 2005].

- La première utilise une fenêtre temporelle décalée. L'activité y est considérée comme constante par sa taille réduite, mais les événements détectés proviennent forcément d'annihilations différentes par sa taille suffisamment élevée.
- La seconde estime le taux de coïncidences aléatoires en connaissant le nombre total de photons détectés par chaque module en utilisant la formule (A-2).

- La dernière estime ce taux à partir de la distribution des coïncidences dans les projections en dehors du patient.

La normalisation

Des différences de sensibilité de détection d'une source uniforme existent d'une LOR à l'autre, d'une part à cause de la forme des systèmes de détections, en général à anneaux circulaires ; d'autre part à cause des différences d'incidence des photons par rapport à la face d'entrée du cristal et de l'efficacité individuelle du système de détection. Les méthodes de correction peuvent être basées sur l'inversion directe des données acquises à partir d'une source irradiant uniformément toutes les lignes de réponse. Elles peuvent aussi être basées sur des mesures traitées par un modèle mathématique reliant le facteur de normalisation pour une LOR entre deux cristaux à l'efficacité intrinsèque des détecteurs et une fonction de la sensibilité dépendant de l'arrangement géométrique des détecteurs.

5) La reconstruction des données

Les données acquises et corrigées lors d'un examen TEP peuvent ensuite être reconstruites afin d'extraire la distribution tridimensionnelle du traceur observé dans l'organisme. Deux grandes catégories d'algorithmes peuvent être utilisées pour réaliser cette étape : les méthodes analytiques et les méthodes itératives. Ces méthodes peuvent être appliquées sur des données mode 3D mais également mode 2D. Etant donné que la reconstruction 3D est très coûteuse en temps de calcul du à un plus grand nombre de données, des méthodes permettant de réarranger les données 3D en données 2D ont été mises en place. Le Fourier Rebinning (FORE) [Defrise et al., 1997] est généralement utilisé.

Les méthodes analytiques

Les méthodes analytiques correspondent généralement à la méthode de rétroprojection des données. Cette méthode est basée comme son nom l'indique, sur la rétroprojection des projections du sinogramme. L'opération P de projection $p(r, \Phi)$ d'une distribution $f(x,y)$ d'un objet selon l'angle azimutal Φ entre la projetée de la LOR dans le plan (x,y) et l'axe x est donnée peut s'écrire comme la transformée de Radon de cette distribution comme le montre l'équation (A-4).

$$p(r, \phi) = P(f(x, y)) = \int_{-\infty}^{+\infty} f(x, y) dr \quad (A-4)$$

L'opération de rétroprojection, dénotée RP approxime l'opération inverse de la projection et permet d'obtenir l'image reconstruite f_{RP} en fonction des projections p acquises (A-5).

$$f_{RP}(x, y) = RP(p(u, \phi)) \quad (A-5)$$

La première manière simpliste de reconstruire l'image est de rétroprojeter simplement les projections sur la LOR correspondante les valeurs de la projection correspondante. Au fur et à mesure des rétroprojections, on voit s'ajouter sur la coupe de l'image, les zones où figurent de l'activité. Il se crée aussi des zones où n'existe aucune activité. Ces résidus déforment l'image résultante comme le montre la Figure A-8.

La seconde manière plus complexe de reconstruire l'image permet de corriger l'apparition des artefacts obtenus par rétroprojection simple. La méthode de rétroprojection filtrée est basée sur le théorème de la coupe centrale.

« La transformée de Fourier 1D (F1) d'une projection d'angle azimutal Φ correspond à la ligne de la transformée de Fourier 2D (F2) de l'image passant par l'origine et d'angle Φ . »

Cette méthode nécessite de multiplier la transformée de Fourier des projections par la valeur absolue d'un filtre rampe. Ce filtre rampe permet d'introduire des valeurs négatives dans les projections qui annulent la composante continue représentant la moyenne du signal. Le filtre rampe permet aussi d'introduire de part et d'autre de l'objet filtré des valeurs négatives. Leur rôle est d'effacer progressivement le signal présent en dehors de l'objet laissé lors des rétroprojections. Le filtrage rampe amplifie également les hautes fréquences du signal. Il est donc nécessaire d'utiliser un filtre passe-bas pour couper ces hautes fréquences de manière à réduire le bruit. La rétroprojection filtrée se résume donc en quatre étapes pour chaque angle de projection Φ :

- Calcul de la transformée de Fourier de la projection $p(r, \Phi)$

- Filtrage de la transformée de Fourier de la projection par le filtre rampe associé à un filtre passe bas
- Calcul de la transformée de Fourier inverse de la transformée de Fourier filtrée
- Rétroprojection de la projection filtrée

La rétroprojection filtrée 3D (Filtered Backprojection, FBP) [Kinahan and Rogers, 1989] est une méthode rapide, facile et assez satisfaisante dans le cas de fixations de traceur relativement homogènes. Elle amplifie cependant généralement le bruit statistique, et génère des artéfacts en étoile. Cette méthode est aussi inadéquate lorsque le nombre d'événements enregistrés par projection est trop faible. En effet, la rétroprojection crée des images biaisées par le manque de projections.

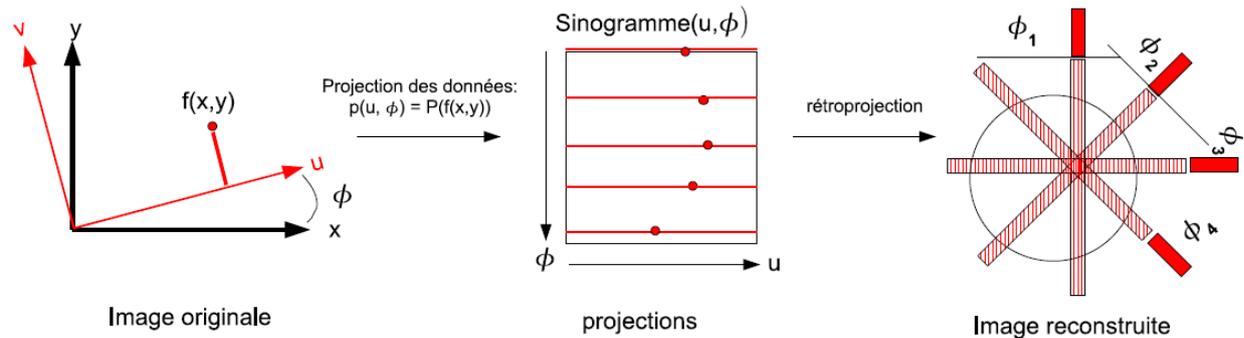


Figure A-8 : Principe de la reconstruction par rétroprojection simple des données

Les méthodes itératives

Les méthodes itératives se basent sur l'hypothèse que les données acquises lors d'un examen TEP, provenant de processus d'émission et d'interaction d'un photon avec la matière, suivent la statistique de Poisson. Les méthodes itératives se basent sur les algorithmes itératifs de type maximum de vraisemblance (Maximum Likelihood) [Shepp and Vardi, 1982]. Ces algorithmes cherchent à maximiser la vraisemblance de l'image f donné par (A-6), par rapport aux données p_k considérées comme des variables aléatoires de Poisson de paramètre \overline{p}_k .

$$prob(p / f) = \prod_k e^{-\overline{p}_k} \frac{\overline{p}_k^{p_k}}{p_k!} \quad (A-6)$$

La reconstruction itérative consiste à réajuster l'image f à l'itération $n+1$ en fonction de son estimée de l'itération précédente à l'aide d'un facteur comparant les données estimées aux données réellement acquises. La formule (A-7) résume cette opération.

$$f_{n+1} = f_n \times s_n \text{ avec } s_n = RP \left(\frac{P_k}{P(f_n)} \right) \quad (A-7)$$

P est l'opérateur de projection donnant les mesures à partir de l'image et RP l'opération de rétroprojection. A chaque itération n , la vraisemblance de l'image augmente et converge vers un maximum global. Cette méthode est marquée par l'absence de critère objectif pour stopper le processus, et implique la propagation du bruit, même si elle fonctionne bien pour des données comptabilisant peu d'événements. La lenteur de la méthode itérative EM peut être palliée en utilisant une technique regroupant les données en sous-ensembles ordonnés de projections (Ordered Subset Expectation Maximization ou OSEM) développé par Hudson et Larkin [Hudson and Larkin, 1994]. En clinique routine sur les scanners Siemens, la méthode Attenuation-Weighted OSEM (AWOSEM) [Comtat et al., 1998] est utilisée. Cette variante de l'algorithme OSEM est pondérée en atténuation. En d'autres termes, la correction d'atténuation est intégrée au processus itératif. Cette méthode permet de mieux répondre à l'hypothèse de données de statistique de Poisson, hypothèse malmenée lors des corrections préliminaires et extérieures à la reconstruction. A l'aide de cette méthode, le bruit statistique des données est plus justement pris en compte, résultant en des images de meilleure qualité.

Annexe B : Rappels sur la théorie des ondelettes

Nous considérons dans cette thèse une *décomposition en ondelettes 3D non décimée* [Shensa, 1992]. Le terme *3D* signifie que nous prenons en compte les caractéristiques de l'image d'un point de vue volumique, dans les trois directions (x, y, z) . Cela semble logique lorsque l'on considère des tumeurs volumiques dans une image 3D.

Le terme le plus important concerne la *décomposition en ondelettes*. Les ondelettes sont des fonctions localisées à la fois dans le domaine fréquentiel et spatial. Elles permettent de visualiser les différents niveaux de résolutions spatiales d'une image, donc de détails, tout en maintenant les distributions de bruit. Commençons d'abord par décrire brièvement la théorie des ondelettes dans le domaine continu et son passage dans le domaine discret.

Toute fonction ψ de carré intégrable, c'est-à-dire appartenant à $L^2(\mathbb{R})$ et de moyenne nulle, ce qui se traduit par $\int_{-\infty}^{+\infty} \psi(t) dt = 0$, peut être une ondelette. Un ensemble de dilatations et de translations d'une même fonction ondelette mère ψ permet la création d'une famille d'ondelettes. La transformée en ondelettes d'une fonction f quelconque est en fait une projection au sens du produit scalaire dans L^2 sur les fonctions de cette famille comme le montre l'équation (B-1).

$$TO(f) = \langle f, \psi_{a,b} \rangle = \frac{1}{a^2} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (\text{B-1})$$

Le coefficient a représente le facteur d'échelle ou la dilatation, et le facteur b le paramètre de position de l'ondelette. En général, les signaux numériques, et dans notre cas les images TEP, sont représentés dans le domaine discret et non pas continu. Dans ce cas là, il est peu confortable d'utiliser la formulation continue des ondelettes. Cela reviendrait à calculer toutes les valeurs possibles de dilatation a et de translation b pour une même fonction mère. Il est donc nécessaire de discrétiser ces paramètres et d'approximer l'intégrale du produit scalaire définissant la transformée en ondelettes continue.

Pour cela, on pose $a = 2^{\frac{j}{\nu}}$, avec ν le nombre d'échelles d'un niveau et j le niveau de décomposition. Cet échantillonnage exponentiel permet de respecter la forme des supports d'ondelettes. Pour le décalage b , l'invariance en translation doit être assurée et le pas d'échantillonnage est choisi proportionnel au facteur d'échelle a , soit $b = kb_0 2^j, k \in \mathbb{Z}$. Lorsque le pas b est échantillonné avec le même pas que la fonction traitée, alors $b_0 = 1$. La famille d'ondelettes discrètes est maintenant exprimée par l'équation B-2.

$$\psi_{j,k}(t) = \frac{1}{2^{\frac{j}{\nu}}} \psi\left(\frac{t}{2^{\frac{j}{\nu}}} - k\right) \quad (\text{B-2})$$

De même, il existe une fonction d'échelle ϕ qui permet d'extraire des coefficients d'approximation de l'image traitée. Celle-ci permet de filtrer l'image originale et de compléter les images de détail obtenues par application de la fonction ondelettes.

Connaissant la fonction d'ondelette ψ , il est maintenant possible d'extraire un filtre passe-haut lui correspondant (B-3).

$$\psi_{j,k}(t) = 2 \sum_n h(n) \psi_{j,k}(t-n) \quad (\text{B-3})$$

Les différents niveaux de décomposition sont modélisés par une dilatation du support du filtre. De même, un filtre passe-bas peut aussi être obtenu à partir de la fonction d'échelle φ liée à la fonction d'ondelettes. La décomposition en ondelettes peut donc être réalisée par l'intermédiaire d'un banc de filtre passe-haut et passe-bas, comme illustrée sur la Figure B-1 en 3D. L'image de départ est filtrée séparément dans les trois directions de l'espace x , y et z par les filtres fréquentiels 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 et une image d'approximation 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.

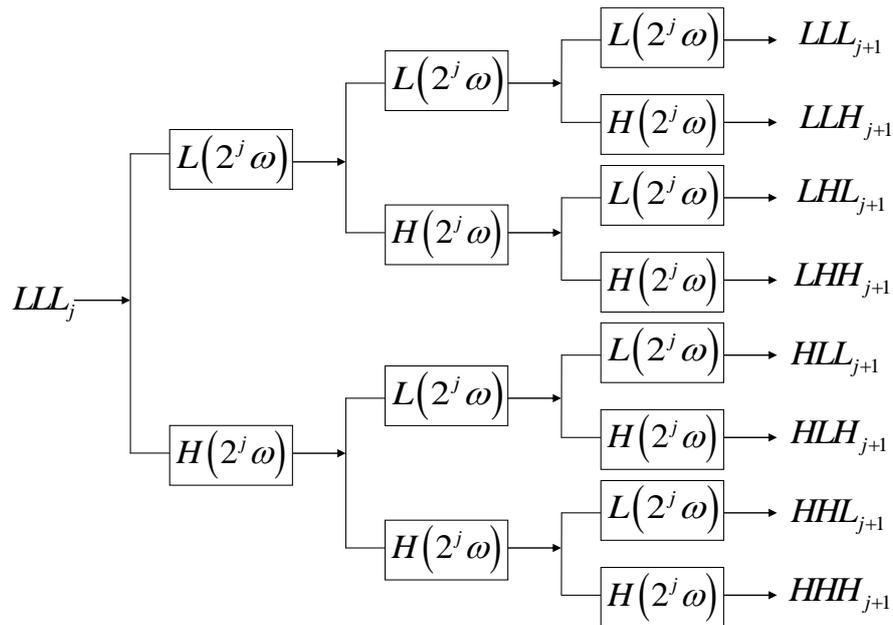


Figure B-1 : Implémentation par banc de filtres d'une décomposition en ondelettes non décimée

Le terme *non décimée* signifie que les images résultantes de la décomposition en ondelettes conservent leur taille d'origine au lieu de subir une décimation d'un facteur 2 comme c'est le cas pour la transformée décimée. Ainsi la navigation entre les différents niveaux de résolution est facilitée et la transformation est connue en chaque pixel de manière à améliorer la précision de la détection. Dans ce cas là, le procédé est isotropique et exclut toute sélection de direction particulière.

Annexe C : Description de la base d'images simulées simplifiée

La première base d'images TEP simulées utilisée pour l'évaluation et la validation du système CAD développé dans cette thèse a été développé au cours de la thèse de Carole Lartizien [Lartizien, 2001]. Elle a été utilisée pour l'optimisation des protocoles d'imagerie TEP corps entier pour la détection des hyperfixations [Lartizien et al., 2004]. Un ensemble d'images saines (sans tumeur) et pathologiques (avec tumeurs) ont été générées par simulation analytique [Comtat et al., 1999] d'un fantôme adapté du MCAT [Tsui et al., 1993] présenté sur la Figure C-1a. L'outil de simulation analytique a été validé pour la géométrie du scanner Siemens/CTI Ecat Exact HR+. Il permet de générer des simulations réalistes sur le plan des propriétés statistiques du bruit tout en tenant compte des effets dégradants l'image TEP corps entier (atténuation, diffusés et décroissance de l'activité...). Son principe repose sur le calcul analytique des projections basées sur les spécifications géométriques du fantôme simulé auquel s'ajoute un niveau de bruit statistique prédéterminé.

Les données simulées utilisées dans cette thèse ont été reconstruites à l'aide de l'algorithme itératif AW-OSEM en utilisant 4 itérations, 16 subsets et un filtre Gaussien post-reconstruction de 10 mm de largeur à mi-hauteur. Les volumes résultants sont de taille 128 x 128 x 225 voxels isotropes de dimension 5,0625 mm x 5,0625 mm x 2,4250 mm. Pour faciliter leur exploitation et ne privilégier aucune direction de l'espace lors de l'application du CAD 3D, les volumes sont rendus isotropes. Les nouvelles images sont ainsi de taille 128 x 128 x 180 voxels de dimension 5,0625 mm dans les trois directions.

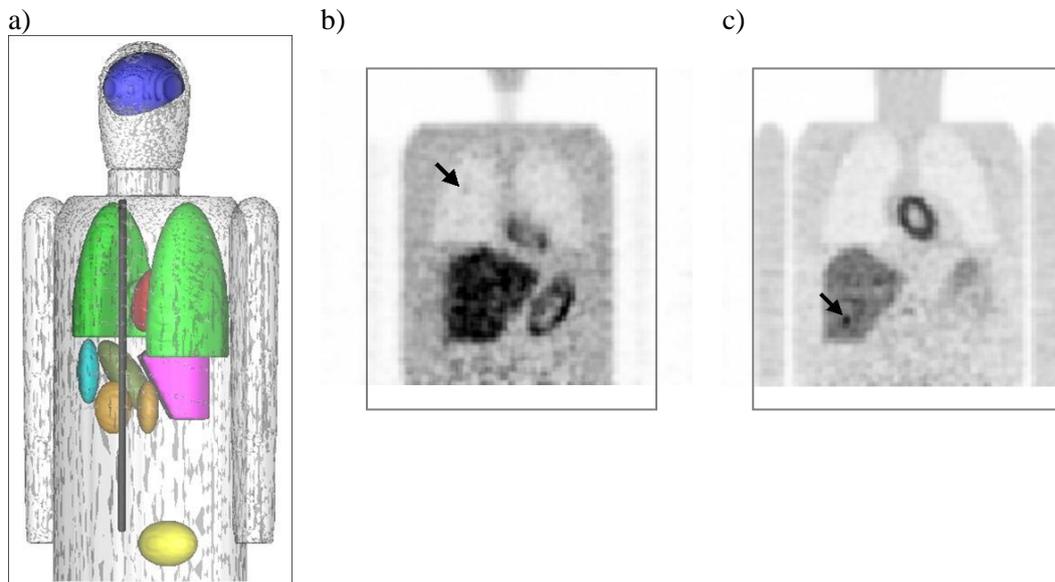


Figure C-1 : Vue 3D du fantôme MCAT modifié a) et de trois coupes coronales présentant une tumeur dans les poumons b), le foie c) fléchée.

Chaque volume TEP pathologique contient sept tumeurs réparties aléatoirement selon une distribution de probabilité multinomiale de moyenne 2,5 dans les poumons et le foie et 2 dans les tissus mous. Les tumeurs sont modélisées dans le fantôme numérique par des sphères de 1 cm de diamètre, soit 2 voxels. Dans l'image reconstruite, elles correspondent à un signal Gaussien d'approximativement 2 voxels de largeur à mi hauteur. Leur activité a été définie selon une étude de calibration humaine [Lartizien et al., 2004] ayant inspirée celle réalisée pour oncoPET_DB (cf. Chapitre 4). Cinq valeurs de contraste ont été extraites de courbes AFROC humaines pour chaque organe par échantillonnage des sensibilités de détection humaines de 10 % à 90 %. L'ensemble de ces

contrastes, résumés dans le Tableau C-1, ont été sélectionnés de manière à ce que les tumeurs soient, d'un côté suffisamment détectables, mais de l'autre, pas trop évidentes. Les Tableau C-2 et Tableau C-3 présentent la répartition des tumeurs dans les différentes gammes de contraste pour les trois organes cibles que sont les poumons, le foie et les tissus mous.

Tableau C-1: Gamme de contrastes définie pour des tumeurs situées dans chaque organe de la base d'images simulées simplifiée à partir des performances de détection humaine

Performances humaines (%)	10	30	50	70	90
Poumons	2,5	6,5	7	7,5	9
Foie	2,5	3	3,25	3,75	4,75
Tissus mous	6,5	7	8	9	10

Tableau C-2 : Répartition du nombre de tumeurs par organe pour chaque contraste dans les 25 images différentes de l'apprentissage et de validation. Le contraste est défini par rapport à la sensibilité humaine correspondante.

Performances humaines (%)	10	30	50	70	90
Poumons	5	15	15	15	4
Foie	12	18	17	8	14
Tissus mous	8	12	10	11	11

Tableau C-3 : Répartition du nombre de tumeurs par organe pour chaque contraste dans les 25 images de test. Le contraste est défini par rapport à la sensibilité humaine correspondante.

Performances humaines (%)	10	30	50	70	90
Poumons	17	11	12	10	18
Foie	14	8	11	19	12
Tissus mous	8	10	10	11	6

Bibliographie

- [Adam et al., 1997] Adam, L.-E., Zaers, J., Ostertag, H. T., H., Belleman, M. E. and Brix, G. (1997). "Performance evaluation of the whole-body PET scanner ECAT EXACT HR following the IEC standard." *IEEE Transactions on Nuclear Science* vol. 44, n° 3(2), pp. 1172-1179.
- [Adams et al., 1998] Adams, S., Baum, R. P., Stuckensen, T., Bitter, K. and Hör, G. (1998). "Prospective comparison of 18F-FDG PET with conventional imaging modalities (CT, MRI, US) in lymph node staging of head and neck cancer." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 25, n° 9, pp. 1255-1260.
- [Agostinelli et al., 2003] Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., et al. (2003). "GEANT4 - A simulation toolkit." *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* vol. 506, n° 3, pp. 250-303.
- [Akhurst and Larson, 1999] Akhurst, T. and Larson, S. M. (1999). "Positron emission tomography imaging of colorectal cancer." *Semin. Oncol.* vol. 26 pp. 577-583.
- [Aristophanous et al., 2007] Aristophanous, M., Penney, B. C., Martel, M. K. and Pelizzari, C. A. (2007). "A Gaussian mixture model for definition of lung tumor volumes in positron emission tomography." *Medical Physics* vol. 34, n° 11, pp. 4223-4235.
- [Aristophanous et al., 2008] Aristophanous, M., Penney, B. C. and Pelizzari, C. A. (2008). "The development and testing of a digital PET phantom for the evaluation of tumor volume segmentation techniques." *Medical Physics* vol. 35, n° 7, pp. 3331-3342.
- [Armato et al., 2004] Armato, S. G., McLennan, G., McNitt-Gray, M. F., Meyer, C. R., Yankelevitz, D., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., et al. (2004). "Lung image database consortium: developing a resource for the medical imaging research community." *Radiology* vol. 232, pp. 739-748.
- [Aubert-Broche et al., 2006] Aubert-Broche, B., Grova, C., Reilhac, A., Evans, A. C. and Collins, D. L. (2006). "Realistic simulated MRI and SPECT databases application to SPECT/MRI registration evaluation". *Medical Image Computing and Computer Assisted Intervention Conference*, Copenhagen, Denmark, Springer Verlag, pp. 330-337.
- [Badel et al., 2007] Badel, J. N., Sarrut, D. and Ginestet, C. (2007). "Prediction of portal dose images with Monte Carlo simulation for dosimetric control in radiotherapy." *Int. Conf. Use of Comp. in Radiotherapy Therapy*.
- [Bae et al., 2005] Bae, K. T., Kim, J. S., Na, Y. H., Kim, K. G. and Kim, J. H. (2005). "Pulmonary nodules: Automated detection on CT images with morphologic matching algorithm – Preliminary results." *Radiology* vol. 236, pp. 286-294.
- [Bangerter et al., 1998] Bangerter, M., Moog, F., Buchmann, I., Kotzerke, J., Griesshammer, M., Hafner, M., Elsner, K., Frickhofen, N., Reske, S. N., et al. (1998). "Whole-body 2-[18F]-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) for accurate staging of Hodgkin's disease." *Annals Oncology* vol. 9, pp. 1117-1122.
- [Banvard, 2002] Banvard, R. A. (2002). "The Visible Human Project® Image Data Set From Inception to Completion and Beyond". *Proceedings CODATA Frontiers of Scientific and Technical Data Track I-D-2: Medical and Health Data*, Montréal, Canada.

- [Bar-Shalom et al., 2001] Bar-Shalom, R., Mor, M., Yefremov, N. and Goldsmith, S. J. (2001). "The value of Ga-67 scintigraphy and F-18 fluorodeoxyglucose positron emission tomography in staging and monitoring the response of lymphoma to treatment." *Seminar Nuclear Medicine* vol. 31, pp. 177-190.
- [Barrett et al., 1993] Barrett, H. H., Yao, J., Rolland, J. P. and Myers K.J." Proc. Natl. Acad. Sci. 90, p.-., 1993 (1993). "Model observers for assessment of image quality." *Proceedings of the National Academy of Sciences of the United States of America* vol. 90, n° 21, pp. 9758–9765.
- [Barrington and O'Doherty, 2003] Barrington, S. F. and O'Doherty, M. J. (2003). "Limitations of PET for imaging lymphoma." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 30, n° suppl.1, pp. 117-127.
- [Baydush et al., 2007] Baydush, A. H., Catarious, D. M., Lo, J. Y. and Floyd, C. E. (2007). "Incorporation of a Laguerre-Gauss channelized Hotelling observer for false-positive reduction in a mammographic mass CAD system." *Journal of Digital Imaging* vol. 20, n° 2, pp. 196-202.
- [Bayes, 1763] Bayes, T. (1763). "An Essay towards solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of London* vol. 53, pp. 370-418.
- [Bazzani et al., 2001] Bazzani, A., Bevilacqua, A., Bollini, D., Brancaccio, R., Campanini, R., Lanconelli, N., Riccardi, A. and Romani, D. (2001). "An SVM classifier to separate false signals from microcalcifications in digital mammograms." *Physics in Medicine and Biology* vol. 46, n° 6, pp. 1651-1663.
- [Bornefalk and Hermansson, 2005] Bornefalk, H. and Hermansson, A. B. (2005). "On the comparison of FROC curves in mammography CAD systems." *Medical Physics* vol. 32, n° 2, pp. 412-417.
- [Boroczky et al., 2006] Boroczky, L., Zhao, L. and Lee, K. P. (2006). "Feature subset selection for improving the performance of false positive reduction in lung nodule CAD." *IEEE Transactions on Information Technology in Biomedicine* vol. 10, n° 3, pp. 504-511.
- [Boser et al., 1992] Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). "A training algorithm for optimal margin classifiers." *Fifth annual workshop on computational learning theory*, pp. 144-152.
- [Boussion et al., 2006] Boussion, N., Hatt, M., Lamare, F., Bizais, Y., Turzo, A., Cheze-Le Rest, C. and Visvikis, D. (2006). "A multiresolution image based approach for correction of partial volume effects in emission tomography." *Physics Medicine Biology* vol. 51, pp. 1857-1876.
- [Brasse et al., 2005] Brasse, D., Kinahan, P. E., Lartzien, C., Comtat, C., Casey, M. and Michel, C. (2005). "Correction methods for random coincidences in fully 3D whole-body PET: impact on data and image quality." *Journal Nuclear Medicine* vol. 46, n° 5, pp. 859-67.
- [Breiman, 1996] Breiman, L. (1996). "Bagging predictors." *Machine Learning* vol. 2, n° 2, pp. 123-140.
- [Breiman et al., 1984] Breiman, L., Friedman, J., Olshen, R. and Stone, C., Eds. (1984). "CART: Classification and Regression Trees". Series "CART: Classification and Regression Trees".
- [Brink et al., 2004] Brink, I., Schumacher, T., Mix, M., Ruhland, S., Stoelben, E., Digel, W., Henke, M., Ghanem, N., Moser, E., et al. (2004). "Impact of [18F]FDG-PET on the primary staging of small-cell lung cancer." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 31, n° 12, pp. 1614-1620.

- [Brown et al., 2003] Brown, M. S., Goldin, J. G., Suh, R. D., McNitt-Gray, M. F., Sayre, J. W. and Aberle, D. R. (2003). "Lung micronodules: automated method for detection at thin-section CT – initial experience." *Radiology* vol. 226, pp. 256–262.
- [Brusseau et al., 2004] Brusseau, E., De Korte, C. L., Mastik, F., Schaar, J. and Van Der Steen, A. F. W. (2004). "Fully automatic luminal contour segmentation in intracoronary ultrasound imaging - a statistical approach." *IEEE Transactions on Medical Imaging* vol. 23, n° 5, pp. 554-566.
- [Buvat et al., 2005] Buvat, I., Castiglioni, I., Feuardent, J. and Gilardi, M.-C. (2005). "Unified description and validation of simulators in PET." *Physics Medecine Biology* vol. 50, n° 2, pp. 329-346.
- [Buvat et al., 2007] Buvat, I., Soret, M. and Foehrenbach, H. (2007). "Clinical comparison of the accuracy of three SUV indices in FDG PET." *Journal Nuclear Medicine* vol. 48, n° (Supplement 2), pp. 44.
- [Cachin et al., 2006] Cachin, F., Kelly, A. and Maublant, J. (2006). "Évaluation de la réponse thérapeutique en cancérologie : place de l'imagerie isotopique." *Bulletin du Cancer* vol. 93, n° 12, pp. 1191-1199.
- [Campadelli et al., 2005] Campadelli, P., Casiraghi, E. and Valentini, G. (2005). "Support vector machines for candidate nodules classification." *Neurocomputing* vol. 68, n° 1-4, pp. 281-288.
- [Carbone et al., 1971] Carbone, P. P., Kaplan, H. S., Musshoff, K., Smithers, D. W. and Tubiana, M. (1971). "Report of the Committee on Hodgkin's disease staging classification." *Cancer Research* vol. 31, pp. 1860-1861.
- [Castiglioni et al., 2005] Castiglioni, I., Buvat, I., Rizzo, G., Gilardi, M. C., Feuardent, J. and Fazio, F. (2005). "A publicly accessible Monte Carlo database for validation purpose in emission tomography." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 32, n° 2, pp. 1234-1239.
- [Castiglioni et al., 1999] Castiglioni, I., Cremonesi, O. G., M.C., Bettinardi, V., Rizzo, G., Savi, A. B., E. and Fazio, F. (1999). "Scatter correction techniques in 3D PET: a monte carlo evaluation." *IEEE Transactions on Nuclear Science* vol. 46, n° 6, pp. 2053-2058.
- [Chakraborty and Winter, 1990] Chakraborty, D. P. and Winter, L. H. (1990). "Free-response methodology: alternate analysis and a new observer-performance experiment." *Radiology* vol. 174, pp. 873-881.
- [Chakraborty and Yoon, 2008] Chakraborty, D. P. and Yoon, H. J. (2008). "Operating characteristics predicted by models for diagnostic tasks involving lesion localization." *Medical Physics* vol. 35, n° 2, pp. 435-445.
- [Chan et al., 2007] Chan, H.-P., Wei, J., Zhang, Y., Moore, R. H., Kopans, D. B., Hadjiiski, L., Sahiner, B., Roubidoux, M. A. and Helvie, M. A. (2007). "Computer-aided detection of masses in digital tomosynthesis mammography: Combination of 3D and 2D detection information". San Diego, CA, United states, SPIE, pp. SPIE.
- [Chan et al., 1987] Chan, H. P., Doi, K., Galhotra, S., Vyborny, C. J., MacMahon, H. and Jokich, P. M. (1987). "Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography." *Medical Physics* vol. 14, pp. 538-548.

- [Chang and Lin, 2001] Chang, C.-C. and Lin, C.-J. (2001). LIBSVM : a library for support vector machines: Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [Cheng and Cui, 2004] Cheng, H. D. and Cui, M. (2004). "Mass lesion detection with a fuzzy neural network." *Pattern Recognition* vol. 37, n° 6, pp. 1189–1200.
- [Cheng et al., 2006] Cheng, H. D., Shi, X. J., Min, R., Hu, L. M., Cai, X. P. and Du, H. N. (2006). "Approaches for automated detection and classification of masses in mammograms." *Pattern Recognition* vol. 39, pp. 646 – 668.
- [Cheson, 2007] Cheson, B. D. (2007). "The international harmonization project for response criteria in lymphoma clinical trials." *Hematology Oncology Clinics of North America* vol. 21, n° 5, pp. 841-854.
- [Cohade et al., 2003] Cohade, C., Osman, M., Leal, J. and Wahl, R. L. (2003). "Direct comparison of (18)F-FDG PET and PET/CT in patients with colorectal carcinoma." *Journal Nuclear Medicine* vol. 44, n° 11, pp. 1797-1803.
- [Cohen et al., 1992] Cohen, A. I., Daubechies, I. and Feauveau, J. C. (1992). "Biorthogonal bases of compactly supported wavelets." *Communications on Pure Applied Mathematics* vol. 45, pp. 485-560.
- [Colavolpe et al., 2008] Colavolpe, C., Fakhry, N., Guedj, E., Tessonier, L., Zanaret, M. and Mundler, O. (2008). "TEP au 18-FDG et cancers ORL : récidence et surveillance post-thérapeutique." *Médecine Nucléaire* vol. 32, n° 8, pp. 399-408.
- [Comtat et al., 1998] Comtat, C., Kinahan, P. E., Defrise, M., Michel, C. and Townsend, D. W. (1998). "Fast reconstruction of 3D PET data with accurate statistical modelling." *IEEE Transactions on Nuclear Science* vol. 45, pp. 1083–1089.
- [Comtat et al., 1999] Comtat, C., Kinahan, P. E., Defrise, M., Michel, C. and Townsend, D. W. (1999). "Simulating whole-body PET scanning with rapid analytical methods". *Proceeding IEEE Nuclear Science Symposium and Medical Imaging Conference*, Seattle, pp. 1260-1264.
- [Corrêa Silva et al., 2005] Corrêa Silva, A., de Paiva, A. C. and de Oliveira, A. C. M. (2005). "Comparison of FLDA, MLP and SVM in Diagnosis of Lung Nodule." *Machine Learning and Data Mining in Pattern Recognition* vol. 3587, pp. 285-294.
- [Cui et al., 2008] Cui, Y., Zhao, B., Akhurst, T. J., Yan, J. and Schwartz, L. H. (2008). "CT-guided, automated detection of lung tumors on PET images". San Diego, CA, United states, SPIE, pp. The International Society for Optical Engineering (SPIE).
- [Dasarathy, 1991] Dasarathy, B. V., Ed. (1991). "Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques". Series "Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques". Los Alamitos, California.
- [Daubechies, 1992] Daubechies, I., Ed. (1992). "Ten lectures on wavelets". Series "Ten lectures on wavelets". Montpellier, Capitol City Press.
- [Defrise et al., 1997] Defrise, M., Kinahan, P. E., Townsend, D. W., Michel, C., Sibomana, M. and Newport, D. F. (1997). "Exact and approximate rebinning algorithms for 3-D PET data." *IEEE Transactions on Medical Imaging* vol. 16, n° 2, pp. 145-158.
- [Depas et al., 2004] Depas, G., De Barys, C. J., G., Hoyoux, C., Dresse, M.-F., Fassotte, M.-F., Paquet, N., Foidart, J., Rigo, P. and Hustinx, R. (2004). "18F-FDG PET in children with

- lymphomas." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 32, n° 1, pp. 31-38.
- [Devijver and Kittler, 1982] Devijver, P. A. and Kittler, J. V., Eds. (1982). "Pattern Recognition: a statistical approach". Series "Pattern Recognition: a statistical approach". Englewood Cliffs NJ: Prentice Hall/International
- [Dittmann et al., 2001] Dittmann, H., Sötkler, M., Kollmannsberger, C., Dohmen, B. M., Baumann, C., Kopp, A., Bares, R., Claussen, C. D., Kanz, L., et al. (2001). "Comparison of 18FDG-PET with CT scans in the evaluation of patients with residual and recurrent Hodgkin's lymphoma." *Oncology reports* vol. 8, n° 6, pp. 1393-1399
- [Doi, 2007] Doi, K. (2007). "Computer-aided diagnosis in medical imaging: historical review, current status and future potential." *Computer Medical Imaging Graphics* vol. 31, n° 4-5, pp. 198-211.
- [Duda and Hart, 1973] Duda, R. O. and Hart, P. E., Eds. (1973). "Pattern Classification and Scene Analysis". Series "Pattern Classification and Scene Analysis".
- [Duhaylongsod et al., 1995] Duhaylongsod, F. G., Lowe, V. J., Patz, E. F., Vaughn, A. L., Coleman, R. E. and Wolfe, W. G. (1995). "Detection of primary and recurrent lung cancer by means of F-18 fluorodeoxyglucose positron emission tomography (FDG PET) " *Journal of Thorac Cardiovascular Surgery* vol. 111, pp. 130-140.
- [Duin and Tax, 2000] Duin, R. P. W. and Tax, D. M. J. (2000). "Experiments with Classifier Combining Rules." *Multiple Classifier System*, pp. 16-29.
- [Efron and Tibshirani, 1997] Efron, B. and Tibshirani, R. (1997). "Improvements on cross-validation: The 0.632+ bootstrap method." *J. Am. Stat. Assoc.* vol. 92, pp. 548-560.
- [El-Naqa et al., 2002] El-Naqa, I., Yongyi, Y., Wernick, M. N., Galatsanos, N. P. and Nishikawa, R. (2002). "A support vector machine approach for detection of microcalcifications." *IEEE Transactions on Medical Imaging* vol. 21, n° 12, pp. 1552-1563.
- [Eschmann et al., 2002] Eschmann, S. M., Friedel, G., Paulsen, F., Budach, W., Harer-Mouline, C., Dohmen, B. M. and Bares, R. (2002). "FDG PET for staging of advanced non-small cell lung cancer prior to neoadjuvant radio-chemotherapy." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 29, n° 6, pp. 804-808.
- [Eubank and Mankoff, 2005] Eubank, W. B. and Mankoff, D. A. (2005). "Evolving role of positron emission tomography in breast cancer imaging." *Seminars in Nuclear Medicine* vol. 35, n° 2, pp. 84-99.
- [Feuardent et al., 2003] Feuardent, J., Soret, M., De Dreuille, O., Foehrenbach, H. and Buvat, I. (2003). "Reliability of SUV estimates in FDG PET as a function of acquisition and processing protocols". Piscataway, NJ 08855-1331, United States, Institute of Electrical and Electronics Engineers Inc., pp. 2877-2881.
- [Fisher, 1925] Fisher, R. A., Ed. (1925). "Statistical Methods for Research Workers". Series "Statistical Methods for Research Workers". Macmillan Pub Co; 15th edition (June 1970)
- [Fisher, 1936] Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." *Annals of Eugenics* vol. 7, pp. 179-188.
- [Frank et al., 1995] Frank, A., Lefkowitz, D., Jaeger, S., Gobar, L., Sunderland, J., Gupta, N., Scott, W., Mailliard, J., Lynch, H., et al. (1995). "Decision logic for retreatment of asymptomatic

- lung cancer recurrence based on positron emission tomography findings " *International Journal of Radiation Oncology Biology Physics* vol. 32, n° 5, pp. 1495-1512.
- [Freer and Ulisse, 2001] Freer, T. W. and Ulisse, M. J. (2001). "Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center." *Radiology* vol. 220, pp. 781–786.
- [Freund and Schapire, 1996] Freund, Y. and Schapire, R. E. (1996). "Experiments with a new boosting algorithm". *Proc. of 13th International Conference on Machine Learning*, pp. 148-156.
- [Freund and Schapire, 1997] Freund, Y. and Schapire, R. E. (1997). "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of Computer and System Sciences* vol. 55, n° 1, pp. 119--139.
- [Fukunaga and Hostetler, 1975] Fukunaga, K. and Hostetler, L. D. (1975). "k-nearest-neighbor Bayes-risk estimation." *IEEE Trans. Inform. Theory* vol. 21, pp. 285-293.
- [Fukunaga and Hayes, 1989] Fukunaga, K. and Hayes, R. R. (1989). "Estimation of classifier performance." *IEEE transactions on pattern analysis and machine intelligence* vol. 11, n° 10, pp. 1087-1101.
- [Garcia-Orellana et al., 2007] Garcia-Orellana, C. J., Gallardo-Caballero, R., Macias-Macias, M. and Gonzalez-Velasco, H. (2007). "SVM and Neural Networks comparison in mammographic CAD". *29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society, EMBC'07*, Lyon, France, Inst. of Elec. and Elec. Eng. Computer Society, pp. 3204-3207.
- [Ge et al., 2005] Ge, J., Wei, J., Hadjiiski, L. M., Sahiner, B., Chan, H.-P., Helvie, M. A. and Zhou, C. (2005). "Computer-aided detection of microcalcification clusters on full-field digital mammograms: Multiscale pyramid enhancement and false positive reduction using an artificial neural network". San Diego, CA, United states, SPIE, pp. 806-812.
- [Ge et al., 2005] Ge, Z. Y., Sahiner, B., Chan, H.-P., Hadjiiski, L. M., Cascade, P. N., Bogot, N., Kazerooni, E. A., Wei, J. and Zhou, C. (2005). "Computer-aided detection of lung nodules: False positive reduction using a 3D gradient field method and 3D ellipsoid fitting." *Medical Physics* vol. 32, pp. 2443–2454.
- [Giger et al., 1987] Giger, M. L., Doi, K. and MaxMahon, H. (1987). "Computerized detection of lung nodules in digital chest radiographs." *Proceeding SPIE* vol. 767, pp. 384-386.
- [Gil-Pita and Yao, 2008] Gil-Pita, R. and Yao, X. (2008). "Evolving edited k-nearest neighbor classifiers." *International Journal of Neural Systems* vol. 18, n° 6, pp. 459-467.
- [Gilbert et al., 2006] Gilbert, F. J., Astley, S. M., McGee, M. A., Gillan, M. G. C., Boggis, C. R. M., Griffiths, P. M. and Duffy, S. W. (2006). "Single Reading with Computer-aided Detection and Double Reading of Screening Mammograms in the United Kingdom National Breast Screening Program." *Radiology* vol. 241, pp. 47-53.
- [Gletsos et al., 2003] Gletsos, M., Mougiakakou, S. G., Matsopoulos, G. K., Nikita, K. S., Nikita, A. S. and Kelekis, D. (2003). "A Computer-Aided Diagnostic System to Characterize CT Focal Liver Lesions: Design and Optimization of a Neural Network Classifier." *IEEE Transactions on Information Technology in Biomedicine* vol. 7, n° 3, pp. 153-162.
- [Gould et al., 2001] Gould, M. K., Maclean, C. C., Kushner, W. G., Rydzak, C. E. and Owens, D. K. (2001). "Accuracy of positron emission tomography for diagnosis of pulmonary nodules and

- mass lesions: a meta-analysis." *Journal of the American Medical Association* vol. 285, pp. 914-924.
- [Grenier, 2005] Grenier, T. (2005). Apport de l'espace des caractéristiques et des paramètres d'échelle adaptatifs pour le filtrage et la segmentation d'image. CREATIS-UMR CNRS 5515-INSERM U 630. Villeurbanne, INSA, Lyon: 167.
- [Guan et al., 2006] Guan, H. Y., Kubota, T., Huang, X. L., Zhou, X. S. and Turk, M. (2006). "Automatic hot spot detection and segmentation in whole body FDG-PET images." *International Conference on Image Processing*, pp. 85-88.
- [Gurcan et al., 2001] Gurcan, M. N., Petrick, N., Sahiner, B., Chan, H. P., Cascade, P. N., Kazerooni, E. A. and Hadjiiski, L. M. (2001). "Computerized lung nodule detection on thoracic CT images: Combined rule-based and statistical classifier for false positive reduction". *Proceedings of SPIE - The International Society for Optical Engineering*, San Diego, CA, United States, SPIE, pp. 686-692.
- [Hamdi et al., 2008] Hamdi, N., Auhmani, K. and Hassani, M. M. R. (2008). "Design of a high-accuracy classifier based on Fisher discriminant analysis: Application to computer-aided diagnosis of microcalcifications". *Proceedings - The International Conference on Computational Sciences and its Applications, ICCSA 2008*, Perugia, Italy, Inst. of Elec. and Elec. Eng. Computer Society, pp. 267-273.
- [Haralick et al., 1973] Haralick, R. M., Shanmugam, K. and Dinstein, I. (1973). "Textural features for image classification." *IEEE Transactions on Systems, Man and Cybernetics* vol. smc 3, n° 6, pp. 610-621.
- [Harrison et al., 1993] Harrison, R. L., Vannoy, S. D., Haynor, D. R., Gillipsie, S. B., Kaplan, M. S. and Lewellen, T. K. (1993). "Preliminary experience with the photon history generator module of a public domain simulation system for emission tomography." *IEEE Nuclear Science Symposium Conference*, pp. 1154-1158.
- [Hatt et al., 2009] Hatt, M., Turzo, A., Roux, C. and Visvikis, D. (2009). "A fuzzy locally adaptive Bayesian segmentation approach for volume determination in PET." *IEEE Transactions on Medical Imaging* vol. to Print.
- [Haykin, 1999] Haykin, S. (1999). *Neural networks: a comprehensive foundation*. Prentice Hall.
- [Heath et al., 2001] Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W. P. and Yaffe, M. J. (2001). "The Digital Database for Screening Mammography." *Proceedings of the Fifth International Workshop on Digital Mammography*, pp. 212-218.
- [Herwig et al., 1999] Herwig, R., Poustka, A. J., Muller, C., Bull, C., Lehrach, H. and O'Brien, J. (1999). "Large-scale clustering of cDNA-fingerprinting data." *Genome Research* vol. 9, n° 11, pp. 1093-1105.
- [Ho, 1998] Ho, T. K. (1998). "The random subspace method for constructing decision forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* vol. 20, n° 8, pp. 832-844.
- [Hodgkin, 1832] Hodgkin, T. (1832). "On some morbid appearances of the absorbent glands and spleen." *Medical and Chirurgical Transactions* vol. 17, pp. 69-97.
- [Hoffman et al., 1999] Hoffman, E., Kletter, K., Diemling, M. B., M. Pfeffel, F. and Perkov, V. (1999). "Positron emission tomography with fluorine-18-2-fluoro-2-deoxyglucose (F18-FDG) does not visualise extranodal B-cell lymphoma of the mucosa-associated lymphoid tissue (MALT)-type." *Annals Oncology* vol. 10, pp. 1185-1189.

- [Hoskin, 2003] Hoskin, P. J. (2003). "PET in lymphoma: what are the oncologist's needs?" *European Journal of Nuclear Medicine and Molecular Imaging* vol. 30, n° suppl. 1, pp. 37-41.
- [Hotelling, 1931] Hotelling, H. (1931). "The generalization of Student's ratio." *Annals of Mathematical Statistics* vol. 2, pp. 360-378.
- [Huang et al., 2007] Huang, X. L., Zhou, X., Jerebko, A., Krshnan, A., Guan, H. Y., Kubota, T. and Potesil, V. (2007). System and method for whole body landmark detection, segmentation and change quantification in digital images. S. corporation. USA.
- [Hudson and Larkin, 1994] Hudson, H. M. and Larkin, R. S. (1994). "Accelerated image reconstruction using ordered subsets of projection data." *IEEE Transactions on Medical Imaging* vol. 13, n° 4, pp. 601-609.
- [Ikedo et al., 2007] Ikedo, Y., Fukuoka, D., Hara, T., Fujita, H., Takada, E., Endo, T. and Morita, T. (2007). "Computerized mass detection in whole breast ultrasound images: Reduction of false positives using bilateral subtraction technique". San Diego, CA, United states, SPIE, pp. SPIE.
- [Ireland et al., 2007] Ireland, R. H., Dyker, K. E., Barber, D. C., Wood, S. M., Hanney, M. B., Tindale, W. B., Woodhouse, N., Hoggard, N., Conway, J., et al. (2007). "Nonrigid image registration for head and neck cancer radiotherapy treatment planning with PET/CT." *International Journal Radiation Oncology Biology Physics* vol. 68, n° 3, pp. 952-957.
- [Jafar et al., 2006] Jafar, I., Hao, Y., Shields, A. F. and Muzik, O. (2006). "Computerized detection of lung tumors in PET/CT images". New York, NY, United states, Institute of Electrical and Electronics Engineers Inc., pp. 2320-2323.
- [Jakoby et al., 2006] Jakoby, B. W., Bercier, Y., Watson, C. C., Rappoport, V. Y., J., Bendriem, B. and Townsend, D. W. (2006). "Physical performance and clinical workflow of a new LSO HI-REZ PET/CT scanner." *Proceeding IEEE Nuclear Science Symposium and Medical Imaging Conference* vol. 5, n° 3130-3134.
- [Jan et al., 2004] Jan, S., Santi, G., Strul, D., Staelens, S., Assiè, K., Autret, D. and al., e. (2004). "GATE: a simulation toolkit for PET and SPECT." *Physics Medecine Biology* vol. 49, n° 4543-4561.
- [Jerusalem et al., 1999] Jerusalem, G., Beguin, Y. and Fassotte, M. F. (1999). "Whole-body positron emission tomography using F-18-fluorodeoxyglucose for post-treatment evaluation in Hodgkin's disease and non-Hodgkin's lymphoma has a higher diagnostic and prognostic value than classical computed tomography scan imaging." *Blood* vol. 94, pp. 429-433.
- [Jerusalem et al., 2001] Jerusalem, G., Beguin, Y., Fassotte, M. F., Naijar, F., Paulus, P., Rigo, P. and Fillet, G. (2001). "Whole-body positron emission tomography using 18F-fluorodeoxyglucose compared to standard procedures for staging patients with Hodgkin's disease." *Haematologica* vol. 86, pp. 266-273.
- [Jerusalem et al., 2003] Jerusalem, G., Hustinx, R., Beguin, Y. and Fillet, G. (2003). "PET scan imaging in oncology." *European Journal of Cancer* vol. 39, n° 11, pp. 1525-1534.
- [Jerusalem et al., 1999] Jerusalem, G., Warland, V., Naijar, F., Paulus, P., Fassotte, M. F., Fillet, G. and Rigo, P. (1999). "Whole body 18F-FDG PET for the evaluation of patients with Hodgkin's disease and non-Hodgkins lymphoma." *Nuclear Medicine Communications* vol. 20, pp. 13-20.
- [Jesneck et al., 2007] Jesneck, J. L., Lo, J. Y. and Baker, J. A. (2007). "Breast Mass Lesions: Computer-aided Diagnosis Models with Mammographic and Sonographic Descriptors." *Radiology* vol. 244, pp. 390-398.

- [Juweid and Cheson, 2005] Juweid, M. E. and Cheson, B. D. (2005). "Role of positron emission tomography in lymphoma." *Journal Clinical Oncology* vol. 23, pp. 4577-4580.
- [Juweid et al., 2007] Juweid, M. E., Stroobants, S., Hoekstra, O. S., Mottaghy, F. M., Dietlein, M., Guermazi, A., Wiseman, G. A., Kostakoglu, L., Scheidhauer, K., et al. (2007). "Use of positron emission tomography for response assessment of lymphoma: consensus of the imaging subcommittee of international harmonization project in lymphoma." *Journal Clinical Oncology* vol. 25, pp. 571-578.
- [Kabani et al., 1998] Kabani, N. J., MacDonald, D. J., Holmes, C. J. and Evans, A. C. (1998). "3D anatomical atlas of the human brain." *Neuroimage* vol. 7, n° 4, pp. 717.
- [Kadrmas, 2009] Kadrmas, J. D. (2009). "The Utah PET Lesion Detection Database". *Proceedings IEEE Medical Imaging Conference*, Orlando, USA, pp. M13-165.
- [Kanakatte et al., 2008] Kanakatte, A., Mani, N., Srinivasan, B. and Gubbi, J. (2008). "Pulmonary tumor volume detection from positron emission tomography images". Sanya, Hainan, China, Inst. of Elec. and Elec. Eng. Computer Society, pp. 213-217.
- [Kim et al., 2005] Kim, J. S., Kim, J. H., Cho, G. S. and Bae, K. T. (2005). "Automated detection of pulmonary nodules on CT images: Effect of section thickness and reconstruction interval – Initial results." *Radiology* vol. 236, pp. 295–299.
- [Kimura and Shridhar, 1991] Kimura, A. and Shridhar, M. (1991). "Handwritten numerical recognition based on multiple algorithms." *Pattern Recognition* vol. 24, n° 10, pp. 969-983.
- [Kinahan and Rogers, 1989] Kinahan, P. E. and Rogers, J. G. (1989). "Analytic 3D image reconstruction using all detected events." *IEEE Transactions on Nuclear Science* vol. 36, n° 1 pt 1, pp. 964-968.
- [Kittler et al., 1998] Kittler, J., Hatef, M., Duin, R. P. W. and Matas, J. M. (1998). "On Combining Classifiers." *IEEE Transactions on Pattern analysis and machine intelligence* vol. 20, n° 3, pp. 226-239.
- [Kohonen, 1982] Kohonen, T. (1982). "Self-organized formation of topologically correct feature maps." *Biological Cybernetics* vol. 43, n° 59-69.
- [Kole et al., 1998] Kole, A. C., Plukker, J. T., Nieweg, O. E. and Vaalburg, W. (1998). "Positron emission tomography for staging of oesophageal and gastrooesophageal malignancy." *Brittanic Journal Cancer* vol. 78, pp. 521-527.
- [Kubota and Espinal, 2000] Kubota, T. and Espinal, F. (2000). "Reaction-diffusion system for hypothesis propagation." *15th International Conference on Pattern Recognition (ICPR'00)* vol. 3, pp. 3547-3550.
- [Kumar et al., 2008] Kumar, R., Dhanpathi, H., Basu, S., Rubello, D., Fanti, S. and Alavi, A. (2008). "Oncologic PET tracers beyond [(18)F]FDG and the novel quantitative approaches in PET imaging." *The Quarterly Journal of Nuclear Medicine Molecular Imaging* vol. 52, n° 1, pp. 50-65.
- [Lamare et al., 2007] Lamare, F., Cresson, T., Savéan, J., Cheze-Le Rest, C., Reader, A. J. and Visvikis, D. (2007). "Respiratory motion correction for PET oncology applications using affine transformation of list mode data." *Physics Medecine Biology* vol. 52, pp. 121-140

- [Lartizien, 2001] Lartizien, C. (2001). Optimisation des protocoles d'imagerie corps entier en tomographie par émission de positons pour la détection des hyperfixations du 18F-FDG en oncologie. Paris, Université Paris XI. Thèse de doctora: 181p.
- [Lartizien et al., 2005] Lartizien, C., Goertzen, A., Reilhac, A., Siegel, S., Costes, N., Magnin, I., Gimenez, G. and Evans, A. C. (2005). "Validation of SORTEO Monte Carlo simulations for the geometry of the Concorde MicroPET R4 system". Puerto rico, Institute of Electrical and Electronics Engineers Inc., pp. 3036-3039.
- [Lartizien et al., 2004] Lartizien, C., Kinahan, P. E. and Comtat, C. (2004). "A lesion detection observer study comparing 2-dimensional versus fully 3-dimensional whole-body PET imaging protocols." *Journal Nuclear Medicine* vol. 45, n° 4, pp. 714-723.
- [Lartizien et al., 2003] Lartizien, C., Reilhac, A., Costes, N., Janier, M. and Sappey-Marinier, D. (2003). "Monte Carlo simulation based design study of a LSO-LuAP small animal PET system." *IEEE Transactions on Nuclear Science* vol. 50, n° 5, pp. 1433-1438.
- [Lartizien et al., 2009] Lartizien, C., Tomei, S., Le Pogam, A., Marache-Francisco, S. and Visvikis, D. (2009). "Impact of Partial Volume Correction in Whole-Body PET Imaging : a Computer-Aided Detection Study". *Proceedings IEEE Medical Imaging Conference*, Orlando.
- [Lartizien et al., 2009] Lartizien, C., Tomei, S. and Marache-Francisco, S. (2009). "Improving performance of automated detection systems for 3D PET oncology imaging by use of resampled training images". *proc. Computer Assisted Radiology and Surgery*, Berlin, Germany.
- [Le Maitre et al., 2009] Le Maitre, A., Segars, W. P., Reilhac, A., Hatt, M., Marache-Francisco, S., Tomei, S., Lartizien, C. and Visvikis, D. (2009). "Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications." *IEEE Transactions on Medical Imaging* vol. To appear.
- [Le Maitre et al., 2009] Le Maitre, A., W.P., S., Reilhac, A., Hatt, M., Tomei, S., Lartizien, C. and Visvikis, D. (2009). "Incorporating patient specific variability in a database of realistic simulated whole body 18F-FDG distributions for oncology applications". *Society of Nuclear Medicine annual meeting*.
- [Le Pogam et al., 2008] Le Pogam, A., Hatt, M., Boussion, N., Guilloteau, D., Baulieu, J.-L., Prunier, C., Turkheimer, F. and Visvikis, D. (2008). "Conditional partial volume correction for emission tomography: A wavelet-based hidden Markov model and multi-resolution approach". *IEEE International Symposium on Biomedical Imaging*, pp. 1319-1322.
- [Leung et al., 2008] Leung, F. H. F., Ling, S. H. and Lam, H. K. (2008). "An improved genetic-algorithm-based neural-tuned neural network." *International Journal of Computational Intelligence and Applications* vol. 7, n° 4, pp. 469-492.
- [Li et al., 2008] Li, H., Thorstad, W. L., Biehl KJ, Laforest, R., Su, Y., Shoghi, K. I., Donnelly, E. D., Low, D. A. and Lu, W. (2008). "A novel PET tumor delineation method based on adaptive region-growing and dual-front active contours." *Medical Physics* vol. 35, n° 8, pp. 3711-3721
- [Li et al., 2001] Li, L., Zheng, Y., Zhang, L. and Clark, R. A. (2001). "False-positive reduction in CAD mass detection using a competitive classification strategy." *Medical Physics* vol. 28, n° 2, pp. 250-258.

- [Li, 2007] Li, Q. (2007). "Recent progress in computer-aided diagnosis of lung nodules on thin-section CT." *Computerized Medical Imaging and Graphics* vol. 31, n° 4-5, pp. 248-257.
- [Li et al., 2008] Li, Q., Li, F. and Doi, K. (2008). "Computerized detection of lung nodules in thin section CT images by use of selective enhancement filters and an automated rule-based classifier." *Academic Radiology* vol. 15, pp. 165-175.
- [Li et al., 2003] Li, Q., Sone, S. and Doi, K. (2003). "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans." *Medical Physics* vol. 30, pp. 2040–2051.
- [Lin et al., 2005] Lin, D.-T., Yan, C.-R. and Chen, W.-T. (2005). "Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system." *Computerized Medical Imaging and Graphics* vol. 29, n° 6, pp. 447-458.
- [Loening and Gambhir, 2003] Loening, A. M. and Gambhir, S. S. (2003). "AMIDE: a free software tool for multimodality medical image analysis." *Journal of Molecular Imaging* vol. 2, n° 3, pp. 131-137.
- [Lu et al., 2009] Lu, L., Wanyu, L. and Xiaoming, S. (2009). "Automated detection of pulmonary nodules in CT images with support vector machines". *Proceedings of SPIE - The International Society for Optical Engineering*, Shenyang, China, SPIE.
- [Maisey et al., 2000] Maisey, N. R., Hill, M. E. and Webb, A. (2000). "Are 18fluorodeoxyglucosepositron emission tomography and magnetic resonance imaging useful in the prediction of relapse in lymphoma residual masses?" *European Journal of Cancer* vol. 36, pp. 200-206.
- [Makela et al., 2003] Makela, T., Pham, Q. C., Clarysse, P., Lotjonen, J., Sipila, O., Hanninen, H., Lauerma, K., Knutti, J., Katila, T., et al. (2003). "A 3D model-based registration approach for the PET, MR and MCG cardiac data fusion." *Medical Imaging Analysis* vol. 7, n° 3, pp. 377-389.
- [Marten and Engelke, 2007] Marten, K. and Engelke, C. (2007). "Computer-aided detection and automated CT volumetry of pulmonary nodules." *European Radiology* vol. 17, n° 4, pp. 888-901.
- [Masotti et al., 2009] Masotti, M., Lanconelli, N. and Campanini, R. (2009). "Computer-aided mass detection in mammography: false positive reduction via gray-scale invariant ranklet texture features." *Medical Physics* vol. 36, n° 2, pp. 311-316.
- [Mckeighen, 1996] Mckeighen, R. E. (1996). "Optimization of broadband transducer designs by use of statistical design of experiments." *IEEE Trans. Ultrasonics Ferroelectrics and Freq. Control* vol. 43, n° 1, pp. 63-70.
- [Meltzer et al., 2000] Meltzer, C. C., Luketish, J. D., Friedman, D., Charron, M., Strollo, D., Meehan, M., Urso, G. K., Dachille, M. A. and Townsend, D. W. (2000). "Whole-body FDG positron emission tomographic imaging for staging oesophageal cancer: comparison with computed tomography." *Clinical Nuclear Medicine* vol. 25, pp. 882-887.
- [Mercer, 1909] Mercer, J. (1909). "Functions of positive and negative type and their connection with the theory of integral equations." *Philos. Trans. Roy. Soc. London* vol. 209, pp. 415-446.
- [Metropolis, 1947] Metropolis, N. (1947). "The Beginning of the Monte Carlo Method " *Los Alamos Science* vol. 15, pp. 125-130.

- [Metropolis and Stanislas Ulam, 1949] Metropolis, N. and Stanislas Ulam, S. (1949). "The Monte Carlo Method." *Journal of the American Statistical Association* vol. 44, n° 247, pp. 335-341.
- [Metz, 1986] Metz, C. E. (1986). "ROC methodology in radiologic imaging." *Investigative Radiology* vol. 21, n° 9, pp. 720-733.
- [Mierswa et al., 2006] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. and Euler, T. (2006). "YALE: Rapid prototyping for complex data mining tasks". *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, United states, Association for Computing Machinery, pp. 935-940.
- [Mikhaeel et al., 2000] Mikhaeel, N. G., Timothy, A. R., Odoherly, M. J., Hain, S. and Maisey, M. N. (2000). "18-FDG-PET as a prognostic indicator in the treatment of aggressive non-Hodgkin's lymphoma, comparison with CT." vol. 39, pp. 543-553.
- [Morrison and Linnett, 1999] Morrison, S. and Linnett, L. M. (1999). "Model based approach to object detection in digital mammography". *IEEE International Conference on Image Processing*, Kobe, Jpn, IEEE, pp. 182-186.
- [Mougiakakou et al., 2007] Mougiakakou, S. G., Valavanis, I. K., Nikita, A. S. and Konstantina, S. N. (2007). "Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers." *Artificial Intelligence in Medicine* vol. 41, n° 1, pp. 25-37.
- [Nandi et al., 2006] Nandi, R. J., Nandi, A. K., Rangayyan, R. M. and Scutt, D. (2006). "Classification of breast masses in mammograms using genetic programming and feature selection." *Medical and biological engineering and computing* vol. 44, pp. 683-694.
- [Nehmeh et al., 2003] Nehmeh, S. A., Erdi, S. E., Rosenzweig, K. E., Schoder, H., Larson, S. M., Squire, O. D. and Humm, J. L. (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* vol. 44, n° 10, pp. 1644-1648.
- [Nie et al., 2006] Nie, Y., Li, Q., Li, F., Pu, Y., Appelbaum, D. and Doi, K. (2006). "Integrating PET and CT information to improve diagnostic accuracy for lung nodules: a semiautomatic computer-aided method." *Journal Nuclear Medicine* vol. 47, n° 7, pp. 1075-1080.
- [O'Doherty et al., 2002] O'Doherty, M. J., MacDonalds, E., Barrington, S. F., Mikhaeel, N. G. and Schey, S. (2002). "Positron emission tomography in the management of lymphomas." *Clinical Oncology* vol. 14, pp. 415-426.
- [Paik et al., 2004] Paik, D. S., Beaulieu, C. F., Rubin, G. D., Acar, B., Jeffrey, R. B., Yee, J., Dey, J. and Napel, S. (2004). "Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT." *IEEE Transactions on Medical Imaging* vol. 23, pp. 661-675.
- [Pan et al., 2005] Pan, T., Mawlawi, O., Nehmeh, S. A., Erdi, Y. E., Luo, D., Liu, H. H., Castillo, R., Mohan, R., Liao, Z., et al. (2005). "Attenuation correction of PET images with respiration-averaged CT images in PET/CT." *Journal Nuclear Medicine* vol. 46, n° 9, pp. 1481-1487.
- [Papadopoulos et al., 2008] Papadopoulos, A., Fotiadis, D. I. and Costaridou, L. (2008). "Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques." *Computers in Biology and Medicine* vol. 10, pp. 1045-1055.

- [Park et al., 2009] Park, S. C., Pu, J. and Zheng, B. (2009). "Improving performance of computer-aided detection scheme by combining results from two machine learning classifiers." *Academic Radiology* vol. 16, n° 3, pp. 266-274.
- [Partridge et al., 2000] Partridge, S., Timothy, A., O'Doherty, M. J., Hain, S. F., Rankin, S. and Mickhaeel, G. (2000). "2-fluorine-18-fluoro-2-deoxyglucose positron emission tomography in the pretreatment staging of Hodgkin's disease: influence on patient management in a single institution. ." *Annals Oncology* vol. 11, pp. 1273-1279.
- [Patlak et al., 1983] Patlak, C. S., Blasberg, R. G. and Fenstermacher, J. D. (1983). "Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data." *Journal of Cerebral Blood Flow and Metabolism* vol. 3, n° 1, pp. 1-7.
- [Pearson, 1901] Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* vol. 2, n° 6, pp. 559-572.
- [Perruchot et al., 2004] Perruchot, F., Reilhac, A., Grova, C., Evans, A. C. and Dagher, A. (2004). "Motion correction of multiframe PET data." *IEEE Nuclear Science Symposium Conference* vol. 5, pp. 3186-3190.
- [Press et al., 1992] Press, W. H., Teukolsky, S. A., William, T., Vetterling, W. T. and Flannery, B. P., Eds. (1992). "Numerical Recipes in C: The Art of Scientific Computing". Series "Numerical Recipes in C: The Art of Scientific Computing".
- [Pudil et al., 1994] Pudil, P., Novovicova, J. and Kittler, J. (1994). "Floating search methods in feature selection." *Pattern Recognition* vol. 15, pp. 1119-1125.
- [Qiao et al., 2006] Qiao, F., Pan, T., Clark, J. W. and Mawlaw, O. R. (2006). "A motion-incorporated reconstruction method for gated PET studies." *Physics Medicine Biology* vol. 51, pp. 3769-3783.
- [Rabiner, 1989] Rabiner, L. R. (1989). "A tutorial on Hidden Markov Models and selected applications in speech recognition." *Proceedings of the IEEE* vol. 77, n° 2, pp. 257-286.
- [Raicu et al., 2007] Raicu, D. S., Varutbangkul, E., Cisneros, J. G., Furst, J. D., Channin, D. S. and Armato Iii, S. G. (2007). "Semantics and image content integration for pulmonary nodule interpretation in thoracic computed tomography". San Diego, CA, United states, SPIE, pp. SPIE.
- [Regentova et al., 2007] Regentova, E., Zhang, L., Zheng, J. and Veni, G. (2007). "Microcalcification detection based on wavelet domain hidden Markov tree model: Study for inclusion to computer aided diagnostic prompting system." *Medical Physics* vol. 34, n° 6, pp. 2206-2219.
- [Reilhac, 2007] Reilhac, A. (2007). Validation et exploitation d'un simulateur TEP Monte Carlo. Institut National des Sciences Appliquées de Lyon. Lyon, Institut National des Sciences Appliquées de Lyon. PhD thesis: 182.
- [Reilhac et al., 2005] Reilhac, A., Batan, G., Michel, C., Grova, C., Tokha, J., Collins, D. L., Costes, N. and Evans, A. C. (2005). "PET SORTEO: Validation and development of a database of simulated PET volumes." *IEEE Transactions on Nuclear Science* vol. 52, n° 5, pp. 1321-1328.
- [Reilhac et al., 2004] Reilhac, A., Rousset, O. G., Comtat, C., Frouin, V., Gregoire, M. C. and Evans, A. C. (2004). "A correction algorithm for partial volume effects in 3D PET imaging: principle and validation." *IEEE Nuclear Science Symposium Conference* vol. 7, pp. 4125-4126.

- [Rioul and Vetterli, 1991] Rioul, O. and Vetterli, M. (1991). "Wavelets and signal processing." *IEEE Signal Processing Mag.*, pp. 14-38.
- [Riyahi-Alam et al., 2004] Riyahi-Alam, N., Ahmadian, A., Tehrani, J. N., Guiti, M., Oghabian, M. A. and Deldari, A. (2004). "Segmentation of suspicious clustered microcalcifications on digital mammograms: Using fuzzy logic and wavelet coefficients". *2004 7th International Conference on Signal Processing Proceedings, ICSP*, Beijing, China, Institute of Electrical and Electronics Engineers Inc., pp. 2228-2230.
- [Rizzo et al., 2005] Rizzo, G., Castiglioni, I., Arienti, R., Cattaneo, G. M., Landoni, C., Artioli, D., Gilardi, M. C., Messa, C., Reni, M., et al. (2005). "Automatic registration of PET and CT studies for clinical use in thoracic and abdominal conformal radiotherapy." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 49, n° 3, pp. 267-279.
- [Rohren et al., 2004] Rohren, E. M., Turkington, T. G. and Coleman, R. E. (2004). "Clinical Applications of PET in Oncology." *Radiology* vol. 231, pp. 305-332.
- [Rousset et al., 2008] Rousset, O. G., Collins, D. L., Rahmim, A. and Wong, D. F. (2008). "Design and implementation of an automated partial volume correction in PET: application to dopamine receptor quantification in the normal human striatum." *Journal Nuclear Medicine* vol. 49, n° 7, pp. 1097-1106.
- [Saad et al., 2008] Saad, A., Hamarneh, G., Moller, T. and Smith, B. (2008). "Kinetic Modeling Based Probabilistic Segmentation for Molecular Images". *Proceeding Medical Image Computing and Computer-Assisted Intervention, MICCAI 2008*, New-York USA.
- [Sachs et al., 1971] Sachs, M. B., Nachmias, J. and ROBSON, J. G. (1971). "Spatial-Frequency Channels in Human Vision." *Journal Optical Society of America* vol. 61, pp. 1176-1186.
- [Sahiner et al., 2000] Sahiner, B., Chan, H.-P., Petrick, N., Wagner, R. F. and Hadjiiski, L. M. (2000). "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size." *Medical Physics* vol. 27, n° 7, pp. 1509-1522.
- [Sahiner et al., 2001] Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A. and Hadjiiski, L. M. (2001). "Improvement of mammographic mass characterization using spiculation measures and morphological features." *Medical Physics* vol. 28, n° 7, pp. 1455-1465.
- [Sajda et al., 2002] Sajda, P., Spence, C. and Pearson, J. (2002). "Learning contextual relationships in mammograms using a hierarchical pyramid neural network." *IEEE Transactions on Medical Imaging* vol. 21, pp. 239-250.
- [Samulski et al., 2007] Samulski, M., Karssemeijer, N., Lucas, P. and Groot, P. (2007). "Classification of mammographic masses using support vector machines and Bayesian networks." *Proceeding SPIE Medical Imaging*.
- [Sankar and Wang, 1996] Sankar, K. and Wang, P. P., Eds. (1996). "Genetic Algorithms for Pattern Recognition". Series "Genetic Algorithms for Pattern Recognition". Duke University, CRC-Press.
- [Segall, 2001] Segall, G. M. (2001). "FDG PET imaging in patients with lymphoma: a clinical perspective." *Journal Nucl Med.* vol. 42, pp. 609-610.
- [Segars et al., 2009] Segars, W., Mahesh, M., Beck, T. J., Frey, E. C. and Tsui, B. M. (2009). "Realistic CT simulation using the 4D XCAT phantom." *Medical Physics* vol. 35, n° 8, pp. 3800-3808.

- [Segars, 2001] Segars, W. P. (2001). Development of a new dynamic NURBS-based cardiac-torso (NCAT) phantom. Chapel Hill, The University of North Carolina. PhD Thesis.
- [Shensa, 1992] Shensa, J. (1992). "The discrete wavelet transform: wedding the à trous and Mallat algorithms." *IEEE Transactions on Signal Processing* vol. 40, pp. 2462-2482.
- [Shepp and Vardi, 1982] Shepp, L. A. and Vardi, Y. (1982). "Maximum likelihood reconstruction for emission tomography." *IEEE Transactions on Medical Imaging* vol. MI-1, n° 2, pp. 113-122.
- [Soret et al., 2007] Soret, M., Bacharach, S. L. and Buvat, I. (2007). "Partial-volume effect in PET tumor imaging." *Journal Nuclear Medicine* vol. 48, n° 6, pp. 932-945.
- [Souvatzoglou et al., 2007] Souvatzoglou, M., Ziegler, S. I., Martinez, M. J., Bush, R., Dzewas, G., Schwaiger, M. and Bengel, F. (2007). "Standardized uptake values from PET/CT images: comparison with conventional attenuation-corrected PET." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 34, n° 3, pp. 405-412.
- [Spaepen et al., 2000] Spaepen, K., Stroobants, S. and Dupont, P. (2000). "The prognostic value of positron emission tomography (PET) with ¹⁸F fluorodeoxyglucose (18FDG) after first-line chemotherapy in non-Hodgkin's lymphoma: 18FDG-PET, a valid alternative to conventional diagnostic methods." *Journal Clinical Oncology* vol. 19, pp. 414-419.
- [Strickland and Hahn, 1996] Strickland, R. N. and Hahn, H. H. (1996). "Wavelet transforms for detecting microcalcifications in mammograms." *IEEE Transactions on Medical Imaging* vol. 15, n° 2, pp. 218-229.
- [Strickland and Hahn, 1997] Strickland, R. N. and Hahn, H. H. (1997). "Wavelet transform methods for object detection and recovery." *IEEE Transactions on Medical Imaging* vol. 6, n° 5, pp. 724-735.
- [Stumple et al., 1998] Stumple, K. D. M., Urbinelli, M. and Steinert, H. C. (1998). "Whole body positron emission tomography using fluorodeoxyglucose for staging of lymphoma." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 25, pp. 721-728.
- [Stute et al., 2009] Stute, S., Necib, H., Grotus, N., Tylski, P., Rehfeld, N. and Buvat, I. (2009). "LuCaS2: Efficient Monte Carlo Simulations of Serial PET Scans for Assessing Detection and Quantification Methods Used in Patient Monitoring". *Proceedings IEEE Medical Imaging Conference*, pp. M13-357.
- [Stute et al., 2008] Stute, S., Tylski, P., Grotus, N. and Buvat, I. (2008). "LuCaS: Efficient Monte Carlo simulations of highly realistic PET tumor images". *Proceeding IEEE Medical Imaging Conference*, Dresden, Germany, pp. 4010-4012.
- [Suckling et al., 1994] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., et al. (1994). "The mammographic images analysis society digital mammogram database." *Exerpta Medica. International Congress* vol. 1069, pp. 375-378.
- [Sugawara et al., 1999] Sugawara, Y., Zasadny, K. R., Neuhoff, K. W. and Wahl, R. L. (1999). "Reevaluation of the standardized uptake value for FDG: variations with body weight and methods for corection." *Radiology* vol. 21, pp. 521-525.
- [Surti et al., 2007] Surti, S. K., A. Werner, M.E., Perkins, A. E. and Kolthammer, J. K., J.S. (2007). "Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities." *Journal Nuclear Medicine* vol. 48, n° 3, pp. 471-480.

- [Swensson, 1996] Swensson, R. G. (1996). "Unified measurement of observer performance in detecting and localizing target objects on images." *Medical Physics* vol. 23, n° 10, pp. 1709-1725.
- [Swets and Pickett, 1982] Swets, J. A. and Pickett, R. M., Eds. (1982). "Evaluation of Diagnostic Systems: Methods from Signal Detection Theory". Series "Evaluation of Diagnostic Systems: Methods from Signal Detection Theory". London.
- [Tabary et al., 2007] Tabary, J., Hugonnard, P. and Mathy, F. (2007). "SINDBAD: a realistic multi-purpose and scalable X-ray simulation tool for NDT applications". *Proceedings International Symposium on DIR and CT*, Lyon.
- [Tan and Ong, 2004] Tan, L. T. H. and Ong, K. L. (2004). "Semi-quantitative measurements of normal organs with variable metabolic activity on FDG PET imaging." *Annals of the Academy of Medicine, Singapore* vol. 33, pp. 183-185.
- [Tang et al., 2008] Tang, X., Xu, X., Zhou, W. and Mao, Y. (2008). "A novel semantic-based CAD system for mammography". Singapore, Singapore, Inst. of Elec. and Elec. Eng. Computer Society, pp. 97-101.
- [Tax et al., 2000] Tax, D. M. J., van Breukelen, M., Duin, R. P. W. and Kittler, J. (2000). "Combining multiple classifiers by averaging or by multiplying?" *Pattern Recognition* vol. 33, n° 9, pp. 1475-1485.
- [Teras and Tolvanen, 2007] Teras, M. and Tolvanen, T. J., J. Williams, J. Knuuti, J (2007). "Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 34, n° 10, pp. 1683-1692.
- [Tokha and Reilhac, 2007] Tokha, J. and Reilhac, A. (2007). "Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method " *Neuroimage* vol. 39, n° 4, pp. 1570-1584.
- [Tomei et al., 2009] Tomei, S., Marache-Francisco, S., Odet, C. and C, L. (2009). "Automatic detection of lesions in 3D PET images using a LDA and SVM: II performances comparison and combination." *submitted in IEEE Transactions on Medical Imaging*.
- [Tomei et al., 2008] Tomei, S., Marache-Francisco, S., Odet, C. and Lartizien, C. (2008). Automatic detection of active nodules in 3D PET oncology imaging using the Hotelling Observer and the Support Vector Machines: a comparison study. *Proceedings IEEE Medical Imaging Conference. Dresden: 5314-5319.*
- [Tomei et al., 2009] Tomei, S., Marache-Francisco, S., Odet, C. and Lartizien, C. (2009). "Automatic detection of lesions in 3D PET images using a LDA and SVM: I description and optimization of the LDA classification method." *Submitted to IEEE Transactions on Medical Imaging*.
- [Tomei et al., 2007] Tomei, S., Odet, C. and Lartizien, C. (2007). "Détection d'hypersignaux en Tomographie par Emission de Positons 3D : Approche statistique basée sur la combinaison d'une transformée en ondelettes avec un test de Hotelling ". *GRETSI*, pp. 869-872.
- [Tomei et al., 2009] Tomei, S., Reilhac, A., Visvikis, D., Boussion, N., Odet, C., Giammarile, F. and Lartizien, C. (2009). "oncoPET_DB: a freely distributed database of realistic simulated whole body [18F]FDG PET images for oncology." *IEEE Transactions on Nuclear Science* vol. (submitted).

- [Tomei et al., 2008] Tomei, S., Reilhac, A., Visvikis, D., Odet, C., Giammarile, F., Mognetti, T. and Lartizien, C. (2008). "Development of a database of realistic simulated whole body [18F]FDG PET images for lymphoma." *Proceedings IEEE Medical Imaging Conference*, pp. 4958-4963.
- [Tourassi et al., 2008] Tourassi, G. D., Ike, R., Singh, S. and Harrawood, B. (2008). "Evaluating the effect of image preprocessing on an information-theoretic CAD system in mammography." *Academic Radiology* vol. 15, n° 5, pp. 626-634.
- [Toussaint, 1974] Toussaint, G. T. (1974). "Bibliography on estimation of misclassification." *IEEE Trans. Inf. Theory* vol. 20, pp. 472-479.
- [Tozaki et al., 2004] Tozaki, T., Fujimoto, T., Sakamoto, S., Matsumoto, K. and Senda, M. (2004). "True positive reduction method for cancer diagnosis based on FDG-PET images." *Proceedings IEEE Medical Imaging Conference* vol. 5, pp. 3022-3025.
- [Tozaki et al., 2003] Tozaki, T., Senda, M., Sakamoto, S. and Matsumoto, K. (2003). "Computer assisted diagnosis method of whole body cancer using FDG-PET images". Barcelona, Spain, Institute of Electrical and Electronics Engineers Computer Society, pp. 1085-1088.
- [Tsui et al., 1993] Tsui, B. M., Terry, J. A. and Gullberg, G. T. (1993). "Evaluation of cardiac cone-beam single photon emission computed tomography using observer performance experiments and receiver operating characteristic analysis." *Investigative radiology* vol. 28, n° 12, pp. 1101-1112.
- [Turkheimer et al., 2008] Turkheimer, F., Boussion, N., Anderson, A. N., Pavese, N., Piccini, P. and Visvikis, D. (2008). "PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets." *Journal Nuclear Medicine* vol. 49, n° 4, pp. 657-666.
- [Tylski et al., 2008] Tylski, P., Dusart, M., Vanderlinden, B. and Buvat, I. (2008). "Assigning statistical significance to tumor changes in patient monitoring using FDG pet". *IEEE International Symposium on Biomedical Imaging* pp. 121-124.
- [Vapnik and Lerner, 1963] Vapnik, V. and Lerner, A. (1963). "Pattern Recognition using Generalized Portrait Method." *Automation and Remote Control* vol. 24.
- [Vapnik, 1995] Vapnik, V. N. (1995). "The nature of statistical learning theory". New-York.
- [Visvikis et al., 2006] Visvikis, D., Lamare, F., Bruyant, P., Boussion, N. and Cheze Le Rest, C. (2006). "Respiratory motion in positron emission tomography for oncology applications : Problems and solutions." *Nuclear instruments & methods in physics research* vol. 569, n° 2, pp. 453-457.
- [Wang et al., 2007] Wang, Z., Lin, J. C., Mao, W., Liu, W., Smith, M. B. and Collins, C. M. (2007). "SAR and temperature: simulations and comparison to regulatory limits for MRI." *Journal Magnetic Resonance* vol. 26, pp. 437-441.
- [Warburg et al., 1930] Warburg, O., Wind, F. and Neglers, E. (1930). "On the metabolism of tumors in the body." *London, Constable*.
- [Ward, 1963] Ward, J. H. (1963). "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* vol. 58, n° 301, pp. 236-244.
- [Weber et al., 2003] Weber, G., Petersen, V., Schmidt, B., Tyndale-Hines, L., Link, T., Peschel, C. and Schwaiger, M. (2003). "Positron Emission Tomography in Non-Small-Cell Lung Cancer:

- Prediction of Response to Chemotherapy by Quantitative Assessment of Glucose Use " *Journal of Clinical Oncology* vol. 21, n° 14, pp. 2651-2657.
- [Wilms, 1865] Wilms, S. (1865). "Cases of enlargement of the lymphatic glands and spleen (or Hodddkin's disease) with remarks." *Guy Hospital Rep.* vol. 11, pp. 56-67.
- [Wong et al., 1997] Wong, W. L., Chevetton, E. B., McGurk, M., Hussain, K., Davis, J., Beaney, R., Baddeley, H., Tierney, P. and Maisey, M. (1997). "A prospective study of PET-FDG imaging for the assessment of head and neck squamous cell carcinoma." *Clinical Otolaryngology and Allied Science* vol. 22, n° 3, pp. 209-214.
- [Workman Jr and Coleman, 2006] Workman Jr, R., B. and Coleman, R. E. (2006). PET in Lymphoma. PET/CT: 55-74.
- [Wu et al., 2006] Wu, Y.-T., Hadjiiski, L. M., Wei, J., Zhou, C., Sahiner, B. and Chan, H.-P. (2006). "Computer-aided detection of breast masses on mammograms: Bilateral analysis for false positive reduction". San Diego, CA, United states, SPIE.
- [Wu et al., 2002] Wu, Y., Ianakiev, K. and Govindaraju, V. (2002). "Improved k-nearest neighbor classification " *Pattern Recognition* vol. 35 pp. 2311–2318.
- [Yeung et al., 1999] Yeung, H., Macapinlac, H., Mazumdar, M. B., M., Finn, R. and Larson, S. M. (1999). "FDG PET in oesophageal cancer: incremental value over computed tomography." *Clinical Positron Imaging* vol. 2, pp. 255-260.
- [Yeung et al., 2002] Yeung, H. W., Sanches, A., Squire, O. D., Macapinlac, H. A., Larson, S. M. and Erdi, Y. E. (2002). "Standardized uptake value in pediatric patients: an investigation to determine the optimum measurement parameter." *European Journal of Nuclear Medicine and Molecular Imaging* vol. 29, pp. 61-66.
- [Ying et al., 2004] Ying, H., Zhou, F., Shields, A. F., Muzik, O., Wu, D. and Heath, E. I. (2004). "A novel computerized approach to enhancing lung tumor detection in whole-body PET images". *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, San Francisco, CA, United states, Institute of Electrical and Electronics Engineers Inc., pp. 1589-1592.
- [Yoshida, 2004] Yoshida, H. (2004). "Local contralateral subtraction based on bilateral symmetry of lung for reduction of false positives in computerized detection of pulmonary nodules." *IEEE Transactions on Biomedical Engineering* vol. 51, n° 5, pp. 778 - 789.
- [Yoshida et al., 2002] Yoshida, H., Dachman, A., Nappi, J., Maceneaney, P. R., D. and Masutani, Y. L., L. (2002). Method for computer-aided detection of three-dimensional lesions, University of Chicago. US 2003/0223627 A1.
- [Yu and Guan, 2000] Yu, S. and Guan, L. (2000). "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films." *IEEE Transactions on Medical Imaging* vol. 19, n° 2, pp. 115-126.
- [Yu, 2000] Yu, S. G., L. (2000). "A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films." *IEEE Transactions on Medical Imaging* vol. 19, n° 2, pp. 115-126.
- [Zaidi et al., 2007] Zaidi, H., Montandon, M.-L. and Meikle, S. (2007). "Strategies for attenuation compensation in neurological PET studies." *Neuroimage* vol. 34, n° 2, pp. 518-541.

- [Zaidi and Morel, 1999] Zaidi, H. S., A.H. and Morel, C. (1999). "An object-oriented monte carlo simulator for 3D cylindrical positron tomographs." *Computer Methods and Programs in Biomedicine* vol. 58, n° 2, pp. 133-145.
- [Zasadny and Whal, 1993] Zasadny, K. R. and Whal, R. L. (1993). "Standardized uptake values of normal tissues in PET with 2-[fluorine-18]fluoro-2-deoxy-D-glucose: variation with body weight and a method for correction." *Radiology* vol. 189, pp. 847-850.
- [Zhang and Jiang, 2007] Zhang, X.-L. and Jiang, C. (2007). "Improved SVM for Learning Multi-Class Domains with ROC Evaluation." *Proc. International Conference on Machine Learning and Cybernetics* vol. 5, pp. 2891-2896.
- [Zhao et al., 2003] Zhao, B. S., Gamsu, G., Ginsburg, M. S., Jiang, L. and Schwartz, L. H. (2003). "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm." *J Applied Clin Med Phys* vol. 4, pp. 248–260.
- [Zhao et al., 2007] Zhao, L., Boroczky, L., Drysdale, J., Agnihotri, L. and Lee, M. C. (2007). "False positive reduction for lung nodule CAD". *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*, San Diego, CA, United states, SPIE, pp. SPIE.
- [Zhen and Chan, 2001] Zhen, L. and Chan, A. K. (2001). "An artificial intelligent algorithm for tumor detection in screening mammogram." *IEEE Transactions on Medical Imaging* vol. 20, n° 7, pp. 559 - 567.
- [Zhuang and Alavi, 2002] Zhuang, H. and Alavi, A. (2002). "18-Fluorodeoxyglucose positron emission tomographic imaging in the detection and monitoring of infection and inflammation." *Semin. Nucl. Med.* vol. 32, pp. 47-59.
- [Zincirkeser et al., 2007] Zincirkeser, S., Sahn, E., Halac, M. and Sager, S. (2007). "Standardized Uptake Values of Normal Organs on 18F-Fluorodeoxyglucose Positron Emission Tomography and Computed Tomography Imaging." *The Journal of International Medical Research* vol. 35, pp. 231-236.
- [Zubal et al., 1994] Zubal, I. G., Harrell, C. R., Smith, E. O., Rattner, Z., Gindi, G. and Hoffer, P. B. (1994). "Computerized three-dimensional segmented human anatomy." *Medical Physics* vol. 21, n° 2, pp. 299-302.

Bibliographie personnelle

Cette thèse a donné lieu aux travaux suivants :

Publications

S. Tomei, S. Marache-Francisco, C. Odet, C. Lartizien, « Automatic detection of lesions in 3D PET images using a LDA and SVM: I description and optimization of the LDA classification model », Soumis pour publication au journal IEEE Transactions on Medical Imaging 2009.

S. Tomei, S. Marache-Francisco, C. Odet, C. Lartizien, « Automatic detection of lesions in 3D PET images using a LDA and SVM: II performances comparison and combination », Soumis pour publication au journal IEEE Transactions on Medical Imaging 2009.

S. Tomei, A. Reilhac, D. Visvikis, N. Boussion, C. Odet, F. Giammarile, C. Lartizien, “oncoPET_DB: a freely distributed database of realistic simulated whole body 18F-FDG PET images for oncology”, Accepté pour publication au journal IEEE Transactions on Nuclear Science 2009.

A. Le Maitre, W.P. Segars, A. Reilhac, M. Hatt, S. Marache-Francisco, S. Tomei, C. Lartizien, D. Visvikis, « Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications », Accepté pour publication au journal IEEE Transactions on Medical Imaging en juillet 2009.

Congrès nationaux et internationaux

C. Lartizien, S. Tomei, V. Maxim, C. Odet, “Combining a wavelet transform with a channelized Hotelling observer for tumor detection in 3D PET”, SPIE Medical Imaging, San Diego, United States, February 2007.

S. Tomei, C. Odet, C. Lartizien, “Détection d'hypersignaux en Tomographie par Emission de Positons 3D: Approche statistique basée sur la combinaison d'une transformée en ondelettes avec un test de Hotelling”, GRETSI 2007, Troyes, France, pages 869-872, September 2007.

S. Tomei, A. Reilhac, D. Visvikis, N. Boussion, C. Odet, F. Giammarile, T. Mognetti, C. Lartizien, ‘Development of a database of realistic simulated whole-body 18F-FDG PET images for lymphoma’, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC), Dresden, Germany, pages 4958-4963, October 2008.

S. Tomei, S. Marache-Francisco, C. Odet, C. Lartizien, “Automatic detection of active nodules in 3D PET oncology imaging using the Hotelling Observer and the Support Vector Machines: a comparison study”, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC), Dresden, Germany, pages 5314-5319, October 2008.

A. Le Maitre, W.P. Segars, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, D. Visvikis, “Incorporating patient specific variability in a database of realistic simulated whole body 18F-FDG distributions for oncology applications”, Society of Nuclear Medicine annual meeting (SNM), Toronto, Canada, June 2009.

C. Lartizien, S. Tomei, S. Marache-Francisco, “Improving performance of automated detection systems for 3D PET oncology imaging by use of resampled training images”, Computer Assisted Radiology and Surgery (CARS), Berlin, Germany, June 2009.

C. Lartzien, S. Tomei, S. Marache-Francisco, A. Le Pogam, D. Visvikis, “Impact of Partial Volume Correction in Whole-Body PET Imaging: a Computer-Aided Detection Study”, IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS-MIC), Orlando, USA, October 2009.

ECOLE DOCTORALE

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	<u>CHIMIE DE LYON</u> http://sakura.cpe.fr/ED206 M. Jean Marc LANCELIN Insa : R. GOURDON	M. Jean Marc LANCELIN Université Claude Bernard Lyon 1 Bât CPE 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72.43 13 95 Fax : lancelin@hikari.cpe.fr
E.E.A.	<u>ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE</u> http://www.insa-lyon.fr/eea M. Alain NICOLAS Insa : C. PLOSSU Secrétariat : M. LABOUNE	M. Alain NICOLAS Ecole Centrale de Lyon Bâtiment H9 36 avenue Guy de Collongue 69134 ECULLY Tél : 04.72.18 60 97 Fax : 04 78 43 37 17 eea@ec-lyon.fr Secrétariat : M.C. HAVGOUDOUKIAN
E2M2	<u>EVOLUTION, ECOSYSTEME, MICROBIOLOGIE, MODELISATION</u> http://biomserv.univ-lyon1.fr/E2M2 M. Jean-Pierre FLANDROIS Insa : H. CHARLES	M. Jean-Pierre FLANDROIS CNRS UMR 5558 Université Claude Bernard Lyon 1 Bât G. Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.26 23 59 50 Fax 04 26 23 59 49 06 07 53 89 13 e2m2@biomserv.univ-lyon1.fr
DISS	<u>INTERDISCIPLINAIRE SCIENCES-SANTE</u> Sec : Safia Boudjema M. Didier REVEL Insa : M. LAGARDE	M. Didier REVEL Hôpital Cardiologique de Lyon Bâtiment Central 28 Avenue Doyen Lépine 69500 BRON Tél : 04.72.68 49 09 Fax : 04 72 35 49 16 Didier.revel@creatis.uni-lyon1.fr
INFOMATHS	<u>INFORMATIQUE ET MATHEMATIQUES</u> http://infomaths.univ-lyon1.fr M. Alain MILLE Secrétariat : C. DAYEYAN	M. Alain MILLE Université Claude Bernard Lyon 1 LIRIS - INFOMATHS Bâtiment Nautibus 43 bd du 11 novembre 1918 69622 VILLEURBANNE Cedex Tél : 04.72. 44 82 94 Fax 04 72 43 13 10 infomaths@bat710.univ-lyon1.fr - alain.mille@liris.cnrs.fr
MATERIAUX	<u>MATERIAUX DE LYON</u> M. Jean Marc PELLETIER Secrétariat : C. BERNAVON	M. Jean Marc PELLETIER INSA de Lyon MATEIS Bâtiment Blaise Pascal 7 avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.43 83 18 Fax 04 72 43 85 28 Jean-marc.Pelletier@insa-lyon.fr
MEGA	<u>MECANIQUE, ENERGETIQUE, GENIE CIVIL, ACOUSTIQUE</u> M. Jean Louis GUYADER Secrétariat : M. LABOUNE	M. Jean Louis GUYADER INSA de Lyon Laboratoire de Vibrations et Acoustique Bâtiment Antoine de Saint Exupéry 25 bis avenue Jean Capelle 69621 VILLEURBANNE Cedex Tél : 04.72.18.71.70 Fax : 04 72 43 72 37 mega@lva.insa-lyon.fr
ScSo	<u>ScSo*</u> M. Lionel OBADIA Insa : J.Y. TOUSSAINT	M. OBADIA Lionel Université Lyon 2 86 rue Pasteur 69365 LYON Cedex 07 Tél : 04.78.69.72.76 Fax : 04.37.28.04.48 Lionel.Obadia@univ-lyon2.fr

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

THESE SOUTENUE DEVANT L'INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON

NOM : TOMEI épouse LE DIGARCHER

DATE de SOUTENANCE : 10 décembre 2009

Prénoms : Sandrine Véronique Colette

TITRE : Détection automatique de fixations radioactives dans les images TEP oncologiques

NATURE : Doctorat

Numéro d'ordre :2009 ISAL-0119

Ecole doctorale : Electronique, Electrotechnique, Automatique

Spécialité :

Sciences de l'Information, des Dispositifs et des Systèmes

Cote B.I.U. - Lyon : T 50/210/19 / et bis

CLASSE :

RESUME :

Le suivi thérapeutique des patients cancéreux est actuellement un axe de recherche primordial. L'imagerie de Tomographie par Emission de Positons (TEP) permet de visualiser l'information fonctionnelle des cellules et notamment leur consommation en FDG, un analogue du glucose, marqué au fluor 18 (18F), dont les cellules tumorales sont particulièrement gourmandes. La TEP s'apprête à devenir un examen de référence dans le suivi pathologique de patients atteints de certains cancers comme le lymphome. Ses performances en termes de détection peuvent cependant être encore améliorées, notamment à travers le développement de méthodes de quantification, d'instrumentation ou de traitement d'images. Ainsi, cette thèse repose sur le développement d'un outil d'aide à la détection (Computer Aided Detection, CAD en anglais) de tumeurs difficilement visibles pour les images TEP 3D corps entier.

Pour ce faire, nous avons suivi la méthodologie standardisée des systèmes de classification supervisée. Nous avons proposé une approche basée sur la combinaison de deux classifieurs, l'analyse discriminante linéaire (LDA en anglais) et les séparateurs à vaste marge (SVM), associés à des caractéristiques descriptives de l'image de type ondelettes. Chaque classifieur fournit une carte 3D de score quantifiant la probabilité des voxels de correspondre à une tumeur. Ces cartes sont ensuite seuillées à l'aide d'une méthode originale basée sur une analyse par courbe FROC pour produire des cartes binaires de détection. Une étape de réduction de faux positifs (RFP) permet de nettoyer ces cartes binaires avant les combiner. Nous avons proposé une stratégie d'évaluation 3D du système, basée sur l'utilisation d'images simulées garantissant la connaissance des caractéristiques des tumeurs cibles. Deux jeux de données de réalisme croissant ont servi à l'évaluation quantitative et qualitative du système. La première base d'images simulées, développée avant la thèse, repose sur le fantôme mathématique MCAT. La seconde base d'images a été développée dans le cadre de cette thèse et présente cent volumes TEP corps entier obtenus par simulation Monte Carlo à partir du fantôme anthropomorphique de Zubal. Elle contient 50 cas sains et 50 cas pathologiques présentant un ensemble de 375 tumeurs sphériques de diamètre et contraste calibrés. Le réalisme de ces images a été évalué à l'aide d'index statistiques et cliniques calculées sur les images simulées et cliniques.

Les résultats obtenus à partir des volumes binaires provenant du système de détection développé se montrent prometteurs. Ils ouvrent la perspective d'enrichir l'information binaire généralement fournie au clinicien par des indices paramétrique quantifiant la pertinence de chaque tumeur détectée.

MOTS-CLES :

Tomographie par émission de positons, Computer Aided Detection, Analyse Discriminante Linéaire, Séparateurs à vaste marge, ondelettes, Simulation

Laboratoire (s) de recherche : CREATIS-LRMN, INSA Lyon

Directeur de thèse:

Christophe Odet, Pr. 61, INSA Lyon, Creatis Lyon, UMR CNRS 5220, U 630 Inserm

Co-directrice de thèse :

Carole Lartizien, CR-CNRS, INSA Lyon, Creatis Lyon, UMR CNRS 5220, U 630 Inserm

Président de jury :

- Rémy Prost, Pr. 61, INSA Lyon, Creatis Lyon, UMR CNRS 5220, U 630 Inserm

Composition du jury :

- Stéphane Canu, Pr. 61, INSA de Rouen, LITIS

- Visvikis Dimitris, DR-INSERM, LATIM INSERM U650

- Francesco Giammarile, PU-PH, Centre Hospitalier Lyon Sud