



HAL
open science

Détection et caractérisation du cancer de la prostate par images IRM 1.5T multiparamétriques

Jérôme Lehaire

► **To cite this version:**

Jérôme Lehaire. Détection et caractérisation du cancer de la prostate par images IRM 1.5T multiparamétriques. Imagerie. Université de Lyon, 2016. Français. NNT : 2016LYSE1174 . tel-01616054

HAL Id: tel-01616054

<https://theses.hal.science/tel-01616054>

Submitted on 13 Oct 2017

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

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



N° d'ordre NNT : 2016LYSE1174

THÈSE DE DOCTORAT DE L'UNIVERSITÉ DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

École Doctorale ED01 : Interdisciplinaire Sciences-Santé (EDISS)

Spécialité de doctorat : traitement d'images

Soutenue publiquement le 04/10/2016, par :

Jérôme Lehaire

**Détection et caractérisation du
cancer de la prostate par images
IRM 1.5T multiparamétriques**

Devant le jury composé de :

Nacim BETROUNI, Chargé de recherche HDR, ONCO-THAI

Bruno GAS, Professeur des universités HDR, ISIR

Paul Michael WALKER, Maître de conférence HDR, LE2I

Olivier BASSET, Professeur des universités HDR, CREATIS

Florence MEGE-LECHEVALLIER, Praticien Hospitalier, CHU Lyon

Rémi FLAMARY, Maître de conférence, Laboratoire Lagrange

Rapporteur

Rapporteur

Rapporteur

Membre du Jury

Membre du Jury

Membre du Jury

Olivier ROUVIÈRE, Professeur-Praticien Hospitalier, LabTAU

Carole LARTIZIEN, Chargée de recherche HDR, CREATIS

Directeur de thèse

Co-directrice de thèse

Résumé

Le cancer de la prostate est actuellement le plus courant en France et la 4^{ème} cause de mortalité par cancer. La méthode diagnostic de référence est la biopsie écho-guidée, effectuée à la suite de deux tests, le PSA et le toucher rectal, souvent insuffisants pour détecter et localiser précisément une lésion. De plus, en raison de la mauvaise qualité de l'imagerie ultrasonore, le cancer est souvent peu évident à discerner et les biopsies ne sont généralement qu'un échantillonnage effectué aléatoirement. Ainsi, elles peuvent manquer leur cible et n'évaluent pas précisément le volume et l'agressivité des cancers détectés.

L'imagerie IRM multi-paramétrique s'impose désormais comme la technique la plus prometteuse pour le diagnostic et la prise en charge du cancer de la prostate. En effet, de nombreuses études montrent que la fusion de différentes séquences permet d'améliorer considérablement les diagnostics de détection, localisation ainsi que l'évaluation de l'agressivité des tumeurs. Ainsi, leur utilisation généralisée en routine clinique permettrait de détecter précocement les cancers, guider les biopsies de manière précise et enfin rendre possible les traitements focaux, alternative des prostatectomies radicales. Néanmoins, l'interprétation visuelle des multiples séquences IRM n'est pas aisée et l'on constate une forte variabilité inter-lecteur dans le diagnostic notamment lorsque les images se contredisent. Dans ces conditions, un fort intérêt s'est porté sur les systèmes d'aide au diagnostic dont le but est d'assister le radiologue dans ses décisions.

Cette thèse présente nos travaux sur la conception d'un Système d'aide à la détection (CADE) dont l'approche finale est de fournir au radiologue une carte de probabilité du cancer dans la zone périphérique de la prostate. Ce CADe repose sur une base d'images IRM multi-paramétrique (IRM-mp) 1.5T de types T2w, dynamique et de diffusion provenant de 49 patients. Une des forces de cette base repose sur une vérité terrain précise issue d'une référence histologique après prostatectomie radicale. Un ensemble de caractéristiques statistiques, structurelles et fonctionnelles est extrait de ces séquences puis utilisées lors d'un processus d'apprentissage supervisé et de classification permettant d'évaluer les performances du CADe. Cette thèse met l'accent sur la détection des cancers mais aussi sur leur caractérisation dans le but de fournir une carte de probabilité corrélée au grade de Gleason des tumeurs. Dans cette optique, nous avons utilisé une méthode d'apprentissage de dictionnaires permettant d'extraire de nouvelles caractéristiques descriptives dont l'objectif est de discriminer chacun des cancers. L'ensemble des caractéristiques est ensuite utilisé en entrée de deux classifieurs, régression logistique et séparateur à vaste marge (SVM), permettant de produire une carte de probabilité du cancer. Nous avons concentré nos efforts sur la discrimination des cancers agressifs (grade de Gleason >6) et fourni une analyse de la corrélation entre probabilités et scores de Gleason. Les résultats montrent de très bonnes performances de détection des cancers agressifs et l'analyse des probabilités conclue sur une forte capacité du système à séparer les cancers agressifs du reste des tissus mais ne permet pas aisément de distinguer chacun des grades de cancer.

La dernière partie présente la constitution d'une nouvelle base d'images IRM-mp 3T ainsi que l'évaluation des performances de recalage d'un logiciel dont le but est de corriger l'erreur d'alignement des séquences extraites ADC, B2000 et DCE sur la séquence de référence T2w. Cette nouvelle base, constituée en parallèle de la thèse, permettra à l'avenir de disposer d'images de meilleure résolution et recalées pouvant potentiellement améliorer les performances de classification dans le cadre d'une approche visant à fournir une carte de probabilité voxel à voxel. L'étude montre que le recalage améliore globalement l'alignement spatiale des séquences IRM et réduit ainsi l'erreur initiale.

Table des matières

Résumé	ii
Sommaire	v
1 Introduction générale	9
I Contexte médical et scientifique	11
2 Le cancer de la prostate	13
2.1 Introduction	13
2.2 Épidémiologie	13
2.3 Anatomie de la prostate	13
2.4 Diagnostic clinique du cancer	15
2.4.1 Classification des cancers de la prostate	15
2.4.2 Examens diagnostiques préliminaires	16
2.4.3 Bilan d'extension	16
2.5 Traitement du cancer	17
2.6 Problématique du diagnostic du cancer de la prostate	18
2.6.1 Le diagnostic	18
2.6.2 Synthèse des faiblesses des techniques actuelles	19
2.6.3 Rôle de l'imagerie dans la cartographie locale du cancer de prostate	19
2.6.4 Une technique d'imagerie très prometteuse : l'IRM	20
2.7 Conclusion	20
3 L'imagerie par résonance magnétique pour le cancer de la prostate	23
3.1 Introduction	23
3.2 Principe général d'acquisition des séquences IRM	23
3.2.1 Image IRM en pondération T2	23
3.2.2 Séquence IRM pondérée en diffusion	27
3.2.3 Image IRM de perfusion DCE	30
3.3 Diagnostic du cancer de la prostate par imagerie IRM multiparamétrique	34
3.4 Conclusion	36
4 Les systèmes d'aide au diagnostic pour l'imagerie du cancer de la prostate	37
4.1 Introduction	37
4.2 Approches par région (CADx) Versus approches par voxel (CAdE)	38
4.3 Méthodologie de développement d'un CAD	39
4.3.1 Recalage	39
4.3.2 Pré-traitement des images	40
4.3.3 Extraction de caractéristiques	41

4.3.4	Réduction de dimension	41
4.3.5	Classification	43
4.3.6	Evaluation des performances en classification	45
4.3.7	Validation d'un modèle d'apprentissage	47
4.4	Etat de l'art des CAD pour le cancer de la prostate	49
4.4.1	Etat de l'art en détection, localisation et évaluation de l'agressivité du cancer de la prostate	49
4.4.2	Tableau de synthèse des CAD	52
4.5	Discussion sur les performances des CAD	54
II Développement d'un CADe pour le diagnostic du cancer de la prostate par analyse d'images IRM multiparamétriques		55
5	Constitution de la base d'images clinique IRM 1.5T	57
5.1	Introduction	57
5.2	Acquisition des images	58
5.2.1	Déroulement de l'examen	58
5.2.2	Paramètres des séquences IRM 1.5T	58
5.3	Analyse des images	59
5.4	Analyse histologique	59
5.5	Corrélation anatomo-radiologique	60
5.6	Constitution de la base d'entraînement du CADe	62
6	Choix méthodologiques pour la réalisation du CADe	65
6.1	Introduction	65
6.2	Acquisition des images	65
6.3	Extraction de caractéristiques	66
6.3.1	Caractéristiques descriptives	67
6.3.2	Caractéristiques fonctionnelles	70
6.4	Sélection de caractéristiques	71
6.5	Caractérisation des voxels	72
6.5.1	Méthodes de caractérisation en imagerie hyperspectrale	73
6.5.2	Apprentissage de dictionnaires	76
6.6	Classification	79
6.6.1	SVM linéaire	79
6.6.2	SVM non linéaire	82
6.6.3	Régression logistique	84
6.6.4	Classification sur base d'apprentissage déséquilibrée	86
6.7	Méthode d'évaluation	88
7	Evaluation des performances des différents schémas de CAD	91
7.1	Introduction	91
7.2	Expérimentations	91
7.3	Caractéristiques sélectionnées	93
7.4	Optimisation des paramètres des classifieurs	94
7.4.1	Paramètres optimaux obtenus sur base d'apprentissage déséquilibrée	94
7.4.2	Paramètres optimaux obtenus sur base d'apprentissage équilibrée	95
7.5	Analyse des performances des différentes combinaisons de CAD	96
7.6	Résultats et comparaison des performances	97
7.6.1	Classification sur base d'apprentissage déséquilibrée	97
7.6.2	Classification sur base d'apprentissage équilibrée	101

7.6.3	Corrélation entre probabilités et score de Gleason	106
7.7	Conclusion	107
III	Recalage d'images et classification	111
8	Création et recalage d'une base d'images IRM 3T	113
8.1	Introduction	113
8.2	Problématique du recalage en classification par approche voxel	114
8.3	Description de la base 3T	115
8.4	Méthodes de recalage	116
8.4.1	Approche générale	116
8.4.2	Construction d'un modèle générique de recalage	116
8.4.3	Modèles de transformation	118
8.5	Choix du modèle	119
8.6	Description de la méthode	120
9	Évaluation des performances de recalage	123
9.1	Base d'évaluation et protocole	123
9.2	Description des métriques	124
9.3	Résultats	124
9.3.1	Résultat global moyenné	125
9.3.2	Matrice image des résultats	126
9.3.3	Histogrammes des écarts de recalage	128
9.3.4	Représentation ponctuelle des écarts de recalage en fonction du décalage initial	129
9.4	Conclusion	132
10	Conclusions et perspectives	133
	Annexes	139
A	Résultats complémentaires sur le CADe	139
A.1	Liste des caractéristiques sélectionnées	139
	Bibliographie	150

Liste des abréviations

AC	Agent de contraste
Acc	Précision
ACP	Analyse en composantes principales
ADC	Coefficient apparent de diffusion
AIF	Fonction d'entrée artérielle
ADL	Analyse discriminante linéaire
RNN	Réseau de neurones artificiels
TA	Temps d'arrivée
AUC	Aire sous la courbe ROC
AUGC	Aire sous la courbe de réhaussement en produit de contraste
CAD	Systèmes d'aide à la décision
CADe	Système d'aide à la détection
CADx	Système d'aide au diagnostic
CLARA-P	Corrélations anatomo-radiologiques en IRM de prostate
CNB	Classifieur naïf de Bayes
CRF	Champs aléatoires conditionnels
CT	Tomodensitométrie
DCE	Réhaussement de contraste dynamique
DCT	Transformée en cosinus discret
DP	Densité de protons
DSC	Indice de Dice
DWI	Imagerie pondérée en diffusion
EEE	Espace extravasculaire/extracellulaire
FMRF	Champs de Markov flous
FN	Faux négatifs
FP	Faux positifs
FOV	champ de vue
Gd-DOTA	Gadotérate méglumine
GE	Graphe embarqué
GLCM	Matrice de co-occurrence des niveaux de gris
GL	Score de Gleason
Hz	Hertz
HES	Hématoxyline-Eosine-Safran
HIFU	Ultrasons focalisés de haute intensité
IM	Information mutuelle
IRM	Imagerie par résonance magnétique
IRM-mp	IRM multi-paramétrique
Kep	Constante de flux entre l'EEE et le compartiment vasculaire
k-PPV	k -plus-proches voisins
Ktrans	Constante de transfert entre le compartiment vasculaire et l'EEE
LDA	Analyse discriminante linéaire

LLE Localement linéaire embarqué
LOO Validation croisée par exclusion d'un élément
LOPO Validation croisée par exclusion d'un patient
M Malin
MCE Entropie croisée moyenne
mg milligramme
min minute
mm millimètre
MRF Champs de Markov flou
MRS Spectroscopie par résonance magnétique
mRMR minimum de redondance, maximum de pertinence
ms milliseconde
MSE Erreur quadratique moyenne
N Normal
Nb Nombre
NFS Numération-Formule-Sanguine
NS non malin mais suspect
PE Pic de réhaussement
PIN Néoplasie intraépithéliale de la prostate
PSA Antigènes spécifiques à la prostate
P-SVM SVM probabiliste
QP Problème quadratique
RBF Fonction à base radiale
RDF Forêt aléatoire de décision
RF Radiofréquence
RL Régression logistique
RMN Résonance magnétique nucléaire
ROC Caractéristique de fonctionnement du récepteur
ROI Région d'intérêt
RVM Machine à vecteur pertinent
s seconde
SE Sensibilité
SFMA Stroma fibro-musculaire
SI Signal
SNR Rapport signal sur bruit
SP Spécificité
SVM Séparateur à vaste marge
SVM-L Séparateur à vaste marge linéaire
SVM-NL Séparateur à vaste marge non-linéaire
T Tesla
TE Temps d'écho
TEMP Tomographie d'émission monophotonique
TEP Tomographie par émission de positron
TR Toucher rectal
TTP Temps d'arrivée au pic
T2-map Cartographie T2
T2-w T2-pondérée
T1-w T1-pondérée
US Ultrason
Ve Volume de l'EEE
VN Vrais négatifs
VOI Volume d'intérêt

VP Vrais positifs
VPN Valeur prédictive négative
VPP Valeur prédictive positive
WI Vitesse de réhaussement
WO Vitesse de lavage
ZC Zone centrale
ZP Zone périphérique
ZT Zone transitionnelle
2-D 2 dimensions
3-D 3 dimensions

Table des figures

2.1	Anatomie de la prostate	14
2.2	Les 3 zones principales de la prostate	15
3.1	Effet des champs magnétiques B_0 et B_1 sur le moment magnétique	24
3.2	Illustration de la relaxation transverse T2 et T2*	26
3.3	Illustration de la séquence multi-échos spin écho	27
3.4	Exemple de séquence T2w : analyse qualitative	28
3.5	Schéma d'acquisition d'une séquence de diffusion	29
3.6	Exemple d'images IRM de diffusion et carte ADC : analyse qualitative	31
3.7	Exemple de séquence IRM DCE et analyse qualitative	32
3.8	Schéma d'une courbe de réhaussement issue d'une séquence IRM dynamique avec injection d'un produit de contraste	33
3.9	Schéma des échanges de l'agent de contraste avec les compartiments avoisinant	34
4.1	Schéma de principe des CADe et Système d'aide au diagnostic (CADx)	39
4.2	Schéma général d'un système de classification	40
4.3	Principe de construction d'une courbe ROC	46
5.1	Préparation de la pièce de prostatectomie	61
5.2	Corrélation anatomo-radiologique	62
6.1	Pipeline du CADe implémenté	66
6.2	Exemples de cartes de caractéristiques de texture	68
6.3	Paramètres fonctionnels semi-quantitatifs	71
6.4	Représentation schématique d'une image hyperspectrale	74
6.5	Exemple de processus de mélange linéaire pour l'imagerie hyperspectrale	74
6.6	Principe du VCA	76
6.7	Hyperplans séparateurs et marges maximales	79
6.8	Fonctions logistiques dans un cas de classification à 2 classes	86
7.1	Représentation des 8 schémas de CAD évalués à partir de la base d'apprentissage déséquilibrée	93
7.2	Tracé du poids des caractéristiques	94

7.3	Exemples de cartes de probabilités générées à partir de la base d'apprentissage déséquilibrée	99
7.4	Exemples de cartes de probabilités générées à partir de la base d'apprentissage déséquilibrée	99
7.5	Exemples de cartes de probabilités générées à partir de la base d'apprentissage équilibrée et SVM linéaire	102
7.6	Exemples de cartes de probabilités générées à partir de la base d'apprentissage équilibrée et SVM non linéaire	103
7.7	Cartes de probabilités sur 3 coupes consécutives	105
7.8	Boxplots des probabilités de cancer en fonction de la classe	108
8.1	Illustration de la nécessité du recalage	115
8.2	Illustration des séquences IRM 3T	117
8.3	Principe du recalage	117
8.4	Les différents modèles de transformations	119
8.5	Description de la méthode par maillage utilisée	120

Liste des tableaux

4.1	Comparaison des études sur les CAD du cancer de la prostate	53
5.1	Paramètres utilisés pour l'imagerie IRM de la prostate à 1.5 T	59
5.2	Détails des catégories de voxels pour la classification voxel à voxel	63
6.1	Tableau des 4 paramètres statistiques du 1er ordre extraits	67
6.2	Caractéristiques issues de la matrice GLCM	69
6.3	Noyaux des gradients de Sobel et de Kirsch	70
6.4	Paramètres de perfusion	71
7.1	Paramètres optimaux des schémas de CAD pour la base d'apprentissage déséquilibrée	95
7.2	Paramètres optimaux des schémas de CAD pour la base d'apprentissage équilibrée	96
7.3	Score de probabilité moyen par ROI 2D (issu d'une unique coupe) des 3 patients illustrés sur les figures 7.3 et 7.4. Valeurs correspondantes à la base déséquilibrée.	98
7.4	Performances de classification sur base d'apprentissage déséquilibrée	100
7.5	Tableau de résultats du test de signe entre chaque combinaison de CAD associés au SVM linéaire et régression logistique sur base déséquilibrée. Les valeurs présentées sont les p-valeurs et les celles en gras sont significatives pour un seuil de 0.05.	100
7.6	Score de probabilité moyen par classe des ROIs 3D sur base d'apprentissage déséquilibrée	101
7.7	Score de probabilité moyen par ROI 2D (issu d'une unique coupe) des 3 patients illustrés sur les figures 7.5 et 7.6. Valeurs correspondantes à la base équilibrée.	103
7.8	Performances de classification sur base d'apprentissage équilibrée	104
7.9	Tableau de résultats du test de signe entre chaque combinaison de CAD associés au SVM non linéaire sur base équilibrée. Les valeurs présentées sont les p-valeurs et les celles en gras sont significatives pour un seuil de 0.05.104	

7.10	Score de probabilité moyen par classe des ROIs 3D sur base d'apprentissage équilibrée	106
8.1	Paramètres utilisés pour l'imagerie IRM de la prostate à 3T	116
9.1	Tableau de résultats de distances, moyenne \pm écart-type, avant et après recalage	125
9.2	Tableau de résultats avant/après recalage en fonction de seuils d'acceptation	128
A.1	Liste des caractéristiques sélectionnées par classifieur	140
A.2	Liste des caractéristiques sélectionnées par classifieur (suite)	141

Introduction générale

Détecter, localiser et caractériser le cancer prostate : un enjeu majeur

En 2012 et dans le monde, le cancer de la prostate est le quatrième plus fréquent et le deuxième chez l'homme dont 1,1 million en ont été diagnostiqués, représentant 15% de la totalité des cancers recensés [40]. On recense cette même année environ 300 000 décès par cancer de la prostate, ce qui en fait la 4^{ième} cause de décès par cancer chez l'homme. Bien que le taux de mortalité soit en diminution de par la multiplication et la précocité des dépistages, les techniques diagnostiques actuelles : toucher rectal, taux de PSA et biopsies souffrent d'un manque de précision. En effet, ces dernières sont la technique de référence pour confirmer ou infirmer la présence du cancer mais représentent des échantillons de prostate, souvent obtenus aléatoirement par écho-guidage, et peuvent ainsi manquer leur cible. De plus elles ne permettent pas d'évaluer précisément le volume de la tumeur.

Promesses de l'IRM multiparamétrique pour le diagnostic du cancer de la prostate

L'imagerie IRM multi-paramétrique (IRM-mp) s'impose depuis peu comme une référence pour la détection, localisation et l'évaluation de l'agressivité des tumeurs. C'est un axe de recherche prometteur pour diagnostiquer et ainsi adapter rapidement un suivi personnalisé des patients. De plus, ses capacités de localisation ouvrent la voie à d'autres techniques de traitement dites focalisées, permettant de traiter localement une tumeur et ne plus systématiquement recourir à l'ablation totale de la prostate.

Problématique de la thèse

L'utilisation large échelle de l'IRM-mp nécessite néanmoins une standardisation à la fois dans la méthode d'acquisition des séquences IRM mais aussi dans la façon d'interpréter les images. En effet, cette dernière reste une tâche complexe et effectuée de manière qualitative par les radiologues. Dans cette optique, les Systèmes d'aide à la décision (CAD) ont connu un essor important dans ce besoin de standardisation et ont pour objectif d'assister le radiologue dans sa prise de décision. Basés sur l'extraction de caractéristiques propres à chacune de ces séquences IRM, les CAD ont pour objectif de détecter, localiser et estimer l'agressivité des lésions. Les CAD sont principalement de deux types : les CADx et les CADe dont les objectifs sont respectivement de fournir un score de probabilité de cancer dans une zone contourée par le radiologue et de fournir une carte de probabilité voxel à voxel de la prostate. La plupart des études et travaux sur les CAD se sont concentrés sur la tâche de détection du cancer de la prostate. Dans cette thèse nous proposons une implémentation d'un CADe dont le but est la détection des cancers agressifs. L'idée clinique est de surveiller plus attentivement les patients atteints de cancers susceptibles d'évoluer rapidement à court terme. Dans le but d'évaluer l'agressivité du cancer, nous proposons une méthode de caractérisation basée sur des dictionnaires appris sur les données et dont l'objectif est de détecter les signatures caractéristiques des constituants de la prostate dont les différents grades de tumeurs. Nous proposons enfin une analyse détaillée des scores de probabilité générés et notamment leur capacité à discriminer les différents grades.

Plan de la thèse

La première partie présente la problématique du cancer de la prostate ainsi que les forces et faiblesses des différentes méthodes diagnostiques. Nous présentons ensuite le principe général de l'IRM, des différentes séquences utilisées dans la suite de nos travaux et l'apport de l'IRM-mp pour le diagnostic. Nous terminons cette partie par un état de l'art des CAD. La seconde partie présente dans un premier temps la base d'images IRM 1.5T utilisée pour réaliser le CADe. Nous détaillons ensuite les différentes étapes méthodologiques de sa réalisation et terminons par une analyse détaillée des résultats. La troisième partie présente nos travaux effectués en parallèle pour la constitution et le recalage d'une nouvelle base d'images IRM 3T. Nous présentons notamment l'apport potentiel de données recalées sur les performances d'un CADe ainsi qu'une étude des performances d'un logiciel de recalage que nous avons choisi et utilisé pour recaler nos données IRM. Nous terminons enfin par un bilan de ces travaux de thèse et des perspectives futures.

I Contexte médical et scientifique

Le cancer de la prostate

2.1 Introduction

Ce chapitre décrit dans un premier temps l'anatomie de la prostate, sa fonction au sein du corps humain, les différentes méthodes diagnostique du cancer ainsi que les possibles traitements. Il est ensuite évoqué leurs limitations avant d'introduire les promesses de l'imagerie IRM pour la prise en charge et le suivi des patients.

2.2 Épidémiologie

En France, en 2011, le cancer de la prostate est la 4^{ieme} cause de mortalité par cancer avec environ 10 000 décès par an et au premier rang des cancers avec environ 70 000 nouveaux cas par an [7, 31, 98]. Son incidence est en hausse constante de par le vieillissement de la population mais aussi grâce à l'amélioration des techniques de dépistage. De plus, le taux de mortalité est en diminution, de l'ordre de 2.5% par an en moyenne, en raison de l'amélioration des traitements. Ainsi, l'âge moyen du diagnostic est de 70 ans, le taux de mortalité et d'incidence sont respectivement de 10 et 100 pour 100 000 hommes. Le facteur de risque principal reste les antécédents familiaux mais aussi certaines caractéristiques ethnogéographiques comme pour les Afro-Antillais par exemple.

2.3 Anatomie de la prostate

La prostate est une glande faisant partie de l'appareil reproducteur et urinaire de l'homme. Elle a la forme d'une châtaigne d'environ 3 centimètres de hauteur et 4 centimètres de

largeur, bien que sa taille soit variable d'un homme à l'autre. Sa fonction principale est la production du liquide prostatique qui vient se mélanger avec les spermatozoïdes en provenance des testicules au moment de l'éjaculation. Elle est située dans le bassin et sous la vessie, entre le pubis en avant et le rectum en arrière, et entoure une partie de l'urètre (cf. figure 2.1). Autour de l'urètre, un ensemble de fibres musculaires est amassé au niveau du col vésical et sous la prostate pour former le sphincter urinaire. Ce dernier, en se contractant ou relâchant permet de contrôler le débit d'urine. La prostate est enveloppée d'une mince couche de tissu fibreux appelée capsule.

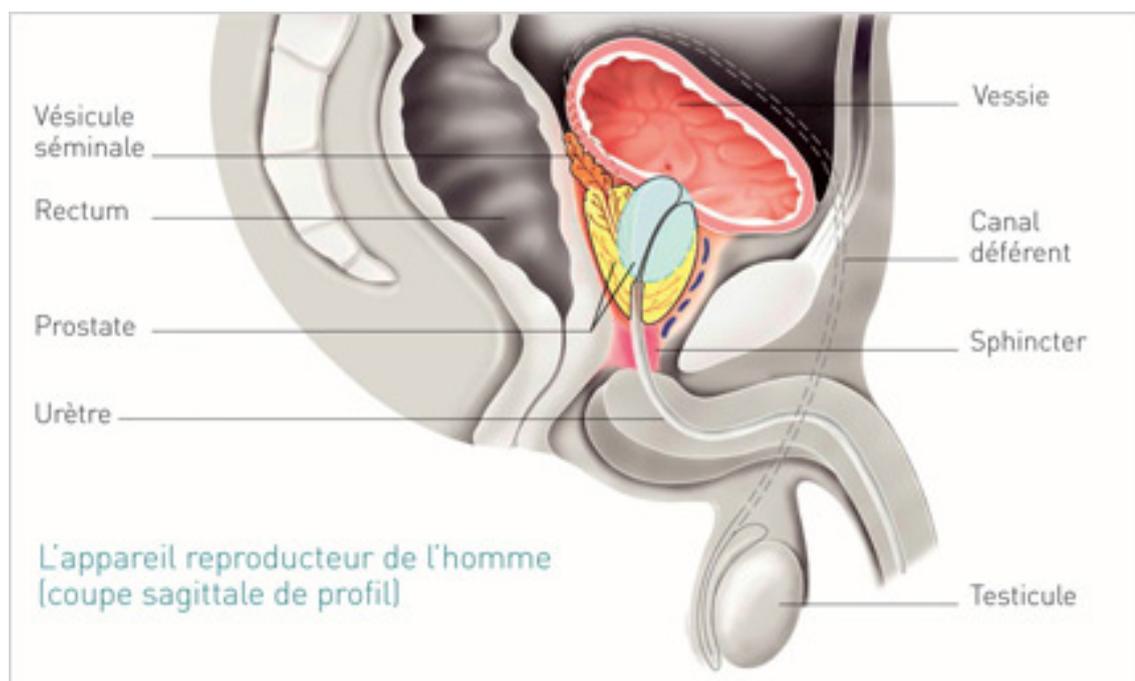


FIGURE 2.1 – Anatomie de la prostate et structures adjacentes, source [Institut national du cancer](#)

La prostate se décompose en 3 zones principales :

- La Zone périphérique (ZP) : la plus proche du rectum, elle est accessible par palpation lors d'un toucher rectal (cf. 2.4.2). C'est dans cette zone que se développent 75% des cancers ;
- la Zone transitionnelle (ZT) : située au milieu de la prostate, entourant l'urètre, sa taille représente environ 5% de la prostate jusqu'à 40 ans puis devient la zone la plus volumineuse ensuite (hypertrophie bénigne). 20% des cancers sont issus de cette zone ;
- la Zone centrale (ZC) : partie entourant les canaux éjaculateurs, elle représente environ 20% du volume de la prostate et environ 5% des cancers sont issus de cette zone. On y trouve les vésicules séminales, deux glandes débouchant dans la prostate pour se relier aux canaux déférents et devenir le canal éjaculateur.

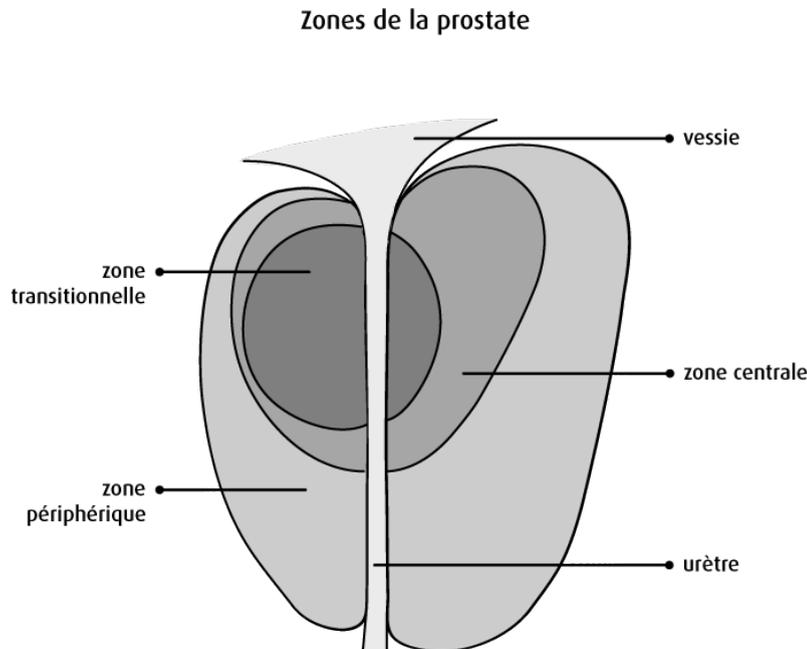


FIGURE 2.2 – Les 3 zones principales de la prostate : ZP, ZT et ZC, source [Institut national du cancer](#)

2.4 Diagnostic clinique du cancer

2.4.1 Classification des cancers de la prostate

Degré d'extension du cancer. Les cancers de la prostate sont traités différemment selon plusieurs critères : la taille de la tumeur primitive (T), la présence ou non de cellules cancéreuses dans les ganglions lymphatiques (N) et la présence ou non de métastases à distance dans d'autres parties du corps (M). Ces trois critères définissent le stade d'avancement du cancer, appelé le degré d'extension. Cette classification selon ces 3 critères est appelée stade TNM.

Score de Gleason. Le score de Gleason, résultat le plus significatif pour conclure sur un traitement [12, 38] est une classification histologique dont le principe repose sur un prélèvement de prostate analysé par un anatomopathologiste. Son score est ensuite attribué en fonction de l'évolutivité et de l'aspect du tissu tumoral, comparé à des tissus prostatiques sains. Ainsi, le score de Gleason évalue l'agressivité d'une tumeur ainsi que sa probabilité de propagation. A noter que cette métrique n'est utilisée que pour les adénocarcinomes, type de tumeur le plus fréquent dans la prostate (95%). L'échelle du grade de Gleason varie entre 1 et 5, assigné en fonction de la différenciation cellulaire. Plus le grade est élevé et plus les cellules semblent anormales. Étant donné qu'une tumeur de grade 1 ou 2 ressemble et se comporte comme des tissus normaux, la plupart des tumeurs de la prostate sont de grade 3 à 5. Le score final de Gleason est obtenu après identification des deux modèles de croissance glandulaire les plus courants dans la tumeur, auxquels on assigne un grade de

1 à 5. Ainsi le score théorique final est la somme de ces deux grades et varie entre 2 et 10. En pratique, ce score est la plupart du temps compris entre 6 et 10. Les tumeurs au dessus du grade 6 sont considérées comme des tumeurs agressives.

Classification de d'Amico. La classification de d'Amico permet la catégorisation du cancer de la prostate en fonction de son risque évolutif à l'aide des critères PSA, score de Gleason et échelle TNM (cf. section 2.4.2). Les 3 catégories sont le cancer localisé à :

- Faible risque : PSA < 10 ng/ml et score de Gleason ≤ 6 et stade clinique T1c ou T2a,
- Risque intermédiaire : PSA compris entre 10 et 20 ng/mL ou score de Gleason de 7 ou stade T2b,
- Risque élevé : PSA > 20 ng/mL ou score de Gleason ≥ 8 ou stade clinique T2c.

2.4.2 Examens diagnostiques préliminaires

Le diagnostic positif de cancer de prostate est établi grâce à plusieurs éléments :

- Tout d'abord l'**interrogatoire initial** permet d'apprécier l'état de santé général et les antécédents familiaux,
- Ensuite l'**examen clinique**, comprenant un Toucher rectal (TR), permet de détecter une éventuelle anomalie dans la zone périphérique,
- un **bilan biologique** apportera des arguments directs par dosage d'antigènes de la prostate (PSA) mais également indirects par l'évaluation de l'état inflammatoire global (NFS),
- et enfin une **biopsie** (prélèvement de tissus prostatiques) affirme le diagnostic de cancer de prostate.

2.4.3 Bilan d'extension

Une fois la présence de cellules cancéreuses confirmée par la biopsie tissulaire, l'enjeu est de déterminer le niveau d'envahissement tumoral par un bilan d'extension complet. Ce bilan d'extension doit permettre, grâce à des examens complémentaires adaptés, de préciser l'extension locale, régionale et à distance (métastatique) du cancer.

On cherche ici dans un premier à évaluer la taille de la tumeur par imagerie (extension locale, "T"). Ensuite, l'IRM cherche à évaluer le bilan d'extension locale (extension régionale "N" : franchissement de la capsule ou atteinte des vésicules séminales). L'atteinte ganglionnaire est cherchée par IRM ou scanner et permet d'évaluer l'extension à distance et donc la présence ou non de métastases ("M").

Ce bilan d'extension permet donc la stadification TNM expliquée précédemment, de laquelle découle la décision thérapeutique.

2.5 Traitement du cancer

Plusieurs solutions peuvent être envisagées pour le traitement du cancer :

- **La chirurgie** dont le but est l'ablation de la prostate et des vésicules séminales lors d'une opération appelée prostatectomie radicale. Elle est généralement pratiquée lorsque le bilan d'extension conclut à une forme de tumeur localisée à risque faible ou intermédiaire. Elle est le premier traitement curatif du cancer de prostate.
- **la radiothérapie externe** a pour objectif de détruire les cellules cancéreuses localisées au niveau de la prostate au moyen de rayons, à très haute énergie, focalisés sur la prostate à irradier. Elle est souvent utilisée pour le cas d'un cancer localisé à risque faible ou intermédiaire ainsi qu'en association à une hormonothérapie pour des cancers à risque élevé. Elle constitue aussi un traitement curatif du cancer de prostate
- **la curiethérapie** consiste à mettre en place des implants radioactifs à l'intérieur de la prostate. Ils émettent un rayonnement qui détruit les cellules cancéreuses de la prostate. Elle est utilisée pour certains cancers à risque faible ;
- **l'hormonothérapie** permet de limiter l'action stimulante de la testostérone sur les cellules cancéreuses, le cancer de la prostate étant hormono-sensible.
- **la surveillance active**, nécessite un dosage du PSA couplé à un toucher rectal deux fois par an associé à une biopsie annuelle. En fonction des résultats, d'autres biopsies peuvent être indiquées tous les 2 à 3 ans. Néanmoins cette surveillance active n'est possible que lorsque l'on est en présence d'un cancer peu agressif confirmé par biopsie. La difficulté tient justement à la détermination de son agressivité, peu évidente avec les moyens de diagnostic courants. (cf. section 3.3).

Avancées récentes dans le traitement du cancer de la prostate. Trois nouvelles techniques ont récemment vu le jour pour le traitement du cancer de la prostate :

- **La cryothérapie** consiste à délivrer un froid intense et concentré à l'extrémité d'aiguilles fines insérées dans la prostate atteinte. Elle détruit les cellules cancéreuses par congélation brutale et répétée,
- **la photothérapie dynamique** utilise l'action conjointe d'un photo-sensibilisant et de la lumière laser. L'illumination de la zone ciblée va provoquer une réaction photodynamique entraînant sa destruction. Cette technique est uniquement préconisée pour les cancers localisés et à faible risque.
- **l'ablathermie**, est un traitement du cancer localisé par Ultrasons focalisés de haute intensité (HIFU). La méthode repose sur l'utilisation d'ultrasons émis par une sonde endorectale et focalisés sur la prostate. Le tissu visé est alors détruit par la chaleur. L'ablathermie, utilise un prototype développé au LabTau (Lyon, France), dont les premiers essais cliniques ont été initiés en 1993 à l'hôpital Édouard Herriot [74].

Ces technologies récentes sont aujourd'hui utilisées pour un traitement global de la prostate. Cependant, elles pourraient être aussi utilisées, au moins en théorie, pour ne détruire

que la zone tumorale. C'est le principe du **traitement focal** qui permettrait de diminuer les effets secondaires du traitement tout en assurant un contrôle satisfaisant de la tumeur. Cependant, un traitement focal rationnel nécessiterait la connaissance précise de la position de la tumeur. Obtenir cette information est un des objectifs de cette thèse.

2.6 Problématique du diagnostic du cancer de la prostate

2.6.1 Le diagnostic

Le diagnostic du cancer de la prostate est généralement basé sur un ensemble d'exams combinant le dosage des Antigènes spécifiques à la prostate (PSA), du toucher rectal et les résultats de biopsies guidées par échographie endorectale. Cette approche diagnostic a néanmoins ses limites. Pris individuellement ou combinés, ces trois tests ne sont pas suffisants pour établir un diagnostic précis et les biopsies manquent cruellement de précision.

Toucher rectal. Le toucher rectal ne permet pas de connaître le volume de la glande, seules les parties postérieures et latérales de la prostate sont accessibles à la palpation (plus de 25% des cancers sont ainsi potentiellement manqués), les cancers de petite taille ne sont pas détectés et le test est opérateur dépendant. le toucher rectal n'obtient qu'une sensibilité moyenne de 37% et une Valeur prédictive positive (VPP) faible lorsque le taux de PSA est compris entre 0 et 3 ng/ml [83].

Taux de PSA. Le taux de PSA augmente progressivement avec l'âge. On estime comme normal un taux de PSA sérique total inférieur à 2,5 ng/ml avant 50 ans, < 3,5 ng/ml entre 50 et 60 ans, < 4,5 ng/ml entre 60 et 70 ans et < à 6,5 ng/ml entre 70 et 80 ans. De plus, il augmente temporairement dans d'autres circonstances : après éjaculation, après réalisation d'un toucher rectal, suite à un exercice physique intense, lors d'une infection urinaire ou prostatite aiguë ou encore lors d'une hypertrophie bénigne de prostate [85]. Ce taux est statistiquement corrélé à la présence d'un cancer de la prostate et le seuil de 4 ng/ml est en général celui retenu comme seuil supérieur d'anormalité. A ce seuil, lors de dépistages précoces et suivis (âge du patient entre 60 et 70 ans), ses performances en sensibilité et spécificité sont de 70% et 88% respectivement. Ainsi, en situation de dépistage ponctuel, la VPP est de l'ordre de 30%, traduisant le fait que parmi les personnes ayant un PSA total > 4 ng/ml, 3 sur 10 sont des vrais positifs et 7 sur 10 sont des faux positifs. De même, la Valeur prédictive négative (VPN) est de l'ordre de 90%, donc sur 10 personnes ayant un taux de PSA < 4 ng/ml, 9 n'en ont effectivement pas [90]. Le test de PSA a donc un taux de sensibilité bien plus élevé que le toucher rectal [24] mais de spécificité très faible (36%). Ainsi, l'utilisation répandue du test du dosage PSA augmente le sur-diagnostic du cancer de la prostate.

Combinés, ces deux tests souffrent aussi de mauvaises performances. Un taux de PSA inférieur au seuil de 4 ng/ml et un toucher rectal normal n'éliminent pas le cancer [22].

Par conséquent, les données de PSA et/ou de toucher rectal peuvent aussi causer des sous-diagnostic du cancer de la prostate.

Les biopsies écho-guidées. En cas de résultat anormal lors des tests précédents, il est impératif de pratiquer une série de biopsies endorectales écho-guidées afin de confirmer ou non la présence du cancer. Leur analyse doit permettre d'obtenir le score de Gleason (donc l'agressivité) et le volume de la tumeur pour conclure sur le traitement à adopter. Cependant, les biopsies ne représentent que des échantillons, souvent obtenus aléatoirement, de la prostate [13, 48]. Une série de biopsies négatives n'élimine donc pas la présence de cancer puisque qu'elles peuvent manquer leur cible. Ainsi lorsque le taux de PSA est anormal et les premières biopsies négatives, plusieurs autres séries de prélèvement peuvent être nécessaires. De plus, même positive une biopsie ne sera représentative que d'une portion de la tumeur, il est donc difficile d'en évaluer son volume de manière fiable basé sur ce seul prélèvement.

Certaines études [20] proposent d'augmenter le nombre de prélèvements pris à chaque séance de biopsie, soit par voie endorectale, soit en utilisant une grille périnéale pour améliorer la localisation 3D des résultats (biopsies « de saturation »). Cette approche peut éliminer la présence de cancer et permet une meilleure évaluation du volume tumoral et du score de Gleason. Cependant, elle augmente le coût et la morbidité des procédures (nécessité d'une anesthésie générale), ainsi que le risque de sur-diagnostic de foyers microscopiques indolents [3, 26].

Par conséquent, les recherches actuelles tendant à améliorer le diagnostic des cancers agressifs et à diminuer la détection des lésions bien différenciées comme les cancers de **GS!** (**GS!**) ≤ 6 qui ont des potentiels métastatiques négligeables [81].

2.6.2 Synthèse des faiblesses des techniques actuelles

En résumé, les techniques courantes de diagnostic présentées précédemment (PSA, TR et biopsies) présentent des faiblesses de par leur manque de sensibilité ou spécificité qui ne se compensent pas une fois combinées. De plus, l'échantillonnage biopsique et le caractère semi-aléatoire de la méthode peut entraîner :

- un sous diagnostic lorsque le PSA anormal et la biopsie est négative. Le médecin est alors confronté à la problématique de quand re-biopsier et quand s'arrêter,
- un sur ou sous-diagnostic de l'agressivité à cause de l'hétérogénéité du foyer,
- une sur ou sous-évaluation du volume tumoral de par le prélèvement en lui-même.

2.6.3 Rôle de l'imagerie dans la cartographie locale du cancer de prostate

Pour surmonter ces difficultés, il faut recourir aux méthodes d'imagerie susceptibles d'estimer le volume, la position et l'agressivité de la tumeur. Dans cette optique, il existe

différents types d'imagerie pour la visualisation de la prostate. Néanmoins, toutes ne permettent pas de détecter une tumeur ni de déterminer son évolution ou son agressivité. Par exemple, la **tomodensitométrie** (CT) est peu utilisée en raison de son faible contraste principalement.

On peut de même citer **l'imagerie ultrasonore** avec sonde endorectale. Cette dernière étant non-invasive et peu coûteuse, elle a longtemps été utilisée comme imagerie principale de la prostate. Néanmoins la technique n'est ni sensible, ni spécifique. Elle reste néanmoins utilisée pour guider les biopsies. Les dernières avancées [6] consistent cependant à utiliser conjointement l'imagerie IRM et ultrasonore de manière à améliorer la précision du guidage.

2.6.4 Une technique d'imagerie très prometteuse : l'IRM

L'imagerie IRM s'est imposée comme une technique prometteuse pour la visualisation du cancer [80]. Elle permet de visualiser précisément, grâce à un excellent contraste tissulaire, la glande et la structure des tissus. De récentes études [45] montrent que l'IRM est la meilleure méthode d'imagerie pour le diagnostic du cancer de la prostate en terme de détection et de localisation (surtout sur les patients avec une persistance des biopsies négatives mais un taux de PSA élevé). Elle s'impose de même comme une référence pour les bilans de pré-traitement ainsi que l'évaluation et le suivi post-traitement [59, 86, 94]. L'European Consensus Meeting [22] suggère d'ailleurs l'utilisation de l'IRM multiparamétrique comprenant les séquences en pondération T2 (T2w), de diffusion (Dw) et une imagerie dynamique après injection de produit de contraste (DCE). L'utilisation combinée de ces 3 séquences améliore considérablement la sensibilité et la spécificité des diagnostics en comparaison du PSA, TR et biopsies.

Un suivi axé sur cette modalité permettrait de mieux prendre en charge un patient en triant dans un premier temps les patients nécessitant un réel suivi et ainsi éviter le sur ou sous-diagnostic.

2.7 Conclusion

Le cancer de la prostate est un problème de santé majeur et se positionne en première position des cancers en France. De plus, les techniques de diagnostics les plus répandues et basées sur le toucher rectal et le taux de PSA souffrent d'un manque de précision. Les biopsies permettant de confirmer ou d'infirmer la présence d'un cancer rajoutent une approximation dans la décision finale. De par ces imprécisions, on relève un fort taux de sur-diagnostic, impliquant un traitement lourd, coûteux et invasif pour le patient. De même, de par l'imprécision des biopsies, des cancers sont aussi manqués. Dans ces conditions, l'image IRM multiparamétrique s'impose peu à peu comme une référence pour la détection, localisation et l'évaluation de l'agressivité des tumeurs. C'est un axe de recherche prometteur pour diagnostiquer le cancer de la prostate et ainsi adapter rapidement un suivi

personnalisé des patients. De plus, ses capacités de localisation ouvrent la voie à d'autres techniques de traitement dites focalisées, permettant de traiter localement une tumeur et ne plus systématiquement recourir à l'ablation totale de la prostate.

L'imagerie par résonance magnétique pour le cancer de la prostate

3.1 Introduction

Le chapitre précédent fait état de la problématique du cancer de la prostate dans son ensemble ainsi que des faiblesses des techniques diagnostiques les plus employées. Depuis peu, l'IRM multi-paramétrique (IRM-mp) s'impose comme une technique pour palier à ces faiblesses dans les phases de suivi du patient. Ce chapitre présente dans un premier temps le principe général de l'IRM et des différentes séquences T2w, ADC et DCE généralement utilisées et préconisées pour le diagnostic du cancer de la prostate. Il est ensuite présenté un état de l'art de l'imagerie IRM-mp dans le contexte du diagnostic du cancer de la prostate. Enfin, ce chapitre termine sur la problématique de l'interprétation combinée des ces différentes modalités d'IRM et le besoin de standardisation donnant ainsi la voie au développement des systèmes d'aide à la décision pour le cancer de la prostate. La présentation générale de l'IRM et des modalités T2, ADC et DCE est tirée de la thèse d'Au Hoang Dinh [36].

3.2 Principe général d'acquisition des séquences IRM

3.2.1 Image IRM en pondération T2

Principe de la relaxation T2. Le principe de l'IRM repose sur les propriétés magnétiques du proton H_+ , composant le noyau d'hydrogène. Ce dernier est un élément de la

molécule d'eau (H_2O), présente en grande quantité dans le corps humain. Son moment magnétique peut être représenté sous la forme d'un vecteur en rotation sur lui-même, caractérisant le spin du proton. A l'état de repos, ces vecteurs ont une orientation aléatoire. La résultante magnétique de tous les spins (le moment magnétique macroscopique M_0 ou l'aimantation) est globalement nulle. Soumis à un champ magnétique B_0 , les spins sont maintenus à un certain angle (appelé angle magique) par rapport à ce champ.

Pour générer une aimantation globale, les spins doivent donc être placés dans un champ magnétique constant et uniforme B_0 (étape de polarisation). Pour l'imagerie médicale, ce champ B_0 varie de 0,5 à 7 Teslas. Sous effet de B_0 , tous les spins décrivent un cône autour de l'axe de B_0 , appelé mouvement de précession (cf. figure 3.1 (a)). Celui-ci est proportionnel à l'intensité du champ magnétique et est caractérisé par la fréquence de précession (aussi appelée fréquence de Larmor ω_0). Le proton possède donc un moment magnétique lié à sa charge en rotation (moment magnétique microscopique μ_0).

Sous l'effet de B_0 , le moment magnétique macroscopique M_0 est alors créé et comporte deux composantes : l'aimantation longitudinale M_z (parallèle à l'axe de B_0) et l'aimantation transversale M_{xy} . À l'état d'équilibre (en absence d'excitation par onde radiofréquence), la composante M_{xy} est nulle du fait du déphasage entre les spins.

Les protons sont ensuite excités par une onde radiofréquence (RF) B_1 (étape d'excitation ou résonance). La résonance se produit lorsque la fréquence de cette onde RF B_1 est exactement égale à la fréquence de Larmor ω_0 . Lors de la résonance, tous les spins sont en phase, produisant ainsi l'apparition d'une aimantation transversale M_{xy} (cf. figure 3.1 (b)).

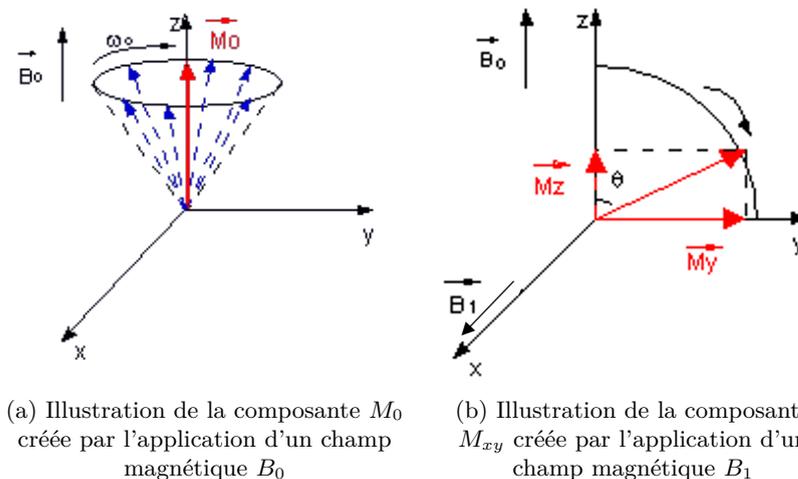


FIGURE 3.1 – (a) L'application d'un champ magnétique B_0 crée une composante longitudinale de l'aimantation M_0 suivant l'axe z, avec M_0 la somme des moments magnétiques de spin de chaque proton (b) L'application d'un courant sinusoïdal à la fréquence de résonance des protons dans une bobine d'axe perpendiculaire à B_0 crée un champ magnétique B_1 qui perturbe les moments magnétiques de spin. L'aimantation résultante M_0 va basculer autour de B_1 . source [Site web laboratoire Navier](#)

Dès l'arrêt de l'excitation de l'onde RF B_1 , les spins se déphasent du fait de leur interaction. En conséquence, l'aimantation M_{xy} décroît rapidement (étape de relaxation). Le temps de relaxation T2 caractérise la perte de l'aimantation transversale dès l'arrêt de l'excitation. La valeur du T2 correspond au moment où l'aimantation transversale atteint 37% (a perdu 63%) de sa valeur initiale (cf. figure 3.2).

C'est l'aimantation transversale M_{xy} qui, en tournant à la fréquence de Larmor ω_0 , va induire dans la bobine qui a servi à l'excitation et qui se positionne maintenant en récepteur, un signal sinusoïdal à la fréquence de Larmor ω_0 . Ce signal est appelé signal de précession libre ou Free Induction Decay (FID). Si le champ B_0 était parfaitement homogène, la constante d'amortissement devrait être T2. Mais, du fait des inhomogénéités de champ, les protons se déphasent plus vite que ne le voudraient leurs interactions au sein du tissu. En conséquence, le signal décroît plus vite que prévu, selon une constante d'amortissement notée T2*. Cette dernière tient donc compte non seulement du T2 (caractéristique du tissu) mais aussi de l'hétérogénéité du champ B_0 .

Ainsi, on souhaiterait dans l'idéal mesurer directement le T2 (décroissance due aux tissus cellulaires) mais c'est impossible car ce signal est "masqué" par inhomogénéité de champ B_0 . On ne mesure alors que le T2* dont la décroissance plus rapide que le T2 va dépendre de la nature du tissu et de l'hétérogénéité du champ B_0 . Afin de s'affranchir de ce phénomène, on va utiliser une séquence spin écho pour ne mesurer que le T2.

Mesure du temps de relaxation T2. Pour mesurer directement le T2, on peut utiliser une séquence « Spin Echo multi écho » en pondération T2. Cette dernière utilise plusieurs impulsions de 180° entre deux impulsions de 90° (cf. figure 3.3). Une séquence d'écho de spin est constituée dans un premier temps d'une impulsion RF de 90° pour basculer l'aimantation dans le plan transversal. Les spins sont alors en phase et l'aimantation transversale M_{xy} est maximale. À la fin de cette impulsion, les spins se déphasent rapidement (par inhomogénéité du champ B_0 et interaction spin-spin illustrés précédemment). Au temps $t=TE/2$, une impulsion de 180° est appliquée pour inverser les déphasages entre les spins. Au temps $t=TE$, les spins sont alors rephasés et l'aimantation M_{xy} est de nouveau maximale induisant un signal d'écho de spin sur la courbe de décroissance du signal T2 qui sera alors directement mesurable.

Ainsi, une cartographie T2 sera reconstruite par un processus de traitement qui combine, pixel par pixel, pour chaque coupe, les images obtenues en multi-écho, selon l'équation :

$$M_{xy} = M_0 \cdot e^{-\frac{TE}{T2}} \quad (3.1)$$

où M_0 désigne la magnétisation du tissu (et représente donc la densité de protons).

Construction de la cartographie T2 de la prostate. Dans les tissus biologiques, en raison de la superposition du champ magnétique local sur les spins, le déphasage entre les protons qui composent les macromolécules est plus rapide (T2 court) que celui entre

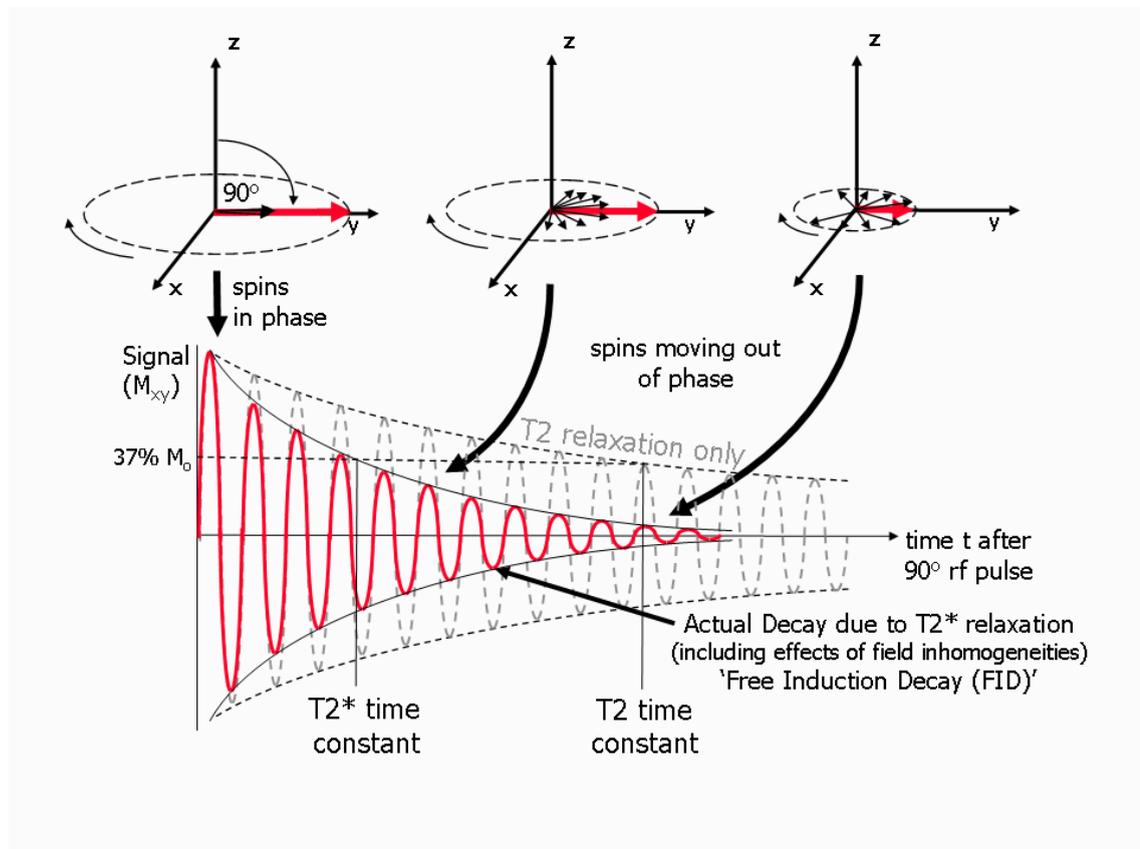


FIGURE 3.2 – Le schéma montre le principe de la relaxation transverse T2 et T2* après qu'une onde RF B_1 ait été appliquée. Dans le premier repère xyz, l'aimantation transverse M_{xy} (flèche rouge) est maximale car les spins précessent en phase. Ensuite, l'amplitude de l'aimantation transverse (le signal détecté) commence à décroître de par le déphasage des moments magnétiques de spins l'un par rapport à l'autre (illustré sur le second repère par les petites flèches noires). Si le champ magnétique B_0 était parfait, le signal mesuré (en noir sur la représentation M_{xy} en fonction du temps) serait uniquement le T2 (caractéristique du tissu) mais de par l'hétérogénéité du B_0 , le signal mesuré est le T2* (en rouge sur la représentation M_{xy} en fonction du temps). Source [Journal of Cardiovascular Magnetic Resonance](#)

les protons mobiles (T2 long, dans l'eau par exemple). Un T2 long se traduit par un signal élevé (blanc) et un T2 court par un signal faible (gris). Par conséquent, dans la zone périphérique de la prostate, les espaces glandulaires prostatiques, riches en liquide, ont un T2 long. Les cancers sont eux plus cellulaires et ont un T2 plus court. Les cancers agressifs sont encore plus cellulaires et devraient avoir le T2 le plus court. Sur l'exemple de cartographie T2 de la prostate (cf. figure 3.4), le tissu d'une ZP normale (entourée en vert) apparaît en blanc et correspond à un T2 long, au contraire du foyer cancéreux (entouré en rouge) apparaissant en gris, correspondant à un T2 court. La zone de transition (entourée en jaune), moins glandulaire que la zone périphérique, a un T2 plus court et est donc composée de gris intermédiaire.

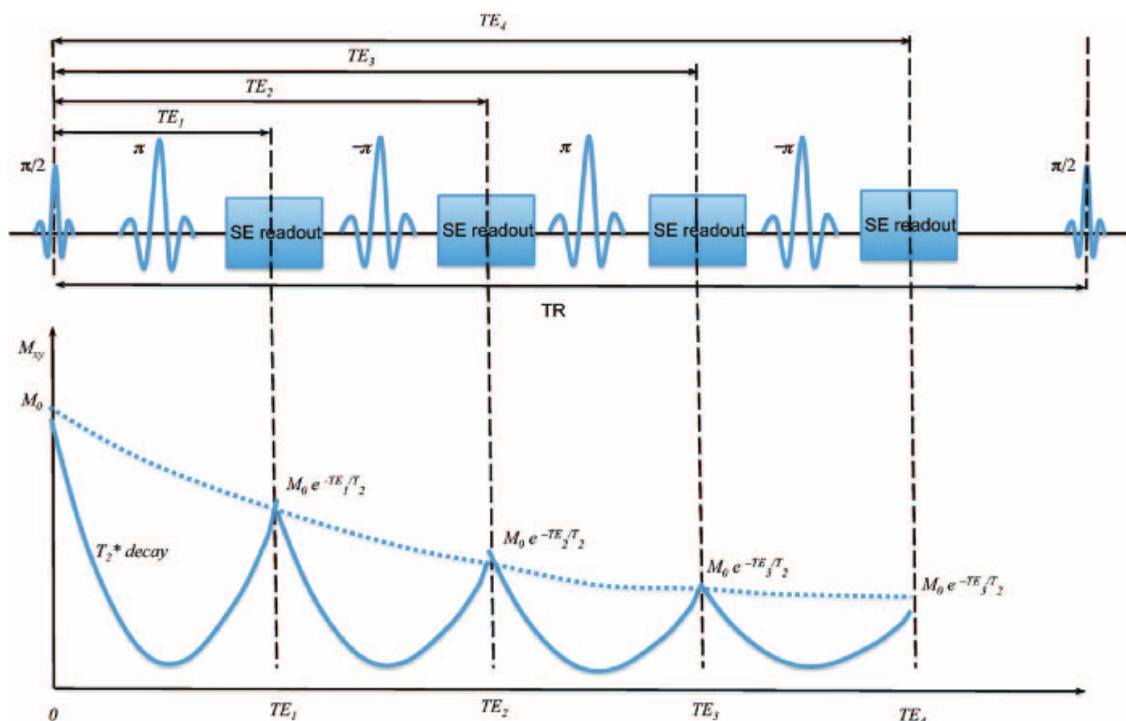


FIGURE 3.3 – Schéma de la séquence multi-écho spin écho (en haut) et la courbe de relaxation T_2 (en bas). La séquence d'écho de spin est constituée dans un premier temps d'une impulsion RF de 90° pour basculer l'aimantation dans le plan transversal. Les spins sont alors en phase et l'aimantation transversale M_{xy} est maximale. À la fin de cette impulsion, les spins se déphasent rapidement (par inhomogénéité du champ B_0 et interaction spin-spin illustrés précédemment). Au temps $t=TE/2$, une impulsion de 180° est appliquée pour inverser les déphasages entre les spins. Au temps $t=TE$, les spins sont alors rephasés et l'aimantation M_{xy} est de nouveau maximale induisant un signal d'écho de spin sur la courbe de décroissance du signal T_2 qui sera alors directement mesurable. source [36]

3.2.2 Séquence IRM pondérée en diffusion

Diffusion moléculaire. A température ambiante les molécules d'eau subissent un mouvement de diffusion aléatoire créé par l'agitation thermique, plus ou moins intense en fonction du milieu biologique dans lesquelles elles se trouvent. Dans un milieu libre, sans obstacles, ce mouvement brownien est isotrope et caractérisé par un coefficient de diffusion D . Dans les tissus organiques, des obstacles tels que les membranes biologiques vont limiter la diffusion de l'eau. Ainsi, lorsque le mouvement des molécules d'eau dans un volume est aisé, l'agitation thermique l'est aussi, on parle alors de diffusion élevée. Lorsque des obstacles s'opposent au mouvement de l'eau, on parle alors de restriction de diffusion. Ainsi, l'objectif des séquences IRM pondérées en diffusion est d'obtenir des images dont le contraste est influencé par les différences de mobilité des molécules d'eau.

Séquence de diffusion. Pour rendre le signal de résonance magnétique sensible à la diffusion moléculaire, on ajoute à une séquence classique des gradients intenses sensibles



FIGURE 3.4 – Exemple de cartographie T2 de la prostate. Sur cet exemple, le tissu d’une ZP normale (entourée en vert) apparaît en blanc et correspond à un T2 long, au contraire du foyer cancéreux (entouré en rouge) qui apparaît en gris, correspondant donc à un T2 court. La ZT (entourée en jaune), moins glandulaire que la ZP, a un T2 plus court et est donc composée de gris intermédiaire. Source : [36]

aux mouvements (appelés gradients de diffusion) [4, 87]. Ces derniers sont appliqués de part et d’autre de l’impulsion 180° de la séquence écho de spin (cf. figure 3.5). Pour les protons immobiles, le déphasage provoqué par le premier gradient est parfaitement compensé par le deuxième, ainsi, le signal issu de la diffusion des molécules d’eau n’est pas atténué. En revanche, les protons mobiles se déphasent lors de l’application du premier gradient et ce déphasage n’est pas compensé par le deuxième. Le signal subit alors une atténuation d’autant plus importante que les mouvements moléculaires sont rapides. Ainsi, les images pondérées en diffusion montrent un hyper-signal dans les régions à diffusion moléculaire restreinte (protons immobiles) et un hypo-signal dans la zone explorée contenant des molécules d’eau à diffusion élevée (protons mobiles). L’effet de diffusion dépend des performances des gradients appliqués et est caractérisé par une valeur de b , exprimé en secondes/mm², déterminée par la relation suivante [87] :

$$b = (\gamma \cdot G \cdot \delta)^2 \cdot (\Delta - \delta/3) \quad (3.2)$$

avec

γ : rapport gyromagnétique

G : amplitude du gradient

δ : durée d'application du gradient

Δ : temps séparant l'application des deux gradients de diffusion

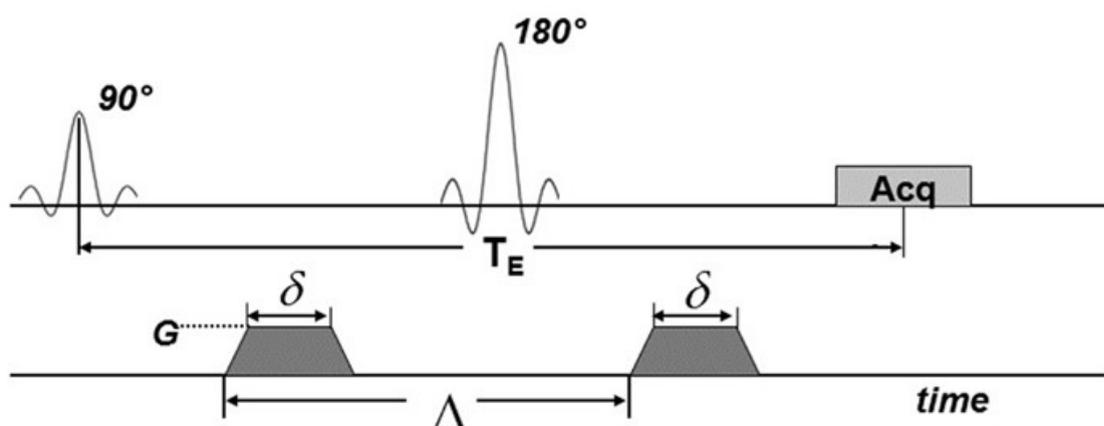


FIGURE 3.5 – Schéma de séquence en diffusion et illustration des paramètres dont le coefficient b dépend. Pour mettre en évidence des mouvements moléculaires microscopiques, on ajoute à une séquence classique des gradients intenses sensibles à la diffusion de l'eau. Ces derniers sont appliqués de par et d'autre de l'impulsion de 180° de la séquence écho de spin. Source : [36]

La valeur de « b » peut varier de 0 à environ 3000 s/mm^2 sur les imageurs actuels. Pour mesurer la diffusion, on utilise dans un premier temps une série de coupes avec une valeur de $b=0$ (sans gradients de diffusion, en pondération T_2) puis une série de coupes avec un facteur de b compris, en général, entre 500 et 3000 s/mm^2 . En augmentant le facteur de b , on augmente la pondération en diffusion ayant pour effet d'améliorer la sensibilité au phénomène de diffusion moléculaire au détriment du rapport signal sur bruit.

Construction de la cartographie ADC. En imagerie IRM, Le paramètre quantitatif permettant d'évaluer l'atténuation d'un signal liée au phénomène de diffusion des molécules d'eau dans les tissus biologiques est appelé Coefficient apparent de diffusion (ADC). On exprime alors cette atténuation S_1 en fonction de ce coefficient par la relation suivante [87] :

$$S_1 = S_0 \cdot e^{-b \cdot ADC} \quad (3.2)$$

avec

b : facteur de gradients de diffusion

ADC : Apparent Diffusion Coefficient

L'ADC s'exprime en mm^2/s (unité opposée à celle de b). Pour calculer l'ADC, il faut utiliser au moins deux acquisitions en imagerie de diffusion : en général une première acquisition sans gradients de diffusion ($b=0$) donnant un signal S_0 et une acquisition comportant des gradients de diffusion avec une valeur de b (entre 0 et 2000 s/mm^2 généralement) de manière à générer le signal S_1 . La valeur d'ADC est alors calculée à l'aide de la relation suivante [87] :

$$ADC = \frac{\log_2(\frac{S_1}{S_0})}{b} \quad (3.2)$$

Ainsi, une cartographie d'ADC sera générée par un processus de traitement qui combine, pixel par pixel, pour chaque coupe, les images obtenues pour des valeurs de b différentes. Il existe plusieurs méthodes pour calculer une carte ADC selon l'utilisation d'un modèle mono ou bi-exponentiel. Dans le cas d'un modèle mono-exponentiel (celui que nous avons utilisé) et bien qu'il s'agisse d'une régression linéaire, il est possible d'utiliser sur l'imageur plus de deux valeurs de b afin de déterminer les paramètres S_0 et ADC . Dans sa lettre, Park et al. [69] revient sur la formulation mathématique du modèle mono-exponentiel et montre qu'au delà de trois valeurs de b utilisées, du bruit est introduit dans le calcul de la carte ADC. En effet, le modèle étant souvent déterminé à l'aide d'une régression linéaire de type moindres carrés, l'utilisation d'une troisième valeur de b n'aura aucune incidence sur la pente ADC du modèle, uniquement sur l'ordonnée à l'origine $\ln(S_0)$.

Sur l'image de la cartographie d'ADC, les zones à diffusion moléculaire lente sont représentées en hypo-signal (ADC faible) alors que les zones à diffusion rapide en hyper-signal. Un tissu prostatique sain, du fait de sa structure glandulaire contenant le liquide séminal, présente une diffusion moléculaire élevée (hyper-signal sur la carte d'ADC). Lorsque le cancer apparaît, les cellules carcinomes se développent très rapidement et occupent l'espace contenant le liquide séminal. Celui-ci introduit une augmentation de densité cellulaire dans le foyer cancéreux. En conséquence, la diffusion moléculaire de cette zone sera réduite (zone en hyper-signal sur les images de diffusion (figure 3.6a) et en hypo-signal sur la cartographie de l'ADC (figure 3.6b)). En principe, plus le score de Gleason d'une tumeur est élevé et plus la densité cellulaire est élevée, entraînant alors une forte restriction de la diffusion moléculaire.

3.2.3 Image IRM de perfusion DCE

Principe. L'étude de la cinétique d'un produit de contraste, injecté par voie intraveineuse dans un tissu biologique, a pour but d'examiner sa structure et sa fonction micro-

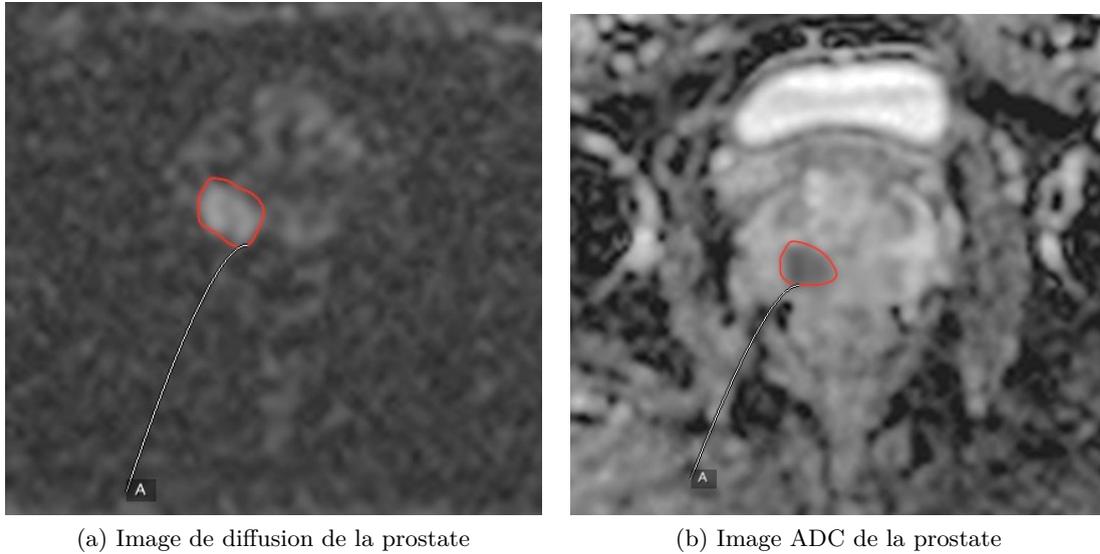


FIGURE 3.6 – Sur l'image de diffusion 3.6a le cancer contouré en rouge apparaît en hypersignal (zone blanche) de par la restriction de la diffusion. Sur l'image ADC 3.6b, par convention, c'est l'inverse, la restriction de la diffusion apparaît en hypo-signal (zone noire). Source [36]

vasculaire. Après injection, l'agent de contraste entre dans les artérioles de la tumeur, traverse ses lits capillaires et est finalement évacué via les veines intra-tumorales. L'image IRM de perfusion, appelée *dynamic contrast-enhanced (DCE) imaging*, consiste à acquérir des images 3D, en pondération T1, tout au long du processus d'injection du produit de contraste afin d'étudier ses caractéristiques cinétiques. A noter que l'agent de contraste paramagnétique le plus utilisé est le gadolinium. En général, une acquisition rapide de quelques secondes permet d'observer l'intensité du signal IRM dans la tumeur. Cette acquisition permet de construire une courbe intensité du signal en fonction du temps dont on extrait une série de paramètres physiologiques renseignant sur la structure et la fonction du système micro-vasculaire de la tumeur. La forme de cette courbe va dépendre de plusieurs facteurs physiologiques comme la perfusion tissulaire, la Fonction d'entrée artérielle (AIF), la surface et la perméabilité capillaire ainsi que le volume de la tumeur.

Analyse qualitative. La méthode d'analyse qualitative ou visuelle des données dynamiques est basée sur l'hypothèse que les vaisseaux tumoraux sont plus nombreux et que leur perméabilité capillaire augmente. Ceci induit un échange rapide de l'agent de contraste entre les capillaires et le tissu tumoral. Les structures tumorales présentent ainsi généralement un rehaussement de contraste précoce, rapide et fort tout de suite après l'injection suivi d'une perte de contraste relativement rapide. Au contraire un tissu sain présente une augmentation de contraste lente et continue (cf. figure 3.7).

Analyse semi-quantitative. Cette approche est issue de l'analyse de la courbe d'évolution de l'intensité du signal au cours du temps dans la tumeur. (cf. figure 3.8). Les

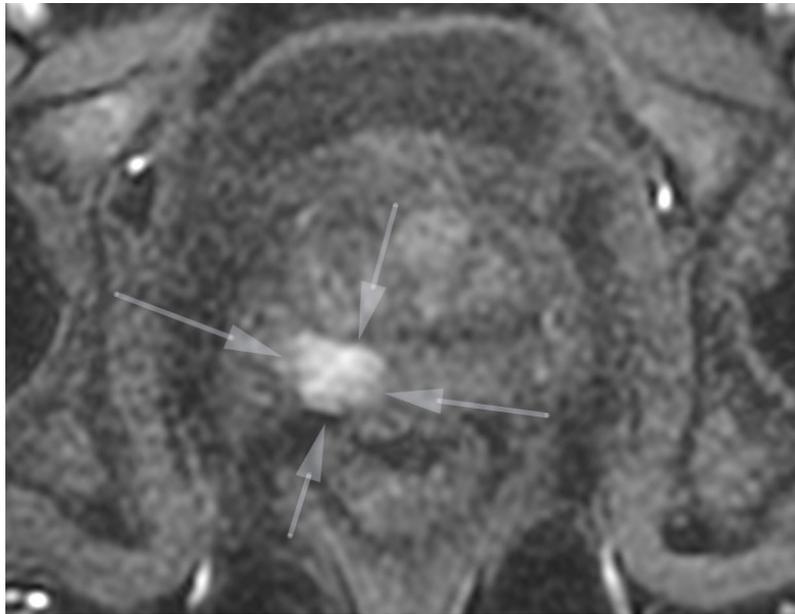


FIGURE 3.7 – Exemple d'analyse qualitative d'une séquence IRM dynamique. Le foyer cancéreux présente un réhaussement de contraste (absorption plus rapide du produit de contraste) en comparaison des tissus sains avoisinant. Source [36]

paramètres qui en sont extraits sont nombreux et dépendent du choix d'un modèle mais les principaux sont :

- Temps d'arrivée (TA) : c'est le temps séparant le début de l'injection du début de réhaussement,
- Vitesse de réhaussement (WI) : indication sur la vitesse d'absorption du produit de contraste par la tumeur,
- Vitesse de lavage (WO) : caractérisation de la cinétique d'élimination du produit de contraste,
- Temps d'arrivée au pic (TTP) : c'est le temps auquel arrive le maximum de réhaussement,
- Pic de réhaussement (PE) : La valeur maximale du signal généralement exprimée en %.

Analyse quantitative. L'analyse quantitative utilise des modèles pharmacocinétiques basés les changements de concentration de l'agent de contraste (gadolinium) dans un tissu et sur l'artère supportant le tissu (fonction AIF). En pratique, il est nécessaire de mesurer le temps de relaxation T1 des tissus afin de calculer la concentration de gadolinium à partir du signal observé. Dans ces modèles, le tissu est modélisé au niveau capillaire en trois compartiments élémentaires : intravasculaire (plasma), extravasculaire-extracellulaire et intracellulaire. La plupart de ces modèles reposent sur le taux d'échange d'agent de contraste entre le plasma et l'espace extracellulaire en supposant qu'il n'y a pas d'échange avec l'espace intracellulaire (figure 3.9). De même, plusieurs modèles sont proposés pour

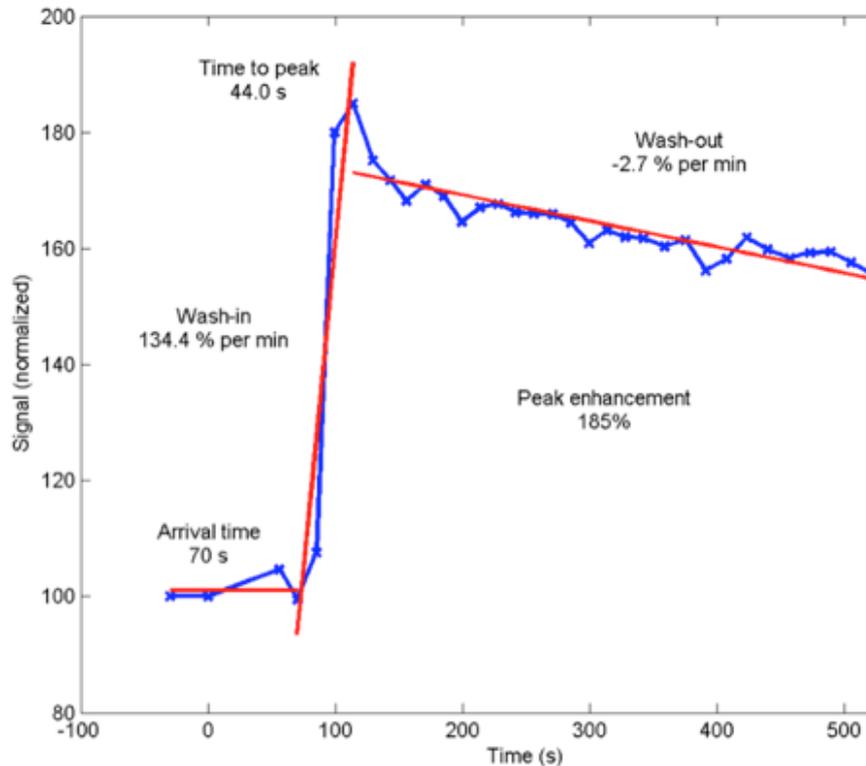


FIGURE 3.8 – Schéma d'une courbe de réhaussement issue d'une séquence IRM dynamique avec injection d'un produit de contraste. Les paramètres illustrés ici sont TA, WI, WO, TTP et PE. Source [36]

extraire des paramètres quantitatifs. On peut citer par exemple [97] :

- K_{trans} : constante de transfert représentant principalement la perméabilité capillaire.
- v_e : fraction de volume de l'espace extravasculaire-extracellulaire,
- $k_{ep} = \frac{K_{trans}}{v_e}$: constante de transfert décrivant le flux d'agent de contraste de l'espace extracellulaire vers le plasma,
- v_p : fraction du volume de plasma.

Angiogenèse du cancer de la prostate. Comme pour la plupart des cancers, les cellules carcinomes de la prostate produisent des facteurs d'angiogénèse affectant la perméabilité et la croissance vasculaire. Par conséquent, le nombre de vaisseaux mais aussi leur perméabilité s'en trouvent augmentés. En général, les vaisseaux tumoraux sont plus perméables, plus nombreux, plus hétérogènes mais aussi plus désorganisés que les vaisseaux normaux. En conséquence, le cancer de la prostate est généralement rapidement contrasté sur l'image IRM (approche qualitative) ; le Wash-in, le K_{trans} plus élevés et le Wash-out plus rapide que pour un tissu normal (approche semi-quantitative et quantitative).

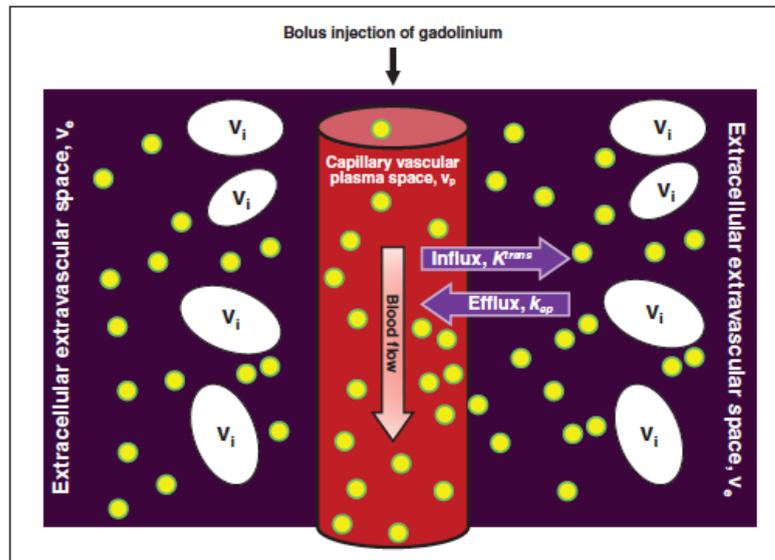


FIGURE 3.9 – Schéma des échanges de l'agent de contraste avec les compartiments avoisinant. Il apparaît le compartiment plasmatique vasculaire (en rouge), le compartiment extravasculaire-extracellulaire représenté par l'espace intra-cellulaire (en blanc) et l'interstitium (en noir) et l'agent de contraste (en jaune). v_e et v_p sont les fractions volumiques de l'extravasculaire-extracellulaire et du plasma respectivement. k_{ep} est la constante de flux de l'espace extravasculaire-extracellulaire vers le plasma. $K_{trans} = k_{ep} \cdot v_e$ est une constante de transfert. Source [36]

3.3 Diagnostic du cancer de la prostate par imagerie IRM multiparamétrique

Cette partie dresse un état de l'art non exhaustif de l'utilisation de l'IRM pour le diagnostic du cancer de la prostate.

Performances des séquences IRM seules. L'utilisation des séquences IRM seules pour le diagnostic n'a qu'un intérêt limité. En effet, des études montrent que l'utilisation de la séquence T2w seule n'a qu'une sensibilité de détection de 25-60% et une spécificité de 57-98% [28, 41, 43, 76]. De plus, on constate une sous-estimation du volume tumoral par rapport à la vérité histologique [43]. Les causes en sont diverses mais l'on constate surtout qu'entre cancers plutôt bien différenciés en termes de score de Gleason et parfois entre tissu sain et cancer, le contraste ne permet pas une différenciation nette. De même l'utilisation de la séquence DCE seule ne permet d'obtenir qu'une sensibilité de 46-60% et une spécificité de 51-97% [28].

L'IRM multiparamétrique pour l'amélioration de la détection et localisation.

En raison des faiblesses des séquences prises seules, l'IRM multiparamétrique, combinant les images T2w, ADC et DCE s'est imposée peu à peu et a amélioré la détection et la localisation tumorale [21, 37]. Deux études récentes [23, 46] sur des images IRM 3T montre

que la combinaison du T2w à l'ADC et/ou au DCE permet d'améliorer la sensibilité de détection du cancer de la prostate dans la zone périphérique de 0.63 à 0.79-0.81. Concernant la localisation de la tumeur, l'association T2w-DCE permet d'obtenir une sensibilité de 0.72-0.91 au lieu de 0.69-0.72 pour le T2w seul. De même, dans une étude [32] sur 49 patients avec prostatectomie radicale comme gold standard, lorsque l'on combine le T2w à la séquence ADC, on passe d'une performance en sensibilité de 0.54 pour le T2w seul à 0.81. Idem pour la spécificité, on note une amélioration de 0.81 à 0.89. Enfin, une étude [42] sur 75 patients avec prostatectomie radicale de même, montre que la combinaison des 3 séquences T2w, DCE et ADC permet d'obtenir une sensibilité de 0.59 contre 0.31, 0.27 et 0.44 respectivement pour les 3 séquences prises individuellement.

L'IRM-mp pour le diagnostic d'agressivité. L'utilisation de l'IRM-mp pour l'identification du grade des lésions est aussi prometteuse. En fonction de leur score de Gleason ainsi que de leur volume, on note une amélioration dans la combinaison des séquences. Une étude récente [103] sur 74 patients, avec prostatectomie radicale, utilisant l'image T2w seule, présente des taux de détection de 79% pour les cancers de grade Gleason 8 et 9 et de 43% pour les Gleason 6 et 7. A titre de comparaison sur une autre étude sur 175 patients [10] avec prostatectomie radicale et utilisant l'IRM-mp, son analyse permet d'identifier l'ensemble des cancers de grade Gleason 8 et 9, 82.5% à 86% des cancers Gleason 7 de plus de 0.5cc et 59.3 à 62.9% pour ceux de moins de 0.5cc. Les cancers de grade Gleason 6 de plus de 2cc sont identifiés à hauteur de 78% et 48.3-62% pour ceux entre 0.5 et 2cc.

L'interprétation simultanée des images n'est pas toujours aisée. L'interprétation images IRM multiparamétriques est actuellement effectuée par une approche qualitative où les radiologues tentent d'analyser simultanément « à l'œil » les 3 séquences IRM. Cette analyse reposant sur les contrastes entre tissus sains et tumoraux présente néanmoins des limites. En effet, les niveaux de gris de l'image IRM et les contrastes dépendent des caractéristiques de l'IRM (notamment la puissance et l'homogénéité du champ magnétique B_0 , l'onde radiofréquence B_1 ou la sensibilité de l'antenne de réception) mais aussi des choix du protocole d'acquisition par l'opérateur (temps de répétition et temps d'écho ayant un impact sur le contraste de l'image) [56]. L'étude de Bratan et al. [10], dont le but est l'évaluation qualitative du signal IRM pour la discrimination des tumeurs, montre que parmi les lésions visibles en IRM, 40% d'entre elles sont bénignes. Bien que les tumeurs agressives semblent mieux détectées « à l'œil », la capacité de l'IRM multiparamétrique à distinguer les tumeurs agressives des autres lésions focales intra-prostatique (bénignes ou cancer indolents) reste problématique [56, 78]. En effet, les séquences IRM peuvent se contredire et l'analyse nécessite un œil expérimenté et donc très subjectif.

Parce qu'il n'est pas possible d'avoir un diagnostic de malignité de certitude, il est recommandé que le radiologue donne à chaque anomalie dans la prostate un score de suspicion de malignité à 5 niveaux (1 : formellement bénin ; 2 : probablement bénin ; 3 : indéterminé ;

4 : probablement malin ; 5 : formellement malin). Même si la communauté radiologique a tenté de minimiser la subjectivité de ce score en introduisant le score PI-RADS [5] [104], la variabilité interlecteur reste grande [95] [79].

Enfin, les protocoles d'acquisition et la façon d'interpréter les images ne sont pas standardisés. En conséquence, les bons résultats obtenus par des groupes spécialisés ne peuvent pas toujours être reproduits ailleurs. La comparaison inter-scanner, intra-scanner ou longitudinale (suivi dans le temps) du patient reste donc difficile à réaliser [56].

Besoin de standardisation. De par les constats dressés précédemment, il est clair que l'IRM-mp est très prometteuse comparée aux moyens diagnostiques actuels. Néanmoins son utilisation à large échelle nécessite une standardisation à la fois dans la méthode d'acquisition des séquences mais aussi dans la façon d'interpréter les images afin de s'affranchir des biais liés aux caractéristiques intrinsèques de l'IRM. Dans ces conditions, les systèmes d'aide au diagnostic (CAD) ont vu le jour dans ce besoin de standardisation. Basés sur l'extraction de caractéristiques propres à chacune de ces séquences IRM, les CAD ont pour objectif ambitieux de détecter, localiser et estimer l'agressivité des lésions. Le chapitre suivant présente l'utilisation des CAD pour le diagnostic du cancer de la prostate et dresse un état de l'art de leurs performances et la manière dont ils sont utilisés par les radiologues.

3.4 Conclusion

Ce chapitre présente les bases de l'IRM ainsi que les propriétés des différentes séquences T2w, ADC et DCE notamment pour la visualisation du cancer de la prostate. L'IRM est une technique d'imagerie en effet très prometteuse pour le diagnostic du cancer de la prostate en comparaison des méthodes actuelles (PSA, toucher rectal et biopsie). L'apport d'une séquence IRM utilisée seule pour le diagnostic est avéré mais reste toutefois limité. En revanche l'utilisation de l'IRM-mp combinant notamment les séquences T2w, ADC et DCE permet d'obtenir de très bonnes performances de détection et localisation ainsi qu'un apport notable dans la différenciation des tumeurs agressives des tissus sains. L'interprétation d'images IRM multiparamétriques reste cependant difficile, effectuée très souvent de manière qualitative donc opérateur dépendant, et le contraste des images dépend fortement des caractéristiques intrinsèques de l'IRM utilisé en pratique. Un besoin de standardisation s'impose, à la fois sur le protocole d'acquisition ainsi que sur la manière d'interpréter. C'est de ce constat que de nombreuses recherches sur les systèmes d'aide au diagnostic tentent de répondre à ce besoin de standardisation.

Les systèmes d'aide au diagnostic pour l'imagerie du cancer de la prostate

4.1 Introduction

Le chapitre précédent donne un aperçu des différentes séquences IRM utilisées pour le diagnostic du cancer de la prostate, leur apport par rapport aux techniques classiques (PSA, Toucher rectal (TR), biopsies) mais aussi le besoin de standardisation dans l'interprétation des images. Dans cette optique, les Systèmes d'aide à la décision (CAD) ont pour objectif d'assister le radiologue dans sa prise de décision. Leur finalité est la détection, la localisation, l'estimation du volume tumoral ainsi que leur différenciation en fonction du score de Gleason. Il est cependant important de préciser que les CAD n'ont pas vocation à remplacer l'avis du radiologue mais bien de le compléter. Son rôle est de l'ordre du second avis, permettant de confirmer ou infirmer un avis initial du praticien, mais aussi d'attirer son attention sur des zones potentiellement pathologiques. Les CAD sont à l'intersection de plusieurs disciplines telles que la médecine, la biologie, la physique d'acquisition des images, l'intelligence artificielle ou encore le traitement d'images. Celles-ci sont toutes complémentaires et l'amélioration des performances d'un CAD passe par l'étude et la recherche dans chacune d'elle. Historiquement, c'est à partir de 1950 [27] que l'utilisation des ordinateurs a été considérée pour l'analyse d'anormalités radiographiques. Ce n'est néanmoins qu'à partir de 1980 que les chercheurs et radiologues ont entamé de vastes efforts pour l'émergence des systèmes d'aide à la décision en utilisant l'ordinateur comme une aide plutôt qu'un système totalement automatisé comme ce fut le cas his-

toriquement pour la détection de lésions sur des radiographies de l'abdomen et sur des mammographies. Depuis, de nombreuses études ont été menées sur l'analyse d'images 2D ou 3D pour la détection ou le diagnostic d'anormalités. Depuis ces premières recherches, il a été montré que les CAD améliorent les performances de diagnostic des radiologues. Chan *et al.* [14] ont montré une amélioration significative de 4% dans la détection du cancer du sein, confirmé plus tard dans d'autres études [19]. De similaires conclusions ont été faites sur le cancer des poumons [50], sur le cancer du colon [71] et sur le cancer de la prostate [33]. Dans le cas du cancer de la prostate, les études publiées par Niaf *et al.* [63] et Hambrock *et al.* [33] notamment ont montré l'apport d'un CAD sur les performances diagnostic de radiologues expérimentés et moins expérimentés. L'aide du CAD permet d'ailleurs aux moins expérimentés de se rapprocher des performances diagnostiques des plus expérimentés. Néanmoins, les CAD pour le cancer de la prostate sont relativement récents en raison de l'utilisation récente de l'imagerie IRM. La première étude a été publiée en 2003 par Chan *et al.* [15] et depuis, ce sont plus de 50 études qui ont été publiées [49], montrant particulièrement l'importance et le rôle prometteur des CAD pour le cancer de la prostate.

La suite de ce chapitre présente dans un premier temps les deux grandes familles de CAD : les CADe et les CADx, leurs particularités ainsi que leur fonctionnement. Il sera notamment développé la méthodologie de conception de ces systèmes ainsi que toutes les étapes nécessaires à leur réalisation. Le chapitre termine sur un état de l'art récent des CAD pour le diagnostic du cancer de la prostate.

4.2 Approches par région (CADx) Versus approches par voxel (CADe)

On distingue principalement deux formes de CAD : le Système d'aide au diagnostic (CADx) et le Système d'aide à la détection (CADe) (cf. figure 4.1). Le CADe fournit généralement une carte de probabilité de présence du cancer dont chaque valeur est générée voxel à voxel. Son rôle est assimilé à une tâche de détection et localisation faisant apparaître des zones suspectes. Le radiologue étudie ensuite cette carte de probabilité pour repérer ces zones suspectes et ainsi confirmer ou infirmer sa première analyse des images IRM. En revanche, le CADx fourni, pour une région suspecte délimitée par le radiologue ou par un processus automatique, un score pouvant être sous la forme d'un descripteur mathématique pour caractériser la lésion ou bien une probabilité de cancer ou d'anormalité. Dans le cas du CADx le radiologue vient pointer la zone qui lui paraît suspecte alors que le CADe fournit une estimation d'où peut se trouver cette zone suspecte. Dans les deux cas, le but reste de réduire les erreurs d'identification et d'interprétation des zones suspectes mais aussi la variabilité de diagnostic d'un radiologue à l'autre.

Bien que les deux approches soient différentes, il est courant de trouver des systèmes automatisés combinant un CADe suivi d'un CADx. En effet, le CADe effectuant une tâche

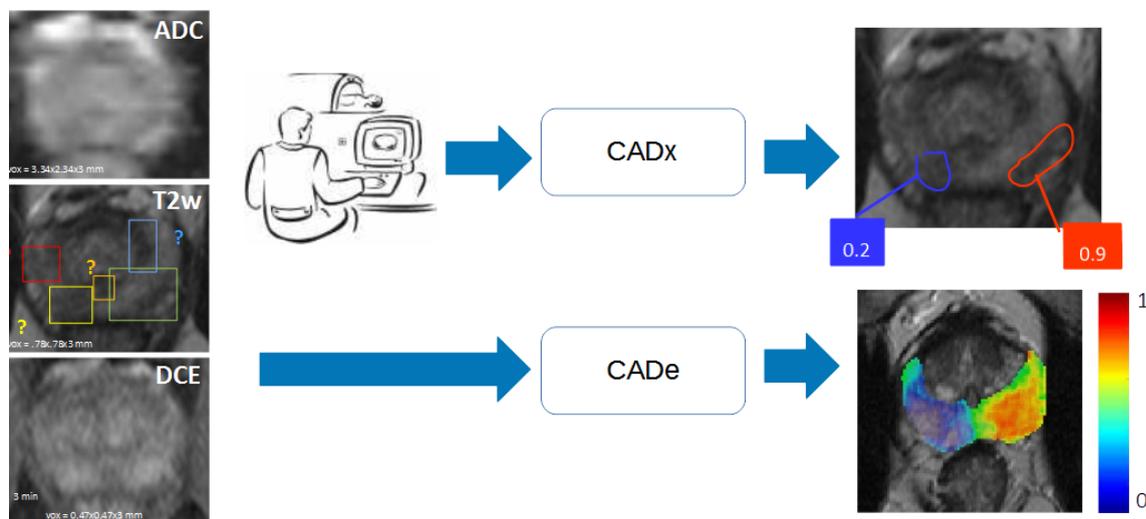


FIGURE 4.1 – Schéma de principe d'un CADe et d'un CADx. Dans le cas d'un CADx, le radiologue interroge le système sur des zones qu'il a contouré lui-même sur l'image IRM et qui lui renvoie un score de suspicion. Dans le cas d'un CADe, c'est une carte de probabilité qui est directement générée (sur la ZP dans cet exemple), le radiologue analyse ensuite cette carte pour identifier les zones suspectes.

de détection, il peut être suivi d'une étape de segmentation automatique permettant de synthétiser l'information provenant du CADe et générer des régions d'intérêts suspectes. Ces zones suspectes sont ensuite filtrées et utilisées en entrée d'un CADx.

4.3 Méthodologie de développement d'un CAD

L'implémentation d'un système CAD peut différer suivant sa nature, CADx ou CADe, et l'automatisation de certaines étapes. On peut néanmoins établir un schéma méthodologique générique, basé sur les systèmes de classification, dont les principales étapes sont détaillées sur le schéma 4.2.

Il est important de noter que la réalisation d'un CAD se déroule en deux principales étapes : la phase d'apprentissage (partie "offline" grisée sur la figure 4.2) et la phase de test dite d'évaluation. La première a pour but d'entraîner le système sur des données d'apprentissage afin de déterminer un modèle générique qui sera utilisé en pratique pour classer les données d'entrée. La seconde partie utilise ce modèle sur des données de test afin d'évaluer les performances du CAD. Chacune des étapes est détaillée dans les sections suivantes.

4.3.1 Recalage

Le recalage utilisé en imagerie médicale possède de nombreuses applications. Il permet de fusionner plusieurs images d'un même patient afin d'exploiter les informations fournies par différentes séquences IRM par exemple. Dans le cadre des CADe, l'objectif est de fournir

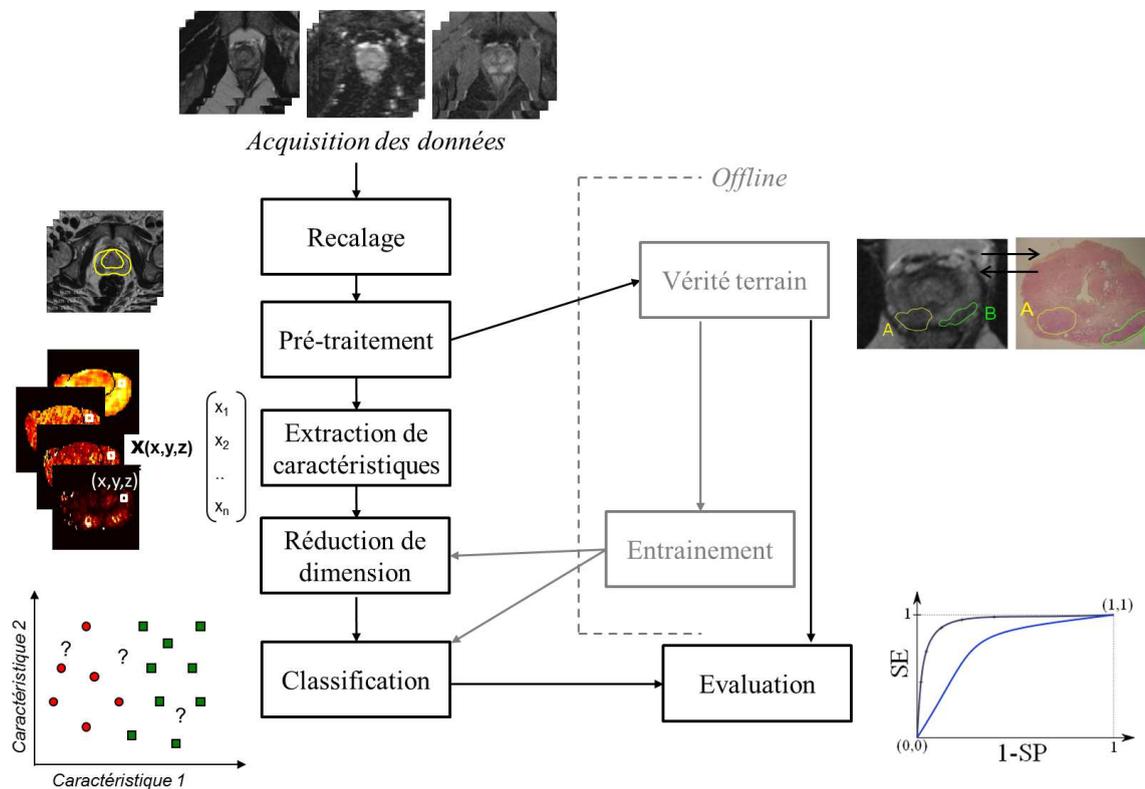


FIGURE 4.2 – Principales étapes d’un système d’aide à la décision. Source : thèse Émilie Niaf [62]

une carte de probabilité de cancer de la prostate voxel à voxel. Cette application nécessite d’aligner les 3 séquences IRM T2w, ADC et DCE de manière à mettre en correspondance les mêmes zones anatomiques. Le mauvais recalage des séquences peut influencer négativement les performances du CAD et introduire du bruit sur les cartes de probabilité. Une description complète de nos travaux de recalage est disponible dans la partie 8.

4.3.2 Pré-traitement des images

Lors de l’acquisition d’une image, un certain nombre de bruits peuvent s’ajouter. Ce bruit est en général aléatoire et dû aux perturbations lors de l’acquisition (électronique, capteur, humidité ou encore température), de l’échantillonnage ou lors de la quantification du signal. La quantification permet de passer d’un signal physique à un signal numérique pour constituer une image. Durant ce processus, des variations d’intensités n’ayant pas de signification anatomique au sein d’un même tissu viennent entacher d’artefacts les images finales. En IRM, la cause la plus fréquente est la sensibilité en émission et réception des antennes radio-fréquence (RF) et parfois appelé inhomogénéité RF. Non corrigés, ce genre d’artefacts vient perturber les algorithmes de segmentation ou de classification. Les principales corrections peuvent être effectuées directement lors de l’examen, sur la console

d'acquisition de l'IRM, à l'aide d'outils implémentés par le constructeur.

4.3.3 Extraction de caractéristiques

L'extraction de caractéristiques est une étape très importante pour les performances d'un CAD. A partir d'images, le principe est de calculer un ensemble de descripteurs, pour chaque voxel ou région, permettant la discrimination du cancer. En imagerie médicale, ces descripteurs peuvent être par exemple le volume, la forme, la texture, l'intensité ou encore des paramètres statistiques ou pharmaco-cinétiques.

4.3.4 Réduction de dimension

Un trop grand nombre de caractéristiques modélisant un problème introduit de la complexité, une augmentation des temps de calcul et une nette baisse de performances du système de classification en présence de données bruitées. Par exemple, ne pas éliminer une caractéristique fortement corrélée avec une autre peut avoir des effets néfastes sur le classifieur en le rendant trop spécifique à la base qui lui a servi d'apprentissage. Ainsi le classifieur perd sa capacité de généralisation et peut ne pas donner suffisamment d'importance à l'information significative.

Pour palier à ces problèmes on utilise la réduction de dimension dont l'objectif est :

- Éviter le sur-apprentissage afin de disposer d'un système permettant de généraliser au mieux,
- Diminuer la complexité algorithmique pour diminuer les temps de calcul,
- Améliorer l'interprétabilité du modèle.

Il existe deux types de méthodes de réduction de dimension :

- La **sélection de caractéristiques** visant à sélectionner parmi l'ensemble de départ les caractéristiques les plus informatives.
- La **transformation des données** consistant à remplacer l'ensemble initial des données par un nouvel ensemble réduit.

Sélection de caractéristiques

Afin de sélectionner les caractéristiques les plus discriminantes de l'ensemble départ, il faut mettre en place une procédure recherche. Les stratégies de recherche sont en général classées en 3 catégories : *exhaustive*, *heuristique* et *aléatoires*. Une stratégie *exhaustive* vise à déterminer le meilleur sous-ensemble en testant toutes les combinaisons possibles. L'inconvénient majeur de cette technique réside dans sa mise en œuvre pratique quasi impossible en présence de nombreuses caractéristiques (2^n combinaisons possibles pour n caractéristiques). Les méthodes *heuristiques* travaillent par paquet de caractéristiques et ne testent évidemment pas tout l'espace possible des combinaisons. On peut citer par exemple les approches *forward*, qui partent d'un ensemble vide puis ajoute au fur et à mesure des caractéristiques ; les approches *backward*, qui partent d'un ensemble complet puis retirent des caractéristiques à chaque itération ; ou encore *stepwise* qui mélange les

deux approches précédentes. Enfin, l'approche aléatoire tire plusieurs sous-ensembles afin de sélectionner le meilleur. Cette dernière approche a pour avantage de limiter les temps de calcul.

Une fois la procédure de recherche déterminée, on doit maintenant choisir une méthode permettant d'évaluer un sous-ensemble de caractéristiques sélectionnées. Les méthodes d'évaluation sont généralement de 3 types : *filter*, *wrapper* et *embedded*. La méthode *filter* utilise comme critère d'évaluation la pertinence d'une caractéristique selon des mesures extraites des données d'apprentissage. On peut par exemple citer le test de student qui calcule une p-valeur sur la significativité de la différence des moyennes d'une caractéristique sur deux populations (deux classes) supposées indépendantes. Les caractéristiques sont alors ordonnées en fonction des valeurs absolues des p-valeurs obtenues à un test de student. L'information mutuelle quant à elle mesure une dépendance entre les distributions de deux populations. Par exemple on peut calculer l'information mutuelle entre la valeur d'une caractéristique calculée pour chaque voxel et les labels de classe associés. Ainsi, plus l'information mutuelle est grande et plus la caractéristique est discriminante. L'inconvénient majeur des méthodes de type *filter* est qu'elles sont calculées "offline", c'est-à-dire sans prendre en compte l'influence de la sélection sur la performance de classification. Les méthodes de type *wrapper* visent à résoudre ce problème en évaluant un sous-ensemble de caractéristiques par sa performance de classification. L'avantage de cette méthode est la possibilité de trouver l'ensemble optimal de caractéristiques maximisant les performances d'un classifieur donné. De plus, elles sont considérées comme étant meilleures que les méthodes de types *filter* [39,66]. Les principaux inconvénients sont la dépendance à un classifieur donné, la complexité et le temps de calcul nécessaire pour la sélection. Les méthodes *embedded* intègrent la sélection de caractéristiques directement lors du processus d'apprentissage. A la différence des méthodes *wrapper* qui utilisent une base d'apprentissage et une base de validation pour sélectionner un sous-ensemble, les méthodes intégrées se servent de tous les exemples d'apprentissage pour sélectionner les caractéristiques les plus pertinentes. Elles sont en général plus rapides que les méthodes *wrapper* car l'apprentissage est réalisé une seule et unique fois et non sur chacun des sous-ensembles.

Transformation des données

Parmi les méthodes de transformation des données, on peut citer l'analyse en composantes principales (ACP). l'ACP permet de trouver un espace de dimension réduite permettant d'expliquer les tendances des variables (caractéristiques). Ainsi, elle consiste principalement à transformer les variables quantitatives de départ, possiblement corrélées, en des variables non corrélées, combinaisons linéaires des variables initiales. Ces nouvelles variables sont appelées axes principaux et sont déterminées de manière à expliquer le maximum de variance dans la distribution initiale des données. Du fait de sa nature même, l'ACP a le principal inconvénient de ne pas expliquer les dépendances non linéaires entre variables. Une autre méthode souvent utilisée est l'analyse discriminante linéaire (LDA).

Lorsque les labels des classes sont à disposition, cette méthode consiste à chercher un nouvel espace de dimension inférieur dans lequel la variance inter-classes est maximisée et la variance intra-classe minimisée.

Le lecteur intéressé pourra se référencer au livre d'Isabelle Guyon [30] et l'étude comparative de [96], donnant une description exhaustive des différentes méthodes de réduction de dimension.

4.3.5 Classification

Une fois les caractéristiques extraites, le meilleur sous-ensemble sélectionné et éventuellement représenté dans un espace de dimension inférieur, un classifieur va attribuer aux données une classe ou une probabilité d'appartenance à une classe. On distingue principalement deux grandes catégories de classifieurs, les supervisés et les non-supervisés.

La classification non-supervisée, aussi appelée *clustering*, vise à regrouper, sans a priori, des ensembles de données en sous-groupes distincts représentatifs d'une classe. Souvent, le nombre de classes est un paramètre d'entrée de l'algorithme.

La classification supervisée d'autre part vise à attribuer une classe à un ensemble de données à partir d'informations a priori. Par exemple, on cherchera à tirer profit d'une vérité terrain acquise préalablement et qui va rentrer en compte pendant la phase d'apprentissage de l'algorithme. Parmi les méthodes supervisées, on peut distinguer plusieurs sous-familles dont principalement : *les modèles génératifs* - visant à estimer les distributions de probabilités a priori à l'aide des exemples d'apprentissage (classifieurs de Bayes, classifieur naïf de Bayes, etc.) - et *les modèles discriminants* - qui estiment directement la fonction de décision (SVM, régression logistique, réseau de neurones, Analyse discriminante linéaire (LDA), arbres de décisions et forêts aléatoires, etc.). Toutes ces méthodes peuvent produire des frontières de décision linéaires ou non linéaires.

Les paragraphes suivant dressent un bref état de l'art des classifieurs utilisés pour l'application des CAD notamment [49]. Le lecteur intéressé pourra se référer aux deux livres [9, 29] contenant une description exhaustive et complète de la plupart des classifieurs utilisés dans beaucoup de domaines d'application.

Méthodes de clustering : parmi celles-ci on peut citer les *k plus proches voisins* (*knn*), un des classifieurs les plus simples de la catégorie des méthodes supervisées. Pour déterminer la classe d'un nouvel exemple, on prendra en compte les k échantillons d'apprentissage les plus "proches" de celui-ci et on lui affectera la classe correspondant à la classe majoritaire de ses points voisins. On peut de même citer l'algorithme non-supervisé du *k-means*, dont le principe repose sur le partitionnement itératif des données en k clusters. Ainsi la classe d'un nouvel exemple est celle correspondant au cluster dont son centre est le plus proche.

Classifieurs linéaires : l'Analyse discriminante linéaire (LDA) peut aussi être utilisée comme un classifieur linéaire (en plus d'être une méthode de réduction de dimension) dont la séparation linéaire optimale entre deux classes est déterminée

en maximisant la variance interclasses et en minimisant la variance intraclasse. De même, le classifieur Régression logistique (RL) (cf. section 6.6.3) peut être utilisée lors d'une classification binaire et fourni en sortie la probabilité qu'une observation appartienne à une classe.

Classifieurs non linéaires : l'analyse quadratique linéaire (LQA) fait partie de la classe des méthodes non linéaires. Contrairement au LDA qui fait l'hypothèse que la matrice de covariance Σ est identique pour toutes les classes, ici une matrice de covariance Σ est calculée pour chaque classe.

Classifieurs probabilistes : le plus célèbre des classifieurs probabilistes est le classifieur naïf bayésien dont l'hypothèse repose sur l'indépendance de chaque dimension de l'espace des caractéristiques. Les probabilités a posteriori du modèle sont souvent estimées à l'aide d'une loi normale.

Classifieurs à méthodes ensemblistes : Ces méthodes reposent sur un partitionnement récursif de l'espace des caractéristiques. Le classifieur le plus célèbre de cette catégorie est *adaboost* dont le principe est de combiner linéairement des classifieurs faibles (*weak learners*) afin d'en construire un plus performant. Un classifieur faible est défini par des performances légèrement meilleures qu'un classifieur donnant sa prédiction aléatoirement. On peut de même citer le classifieur *forêt aléatoire* (*random forest*) qui utilise une combinaison d'arbres de décision pour construire une décision finale. L'inconvénient de cette méthode est qu'elle dépend largement de la graine utilisée et du nombre d'arbres, fixé a priori (ou par grid search).

Méthode à noyaux : ces méthodes utilisent une distribution (la plupart du temps gaussienne) multivariée pour représenter n'importe quel point d'apprentissage.

Méthode à noyaux parcimonieux : Ces méthodes sont basées sur la minimisation du risque empirique (coût de mauvaise classification). Celles-ci attribuent une classe à un échantillon sur la simple observation de quelques points d'apprentissage. Le plus célèbre des classifieurs appartenant à cette classe est le SVM (cf. section 6.6.1) dont l'objectif est de trouver un hyperplan séparateur linéaire, permettant de séparer deux classes, tel que la marge entre les deux soit maximisée. Le SVM peut aussi être utilisé en tant que classifieur non-linéaire par le biais de l'astuce du noyau (*kernel trick*) consistant à projeter les données d'entrée dans un espace de dimension supérieur afin de trouver une séparatrice optimale. Des noyaux gaussien, RBF ou polynomiaux peuvent être utilisés pour séparer les données dans l'espace de dimension supérieur. On peut aussi citer le classifieur Machine à vecteur pertinent (RVM), dérivé d'un process gaussien avec une fonction de covariance spécifique.

Réseaux de neurones : Basés également sur la minimisation du risque empirique, les réseaux de neurones reposent sur l'activation de fonctions élémentaires appelées neurones pour construire une fonction d'activation plus complexes formée à partir de ces neurones élémentaires. Ce classifieur est un des plus performants mais souffre

de difficultés de mise en place à cause de leur grand nombre de paramètres et d'un temps d'apprentissage très important.

Classifieurs à modèles graphiques : les champs aléatoires de Markov (MRF : Markov Random Field) reposent aussi sur un partitionnement récursif de l'espace des caractéristiques et sont particulièrement utilisés pour des tâches de segmentation de lésions par exemple.

4.3.6 Evaluation des performances en classification

Bien que difficilement comparables les uns aux autres (cf. section 4.5), les CAD sont évalués selon une méthodologie qui définit la métrique utilisée pour rendre compte des performances ainsi que la manière dont la base d'images va être partitionnée pour l'apprentissage et le test.

Vrais/faux positifs (VP/FP) et vrais/faux négatifs (VN/FN)

Dans le cadre d'une classification binaire dont la sortie prend les valeurs $y_i = \{+1; -1\} = \{\text{cancer}/\text{non cancer}\}$, on mesure la capacité du système de classification à prédire ou non la présence d'un cancer à l'aide de quatre indicateurs :

- Le vrai positif (VP) indique la détection d'un cancer par le classifieur alors qu'il s'agit effectivement d'un cancer,
- le faux positif (FP) indique la détection d'un cancer qui n'en est en réalité pas un,
- le vrai négatif (VN) indique la non détection d'un cancer par le classifieur alors qu'il s'agit effectivement d'un tissu non cancer,
- et enfin le faux négatif (FN) indiquant la classification en non cancer alors qu'il s'agit d'un cancer.

Sensibilité/Spécificité

A l'aide des indicateurs VP, FP, VN et FN précédents, on définit deux métriques : la sensibilité et la spécificité du classifieur. La sensibilité (Se) définie par $Se = \frac{VP}{VP+FN}$, caractérise la proportion de cas correctement étiquetés comme pathologique sur le nombre total de tissus pathologiques. La spécificité définie par $Spe = \frac{VN}{VN+FP}$, caractérise la proportion de cas correctement étiquetés comme non pathologiques sur le nombre total de tissus non pathologiques.

Il est à noter que ces deux métriques sont complémentaires et dépendent d'un seuil réglant le compromis entre sensibilité et spécificité. En effet, un classifieur très sensible peut avoir tendance à tout étiqueter comme cancer au détriment de la spécificité et vice-versa. C'est ce seuil qu'il est important de bien régler pour trouver le meilleur compromis sensibilité/spécificité répondant aux exigences d'un problème. Un couple (Se, Spe) se limite donc à comparer les performances pour un niveau de confiance ou seuil donné.

Courbe ROC : Receiver Operating Characteristic

La courbe ROC, appelée "Receiver Operating Characteristic" ou caractéristique de fonctionnement du récepteur est une mesure de la performance d'un classifieur binaire. La courbe ROC se construit en 2 dimensions avec la sensibilité sur l'axe des ordonnées et 1-spécificité sur l'axe des abscisses. Elle permet de synthétiser graphiquement ses performances, c'est-à-dire l'ensemble des compromis entre sensibilité et spécificité, pour différents niveaux de certitudes (seuils). En effet, la sortie d'un classifieur binaire est majoritairement sous forme continue représentant soit un score de probabilité d'appartenance à une classe ou soit une distance entre deux classes. Dans le cas du SVM par exemple (cf. section 6.6.1), le classifieur retourne pour chaque point à classer, sa distance à l'hyperplan séparateur. La classe du point est donc donnée en fonction de sa position par rapport à l'hyperplan. Ainsi, faire varier les différents niveaux de certitude du classifieur revient à décaler l'hyperplan séparateur et pour chacune de ses positions, calculer sensibilité et spécificité et reporter le point sur la courbe. De ce tracé, on peut tirer une mesure synthétique de performance, appelée Aire sous la courbe ROC (AUC), représentant l'aire sous cette courbe ROC. Celle-ci, comprise entre 0 et 1, représente la probabilité qu'un exemple positif ait un score supérieur à un exemple négatif. Ainsi une AUC de 0.5 représente un classifieur dont la sortie peut s'apparenter à une prédiction aléatoire. Le classifieur parfait est donc celui obtenant un score d'AUC de 1, c'est-à-dire celui dont on a réussi à trouver un seuil (ou position de l'hyperplan) permettant de classer correctement tous les exemples. Intuitivement le point optimal est celui le plus proche du coin supérieur gauche, donnant le meilleur compromis entre sensibilité et spécificité.

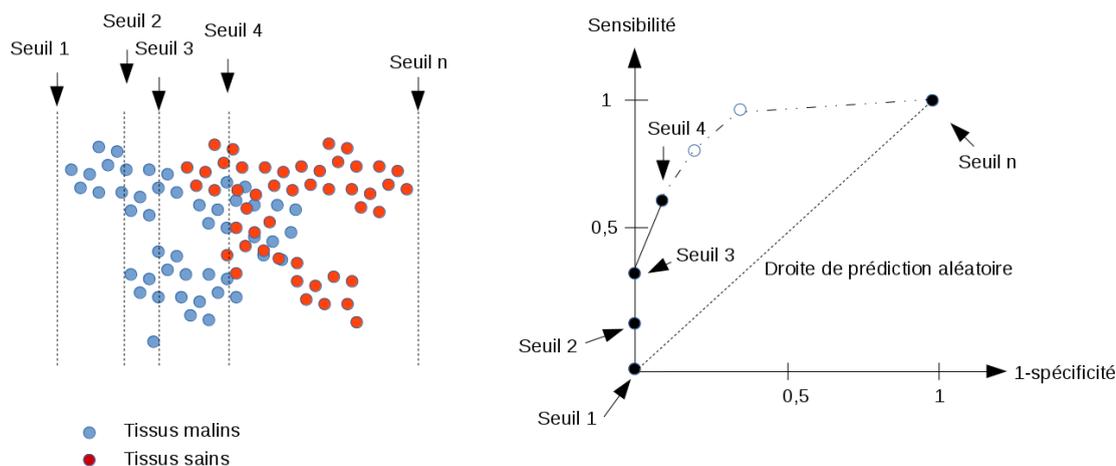


FIGURE 4.3 – Principe de construction d'une courbe ROC

La figure 4.3 représente un exemple de sortie d'un classifieur binaire (séparation tissus sains / tissus malins), le SVM par exemple. L'ensemble des seuils ou positions de l'hyperplan séparateur est représenté par les lignes discontinues notées seuil 1 à seuil n . Pour chaque valeur de seuil on compte le nombre vrais positifs (VP), vrais négatifs (VN), faux positifs (FP) et faux négatifs (FN) afin de calculer sensibilité et spécificité. Ces seuils sont ensuite

reportés sur la courbe ROC et illustrés sur la courbe $Se = f(Spe)$.

Courbe FROC : Free-Response Receiver Operating Characteristic

La courbe ROC peut être qualifiée de méthode d'analyse pixel à pixel. Bien qu'elle puisse être utilisée lorsque l'on est en présence de régions, il est parfois plus judicieux d'utiliser une courbe FROC, adaptée à l'évaluation des performances d'un classifieur basé sur la classification de régions (agrégat de pixels). La courbe FROC est obtenue en remplaçant l'axe des abscisses (1-specificité) par le nombre de faux positifs par image ou patient. Dans ce cas, la définition d'une région détectée est requise et l'approche typique est de considérer une région comme détectée lorsqu'elle a au moins 50% de recouvrement avec la vérité terrain. L'avantage de cette méthode est qu'elle donne un bon aperçu des performances d'un système en fonction du nombre de fausses détections que l'on s'accorde à faire. Le principal inconvénient de cette analyse est qu'elle nécessite d'introduire un jugement subjectif.

Indice de dice

En complément d'une mesure de recouvrement entre vérité terrain et région détectée, l'indice de Dice est souvent calculé pour évaluer la précision de localisation de la lésion. Ce ratio consiste en un ratio entre deux fois le nombre de pixels en commun divisé par la somme des pixels dans la lésion (A) et dans la zone détectée (B) par le classifieur. Il est défini par :

$$DSC(A, B) = \frac{2 | A \cap B |}{| A | + | B |} \quad (4.0)$$

où $| \cdot |$ désigne l'aire de la surface (nombre de voxels).

Moyenne géométrique

La moyenne géométrique, noté $Gmean$ et définie par $Gmean = \sqrt{Se \times Spe}$ est une métrique utilisée principalement lorsque l'on est en présence d'une base de données déséquilibrée. Elle permet en effet de mesurer l'équilibre entre sensibilité et spécificité pour chaque point de la courbe ROC. En présence d'un jeu de données déséquilibré, le point optimal obtenu par analyse de la courbe ROC (point le plus proche du coin supérieur gauche) peut ne pas être le meilleur compromis. [65]

4.3.7 Validation d'un modèle d'apprentissage

La validation d'un modèle d'apprentissage joue un rôle important pour évaluer les performances d'un CAD. En effet, il existe plusieurs méthodes ayant chacune leurs avantages et inconvénients mais leur choix va surtout s'orienter en fonction de la base de départ, notamment du nombre d'exemples en notre possession pour valider le modèle.

Validation croisée k -fold

La validation croisée (k -fold) consiste à diviser la base d'origine en k sous-ensembles où les échantillons sont choisis aléatoirement et sans remise. Ensuite, un sous ensemble est choisi pour validation, les sous-ensembles restant pour l'apprentissage. La classification est ensuite répétée jusqu'à ce que tous les sous-ensembles aient été utilisés pour la validation. Le résultat final est donc la moyenne des résultats sur chacun des sous-ensembles de validation. Dans la littérature la valeur de k choisie est généralement de 3 à 5. Cette méthode de validation est considérée comme la plus adaptée et la moins biaisée mais nécessite un nombre de patients assez large, ce qui n'est généralement pas le cas dans les applications médicales.

Leave-One-Out (LOO) ou Leave-One-Patient-Out (LOPO)

De par l'inconvénient du nombre limité de patients dans la base, on adopte plus généralement une stratégie de type Validation croisée par exclusion d'un patient (LOPO). Le LOPO est un cas particulier du k -fold lorsque k est égal au nombre de patients (ou égal au nombre d'exemples d'apprentissage pour un Validation croisée par exclusion d'un élément (LOO)). Le principe est donc de ne garder qu'un patient pour le test et d'apprendre sur le reste en répétant l'opération jusqu'à ce que tous les patients aient été utilisés pour la validation. Cette stratégie permet donc d'entraîner le classifieur sur un nombre représentatif de données même en présence d'une base de taille réduite. Cependant les LOO et LOPO souffrent de large variance et sont considérés comme biaisés. Néanmoins, plusieurs études proposent de corriger ce biais qui sous-estimerait néanmoins les performances réelles d'un classifieur [91, 97].

Re-substitution

Cette méthode utilise toutes les données disponibles pour l'apprentissage et teste le modèle du classifieur sur chacune de ces mêmes données (les ensembles d'apprentissage et de test ne sont pas disjoints). Cette méthode, simple à mettre en œuvre, est souvent critiquée parce qu'elle ne prend pas en considération le comportement de l'algorithme lorsqu'un cas de figure inconnu est rencontré. De plus, cette méthode donne une estimation biaisée positivement des performances en généralisation (ie. mesurée si le test se faisait sur des données de taille infinie).

Bootstrap

La validation par méthode de Bootstrap est généralement utilisée lorsque la base d'exemple est trop restreinte. Le principe est de répliquer les données disponibles issues d'une base d'origine pour créer plusieurs autres bases d'apprentissage de taille N . La base de test est toujours la base d'origine mais le score final est évalué en moyennant des résultats obtenus pour chaque nouvelle base de taille N . Cette méthode permet notamment d'estimer le biais

positif introduit par la re-substitution et de s'approcher ainsi davantage des performances en généralisation.

4.4 Etat de l'art des CAD pour le cancer de la prostate

Comme indiqué en introduction, il existe deux types de CAD : (1) les CADe où tous les voxels sont considérés comme de possibles lésions et la sortie est une classification par voxel sous forme de carte de probabilité de cancer et (2) les CADx dont l'approche est la classification de régions, sous-entendu un agrégat de voxels obtenu manuellement ou automatiquement, et la sortie un score de malignité pour chaque région suspecte. De plus, les CADe sont parfois une première étape de sélection de voxels et de zones suspectes afin d'alimenter un CADx par la suite. La partie suivante dresse un état de l'art récent (postérieure à l'année 2012) des différents CADe similaires à nos travaux. Nous détaillons, pour chaque étude, la description de la tâche effectuée par le CAD (détection, localisation, évaluation de l'agressivité), la base d'image utilisée, le protocole de constitution de la vérité terrain, les grandes étapes méthodologiques ainsi qu'une analyse des résultats. La dernière partie propose un tableau de synthèse des CADx, CADe et combinaison CADx/CADe basés sur des images IRM multiparamétriques. Un état de l'art plus ancien peut être consulté dans la thèse d'*Emilie Niaf* [64]). De même, l'article de synthèse de *Lemaître et al.* [49] présente un état de l'art très complet des 10 dernières années ainsi qu'une présentation détaillée des différentes étapes de traitement des données lors de la conception d'un CAD.

4.4.1 Etat de l'art en détection, localisation et évaluation de l'agressivité du cancer de la prostate

Dans leur étude de 2012, *Viswanath et al.* [99] ont implémenté un CAD basé sur des IRM 3T T2w, ADC et DCE de prostates in-vivo de 12 patients. La vérité terrain est obtenue après prostatectomie et corrélation entre l'histologie et l'IRM (cf. section 5 pour les détails sur la prostatectomie). Les séquences IRM sont dans un premier temps recalées puis un ensemble de caractéristiques statistiques, fonctionnelles ainsi que des informations de texture sont extraites des 3 séquences. Le coeur de leur travail repose sur une méthode de réduction de dimension capable de tirer de l'information non corrélée issue des multiples caractéristiques. Cette étude a aussi pour vocation de fournir une méthode générale de réduction de dimension quelque soit l'application ou la modalité d'acquisition de la séquence IRM. Une fois les données réduites, ils utilisent un classifieur de type *arbre probabiliste* dans le but de détecter le cancer de la prostate au niveau voxelique suivi d'un champ de Markov aléatoire pour régulariser spatialement les résultats. Ils obtiennent un score d'AUC=0.77±0.16, significativement meilleur que sans leur méthode de réduction de dimension (AUC=0.67±0.21) ou en utilisant individuellement les caractéristiques de chacune des séquences IRM (AUC=0.62±0.22, 0.62±0.14 et 0.65±0.21 pour T2w, DCE et

ADC respectivement).

Dans leur étude publiée en 2012, *Vos et al.* [101] ont implémenté un système CAD automatisé à plusieurs étapes sur une base de 200 patients, dont la majorité a subi une prostatectomie radicale avec corrélation IRM-histologie en gold standard. D'autres patients ont néanmoins été ajoutés sans ce gold standard. Chaque donnée patient est constitué de 4 séquences IRM 3T : T2w, PDW, ADC et DCE. La première repose sur la détection de zones suspectes (blob detection) en exploitant la carte ADC uniquement. En effet, les bonnes performances de l'ADC en sensibilité en fait un outil intéressant pour détecter l'ensemble des lésions suspectes. Ces dernières sont triées en utilisant des critères a priori tels que le diamètre ou l'intensité moyenne du signal dans la région (calculée à la fois sur l'ADC et le T2w). Une deuxième étape de segmentation permet de restreindre la sélection des zones suspectes à la zone de la prostate puis une analyse d'histogramme permet ensuite d'en extraire des caractéristiques. Enfin une étape de classification supervisée à deux étages, basées sur un classifieur LDA, permet de fournir un score final. Leurs performances lors de la première étape de détection basée sur l'ADC atteignent 92% de sensibilité. L'étape de classification est ensuite évaluée en sensibilité par l'utilisant d'une courbe FROC avec des seuils de 1, 3 et 5 faux positifs par patient. En fonction du volume de la tumeur, un grade lui est affecté et les scores déclinés en fonction de celui-ci. Ils obtiennent des sensibilités de l'ordre de 0.48, 0.73 et 0.88 pour 1, 3, 5 FP respectivement pour les tumeurs de haut grade, 0.41, 0.65 et 0.74 pour tous grades confondus et 0.27, 0.40, 0.68 pour les tumeurs de bas grade. Ils concluent leur article sur le fort potentiel de leur méthode à détecter les tumeurs de haut grade (0.88), et ainsi assister les radiologistes dans la détection et le guidage des biopsies qui s'annonce comme une amélioration en comparaison de la biopsie systématique.

En 2013, *Tiwari et al.* [92] ont implémenté un CAD dont le double objectif est de différencier régions bénignes des cancers puis de discriminer cancers agressifs des cancers non agressifs. L'étude est réalisée sur 29 patients ayant subi une prostatectomie radical. Un ensemble de caractéristiques structurelles et métaboliques est extrait des images IRM 1.5T T2w et spectroscopie des 29 patients puis une méthode de réduction de dimension est appliquée avant une phase de classification à deux étapes. La première étape, destinée à identifier les régions suspectes, utilise un classifieur de type forêt aléatoire couplé à un algorithme basé sur des chaînes de Markov aléatoires afin de contraindre spatialement les données. Ces régions suspectes sont ensuite directement classifiées en cancer de haut ou bas grade. La méthode, appelée SeSMIK-GE, est évaluée en comparaison d'une classification sur les séquences IRM prises individuellement mais aussi en comparaison d'une simple concaténation directe des caractéristiques des 2 séquences IRM. Les score obtenus sont 0.89(SeSMIK-GE), 0.54(T2w), 0.61(spectroscopie) et 0.64(concaténation directe) pour la discrimination des régions bénignes des cancers. Pour la discrimination des cancers de haut grade des cancers de bas grade, les score obtenus sont 0.84(SeSMIK-GE), 0.54(T2w), 0.59(spectroscopie) et 0.62(concaténation directe).

Dans leur étude, *Litjens et al.* [51] ont implémenté un CAD automatisé à deux étages

(combinaison d'un CADe suivi d'un CADx) et l'ont testé sur sa capacité à différencier tout grade de tumeur des zones saines ainsi que les tumeurs de haut grade (seconde partie du score de gleason > 3) des autres zones toutes confondues (ROI saines et tumeurs dont la seconde partie du score de gleason < 3). La base d'images IRM, dont les séquences sont le T2w, l'ADC le DCE le PD-W (proton density-weighted), est constituée de 347 patients dont la vérité terrain est confirmée par biopsie uniquement. Dans leur premier étage de détection de zones suspectes, ils extraient des caractéristiques de type intensité, anatomique, pharmacocinétique, texture et forme pour être classées à l'aide d'un classifieur de type arbre de forêts aléatoires. La sélection des zones suspectes se fait via un détecteur de maximums locaux. La seconde étape de classification des zones suspectes fait intervenir l'extraction de nouvelles caractéristiques de type statistiques, contraste local, forme et symétrie associés à un classifieur LDA ou arbre de forêts aléatoires. Dans la tâche de discrimination du cancer de la prostate des zones saines, ils obtiennent un score d'AUC de 0.81. Avec une métrique d'évaluation FROC pour un seuil de 1 FP par image, le CAD obtient un score en sensibilité de 0.75 pour la détection de tout type de lésion et 0.83 pour la détection de tumeurs de haut grade. Ils comparent ensuite les performances du CAD à ceux d'un radiologue expérimenté et concluent sur des performances similaires à haute spécificité (0.88) alors que pour de plus basses spécificités le radiologue obtient de biens meilleurs performances. De plus l'ajout de deux étapes de sélection de voxels et de candidats améliore les performances de classification par lésion (de 0.55 à 0.75 de sensibilité au seuil de 1 FP par patient) et par patient (de 0.72 à 0.81 en AUC). Les auteurs concluent cette étude sur les possibles applications de ce CAD notamment en tant que second observateur puisqu'ils obtiennent les mêmes performances à haute spécificité, ce qui est typiquement le cas puisque la plupart des sujets sont sains et ne pas générer trop de faux positifs et donc de sur-diagnostic est un objectif désormais. La deuxième application du CAD perçue par les auteurs est son utilisation en tant que filtre réglé à très haute sensibilité ($Se = 0.96$) pour ne rater que très peu de cas pathologiques. A ce seuil de sensibilité le CAD présente une spécificité comprise entre 0.15 et 0.40 donc potentiellement 15 à 40% des cas ne nécessiteraient pas d'intervention humaine et donc réduirait la charge de travail.

Dans leur étude, *Garcia-Molina et al.* [25] proposent un CAD basé sur un apprentissage dit "incrémental", c'est-à-dire qu'à l'ajout de nouvelles données, le système n'a pas besoin de réapprendre un modèle complet. Cette solution permet d'éviter ce qu'ils appellent le phénomène de "désapprentissage" qui intervient lorsque l'on doit refaire tourner un modèle complet lorsque de nouvelles arrivent et ainsi potentiellement "perdre" les informations précédemment apprises. Leur étude est testé sur une base de 12 patients, avec chacun 3 séquences IRM 3T de type T2w, DCE plasma flow et DCE mean tansit time. La vérité est obtenue par corrélation IRM-histopathologie. Les caractéristiques extraites des images sont de type structurelles, fonctionnelles et statistiques. Le classifieur utilisé est un ensemble SVM, basé sur un one-class SVM intégré dans une stratégie de recherche de sous-espace optimal de points et de caractéristiques. Ils obtiennent un score d'AUC de 0.87 dans la

tâche de discrimination du cancer contre les tissus sains. Évalué par rapport à des radiologues expert, leur système obtient de moins bonnes performances en Valeur prédictive positive (VPP) (0.79 versus 0.42) mais de meilleures performances en sensibilité (0.69 versus 0.49). Ils concluent sur l'intérêt de leur méthode en mettant en avant ces performances en sensibilité pouvant considérablement faciliter l'analyse multiparamétrique visuelle des 3 séquences pour la détection des zones suspectes.

Lors de nos précédents travaux, *Niaf et al.* [64] ont implémenté un CADx dont le but est la discrimination des tumeurs malignes des zones saines. La base utilisée est composée de 30 patients avec images IRM 1.5T de type T2w, ADC et DCE dont la vérité terrain est obtenue après corrélation IRM-histologie. Les caractéristiques extraites sont de type structurelles, statistiques et fonctionnelles et les classifieurs testés le k -plus-proches voisins (k -PPV), SVM non-linéaire, LDA et le classifieur naïf bayésien. L'objectif était d'obtenir le meilleur sous-ensemble de caractéristique associé à un des classifieurs suivant une méthodologie de sélection de caractéristiques basée sur le t-test, l'information mutuelle, les critères de maximum de redondance et maximum de pertinence. Le meilleur schéma obtenu est composé du SVM linéaire combiné aux caractéristiques sélectionnées par le t-test. Les performances obtenues en AUC sont de 0.89(0.81-0.94) pour la discrimination des cancers versus tissus non-malins et 0.82(0.73-0.90) pour la discrimination des cancers versus tissus suspects. Une comparaison avec les performances de deux radiologues (>15 ans et >2 ans d'expérience) indique que le CAD obtient des résultats similaires au radiologue le moins expérimenté et pourrait faire fonction d'assistant en second avis.

A la suite de ces premiers travaux, *Niaf et al.* [63] ont étudié l'influence du CAD sur la décision finale du radiologue. Axé sur la même base IRM 1.5T présentée précédemment mais sur 49 patients, l'étude compare les performances de 12 radiologues, scindés en deux groupes inexpérimentés et expérimentés, après deux sessions de lecture et d'annotation des images avec une troisième session de lecture aidée par le CAD. L'annotation consistait à attribuer un score, compris entre 0 et 4, aux ROIs (saines, suspectes et malignes) sélectionnées pour l'étude. Lors de la première session, les scores d'AUC obtenus sont de 0.83 en moyenne pour les radiologues junior et sénior. Lors de la session 2, les scores obtenus sont en moyenne de 0.83 et 0.81 (radiologues sénior et junior respectivement) et enfin lors de la session 3, les scores obtenus sont 0.87 et 0.85 respectivement. Bien que les différences d'AUC entre chaque session ne soit pas statistiquement significatives, cette étude a clairement montré le potentiel du CAD en second avis de radiologues inexpérimentés.

4.4.2 Tableau de synthèse des CAD

Le tableau 4.1 dresse un état de l'art non-exhaustif des principales études sur les CAD de la prostate. Pour chaque étude, il est détaillé les séquences IRM-mp utilisées, la taille de la base (en nombre de patients), le champ B_0 , les séquences IRM utilisées, la méthode permettant d'obtenir la vérité terrain, le type de CAD (CADe ou CADx), la zone de la prostate sujet à la classification, le type de classifieur utilisé et les résultats.

Publication	nb patients	Séquences IRM	Champ B_0	Vérité terrain	Type de CAD	Région	Classifieur	Performance
<i>Artan et al.</i> 2010 [2]	21	T2w, ADC, DCE	1.5T	Prostatectomie	CADe	ZP	SVM,CRF	AUC=0.79±0.12
<i>Chan et al.</i> 2003 [16]	15	T2w, ADC, PDW, T2 map	1.5T	Biopsie	CADe	ZP	FLD	AUC=0.839±0.064
<i>Garcia-Molina et al.</i> 2014 [25]	12	T2w, DCE-PF, DCE-TT	3T	Prostatectomie	CADe	ZP	Ensemble SVM	AUC=0.864±0.045
<i>Khalvati et al.</i> 2015 [44]	20	T2w, ADC, CDI	3T	Prostatectomie	CADe+CADx	PT	SVM	AUC=0.90, IC [0.88 0.93]
<i>Litjens et al.</i> 2014 [51]	347	T2w, PDW, DCE, ADC	3T	Biopsie	CADe+CADx	PT	Random forest	AUC=0.89
<i>Liu et al.</i> 2011 [53]	20	T2w, ADC, DCE	1.5T	Biopsie	CADe	ZP	SVM	AUC=0.89
<i>Liu et al.</i> 2013 [52]	54	T2w, ADC, DCE	3T	Biopsie	CADx	PT	SVM	AUC=0.82, IC [0.71 0.93]
<i>Moradi et al.</i> 2012 [60]	29	DT, DCE	3T	Biopsie	CADx	PT	SVM	AUC=0.96
<i>Niaf et al.</i> 2012 [64]	30	T2w, ADC, DCE	1.5T	prostatectomie	CADx	ZP	SVM	AUC=0.89,IC [0.81 0.94]
<i>Peng et al.</i> 2013 [70]	48	T2w, ADC, DCE	3T	Prostatectomie	CADx	PT	LDA	AUC=0.95±0.02
<i>Puech et al.</i> 2009 [75]	84	DCE	1.5T	Biopsie	CADx	ZP	arbre de décision	AUC=0.77
<i>Shah et al.</i> 2012 [84]	31	T2w, ADC, DCE	3T	Prostatectomie	CADe	ZP	SVM	F-measure=0.89
<i>Tiwari et al.</i> 2012 [93]	36	T2w, MRS	1.5T	Prostatectomie	CADe	PT	Radom forest	AUC=0.89±0.02
<i>Tiwari et al.</i> 2013 [92]	29	T2w, MRS	1.5T	Prostatectomie	CADe	PT	SeSMiK-GE+random forest	AUC=0.89±0.09
<i>Viswanath2011</i> 2011 [99]	12	T2w, DCE	3T	Biopsie	CADe+CADx	PT	random forest	AUC=0.815±0.16
<i>Vos et al.</i> 2010 [102]	34	T2w, DCE	1.5T	Prostatectomie	CADx	ZP	SVM	AUC=0.89, IC [0.81 0.95]
<i>Vos et al.</i> 2012 [101]	200	T2w, ADC, PDW, DCE	3T	Biopsie	CADe+CADx	ZP	LDA	FROC=0.88 à FP=5

TABLE 4.1 – Comparaison des études sur les CAD du cancer de la prostate. Les acronymes sont PDW : Proton density weighted, CDI : correlated diffusion imaging (CDI) [105], PT : prostate totale, CRF : conditional random field, DCE-TT : dynamic ontrast enhanced time-transit, DCE-PF : dynamic ontrast enhanced plasma flow, DT : tenseur de diffusion, MRS : spectroscopie

4.5 Discussion sur les performances des CAD

Une comparaison difficile. Malgré une méthodologie de développement plus ou moins standardisée, on se rend compte qu'il est difficile de comparer les performances de deux CAD. En effet, d'une étude à l'autre il existe une forte variabilité sur les séquences IRM utilisées, sur leur protocole d'acquisition, sur le champ B_0 utilisé (impact sur la résolution de l'image) ainsi que sur la constitution de la vérité terrain. Cette dernière, si obtenue via prostatectomie radicale suivie d'une corrélation entre l'histologie et l'IRM, aura forcément plus de poids que celle obtenue uniquement par une biopsie ne permettant pas d'accéder à une information précise sur le volume et la localisation de la tumeur. De plus, la taille de la base aura aussi un impact notable puisqu'une base d'une dizaine de patients sera beaucoup moins significative statistiquement qu'une base de 50 patients. Enfin, les méthodes de validation des CAD ainsi que les métriques de performances sont parfois différentes et rendent la comparaison une fois de plus difficile.

Utilisations potentielles des CAD. A travers cet état de l'art, on peut se rendre compte de l'application potentielle des CAD. Dans un premier temps, le type de CAD renseigne si l'objectif est la détection de zones suspectes analysées par un radiologue (CADe), ou bien l'obtention d'un score associé à une région suspecte (CADx). Certains auteurs tentent d'implémenter des solutions quasi automatiques (CADe+CADx) allant de la phase de détection des zones suspectes à leur classification. On notera de même que la plupart des études tente de discriminer les zones pathologiques des tissus sains. Néanmoins, depuis peu, des études tentent d'incorporer la notion d'agressivité du cancer dans leur tâche de discrimination et comparent leurs résultats à une tâche classique cancer versus tissus sains. Réserver un traitement aux patients atteints de cancer agressif est aussi un enjeu majeur dans la réduction du sur-diagnostic. Enfin, il est aussi à noter que certaines études évaluent les performances du CAD en second avis ou face à un radiologue. On remarque que ses performances ne sont jamais supérieures à celles d'un radiologue expérimenté. Bien que le but ultime serait de pouvoir détecter, localiser et évaluer précisément l'agressivité d'un cancer, les conclusions sur son apport potentiel sont plutôt orientées à ce jour vers une utilisation en second avis notamment pour améliorer les performances du radiologue inexpérimenté. Sa deuxième utilisation potentielle est celle d'un outil de pré-analyse, réglé sur un seuil de sensibilité très haut (de manière à détecter tous les cancers) tout en éliminant quelques zones suspectes. Cette dernière utilisation pourrait permettre un gain de temps dans l'analyse des images.

II Développement d'un CADe pour le diagnostic du cancer de la prostate par analyse d'images IRM multiparamétriques

Constitution de la base d'images clinique IRM 1.5T

5.1 Introduction

En vue de la construction d'un système d'aide au diagnostic pour le cancer de la prostate, une base d'images IRM 1.5T a été initiée en 2008 par le professeur Olivier Rouvière dans le cadre du projet Corrélations anatomo-radiologiques en IRM de prostate (CLARA-P). La constitution de cette base de 49 patients est soumise à un protocole standardisé. En effet, chaque patient dont le cancer est avéré et confirmé par biopsie subit systématiquement, avant prostatectomie radicale, un examen IRM permettant d'acquérir les 3 séquences IRM T2-w, ADC et DCE (cf. section 3.2). Une première analyse en aveugle est effectuée et les zones suspectes sont contourées sur les séquences IRM. A la suite de la prostatectomie, chaque pièce histologique est analysée et marquée par un anatomopathologiste. Une deuxième lecture a posteriori mettant en regard les pièces histologiques et la première analyse en aveugle, permet ensuite d'établir une vérité terrain constituée de la localisation précise des tissus sains, suspects et des foyers cancéreux dont leur agressivité est caractérisée par leur score de Gleason. Historiquement, c'est à partir de la base présentée dans cette section que la thèse d'Emilie Niaf [62] a abouti au développement d'un Système d'aide au diagnostic (CADx), basé sur une approche région, permettant la discrimination du cancer versus tissus sains ou suspects. Dans le cadre de cette nouvelle thèse, nous avons donc repris ce travail et cette même base pour le développement d'un Système d'aide à la détection (CADE), dont l'objectif est de détecter, localiser et évaluer l'agressivité des

cancers. La description du protocole d'acquisition et d'annotation des images est tirée de la thèse d'Emilie Niaf [62].

5.2 Acquisition des images

5.2.1 Déroulement de l'examen

Tous les examens IRM de la prostate sont exécutés selon un protocole standardisé, sur deux scanner clinique à 1.5 T (Siemens Magnetom Symphony MR 2004A, Siemens Medical Systems, Erlangen, Allemagne) et 3T (General Electric) en utilisant une antenne pelvienne multi-canaux .

Les paramètres des séquences sont ceux utilisés en routine clinique à l'hôpital Édouard Herriot de Lyon (France) par le Pr. Rouvière pour l'analyse post-biopsies (et dans certains cas pré-biopsies) de la prostate. Dans l'ordre, ce sont d'abord les images en T2-pondérées (T2-w) qui sont acquises, suivies des images pondérées en diffusion (DWI) et enfin les images de perfusion (DCE). Le temps total de l'examen est approximativement de 30 minutes.

5.2.2 Paramètres des séquences IRM 1.5T

Les images T2-w turbo-spin-écho sont obtenues dans les plans axial, sagittal et coronal. Les 24 coupes axiales contiguës couvrent entièrement la glande [temps de répétition (T_R) : 7750 ms ; temps d'écho (T_E) : 109 ms ; champ de vue (FOV) : 200×200 mm ; matrice : 256×256 ; nombre de coupes : 24 ; épaisseur des coupes : 3 mm, sans espacement].

Les données de la séquence DWI sont ensuite acquises [T_R : 4800 ms ; T_E : 90 ms ; FOV : 300×206 mm ; matrice : 128×88 ; nombre de coupes : 24 ; épaisseur des coupes : 3 mm, sans espacement ; angle de bascule (flip angle) : 90° ; valeurs de b : 0 and 600 s/mm²].

Enfin, une séquence DCE acquise en T1-w et écho de gradient 3-D fat-saturated est obtenue dans le plan axial avant et après injection de produit de contraste (bolus injecté en intraveineuse à une vitesse de 3 ml/s, à une dose de 0.1 mmol/kg de gadotérate méglumine _ Gd-DOTA, Dotarem, Guerbet, Roissy, France) [T_R : 5.38 ms ; T_E : 2.73 ms ; angle de bascule : 10° ; FOV : 210×240 mm ; matrice : 256×135 ; nombre de coupes : 24 ; épaisseur des coupes : 3 mm, sans espacement ; temps d'acquisition : 15 s, répétée 12 fois, pour une durée totale de 3 min].

L'épaisseur de coupe de 3 mm a été choisie de manière identique pour toutes les séquences IRM réalisées pour permettre une comparaison directe coupe à coupe entre séquences. Les paramètres des différentes séquences utilisées sont synthétisés dans le tableau 5.1. Une description détaillée des spécificités de chaque séquence est donnée au chapitre 3.2.

Séquence	T_R (ms)	T_E (ms)	Angle bascule (°)	Epaisseur coupe (mm)	Matrice (vox)	Nb. coupes	FOV (mm)	dimension voxel (mm)	temps échantillon- nage (s)
T2-w	7750	109	180	3	256×256	24	200×200	.78x.78x3	-
DCE	5.38	2.73	10	3	256×135	24	210×240	.47x.47x3	15
DWI	4800	90	90	3	128×88	24	300×206	2.34x2.34x3	-

TABLE 5.1 – Paramètres utilisés pour l'imagerie de la prostate par IRM à 1.5T

5.3 Analyse des images

Deux radiologues, sénior et junior, analysent en aveugle les images IRM des patients et contournent les régions jugées suspectes. Lorsqu'un tissu d'intérêt est visible sur plusieurs coupes consécutives (mêmes caractéristiques image), il est contourné sur chacune des coupes sur lesquelles il est visible. Dans la suite, le terme ROI se rapporte donc à l'ensemble de ces contours 2-D réalisés pour délimiter un tissu cible. Un score purement subjectif (score dit de Likert) a été utilisé puisque les critères des scores PI-RADS n'avaient pas encore été décrits quand la base prospective a été débutée, correspondant au degré de suspicion de malignité est affecté à chaque ROI mise en évidence [22] :

- 0 : bénignité certaine,
- 1 : probablement bénin,
- 2 : intermédiaire,
- 3 : probablement malin,
- 4 : malignité certaine.

5.4 Analyse histologique

Préparation des pièces de prostatectomie

Tous les patients considérés dans cette étude ont subi une prostatectomie radicale. Chaque pièce de prostatectomie est traitée selon la technique standard dite de Stanford schématisée sur la figure 5.1c et dont les étapes sont détaillées ci-dessous.

Orientation de la pièce de prostatectomie : La pièce est orientée vésicule séminales en haut et en arrière, face rectale plane, face antérieure convexe et apex en pointe vers le bas. Le lobe droit et le lobe gauche sont encrés respectivement en jaune et noir, la face antérieure en rouge (voir illustration 5.1b).

Fixation de la pièce de prostatectomie : La pièce est fixée pendant 24h au minimum dans un fixateur formolé (formol tamponné à 10).

Séparation des vésicules séminales et de la base : Les vésicules séminales (en bleu et rouge sur l'illustration 5.1c) sont coupées au niveau de la base selon un

axe horizontal et séparées de la prostate.

Conisation : Des conisations transversales des portions distales de l'apex (en rose sur l'illustration 5.1c) et du col vésical (en vert) d'épaisseur 3-5 mm sont pratiquées.

Découpe de la pièce de prostatectomie : Le reste de la glande (en violet sur l'illustration 5.1c) est ensuite coupé en totalité en tronçons de 6-mm d'épaisseur (illustration 5.1d) selon un plan perpendiculaire à la surface rectale, correspondant au plan utilisé pour l'imagerie Imagerie par résonance magnétique (IRM), à l'aide d'une machine dédiée (illustration 5.1d) réalisée au service d'histopathologie de l'hôpital Edouard Herriot.

Les tranches sont mises dans des cassettes et refixées pendant au moins 48 heures avant de subir une étape de déshydratation (dans des bains de formol et de méthyl) puis d'être inclus en paraffine.

Enfin, les blocs de paraffine obtenues sont ensuite coupés au microtome (illustration 5.1f). Les rubans obtenus sont étalés sur lames et colorés au HES (Hématoxyline-Eosine-Safran) (illustration 5.1g).

Analyse des coupes histologiques

Une anatomo-pathologiste analyse au microscope l'ensemble de la pièce histologique issue de la prostatectomie radicale. A ce stade, les résultats de l'imagerie pré-chirurgicale ne sont pas communiqués. Sur chacune des lames histologiques où elles sont visibles, toutes les zones malignes sont contourées. Les tumeurs ne sont considérées comme telles que si leur surface dans le plan est au minimum de 2 mm x 2 mm, si elles sont visibles sur au moins 2 coupes et ont un score de Gleason ≥ 5 . Les lésions malignes séparées de moins de 1 mm l'une de l'autre dans le même plan, avec la même architecture et le même score de Gleason ont été considérées comme faisant partie de la même tumeur. Les régions de tissus remarquables et leurs caractéristiques (prostatite, inflammation, PIN, etc) sont également mises en évidence.

A la fin de l'analyse, un jeu de coupes séparées de 3-mm est sélectionné parmi l'ensemble des lames, permettant ainsi une comparaison directe avec les coupes IRM pour l'analyse de la corrélation anatomo-radiologique (voir photo 5.1g).

5.5 Corrélation anatomo-radiologique

Les données issues de l'analyse en aveugle des images IRM et celles issues de la lecture des lames histologiques sont revues lors d'une troisième séance de lecture par les deux radiologues et l'anatomo-pathologiste travaillant en consensus. L'analyse de la pièce de prostatectomie, utilisée comme "vérité terrain", permet la délimitation *a posteriori*, sur les images IRM (voir la figure 5.2), des zones de :

- tissu malin (Vrais positifs (VP) ou Faux négatifs (FN) de l'analyse radiologique), noté **{M}** pour tissu **Malin**,

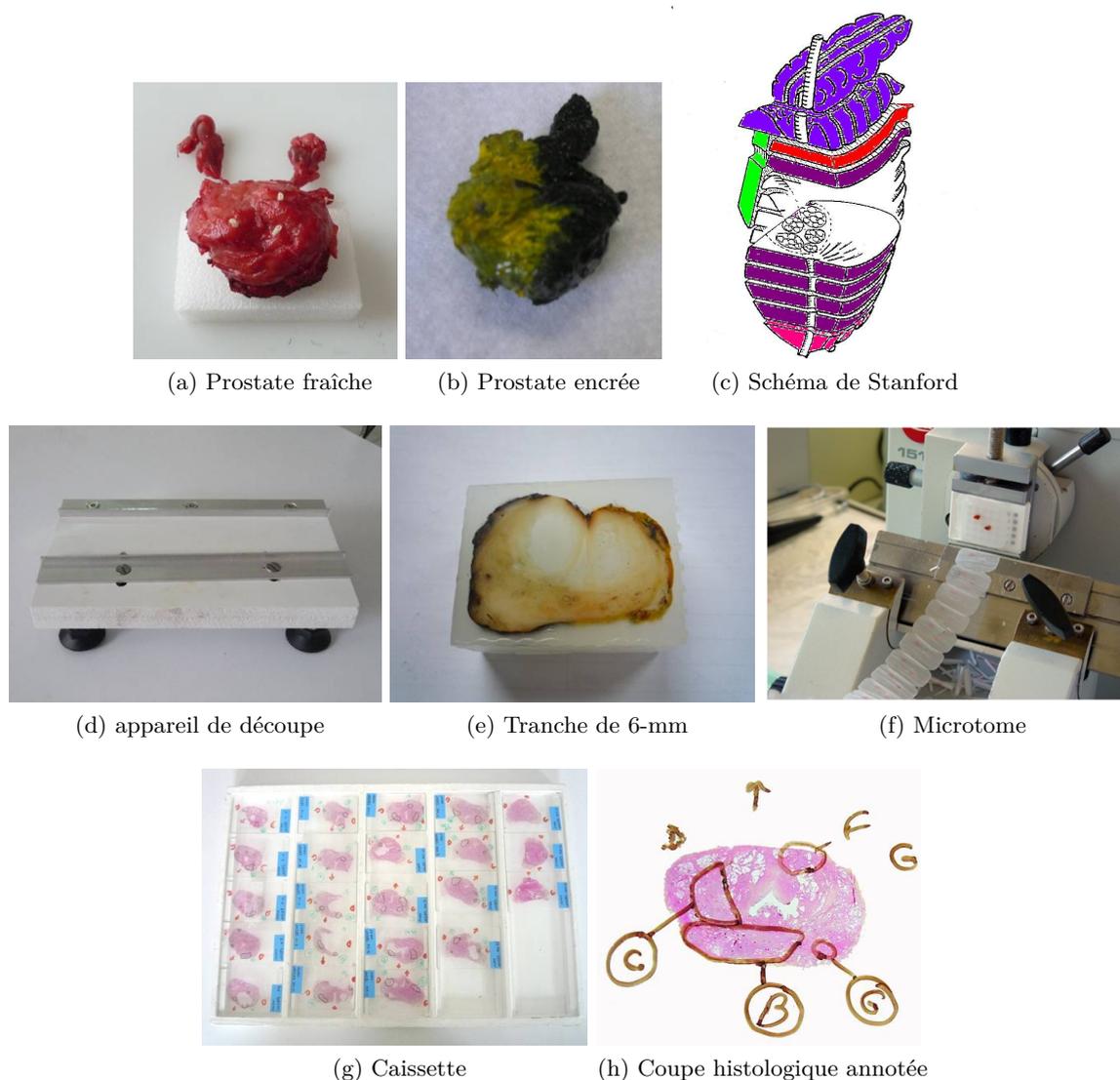


FIGURE 5.1 – Différentes étapes de la préparation de la pièce de prostatectomie. (a) Prostate fraîche. (b) Prostate après orientation et encre (face antérieure en rouge, lobe droit en jaune, lobe gauche en noir). La prostate est ensuite fixée dans du formol et conisée en suivant le (c) schéma de coupe de Stanford. (d) Un appareil dédié est utilisé pour découper la prostate perpendiculairement à la surface rectale de l'apex à la base en (e) tranches de 6-mm d'épaisseur. Les tronçons sont ensuite mis en cassettes et inclus dans la paraffine. Ces blocs de paraffine sont découpés au (f) microtome tous les 500 microns jusqu'à épuiser le bloc au moins de moitié. On étale alors les rubans obtenus sur lames. Après coloration au HES, on obtient ainsi un (g) ensemble de coupes (au moins tous les 3 mm). Chaque (h) coupe histologique est alors analysée au microscope et annotée par un anatomo-pathologiste. Source [62]

- tissu bénin présentant un signal anormal sur les images IRM (Faux positifs (FP) ou Vrais négatifs (VN) de l'analyse radiologique), noté **{NS}** : pour **N**on malin mais **S**uspect,
- tissu bénin présentant une apparence normale sur les images IRM (VN de l'analyse radiologique), noté **{N}** : pour **N**ormal

A noter que, suivant la référence histologique, la lésion entière est contourée et non pas seulement la région de forte anormalité.

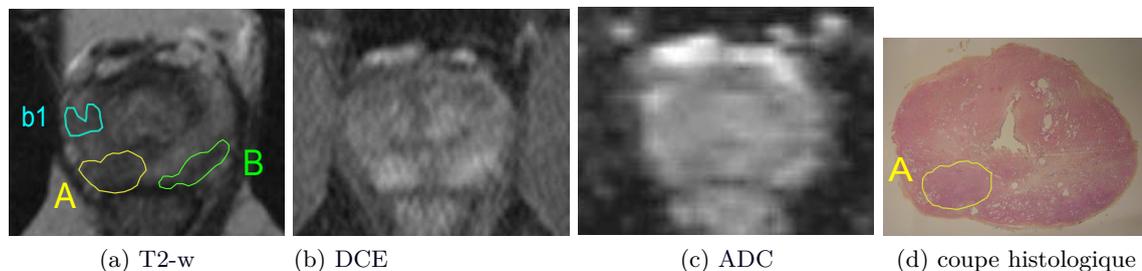


FIGURE 5.2 – IRM de Prostate : (a) T2-w axial, (b) DCE (45 s après injection de Gd-DOTA) et (c) carte ADC obtenue sur un patient âgé de 53 ans avec (d) la coupe histologique numérisée correspondante. Les cancers vérifiés à l'histologie (lettres majuscules A et B) sont reportés sur les images IRM ainsi que les tissus jugés visuellement suspects (lettre minuscule b1). Des Région d'intérêt (ROI) de tissus sains sont extraites automatiquement pour couvrir le reste de la Zone périphérique (ZP). Source [62]

5.6 Constitution de la base d'entraînement du CADe

Les données présentées précédemment sont issues des travaux de thèse d'Emilie Niaf dont l'objectif était la classification binaire de zones d'intérêts (ROIs) en cancer/non cancer. Les ROIs étaient catégorisées en : tissus normaux (N), tissus normaux suspects (NS) et tissus malins (M). Dans le cadre de cette thèse, nous avons choisi d'implémenter un CADe, basé sur une approche de classification voxel à voxel où tous les voxels issus de la base ont été utilisés. De plus, nous souhaitons caractériser le cancer en obtenant des scores de probabilité corrélés au score de Gleason de la tumeur. Dans cette optique, les ROIs malignes (M) ont été plus finement catégorisées selon leur grade de Gleason allant de 6 à 9 (GL6, GL7, GL8 et GL9).

Le tableau 5.2 synthétise la répartition des voxels par classe ainsi que le pourcentage représenté sur le nombre total de voxels issus des 49 patients de la base. Une ROI 2D correspond donc à un contour sur une coupe IRM et une ROI 3D correspond à une même tumeur ou zone suspecte dont les contours peuvent se situer sur plusieurs ROIs 2D consécutives. De plus, tous les voxels restant ont été étiquetés comme normaux (N) contrairement au travail d'Emilie Niaf où des zones d'intérêts avaient été sélectionnées aléatoirement dans les voxels restant.

Classe	ROIs 3D	ROIs 2D	Nb voxels
Gleason 9	2	24	7242 (1.6%)
Gleason 8	7	50	15871 (3.4%)
Gleason 7	34	202	33792 (7.4%)
Gleason 6	28	42	7747 (1.7%)
NS	51	112	31763 (6.9%)
N	X	X	359321 (78.3%)

TABLE 5.2 – Tableau récapitulatif du nombre de ROIs 2D (contourées par coupe), 3D (ROI anatomique présente sur plusieurs coupes), de la répartition des voxels par classe ainsi que leur proportion sur la base totale. La classe (N) correspond aux tissus sains, (NS) aux tissus normaux suspects et le reste de classes correspond aux cancers d'agressivité croissante de grade Gleason 6 à 9.

Choix méthodologiques pour la réalisation du CADe

6.1 Introduction

Ce chapitre détaille les choix que nous avons effectués à chaque étape d'implémentation du Système d'aide à la détection (CADe). Le schéma de la figure 6.1 détaille les principales étapes ainsi que les dimensions des différents vecteurs de caractéristiques mises en jeu. A noter que les caractéristiques descriptives et fonctionnelles ont été directement réutilisées suite aux travaux d'Emilie Niaf [62] sur l'implémentation d'un CADx.

6.2 Acquisition des images

Après acquisition des images IRM multi-paramétrique (IRM-mp) en vue de la constitution de la base détaillée dans le chapitre 5, celles-ci sont normalisées afin de pouvoir comparer les données des différents patients. Le signal T2-w moyen de l'urine présente dans la vessie est d'abord mesuré sur l'image puis le rapport *signal brut T2-w / signal moyen de l'urine* est ensuite calculé et utilisé comme valeur d'intensité pour chaque voxel des images T2-w. De même, le rehaussement relatif du signal $\Delta S = \frac{SI(t) - SI(t=0)}{SI(t=0)}$ est utilisé pour l'image T1-w au lieu du signal brut. Ce rehaussement relatif joue le rôle de correction des variations de sensibilité de l'antenne RF au cours des différentes acquisitions. Enfin, Les cartes ADC ne nécessitent pas de correction car elles sont directement des données quantitatives.

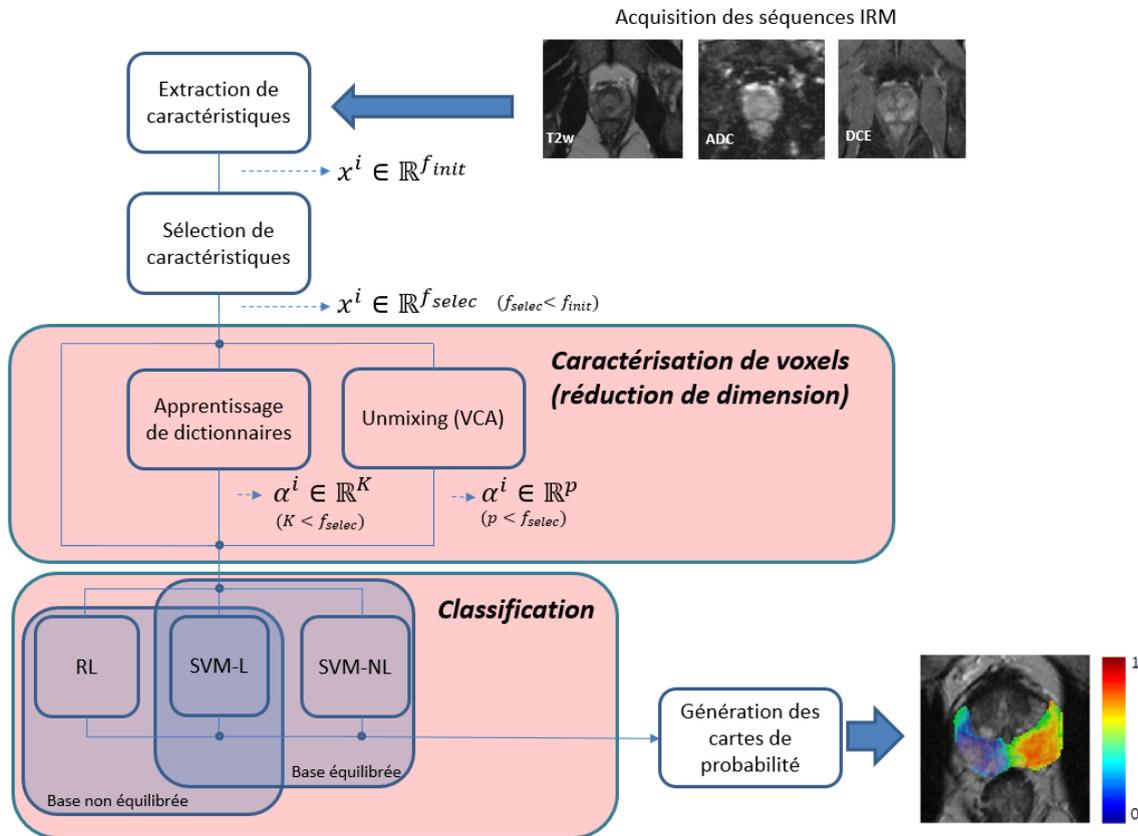


FIGURE 6.1 – Représentation de la méthodologie de développement du CADE. Après acquisition des séquences IRM et création de la vérité terrain, la première étape extrait un vecteur de 110 caractéristiques par voxel x^i . Une phase de sélection permet ensuite de passer d'un total de f_{init} à f_{selec} caractéristiques par vecteur. Ces caractéristiques sont ensuite soit directement utilisées pour la classification ou soit réutilisées dans une étape réduction de dimension par le biais de deux méthodes : VCA et apprentissage de dictionnaires. De la phase de caractérisation des voxels, des coefficients α^i sont extraits (de dimensions K et $K < f_{selec}$) et vont être utilisés en entrée des classifieurs RL et SVM-L sur la base d'apprentissage déséquilibrée, et SVM-L et SVM-NL sur la base d'apprentissage équilibrée. Les sorties des classifieurs sont ensuite utilisées pour générer des cartes de probabilité du cancer

6.3 Extraction de caractéristiques

Cette étape d'extraction de caractéristiques a pour but, pour chaque voxel de chaque séquence IRM, d'extraire une information discriminante dont sa nature peut être descriptive ou fonctionnelle. A la fin de cette étape, chaque voxel sera représenté par un vecteur $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^f]$, dont chaque composante x_i^f représente une caractéristique f extraite au niveau du voxel i . Ces vecteurs de caractéristiques sont ensuite soit directement utilisés pour la classification ou bien réduits une nouvelle fois de dimension lors d'une de étape caractérisation de l'agressivité du cancer. Au total ce sont 110 caractéristiques qui ont été extraites des 3 séquences IRM.

6.3.1 Caractéristiques descriptives

Les caractéristiques descriptives sont extraites des séquences IRM T2w, ADC, DCE et sont de 4 types :

- niveaux de gris brutes de l'image,
- paramètres statistiques du premier et second ordre,
- des gradients calculés sur les niveaux de gris.

Pour la séquence DCE, les paramètres statistiques ainsi que les gradients sont calculés sur un unique temps d'acquisition ($t=1$ min) correspondant au temps moyen d'arrivée au pic de rehaussement $SI_{95\%}$ des tissus malins et présentant donc un meilleur contraste. En revanche le niveau gris brute de l'image n'a pas été utilisé en tant que caractéristique.

Caractéristiques niveaux de gris de l'image. La valeur d'intensité en niveau de gris du voxel est directement exploitée comme caractéristique discriminante. Néanmoins, comme expliqué en section 6.2, les niveaux de gris des séquences T2w et DCE sont normalisées avant utilisation en tant que caractéristique.

Caractéristiques statistiques du premier ordre. Ces caractéristiques, détaillées dans le tableau 6.1, sont calculées pour chaque voxel de l'image sur une fenêtre locale glissante 9×9 . Ce sont la moyenne, médiane, écart-type et déviation absolue moyenne. Ces valeurs sont donc indépendantes de la répartition spatiale des pixels.

Caractéristique	Description
Moyenne	$\frac{1}{ \nu(x) } \sum_{y \in \nu(x)} I(y)$
Médiane	valeur m telle que le nombre de voxels de valeur $\geq m$ soit égal au nombre de voxels de valeur $\leq m$ dans la fenêtre $\nu(x)$
Écart-type	$\sqrt{\frac{1}{ \nu(x) } \sum_{y \in \nu(x)} (I(y) - I(x))^2}$
Déviation absolue moyenne	$ I(x) - (\frac{1}{ \nu(x) } \sum_{y \in \nu(x)} I(y)) $

TABLE 6.1 – Tableau des 4 paramètres statistiques du 1er ordre extraits sur une fenêtre $\nu(x)$ de dimension 9×9 voxels centrée sur le voxel d'intérêt x . $I(y)$ correspond à la valeur en niveau de gris de l'image I pour le voxel $y \in \nu(x)$

Paramètres statistiques du second ordre. Ces caractéristiques du second ordre sont essentiellement des paramètres de texture. En effet, les paramètres statistiques du premier ordre ne fournissent aucune information concernant la nature périodique d'une texture présente dans l'image. De plus, les tissus que l'on souhaite détecter et caractériser (tissus sains, tissus suspects, cancers non-agressifs et agressifs) présentent potentiellement un caractère périodique qu'aucun descripteur du premier ordre ne peut caractériser. Afin de tirer profit de ces informations, des paramètres de textures sont calculés et sont issus de la

matrice de co-occurrence GLCM (Gray-Level Co-occurrence Matrix) des niveaux de gris des voxels. Cette matrice est de taille $N_g \times N_g$ où N_g représente le nombre de niveaux de gris présents dans l'image. Un élément $GLCM_{(\theta,d)}(i,j); (i,j) = 1 \dots N_g$ de cette matrice est obtenu en calculant le nombre de fois qu'un pixel de niveau de gris i est voisin d'un pixel de niveau de gris j à une distance $d = (d_x, d_y) = (d \cos(\theta), d \sin(\theta))$. Ce nombre d'occurrences est ensuite normalisé par le nombre de comparaisons et noté $p_{\theta,d}(i,j)$. Ainsi $p_{\theta,d}(i,j)$ représente une estimation de la probabilité de transition d'un niveau de gris à un autre selon une direction orientée par l'angle θ et une distance d en voxels. Cet angle θ a été limité aux valeurs $0^\circ, 45^\circ, 90^\circ$ et 135° à une distance $d = 1$ voxel. Enfin, la probabilité normalisée $p_d = \frac{\sum_{\theta} p_{\theta,d}}{4}$ est utilisée afin s'affranchir de la direction de l'angle θ . Les 22 paramètres de texture [34] utilisés sont résumés dans le tableau 6.2. La figure 6.2 donne un exemple visuel de quatre de ces caractéristiques.

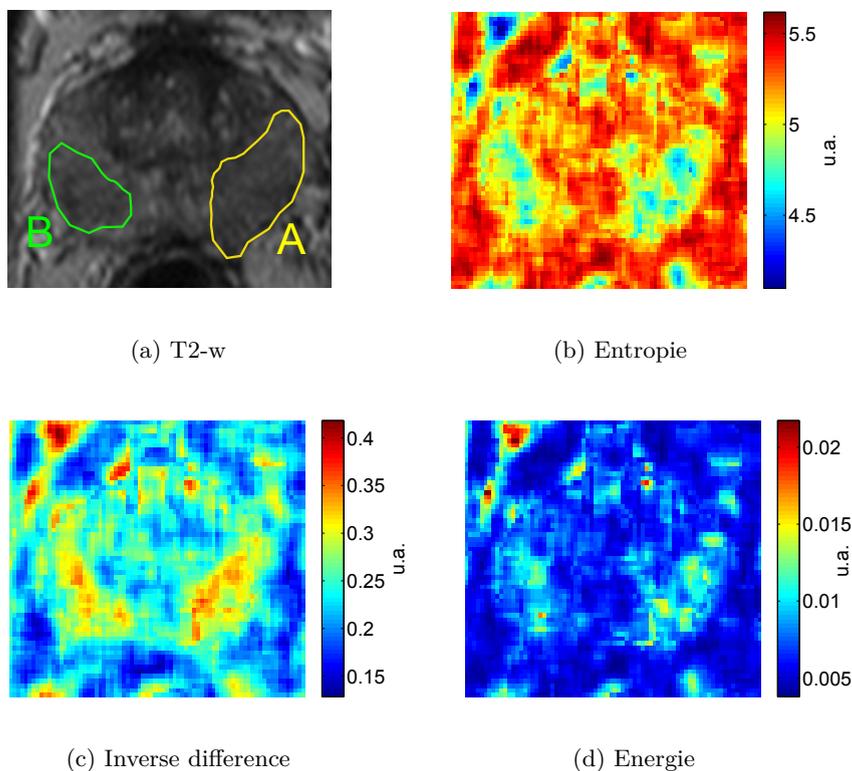


FIGURE 6.2 – Exemples de cartes de caractéristiques de texture : (a) Image T2-w normalisée avec les cancers A et B contourés par les experts, (b) Carte des mesures d'entropie, (c) Carte d'Inverse difference, (d) carte des mesures d'énergie. Les régions malignes A et B sont mises en évidence par les caractéristiques extraites.

Paramètres de gradients. Les 3 opérateurs de gradients que nous avons utilisés sont les filtres de Sobel, Kirsch et un gradient numérique (approche par différence finie). Les gradients directionnels de Sobel sont évalués dans les trois directions principales (horizontal, vertical, diagonal) donnant les caractéristiques G_1 à G_4 ainsi que la magnitude G_{norm} .

Caractéristique	Description
Autocorrélation	$\sum_{ij} ij p(i, j)$
Contrast	$\sum_{ij} i - j ^2 p(i, j)$
Correlation	$\frac{\sum_{ij} (i - \mu_x)(j - \mu_y) p(i, j)}{\sigma_x \sigma_y}$
Correlation	$\frac{\sum_{ij} ij p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Cluster Prominence	$\sum_{ij} (i + j - \mu_x - \mu_y)^4 p(i, j)$
Cluster Shade	$\sum_{ij} (i + j - \mu_x - \mu_y)^3 p(i, j)$
Dissimilarity	$\sum_{ij} i - j p(i, j)$
Energy	$\sum_{ij} p(i, j)^2$
Entropy	$-\sum_{ij} p(i, j) \log(p(i, j) + \epsilon) = H_{XY}$
Homogeneity	$\frac{\sum_{ij} p(i, j)}{1 + i - j ^2}$
Maximum probability	$\max(p(i, j))$
Sum of squares : Variance	$\sum_{ij} (i - \mu)^2 p(i, j)$
Sum average	$\sum_{k=2}^{2N} k p_{i+j}(k)$
Sum variance	$\sum_{k=2}^{2N} (k - H_{X+Y}) p_{i+j}(k)$
Sum entropy	$-\sum_{k=2}^{2N} p_{i+j}(k) \log(p_{i+j}(k) + \epsilon) = H_{X+Y}$
Difference variance	$\sum_{k=0}^{N-1} k^2 p_{i-j}(k)$
Difference entropy	$-\sum_{k=0}^{N-1} p_{i-j}(k) \log(p_{i-j}(k) + \epsilon)$
Information measure of correlation 1	$\frac{H_{XY} - H_{XY1}}{\max(H_X, H_Y)}$
Information measure of correlation 2	$\sqrt{1 - \exp^{-2(H_{XY2} - H_{XY})}}$
Inverse difference	$\sum_{ij} \frac{p(i, j)}{1 + i - j }$
Normalized inverse difference	$\sum_{ij} \frac{p(i, j)}{1 + i - j /N}$
Normalized inverse difference moment	$\sum_{ij} \frac{p(i, j)}{1 + (i - j /N)^2}$

TABLE 6.2 – Liste des 22 caractéristiques de texture issues de la matrice GLCM, où : $p(i, j)$ désigne la probabilité de transition d'un niveau de gris i à un autre j , p est la matrice GLCM normalisée. μ_x, μ_y, σ_x et σ_y désignent les moyennes et écarts-type des densités partielles de probabilité p_x et p_y . p_{i+j} (respectivement p_{i-j}) désigne la somme des probabilités $p(i, j)$ dont la somme (respectivement la différence) des coordonnées vaut $i + j$ (respectivement $i - j$). H_X et H_Y désignent les entropies des densités de probabilité partielles p_x et p_y . $H_{XY1} = \sum_{ij} p(i, j) \log(p_x(i)p_y(j))$ et $H_{XY2} = \sum_{ij} p_x(i)p_y(j) \log(p_x(i)p_y(j))$ sont deux estimations d'entropie jointe (i.e. d'information mutuelle).

Pour le filtre de Kirsh, seul le maximum G_{max} selon 8 directions G_1 à G_8 , correspondantes à des incréments de 45° de 0 à 360° , a été retenu. Le tableau 6.3 présente les 8 paramètres extraits.

Gradient	Description
Sobel	$G_1 = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, G_2 = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, G_3 = \begin{bmatrix} 0 & +1 & +2 \\ -1 & 0 & +1 \\ -2 & -1 & 0 \end{bmatrix}, G_4 = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & +1 \\ 0 & +1 & 2 \end{bmatrix},$ $\mathcal{G}_{norm} = \sqrt{\sum_{i \in Sobel} \mathcal{G}_i^2}$
Kirsch	$\mathcal{G}(y)_{max} = \max_{i \in Kirsch} \mathcal{G}(y)_i,$
Différences finies centrées	$\delta_x \mathcal{I}(x, y) = \frac{\mathcal{I}(x+1, y) - \mathcal{I}(x-1, y)}{2} \quad \delta_y \mathcal{I}(x, y) = \frac{\mathcal{I}(x, y+1) - \mathcal{I}(x, y-1)}{2}$

TABLE 6.3 – Les images de gradient de Sobel et de Kirsch \mathcal{G}_i sont obtenues par filtrage de l'image source \mathcal{I} par les noyaux donnés ci-dessus. Pour chaque noyau G_i , l'image résultante \mathcal{G}_i est obtenue par : $\mathcal{G}_i = G_i \star \mathcal{I}$ où \star est l'opérateur de convolution. Des images caractérisant la magnitude et l'amplitude du gradient sont également calculées : $\mathcal{G}_{norm} = \sqrt{\sum_{i \in Sobel} \mathcal{G}_i^2}$ et \mathcal{G}_{max} telle que $\mathcal{G}(y)_{max} = \max_{i \in Kirsch} \mathcal{G}(y)_i$, pour tout pixel $y \in \mathcal{I}$.

6.3.2 Caractéristiques fonctionnelles

Des caractéristiques fonctionnelles semi-quantitatives, obtenues par analyse directe de la forme de la courbe de rehaussement en agent de contraste (courbe tracée à partir des images DCE) ont été extraites des séquences dynamiques DCE.

L'analyse de la courbe de rehaussement dynamique du signal permet une première description simplifiée de la distribution de l'Agent de contraste (AC) [68]. Elle fournit des paramètres empiriques tels que la pente, le temps d'arrivée de l'AC, sa pente d'élimination, le temps où le rehaussement est maximal. Pour extraire ces caractéristiques, on s'intéresse directement aux variations du signal T1-pondérée (T1-w) après injection du bolus (cf. figure 6.3). Les principales caractéristiques pouvant être extraites sont les suivantes :

Valeurs de signal remarquables :

- signal de base, "baseline" (SI_0 , acquis au temps 0),
- pic du contraste (réhaussement maximal, SI_{max}),
- valeur à 95% du pic ($SI_{95\%}$),
- valeur à 5% du pic (SI_{onset} , début du réhaussement),
- amplitude ($A = SI_{95\%} - SI_0$),
- signal de fin d'acquisition (SI_{end}).

Temps remarquables :

- temps d'arrivée au pic (T_{pic}),
- temps d'arrivée à 95% du pic (T_{max}),
- temps à 5% du pic (T_{onset}),
- temps de perfusion ($T_{95\%} - T_{onset}$).

Aire sous la courbe : Aire sous la courbe de rehaussement en produit de contraste (AUGC, de l'anglais Area Under the Gadolinium Curve).

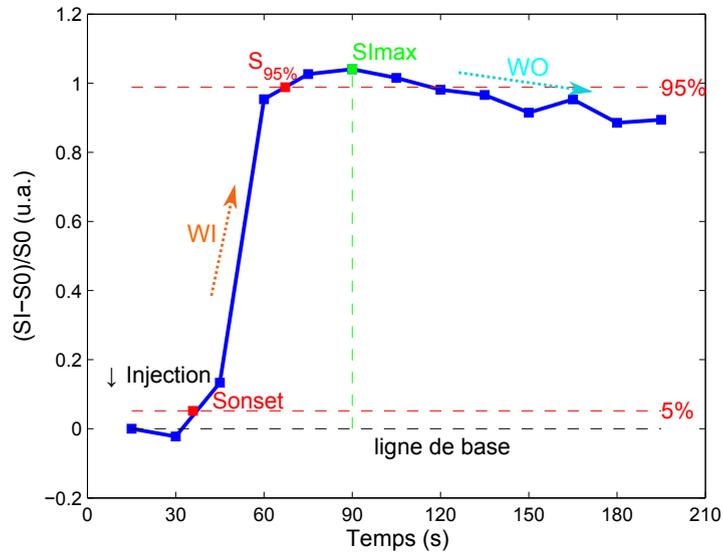


FIGURE 6.3 – Un exemple de courbe temps-intensité du signal d'un pixel en IRM DCE. Quelques paramètres fonctionnels semi-quantitatifs extraits de la courbe de réhaussement de l'AC sont représentés : SI_0 , SI_{onset} , $SI_{95\%}$, SI_{max} .

Dans le cadre de cette thèse, les paramètres semi-quantitatifs que nous avons utilisés sont donnés dans le tableau 6.4.

Caractéristique	Description
SI_{95}	intensité du signal acquis à $t=95\%$ du pic de réhaussement
SI_{onset}	valeur à 5% du pic
A	amplitude du rehaussement
SI_{end}	valeur du signal en fin d'acquisition
T_{pic}	temps d'arrivée au pic
AUGC	aire sous la courbe de rehaussement

TABLE 6.4 – Liste des 6 paramètres de perfusion semi-quantitatifs extraits des images DCE

6.4 Sélection de caractéristiques

Notre choix de schéma de sélection de caractéristiques s'est porté sur une méthode hybride de type *embedded* dont les caractéristiques sont sélectionnées par *filtrage* à l'aide d'un classifieur de type SVM linéaire [77] décrit dans la section 6.6.1. L'idée principale est de sélectionner, parmi les $f = 110$ caractéristiques descriptives et fonctionnelles initialement extraites des séquences IRM, les k caractéristiques permettant de maximiser le score d'AUC d'un SVM linéaire. Soit la tâche binaire de classification consistant à

séparer les voxels appartenant aux classes (N),(NS),(GL6) des voxels appartenant aux classes (GL7),(GL8) et (GL9) correspondant aux cancers agressifs. La méthode de sélection consiste dans un premier temps à réaliser un apprentissage en Validation croisée par exclusion d'un patient (LOPO) (cf section 4.3.7) sur l'ensemble des données avec un vecteur de caractéristiques image de taille $f = 110$. Suite à cette première étape, le meilleur paramètre C du SVM est sélectionné, i.e. celui maximisant l'AUC moyennée sur les patients pour chaque valeur de C . Ensuite, les poids $\|\mathbf{w}^j\|^2$ de l'hyperplan séparateur (obtenu avec la précédente valeur optimale de C) sont triés par ordre décroissant. Dans le cadre du SVM linéaire, ces poids peuvent être assimilés à l'influence d'une caractéristique dans le résultat final de classification. En effet, la classe y d'un voxel x est donnée par $y = \text{sign}(\sum_{j=1}^f w_j^t x_j + b)$ avec j l'indice de la $j^{\text{ème}}$ poids de l'hyperplan séparateur ou le poids de la $j^{\text{ème}}$ caractéristique. Ainsi, on voit que plus le poids de la $j^{\text{ème}}$ caractéristique sera élevé et plus celle-ci aura de l'influence dans le résultat de classification. Ensuite, suivant une procédure de recherche de type "forward" pour trouver le meilleur sous-ensemble de caractéristiques, nous relançons une série de phases d'apprentissages avec le meilleur C obtenu précédemment mais en faisant varier le nombre de caractéristiques, rangées par ordre d'importance de poids, de $f = 5$ à $f = 110$. Le nombre optimal de caractéristique est celui qui maximise une nouvelle fois l'AUC moyennée sur les patients.

6.5 Caractérisation des voxels

Cette étape a pour but de caractériser les voxels dans l'objectif d'évaluer l'agressivité des tumeurs. Le principe est de projeter les caractéristiques images décrites précédemment dans un espace de dimension inférieur. Ce nouvel espace de représentation a pour objectif de mieux décrire les différentes classes et ainsi améliorer le pouvoir discriminant de la chaîne de classification. Dans cette optique nous avons utilisé deux schémas : *Vertex Component Analysis* [61], qui est une méthode de base en imagerie hyperspectrale (HSI) et une méthode globale d'*apprentissage de dictionnaire* [55]. Ces deux approches reposent sur une hypothèse commune : une décomposition linéaire sur des éléments (atomes) d'une base D (dictionnaire). L'objectif est d'apprendre les éléments \mathbf{d}_k de D ainsi que leur poids α_k correspondant.

En gardant pour hypothèse que la signature d'un voxel est le résultat d'un mélange linéaire de différents types de matériaux de base (eau, sang, tissus normaux, cancers agressifs, cancers non agressifs, etc.), on souhaite caractériser le voxel, i.e. quantifier la fraction de chaque élément de base du dictionnaire et classer le voxel dans la classe correspondant à l'élément dominant. Par analogie au formalisme de séparation de source (unmixing) en HSI, un voxel $\mathbf{x} \in \mathbb{R}^f$ peut être exprimé par une combinaison linéaire d'un nombre fini d'éléments de base constituant un dictionnaire $D \in \mathbb{R}^{f \times K}$ de la manière suivante :

$$\hat{\mathbf{x}} = \sum_{k=1}^K \alpha^k \mathbf{d}_k = \mathbf{D} \cdot \boldsymbol{\alpha} \quad (6.1)$$

où K est la taille du dictionnaire, f représente le nombre de caractéristiques "brutes" extraites des images IRM multiparamétriques, \mathbf{d}_k les éléments du dictionnaire (i.e. les colonnes de D et $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_K]$ sont les coefficients représentant les fractions de chaque élément de base du voxel \mathbf{x} . N'ayant aucune indication ou hypothèse sur le nombre d'éléments de base présents dans la prostate, nous avons utilisé une méthode d'apprentissage de dictionnaires basée sur l'estimation conjointe des éléments de base \mathbf{d}_k et de leur coefficient α_k afin de construire un dictionnaire de référence D . L'intérêt de cette représentation est donc de représenter le contenu d'une image au niveau voxelique mais aussi de réduire la dimension du vecteur de caractéristique \mathbf{x} en considérant uniquement les α^k pour la partie classification. Ainsi, à la fin de cette étape, un voxel $\mathbf{x} \in \mathbb{R}^f$ sera représenté par $\hat{\mathbf{x}} \in \mathbb{R}^k$ avec $k < f$.

6.5.1 Méthodes de caractérisation en imagerie hyperspectrale

Une image hyperspectrale est par définition acquise à plusieurs longueurs d'onde λ_j avec $j = 1 \dots f$ et f le nombre de bandes spectrales. Ainsi, chaque pixel d'une image hyperspectrale peut-être représenté par un vecteur $\mathbf{x} \in \mathbf{R}^f$ dont chaque composante correspond au niveau de gris de l'image acquise dans la bande spectrale λ_j . Par analogie aux vecteurs de caractéristiques descriptives et fonctionnelles $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^f]$ directement issues des images, chaque longueur d'onde λ_j représente donc une caractéristique x_i^f . Ces différents niveaux de gris pour un même pixel à plusieurs longueurs d'onde peuvent être le résultat d'un mélange linéaire ou non-linéaire de signatures de différents matériaux de base (eau, végétation, plusieurs types de sols, etc..) dont on souhaite en identifier les proportions. La figure 6.4 illustre ces propos à l'aide de la représentation schématique d'une image hyperspectrale dont les strates représentent différentes bandes spectrales d'acquisition. Sur les trois pixels de l'image, deux représentent un mélange de sols, roches et végétations, le troisième est un pixel pure représentant de l'eau. Ainsi, la visualisation de leur spectre de reflectance en fonction de la longueur d'onde permet éventuellement de distinguer une tendance caractéristique de ces types de milieux imagés. Cette problématique, appelée *unmixing* ou séparation de sources, est notamment rencontrée dans le domaine de l'imagerie satellite hyperspectrale.

Ainsi, comme représenté sur la figure 6.5 et dans le cas linéaire, un pixel \mathbf{x} d'une image hyperspectrale peut se décomposer sous la forme $\hat{\mathbf{x}} = \sum_{k=1}^K \mathbf{d}_k \alpha_k$ avec \mathbf{d}_k le spectre de reflectance (aussi appelé signature spectrale) d'un des matériaux de base contenu dans l'image et α_k la proportion de ce matériau de base pour le voxel \mathbf{x} .

Parmi les différents algorithmes de séparation de sources, on peut citer la méthode Vertex Component Analysis (VCA) [61], une des premières utilisée en HSI, dont l'hypothèse forte

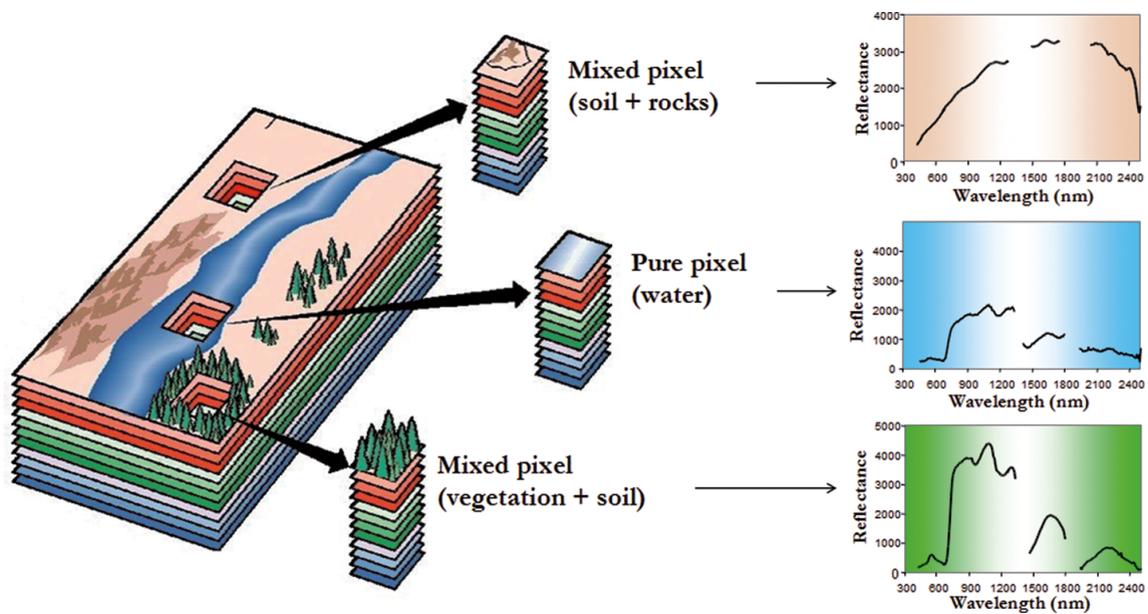


FIGURE 6.4 – Représentation schématique d’une image hyperspectrale contenant un pixel pur et deux pixels résultants chacun d’un mélange de deux matériaux. Les spectres de réflectance représentent donc les valeurs que prennent ces pixels pour chacune des longueurs d’onde λ_j . Source [8]

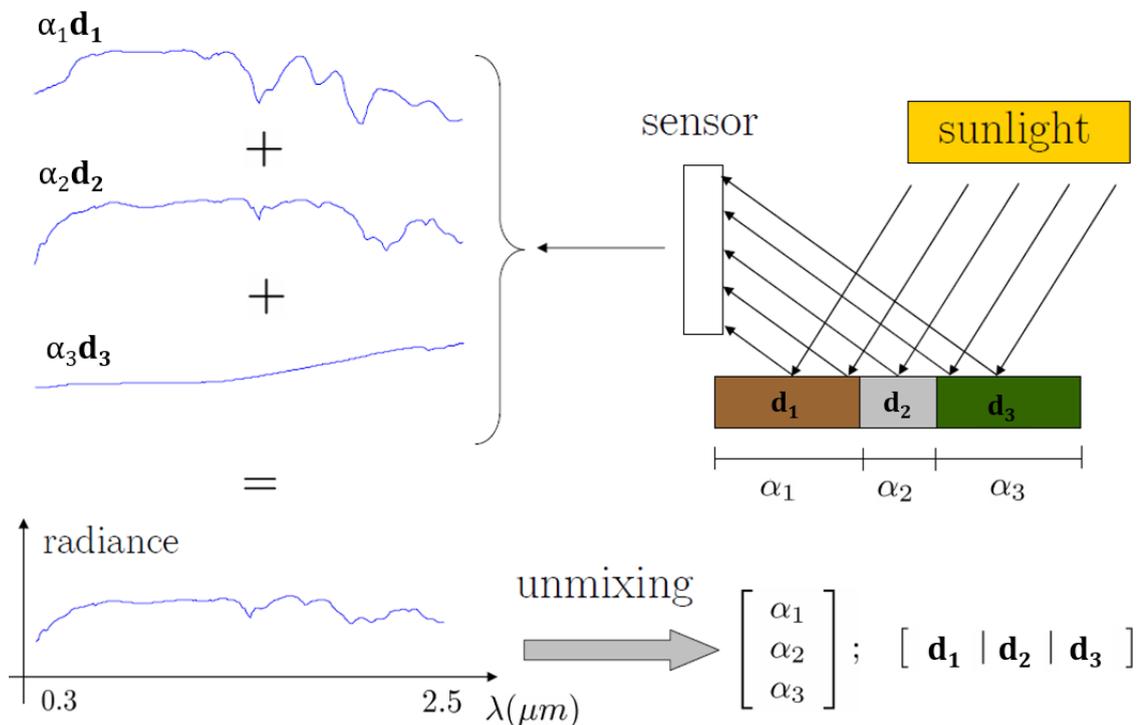


FIGURE 6.5 – Exemple de processus de mélange linéaire pour l’imagerie hyperspectrale : $\alpha_1, \alpha_2, \alpha_3$ sont les coefficients associés aux spectres de base $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ représentant des matériaux purs. Source [8]

repose sur la présence dans l'image de pixels \mathbf{x} dit "purs", i.e. ne résultant d'aucun mélange linéaire ou non-linéaire, et censés représenter la signature spectrale \mathbf{d}_k d'un matériau de base. De plus, afin de respecter des contraintes physiques, la méthode impose $0 < \alpha_k < 1$ et $\sum_{k=1}^K \alpha_k = 1$, c'est-à-dire une contrainte de positivité sur la proportion du matériau de base ainsi qu'une contrainte de somme à 1 de telle manière qu'on ait identifié tous les éléments contenus dans l'image. Afin d'identifier ces pixels "purs" et ainsi en déduire leur signature spectrale \mathbf{d}_k , l'ensemble des pixels représentés par un vecteur $\mathbf{x} \in \mathbb{R}^f$ sont projetés dans un espace de dimension f . De par les contraintes de positivité et de somme à 1, chaque pixel projeté appartient donc à un simplexe (triangle à f dimensions) dont les K sommets sont les pixels purs (cf. figure 6.6). Dans la suite de cette section, $\mathbf{x}_i \in \mathbb{R}^f$ représente un voxel élémentaire de l'image. L'algorithme 1 présente le principe général du VCA.

Input : K ; $x := [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$; (avec $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^f]$)
Result : D, \mathbf{d}_k
 $\hat{D} := 0$ ($D \in \mathbb{R}^{f \times K}$) ;
for $k := 1$ **to** K **do**
 $y := l^\top x$;
 $i := \arg \max_{i=1..n} ||y(i)||$;
 $\mathbf{d}_k := \mathbf{x}_i$;
 $[\hat{D}]_{:,k} := \mathbf{d}_k$;
 $l := \text{orthonormal}([\hat{D}]_{:,1:k})$;
 (génère un vecteur l orthonormal au sous-espace engendré par $[\hat{D}]_{:,1:k}$)
end

Algorithme 1 : Algorithme simplifié du Vertex Component Analysis. Les données d'entrée sont K le nombre d'atomes du dictionnaire et $x = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ un ensemble de n voxels de taille $\mathbf{x}_i \in \mathbb{R}^f$. A chaque itération k , on détermine un atome \mathbf{d}_k du dictionnaire par recherche de l'indice du voxel correspondant au maximum (pixel pur) dans un sous-espace engendré par chaque estimation intermédiaire $[\hat{D}]_{:,1:k}$ du dictionnaire D .

Ainsi, après projection, un voxel $\mathbf{x}_i \in \mathbb{R}^f$ s'exprime par $\hat{\mathbf{x}}_i = D \cdot \alpha$ avec $D = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]$ une matrice constituée des signatures des composants de base et $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]$ les fractions de ces éléments de base. Les pixels purs sont ensuite identifiés par recherche de maximums suivant chaque axe K représentant le nombre a priori d'éléments de base dans l'image.

Par transposition à notre problème de caractérisation des lésions en différents grades de cancers, nous avons utilisé VCA afin d'identifier les composants de base correspondant aux classes N, NS, Gleasons 6, 7, 8 et 9.

Néanmoins, plusieurs facteurs communs entre l'imagerie hyperspectrale et IRM multiparamétriques rendent la tâche de classification complexe :

- le nombre élevé de caractéristiques extraites et de bandes spectrales,
- Le faible nombre de données d'entraînement parfaitement étiquetées,

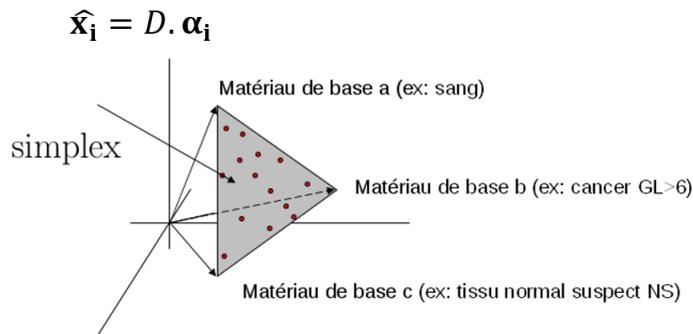


FIGURE 6.6 – Principe du VCA : les voxels $\mathbf{x}_i \in \mathbb{R}^f$ sont projetés dans un espace de dimension f (le nombre de bandes spectrales) dans lequel est recherché un sous-espace orthogonal de dimension K , le nombre hypothétique de composants de base. Ainsi, après projection, un voxel s'exprime dans un espace $\mathbf{x}_i \in \mathbb{R}^K$ par $\hat{\mathbf{x}}_i = D \cdot \boldsymbol{\alpha}_i$ avec D un dictionnaire constituée des signatures des composants de base et α_i les fractions de ces éléments de base pour chaque voxel i . Source : [8]

- le mélange linéaire et non-linéaire des constituants de base,

En effet, le nombre élevé de caractéristiques et donc de dimensions rend les espaces de projection particulièrement complexes pour les algorithmes de classification (cf. section 4.3.4) et tombent ainsi dans des problèmes dits de "curse of dimensionality", connus en apprentissage supervisé. De même, l'hypothèse de présence d'un pixel pur dans le jeu d'apprentissage est une limitation très forte. En effet, dans notre cas il ne nous est pas possible de s'assurer qu'une zone étiquetée comme saine par exemple ne comporte pas de pixels appartenant à une autre classe, connue ou non. Cette dernière contrainte nous amène donc directement à la troisième, l'appartenance de l'algorithme aux méthodes linéaires limite donc son application à des données pas ou peu "mélangées". Enfin, il est important de noter que la contrainte de positivité des constituants de base \mathbf{d}_k et des fractions d'abondance \mathbf{d}_k imposée par la méthode *VCA* n'est pas initialement respectée dans le cas des données de prostate. Afin de la respecter, les données ont été normalisées entre 0 et 1, ce qui a pour effet d'obtenir des constituants de base non interprétables physiquement contrairement aux données hyperspectrales.

6.5.2 Apprentissage de dictionnaires

L'idée est d'utiliser un dictionnaire dont les éléments de base (appelés atomes dans la terminologie de l'apprentissage de dictionnaires) sont estimés à partir des caractéristiques descriptives et fonctionnelles décrites précédemment.

Comme expliqué précédemment, *VCA* fait l'hypothèse de la présence de pixels "purs" dans l'ensemble d'apprentissage. Cette hypothèse est très certainement fautive de par les non-linéarités introduites lors de l'acquisition des images IRM ainsi que lors du calcul des caractéristiques descriptives et fonctionnelles. Néanmoins, les méthodes linéaires basées sur l'apprentissage de dictionnaires parcimonieux sont de prometteuses alternatives aux

méthodes linéaires basées sur l'hypothèse de la présence de pixels pures. Bien qu'appartenant à cet ensemble de méthodes dites linéaires, celles-ci sont plus robustes et permettent l'estimation des éléments du dictionnaire sans présence de pixels purs dans les données [8]. En effet, elles peuvent aussi encoder des variations non-linéaires car les éléments de base du dictionnaire sont directement appris de la de départ et plusieurs éléments de base peuvent représenter un unique et même tissu.

Suivant le modèle général de l'équation 6.2, les méthodes d'apprentissage de dictionnaires parcimonieux (DL) sont basées sur l'optimisation d'une fonction de coût dont le but est de trouver une représentation d'un dictionnaire D représentant au mieux les données tout en encourageant une représentation parcimonieuse. La forme générale du problème d'optimisation est décrit par :

$$+ + \min_{\mathbf{D} \in \mathcal{C}, \alpha_i \in \mathbb{R}^k} \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \quad (6.2)$$

où n est le nombre de voxels de l'ensemble de données d'apprentissage, λ un paramètre de régularisation de la parcimonie et \mathcal{C} est un ensemble convexe dans lequel se trouvent les éléments \mathbf{d}_k du dictionnaire.

Notion de parcimonie. D'une manière générale, la parcimonie en traitement du signal a pour but de décrire un vecteur de manière "économique", i.e. le représenter à l'aide de quelques coefficients non-nuls, les autres étant supposés nuls ou de faible valeur. Les applications sont nombreuses, notamment en compression de données, où des vecteurs parcimonieux occupent donc beaucoup moins d'espace en mémoire puisqu'ils ne sont représentés que par quelques coefficients non-nuls. En optimisation, la parcimonie permet un gain de temps dans les calculs ainsi que d'une meilleure approche des solutions. Intuitivement, la norme l_0 , consistant à compter le nombre de coefficients non nuls, est la mesure de parcimonie la plus adéquate mais aussi une des plus contraignantes lors de la minimisation d'une fonction. Afin de relâcher ces contraintes, on cherche plutôt à mesurer la vitesse de décroissance à l'aide de la norme l_1 ou d'une quasi-norme l_q où $\|x\|_q = \left(\sum_{i=1}^N |\mathbf{x}_i|^q\right)^{1/q}$ avec $0 < q < 1$. Dans l'équation 6.2, la norme l_1 est utilisée comme mesure de parcimonie et introduite par le terme $\lambda \|\alpha_i\|_1$. Cette norme l_1 est la plus utilisée dans le domaine de l'optimisation convexe et a prouvé son efficacité à approcher la solution la plus parcimonieuse l_0 sans ses contraintes.

Algorithme d'apprentissage de dictionnaires. Dans le but de résoudre le problème 6.2, nous avons choisi d'utiliser une méthode d'apprentissage de dictionnaires dite "online" proposée par Mairal et al. [55] et décrite par l'algorithme 2. Il s'agit d'une méthode dite "online" car elle traite, à chaque itération de l'algorithme, un échantillon x_t , de taille fixée. Cette manière de procéder est particulièrement intéressante en termes de temps de calculs et en présence d'une base d'apprentissage volumineuse. L'approche employée ici

estime alternativement α et D . A chaque itération t , un échantillon x_t de taille fixe est tirée aléatoirement et l'algorithme estime dans un premier temps α_t , les coefficients de décomposition de l'échantillon x_t , avec le dictionnaire D_{t-1} acquis à l'itération précédente. Le dictionnaire D_t est ensuite mis à jour à l'aide des coefficients α_t et du dictionnaire D_{t-1} . On notera également les matrices A_t et B_t , dont le rôle est de conserver l'information contenue dans les coefficients α_t du cycle précédent. Cette information est mise à profit dans le calcul de la mise à jour du dictionnaire D_t . L'algorithme 2 présente le principe de la méthode pour un unique échantillon (voxel) x_t tiré à chaque itération t .

```

Input :  $\mathbf{x}_t \in \mathbb{R}^f \sim p(\mathbf{x})$  ;  $\lambda \in \mathbb{R}$  ;  $D_0 \in \mathbb{R}^{f \times k}$  ;  $T$ 
 $A_0 \leftarrow 0$  ,  $B_0 \leftarrow 0$ 
for  $t := 1$  to  $T$  do
    Draw  $x_t$  from  $p(x)$  ;
    Calcul des coefficients :
     $\alpha_t = \underset{\alpha \in \mathbb{R}^k}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_t - \mathbf{D}_{t-1} \alpha\|_2^2 + \lambda \|\alpha\|_1$  ;
     $A_t \leftarrow A_{t-1} + \alpha_t \alpha_t^\top$  ;
     $B_t \leftarrow B_{t-1} + x_t \alpha_t^\top$  ;
    Calcul de  $D_t$  :
     $D_t = \underset{D \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{t} \sum_{i=1}^t \frac{1}{2} \|\mathbf{x}_i - \mathbf{D} \alpha_i\|_2^2 + \lambda \|\alpha\|_1$  ;
     $D_t = \underset{D \in \mathcal{C}}{\operatorname{argmin}} \frac{1}{t} (\frac{1}{2} \operatorname{Tr}(D^\top D A_t) - \operatorname{Tr}(D^\top B_t))$  ;
end
return  $D_T$ 
    
```

Algorithme 2 : "Online dictionary learning for sparse coding" par Mairal et al. [55] ; D_0 est le dictionnaire initial, x_t est l'échantillon tiré aléatoirement suivant la loi $p(x)$ dans la base d'apprentissage, T est le nombre d'itérations, λ le paramètre de régularisation de la parcimonie.

Extension mini-batch. Le principal avantage de cette méthode, dont la convergence est prouvée, est son adaptabilité à de grands volumes. Afin d'accélérer le processus de convergence, cette méthode permet de choisir aléatoirement à chaque itération t , un échantillon x_t de plusieurs voxels. Nous avons réglé ce paramètre, appelé mini-batch extension, à 512 voxels et le nombre d'itérations à $T = 100$. En effet, Mairal et al [54] ont montré que le gain en performance est négligeable lorsque l'on augmente considérablement ces 2 paramètres, au détriment du temps de calcul.

Application à notre base d'apprentissage et à la classification. La méthode décrite précédemment permet d'estimer un dictionnaire D sur un ensemble d'échantillons d'apprentissage. Ensuite, pour la phase de classification, nous ré-estimons, à partir de D , les coefficients α de chacun des voxels de notre base d'apprentissage. Dans le cadre de nos travaux, nous avons choisi d'estimer deux types de vecteurs de coefficients, α_{DL} et α_{ssDL} ,

issus de deux dictionnaires différents. le premier type de vecteur est issu d'un dictionnaire global D estimé par tirage aléatoire sur les voxels de la base d'apprentissage. Le second vecteur α_{ssDL} est calculé sur un dictionnaire D construit par concaténation de dictionnaires élémentaires D_i estimés individuellement sur chacune des 4 classes (N),(NS),(Gleason 6) et (Gleason >6). Ainsi ce deuxième vecteur de coefficients a pour but d'introduire la notion de classe dans l'estimation du dictionnaire D , rendant ainsi la démarche semi-supervisée.

6.6 Classification

6.6.1 SVM linéaire

Les deux sous-parties suivantes traitant de la formulation théorique des SVM est tirée des livres [82, 88]. Les machines à vecteurs de support (SVM) font partie des modèles de classification supervisée binaire dont le but est de séparer deux jeux de données, appartenant à deux classes différentes, par un hyperplan séparateur. Ils reposent sur deux principes fondamentaux : la notion de *marge maximale* et celle de *Kernel trick*. pour la première notion, la marge est la distance entre l'hyperplan séparateur et les points les plus proches, appelés vecteurs support. En effet, l'hyperplan séparateur optimal est celui qui maximise la marge.

La deuxième notion de kernel trick vise à traiter les problèmes non linéairement séparables en transformant l'espace de représentation des données d'entrée en un espace de plus grande dimension, voire infini, dans lequel il serait possible de séparer linéairement les données. Cette "astuce" est réalisée à l'aide de fonctions noyaux, dont l'avantage est qu'il n'est explicitement pas nécessaire de connaître la transformation à appliquer pour changer d'espace. Ces fonctions transforment un produit scalaire impossible en une simple évaluation ponctuelle de fonction.

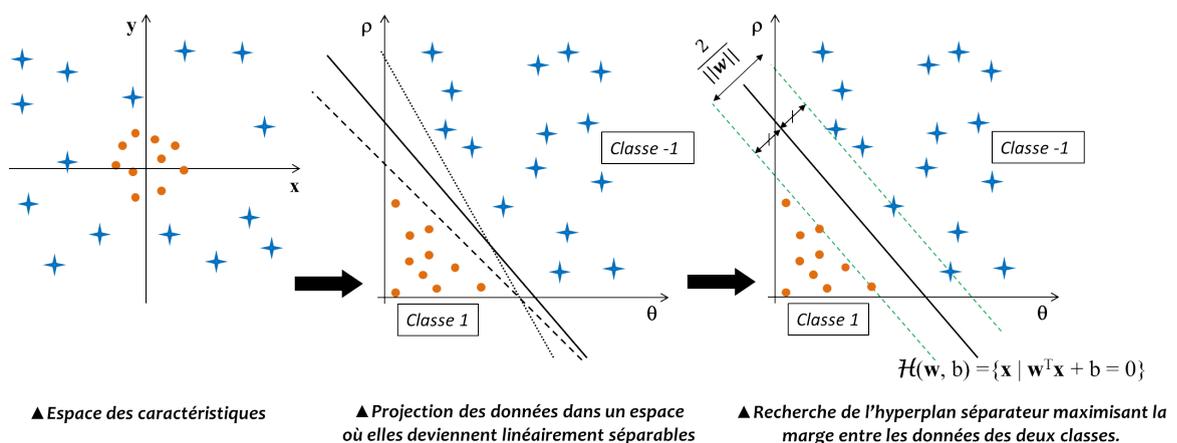


FIGURE 6.7 – Schéma de principe des SVM. Source [62]

Formulation primale

Soit un jeu de données d'apprentissage $(\mathbf{x}_i)_{i=1..n}$ avec $\mathbf{x}_i \in \mathbb{R}^f$, f étant le nombre de caractéristiques, et leur classe associée $y_i = \{-1; +1\}$. On cherche les paramètres (\mathbf{w}, b) d'un hyperplan séparateur h des données \mathbf{x}_i pouvant s'exprimer par $h(\mathbf{x}_i) = \mathbf{w}^\top \mathbf{x}_i + b = 0$. \mathbf{w} est un vecteur normal à l'hyperplan et $\frac{b}{\|\mathbf{w}\|}$ le décalage à l'origine de l'hyperplan. La classe d'un exemple \mathbf{x}_i est alors donnée par le signe de $h(\mathbf{x}_i)$, soit $+1$ si $h(\mathbf{x}_i) \geq 0$ et -1 sinon. On remarque que la condition de séparabilité linéaire des données peut s'exprimer par : $\forall i$, on a $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 0$.

Afin de respecter la condition de séparabilité précédente, les SVM cherchent à maximiser la marge (cf figure), c'est-à-dire maximiser la plus petite distance entre un point \mathbf{x}_i de l'espace d'entrée et l'hyperplan séparateur. Cette marge se caractérise par une marge positive et négative, c'est-à-dire les distances aux points de classe $+1$ et -1 se trouvant les plus proches de l'hyperplan séparateur. Ces points sont appelés *vecteurs supports* et notés \mathbf{x}_{sup}^+ et \mathbf{x}_{sup}^- respectivement.

De manière générale, la distance d'un point \mathbf{x}_i à l'hyperplan séparateur est donnée par sa projection orthogonale sur l'hyperplan :

$$\frac{|\mathbf{w}^\top \mathbf{x}_i + b|}{\|\mathbf{w}\|} \quad (6.3)$$

pouvant aussi s'exprimer par :

$$\frac{y_i(\mathbf{w}^\top \mathbf{x}_i + b)}{\|\mathbf{w}\|} \quad (6.4)$$

puisque l'hyperplan recherché sépare linéairement les données. Ce problème de maximisation de la marge peut donc s'écrire sous la forme :

$$\arg \max_{\mathbf{w}, b} \left\{ \frac{1}{\|\mathbf{w}\|} \min_k [y_i(\mathbf{w}^\top \mathbf{x}_i + b)] \right\} \quad (6.5)$$

En normalisant les paramètres \mathbf{w}, w_0 de l'hyperplan de telle manière que les points se trouvant sur les marges positives et négatives satisfassent :

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_{sup}^+ + b = 1 \\ \mathbf{w}^\top \mathbf{x}_{sup}^- - b = -1 \end{cases} \quad (6.6)$$

on a $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$ et la marge totale peut désormais s'exprimer par $\frac{1}{\|\mathbf{w}\|}$. Le problème d'optimisation peut alors s'exprimer sous la forme :

$$\begin{cases} \text{minimiser} & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{sous contraintes} & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 \end{cases} \quad (6.7)$$

Marge souple

Le problème précédent est beaucoup trop contraint pour être résolu directement. En effet on n'autorise aucun exemple d'apprentissage à résider dans la marge, de par une certaine non linéarité dans les données ou une mauvaise labelisation au moment de la constitution de la vérité terrain des données. On introduit alors la notion de "*marge souple*" dont l'idée est d'autoriser certains exemples à résider dans la marge à l'aide d'une variable ressort \mathbf{x}_i . La contrainte de bonne classification du système 6.7 devient alors $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i$. On choisit ensuite de plus ou moins pénaliser les variables ressort élevées à l'aide de la constante C (avec $C \geq 1$), appelée coût de mauvaise classification, qui permet de contrôler le compromis entre nombre d'erreurs et largeur de la marge. En effet, plus on pénalisera les erreurs ξ_i par une valeur de C élevée et moins la marge sera souple. L'idée est d'obtenir un classifieur ayant une bonne capacité de généralisation. Avec ces nouvelles contraintes le problème 6.7 peut se réécrire de la manière suivante :

$$\begin{cases} \text{minimiser} & \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \\ \text{sous contraintes} & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall i \in 1 \dots n, \\ & \xi_i \geq 0, \quad \forall i \in 1 \dots n \end{cases} \quad (6.8)$$

Formulation duale

Le problème primal précédent peut se mettre sous la forme d'un problème dual dans le cas où la fonction objectif $\frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$ ainsi que les contraintes $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i$ et $\xi_i \geq 0$ sont strictement convexes (conditions de Kuhn-Tucker). Si on note \mathcal{L} le lagrangien associé au problème d'optimisation, \mathcal{L} est défini par la somme de la fonction objectif et d'une combinaison linéaire des contraintes dont les coefficients sont appelés multiplicateurs de Lagrange ou variables duales :

$$\mathcal{L}(\mathbf{w}, b, \xi, \alpha, \beta) = \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i(\mathbf{w}^\top \mathbf{x}_i + b) - (1 - \xi_i)) - \sum_{i=1}^n \beta_i \xi_i \quad (6.9)$$

Le problème revient donc à minimiser le lagrangien précédent par rapport à \mathbf{w} et w_0 et maximisé par rapport à α et β ($\alpha \geq 0, \beta \geq 0$) avec $\alpha_i, \beta_i, \forall i \in 1 \dots n$, les multiplicateurs de Lagrange. En annulant les dérivées partielles du lagrangien, selon les conditions de Kuhn-Tucker, par rapport à chaque variable primaire \mathbf{w}, b et ξ , on obtient les conditions d'optimalité suivantes :

$$\begin{cases} \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \\ Ce = \alpha + \beta \end{cases} \quad (6.10)$$

avec $e = \underbrace{[1 \dots 1]}_{n \text{ fois}}^\top$, le vecteur unité.

Sous ces conditions d'optimalité, le problème d'optimisation prend la forme :

$$\begin{cases} \max_{\alpha} & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^n \alpha_i \\ \text{sous contraintes} & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n \\ \text{et} & \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (6.11)$$

Avec les α^* optimaux, l'hyperplan séparateur peut s'écrire sous la forme :

$$h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b = \sum_{i=1}^n \alpha_i^* y_i (\mathbf{x}_i^\top \mathbf{x}) + b \quad (6.12)$$

Ces résultats traduisent une propriété fondamentale dans la théorie des SVM. La troisième condition d'optimalité impose que seuls les points x_s qui sont sur les hyperplans frontière M^\pm tels que $\mathbf{w}^\top x_s + b = \pm 1$ ou à l'intérieur de la marge (de par la marge souple) $-1 < \mathbf{w}^\top x_s + b < 1$ participent à la classification. Ces points sont les vecteurs supports et l'on peut donc aussi en déduire que le niveau de complexité des SVM ne dépend que du nombre de vecteurs supports dans la base d'apprentissage. L'ajout d'un point non vecteur support ne modifiera pas l'hyperplan optimal. De plus, on remarque que la solution ne dépend que du produit scalaire entre les points \mathbf{x}_i et les vecteurs supports x_s , une propriété importante pour les SVM non linéaires expliqués dans la partie suivante.

6.6.2 SVM non linéaire

Les SVM linéaires présentées précédemment ne permettent de résoudre que des problèmes linéairement séparables malgré l'introduction de la marge souple. Afin de résoudre les problèmes non linéairement séparables, l'idée est de se placer dans un espace de plus haute dimension dans lequel il est probable de trouver un hyperplan permettant de séparer linéairement les données. Ce concept est expliqué sur la figure, un problème non linéairement séparable dans un espace à 2 dimensions peut l'être dans un espace à 3 dimensions au moyen d'une transformation non-linéaire ϕ . Ce nouvel espace est appelé espace de redes-

cription.

Le problème d'optimisation dual précédent peut se réécrire en introduisant la fonction noyau ϕ :

$$\left\{ \begin{array}{l} \max_{\alpha} \quad -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} + \sum_{i=1}^n \alpha_i \\ \text{avec} \quad 0 \leq \alpha_i \leq C \\ \text{et} \quad \sum_{i=1}^n \alpha_i y_i = 0. \end{array} \right. \quad i = 1, \dots, n \quad (6.13)$$

L'équation de l'hyperplan séparateur devient donc :

$$h(\mathbf{x}) = \mathbf{w}^{\top} \phi(\mathbf{x}) + b = \sum_{i=1}^n \alpha_i y_i \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}) \rangle_{\mathcal{H}} + b. \quad (6.14)$$

et vérifie toujours $y_i h(\mathbf{x}_i) \geq 0$ comme contrainte de bonne classification. On remarque que la nouvelle formulation de l'hyperplan à travers la fonction ϕ implique un produit scalaire entre vecteurs dans l'espace de redescription. Ce produit scalaire peut être problématique dans la mesure où l'idée est de se placer dans un espace de plus grande dimension, potentiellement très grand, impliquant un coût énorme en termes de temps de calculs. Pour palier à ce problème, on introduit le "*kernel trick*" ou "astuce du noyau", dont l'objectif est de trouver une fonction noyau k qui se comporte comme un produit scalaire dans l'espace de redescription et vérifie ainsi $k = \phi(\mathbf{x}_i)^{\top} \cdot \phi(\mathbf{x})$. Cette astuce rend donc possible le produit scalaire dans l'espace de redescription en ne nécessitant que l'évaluation de la fonction noyau.

On obtient alors le problème d'optimisation final suivant :

$$\left\{ \begin{array}{l} \max_{\alpha} \quad -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^n \alpha_i \\ \text{avec} \quad 0 \leq \alpha_i \leq C \\ \text{et} \quad \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right. \quad i = 1, \dots, n \quad (6.15)$$

et l'expression de l'hyperplan séparateur h devient, en fonction de la fonction noyau :

$$h(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b. \quad (6.16)$$

En pratique on ne connaît pas directement la transformation ϕ , on cherche à construire un noyau k qui doit donc correspondre à un produit scalaire dans l'espace de redescription.

D'après les conditions du théorème de Mercer, le noyau k doit être symétrique et semi-défini positif.

Il existe un certain nombre de noyaux usuels utilisés couramment dans la pratique, on peut citer par exemple :

- le noyau linéaire : $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$
- le noyau polynomial : $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + 1)^d$,
- le noyau gaussien : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$.

Le noyau linéaire n'est en fait qu'une généralisation du SVM linéaire. Dans nos travaux et résultats présentés dans la suite, nous avons utilisés le noyau gaussien dont le paramètre à optimiser est $\gamma = \frac{1}{2\sigma^2}$.

6.6.3 Régression logistique

Lors de nos travaux nous avons utilisé un second classifieur, la régression logistique, comme élément de comparaison au SVM. L'avantage principal de ce classifieur est qu'il permet directement d'estimer des probabilités, donc de sortir un résultat $p(y_i|\mathbf{x}_i) \in [0; 1]$ à l'aide d'une fonction logistique. Le lecteur intéressé pourra se référer au livre [35], fournissant une explication détaillée sur la régression logistique et dont l'explication suivante en est tirée.

Position du problème

Un classifieur de Bayes est basé sur une comparaison de probabilités où une donnée \mathbf{x}_i sera affectée à la classe de plus forte probabilité a posteriori. La règle de décision D pour classer un échantillon \mathbf{x}_i peut ainsi être notée $D(\mathbf{x}_i) = \underset{y \in Y}{\operatorname{arg\,max}} P(y|\mathbf{x}_i)$ avec $y \in Y = y_1, y_2, \dots, y_c$ l'ensemble des classes possibles. D'après les lois de Bayes, il ne nécessite que la connaissance des probabilités a posteriori pour appliquer la règle de décision. Par exemple, dans un cas binaire la règle de décision peut s'exprimer par :

$$D(\mathbf{x}_i) \begin{cases} y_1 & \text{si} & \frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)} > 1 \\ y_2 & \text{sinon} \end{cases} \quad (6.17)$$

Ces probabilités a posteriori peuvent s'exprimer en fonction des probabilités a priori $P(y_i)$ ainsi que des probabilités conditionnelles $P(\mathbf{x}_i|y_i)$:

$$\frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)} = \frac{P(y_1)}{P(y_2)} \cdot \frac{P(\mathbf{x}_i|y_1)}{P(\mathbf{x}_i|y_2)} \quad (6.18)$$

Ainsi, dans le cadre d'un classifieur de Bayes, on estime les lois des probabilités conditionnelles et a priori afin d'estimer les probabilités a posteriori.

Dans le cadre de la régression logistique, on veut estimer directement le rapport des probabilités a posteriori à l'aide d'une modélisation linéaire.

Modèle logistique

Dans le cadre de la régression logistique et pour un problème à K classes, on estime non pas directement $\frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)}$ mais plutôt $\log\left(\frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)}\right)$. L'application du logarithme sur le rapport des probabilités n'affecte pas la règle de décision puisque la fonction log est monotone. Ainsi, on cherche à estimer ce rapport de probabilités par un modèle linéaire prenant la forme :

$$\begin{cases} \log\left(\frac{P(y_i=1|\mathbf{x}_i)}{P(y_i=K|\mathbf{x}_i)}\right) = \beta_{10} + \beta_1^\top x \\ \log\left(\frac{P(y_i=2|\mathbf{x}_i)}{P(y_i=K|\mathbf{x}_i)}\right) = \beta_{20} + \beta_2^\top x \\ \dots \\ \log\left(\frac{P(y_i=K-1|\mathbf{x}_i)}{P(y_i=K|\mathbf{x}_i)}\right) = \beta_{(K-1)0} + \beta_{K-1}^\top x \end{cases} \quad (6.19)$$

Dans le cas d'une classification binaire, on aura $\log\left(\frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)}\right) = \beta_0 + \beta_1^\top x$ et de par la somme à 1 des probabilités $P(y_1|\mathbf{x}_i)$ et $P(y_2|\mathbf{x}_i)$ on peut écrire :

$$\frac{P(y_1|\mathbf{x}_i)}{P(y_2|\mathbf{x}_i)} = \exp^{\beta_0 + \beta_1^\top x} \quad (6.20)$$

et

$$\begin{cases} P(y_1|x) = \frac{\exp^{\beta_0 + \beta_1^\top x}}{1 + \exp^{\beta_0 + \beta_1^\top x}} \\ P(y_2|x) = \frac{1}{1 + \exp^{\beta_0 + \beta_1^\top x}} \end{cases} \quad (6.21)$$

Estimation des paramètres du modèle

On souhaite désormais estimer les paramètres β_i du modèle logistique présenté ci-dessus. Si on note f le nombre de caractéristiques (images, fonctionnelles ou parcimonieuses) extraites pour un voxel $\mathbf{x} \in \mathcal{R}^f$, on souhaite déterminer le vecteur de paramètre $\theta \in \mathcal{R}^f = [\beta_0, \beta_0, \dots, \beta_f]$ permettant de construire le modèle logistique $p(\mathbf{x}) = \beta_0 + \sum_{j=1}^f \beta_j x^j$. Pour cela on utilise une estimation basée sur le maximum de vraisemblance. Si on note $(\mathbf{x}_i)_{1\dots n}$ le nombre de voxels de la base d'apprentissage, on définit la vraisemblance par :

$$\begin{cases} \Lambda(x_1, \dots, x_n) = \prod_{i=1}^n P(y_i, \mathbf{x}_i) \\ \Lambda(\theta, x_1, \dots, x_n) = \prod_{i=1}^n p(y_i|\mathbf{x}_i, \theta) \cdot p_X(\mathbf{x}_i) \end{cases} \quad (6.22)$$

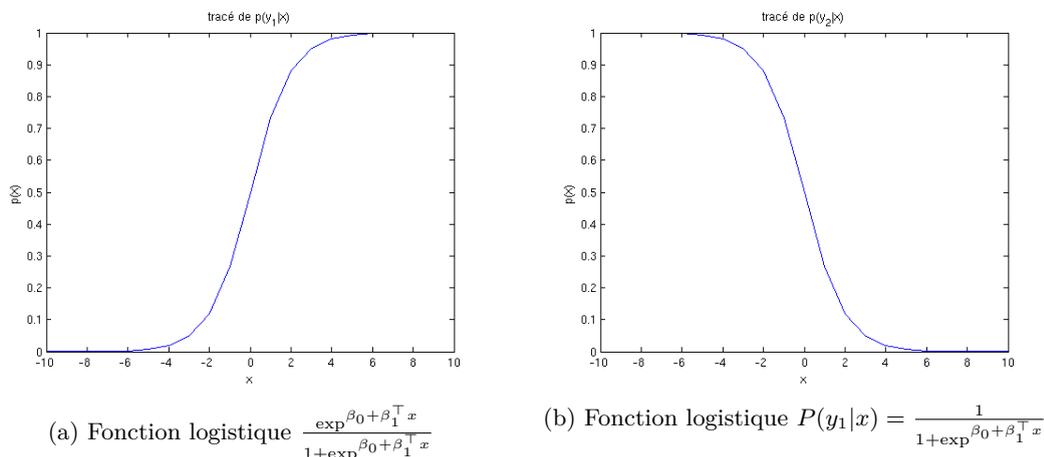


FIGURE 6.8 – Tracé des fonctions logistiques dans un cas de classification à 2 classes

et le log-vraisemblance par :

$$\mathcal{L}(x_1, \dots, x_n) = \sum_{i=1}^n \log(p(y_i|\mathbf{x}_i, \theta)) + \sum_{i=1}^n \log(p_X(\mathbf{x}_i)) \quad (6.23)$$

Le deuxième terme du log-vraisemblance étant indépendant de θ , le vecteur de coefficients θ est déterminé en maximisant le critère $J = \sum_{i=1}^n \log(p(y_i|\mathbf{x}_i, \theta))$. Plusieurs méthodes d'optimisation peuvent être utilisés dont les méthodes de Newton ou la descente de gradient.

Classification

Une fois le vecteur de paramètre θ obtenu, on a l'estimation des probabilités $p(y_1|\mathbf{x}_i) = \frac{\exp(\beta_0 + \sum_{j=1}^f \beta_j x_i^j)}{1 + \exp(\beta_0 + \sum_{j=1}^f \beta_j x_i^j)}$ et $p(y_1|\mathbf{x}_i) = 1 - p(y_0|\mathbf{x}_i)$. On peut alors construire une règle de décision pour \mathbf{x}_i quelconque en l'affectant à la classe y_1 si $p(y_1|\mathbf{x}_i) > 0.5$ ou y_2 sinon.

6.6.4 Classification sur base d'apprentissage déséquilibrée

Comme détaillé dans la partie traitant de la constitution de la vérité terrain, La base d'apprentissage originale est composée de 6 classes de tissus (N), (NS), (GL6), (GL7), (GL8) et (GL9). De ces classes nous avons constitué un problème de classification binaire en concaténant les classes (N), (NS), (GL6) pour obtenir la classe -1 (ou classe des tissus sains ou peu agressifs) et les classes (GL7), (GL8) et (GL9) pour obtenir la classe $+1$ (ou classe des cancers agressifs). Ce problème de classification binaire est intrinsèquement déséquilibré puisque sur les voxels des 49 patients de la base, 401 919 sont de la classe -1 et 56 845 sont de la classe $+1$, ce qui représente un ratio de déséquilibre $\frac{398831}{56905} = \frac{7}{1}$. Un problème majeur lorsque l'on traite des données déséquilibrées est la tendance du classifieur à construire une frontière de décision biaisée vers la classe minoritaire.

Effet du déséquilibre sur la frontière de décision.

La plupart des algorithmes de classification partent du principe que les données sont équilibrées ou ayant des coûts de mauvaise classification égaux. La conséquence pour ces algorithmes est qu'ils échouent à fournir en sortie une distribution représentative des données ainsi qu'un taux de bonne classification convenable et non biaisé. Dans le cas du précis du SVM, il existe plusieurs causes de pertes de performance avec des données déséquilibrées. Premièrement, on peut voir mathématiquement, dans le problème d'optimisation primal 6.7, que minimiser le terme $|\frac{\mathbf{w}}{2}|^2$ est équivalent à maximiser la marge alors que le second terme $C \sum \xi$ minimise l'erreur de classification. C spécifie le compromis entre maximisation de la marge et minimisation de l'erreur. Dans le cas d'un C petit, le SVM apprendra à classer tous les exemples dans la classe majoritaire puisque l'erreur cumulée sera globalement faible. En effet on ne dispose que de peu d'exemples de la classe minoritaire et donc ces derniers n'auront que peu de poids dans la construction de l'hyperplan séparateur. Dans le pire des cas, en présence d'un ratio de déséquilibre très élevé ($> \frac{50}{1}$), un SVM peut apprendre à classer tous les exemples dans la classe majoritaire, considérant ainsi la classe minoritaire comme du bruit, tout en obtenant des taux de classifications élevés. Ensuite, la seconde source de problème est le ratio déséquilibré de vecteurs supports appartenant aux deux classes. En effet, plus ce ratio de déséquilibre augmente et plus on retrouve ce déséquilibre dans les vecteurs supports [106]. Dans ce cas, le voisinage d'un point test à classer proche de la frontière de décision a plus de chance d'être dominé par les vecteurs de supports de la classe majoritaire et ainsi classer ce point dans la classe minoritaire. Néanmoins, de par les conditions de Kuhn-Tucker ($\sum \alpha_i y_i > 0$) (cf. section 6.6.1), la somme des α associés à la classe majoritaire doit être égale à la somme des α associés à la classe minoritaire. Et puisque dans cette situation il y a moins de vecteurs supports de la classe minoritaire que la classe majoritaire, les poids α^+ seront donc plus grand que les α^- de manière à respecter la contrainte de somme à 0. Ainsi cette contrainte aura tendance tout de même à décaler l'hyperplan de manière à prendre en compte le déséquilibre lors de l'apprentissage. Cette propriété explique aussi pourquoi les SVM arrivent à obtenir des performances correctes en présence de jeu de données modérément déséquilibrés comparé à d'autres classifieurs.

Stratégie de sous-échantillonnage aléatoire.

Dans le but de contrer le problème décrit précédemment, Plusieurs approches sont possibles, par exemple utiliser deux couts de mauvaise classification C^+ et C^- des classes majoritaire et minoritaire respectivement en rajoutant du poids sur C^+ de manière à avantager les exemples de la classe minoritaire. Nous avons choisi d'adopter une stratégie de sous-échantillonnage aléatoire en constituant deux nouvelles bases d'apprentissage. La première est le résultat d'un tirage aléatoire de 30 000 voxels pour la classe -1 (10 000 voxels pour chacune des classes (N), (NS) et (GL6)) et 10 000 voxels pour la classe $+1$ (classes (GL7), (GL8) et (GL9)). Cette première base possède donc un ratio de déséquilibre

$\frac{30000}{10000} = \frac{3}{1}$. Le pourcentage de voxels tirés par classe est :

- $\%(N) = \frac{10000}{359321} = 2.7\%$,
- $\%(NS) = \frac{10000}{31763} = 3\%$,
- $\%(GL6) = \frac{10000}{10835} = 99\%$,
- $\%(GL7) = \frac{3333}{33732} = 9.8\%$,
- $\%(GL8) = \frac{3333}{15871} = 21\%$,
- $\%(GL9) = \frac{3333}{7242} = 46\%$,

Malgré son caractère basique, le sous-échantillonnage aléatoire est une méthode utilisée, sous différentes formes et variantes, dans nombres de problèmes de classification nécessitant un rééquilibrage des classes [1]. Celle-ci améliore sensiblement les performances initiales d'un classifieur biaisé par ce déséquilibre mais a surtout l'avantage de réduire la taille du jeu de données initial et donc les temps de calcul.

6.7 Méthode d'évaluation

Dans le but d'évaluer les performances du classifieur, nous avons utilisé les métriques AUC, sensibilité, spécificité et moyenne géométriques décrites dans la section 4.3.6. Ensuite, dans le double objectif de réduire considérablement les temps de calcul et le déséquilibre du nombre d'exemples d'apprentissage par classe, nous avons constitué deux bases issues des 49 mêmes patients. La première conserve la répartition des exemples par classe et la seconde a été obtenue en sous-échantillonnant aléatoirement chacune des classes afin de rééquilibrer les données.

Évaluation sur base d'apprentissage déséquilibrée

Pour cette base, l'apprentissage permettant l'extraction des caractéristiques α_{VCA} , α_{DL} et α_{ssDL} a été effectué sur l'ensemble des 49 patients. Nous avons ensuite réalisé l'apprentissage du modèle des classifieurs suivant un LOPO classique (cf. section 4.3.7), basé sur 35 des 49 patients de la base. Ces 35 patients sont ceux ayant à la fois des voxels de la classe des tissus non agressifs (N, NS et GL6) et des voxels de la classe des tissus agressifs (GL7, GL8 et GL9). Cette base n'a été utilisée qu'en association des classifieurs SVM linéaires et régression logistique linéaire.

Évaluation sur base d'apprentissage équilibrée

Pour constituer la base d'apprentissage équilibrée, nous avons utilisé un sous-échantillonnage aléatoire (cf. section 6.6.4) pour constituer une base de 40 000 voxels issus des 49 patients initiaux. Ici, l'apprentissage permettant l'extraction des caractéristiques α_{VCA} , α_{DL} et α_{ssDL} et l'apprentissage du modèle des classifieurs a été effectuée sur cette base de 40 000 exemples. Dans cette configuration, pour chaque valeur des paramètres à optimiser, nous apprenons une unique fois sur les mêmes 40 000 exemples et testons sur les 35 patients de la base déséquilibrée. Ainsi, certains voxels sont utilisés à la fois pour l'apprentissage et

le test mais ceux-ci ne représentent que 3% du nombre total de voxels. Néanmoins, nous n'avons utilisé ce processus qu'afin de déterminer les paramètres optimaux de classification. En effet, les scores présentés dans le tableau 7.8 sont déterminés en utilisant un LOPO : pour chaque patient test, les éventuels voxels de ce dernier sont retirés de la base de 40 000 exemples puis un apprentissage est réalisé sur le reste avec les paramètres optimaux déterminés précédemment. Cette base n'a été utilisée qu'avec les classifieurs SVM linéaire et non linéaire.

Evaluation des performances des différents schémas de CAD

7.1 Introduction

Ce chapitre présente les résultats de classification obtenus par les différentes combinaisons de CAD sur une base de 49 patients annotés (cf. chapitre 5). Ces résultats sont présentés en deux parties distinctes : l'évaluation sur une base dite déséquilibrée où toutes les données sont utilisées, préservant ainsi la répartition initiale du nombre d'exemples par classe, et sur une base dite équilibrée où le nombre d'exemples par classe est identique. L'objectif de cette seconde base d'apprentissage est à la fois de réduire l'effet du déséquilibre ainsi que d'améliorer les performances du CAD sur des données non-linéairement séparables. Les combinaisons de CAD sont basées sur 4 jeux de caractéristiques (décrites dans la section 6) de type images (descriptives et fonctionnelles) et issues d'une réduction de dimension associés aux classifieurs régression logistique et SVM. L'aire sous la courbe ROC (AUC) est utilisée comme mesure principale de performance et de comparaison.

7.2 Expérimentations

Dans le cadre de ces expérimentations, on considère une tâche de classification binaire, telle que la sortie du classifieur prend les valeurs $y_i \in \{-1; +1\}$ où -1 représente la classe des voxels de type N , NS et Gleason 6 et $+1$ celle des voxels appartenant aux tumeurs agressives, notée $Gleason > 6$. Malgré le caractère multiclassés du problème de classification, nous avons construit un unique classifieur binaire. En effet, l'objectif premier

est d'évaluer si une unique fonction de prédiction est capable d'encoder non seulement la présence d'un cancer mais aussi son degré d'agressivité (score de Gleason). Dans ce but, 3 classifieurs ont été évalués : le SVM linéaire et non linéaire dont les bonnes performances de classification sur des images IRM multiparamétriques de prostate ont été démontrées [64] et la régression logistique permettant de fournir directement en sortie un score de probabilité pour chaque voxel.

Ces 3 classifieurs ont été utilisés en combinaison de 4 types de vecteurs de caractéristiques notés *rawfeat*, α_{VCA} , α_{DL} et α_{ssDL} . Le premier vecteur *rawfeat* représente les 110 caractéristiques issues des 3 séquences IRM T2w, ADC et DCE. Le second vecteur α_{VCA} représente les coefficients de décomposition (abondances) obtenus par le modèle linéaire du VCA (cf. section 1). Ensuite, deux types de vecteurs de coefficients, α_{DL} et α_{ssDL} , sont déterminés par l'intermédiaire de la méthode d'apprentissage de dictionnaires (cf. section 2). α_{DL} représente les coefficients (notés α^k dans la formule 6.1) d'un dictionnaire global appris sur toutes les données sans a priori de classe. Le second vecteur, noté α_{ssDL} , est issu de la concaténation de 4 dictionnaires élémentaires, chacun appris sur les voxels des 4 classes N, NS, Gleason 6 et Gleason >6.

Les expérimentations présentées par la suite sont scindées en deux grandes parties : celles traitant des données déséquilibrées et celles traitant des données équilibrées par échantillonnage aléatoire. Afin de déterminer les paramètres optimaux des classifieurs nous avons utilisées deux approches différentes en fonction de la base d'apprentissage équilibrée ou déséquilibrée. (cf. section 6.7)

Expérimentations sur base d'apprentissage déséquilibrée. Dans cette première partie 8 expérimentations (cf. figure 7.1), basées sur la combinaison de 4 types de vecteurs de caractéristiques *rawfeat*, α_{VCA} , α_{DL} , et α_{ssDL} et des classifieurs SVM linéaire et régression logistique, ont été menées sur base la d'apprentissage déséquilibrée (cf. chapitre 6.6.4). De même le dictionnaire a été estimé sur cette même base déséquilibrée de 455 000 voxels.

Expérimentations sur base équilibrée. Dans cette seconde partie, 6 expérimentations, basées sur les vecteurs de caractéristiques *rawfeat*, α_{DL} et α_{ssDL} et associés aux classifieurs SVM linéaire et non linéaire, ont été menées sur des données réduites (10^3 voxels par classe soit 4×10^3 voxels au total) et partiellement rééquilibrées (ratio=3). Suite aux résultats de la première expérimentation (cf. section 7.6.1) sur les données déséquilibrées, nous avons en effet choisi d'éliminer le vecteur de caractéristiques α_{VCA} , en raison de performances inférieures aux autres vecteurs de caractéristiques. Nous avons de même choisi d'éliminer le classifieur régression logistique en raison de faibles performances en linéaires mais surtout afin d'approfondir la classification non-linéaire en orientant notre choix sur le SVM non linéaire. Les performances de ce dernier sont présentées en regard du SVM linéaire.

Contrairement au SVM linéaire, l'évaluation du SVM non linéaire n'était raisonnablement

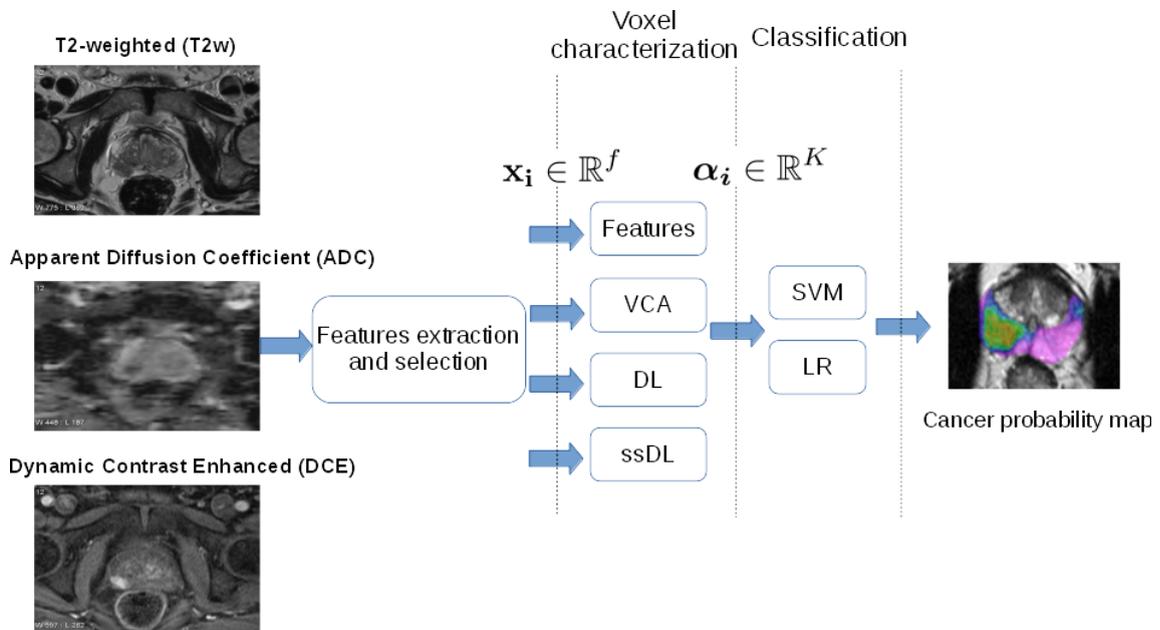


FIGURE 7.1 – Représentation des 8 schémas de CAD évalués à partir de la base d’apprentissage déséquilibrée pour les vecteurs de caractéristiques : $rawfeat$, α_{DL} and α_{ssDL} associés au classifieur SVM linéaire et régression logistique

pas possible en pratique sur l’ensemble des données déséquilibrées en raison du temps de calcul considérable pour l’apprentissage d’un modèle. En effet l’implémentation sous *libsvm* [17] du SVM non-linéaire ne permet pas la gestion de données large échelle. Par exemple, pour de faibles valeurs du paramètre C (< 10), le temps de calcul pour apprendre le modèle du SVM linéaire sur 455 000 exemples est de l’ordre de la seconde alors qu’il faut compter près de 4h pour un SVM non-linéaire. Pour de fortes valeurs de C , le modèle non linéaire n’a pas convergé après 200 heures CPU, rendant donc impossibles les expérimentations couplant SVM non-linéaire et base initiale déséquilibrée.

De même, le dictionnaire a été estimée sur cette même base de 40 000 voxels conduisant donc à des atomes de base et des coefficients différents de celui estimé sur la base déséquilibrée. Dans la suite de ce chapitre les notations α_{DL} et α_{ssDL} correspondent aux vecteurs de caractéristiques issus de la base d’apprentissage déséquilibrée et α_{DL-US} et $\alpha_{ssDL-US}$, où US signifie *undersampled*, à ceux issus de la base équilibrée.

7.3 Caractéristiques sélectionnées

La méthode décrite dans la section 6.4 nous a permis de sélectionner un sous-ensemble optimal de caractéristiques pour les classifieurs SVM linéaire et régression logistique à l’aide du vecteur de caractéristiques $rawfeat$ uniquement. Cette étape de sélection a été réalisée une unique fois sur la base déséquilibrée. Dans les deux cas, le nombre optimal de caractéristiques est $f = 70$. Le tableau A en annexe présente la liste des 70 caractéristiques sélectionnées par les deux classifieurs. On remarque dans un premier temps que celles-ci

sont complètement différentes d'un classifieur à l'autre mais les trois types d'image IRM T2w, ADC et DCE sont sélectionnées. Cela valide l'approche multiparamétrique pour les systèmes d'aide à la décision pour la détection du cancer de la prostate. On peut aussi remarquer pour le classifieur SVM, celui donnant les meilleurs résultats (cf. section 7.6.2), que les paramètres de texture sont les plus sélectionnés.

La figure 7.2 présente l'évolution des poids normalisés (par rapport à la valeur maximale) et triés par ordre décroissant. On remarque qu'environ 90% de la valeur des poids sont contenus dans les 11 premières caractéristiques. Malgré cette tendance, le nombre optimal obtenu décrit dans la section 7 est $f=70$ caractéristiques.

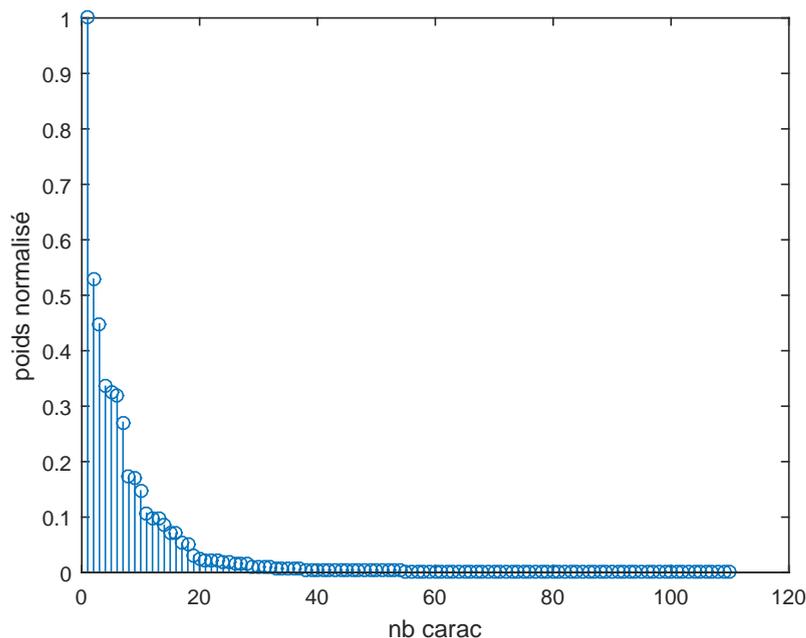


FIGURE 7.2 – Représentation du poids normalisé des caractéristiques dans le processus de sélection. Ces poids sont obtenus en moyennant la valeur absolue des poids des caractéristiques obtenus pour chaque étape du LOPO pour le schéma de CAD basé sur *rawfeat* associé au classifieur SVM linéaire. Ces poids sont normalisés en valeur maximale.

7.4 Optimisation des paramètres des classifieurs

7.4.1 Paramètres optimaux obtenus sur base d'apprentissage déséquilibrée

Dans le but de déterminer le nombre optimal K d'éléments de base du dictionnaire, la valeur du paramètre de régularisation λ (parcimonie), le nombre K de endmembers du VCA, le coût C de mauvaise classification du SVM linéaire et le paramètre de régularisation λ_{RL} de la fonction de minimisation du log-likelihood de la régression logistique, une recherche en grille a été lancée avec pour critère de maximisation l'AUC. Cette recherche

en grille est réalisée par le biais d'un LOPO (cf. section 4.3.7). La courbe ROC est calculée sur la base d'une classification voxel à voxel, signifiant donc que le voxel est classé dans la classe positive si le score fourni par le CAD dépasse un certain seuil de décision indiquant la position du voxel dans une région d'intérêt de la classe positive.

Dans le cas du VCA, le nombre optimal K de endmembers a été déterminé dans un intervalle de 1 à 15 par pas de 1 élément. Concernant la première méthode d'apprentissage de dictionnaires donnant les vecteurs de caractéristiques α_{DL} , Nous avons fait varier la taille K_{DL} du dictionnaire entre 5 et 60 atomes de base par pas de 5 atomes. Comme expliqué dans la section 2, la seconde méthode d'apprentissage de dictionnaires donnant les vecteurs α_{ssDL} apprend séparément un nombre fixé de dictionnaires pour chacune des classes N , NS , Gleason 6 et $Gleason > 6$. Le dictionnaire final est construit par concaténation des éléments appris sur chaque classe. Le nombre de dictionnaire est fixé à k pour les classes NS, Gleason 6 et Gleason >6 puis à $3k$ pour la classe N dans le but de prendre en compte a priori la plus forte variabilité de ce groupe et ainsi de mieux représenter les différents constituants présents dans les tissus sains. L'entier k varie de 1 à 4 de telle sorte que la taille globale K_{ssDL} du dictionnaire varie de 6 à 24 atomes. Dans les deux cas précédents, le paramètre de régularisation λ a été choisi dans un intervalle compris entre $[10^{-5}; 10^1]$. Enfin, le paramètre C du SVM a été choisi dans un intervalle compris entre $[10^{-5}; 10^5]$. Le tableau 7.1 synthétise les paramètres optimaux déterminés.

Vecteurs de caractéristiques	Classifieur	paramètres optimaux
$rawfeat$	SVM-L	$C = 1e^3$
	RL	$\lambda_{RL} = 1e^{-6}$
α_{VCA}	SVM-L	$K=6, C=1e^3$
	RL	$K=8, \lambda_{RL} = 1e^{-3}$
α_{DL}	SVM-L	$K=60, \lambda = 1e^{-5}, C = 10e^3$
	RL	$K=30, \lambda = 0.13, \lambda_{RL} = 1e^{-6}$
α_{ssDL}	SVM-L	$K=[9 \ 3 \ 3 \ 3], \lambda = 1e^{-5}, C = 1e^3$
	RL	$K=[12 \ 3 \ 3 \ 3], \lambda = 3.98, \lambda_{RL} = 1e^{-3}$

TABLE 7.1 – Paramètres optimaux des schémas de CAD pour la base d'apprentissage déséquilibrée

7.4.2 Paramètres optimaux obtenus sur base d'apprentissage équilibrée

Les paramètres optimaux des classifieurs sur la base d'apprentissage équilibrée ont été déterminées suivant la méthodologie décrite dans la section 6.7. L'étape d'extraction de caractéristiques basée sur les deux méthodes d'apprentissage de dictionnaire a été répétée afin d'extraire les deux vecteurs α_{DL-US} et $\alpha_{ssDL-US}$. Les valeurs de C , λ et K ont été déterminées dans les mêmes plages de valeurs que dans la section précédente pour la base d'apprentissage déséquilibrée, les valeurs du noyau gaussien γ a été sélectionné dans l'intervalle compris entre $[2^{-8}; 2^8]$.

Vecteurs de caractéristiques	Classifieur	paramètres optimaux
<i>rawfeat</i>	SVM-L	$C=7e^4$
	SVM-NL	$C=1.29e3, \gamma=0.0098$
α_{DL}	SVM-L	$K=50, \lambda = 0.0398, C=1e^{-4}$
	SVM-NL	$K=50, \lambda = 0.01, C=100, \gamma=0.12$
α_{ssDL}	SVM-L	$K=[12\ 4\ 4\ 4], \lambda = 2.15e^{-4}, C=10e4$
	SVM-NL	$K=[9\ 6\ 6\ 6], \lambda = 0.3162, C=100, \gamma=0.25$

TABLE 7.2 – Paramètres optimaux des schémas de CAD pour la base d’apprentissage équilibrée

7.5 Analyse des performances des différentes combinaisons de CAD

Génération d’une carte de probabilité du cancer de la prostate. L’optimisation des paramètres des classifieurs (cf. section 6.7) a permis de déterminer les meilleures configurations pour les différentes combinaisons de CAD : 8 basées sur le SVM linéaire et régression logistique appris sur la base d’apprentissage déséquilibrée avec les vecteurs *rawfeat*, α_{VCA} , α_{DL} et α_{ssDL} et 6 autres basées sur les SVM linéaire et non-linéaire mais sur la base d’apprentissage équilibrée avec les vecteurs *rawfeat*, α_{DL-US} et $\alpha_{ssDL-US}$. Ensuite, la meilleure configuration de paramètres a permis de construire une carte de probabilité de cancer où chaque valeur de voxel est la probabilité d’appartenance à la classe pathologique. Cette probabilité est obtenue en appliquant un modèle de Platt [73] à la distance à l’hyperplan fournie par la sortie du SVM ou de la régression logistique.

Métriques et test de signe. Les performances de ces schémas de CAD ont été évaluées à l’aide de métriques incluant, sensibilité (*Se*), spécificité (*Spe*), moyenne géométrique (*gmean*) et l’aire sous la courbe ROC (*AUC*). Ces métriques sont détaillées dans la section 4.3.6 et ont été calculées voxel à voxel pour chacun des 35 patients test et ensuite moyennées sur les 35 patients. Un test d’hypothèse non paramétrique, le test de signe bilatéral, a été conduit de manière à comparer les performances de deux schémas de CAD, en terme d’AUC obtenue pour chacun des 35 patients. Le niveau de significativité pour rejeter l’hypothèse nulle a été fixé à une p-valeur=0.05. Nous présentons ensuite un tableau de résultats de probabilités moyennes par ROI 3D et par classe. Dans cette optique, les valeurs de probabilités moyennes par coupe (ROI 2D) ont été utilisées afin de calculer un score sur l’ensemble de la lésion pouvant être présente sur plusieurs coupes. Le score final présenté est la moyenne de toutes ces probabilités par ROI 3D et par classe.

Effet du tirage aléatoire sur la reproductibilité des résultats. Dans le but d’évaluer l’impact sur les performances de classification de la méthode de sous-échantillonnage aléatoire utilisée pour générer la base de 40 000 voxels, nous avons conduit un test visant à générer 10 bases aléatoires de 40 000 voxels puis de lancer une même simulation basée

sur un SVM non linéaire et les vecteurs de caractéristiques $rawfeat$, α_{DL-US} et $\alpha_{ssDL-US}$. Les paramètres optimaux sont les mêmes que ceux obtenus dans la section précédente. L'écart-type et la moyenne ont été calculés sur les 10 résultats d'AUC et pour chaque combinaison.

Corrélation entre score de Gleason et score de probabilité. Finalement, bien que les expérimentations de classification binaire menées ont pour but de séparer les cancers agressifs des autres tissus, nous avons étudié si les valeurs de probabilités fournies par les différents schémas de CAD peuvent être indicateurs de l'agressivité du cancer. Ceci a été réalisé en moyennant les valeurs de probabilité dans une chacune des régions d'intérêts contourées par le radiologue puis en extrayant des statistiques du premier ordre sur ces valeurs de probabilité par région en fonction de leur classe pathologique (N,NS,Gleason 6 et Gleason >6). Cette analyse a été conduite sur la base équilibrée, produisant les meilleurs résultats, pour les SVM linéaires et non linéaires associés aux vecteurs $rawfeat$, α_{DL-US} et $\alpha_{ssDL-US}$.

7.6 Résultats et comparaison des performances

7.6.1 Classification sur base d'apprentissage déséquilibrée

Choix des illustrations. Tout au long de ce chapitre, l'analyse des cartes de probabilités de cancer a été réalisée sur 3 patients dont on a sélectionné une coupe parmi les 24 disponibles. Le premier patient présente deux cancers agressifs, de grade Gleason 8 en orange et Gleason 7 en jaune, et a été choisi pour illustrer la manière dont les différents schémas de CAD se sont comportés pour différencier deux cancers agressifs de grades proches. Le second patient choisi présente un cancer agressif de grade maximal (Gleason 9) présent dans nos données et une région suspecte mais bénigne (NS), de manière à mettre en avant le comportement des CAD face à ce que l'on attendrait comme deux extrêmes en termes de scores de probabilités du cancer. Enfin, le troisième patient présente un cancer non agressif (Gleason 6), un tissu suspect mais bénin (NS) et un cancer agressif (Gleason 8) et a été sélectionné afin d'analyser la plage de probabilités obtenues face à deux régions non agressives (NS et Gleason 6) et un cancer agressif (Gleason 8) et ainsi sa capacité à les discerner individuellement.

Analyse des cartes de probabilités de 3 patients. Les figures 7.3 et 7.4 illustrent des exemples de cartes de probabilité de cancer, (colonnes 2 à 4) superposées sur des coupes transverses IRM T2w de 3 patients, issues des schémas de CAD dont les vecteurs de caractéristiques sont $rawfeat$, α_{VCA} , α_{DL} et α_{ssDL} associés au *SVM linéaire* (figure 7.3) et *Régression logistique linéaire* (figure 7.4). Les valeurs de probabilité moyenne (probabilité sur ROI 2D correspondant à la coupe présentée) à l'intérieur de chacune de ces régions illustrées sur les cartes sont synthétisées dans le tableau 7.3.

Vecteur	Classifieur	Patient 1			Patient 2		Patient 3		
		NS	GL 7	GL 8	NS	GL 9	NS	GL 6	GL 8
Rawfeat	SVM-L	0.16	0.4	0.55	0.25	0.51	0.29	0.4	0.4
	RL	0.26	0.35	0.39	0.25	0.34	0.32	0.39	0.37
VCA	SVM-L	0.2	0.4	0.5	0.21	0.36	0.32	0.49	0.45
	RL	0.25	0.35	0.39	0.24	0.39	0.32	0.39	0.37
DL	SVM-L	0.22	0.41	0.55	0.25	0.51	0.25	0.41	0.36
	RL	0.33	0.32	0.34	0.3	0.34	0.27	0.36	0.36
ssDL	SVM-L	0.21	0.34	0.51	0.21	0.34	0.21	0.34	0.31
	RL	0.27	0.27	0.28	0.28	0.28	0.26	0.29	0.28

TABLE 7.3 – Score de probabilité moyen par ROI 2D (issu d’une unique coupe) des 3 patients illustrés sur les figures 7.3 et 7.4. Valeurs correspondantes à la base déséquilibrée.

Pour les cartes de probabilités associées au SVM linéaire, les deux premiers patients illustrent la bonne performance des CAD pour discriminer le cancer agressif (Gleason >6) des autres classes (N , NS et Gleason 6). L’analyse visuelle des cartes paramétriques indiquent que les trois lésions (Gleason 7, Gleason 8 pour le patient 1 et Gleason 9 pour le patient 2) ont une plus forte probabilité d’être des cancers agressifs que la lésions de classe NS . De même, la lésion de grade Gleason 9 apparait plus suspecte que les lésions Gleason 7 et Gleason 8 respectivement. Cette tendance est confirmée par les valeurs de probabilité moyenne (cf tableau 7.3) à l’intérieur de chaque ROI 2D : pour le patient 1, les scores moyens des classes des tumeurs de grade Gleason 7 et Gleason 8 sont systématiquement supérieurs à ceux de la classe des tissus NS sauf pour le classifieur RL associé au vecteur α_{DL} . Pour le patient 2, de même, les probabilités moyennes de la classe Gleason 9 sont toute supérieures à celle de la classe NS sauf pour le vecteur α_{ssDL} et régression logistique. Pour ces 2 premiers patients le vecteur *rawfeat* associé au SVM linéaire obtiennent les meilleurs performances.

Le troisième patient illustre la difficulté des CAD à discriminer distinctement les classes NS , Gleason 6 et Gleason 8. Ceci est confirmé par le tableau 7.3 : L’écart maximal de probabilité entre la classe NS et Gleason 6 est obtenu par le vecteur α_{DL} et le SVM linéaire (0.25 vs 0.41). En revanche, il est impossible de distinguer Gleason 8 du Gleason 6 puisque toutes les probabilités moyennes du Gleason 8 sont inférieurs ou égales à celle du Gleason 6 et autour de 0.4. On remarque donc que le Gleason 6 est significativement surestimé.

On peut remarquer que les classifieurs n’obtiennent au mieux que des probabilités aux alentours de 0.5 et ne permettent pas de mettre suffisamment en avant les lésions. Une possible explication de ces faibles performances peut provenir de l’hypothèse de distribution linéaire des données que les classifieurs tentent de modéliser. Cette hypothèse de linéarité pourrait ne pas être valide pour nos données représentées un espace de dimension f . Ce résultat pourrait donc souligner une limitation des modèles linéaires.

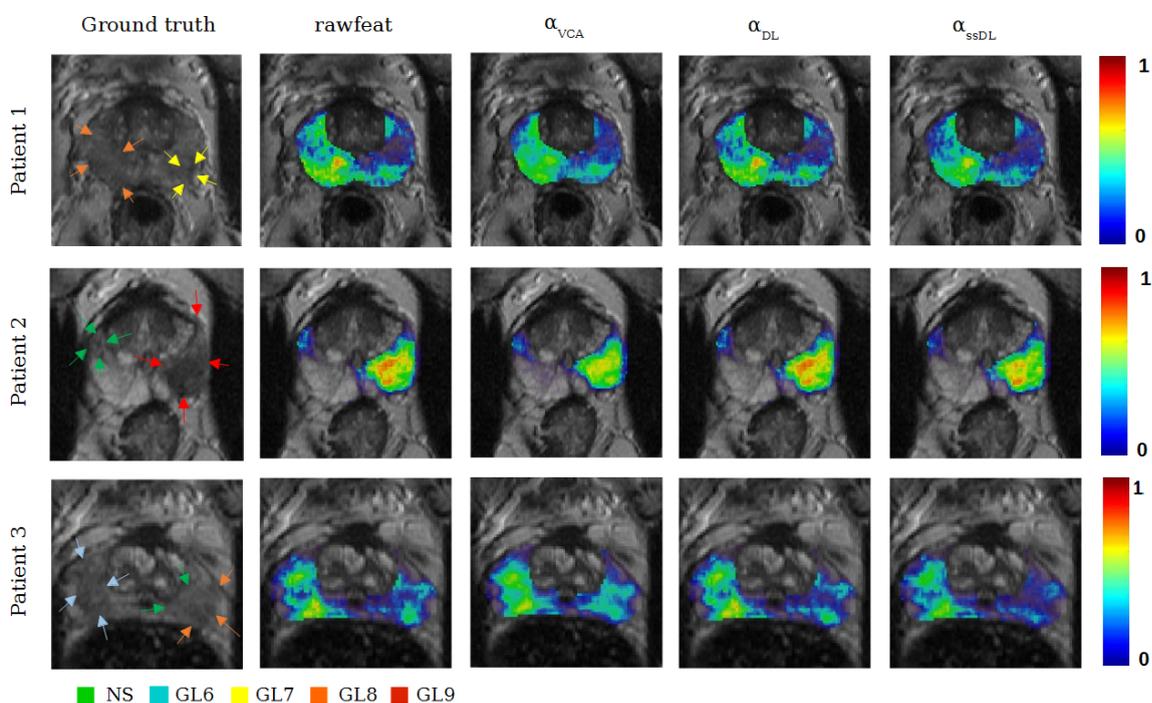


FIGURE 7.3 – Exemples de cartes de probabilités générées à partir de la base d’apprentissage déséquilibrée avec les vecteurs de caractéristiques : $rawfeat$, α_{VCA} , α_{DL} et α_{ssDL} associés au classifieur SVM linéaire

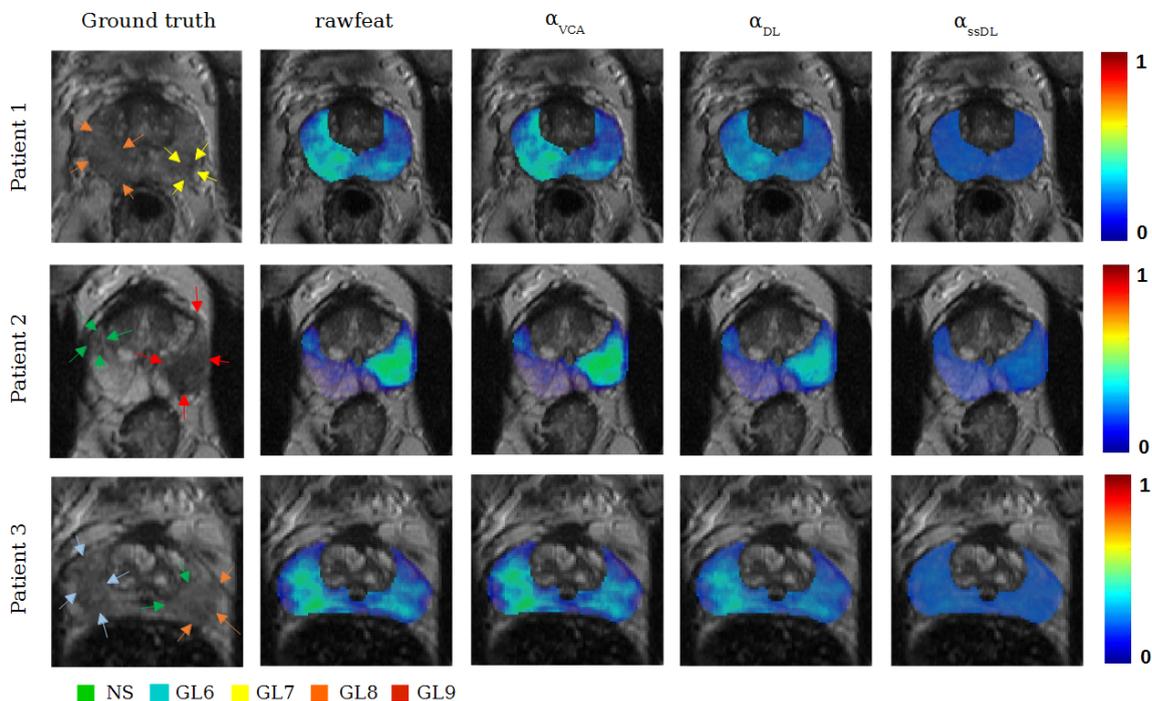


FIGURE 7.4 – Exemples de cartes de probabilités générées à partir de la base d’apprentissage déséquilibrée avec les vecteurs de caractéristiques : $rawfeat$, α_{VCA} , α_{DL} et α_{ssDL} associés au classifieur Régression logistique

Analyse générale des performances. Le tableau 7.4 présente les résultats de classification des 8 schémas de CAD basés sur les vecteurs de caractéristiques $rawfeat$, α_{VCA} , α_{DL} et α_{ssDL} associés aux classifieurs *SVM linéaire* et *régression logistique linéaire*. Toutes les méthodes obtiennent des scores d'AUC compris entre 0.77 et 0.79, les meilleurs scores de 0.79 sont obtenus pour les vecteurs de caractéristiques $rawfeat$, et α_{DL} associés au SVM linéaire. La valeur de sensibilité résulte de l'application d'un seuil de zéro sur les fonction de décision des deux classifieurs. On constate qu'elle est relativement basse ($0.06 \leq Se \leq 0.40$) comparée aux scores de spécificité ($0.15 \leq Spe \leq 0.98$). Cette tendance est très probablement causée par le déséquilibre des classes (ratio 7 :1) lorsque la classe majoritaire est la classe négative. Les classifieurs ont tendance à être biaisés naturellement vers la classe majoritaire (cf. section 6.6.4). Un test de signe (cf. tableau 7.4) entre toutes les paires de schémas de CAD indique que $rawfeat$ associé au SVM linéaire est plus performant que α_{VCA} associé au classifieur RL et α_{ssDL} associés aux deux classifieurs SVM et RL. On peut de même remarquer que α_{DL} associé au classifieur RL est significativement meilleur que α_{VCA} ainsi que $\alpha_{DL}+SVM-L$ et $\alpha_{ssDL}+SVM-L$ sont significativement meilleurs que $\alpha_{ssDL}+RL$. Toutes les autres combinaisons de CAD ne produisent aucune différence significative sur les résultats.

Vecteur de caractéristiques	Classifieur	Se	Spe	Gmean	AUC
$rawfeat$	SVM-L	0.40 ± 0.24	0.89 ± 0.10	0.55 ± 0.20	0.79 ± 0.11
	RL	0.40 ± 0.24	0.89 ± 0.10	0.55 ± 0.20	0.78 ± 0.11
α_{VCA}	SVM-L	0.06 ± 0.11	0.98 ± 0.05	0.17 ± 0.17	0.77 ± 0.11
	RL	0.11 ± 0.14	0.89 ± 0.06	0.25 ± 0.20	0.77 ± 0.11
α_{DL}	SVM-L	0.39 ± 0.23	0.90 ± 0.10	0.55 ± 0.20	0.79 ± 0.11
	RL	0.39 ± 0.23	0.90 ± 0.10	0.55 ± 0.20	0.78 ± 0.11
α_{ssDL}	SVM-L	0.33 ± 0.24	0.92 ± 0.09	0.50 ± 0.22	0.78 ± 0.11
	RL	0.33 ± 0.24	0.92 ± 0.09	0.50 ± 0.22	0.78 ± 0.11

TABLE 7.4 – Performances de classification sur base d'apprentissage déséquilibrée

X	$rawfeat$ SVM-L	$rawfeat$ RL	α_{VCA} SVM-L	α_{VCA} RL	α_{DL} SVM-L	α_{DL} RL	α_{ssDL} SVM-L	α_{ssDL} RL
$rawfeat$ SVM-L	X	0.175	0.311	0.090	0.090	0.311	0.017	0.002
$rawfeat$ RL	0.175	X	0.09	0.041	0.736	0.5	1.0	0.175
α_{VCA} SVM-L	0.311	0.090	X	0.500	0.500	0.041	0.500	0.736
α_{VCA} RL	0.090	0.041	0.5	X	0.090	0.041	0.311	1.0
α_{DL} SVM-L	0.090	0.736	0.500	0.09	X	0.5	0.175	0.001
α_{DL} RL	0.311	0.5	0.041	0.041	0.500	X	1.000	0.175
α_{ssDL} SVM-L	0.017	1.0	0.5	0.311	0.175	1.0	X	0.017
α_{ssDL} RL	0.002	0.175	0.736	1.0	0.001	0.175	0.017	X

TABLE 7.5 – Tableau de résultats du test de signe entre chaque combinaison de CAD associés au SVM linéaire et régression logistique sur base déséquilibrée. Les valeurs présentées sont les p-valeurs et les celles en gras sont significatives pour un seuil de 0.05.

Analyse des performances par probabilité moyenne dans chacune des lésions. Le tableau 7.6 présente les scores de probabilités moyens par ROI 3D et par classe. La première constatation ici est que le classifieur régression logistique obtient quasi systématiquement des scores inférieurs à ceux du SVM linéaire pour les tumeurs agressives (Gleason

>6). Si l'on peut constater une faible tendance linéaire des scores de probabilités entre les classes NS, Gleason 6, Gleason 7 et Gleason 8, il est en revanche très difficile de distinguer la classe Gleason 9 de la classe Gleason 8. D'après ce tableau, la meilleure combinaison de CAD semble être le SVM linéaire associé au vecteur α_{DL} dont les probabilités moyennes par classe sur l'ensemble des ROIs 3D sont 0.29, 0.30, 0.42, 0.47 et 0.50 pour les classes NS, Gleason 6, Gleason 7, Gleason 8 et Gleason 9 respectivement.

Vecteur de caractéristiques	Classifieur	NS	Gleason 6	Gleason 7	Gleason 8	Gleason 9
$rawfeat$	SVM-L	0.29 ± 0.12	0.32 ± 0.16	0.42 ± 0.14	0.50 ± 0.17	0.47 ± 0.13
	RL	0.30 ± 0.05	0.30 ± 0.06	0.34 ± 0.06	0.38 ± 0.08	0.36 ± 0.08
α_{VCA}	SVM-L	0.30 ± 0.11	0.31 ± 0.13	0.39 ± 0.12	0.47 ± 0.17	0.44 ± 0.16
	RL	0.30 ± 0.05	0.30 ± 0.07	0.34 ± 0.06	0.39 ± 0.09	0.36 ± 0.08
α_{DL}	SVM-L	0.29 ± 0.10	0.30 ± 0.14	0.42 ± 0.14	0.47 ± 0.18	0.50 ± 0.20
	RL	0.29 ± 0.03	0.30 ± 0.03	0.32 ± 0.03	0.34 ± 0.05	0.34 ± 0.05
α_{ssDL}	SVM-L	0.28 ± 0.10	0.28 ± 0.13	0.38 ± 0.13	0.44 ± 0.17	0.46 ± 0.17
	RL	0.28 ± 0.01	0.28 ± 0.01	0.28 ± 0.01	0.29 ± 0.01	0.28 ± 0.02

TABLE 7.6 – Score de probabilité moyen par classe des ROIs 3D sur base d'apprentissage déséquilibrée

De par ces premières constatations où l'hypothèse de linéarité sur les données semble ne pas être adéquate, nous avons choisi de nous concentrer pour la suite des travaux sur une hypothèse de non linéarité exploitée avec le SVM non linéaire sur une base équilibrée. Bien qu'il aurait été de même possible d'explorer la piste du classifieur régression logistique non linéaire nous avons décidé d'exclure ce classifieur dans la suite de nos travaux en raison de premiers résultats décevants. Enfin nous avons exclu le vecteur α_{VCA} de par ses performances en deçà des autres vecteurs de caractéristiques.

7.6.2 Classification sur base d'apprentissage équilibrée

Analyse de cartes de probabilités de 3 patients pour le SVM linéaire et non-linéaire. Les résultats présentés ici sont issus des 3 mêmes patients précédent. La figure 7.5 présente des cartes de probabilités issues des schémas de CAD dont les vecteurs de caractéristiques sont $rawfeat$, α_{DL-US} et $\alpha_{ssDL-US}$ associés au *SVM linéaire*. La comparaison avec la figure 7.3 indique que ces cartes de probabilité semblent similaires avec celles obtenues avec les schémas de CAD sur jeu de données déséquilibrées. Une faible diminution des performances peut néanmoins être observée, notamment pour la détection de la lésion de grade Gleason 8 du patient 3 et la lésion Gleason 7 du patient 1. Le taux de déséquilibre des classes ne semble donc pas prédominant avec le SVM linéaire.

La figure 7.6 illustre les cartes de probabilités obtenues avec les vecteurs de caractéristiques $rawfeat$, α_{DL-US} et $\alpha_{ssDL-US}$ associés au *SVM non-linéaire*. Une première observation concerne les moins bonnes performances du vecteur de caractéristiques $\alpha_{ssDL-US}$ basé sur un apprentissage semi-supervisé de dictionnaires comparé aux autres schémas. En effet cette configuration échoue à détecter la lésion de grade Gleason 7 sur le patient 1 alors que les lésions NS et Gleason 6 des patients 2 et 3 respectivement sont visuellement apparentés

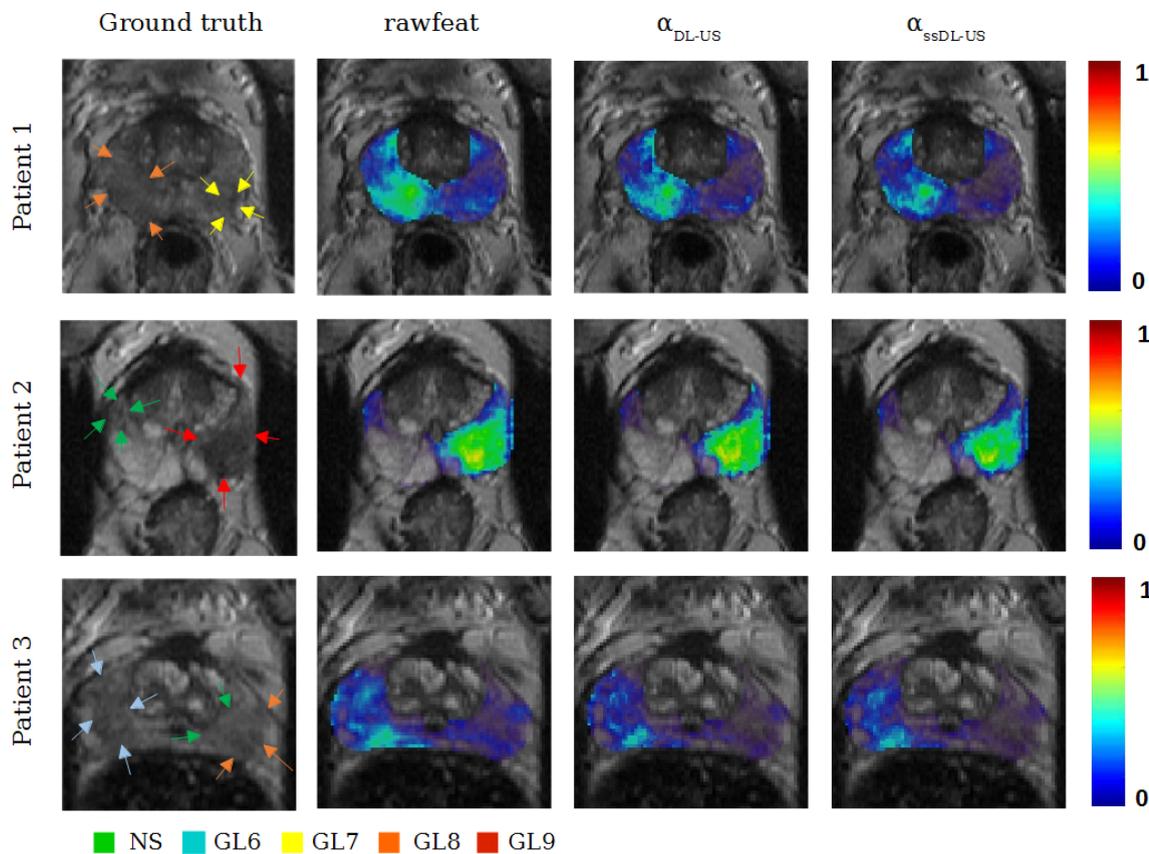


FIGURE 7.5 – Exemples de cartes de probabilités générées à partir de la base d’apprentissage équilibrée avec les vecteurs de caractéristiques : $rawfeat_{US}$, α_{DL-US} and $\alpha_{ssDL-US}$ associés au classifieur SVM linéaire

à des cancers. Le tableau 7.7 permet d’observer que les deux schémas de CAD basés sur les vecteurs de caractéristiques $rawfeat$, α_{DL-US} détectent correctement les cancers agressifs (Gleason >6) avec des probabilités élevées : 0.80 ($rawfeat$) et 0.74 (α_{DL-US}) pour la lésion de grade Gleason 9, 0.85 et 0.81 pour la lésion Gleason 8 du patient 1, 0.73 et 0.58 pour la lésion Gleason 8 du patient 3 puis 0.56 et 0.62 pour la lésion Gleason 7. Les valeurs de probabilités correspondantes aux lésions NS sont < 0.05 . Contrairement aux autres schémas de CAD basés sur le SVM linéaire (cf. figures 7.3 et 7.5), ces deux schémas de CAD détectent correctement la lésion Gleason 8 du patient 3 comme un cancer tout en classifiant correctement les lésions NS et Gleason 6 dans la classe des tissus sains. Cette tendance démontre de même que les différents types de cancers ne sont probablement linéairement séparables dans l’espace de départ. En conclusion, on peut remarquer que les schémas de CAD basés sur $rawfeat$ et α_{DL-US} associés au SVM non-linéaire obtiennent de très bonnes performances dans la tâche de séparation des lésions agressives (Gleason >6) du reste (N,NS,Gleason 6).

Analyse générale des performances. Le tableau 7.8 présente les résultats de classification des 6 schémas de CAD basés sur les vecteurs de caractéristiques $rawfeat$, α_{DL-US}

Vecteur	Classifieur	Patient 1			Patient 2		Patient 3		
		NS	GL 7	GL 8	NS	GL 9	NS	GL 6	GL 8
Rawfeat	SVM-L	0.15	0.28	0.44	0.19	0.32	0.20	0.23	0.25
	SVM-NL	0.04	0.56	0.85	0.05	0.8	0.07	0.01	0.73
DL	SVM-L	0.15	0.26	0.48	0.19	0.35	0.20	0.23	0.22
	SVM-NL	0.01	0.62	0.81	0.02	0.74	0.03	0.01	0.58
ssDL	SVM-L	0.19	0.31	0.45	0.20	0.34	0.17	0.22	0.23
	SVM-NL	0.01	0.32	0.45	0.03	0.31	0.02	0.01	0.22

TABLE 7.7 – Score de probabilité moyen par ROI 2D (issu d’une unique coupe) des 3 patients illustrés sur les figures 7.5 et 7.6. Valeurs correspondantes à la base équilibrée.

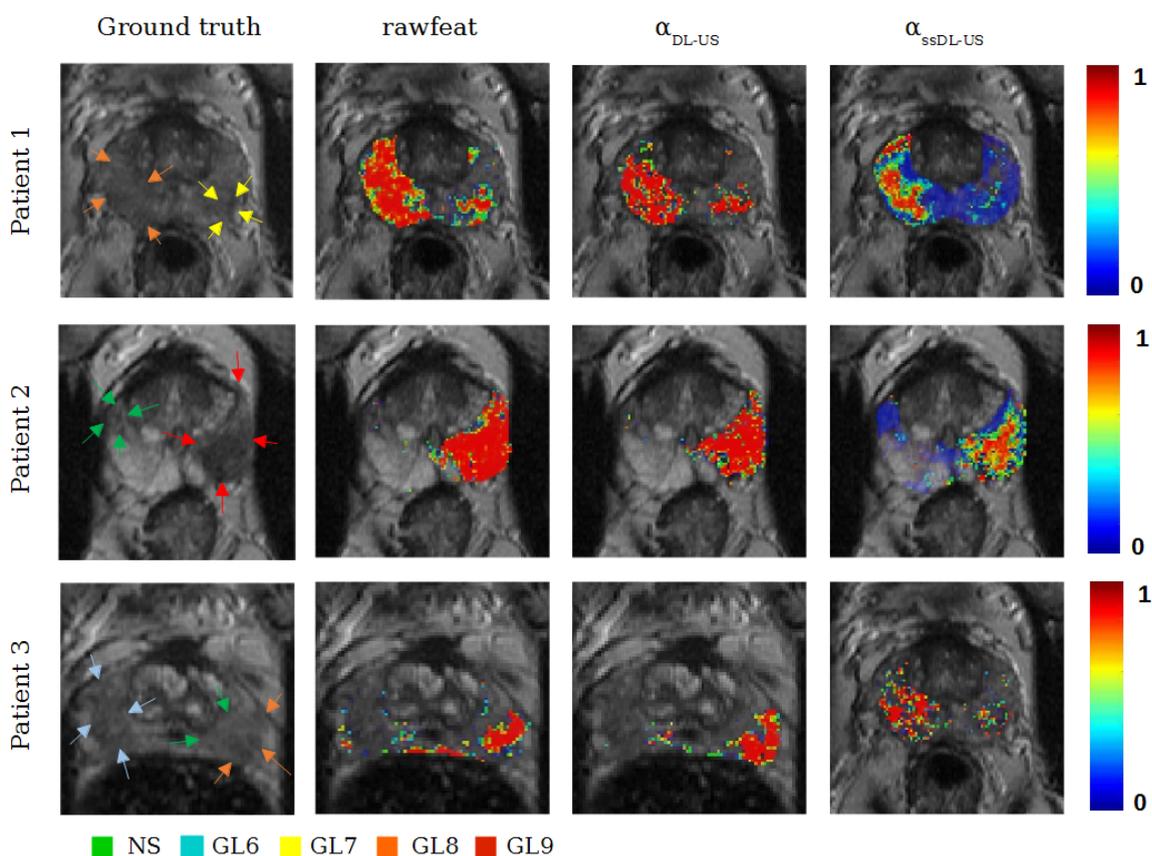


FIGURE 7.6 – Exemples de cartes de probabilités générées à partir de la base d’apprentissage équilibrée avec les vecteurs de caractéristiques : $rawfeat$, α_{DL-US} and $\alpha_{ssDL-US}$ associés au classifieur SVM non linéaire

et $\alpha_{ssDL-US}$ et les classifieurs SVM linéaire et non-linéaire. La combinaison du vecteur α_{DL-US} avec le SVM non-linéaire présente les meilleures performances de classification avec une valeur d’AUC proche de 0.92. Le test de signe basé sur l’AUC (cf tableau 7.9) confirme que cette configuration présente statistiquement de meilleures performances que tous les autres schémas de CAD (p-valeurs < 0.001). La comparaison des performances obtenues avec le SVM linéaire pour les données déséquilibrées (cf. tableau 7.4) et équilibrées

(cf. tableau 7.8) montre que la stratégie de sous-échantillonnage aléatoire pour ré-équilibrer les classes ne permet pas d’obtenir de meilleures performances (valeur moyenne d’AUC de 0.78 et 0.79 pour les bases équilibrées et déséquilibrées respectivement) et ce quelque soit le vecteur de caractéristiques. Le ratio entre sensibilité (valeur moyenne < 0.1) et spécificité (> 0.98) est de plus diminué comparé au tableau 7.4 des performances sur données déséquilibrées. Cependant, les performances obtenues avec le SVM non-linéaire par la suite sont bien meilleures que celles obtenues avec le SVM linéaire, quelque soit le vecteur de caractéristiques, laissant donc penser que la distribution non-linéaire des caractéristiques est la principale difficulté ne pouvant être résolue par le SVM linéaire même en présence de données équilibrées. Ce résultat est confirmé par l’analyse statistique basée sur le test de signe (p-valeurs < 0.001) excepté pour le vecteur $\alpha_{ssDL-US}$ qui obtient des performances statistiquement similaires pour les 3 configurations basées sur le modèle linéaire du SVM.

Vecteurs de caractéristiques	Classifieur	Se	Spe	Gmean	AUC
$rawfeat_{US}$	SVM-L	0.11 ± 0.15	0.98 ± 0.05	0.25 ± 0.20	0.78 ± 0.12
	SVM-NL	0.74 ± 0.10	0.86 ± 0.05	0.79 ± 0.06	0.88 ± 0.04
α_{DL-US}	SVM-L	0.04 ± 0.08	0.99 ± 0.03	0.14 ± 0.15	0.78 ± 0.12
	SVM-NL	0.73 ± 0.07	0.94 ± 0.03	0.83 ± 0.03	0.92 ± 0.03
$\alpha_{ssDL-US}$	SVM-L	0.10 ± 0.11	0.98 ± 0.04	0.25 ± 0.18	0.78 ± 0.12
	SVM-NL	0.32 ± 0.11	0.97 ± 0.54	0.10 ± 0.07	0.80 ± 0.07

TABLE 7.8 – Performances de classification sur base d’apprentissage équilibrée

X	rawfeat SVM-L	rawfeat SVM-NL	α_{DL} SVM-L	α_{DL} SVM-NL	α_{ssDL} SVM-L	α_{ssDL} SVM-NL
rawfeat SVM-L	X	2.09e-09	0.0018	5.82e-11	1e-3	0.04
rawfeat SVM-NL	2.09e-09	X	2.23e-05	3.67e-08	3.67e-08	5.82e-11
α_{DL} SVM-L	0.0018	2.23e-05	X	2.09e-09	1.00	1.00
α_{DL} SVM-NL	5.82e-11	3.67e-08	2.09e-09	X	5.82e-11	5.82e-11
α_{ssDL} SVM-L	1e-3	3.67e-08	1.00	5.82e-11	X	0.73
α_{ssDL} SVM-NL	0.04	5.82e-11	1.00	5.82e-11	0.73	X

TABLE 7.9 – Tableau de résultats du test de signe entre chaque combinaison de CAD associés au SVM non linéaire sur base équilibrée. Les valeurs présentées sont les p-valeurs et les celles en gras sont significatives pour un seuil de 0.05.

Illustration de la détection du cancer sur 3 coupes consécutives. La figure 7.7 présente, pour un même patient et sur 3 coupes consécutives, les cartes de probabilités obtenues à l’aide des vecteurs de caractéristiques $rawfeat_{US}$, α_{DL-US} et $\alpha_{ssDL-US}$ associés au classifieur SVM non-linéaire. Dans cet exemple, le cancer de grade *Gleason9* apparaît sur la vérité terrain des coupes 16 et 17. Il n’y a en revanche pas de cancer avéré sur la vérité terrain de la coupe 18. On remarque que le schéma de CAD basé sur le vecteur $\alpha_{ssDL-US}$ échoue à faire apparaître la lésion quelque soit la coupe étudiée. En revanche, une lésion est détectée sur les 3 coupes avec le vecteur $rawfeat_{US}$ et uniquement sur les coupes 16 et 17 pour le vecteur α_{DL-US} .

Il est intéressant de noter qu’une large zone uniforme de probabilités proches de 1 (zone rouge) est mise en valeur par le vecteur $rawfeat_{US}$ sur les 3 coupes. Une lésion apparaît

néanmoins sur la coupe 18, en désaccord avec la vérité terrain. Le CAD basé sur le vecteur α_{DL-US} affiche une détection dont les probabilités sont proche de 1 sur les coupes 16 et 17 mais beaucoup plus bruitée dans la zone de la lésion. En revanche, en accord avec la vérité terrain, il ne fait pas apparaître de lésion sur la coupe 18. Sur cet exemple on peut conclure que le vecteur $rawfeat_{US}$ semble plus sensible que le vecteur α_{DL-US} mais au détriment d'une fausse détection de lésion sur la coupe 18.

On peut de même s'intéresser à la spécificité des deux schémas de CAD précédents, notamment leur capacité à mettre en avant la forme de la lésion conformément à la vérité terrain. Sur la coupe 18, on remarque que le CAD basé sur le vecteur α_{DL-US} met avant une lésion dont la forme est très proche de la vérité terrain alors que la forme mise en avant par le vecteur $rawfeat_{US}$ est beaucoup plus grossière. Dans ces conditions, le vecteur α_{DL-US} semble plus spécifique que le vecteur $rawfeat_{US}$.

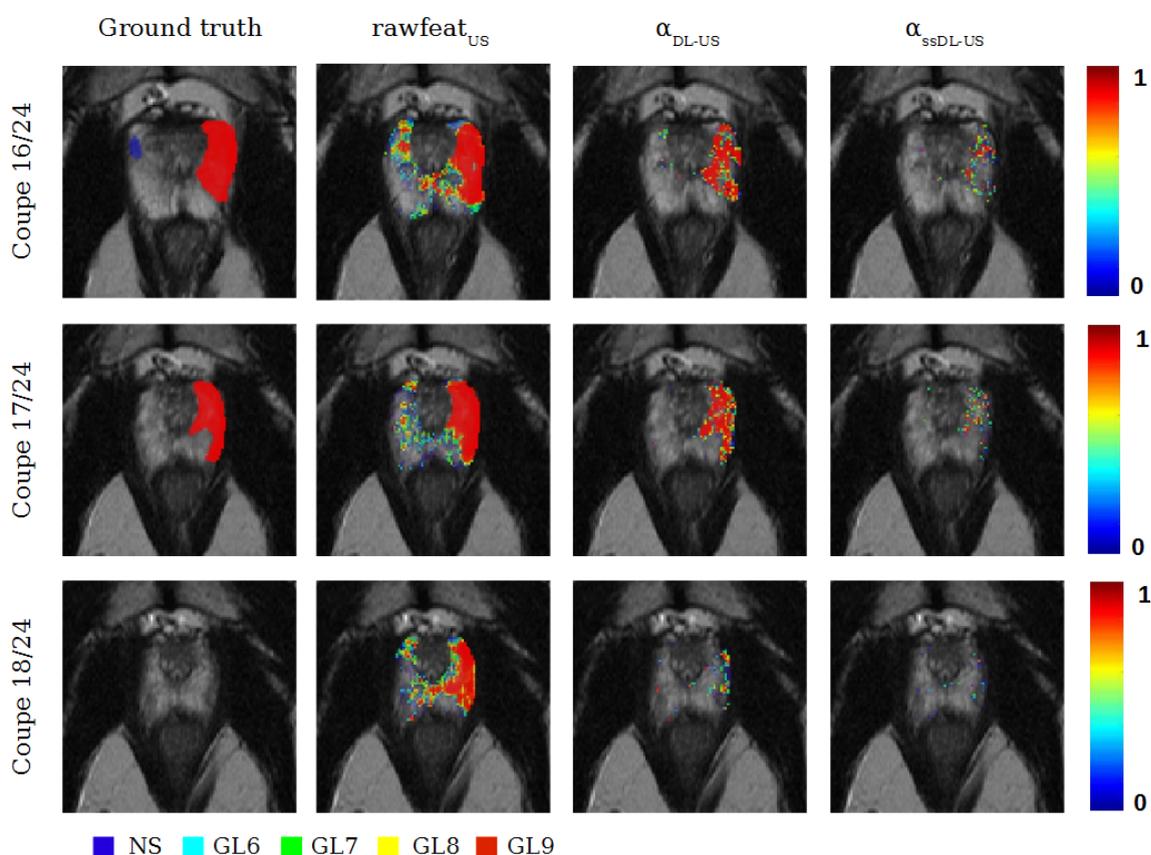


FIGURE 7.7 – Illustration de cartes de probabilités générées sur 3 coupes consécutives (16, 17 et 18 sur 24 coupes) avec les vecteurs de caractéristiques $rawfeat_{US}$, α_{DL-US} and $\alpha_{ssDL-US}$ associés au classifieur SVM non-linéaire. La colonne *Ground truth* correspond à la vérité terrain.

Analyse des performances par probabilité moyenne dans chacune des lésions.

Le tableau 7.10 présente les scores de probabilités moyens par ROI 3D et par classe sur la base équilibrée. De manière analogue au tableau 7.6, chaque valeur présentée dans ce

tableau est la moyenne des scores obtenus sur chacune des ROIs 3D et par classe. Les premières observations confirment les conclusions du paragraphe précédent : le SVM non linéaire permet d’obtenir des valeurs moyennes de probabilités permettant de beaucoup mieux discriminer tissus agressifs et tissus non agressifs. Ainsi les deux meilleures combinaisons de CAD sont le SVM non linéaire associé aux vecteurs *rawfeat*, α_{DL-US} . Les scores moyens obtenus sont 0.09, 0.07, 0.72, 0.76 et 0.79 pour le vecteur *rawfeat* et 0.04, 0.03, 0.63, 0.68 et 0.79 pour le vecteur α_{DL-US} (score pour les classes NS, Gleason 6, Gleason 7, Gleason 8 et Gleason 9 respectivement). De même pour ces deux meilleures combinaisons de CAD, bien que l’on note une nette différence dans les probabilités entre cancers agressifs (Gleason >6) et reste des tissus (NS, Gleason 6), il est difficile de discerner individuellement chacun des tissus puisque l’écart maximal de probabilité est de 0.11 entre Gleason 7 et Gleason 9 pour le vecteur α_{DL-US} .

Vecteur de caractéristiques	Classifieur	NS	Gleason 6	Gleason 7	Gleason 8	Gleason 9
<i>rawfeat</i>	SVM-L	0.22 ± 0.08	0.24 ± 0.10	0.32 ± 0.09	0.39 ± 0.14	0.37 ± 0.07
	SVM-NL	0.09 ± 0.06	0.07 ± 0.06	0.72 ± 0.09	0.76 ± 0.13	0.79 ± 0.07
α_{DL}	SVM-L	0.23 ± 0.07	0.24 ± 0.09	0.33 ± 0.10	0.40 ± 0.17	0.44 ± 0.13
	SVM-NL	0.04 ± 0.06	0.03 ± 0.05	0.63 ± 0.07	0.68 ± 0.17	0.79 ± 0.06
α_{ssDL}	SVM-L	0.24 ± 0.08	0.25 ± 0.09	0.32 ± 0.09	0.38 ± 0.13	0.40 ± 0.09
	SVM-NL	0.04 ± 0.04	0.03 ± 0.04	0.30 ± 0.08	0.27 ± 0.13	0.41 ± 0.13

TABLE 7.10 – Score de probabilité moyen par classe des ROIs 3D sur base d’apprentissage équilibrée

Effet du tirage aléatoire sur la reproductibilité des résultats. Dans le but d’estimer la reproductibilité de la méthode de sous-échantillonnage aléatoire, nous avons évalués la performance en AUC des schémas de CAD basés sur le SVM non-linéaire avec 10 séries de 40 000 exemples d’apprentissage tirés aléatoirement dans la base originale de 455000 points d’apprentissage (cf. section 5.2). Les valeurs moyennes d’AUC sont de 0.88 ± 0.002 , 0.92 ± 0.002 , 0.79 ± 0.003 pour les vecteurs *rawfeat*, α_{DL-US} et $\alpha_{ssDL-US}$ respectivement. La comparaison avec les performances obtenues dans le tableau 7.8 confirme que la sélection aléatoire de 40 000 points permet de générer une base suffisante à modéliser la distribution des données.

7.6.3 Corrélation entre probabilités et score de Gleason

Les résultats précédents montrent que le meilleur schéma de CAD pour discriminer les cancers agressifs des tissus sains, sains mais suspects et cancer non agressifs est celui associant le vecteur de caractéristiques α_{DL-US} , issu de l’apprentissage de dictionnaires, associé au SVM non linéaire. Dans cette partie on cherche à déterminer si les probabilités fournies par les différentes sorties des CAD sont corrélées avec le score de Gleason, et plus précisément avec les cancers agressifs (Gleason >6). La figure 7.8 présente les boxplots de différentes combinaisons de CAD sur la base équilibrée. La construction d’un boxplot est réalisée en moyennant dans un premier les scores de probabilités à l’intérieur de chacune

des ROI 2D de classe NS et Gleason 6 à Gleason 9. Ainsi, l'abscisse correspond aux classes des tissus et l'ordonnée aux score de probabilités moyens par ROI 2D. Un boxplot est représenté par un rectangle bleu dont les parties extrêmes correspondent aux 25e et 75e percentiles et la ligne horizontale rouge à la médiane. Les points en dehors de ce rectangle bleu sont considérés comme des valeurs aberrantes, c'est à dire en dehors d'une distribution suivant une loi normale à $\pm 2.7\sigma$.

Concernant les deux schémas de CAD donnant les meilleurs performances, ie *rawfeat* et α_{DL-US} associés au SVM non linéaire, les boxplots montrent très nettement le pouvoir discriminant du CAD à séparer cancer agressifs (Gleason >6) des autres tissus. De plus, Les valeurs de probabilités montrent une légère tendance à l'augmentation en fonction des grades de Gleason, notamment pour les lésions Gleason 7, Gleason 8 et Gleason 9 avec des valeurs médianes de 0.73, 0.81 et 0.80 pour *rawfeat* et 0.7, 0.72 et 0.78 pour α_{DL-US} . Cette différence n'est cependant pas statistiquement significatif. La transformation non-linéaire appliquée aux caractéristiques par le biais du noyau gaussien du SVM non-linéaire ne garantie pas qu'un échantillon d'un tissu agressif sera projeté plus loin de la frontière de décision qu'un tissu cancer moins agressif.

7.7 Conclusion

Comparaison aux CADe pour la détection du cancer. Ce chapitre présente les résultats de différents schémas de système d'aide à la décision dont le but est de discriminer les cancers agressifs (Gleason >6) des cancers non agressifs (Gleason 6), tissus sains (N) et suspects (NS). La meilleur configuration de CAD obtient un score d'AUC de 0.92. Elle est basée sur un classifieur SVM non linéaire associé à un vecteur de caractéristiques issu de l'apprentissage d'un dictionnaire sur une série de données statistiques et semi-quantitative extraites d'images IRM T2w, ADC et DCE. La comparaison avec l'état de l'art des CADe montre que nos performances rivalisent favorablement avec les systèmes d'aide au diagnostic proposés par [16] (AUC=0.83), [100] (AUC=0.82), [2] (AUC=0.79), [67] (AUC=0.82), [102] (AUC=0.89), [93] (AUC=0.89) et [25] (AUC=0.86). Cependant les comparaisons entre études restent difficiles étant données les différences de type et de taille de base d'images IRM ainsi que les différentes métriques et méthodes d'évaluation pour l'analyse des performances.

Comparaison à l'état de l'art des CADe pour l'évaluation de l'agressivité du cancer. Récemment, sur une problématique similaire à la notre pour la caractérisation du cancer, [51] et [101] ont présenté des performances de diagnostic en fonction de l'agressivité du cancer. [101] ont proposé un CAD à plusieurs étapes dont la première sélectionne les régions suspectes dans l'image à l'aide d'un détecteur de formes circulaires suivi d'une classification utilisant l'analyse linéaire discriminante (LDA) en 2 autres étapes. Le CAD classe des clusters de points et sa performance est évaluée à l'aide d'une courbe FROC nécessitant ainsi une règle de labellisation des clusters à classifier [72]. Leur CAD montre

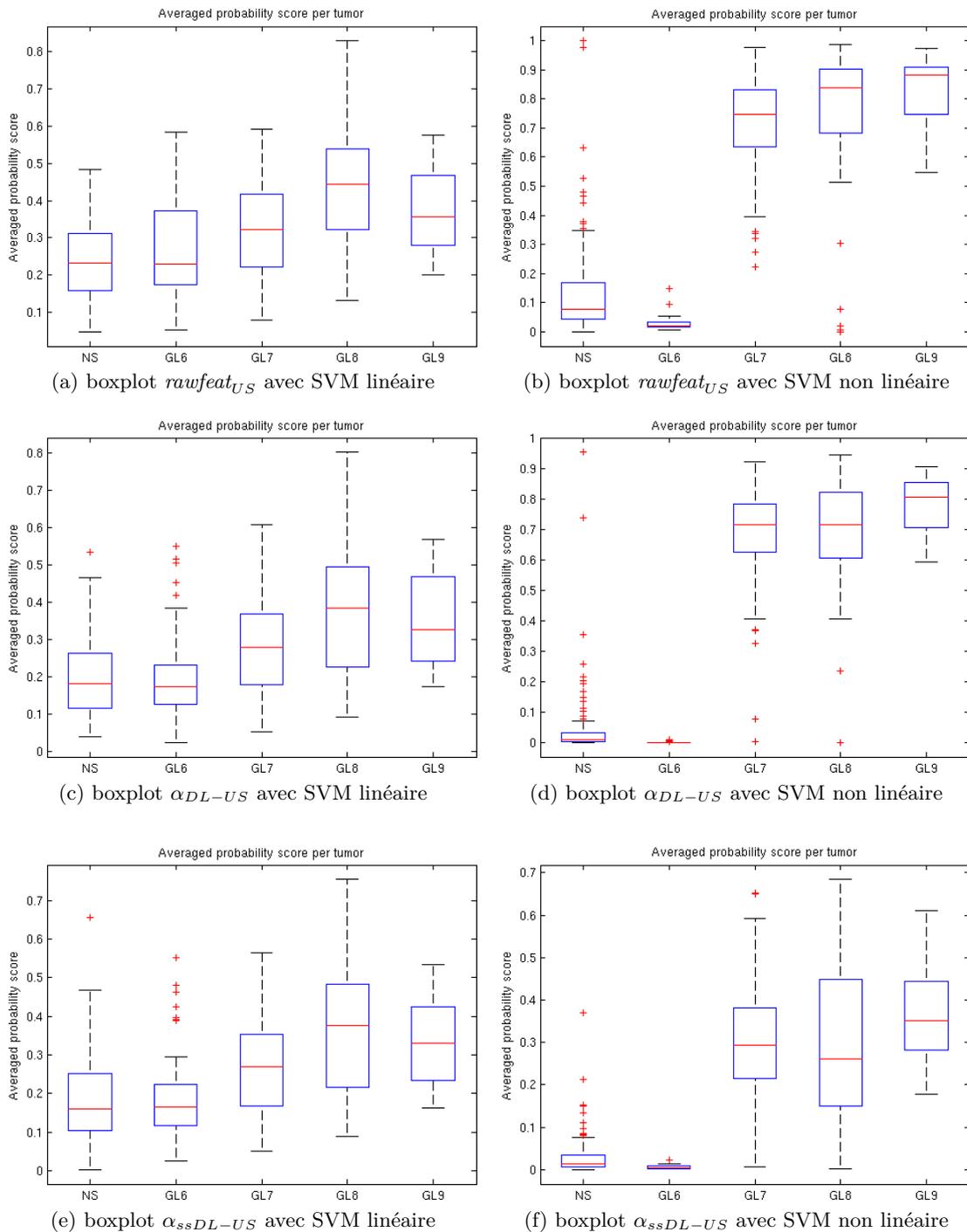


FIGURE 7.8 – Boxplots représentant en ordonnées les valeurs de probabilité dans les tumeurs et les classes de tissus en abscisses pour la base d’apprentissage équilibrée (ratio 3). La colonne de gauche correspond au SVM linéaire et celle de droite au SVM non linéaire. Les 3 lignes correspondent aux 3 types de vecteurs de caractéristiques : $rawfeat_{US}$, α_{DL-US} et $\alpha_{ssDL-US}$

de bonnes performances pour détecter des tumeurs de différents score de Gleason (Se=0.74

pour un taux de 5 faux positifs par patient). Leur sensibilité de détection des cancers agressifs (Gleason >6) et de grade intermédiaire (Gleason 6) est en effet très élevée ($Se=0.88$ et $Se=0.74$ respectivement) mais bien moindre ($Se=0.6$) pour le reste (Gleason <6). [51] ont récemment évalué un CADe à deux étages avec pour double objectif la détection de tout type de cancers et cancers agressifs (Première ou seconde partie du score de Gleason >3). Leur CADe fonctionne de même avec une première étape de détection de zones suspectes avant d'être envoyées à une seconde étape de classification. Ils obtiennent des scores d'AUC de 0.81 et 0.83 pour la détection de tout type de cancers et de cancers agressifs (Première ou seconde partie du score de Gleason >3) respectivement. La comparaison avec les résultats de notre CADe à un seul étage reste difficile de par la différence dans les métriques utilisées pour l'évaluation (voxel à voxel, par région ou cluster). Il est néanmoins possible de les comparer à l'étude de [51] après la première étape de classification voxel à voxel. Ils obtiennent des score d'AUC de 0.89, donc en deçà de nos performances ($AUC=0.92$). Ainsi, notre étude pourrait s'apparenter à une très bonne étape de sélection de voxels pour un CADe multi-étage. Comme mentionné par [51], la performance finale du CADe est fortement impactée par cette première étape de sélection de voxels.

Analyse de la méthode d'apprentissage de dictionnaire et perspectives d'amélioration. Notre meilleur résultat de CADe ($AUC=0.92$) est obtenu avec le vecteur de caractéristiques α_{DL-US} provenant d'une méthode d'apprentissage de dictionnaires. Bien qu'elle soit statistiquement meilleure que tous les autres schémas de CADe, y compris celui avec le vecteur $rawfeat_{US}$, cette différence n'est pas clairement visible sur les exemples de cartes de probabilité de la figure 7.6. Cette figure met aussi en avant les moins bonnes performances du schéma de CADe basé sur le vecteur $\alpha_{ssDL-US}$, lui aussi basé sur la même méthode d'apprentissage de dictionnaire. Une possible explication réside dans le fait que les éléments du dictionnaire ont été appris séparément sur chacune des classes (N, NS, Gleason 6 et Gleason >6) puis concaténés, faisant ainsi potentiellement apparaître des atomes de base corrélés. Cette corrélation pourrait expliquer le plus faible pouvoir discriminant des coefficients $\alpha_{ssDL-US}$. Une manière d'améliorer les performances obtenues avec la méthode d'apprentissage de dictionnaires serait de restreindre les éléments du dictionnaire à être tous orthogonaux les uns aux autres, en particulier pour la méthode semi supervisée basée sur $\alpha_{ssDL-US}$, afin de limiter la corrélation entre les éléments du dictionnaire. Une autre approche prometteuse serait de regrouper l'étape d'apprentissage du dictionnaire avec celle du classifieur dans une seule et même fonction d'optimisation comme pratiqué par [54]. De plus, ces dictionnaires sont estimés à partir des caractéristiques images déjà extraites, qui ont prouvé leur pouvoir discriminant pour les tâches de classification cancer VS non cancer et cancer agressif VS reste des tissus. Néanmoins, elles semblent ne pas capter précisément les différences entre cancers. Une idée serait d'apprendre un dictionnaire overcomplete (nombre d'atomes \gg nombre de caractéristiques) directement sur des patches des images IRM de manière à possiblement capter des informations plus fines sur l'agressivité du cancer.

Perspectives pour la corrélation des scores de probabilité avec le Gleason. Le principal objectif de cette thèse était de construire un CADe capable de discriminer cancers agressifs et lésions suspectes. Le second objectif était d'évaluer si une unique fonction de prédiction était capable d'encoder non seulement la certitude de la présence d'un cancer mais aussi son degré d'agressivité (score de Gleason). Les résultats de boxplots affichés sur la figure 7.8 ainsi que sur les cartes de probabilités de la figure 7.6 indiquent que les valeurs de probabilités des 3 cancers agressifs (Gleason 7, Gleason 8 et Gleason 9) sont très proches et donc peu discernables les uns des autres. Un travail supplémentaire est nécessaire afin de raffiner ce modèle. Puisque le modèle est considéré comme ordinal (les classes sont rangés des tissus normaux aux tissus agressifs), il serait intéressant de tester un modèle de régression logistique multiclassés. Néanmoins, cette approche nécessiterait de se poser la question de comment fusionner les différentes cartes de probabilités obtenues pour chacune des classes de manière à améliorer la performance de diagnostic du radiologue.

Evaluation clinique du CADe. Dans nos précédents travaux [64], un CADx avait été évalué par des radiologues de manière à quantifier l'apport de ce CADx et avait clairement mis en avant l'amélioration des performances de diagnostic lorsque ce dernier est utilisé en second avis afin de confirmer la présence d'une zone suspecte. Ce même travail doit être mis en œuvre sur ce CADe afin de quantifier son apport et surtout de déterminer à quel moment utiliser ce CADe, c'est-à-dire en tant qu'outil de sélection des voxels suspect ou comme dans le cas précédent pour confirmer ou infirmer un premier diagnostic.

III Recalage d'images et classification

Création et recalage d'une base d'images IRM 3T

8.1 Introduction

Ce chapitre présente les résultats de travaux de recalage de séquences IRM 3T Coefficient apparent de diffusion (ADC), Réhaussement de contraste dynamique (DCE) et B2000 sur la séquence IRM T2-pondérée (T2-w). Bien qu'au cours de cette thèse nos travaux ont été effectués sur des IRM 1.5T, nous avons initié en parallèle la construction d'une nouvelle base IRM 3T, enrichie de la séquence de diffusion B2000. Ces nouvelles données seront intégrées dans la base CLARA-P, après avoir été annotées selon le protocole décrit dans le chapitre 5. La résolution spatiale étant améliorée par rapport aux images de la base 1.5T, il est apparu important d'évaluer l'intérêt du recalage des différentes séquences d'un même patient sur une image de référence (l'image T2-w) afin de corriger d'éventuels mouvements au cours de l'acquisition ainsi que certains phénomènes de distorsions.

Il est présenté dans un premier temps la problématique de la classification supervisée sur des données non recalées ainsi que la méthodologie générale de construction d'un système de recalage. Nous présenterons ensuite le principe général de la méthode de recalage implémentée par le logiciel ProMap, développé par le société Koelis, que nous avons utilisé et répondant particulièrement bien au recalage de séquences IRM présentant de fortes hétérogénéités de contraste et d'aspect. Enfin, nous terminerons par une analyse détaillée de ses performances sur une base de 40 patients.

8.2 Problématique du recalage en classification par approche voxel

Dans le cadre des systèmes d'aide au diagnostic pour le cancer de la prostate, Il existe deux approches de classification : l'approche ROI visant à attribuer une classe à une région ou objet, sous-entendu un ensemble agrégé de voxels, ou l'approche que nous avons suivie, celle visant à classer le voxel lui-même. Dans le cadre d'un examen IRM multi-modalités de la prostate, le patient est soumis à un protocole d'acquisition pouvant durer jusqu'à 30 minutes. Ce temps d'acquisition introduit invariablement du mouvement et par conséquent un non alignement spatial des séquences IRM.

A l'origine de ces mouvements, on notera celui du patient sur la table d'acquisition mais aussi le remplissage de sa vessie ou encore l'afflux d'air dans le rectum. On constate alors généralement une déformation des bords de la prostate ou un simple décalage de référence sur les images IRM. De plus, on constate une distorsion géométrique de la prostate de part la nature même des séquences IRM, notamment les séquences de diffusion (ADC et B200).

De par ces principales variabilités d'acquisition, la comparaison d'un même voxel d'une image à une autre n'est pas forcément aisée. Ainsi, suivant la séquence IRM prise pour référence afin de constituer la vérité terrain, un même voxel peut se voir attribuer deux classes différentes et ainsi rendre moins précise la tâche de classification notamment aux frontières des zones d'intérêt. Le modèle de la classification peut alors s'en trouver perturbé et générer des cartes de probabilités de cancer bruitées : il est donc généralement préconisé de recalibrer les séquences IRM entre elles avant la phase d'apprentissage.

Néanmoins, il est important de nuancer l'impact du recalage dans le cadre de notre CADE actuel. Ce dernier n'étant pas dans une problématique de détection de petites lésions qui pourraient ne pas apparaître à cause d'un décalage des images de quelques voxels. Nous nous sommes plutôt concentrés jusqu'à maintenant sur la détection de tendances et l'effet principal du recalage concerne les voxels situés en périphérie de la tumeur.

La problématique du recalage est illustrée sur la figure 8.1. Les 3 séquences IRM T2w, ADC et DCE imagent la prostate du même patient durant le protocole d'acquisition. Sur l'image IRM T2w prise comme référence, sont tracés un landmark circulaire (noté L1) représentant un kyste et un landmark curvilinéaire (noté L6) délimitant le bord postérieur de la prostate. Ces 2 landmarks sont directement reportés sur l' image ADC et DCE respectivement, c'est-à-dire aux mêmes coordonnées que le T2w de manière à étudier la correspondance spatiale entre les landmarks.

Pour le landmark L1, on voit une nette délimitation sur l'image T2w alors que sur l'image ADC, les frontières du kyste sont nettement moins visibles. Ce landmark illustre la variabilité de visibilité d'une même zone anatomique sur deux séquences. Enfin, le landmark L6 illustre un mouvement au niveau postérieur de la prostate, on constate ainsi un non-alignement des bords de la prostate.

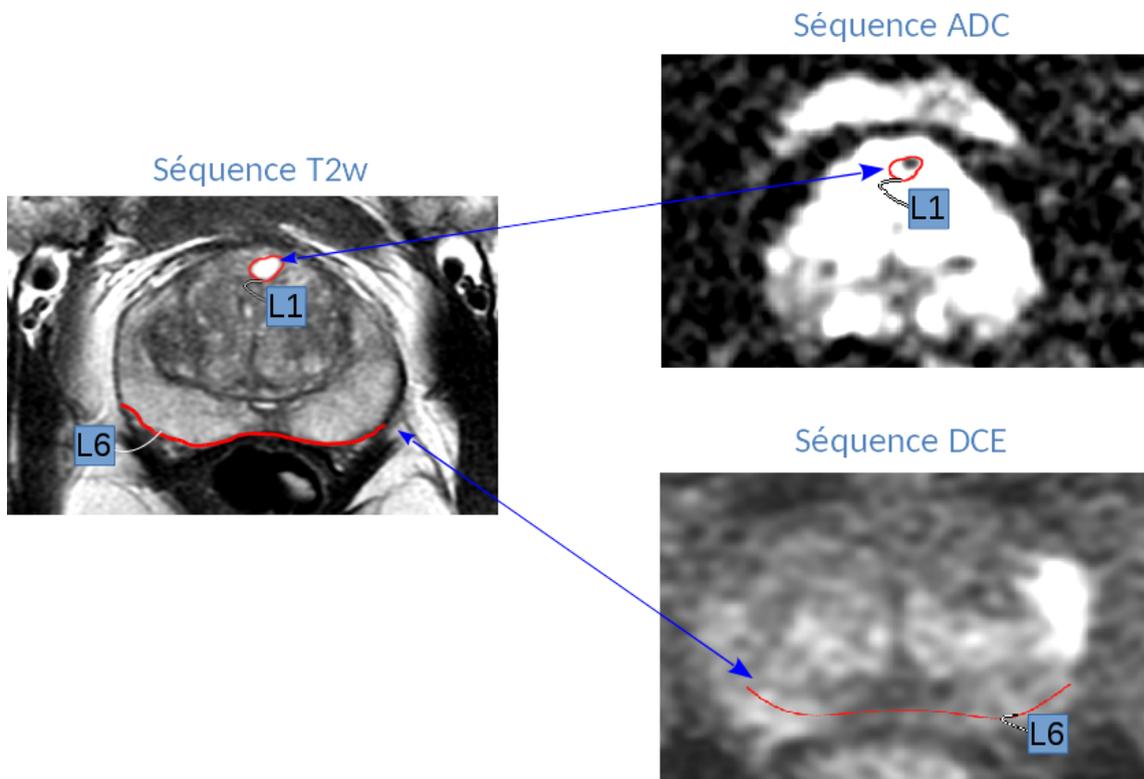


FIGURE 8.1 – Illustration de la nécessité du recalage

8.3 Description de la base 3T

Comme évoqué précédemment, l'objectif est le recalage d'images IRM multi-paramétriques 3T. Ces paramètres, illustrés par la figure 8.2, sont le T2-w, le B2000 et le DCE à 32 séquences temporelles. Dans cette nouvelle base, nous avons choisi d'ajouter la modalité une image IRM de diffusion (cf. section 3.2.2) pour un coefficient spécifique $b=2000$. En effet, plusieurs études [47, 89] ont montré le potentiel de l'imagerie de diffusion pour la détection du cancer de la prostate et notamment pour cette valeur particulière de b . La modalité T2-w est choisie comme référence car cette séquence contient le plus d'information anatomique et les meilleures caractéristiques de résolution spatiale et contraste. De plus, nous avons choisi de recalibrer les 32 séquences DCE entre elles de manière à corriger la variabilité intra-séquence. En effet, la durée d'acquisition de la séquence dynamique étant de l'ordre de 3 minutes (7 secondes par séquence), la probabilité de mouvement du patient au cours de l'examen est donc forte. Les paramètres principaux des séquences IRM 3T sont décrits dans le tableau 8.1.

Séquence	T_R (s)	T_E (s)	Angle bascule (°)	Épaisseur coupe (mm)	Matrice (vox)	Nb. coupes	FOV (mm)	dimension voxel (mm)	temps échantillon- nage (s)
T2-w	5	0.104	180	3	512×512	24	200×200	.42x.42x3	-
DCE	1.70	3.93	10	3	256×256	24	210×240	.93x.93x3	7
ADC	5	0.09	90	3	256×256	24	300×206	1.48x1.48x3	-
B2000	5	0.09	90	3	256×256	24	300×206	1.48x1.48x3	-

TABLE 8.1 – Paramètres utilisés pour l'imagerie de la prostate par IRM à 3T

8.4 Méthodes de recalage

8.4.1 Approche générale

Le recalage d'images consiste à mettre en correspondance deux ou plusieurs images afin de pouvoir les comparer ou extraire de l'information dont l'analyse requiert une similarité spatiale entre les images. A la fin de la procédure, les images se trouvent dans un même repère géométrique et le résultat final peut être soit sous forme d'un vecteur décrivant la transformation mettant en correspondance les informations données par deux images ou bien dans notre cas le résultat est directement l'image résultante de la mise en correspondance.

Cette étape de recalage est très fréquente en imagerie médicale, surtout lorsque la comparaison d'images multi-modales, multi-paramétriques ou encore inter ou intra-patient est nécessaire. Par exemple, dans le cadre du cancer de la prostate et du guidage des biopsies vers les foyers suspects, il est souvent nécessaire de recalibrer une image de type IRM T2w et une image de type ultrasonore afin d'améliorer la précision du guidage [11]. La figure 8.3 illustre ces propos.

Mathématiquement, le recalage peut se formuler par :

$$\hat{t} = \arg \max_{t \in \mathcal{T}} D(S, C, t) \quad (8.1)$$

Il consiste donc à trouver la transformation idéale \hat{t} , au sens d'une mesure de distance D , permettant de superposer l'image source S sur l'image cible C .

8.4.2 Construction d'un modèle générique de recalage

Il existe différentes manières de poser mathématiquement un problème de recalage entre images afin d'obtenir la transformation \hat{t} décrite ci-dessus. Néanmoins il se composera invariablement de quatre critères principaux dépendant directement de la nature des images à recalibrer et des conditions d'acquisitions. Ces critères sont :

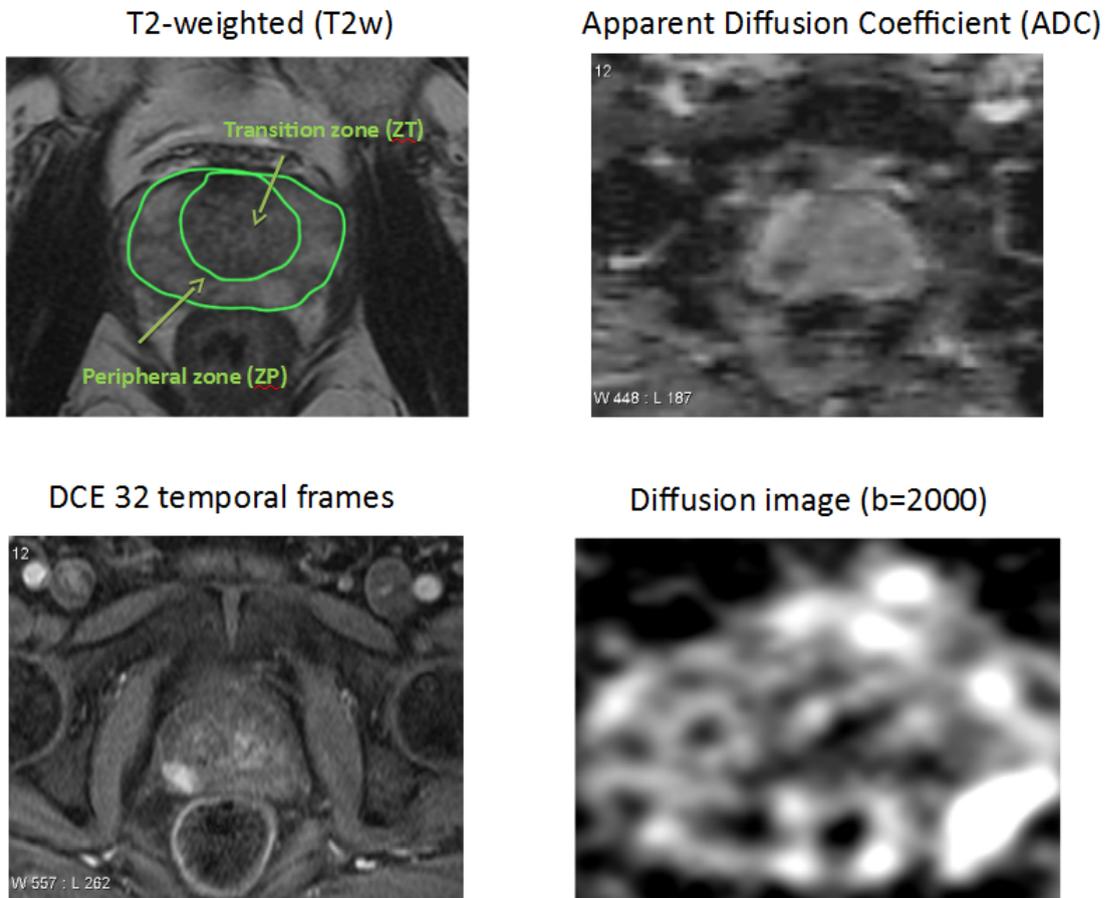


FIGURE 8.2 – Illustration des séquences IRM 3T

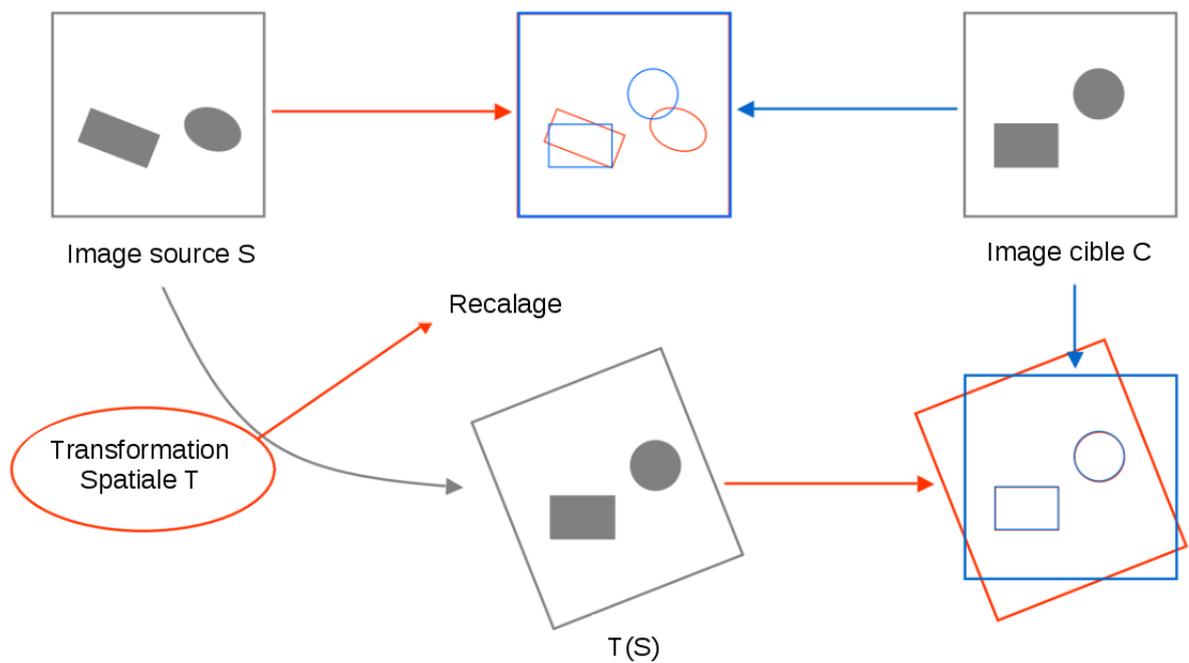


FIGURE 8.3 – Principe du recalage. Source [18]

- **Les primitives** : quelles sont les informations images guidant le système de recalage ? Ces primitives peuvent être de type géométrique (contours, sommets, amers, surfaces, etc...), iconiques (luminances de l'image) ou hybrides (approche géométrique + iconique)
- **La similarité** : comment estimer la correspondance entre les points ? La similarité dépend directement du choix du type de similarité (géométrique, iconique ou hybride). On peut citer par exemple la minimisation d'une distance pour l'approche géométrique ou la minimisation de critères statistiques (e.g information mutuelle) pour l'approche iconique.
- **Le modèle de transformation** : comment appairer les points sources aux points cibles ? On retrouve les modèles de transformation dits linéaires (transformation affines et rigides) ou non-linéaires (transformation élastique). La figure 8.4 illustre ces différents modèles.
- **La stratégie d'optimisation** : comment converger vers une solution ? On peut citer par exemple une stratégie basée sur une descente de gradient ou une minimisation de type moindres carrés. Le choix de la stratégie d'optimisation dépendra principalement des choix précédents.

Recalage basé sur des primitives géométriques. Ce type de recalage est basé sur l'extraction des primitives géométriques de deux images à recalcr. Dans un premier temps, on extrait des informations de type points d'intérêt, contours ou surface sur les deux images. Ensuite, ces informations sont mises en correspondances par un processus d'appariement puis triées afin de ne garder que les primitives les plus corrélées entre les deux images. Cette étape est d'ailleurs cruciale quant à la qualité finale du recalage.

Recalage basé sur des primitives iconiques. Contrairement au recalage basé sur des primitives géométriques, l'étape d'extraction de caractéristiques est éliminée et l'on considère uniquement une mesure de similarité sur l'intensité des voxels des deux images à mettre en correspondance. La mesure de similarité et la stratégie d'optimisation sont donc primordiales ici.

Recalage hybride. Afin de compenser les imprécisions dues à l'extraction de primitives géométriques, l'idée est ici de combiner l'approche géométrique avec l'approche iconique de manière temporaire, spatiale ou simultanée. L'approche temporaire permet notamment de raffiner les premiers résultats obtenus avec l'approche géométrique alors que les approches spatiales et simultanées présentent toujours l'inconvénient d'être partiellement dépendantes de la précision des primitives géométriques.

8.4.3 Modèles de transformation

On recense principalement trois modèles de transformations permettant de passer d'une image à une autre : rigide, affine et élastique (cf. figure 8.4). Les deux premiers font

partie de la famille des transformations linéaires appliquées à l'image alors que le recalage élastique appartient à celle des transformations non-linéaires.

Le recalage rigide introduit une rotation R et une translation t de l'image d'origine I pour obtenir l'image J . Cette transformation possède 3 degrés de liberté conserve les angles et les distances.

Le recalage affine introduit une translation, une rotation, une homotétie, un changement d'échelle et un cisaillement (6 degrés de liberté). Cette transformation conserve néanmoins les droites parallèles.

Le recalage élastique est une transformation non linéaire autorisant 9 degrés de liberté et ainsi toutes les déformations possibles. Cette transformation ne conserve donc pas les droites parallèles. Elle est particulièrement utilisée en imagerie médicale lorsque l'on souhaite corriger des déformations liées à l'acquisition et lorsque des structures anatomiques sont en mouvement.

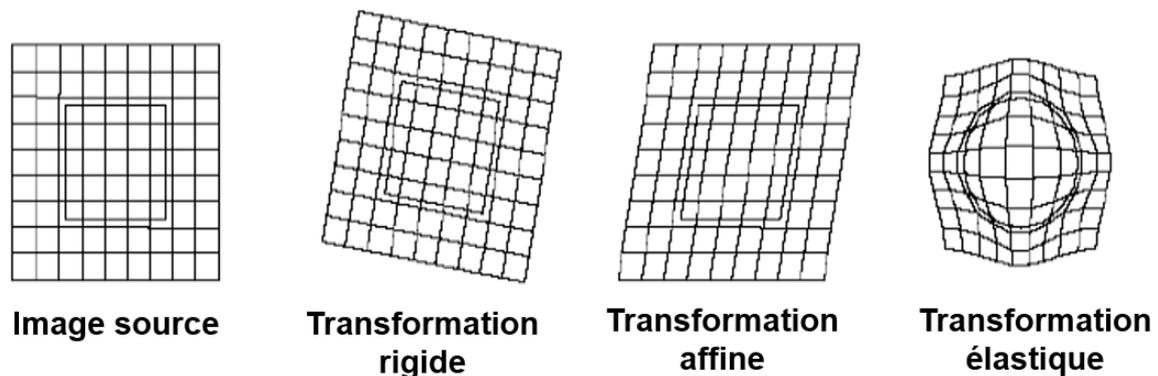


FIGURE 8.4 – Les différents modèles de transformations. Source [18]

8.5 Choix du modèle

De par la nature même des modalités d'image IRM que nous utilisons, l'approche iconique est à exclure dans notre cas de figure. En effet si l'on se réfère à la figure 8.2, on voit très clairement qu'utiliser une information basée sur l'intensité image entre le T2w et le B2000 ne donnerait pas de résultat convenable, cette dernière est en effet très fortement bruitée. Une approche de type géométrique est recommandée ici car la forme de la prostate est globalement conservée d'une séquence à l'autre. Concernant le modèle de transformation, un recalage élastique est ici préconisé afin de corriger la déformation des bords de la prostate. De par ces constats nous avons opté pour le logiciel développé par la société Koelis, ProMap© [11, 57]. Actuellement utilisé en routine clinique dans le service de radiologie urinaire de l'hôpital Edouard Herriot de Lyon, il permet un recalage élastique d'une séquence IRM T2w avec image ultrasonore dans le but de guider les biopsie dans la prostate. La version recherche du logiciel que nous avons utilisé permet un recalage entre plusieurs modalités d'IRM : T2w, ADC, B2000 et DCE multi-séquences. Malgré la diver-

sité de l'état de l'art en recalage nous avons choisi ce logiciel car il répond correctement à notre cahier des charges et son efficacité pour le recalage IRM T2w/ US a été prouvée. Ce logiciel nous permettra in fine d'avoir un recalage complet de toutes les séquences IRM sur l'image T2-w de référence, elle-même recalée sur l'image ultrasonore. L'une des applications visées est de pouvoir superposer la carte probabiliste estimée par notre système CAD sur l'image ultrasonore, afin d'améliorer le ciblage des zones suspectes au cours de la biopsie échoguidée.

8.6 Description de la méthode

La figure 8.5 donne une vue d'ensemble de la méthode implémentée par le logiciel ProMap©. Le process global de recalage se déroule en deux principales étapes : un pré-alignement rigide suivi d'un recalage élastique entre la séquence IRM de référence T2w et les séquences ADC, B2000 et DCE. Une description exhaustive des méthodes utilisées peut être trouvée aux références [6, 11, 57, 58].

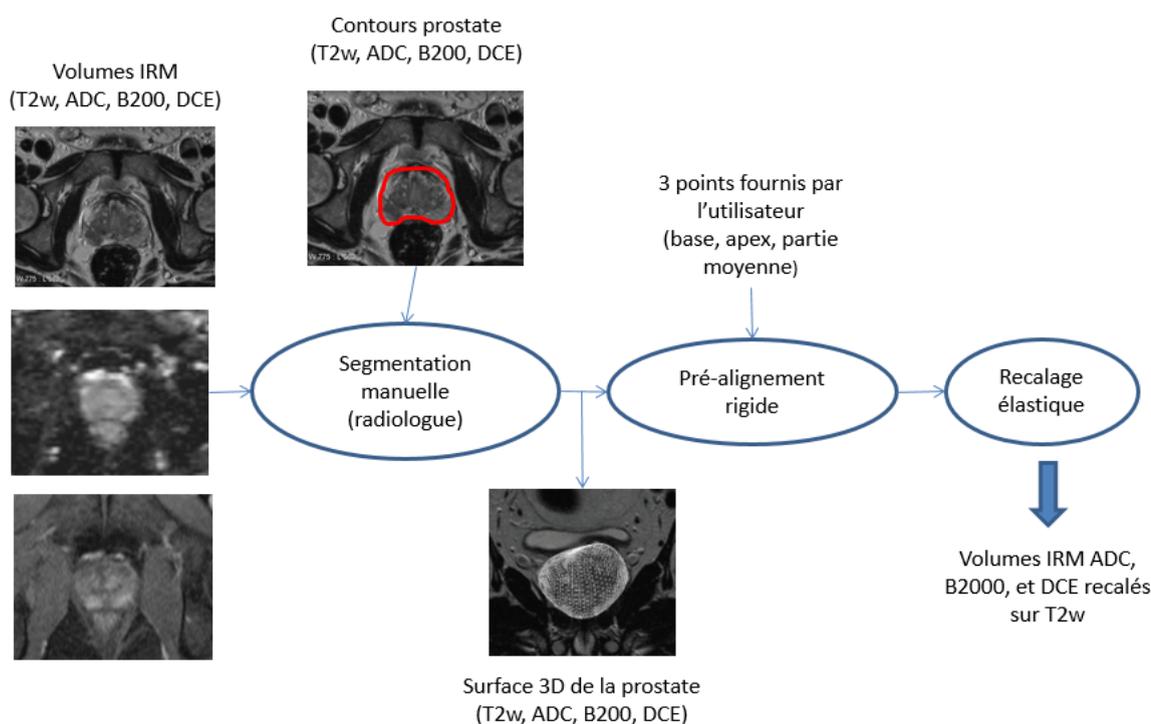


FIGURE 8.5 – Description de la méthode par maillage utilisée

Segmentation de la prostate. Lors de cette première étape, les volumes 3D des séquences IRM T2w, ADC, B2000 et DCE sont segmentées manuellement à l'aide des contours de prostate, tracés sous OSIRIX© par un radiologue expert.

Recalage rigide. Ce recalage rigide est un pré-alignement rigide. Celui-ci est réalisé à l'aide de 3 landmarks fournis par l'utilisateur du logiciel. Ces trois landmarks identifient

la base, l'apex et la partie moyenne de la prostate et sont indiqués pour chacune des séquences IRM. Le recalage rigide est ensuite appliqué entre les maillages de la séquence T2w (référence) et la séquence IRM à recaler (ADC ou DCE), obtenus lors de la première étape. Ce recalage rigide vise à minimiser itérativement, par descente de gradient, la distance entre un sommet m_i et les points s_j de la surface S des maillages obtenus à la première étape de segmentation.

Cette distance, appelée "iterative closest points" (ICP) est définie par :

$$D_{ICP} = \sum_i \|m_i - s_i\| \quad (8.2)$$

La stratégie consiste à rechercher le correspondant de chaque point de la scène parmi les points du modèle puis de minimiser une distance quadratique entre les points appariés.

Recalage élastique. Une fois le pré-alignement rigide effectué, une dernière étape de recalage élastique vise à déformer les maillages afin de minimiser la distance D entre les sommets s_j du maillage de référence, appelé R, et les projections de ces sommets s_j sur la surface du maillage à recaler, appelé S. La déformation élastique entre les deux maillage peut-être estimée en résolvant le problème de minimisation suivant :

$$\phi^* = \underset{\phi}{\operatorname{argmin}}(D[S, C, \phi]) \quad (8.3)$$

où D correspond à une mesure de distance entre les maillages S et C par la transformation ϕ .

Évaluation des performances de recalage

9.1 Base d'évaluation et protocole

Ce travail d'évaluation des performances de recalage a été initié au début de la thèse puis repris et finalisé par Fabrice Prost-Boucle. Les performances du logiciel décrit dans la partie 8, on été évaluées sur une base de 41 patients ayant subi un examen IRM 3T fournissant les 3 séquences T2w, B2000 et DCE.

Le protocole d'annotation et d'évaluation est le suivant :

- a) 6 repères, nommés ROI_{Ref} , sont positionnés sur les séquences T2 de chaque patient :
 - 3 repères circulaires marquaient des éléments visibles sur les 3 séquences (T2, B2000, DCE) : col, kystes, et/ou nodule adénomateux en fonction des patients.
 - 3 repères curvilinéaires repérant la face postérieure de la prostate à la base, la PM et l'apex.
- b) Les 6 ROI_{Ref} sont ensuite transférés sur les images B2000 et 3 séquences dynamiques (correspondant aux temps d'acquisition 7, 15 et 32 permettant de couvrir l'intégralité du temps d'acquisition). Ce transfert était rendu possible car les séquences ont même centre et même positionnement dans les coupes de 3 mm. Si les ROI_{Ref} ne tombent pas sur la zone d'intérêt correspondante, alors ils sont déplacés manuellement pour faire correspondre pour produire un nouveau jeu de 6 repères nommés $ROI_{norecal}$;
- c) L'opération b) est répétée après recalage élastique des séquences DCE et DWI sur le T2-W par le logiciel Koelis, pour donner un dernier jeu de 6 repères, ROI_{recal} .
- d) On mesure ensuite les écarts entre ROI_{Ref} et $ROI_{norecal}$ puis entre ROI_{Ref} et ROI_{recal} .

9.2 Description des métriques

Les métriques choisies et typiquement utilisées pour l'évaluation d'algorithmes de recalage en IRM [11] sont :

- La distance euclidienne entre les barycentres des landmarks circulaires
- Le décalage des landmarks curvilignes sur les bords postérieurs de la prostate

Calcul de la distance euclidienne entre landmarks circulaires. Dans le cas général ou l'on assimile le landmark circulaire contouré à un polygone irrégulier, sa surface A est exprimée par : $A = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$ avec n le nombre de points du polygone, la paire (x_i, y_i) les coordonnées d'un sommet du polygone. Le barycentre $G = (G_x; G_y)$ du polygone s'exprime alors par :

$$\begin{cases} G_x = \frac{1}{6A} \sum_{i=0}^{n-1} (x_i + x_{i+1})(y_{i+1} x_i - x_{i+1} y_i) \\ G_y = \frac{1}{6A} \sum_{i=0}^{n-1} (y_i + y_{i+1})(y_{i+1} x_i - x_{i+1} y_i) \end{cases} \quad (9.1)$$

avec G_x et G_y , l'abscisse et l'ordonnée respectives du barycentre du landmark. Ainsi, on calcul la distance euclidienne D_{bary} entre deux barycentres, avant et après recalage, par :

$$D_{bary} = \sqrt{(G_x - G_{x_{rec}})^2 + (G_y - G_{y_{rec}})^2} \quad (9.2)$$

avec $(G_x; G_y)$ et $(G_{x_{rec}}; G_{y_{rec}})$ les coordonnées de deux barycentres des deux landmarks circulaires avant et après recalage respectivement.

Mesure du décalage des landmarks curvilignes. Le décalage des landmarks curvilignes sur les bords postérieurs de la prostate est quant à lui évalué par un décalage absolu entre les landmarks de la prostate. En effet, le radiologue reporte sur la séquence recalée le même landmark curvilinéaire sans en altérer sa forme. Ainsi on exprime la distance D_{curv} par

$$D_{curv} = \sqrt{(x - x_{rec})^2 + (y - y_{rec})^2} \quad (9.3)$$

avec $(x; y)$ et $(x_{rec}; y_{rec})$ les coordonnées de deux points identiques du landmark curvilinéaire avant et après recalage respectivement.

9.3 Résultats

Cette section présente les résultats de recalage par des tableaux et histogrammes de distances mais aussi par des moyens visuels permettant plus aisément la comparaison des résultats. Lorsqu'elle est mentionnée, la notion de distance renvoie systématiquement à

celle entre des landmarks appartenants aux séquences IRM B2000 ou DCE et la référence T2w. La même distance est calculée avant et après recalage afin d'effectuer la comparaison. Il est à noter que nous ne présentons pas les résultats de recalage de la séquence ADC puisque les séquences ADC et B2000 sont spatialement équivalentes (ADC est construite à partir de plusieurs coefficients de diffusion b) et les résultats de recalage identiques.

9.3.1 Résultat global moyenné

Le tableau 9.1 présente un résultat global, pour la totalité des 41 patients, où toutes les distances par type de séquence IRM et landmark ont été moyennées avant et après recalage. ces premiers résultats montrent que le recalage permet une meilleure superposition des bords postérieures de la prostate pour la séquence B2000. En effet, on note une amélioration de 1.1mm sur la distance moyenne ($1.76 \pm 4.16\text{mm}$ et $0.61 \pm 2.57\text{mm}$ avant et après recalage respectivement). Ce résultat n'est pas surprenant, c'est effectivement dans cette région que la prostate subie la plus forte déformation et donc là où le recalage est le plus attendu pour corriger cette erreur. On note de même une sensible amélioration de l'ordre de 0.7mm pour les landmarks circulaires ($0.97 \pm 2.44\text{mm}$ et $0.25 \pm 1.67\text{mm}$ avant et après recalage respectivement).

En ce qui concerne la séquence DCE, on notera dans un premier temps qu'avant recalage, les distances moyennes sont toutes inférieures à celles de la séquence B2000. Ceci confirme le fait que la séquence B2000 est la plus difficile à recalculer du fait de sa nature très bruitée. Pour les landmarks curvilignes, on ne note pas d'amélioration moyenne après recalage ($0.83 \pm 1.33\text{mm}$ et $0.83 \pm 2.53\text{mm}$) et une faible amélioration de 0.13mm ($0.45 \pm 0.9\text{mm}$ et $0.32 \pm 1.63\text{mm}$ avant et après recalage respectivement) pour les landmarks circulaires. Ces premiers résultats donnent un aperçu global de l'erreur moyenne mais ne sont pas informatifs de la qualité du recalage. Il faut tout d'abord prendre en compte la taille du voxel qui est de 1.48 mm pour la séquence B2000 et 0.93 pour la séquence DCE. En effet, pour un même landmark, si sa distance est inférieure à la précision du voxel avant recalage et reste inférieure après recalage, il n'est pas aisé de conclure sur une amélioration ou une dégradation apportée par l'algorithme de recalage. De plus, ces résultats étant synthétisés par une moyenne globale, on ne dispose pas d'information concernant la capacité du recalage à diminuer considérablement une erreur ou au contraire sa propension à augmenter l'erreur dans certains cas.

	B2000		DCE	
	avant rec.(mm)	après rec.(mm)	avant rec.(mm)	après rec.(mm)
Barycentres	0.97 ± 2.44	0.25 ± 1.67	0.45 ± 0.90	0.32 ± 1.63
Curvilignes	1.76 ± 4.16	0.61 ± 2.57	0.83 ± 1.33	0.83 ± 2.53

TABLE 9.1 – Tableau de résultats de distances, moyenne \pm écart-type, avant et après recalage entre séquences DCE / B2000 et T2w

9.3.2 Matrice image des résultats

La figure 9.1 présente les résultats globaux de recalage sous la forme d'une matrice image, pour des seuils d'erreur d'acceptation après recalage de 1, 2, 3 et 4mm respectivement aux figures (a), (b) et (c). Les 4 colonnes principales de cette matrice (séparées par un trait épais noir) représentent les séquences B2000 et DCE acquises aux temps 7, 15 et 24. Chacune de ces 4 colonnes est ensuite subdivisée en 6 "barrettes" représentant les landmarks L1 à L6 respectivement. Chaque ligne représente un patient. Ainsi, on dispose de 4 code couleurs :

- *Vert foncé* : avant et après recalage la distance reste $<$ seuil d'acceptation, dans ce cas le process recalage n'a pas d'influence notable.
- *Vert clair* : la distance avant recalage est $>$ seuil d'acceptation et $<$ seuil d'acceptation après recalage, il s'agit du cas de figure idéal où le process de recalage améliore l'erreur pour la rendre acceptable suivant le seuil d'acceptation.
- *Rouge* : avant et après recalage la distance reste $>$ seuil d'acceptation, le process de recalage ne parvient pas à rendre l'erreur acceptable.
- *Orange* : la distance avant recalage est $<$ seuil d'acceptation et $>$ seuil d'acceptation après recalage, il s'agit du pire cas, celui où le process de recalage fait basculer une erreur initialement acceptée en une erreur rejetée selon le critère du seuil d'acceptation.

Les matrices images présentées sur la figure 9.1 donnent une indication globale des performances de recalage et notamment sur la capacité de l'algorithme à correctement corriger une erreur ou au contraire à dégrader l'alignement initial entre deux maillages. Le tableau 9.2 résume les informations données dans la matrice image pour différents seuil d'acceptation. Chaque ligne du tableau représente les résultats pour un seuil (en mm) et chaque colonne une séquence IRM. Chaque colonne par séquence est re-sous-divisée en deux colonnes faisant apparaître un nombre total de landmarks (couleur verte clair + rouge sur la matrice image) initialement supérieur au seuil d'acceptation dans la colonne "avant rec" et le nombre total de landmarks (couleur orange+rouge sur la matrice image) toujours supérieurs au seuil d'acceptation après recalage . Les chiffres entre parenthèses indiquent la décomposition par type de landmarks (barycentres/curvilignes).

Un des premiers constats de ce tableau est que la majorité des landmarks en défaut sont les landmarks curvilignes, ce qui indique que la déformation se trouve principalement sur les contours de la prostate et confirme les conclusions du tableau 9.1. On constate de plus que le recalage fonctionne mieux sur la séquence B2000. En effet, si l'on considère la première ligne du tableau pour le seuil d'1 mm, on voit que l'on passe de 62 landmarks rejetés (15 landmarks barycentres + 47 landmarks curvilignes) à 25 (6 landmarks barycentres + 19 landmarks curvilignes) soit 37 landmarks corrigés sur 65. En revanche si l'on considère la séquence DCE Phase 7, on voit que l'on passe de 46 landmarks (10+36) à 40 landmarks (9+31) soit 6 landmarks corrigés sur 46. Cette tendance à mieux recalibrer la séquence B2000 se retrouve quel que soit le seuil d'acceptation et pourrait s'expliquer par son meilleur

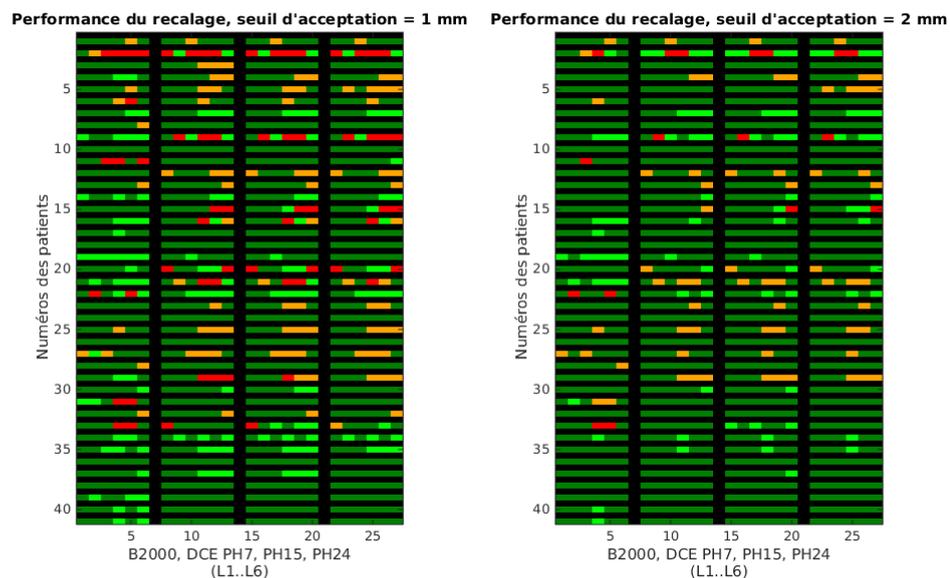
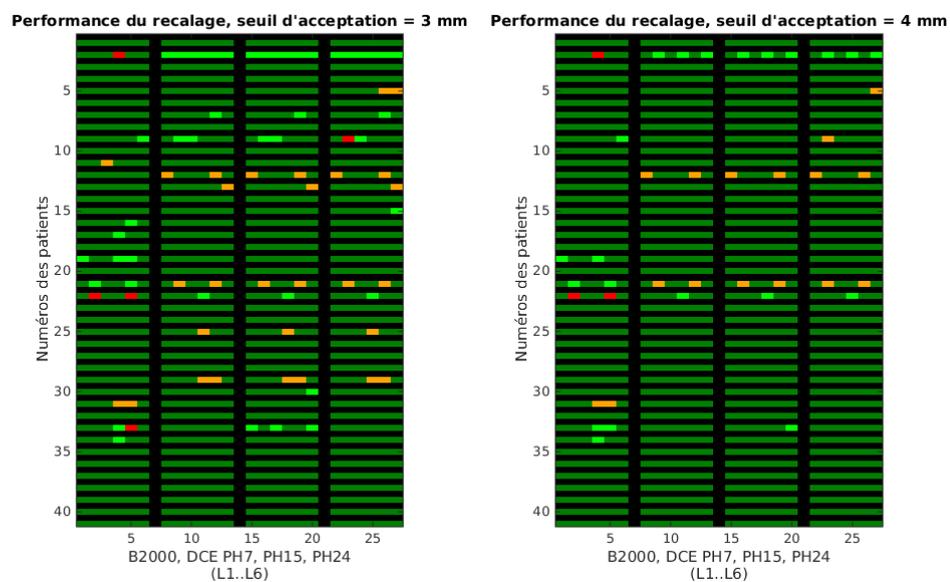
(a) $\text{delta}=\text{f}(\text{init})$ landmarks circulaires B2000.(b) $\text{delta}=\text{f}(\text{init})$ délinéations B2000.(c) $\text{delta}=\text{f}(\text{init})$ landmarks circulaires DCE(d) $\text{delta}=\text{f}(\text{init})$ délinéations DCE

FIGURE 9.1 – Représentation des résultats de recalage pour chaque type de landmark et par séquence : courbe $\text{delta} = \text{décalage initial} - \text{décalage final}$ en fonction du décalage initial

seuil (mm)	B2000		DCE Ph7		DCE Ph15		DCE Ph24	
	avant rec.	après rec.						
1	62 (15/47)	25 (6/19)	46 (10/36)	40 (9/31)	49 (11/38)	39 (9/30)	38 (8/30)	42 (10/32)
2	27 (6/21)	15 (5/10)	22 (6/16)	21 (6/15)	27 (8/19)	21 (6/15)	23 (5/18)	25 (7/18)
3	14 (3/11)	7 (2/5)	10 (5/5)	8 (2/6)	14 (7/7)	8 (2/6)	11 (5/6)	11 (3/8)
4	11 (3/8)	5 (1/4)	4 (1/3)	4 (2/2)	5 (1/4)	4 (2/2)	4 (1/3)	6 (3/3)

TABLE 9.2 – Tableau de résultats avant/après recalage en fonction de seuils d’acceptation. Chaque ligne représente les résultats pour un seuil (en mm) et chaque colonne une séquence IRM. Chaque colonne est subdivisée en deux colonnes faisant apparaître un nombre total de landmarks (couleurs verte et rouge sur la matrice image) initialement supérieur au seuil d’acceptation dans la colonne "avant rec" et le nombre total de landmarks (couleur orange+rouge sur la matrice image) toujours supérieurs au seuil d’acceptation après recalage. Les chiffres entre parenthèses indiquent la décomposition par type de landmarks (barycentres/curvilignes)

contraste pour la distinction des bords de la prostate. Les imprécisions de contourage sur la séquence DCE pourraient aussi être un facteur expliquant ces moins bonnes performances. En conclusion, de par ces résultats pour un seuil d’acceptation choisi, si l’on considère qu’un patient doit être exclu si au moins 1 landmark, toutes séquences confondues, reste supérieur au seuil d’acceptation après recalage (landmarks rouges) alors le nombre de patients conservés est 29, 35, 37 et 39, pour les seuils de 1,2,3 et 4mm respectivement.

9.3.3 Histogrammes des écarts de recalage

Dans cette section, il est présenté 4 histogrammes de répartition des landmarks en fonction de $\delta = \text{décalage initial} - \text{décalage final}$. Les deux premiers histogrammes (a) et (b) sont associés à la séquence B2000 pour les landmarks circulaires et curvilignes respectivement alors que les deux suivants (c) et (d) représentent de la même manière la séquence DCE toute phases confondues (Phases 7, 15 et 24). L’analyse de ces histogrammes fait ressortir les conclusions évoquées dans la partie 9.3.2, c’est-à-dire le fait que les résultats de recalage sont globalement meilleurs pour le B2000 et pour les landmarks circulaires.

En effet, si l’on considère la figure (a), on remarque une dissymétrie au tour de 0, indiquant donc $\text{décalage final} < \text{décalage initial}$, c’est-à-dire que le recalage a globalement amélioré le décalage initial. Pour les landmarks curvilignes de la séquence B2000 de la figure (b), la dissymétrie est moins nette et donc la distribution est plus large. Cette tendance met en évidence le bruit de contourage ou la vulnérabilité de la prostate aux mouvements du rectum ou de la vessie. Pour la séquence DCE et les landmarks circulaires représentés sur la figure (c), on remarque une tendance à la symétrie autour de 0 tout en ayant un peu plus de points positifs que négatifs, indiquant donc une amélioration peu efficace du recalage. Enfin, la figure (d) représente les landmarks curvilignes de la séquence DCE et une nette symétrie indiquant donc autant d’amélioration que de dégradations. Cette dernière figure semble confirmer que les bords postérieurs de la prostate sont mieux contrastés sur la séquence B2000, expliquant donc les meilleurs résultats de recalage.

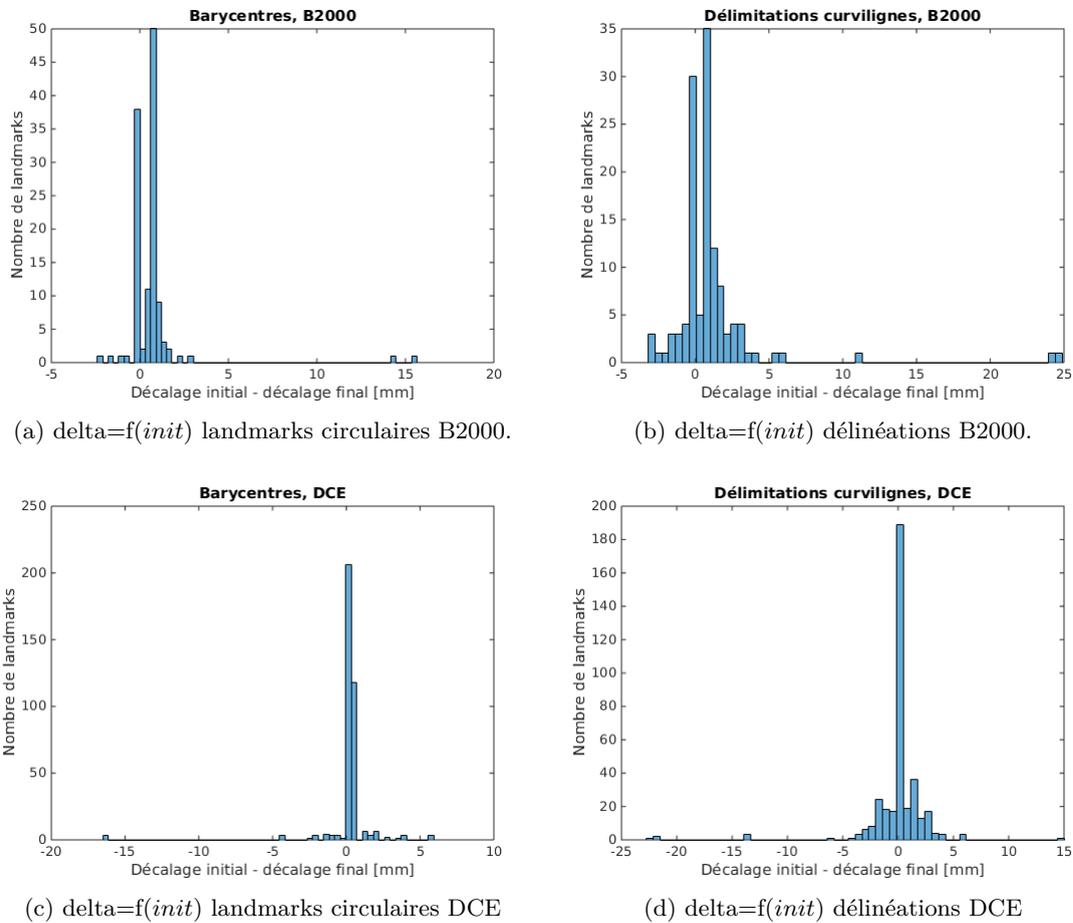
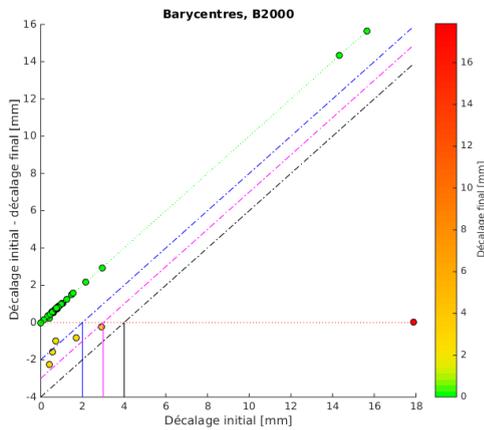


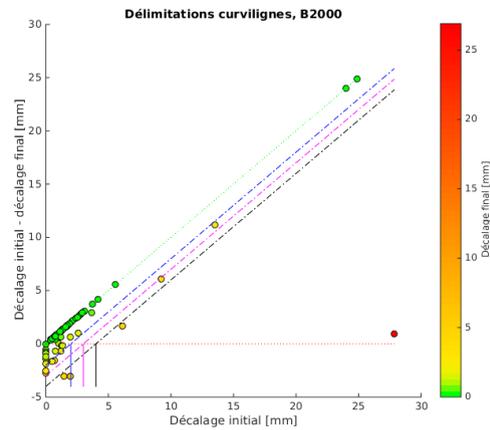
FIGURE 9.2 – Représentation des résultats de recalage pour chaque type de landmark et par séquence : courbe $\delta = \text{décalage initial} - \text{décalage final}$ en fonction du décalage initiale.

9.3.4 Représentation ponctuelle des écarts de recalage en fonction du décalage initial

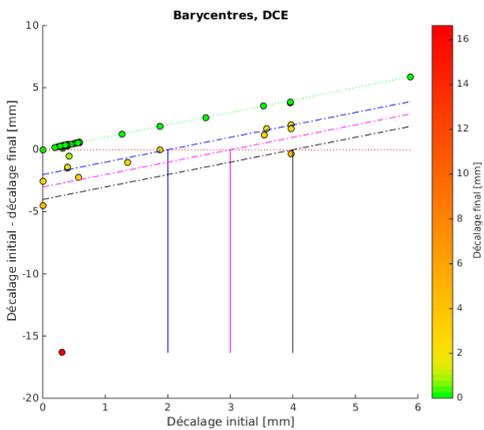
Cette dernière partie apporte une représentation exhaustive de l'influence du recalage sur chaque landmark. L'idée est de pouvoir quantifier de combien le recalage a amélioré ou dégradé le décalage initial. Chaque figure représente donc le δ (décalage initial - décalage final) en fonction du décalage initial pour chaque séquence et chaque type de landmark. On observe dans chaque figure deux zones principales, la zone dite "constructive" et la zone dite "destructive" correspondant aux zones où le recalage améliore et dégrade, respectivement, le décalage initial. La droite optimale verte représente le cas où le décalage final est nul donc celui pour lequel le recalage a entièrement corrigé l'erreur initiale. Ainsi, plus les points s'éloignent de cette droite et moins le recalage est performant. De plus, les 3 droites obliques de couleurs bleu, rose et vert représentent les zones d'acceptation ou de rejet correspondant aux seuils 2, 3 et 4 mm respectivement. Ainsi, par exemple, si un point est compris entre la droite verte et la droite rose, alors il sera accepté selon le critère de 3



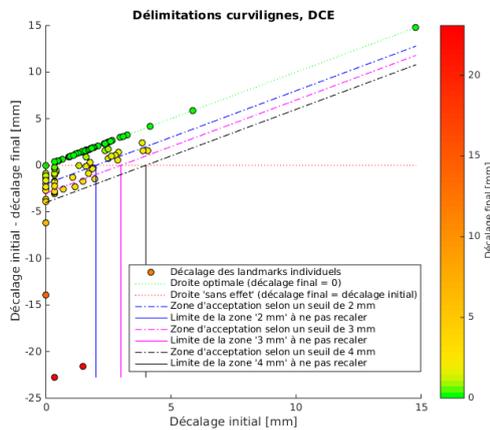
(a) $\delta = f(\text{init})$ landmarks circulaires B2000.



(b) $\delta = f(\text{init})$ délimitations B2000.



(c) $\delta = f(\text{init})$ landmarks circulaires DCE



(d) $\delta = f(\text{init})$ délimitations DCE

FIGURE 9.3 – Représentation des résultats de recalage pour chaque type de landmark et par séquence : courbe $\delta = \text{décalage initial} - \text{décalage finale}$ en fonction du décalage initiale. La légende sur la première figure s'applique aux 3 suivantes.

mm. Enfin les intervalles entre 0 et les droites verticales bleue, rose, et noire représentent les zones où le décalage initial est inférieur aux seuils 2, 3 et 4 mm, c'est-à-dire qu'il vaut mieux ne pas recalcer pour ne pas éventuellement introduire de détérioration. Ainsi, on peut facilement visualiser pour un landmark donné, selon sa position dans la zone constructive ou destructive, si le recalage a amélioré ou détérioré le décalage initial.

Pour les landmarks B2000 circulaires représentés sur la figure (a), on peut voir que la plupart des points sont proches de la droite optimale verte et initialement < 2 mm (droite verticale bleue). Ces points ne nécessitaient donc pas de recalage selon le critère 2 mm par exemple. On peut toutefois visualiser que 4 points de couleur jaune se situent dans la zone de destruction, indiquant donc que ces points ont été détériorés par le recalage mais restent tout de même en deçà du seuil d'acceptation de 2mm.

Pour les landmarks B2000 curvilignes représentés sur la figure (b), on se rend compte que la plupart des points sont là aussi initialement < 2 mm. Globalement, le recalage améliore sensiblement les résultats puisque la quasi totalité des points se retrouve entre la droite oblique verte et rose. On remarque néanmoins beaucoup de points pour lesquels le recalage les a dégradés (points appartenant à la zone de destruction).

Concernant les landmarks DCE circulaires représentés sur la figure (c), on remarque dans un premier temps que les points sont plus dispersés et ne sont pas autant initialement regroupés en dessous du seuil de 2mm (droite verticale bleue), preuve d'une nécessité de recalage pour certains points. On constate ensuite une amélioration après recalage puisque la plupart des points se retrouvent entre la droite oblique rose et la droite oblique optimale verte. On constate néanmoins aussi un certain nombre de points dans la zone destructive.

Enfin pour landmarks DCE curvilignes représentés sur la figure (d), on peut voir que la plupart sont initialement inférieurs aux seuils de 2 et 3mm (droites verticales bleue et rose). On remarque qu'après recalage, une grosse proportion de landmarks ont été dégradés et apparaissent dans la zone destructive. Ici l'apport du recalage est difficile à évaluer et semble avoir autant apporté d'améliorations que de dégradations. De par cette tendance, on se rend compte qu'il est préférable de ne pas recalcer les landmarks initialement inférieurs au seuil d'acceptation voulu, sous peine de probable dégradation.

Ici encore on se rend compte que la plupart des erreurs sont pour les landmarks curvilignes et notamment la séquence DCE. Ces résultats montrent que le recalage améliore globalement l'erreur initiale quelque soient les séquences ou les types de landmarks. On remarque néanmoins que certaines erreurs sont au contraire amplifiées, notamment pour les landmarks curvilignes. Ce comportement pourrait être gênant dans la mesure où l'on décide de recalcer systématiquement les séquences d'un patient ou de les recalcer lorsqu'au moins un landmark est supérieur à un seuil d'acceptation. En effet, il est difficile d'évaluer a posteriori l'effet de la dégradation introduit par le recalage sur la classification finale des images.

9.4 Conclusion

Cette première étape préliminaire au travail de classification au niveau pixelique, a permis d'évaluer les performances du logiciel de recalage élastique ProMap de la société Koe-lis. Celui-ci implémente un algorithme de recalage élastique basé sur la déformation de maillages 3D obtenus préalablement par une étape de segmentation des séquences IRM T2w, ADC, B2000 et DCE. L'objectif était d'évaluer le recalage des séquences B2000 et DCE sur la séquence référence T2w. Basée sur une base images de 41 patients, tous anno-tés par des landmarks anatomiques, de type circulaire (kystes, nodules adénomateux,...) et curvilignes (bords postérieurs de prostate), avant et après recalage, l'évaluation a permis de mettre en avant la capacité du recalage à globalement améliorer le décalage initial. A l'aide d'un certains nombres de seuils d'acceptation d'erreur (de 1 à 4mm), avant et après recalage, on peut conclure sur une meilleur efficacité du recalage pour les séquences B2000 et notamment les landmarks circulaires. De plus, on constate une tendance du recalage à détériorer certains landmarks dont l'erreur initiale était inférieure aux seuils d'accepta-tions donnés et notamment pour les séquences DCE. Cette tendance peut éventuellement s'expliquer par le mouvement (circulation d'air dans le rectum et remplissage de sa ves-sie) introduit pendant la durée de l'acquisition des séquences dynamiques (>3 minutes), mais aussi par l'imprécision de contourage des bords de la prostate en raison de la nature moins contrastée des images dynamiques (DCE) par rapport aux séquences de diffusion (ADC, B2000). Néanmoins, Cette étude ne permet pas de démontrer l'intérêt du recalage pour notre application. Il faudrait estimer l'impact de ce recalage sur les performances de classification. A priori, on peut penser que l'influence majeure concernerait : 1) la détec-tion des petites lésions sur les bordures postérieures de la prostate. Comme souligné en début de chapitre, les lésions sont plutôt volumineuses dans notre application, 2) sur la quantification de l'agressivité tumorale. Il faudrait étudier comment le recalage influence la mesure du score CAD et sa corrélation avec le score de Gleason.

Conclusions et perspectives

Conclusions des chapitres

Nous avons présenté un Système d'aide à la détection (CADe) pour le cancer de la prostate basé sur des images IRM multi-paramétrique (IRM-mp) T2w, ADC et DCE. L'objectif de ce CAdE est la détection, localisation des cancers agressifs ainsi que de fournir des cartes de probabilités corrélées au grade d'agressivité du cancer.

Dans la première partie nous avons présenté le cancer de la prostate dans son ensemble ainsi que les différentes techniques diagnostiques de référence que sont le toucher rectal, le taux de PSA et les biopsies écho-guidées. Ces techniques souffrent toutes d'un manque de précision induisant des sur ou sous-diagnostic, impliquant un traitement lourd, coûteux et invasif pour le patient.

Dans ces conditions, l'image IRM-mp s'impose peu à peu comme une référence pour la détection, localisation et l'évaluation de l'agressivité des tumeurs. L'apport d'une séquence IRM utilisée seule pour le diagnostic est avéré mais reste toutefois limité. En revanche, l'utilisation de l'IRM-mp combinant notamment les séquences T2w, ADC et DCE permet d'obtenir de très bonnes performances de détection et localisation ainsi qu'un apport notable dans la différenciation des tumeurs agressives des tissus sains. L'interprétation d'images IRM-mp reste cependant difficile, effectuée de manière qualitative et le contraste des images dépend fortement des caractéristiques intrinsèques de l'IRM utilisé en pratique. De ce constat, un besoin de standardisation s'impose à la fois sur le protocole d'acquisition ainsi que sur la manière d'interpréter les images. Les Systèmes d'aide à la décision (CAD) sont donc une proposition de réponse à ce besoin de standardisation. Ces derniers ont pour objectif de fournir un score de probabilité au niveau d'une région d'intérêt préalable-

ment contourée par le radiologue (approche CADx) ou bien au niveau voxelique (approche CADe). Bien que très prometteurs, ces systèmes sont difficilement comparables les uns aux autres. L'état de l'art effectué dans cette partie montre ainsi une très grande variabilité à la fois dans la technique d'acquisition des images IRM, sur le type de vérité terrain (histologique ou non) et sur les méthodes de validation des performances des CAD. Concernant l'utilisation potentielle des CAD en routine clinique, les conclusions sur son apport potentiel sont plutôt orientées vers une utilisation en second avis notamment pour améliorer les performances d'un radiologue inexpérimenté. Le CAD pourrait également servir d'outil de pré-analyse, réglé sur un seuil de sensibilité très haut, de manière à filtrer les résultats. Cette dernière utilisation pourrait permettre un gain de temps dans l'analyse des images. La seconde partie présente dans un premier temps la base IRM-mp utilisée pour l'implémentation du CADe. Constituée de 49 patients ayant subi une série d'IRM-mp suivie d'une prostatectomie radicale, celle-ci a été réalisée suivant un protocole standardisée permettant de mettre en corrélation les pièces de prostatectomie et les séquences IRM. Ce protocole a permis d'établir une vérité terrain incluant tissus sains, suspects mais sains et différents grades de cancer du Gleason 6 au Gleason 9. La force de cette base repose sur un protocole permettant d'obtenir une vérité terrain rigoureuse et non systématique dans l'état de l'art des CAD.

Il est ensuite présenté les différents choix méthodologiques d'implémentation du CADe : choix et sélection des caractéristiques discriminantes, méthodes de caractérisation des voxels pour déterminer l'agressivité du cancer, classifieurs supervisés, validation des paramètres optimaux et enfin métriques de performances. Bien que l'état de l'art montre une grande hétérogénéité dans l'implémentation des CAD, nous avons suivi un schéma standardisé de manière à minimiser les possibles biais à chaque étape.

Enfin, cette partie se termine sur l'évaluation des performances du CADe dont le but est de discriminer cancer agressifs du reste des tissus. La meilleur configuration de CAD est basée sur un classifieur SVM non linéaire associé à un vecteur de caractéristiques issu de l'apprentissage d'un dictionnaire sur une série de données statistiques et semi-quantitative extraites d'images IRM T2w, ADC et DCE. Bien que, comme évoqué précédemment, la comparaison soit difficile, notre CADe obtient des performances du même ordre que les meilleurs résultats de l'état de l'art pour la discrimination des cancers agressifs. Le second objectif était d'évaluer si une unique fonction de prédiction était capable d'encoder non seulement la certitude de la présence d'un cancer mais aussi son degré d'agressivité (score de Gleason). Les scores et cartes de probabilités indiquent que les valeurs moyennes prises par 3 cancers agressifs (Gleasons 7, 8 et 9) sont très proches et donc peu discernables les uns des autres.

La troisième et dernière partie présente nos travaux effectués en parallèle pour la constitution et le recalage d'une nouvelle base d'images IRM-mp 3T de type T2w, ADC, B2000 et DCE. Nous présentons dans un premier temps l'apport potentiel de données recalées sur les performances d'un CADe ainsi qu'une étude d'un logiciel de recalage que nous avons choisi et utilisé pour recalibrer nos données. Cette dernière montre que les séquences B2000 et

notamment les landmarks circulaires sont les mieux corrigés. On constate néanmoins que le recalage est inefficace sur certains landmarks notamment pour la séquence DCE. Cette tendance peut éventuellement s'expliquer par du mouvement introduit pendant la durée de l'acquisition des séquences dynamiques, mais aussi par l'imprécision de contourage des bords de la prostate en raison la nature moins contrastée des images dynamiques (DCE) par rapport aux séquences de diffusion (ADC, B2000).

Limitations et perspectives

Notre meilleur résultat de CADe est obtenu avec une méthode d'apprentissage de dictionnaires. Statistiquement meilleure que tous les autres schémas de CADe, cette différence n'est pas clairement visible sur les exemples de cartes de probabilités. Les dictionnaires sont estimés à partir des caractéristiques images déjà extraites, qui ont prouvé leur pouvoir discriminant pour les tâches de classification cancer VS non cancer et cancer agressif VS reste des tissus. Néanmoins, elles semblent ne pas capter précisément les différences entre cancers. Une idée serait d'apprendre un dictionnaire overcomplete (nombre d'atomes » nombre de caractéristiques) directement sur des patches d'images IRM [54] de manière à capter des informations plus fines sur l'agressivité du cancer.

Concernant la difficulté à discerner les différents grades de cancers, il serait intéressant de tester un modèle de régression logistique multiclassées. Cette approche nécessiterait néanmoins de fusionner les différentes cartes de probabilités obtenues pour chacune des classes de manière à améliorer la performance de diagnostic du radiologue.

Enfin, dans de précédents travaux [64], un CADx avait été évalué par des radiologues de manière à quantifier son apport en routine clinique. L'étude avait souligné l'amélioration des performances de diagnostic lorsque ce dernier est utilisé en second avis afin de confirmer la présence d'une zone suspecte. Ce même travail doit être mis en œuvre afin de quantifier l'apport du CADe et déterminer une possible utilisation en tant qu'outil de sélection des voxels suspect ou pour confirmer ou infirmer un premier diagnostic.

Publications de l'auteur

Revue Internationale avec comité de lecture

- [publié] A. Hoang Dinh, C. Melodelima, R. Souchon, J. Lehaire, F. Bratan, F. Mège-Lechevallier, A Ruffion, S. Crouzet, M Colombel and O. Rouvière, *Quantitative Analysis of Prostate Multiparametric MR Images for Detection of Aggressive Prostate Cancer in the Peripheral Zone : a multi-imager study*, Radiology, 280(1) :117-127, 2016

Congrès internationaux

- R. Aljundi, J. Lehaire, F. Prost-Boucle, O. Rouvière and C. Lartizien, *"Transfer Learning for Prostate Cancer Mapping Based on Multicentric MR imaging databases*, International Conference on Machine Learning (ICML), Lille, France, 2015
- J. Lehaire, R. Flamary, O. Rouvière, and C. Lartizien, *Computer-aided diagnostic system for prostate cancer detection and characterization combining learned dictionaries and supervised classification*, IEEE International Conference on Image Processing (ICIP), pp. 2251-2255, Paris, France, 10/2014
- J. Lehaire, C. Lartizien, E. Niaf, M. Baumann, F.Bratan, and O. Rouvière, *Evaluation of a 3D MR/MR elastic registration method for prostate cancer imaging*, European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), abstract, Toulouse, France, 10/2013

Annexes

Résultats complémentaires sur le CADe

A.1 Liste des caractéristiques sélectionnées

feature rank	SVM-L	Régression logistique
1	ADC Haralick savgh	T2 grey levels
2	T2 Haralick savgh	ADC grey levels
3	ADC Haralick dissi	T2 1-st Order StatisticsMedian
4	ADC Haralick homom	T2 1-st Order Statistics Mean
5	T2 Haralick dissi	T2 1-st Order Statistics std
6	ADC Haralick homop	DYN quantitative Ktrans
7	ADC Haralick senth	DYN Haralick entro
8	ADC Haralick maxpr	DYN Haralick dissi
9	ADC Haralick svarh	DYN semi-quantitative Tonset
10	ADC Haralick energ	DYN Haralick energ
11	DYN semi-quantitative Tonset	DYN semi-quantitative SIend
12	ADC Haralick indnc	T2 Haralick cprom
13	T2 1-st Order Statistics std	T2 Gradient sobel1
14	ADC Haralick idmnc	T2 Gradient kirsch
15	T2 Haralick senth	T2 Haralick cshad
16	T2 1-st Order Statistics Median	DYN quantitative ve
17	ADC grey levels	DYN Haralick homop
18	DYN Haralick contr	DYN Haralick svarh
19	ADC 1-st Order Statistics std	T2 1-st Order Statistics var
20	DYN semi-quantitative SIend	ADC Haralick cprom
21	T2 Haralick svarh	ADC Haralick homom
22	ADC 1-st Order Statistic sMedian	ADC Haralick homop
23	T2 Haralick homop	T2 Haralick coop
24	ADC Haralick cprom	T2 Haralick corrm
25	DYN Haralick maxpr	T2 Haralick contr
26	ADC Haralick inflh	T2 Haralick dvarh
27	T2 Haralick indnc	T2 Haralick indnc
28	T2 Haralick denth	DYN Haralick senth
29	T2 Haralick autoc	ADC Haralick maxpr
30	T2 Haralick sosvh	ADC Haralick entro
31	DYN semi-quantitative Tmax	ADC Haralick senth
32	DYN semi-quantitative SIonset	ADC Haralick inf2h
33	ADC Haralick autoc	DYN Haralick autoc
34	ADC Haralick inf2h	T2 Gradient sobel5
35	ADC Haralick sosvh	DYN semi-quantitative SImax
36	T2 grey levelsT2 w	ADC Gradient sobel3
37	T2 Haralick cprom	DYN 1-st Order Statistics Median
38	T2 1-st Order Statistics var	DYN 1-st Order Statistics Mean
39	ADC Gradient sobel4	T2 Haralick idmnc
40	DYN Haralick homom	ADC 1-st Order Statistics Median

TABLE A.1 – Liste des caractéristiques sélectionnées par classifieur

feature rank	SVM-L	Régression logistique
41	ADC Gradient sobel1	T2 Gradient sobel3
42	DYN semi-quantitative SImax	ADC Haralick inf1h
43	T2 Gradient sobel5	ADC Gradient grad2
44	ADC Gradient kirsch	ADC Haralick energ
45	ADC Gradient grad2	ADC Haralick idmnc
46	DYN Haralick autoc	T2 Haralick energ
47	ADC Haralick cshad	DYN Haralick contr
48	DYN semi-quantitative WO	T2 Gradient grad2
49	T2 Haralick homom	ADC Haralick denth
50	T2 Haralick contr	ADC Haralick indnc
51	T2 Haralick dvarh	DYN semi-quantitative WO
52	DYN Haralick cshad	ADC Gradient sobel5
53	T2 Haralick entro	T2 Haralick maxpr
54	ADC Haralick dvarh	ADC Gradient sobel4
55	ADC Haralick contr	DYN Haralick cshad
56	ADC 1-st Order Statistics var	T2 Haralick dissi
57	DYN Haralick sosvh	DYN 1-st Order Statistics var
58	DYN semi-quantitative Area	ADC Haralick cshad
59	T2 Gradient grad1	ADC 1-st Order Statistics Mean
60	DYN Haralick corrm	DYN semi-quantitative A
61	DYN Haralick energ	DYN Haralick corrm
62	DYN Haralick senth	ADC 1-st Order Statistics var
63	DYN Haralick svarh	DYN Haralick sosvh
64	ADC Gradient sobel3	DYN semi-quantitativeWI
65	T2 Haralick inf1h	ADC 1-st Order Statistics std
66	ADC Gradient grad1	DYN semi-quantitative Ta
67	ADC 1-st Order Statistics Mean	DYN Haralick savgh
68	ADC Haralick denth	DYN Haralick homom
69	DYN Haralick entro	T2 Gradient sobel2
70	T2 Gradient sobel3	ADC Haralick svarh

TABLE A.2 – Liste des caractéristiques sélectionnées par classifieur (suite)

Bibliographie

- [1] R. Akbani, S. Kwek, and N. Japkowicz. to Imbalanced Datasets. In *European Conference on Machine Learning*, pages 39–50, 2004.
- [2] Y. Artan, M. A. Haider, D. L. Langer, T. H. van der Kwast, A. J. Evans, Yongyi Yang, M. N. Wernick, J. Trachtenberg, and I. S. Yetik. Prostate Cancer Localization With Multispectral MRI Using Cost-Sensitive Support Vector Machines and Conditional Random Fields. *IEEE Transactions on Image Processing*, 19(9) :2444–2455, sep 2010.
- [3] R. Ashley, B. Inman, J. Routh, and M. Blute. Reassessing the diagnostic yield of saturation biopsy of the prostate. *European Urology*, 53(5) :976–981, 2008.
- [4] R. Bammer. Basic principles of diffusion-weighted imaging. *European journal of radiology*, 45(3) :169–184, 2003.
- [5] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, and J. J. Fütterer. ESUR prostate MR guidelines 2012. *European Radiology*, 22(4) :746–757, 2012.
- [6] M. Baumann, P. Mozer, V. Daanen, and J. Troccaz. Prostate biopsy tracking with deformation estimation. *Medical Image Analysis*, 16(3) :562–576, apr 2012.
- [7] F. Binder-Foucard, A. Belot, and P. Delafosse. Estimation nationale de l’incidence et de la mortalité par cancer en France entre 1980 et 2012. 2013.
- [8] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2) :354–379, 2012.
- [9] C. M. Bishop. *Pattern Recognition and Machine Learning*, volume 16. 2007.
- [10] F. Bratan, E. Niaf, C. Melodelima, A. L. Chesnais, R. Souchon, F. Mège-Lechevallier, M. Colombel, and O. Rouviere. Influence of imaging and histological factors on prostate cancer detection and localisation on multiparametric MRI : a prospective study. *European radiology*, 23(7) :2019–2029, 2013.
- [11] L. Brolis, P. Mozer, J. Troccaz, and M. Baumann. Accuracy evaluation of Koelis elastic fusion of MR and 3D TRUS prostate image. 2012.
- [12] a. Buhmeida, S. Pyrhönen, M. Laato, and Y. Collan. Prognostic factors in prostate cancer. *Diagnostic pathology*, 1(4) :4, 2006.
- [13] M. a. Bulbul, Y. El-Hout, M. Haddad, A. Tawil, A. Houjaij, N. Bou Diab, and O. Darwish. Pathological correlation between needle biopsy and radical prostatectomy specimen in patients with localized prostate cancer. *Canadian Urological Association journal = Journal de l’Association des urologues du Canada*, 1(3) :264–6, 2007.

- [14] H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis : an ROC study. *Radiology*, 212(3) :817–27, 1999.
- [15] I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging ; a multichannel statistical classifier. *Medical physics*, 30(9) :2390–2398, 2003.
- [16] I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging ; a multichannel statistical classifier. *Medical physics*, 30(9) :2390–2398, 2003.
- [17] C.-C. Chang and C.-J. Lin. Libsvm. *ACM Transactions on Intelligent Systems and Technology*, 2(3) :1–27, 2011.
- [18] O. Coulon. Analyse d ' images 5 - Introduction au recalage Problématique.
- [19] J. C. Dean and C. C. Ilvento. Improved Cancer Detection Using Computer-Aided Detection with Diagnostic and Screening Mammography : Prospective Study of 104 Cancers. *American Journal of Roentgenology*, 187(1) :20–28, 2006.
- [20] N. B. Delongchamps and G. P. Haas. Saturation biopsies for prostate cancer : current uses and future prospects. *Nature reviews. Urology*, 6(12) :645–52, 2009.
- [21] N. B. Delongchamps, M. Rouanne, T. Flam, F. Beuvon, M. Liberatore, M. Zerbib, and F. Cornud. Multiparametric magnetic resonance imaging for the detection and localization of prostate cancer : combination of T2-weighted, dynamic contrast-enhanced and diffusion-weighted imaging. *BJU International*, 107(9) :1411–1418, 2011.
- [22] L. Dickinson, H. U. Ahmed, C. Allen, J. O. Barentsz, B. Carey, J. J. Futterer, S. W. Heijmink, P. J. Hoskin, A. Kirkham, A. R. Padhani, R. Persad, P. Puech, S. Punwani, A. S. Sohaib, B. Tombal, A. Villers, J. Van Der Meulen, and M. Emberton. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer : Recommendations from a European consensus meeting. *European Urology*, 59(4) :477–494, 2011.
- [23] J. J. Fütterer, S. W. T. P. J. Heijmink, T. W. J. Scheenen, J. Veltman, H. J. Huisman, P. Vos, C. A. H. de Kaa, J. A. Witjes, P. F. M. Krabbe, A. Heerschap, and J. O. Barentsz. Prostate Cancer Localization with Dynamic Contrast-enhanced MR Imaging and Proton MR Spectroscopic Imaging. *Radiology*, 241(2) :449–458, 2006.
- [24] J. Galić, I. Karner, L. Cenan, A. Tucak, I. Hegedus, J. Pasini, M. Bilandzija-Peranović, and S. Mihaljević. Comparison of digital rectal examination and prostate specific antigen in early detection of prostate cancer. *Collegium antropologicum*, 27 Suppl 1(1) :61–6, 2003.
- [25] J. F. García Molina, L. Zheng, M. Sertdemir, D. J. Dinter, S. Schönberg, and M. Rädle. Incremental Learning with SVM for Multimodal Classification of Prostatic Adenocarcinoma. *PLoS ONE*, 9(4) :e93600, jan 2014.
- [26] G. Giannarini, R. Autorino, and G. di Lorenzo. Saturation Biopsy of the Prostate : Why Saturation Does Not Saturate. *European Urology*, 56(4) :619–621, 2009.
- [27] M. L. Giger, H.-P. Chan, and J. Boone. Anniversary paper : History and status of CAD and quantitative image analysis : the role of Medical Physics and AAPM. *Medical physics*, 35(12) :5799–5820, 2008.

-
- [28] N. Girouin, F. Mège-Lechevallier, A. Tonina Senes, A. Bissery, M. Rabilloud, J.-M. Maréchal, M. Colombel, D. Lyonnet, and O. Rouvière. Prostate dynamic contrast-enhanced MRI with simple visual diagnostic criteria : is it reasonable? *European radiology*, 17(6) :1498–509, 2007.
- [29] I. Guyon. *Pattern classification*, volume 1. 1998.
- [30] I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh. *Feature Extraction, Foundations and Applications*, volume 207. 2006.
- [31] G. P. Haas, N. Delongchamps, O. W. Brawley, C. Y. Wang, and G. de la Roza. The worldwide epidemiology of prostate cancer : perspectives from autopsy studies. *The Canadian journal of urology*, 15(1) :3866–3871, 2008.
- [32] M. a. Haider, T. H. Van Der Kwast, J. Tanguay, A. J. Evans, A. T. Hashmi, G. Lockwood, and J. Trachtenberg. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *American Journal of Roentgenology*, 189(2) :323–328, 2007.
- [33] T. Hambrock, P. C. Vos, and C. a. H. V. D. Kaa. Prostate cancer : Computer-aided Diagnosis with Multiparametric 3-T MR Imaging — Effect on Observer. *Radiology*, 266(2) :521–530, 2013.
- [34] R. Haralick, K. Shanmugan, and I. Dinstein. Textural features for image classification, 1973.
- [35] T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning : Data Mining, Inference, and Prediction, Second Edition. In *The elements of statistical learning : data mining, inference, and prediction*, pages 501–520. 2011.
- [36] A. Hoang Dinh. *Développement et Évaluation Des paramètres Quantitatifs de l'IRM de la Prostate*. PhD thesis, Université Lyon, 2015.
- [37] C. C. M. a. Hoeks, J. J. O. Barentsz, T. Hambrock, D. Yakar, D. M. Somford, S. W. T. P. J. Heijmink, T. W. J. Scheenen, P. C. Vos, H. Huisman, I. M. van Oort, J. A. Witjes, A. Heerschap, and J. J. Fütterer. Prostate cancer : multiparametric MR imaging for detection, localization, and staging. *Radiology*, 261(1) :46–66, 2011.
- [38] K. Hospital, C. Hospital, L. Egevad, T. Granfors, L. Karlberg, A. Bergh, and P. Stattin. Prognostic value of the Gleason score in prostate cancer. *BJU International*, 89(6) :538–542, 2002.
- [39] Z. Huang, M. Thint, and Z. Qin. Question Classification using Head Words and their Hypernyms. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, number October, pages 927–936, 2008.
- [40] Iarc. GLOBOCAN 2012 : Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. *Globocan*, 2012 :3–6, 2012.
- [41] S. Ikonen, P. Kärkkäinen, L. Kivisaari, J. O. Salo, K. Taari, T. Vehmas, P. Tervahartiala, and S. Rannikko. Endorectal magnetic resonance imaging of prostatic cancer : Comparison between fat-suppressed T2-weighted fast spin echo and three-dimensional dual-echo, steady-state sequences. *European Radiology*, 11(2) :236–241, 2001.
- [42] S. Isebaert, L. Van Den Bergh, K. Haustermans, S. Joniau, E. Lerut, L. De Wever, F. De Keyzer, T. Budiharto, P. Slagmolen, H. Van Poppel, and R. Oyen. Multiparametric MRI for prostate cancer localization in correlation to whole-mount histopathology. *Journal of Magnetic Resonance Imaging*, 37(6) :1392–1401, 2013.

- [43] G. J. Jager, E. T. Ruijter, C. a. van de Kaa, J. J. de la Rosette, G. O. Oosterhof, J. R. Thornbury, and J. O. Barentsz. Local staging of prostate cancer with endorectal MR imaging : correlation with histopathology. *AJR. American journal of roentgenology*, 166(4) :845–52, 1996.
- [44] F. Khalvati, A. Wong, and M. A. Haider. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Medical Imaging*, 15(1) :27, 2015.
- [45] B. Kim, R. H. Breau, D. Papadatos, D. Fergusson, S. Doucette, I. Cagiannos, and C. Morash. Diagnostic accuracy of surface coil magnetic resonance imaging at 1.5 T for local staging of elevated risk prostate cancer. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada*, 4(4) :257–62, aug 2010.
- [46] C. K. Kim, B. K. Park, and B. Kim. Localization of prostate cancer using 3T MRI : comparison of T2-weighted and dynamic contrast-enhanced imaging. *Journal of Computer Assisted Tomography*, 30(1) :7–11, 2006.
- [47] C. K. Kim, B. K. Park, and B. Kim. High-b-value diffusion-weighted imaging at 3 T to detect prostate cancer : Comparisons between b values of 1,000 and 2,000 s/mm². *American Journal of Roentgenology*, 194(1) :33–37, 2010.
- [48] C. R. King and J. P. Long. Prostate biopsy grading errors : a sampling problem ? *Int J Cancer.*, 90(6) :326–30, 2000.
- [49] G. Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau. Computer-Aided Detection and diagnosis for prostate cancer based on mono and multi-parametric MRI : A review. *Computers in Biology and Medicine*, 60 :8–31, 2015.
- [50] F. Li, M. Aoyama, J. Shiraishi, H. Abe, Q. Li, K. Suzuki, R. Engelmann, S. Sone, H. MacMahon, and K. Doi. Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy. *American Journal of Roentgenology*, 183(November) :1209–1215, 2004.
- [51] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Transactions on Medical Imaging*, 33(5) :1083–1092, may 2014.
- [52] P. Liu, S. Wang, B. Turkbey, K. Grant, P. Pinto, P. Choyke, B. J. Wood, and R. M. Summers. A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels. In *SPIE*, volume 8670, pages 1–6, 2013.
- [53] X. Liu and I. S. Yetik. Automated prostate cancer localization without the need for peripheral zone extraction using multiparametric MRI. *Medical physics*, 38(6) :2986–2994, 2011.
- [54] J. Mairal, F. Bach, and J. Ponce. Task-Driven Dictionary Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4) :791–804, apr 2012.
- [55] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th International Conference on Machine Learning*, pages 1–8, 2009.
- [56] H. L. Margaret Cheng, N. Stikov, N. R. Ghugre, and G. a. Wright. Practical medical applications of quantitative MR relaxometry. *Journal of Magnetic Resonance Imaging*, 36(4) :805–824, 2012.

-
- [57] S. Martin, M. Baumann, V. Daanen, and J. Troccaz. MR prior based automatic segmentation of the prostate in TRUS images for MR/TRUS data fusion. *2010 IEEE International Symposium on Biomedical Imaging : From Nano to Macro*, pages 640–643, 2010.
- [58] S. Martin, V. Daanen, and J. Troccaz. Atlas-based prostate segmentation using an hybrid registration. *International Journal of Computer Assisted Radiology and Surgery*, 3(6) :485–492, sep 2008.
- [59] Y. Mazaheri, A. Shukla-Dave, A. Muellner, and H. Hricak. MRI of the prostate : Clinical relevance and emerging applications. *Journal of Magnetic Resonance Imaging*, 33(2) :258–274, 2011.
- [60] M. Moradi, S. E. Salcudean, S. D. Chang, E. C. Jones, N. Buchan, R. G. Casey, S. L. Goldenberg, and P. Kozlowski. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *Journal of Magnetic Resonance Imaging*, 35(6) :1403–1413, 2012.
- [61] J. M. P. Nascimento and J. M. Bioucas Dias. Vertex component analysis : A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4) :898–910, 2005.
- [62] E. NIAF. *Aide au diagnostic du cancer de la prostate par IRM multi-paramétrique : une approche par classification supervisée*. PhD thesis, 2012.
- [63] E. Niaf, C. Lartizien, F. Bratan, L. Roche, M. Rabilloud, F. Mège-Lechevallier, and O. Rouviere. Prostate Focal Peripheral Zone Lesions : Characterization at Multiparametric MR Imaging—Influence of a Computer-aided Diagnosis System. *Radiology*, 271(3) :761–769, 2014.
- [64] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine and Biology*, 57(12) :3833–3851, jun 2012.
- [65] H. Núñez, L. Gonzalez-Abril, and C. Angulo. A post-processing strategy for SVM learning from unbalanced data. *ESANN 2011 proceedings, 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, (April) :195–200, 2010.
- [66] F.-b. Optimizations. TCM-KNN Scheme for Network Anomaly Detection Using. In *ACM symposium on Applied computing*, pages 2103–2109, New York, New York, USA, 2008.
- [67] S. Ozer, D. L. Langer, X. Liu, M. a. Haider, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, and I. S. Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*, 37(4) :1873, 2010.
- [68] A. R. Padhani and J. E. Husband. Dynamic contrast-enhanced MRI studies in oncology with an emphasis on quantification, validation and human studies. *Clinical radiology*, 56(8) :607–620, 2001.
- [69] M. Y. Park and J. Y. Byun. Understanding the mathematics involved in calculating apparent diffusion coefficient maps. *American Journal of Roentgenology*, 199(6) :2012, 2012.
- [70] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto. Quantitative analysis of multiparametric prostate MR images : differentiation between prostate cancer and normal tissue

- and correlation with Gleason score—a computer-aided diagnosis development study. *Radiology*, 267(3) :787–796, 2013.
- [71] N. Petrick, M. Haider, R. M. Summers, S. C. Yeshwant, L. Brown, E. M. Iuliano, A. Louie, J. R. Choi, and P. J. Pickhardt. CT colonography with computer-aided detection as a second reader : observer performance study. *Radiology*, 246(1) :148–156, 2008.
- [72] N. Petrick, B. Sahiner, S. G. Armato, A. Bert, L. Correale, S. Delsanto, M. T. Freedman, D. Fryd, D. Gur, L. Hadjiiski, Z. Huo, Y. Jiang, L. Morra, S. Paquerault, V. Raykar, F. Samuelson, R. M. Summers, G. Tourassi, H. Yoshida, B. Zheng, C. Zhou, and H.-P. Chan. Evaluation of computer-aided detection and diagnosis systems. *Medical Physics*, 40(8) :087001, aug 2013.
- [73] J. C. Platt. Probabilities for SV Machines. *Advances in large margin classifiers*, pages 61–74, 2000.
- [74] L. Poissonnier, J. Y. Chapelon, O. Rouviere, L. Curiel, R. Bouvier, X. Martin, J. M. Dubernard, and A. Gelet. Control of prostate cancer by transrectal HIFU in 227 patients. *Eur Urol*, 51(2) :381–387, 2007.
- [75] P. Puech, E. Potiron, L. Lemaitre, X. Leroy, G.-P. Haber, S. Crouzet, K. Kamoi, and A. Villers. Dynamic contrast-enhanced-magnetic resonance imaging evaluation of intraprostatic prostate cancer : correlation with radical prostatectomy specimens. *Urology*, 74(5) :1094–9, 2009.
- [76] S. F. Quinn, D. A. Franzini, T. A. Demlow, D. R. Rosencrantz, J. Kim, R. M. Hanna, and J. Szumowski. MR imaging of prostate cancer with an endorectal surface coil technique : correlation with whole-mount specimens. *Radiology*, 190(2) :323–327, 1994.
- [77] A. Rakotomamonjy. Variable selection using SVM-based criteria. *Journal of Machine Learning Research*, 3 :1357–1370, 2003.
- [78] A. Rosenkrantz, S. Kim, and R. Lim. 51. Prostate cancer Localization Using Multiparametric MR Imaging : Comparison of Prostate Imaging reporting and Data System (PI-RADS) and Likert Scales. *Radiology*, 269(2) :482–492, 2013.
- [79] A. B. Rosenkrantz, L. A. Ginocchio, D. Cornfeld, A. T. Froemming, R. T. Gupta, B. Turkbey, A. C. Westphalen, J. S. Babb, and D. J. Margolis. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon : A Multicenter Study of Six Experienced Prostate Radiologists. *Radiology*, 000(0) :152542, 2016.
- [80] O. Rouvière. Cancer de prostate : quelle imagerie en 2012 ?, 2012.
- [81] R. Rubin. Researchers Look to MRI and Biomarkers to Help Improve Detection of Aggressive Prostate Cancers. *Jama*, 313(7) :654–6, 2015.
- [82] B. Scholkopf and A. J. Smola. *Learning With Kernels : Support Vector Machines, Regularization, Optimization, and Beyond*, volume 129. 2002.
- [83] F. H. Schroder, A. B. Kruger, J. Rietbergen, R. Kranse, P. v. d. Maas, P. Beemsterboer, and R. Hoedemaeker. Evaluation of the Digital Rectal Examination as a Screening Test for Prostate Cancer. *JNCI Journal of the National Cancer Institute*, 90(23) :1817–1823, 1998.
- [84] V. Shah, B. Turkbey, H. Mani, Y. Pang, T. Pohida, M. J. Merino, P. A. Pinto, P. L. Choyke, and M. Bernardo. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Medical physics*, 39(7) :4093–103, 2012.

-
- [85] S. F. Shariat and P. I. Karakiewicz. Screening for prostate cancer in 2007 : the PSA era and its challenges are not over. *European urology*, 53(3) :457–460, 2008.
- [86] A. Shukla-Dave and H. Hricak. Role of MRI in prostate cancer detection. *NMR in Biomedicine*, 27(1) :16–24, 2014.
- [87] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements : spin echoes in the presence of a time-dependant field gradient. *The Journal of chemical physics*, 42(1) :5, 1965.
- [88] J. A. K. Suykens, M. Signoretto, and A. Argyriou. *Regularization, Optimization, Kernels, and Support Vector Machines*. The MIT Press ; 1st edition, 2015.
- [89] T. Tamada, N. Kanomata, T. Sone, Y. Jo, Y. Miyaji, H. Higashi, A. Yamamoto, and K. Ito. Magnetic Resonance Imaging in Prostate Cancer at 3 Tesla : Comparison with 1 , 000 s / mm ² for Tumor Conspicuity and Discrimination of Aggressiveness. *PLoS ONE*, 9(5) :3–10, 2014.
- [90] P.-n. Teng, N. W. Bateman, K. M. Darcy, C. A. Hamilton, G. L. Maxwell, C. J. Bakkenist, T. P. Conrads, O. Service, W. Reed, N. Military, and I. F. Hospital. HHS Public Access. *Gynecol Oncol.*, 136(3) :554–561, 2015.
- [91] R. J. Tibshirani and R. Tibshirani. A bias correction for the minimum error rate in cross-validation. *The Annals of Applied Statistics*, 3(2) :822–829, jun 2009.
- [92] P. Tiwari, J. Kurhanewicz, and A. Madabhushi. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Medical image analysis*, 17(2) :219–35, 2013.
- [93] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi. Multi-modal wavelet embedding representation for data combination (MaWERiC) : integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR in Biomedicine*, 25(4) :607–619, apr 2012.
- [94] B. Turkbey and P. Albert. Imaging Localized Prostate Cancer : Current Approaches and New Developments. *American Journal of Roentgenology*, 192(6) :1471–1480, 2009.
- [95] T. Vaché, F. Bratan, F. Mège-Lechevallier, S. Roche, M. Rabilloud, and O. Rouvière. Characterization of Prostate Lesions as Benign or Malignant at Multiparametric MR Imaging : Comparison of Three Scoring Systems in Patients Treated with Radical Prostatectomy. *Radiology*, 272(2) :446–55, 2014.
- [96] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik. Dimensionality Reduction : A Comparative Review. *Journal of Machine Learning Research*, 10 :1–41, 2009.
- [97] S. Varma and R. Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7 :91, jan 2006.
- [98] S. Vieillot, P. Fenoglietto, N. Aillères, M.-H. Hay, J.-B. Dubois, and D. Azria. Cancer de la prostate. Technical Report 0, Haute Autorité de Santé, 2010.
- [99] S. Viswanath, B. N. Bloch, J. Chappelow, P. Patel, N. Rofsky, R. Lenkinski, E. Genega, and A. Madabhushi. Enhanced multi-protocol analysis via intelligent supervised embedding (EMPrAvISE) : detecting prostate cancer on multi-parametric MRI. In *SPIE*, volume 7963, pages 79630U–79630U–15, 2011.
- [100] S. Viswanath, B. N. Bloch, M. Rosen, J. Chappelow, R. Toth, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur, and A. Madabhushi. Integrating structural and

- functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 Tesla MRI. In N. Karssemeijer and M. L. Giger, editors, *SPIE*, volume 7260, pages 72603I–72603I–12, feb 2009.
- [101] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. *Physics in Medicine and Biology*, 57(6) :1527–1542, mar 2012.
- [102] P. C. Vos, T. Hambrock, J. O. Barentsz, and H. J. Huisman. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Physics in medicine and biology*, 55(6) :1719–1734, 2010.
- [103] L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, K. Kuroiwa, and H. Hricak. Assessment of biologic aggressiveness of prostate cancer : correlation of MR signal intensity with Gleason grade after radical prostatectomy. *Radiology*, 246(1) :168–176, 2008.
- [104] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, H. C. Thoeny, and S. Verma. PI-RADS Prostate Imaging - Reporting and Data System : 2015, Version 2. *European Urology*, 69(1) :16–40, 2016.
- [105] A. Wong, J. Glaister, A. Cameron, and M. Haider. Correlated diffusion imaging. *BMC medical imaging*, 13 :26, 2013.
- [106] G. Wu and E. Chang. Class-boundary alignment for imbalanced dataset learning. In *The Twentieth International Conference on Machine Learning (ICML)*, number 1, pages 49–56, Washington DC, 2003.

TITRE EN FRANCAIS

Détection et caractérisation du cancer de la prostate par images IRM 1.5T multiparamétriques

RESUME EN FRANCAIS

Le cancer de la prostate est le plus courant en France et la 4^{ème} cause de mortalité par cancer. L'imagerie IRM multiparamétrique est une technique d'imagerie prometteuse pour le diagnostic et la prise en charge de ce cancer. Néanmoins, l'interprétation visuelle des multiples séquences IRM n'est pas aisée. Dans ces conditions, un fort intérêt s'est porté sur le développement de systèmes d'aide au diagnostic (CAD) dont le but est d'assister le radiologue dans ses décisions. Cette thèse présente la conception d'un système CAD qui fournit au radiologue une cartographie probabiliste des localisations cancéreuses dans la zone périphérique de la prostate. Ce CAD repose sur une base d'images IRM multi-paramétrique (IRM-mp) 1.5T, comprenant pour chacun des 49 patients inclus, une image pondérée en T2, une série d'images de perfusion en pondération T1 et une image de diffusion. La localisation des lésions suspectes est obtenue par la confrontation des annotations de chaque image réalisées par deux radiologues avec l'analyse des coupes histologiques des pièces de prostate. Pour cette étude, nous avons utilisé une méthode d'apprentissage de dictionnaires permettant d'extraire de nouvelles caractéristiques descriptives dont l'objectif est de discriminer chacun des cancers. Cet ensemble de descripteurs, obtenu à l'échelle du voxel, est utilisé pour apprendre un modèle de classification permettant de prédire la classe de chaque voxel et de produire ainsi une cartographie de la probabilité de présence du cancer. Deux méthodes de classification supervisée ont été comparées, la régression logistique et le séparateur à vaste marge (SVM). Dans cette étude, nous avons concentré nos efforts sur la discrimination des cancers agressifs (Gleason >6) et fourni une analyse de la corrélation entre probabilités et scores de Gleason. Le système développé montre de très bonnes performances de détection des cancers agressifs mais ne permet pas, en revanche, de distinguer les grades intermédiaires (Gleasons 6, 7, 8 et 9) du cancer.

TITRE EN ANGLAIS

Prostate cancer detection and characterization based on 1.5T multiparametrics MRI

RESUME EN ANGLAIS

Prostate cancer is the most frequent and the fourth leading cause of mortality in France. Actual diagnosis methods are often insufficient in order to detect and precisely localize cancer. Multiparametrics MRI is now one of the most promising method for accurate follow-up of the disease. However, the visual interpretation of MRI is not easy and it is shown that there is strong variability among expert radiologists to perform diagnosis, especially when MR sequences are contradictory. Under these circumstances, a strong interest is for Computer-aided diagnosis systems (CAD) aiming at assisting expert radiologist in their final decision. This thesis presents our work toward the conception of a CADe which final goal is to provide a cancer probability map to expert radiologist. This study is based on a rich dataset of 49 patients made of T2w, dynamic and diffusion MR images. The ground truth was obtained through strict process of annotations and correlation between histology and MRI. This thesis focuses both for cancer detection and characterization in order to provide a cancer probability map correlated to cancer aggressiveness (Gleason score). To that end we used a dictionary learning method to extract new features to better characterize cancer aggressiveness signatures as well as image features. Those features are then used as an input to Support Vector Machines (SVM) and Logistic Regression (LR) classifiers to produce a cancer probability map. We then focused on discriminating aggressive cancers (Gleason score >6) from other tissues and provided an analysis of the correlation between cancer aggressiveness and probabilities. Our work conclude on a strong capability to distinguish aggressive cancer from other tissues but fails to precisely distinguish intermediate grades of cancers (Gleasons 6, 7, 8 and 9).

MOTS-CLES

Imagerie par résonance magnétique multi-paramétrique, Cancer de la prostate, Systèmes d'aide à la décision, Apprentissage supervisé, Séparateurs à vaste marge, Apprentissage de dictionnaires

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE

INSERM, U1032, LabTau, Lyon, F-69003, France, 151, cours Albert Thomas 69424 LYON Cedex 03. Université de Lyon, CREATIS ; CNRS UMR5220 ; Inserm U1044 ; INSA-Lyon ; Université Lyon 1, 7 Av. Jean Capelle, 69621 VILLEURBANNE, France.