

ACADÉMIE DE MONTPELLIER
UNIVERSITÉ MONTPELLIER II
Sciences et Techniques du Languedoc

THÈSE

présentée au Laboratoire d'Informatique de Robotique
et de Microélectronique de Montpellier pour
obtenir le diplôme de doctorat

Spécialité : **Informatique**
Formation Doctorale : **Informatique**
École Doctorale : **Information, Structures, Systèmes**

Caractérisation des lésions hépatiques focales sur des acquisitions scanner multiphasiques

par

Auréline QUATREHOMME

Version du 23 septembre 2014

Directeur de thèse

William PUECH, Professeur LIRMM, Université Montpellier II

Co-Directeur de thèse

Gérard SUBSOL, Chargé de Recherche CNRS LIRMM, Université Montpellier II

Rapporteurs

Frederic PRECIOSO, Professeur Laboratoire d'Informatique, Signaux et Systèmes, Université de Nice Sophia-Antipolis

Christian ROUX, Professeur Laboratoire de Traitement de l'Information Médicale, Telecom Bretagne

Examineurs

Khalifa DJEMAL, Maître de conférences Laboratoire IBISC, Université d'Évry Val d'Essonne

Gilles GESQUIÈRE, Professeur LIRIS, Université Lumière Lyon 2

Denis HOA, Docteur en médecine Imaios



Table des matières

Table des matières	i
1 Introduction	1
1.1 L'apport de l'imagerie médicale radiologique	2
1.2 Vers le Diagnostic Assisté par Ordinateur	4
1.3 Application à la caractérisation des lésions hépatiques	5
1.4 Plan du manuscrit	9
I État de l'art	11
2 Descripteurs visuels	13
2.1 Présentation générale	14
2.2 Caractérisation des niveaux de gris ou des couleurs	14
2.3 Caractérisation de la texture	16
2.3.1 Définition de la texture	17
2.3.2 Descripteurs fondés sur les statistiques	17
2.3.3 Descripteurs fondés sur des modèles	22
2.3.4 Descripteurs fondés sur une représentation fréquentielle	23
2.3.5 Problématiques liées à la texture	27
2.3.6 Synthèse	27
2.4 Caractérisation de la forme	27
2.5 Conclusion	29
3 Classification supervisée	31
3.1 Présentation générale	32

3.2	Principaux algorithmes de classification supervisée	33
3.2.1	Arbres de décision	33
3.2.2	Réseaux de neurones	35
3.2.3	Support Vector Machine	35
3.2.4	Sac de mots ("bag of words")	36
3.2.5	Synthèse des méthodes présentées	37
3.2.6	Apprentissage par ensembles	38
3.3	Evaluation des performances	39
3.3.1	Techniques d'évaluation	39
3.3.2	Mesures d'évaluation	40
3.3.3	Le cas des effectifs de classe déséquilibrés	42
3.4	Conclusion	42
4	Caractérisation des lésions hépatiques dans des images scanner	43
4.1	Caractérisation des lésions hépatiques dans des acquisitions monophasiques	44
4.1.1	Objectifs	44
4.1.2	Données	46
4.1.3	Méthodes	46
4.1.4	Résultats	47
4.1.5	Synthèse des travaux	47
4.2	Caractérisation des lésions hépatiques sur acquisitions multiphasiques	48
4.2.1	Un système de Diagnostic Associé par Ordinateur entièrement auto- matique	48
4.2.2	Une méthode de classification en 3 classes	49
4.2.3	Une méthode de classification en 4 classes utilisant des descripteurs temporels	50
4.2.4	Synthèse des travaux	53
4.3	Conclusion	54
II	Contributions	55
5	Mise en place des données et développement d'une première méthode de ca- ractérisation sur des examens monophasiques	57
5.1	Introduction	58
5.2	Description des données	58
5.2.1	Définition clinique des classes	58
5.2.2	Description visuelle des classes	60
5.2.3	Sélection et récupération des images médicales	62
5.2.4	Régions d'intérêt	62
5.2.5	Présentation des bases de données utilisées	63

5.3	Description de la méthode	64
5.3.1	Principe	64
5.3.2	Programmation	66
5.4	Analyse des résultats	66
5.4.1	Présentation	66
5.4.2	Discussion	67
5.5	Conclusion	68
6	Examens scanner multiphasiques	69
6.1	Introduction et comparaison avec les systèmes existants	70
6.2	Amélioration des résultats par la prise en compte du multiphase	72
6.2.1	Analyse des résultats	72
6.3	Comparaison avec une classification par des experts	74
6.4	Problème de la normalisation des intensités	77
6.4.1	Principe	79
6.4.2	Influence sur la dynamique temporelle des intensités	79
6.4.3	Influence sur la classification	82
6.5	Conclusion	83
7	Intégration de nouveaux descripteurs visuels	85
7.1	Introduction	86
7.2	Un descripteur visuel plus rapide pour la texture	86
7.2.1	Temps de calcul	86
7.2.2	Matrice de co-occurrence	86
7.2.3	Alternatives	87
7.2.4	Expérimentations	89
7.2.5	Synthèse autour de la matrice de co-occurrence	91
7.3	Des descripteurs visuels adaptés au multi-phase	91
7.3.1	Descripteurs temporels issus des travaux existants	92
7.3.2	Profils temporels	92
7.3.3	Cartes de densité	95
7.4	Conclusion	99
8	Expérimentations autour la caractérisation de lésions hépatiques	101
8.1	Introduction	102
8.2	Pré-traitement des données	102
8.3	Sélection d'un sous-ensemble d'attributs	103
8.3.1	Proposition d'un sous-ensemble de groupes de descripteurs	103
8.3.2	Sélection d'un sous-ensemble d'attributs à partir de la matrice de corrélation	105
8.3.3	Synthèse autour de la sélection	105

8.4	Introduction d'une classe "inconnue"	106
8.4.1	Base de données 3 sans abcès ni HNF	107
8.4.2	Reconnaissance d'une classe	107
8.4.3	Caractérisation clinique sain - pathologique - cancéreux	108
8.5	Conclusion	109
9	Conclusion et Perspectives	111
9.1	Résumé de nos contributions	112
9.2	Travaux futurs	114
	Listes des publications	117
	Bibliographie	119
	Table des figures	125
	Liste des tableaux	126

Introduction

Préambule

Ce chapitre présente le contexte collaboratif et scientifique dans lequel s'inscrivent ces travaux. Ces dernières années ont vu les contenus visuels prendre une part de plus en plus importante dans notre société, et ce, dans tous les domaines. La médecine n'y fait pas exception. Nous allons tout d'abord nous intéresser aux avancées technologiques en radiologie, avant de nous pencher sur les applications de l'imagerie médicale, et enfin d'évoquer le contexte industriel entourant ce projet de thèse qui porte sur la caractérisation automatique des lésions focales hépatiques sur des images médicales provenant de scanners.

Contents

1.1	L'apport de l'imagerie médicale radiologique	2
1.2	Vers le Diagnostic Assisté par Ordinateur	4
1.3	Application à la caractérisation des lésions hépatiques	5
1.4	Plan du manuscrit	9

1.1 L'apport de l'imagerie médicale radiologique

Les premières technologies d'acquisition d'images médicales sont apparues dans les années 1900 avec la découverte des rayons X par Röntgen. L'image présentée dans la Figure 1.1 a été publiée dans son article fondateur [Rö95].

Une nette accélération des évolutions matérielles est observable depuis la fin du XXème siècle. Le lecteur pourra se référer à [TB13] pour plus de détails sur ce sujet.

L'objectif de toutes les techniques d'acquisition est de former une représentation visuelle d'un ensemble d'informations médicales qui soit synthétique et compréhensible pour l'être humain. Selon l'objectif clinique, l'examen sera structurel (si on s'intéresse à l'anatomie) ou fonctionnel (si on s'intéresse au fonctionnement des organes).

Les médecins font appel à l'imagerie dans de nombreuses situations :

- afin d'établir un diagnostic
- dans le but d'évaluer la sévérité et/ou l'évolution d'une pathologie
- avec pour objectif de mesurer l'efficacité d'un traitement
- pour obtenir les informations nécessaires en vue d'une opération
- afin de faire avancer les connaissances sur la physiologie ou les pathologies

Ce sont les radiologues qui sont spécialisés dans l'acquisition et l'interprétation des images médicales. Ces dernières années, leur nombre, tout comme celui du nombre d'examen radiologiques pratiqués, a augmenté de manière significative : tous les autres spécialistes médicaux sont demandeurs de leurs compétences.

Bien entendu, le choix de la modalité d'acquisition dépendra du besoin clinique : certains types d'examen seront plus adaptés à des parties du corps spécifiques, mais tout à la fois la même partie anatomique peut être acquise selon différentes modalités. Le point essentiel de l'imagerie médicale est que toute exposition présente un risque : le principe de précaution préconise des doses de rayonnement les plus faibles possibles.

Certaines modalités exposent le patient à des rayonnements (rayons X), d'autres non (Imagerie par Résonance Magnétique ou IRM), mais les coûts financiers et les contre-indications ne sont pas les mêmes. Par exemple il n'est pas possible pour un patient portant du métal, stimulateur cardiaque ou autre, de passer une IRM.

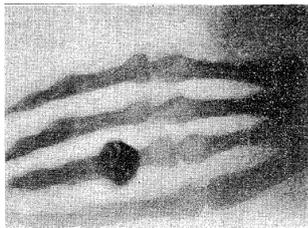


FIGURE 1.1 – *L'une des premières images radiologiques réalisée sur une main humaine [Rö95]*

Les questions que se pose le radiologue afin de déterminer une modalité et un protocole d'acquisition ont été posées par l'International Commission for Radiation Protection (ICRP) ¹ :

- **Principe de justification** : l'acte médical apporte-t-il un bénéfice supérieur au risque supposé ?
- **Principe d'optimisation** : comment ajuster le protocole d'acquisition (modalité, dose de rayon et/ou de produit de contraste, timing) afin d'obtenir les informations souhaitées (donc ni trop, ni trop peu) pour établir un diagnostic ? Cette question étant encadrée par celle qui suit.
- **Principe de limitation** : comment limiter au maximum le rayonnement d'une source donnée reçu par un individu ?

Beaucoup de problématiques se regroupent autour de l'imagerie médicale, dont voici les principales :

L'acquisition : de nombreuses recherches sont menées afin d'obtenir des examens plus précis, de meilleure qualité, moins nocifs, et également pour s'affranchir de certaines contraintes comme la sédentarité de l'équipement (par exemple un appareil à ultrasons permettant de réaliser des échographies branché sur un smartphone).

Le format des données : le DICOM (Digital Imaging and COmmunications in Medicine) est le standard défini en 1985 par l'American College of Radiology et diffusé universellement chez tous les constructeurs de matériel d'imagerie médicale ². Il permet d'insérer dans un seul fichier les données visuelles de l'image ainsi que d'autres informations comme des détails sur le patient, le matériel et le protocole employés, la matrice image elle-même. Il définit également le protocole de transmission de ces données. Différents encodages des images sont possibles, entre 8 et 12 bits.



Le stockage et l'accès aux examens : étant donné le format volumineux des images médicales évoqué ci-dessus, le nombre croissant d'examens d'imagerie pratiques et d'images acquises lors de chaque examen grâce à l'amélioration des machines, l'archivage ainsi que l'accès aux examens des patients est un problème à part entière, que tentent de résoudre les systèmes nommés PACS (Picture Archiving and Communication System). Les données sont stockées sur des serveurs auxquels accèdent de façon sécurisée tous les périphériques nécessaires : les machines d'acquisition, les ordinateurs des établissements de santé et les imprimantes. Les examens sont généralement compressés après un certain laps de temps. Ainsi, lorsqu'un radiologue

1. www.icrp.org

2. <http://medical.nema.org>

souhaite accéder à un examen ancien, le système le décompresse avant de le mettre à disposition de l'ordinateur du médecin.

La visualisation : à partir de l'ensemble des coupes 2D d'une acquisition scanner, il est par exemple intéressant d'obtenir une reconstruction volumique de ce que l'on souhaite examiner.

La compression des images médicales : la taille importante des données d'imagerie médicale pose également la question de la compression, qui fait débat. En effet, si elle résout de nombreux soucis, notamment de stockage, la question de l'apparition d'éventuels artefacts sur les images qui pourraient compromettre l'analyse du médecin se pose.

Les systèmes de diagnostic assisté par ordinateur : nous les verrons plus en détail dans la prochaine section.

1.2 Vers le Diagnostic Assisté par Ordinateur

Le Diagnostic Assisté par Ordinateur, ou Computer Aided Diagnosis (CAD) en anglais, regroupe les procédures informatiques permettant d'assister les médecins dans l'interprétation des images médicales (consulter [Doi07] pour un état de l'art). Le but n'étant évidemment pas de remplacer le radiologue, dont l'expertise reste indispensable, mais plutôt de le guider et de lui faire gagner du temps.

Le CAD est donc à la frontière de plusieurs domaines : l'intelligence artificielle, le traitement d'images et la médecine. Cette pluralité est à la fois une richesse et une difficulté : comment combler l'écart entre la pratique clinique et l'informatique ?

En s'intéressant au raisonnement des cliniciens conduisant à l'établissement d'un diagnostic, on se rend compte que leur cheminement est relativement intuitif et basé sur l'expérience, malgré quelques points clairement identifiés. Le praticien n'est pas en mesure de décrire tous les éléments qu'il a synthétisé de façon inconsciente pour prendre sa décision. C'est un processus difficilement reproductible par un algorithme simple, il est donc naturel de se tourner vers des techniques d'intelligence artificielle pour tenter de le reproduire informatiquement.

Plusieurs objectifs co-existent pour ces systèmes d'assistance :

- **Recherche par le contenu :** (Content-Based Image Retrieval ou CBIR en anglais) dans une base de données, quelle est l'image la plus proche de l'image requête ?
- **Reconnaissance :** une image correspond-t-elle à un modèle donné ?
- **Classification :** quelle est la classe représentative d'une image ?
- **Détection et segmentation :** l'image contient-elle un objet donné, et si oui où se situe-t-il et quelle forme a-t-il ?
- **Alignement et recalage :** comment transformer une image afin de correspondre à une autre, en partie ou en totalité ?

- **Approximation et simplification** : comment construire une image similaire à l'originale mais basée sur moins d'éléments ?

Cette discipline est relativement récente. Elle a bénéficié de belles avancées, comme par exemple le diagnostic de cancer du sein dans les mammographies et du cancer du poumon sur les images scanner. En effet, ces pathologies sont des priorités du secteur de la santé en raison de leur impact sur la société actuelle.

Pour ces exemples précis, des systèmes complets sont déjà en place. Un exemple parmi d'autres, la société iCAD³ qui commercialise des solutions d'aide au diagnostic depuis 2002 vendues dans le monde entier, suite à un projet de recherche de l'Université de Chicago.

La tendance s'est maintenant répandue à d'autres organes ou pathologies, notamment le cerveau (problématiques autour d'Alzheimer ou du vieillissement) et le foie. C'est ce dernier, représentant une problématique de santé majeure en raison du cancer hépatique, qui nous intéresse ici.

1.3 Application à la caractérisation des lésions hépatiques

Un certain nombre d'enjeux cliniques autour de l'imagerie du foie co-existent, dont le plus important est le dépistage du cancer primaire du foie qui peut emporter les patients en seulement quelques mois. Une étude mondiale⁴ effectuée par l'International Agency for Research on Cancer (IARC) révèle que le cancer du foie est le cinquième cancer en termes de fréquence chez les hommes, septième chez les femmes. Il est la troisième cause de mortalité mondiale par le cancer. En effet, il est fatal dans 93% des cas. Le nombre de décès dans le monde entier a ainsi été estimé à 694 000 en 2008. Une des raisons de cette mortalité est la découverte souvent tardive du cancer de ce type.

Il est donc important de pouvoir être capable de distinguer des lésions des tumeurs malignes (Carcinome Hépatocellulaire (CHC) qui sont les tumeurs primitives du foie, métastases les tumeurs secondaires) de lésions bénignes comme les kystes.

Les examens d'imagerie médicale pratiqués autour du foie sont des explorations ultrasonores de type échographie ou écho-endoscopie, des IRM ou des scanners. De nombreux détails sur cet aspect clinique sont exposés dans [BL05]. L'échographie est une méthode peu invasive et de faible coût, qui peut se réaliser au lit même du patient. C'est d'ailleurs dans la plupart des cas l'examen de première intention. Le scanner est plus contraignant puisqu'il oblige d'une part à prévoir l'examen, d'autre part le patient à un jeûne d'une douzaine d'heures. De plus, le patient est soumis à des radiations. Mais il est de coût moyen est ne dure que 5 minutes. L'IRM, quant à lui, n'irradie pas le patient, mais a un coût plus élevé et l'examen dure une trentaine de minutes. Il permet au praticien de voir de sub-

3. www.icamed.com

4. <http://globocan.iarc.fr>

tiles variations dans les tissus mous. La Figure 1.2 présente le foie et de façon plus générale l'abdomen acquis selon ces trois modalités.

Notre travail se restreint aux examens scanner, qui de par leur coût moins élevé sont plus répandus que l'IRM. Les travaux existants sur le sujet sont également effectués sur cette modalité d'images. La Tomographie acquiert une série importante d'images 2D de rayons X, prises autour d'un axe de rotation unique.

Les images acquises sont au format DICOM. Elles possèdent une dynamique de plus de



FIGURE 1.2 – Echographie (gauche), scanner (milieu) et IRM (droite, couleurs artificielles) de l'abdomen

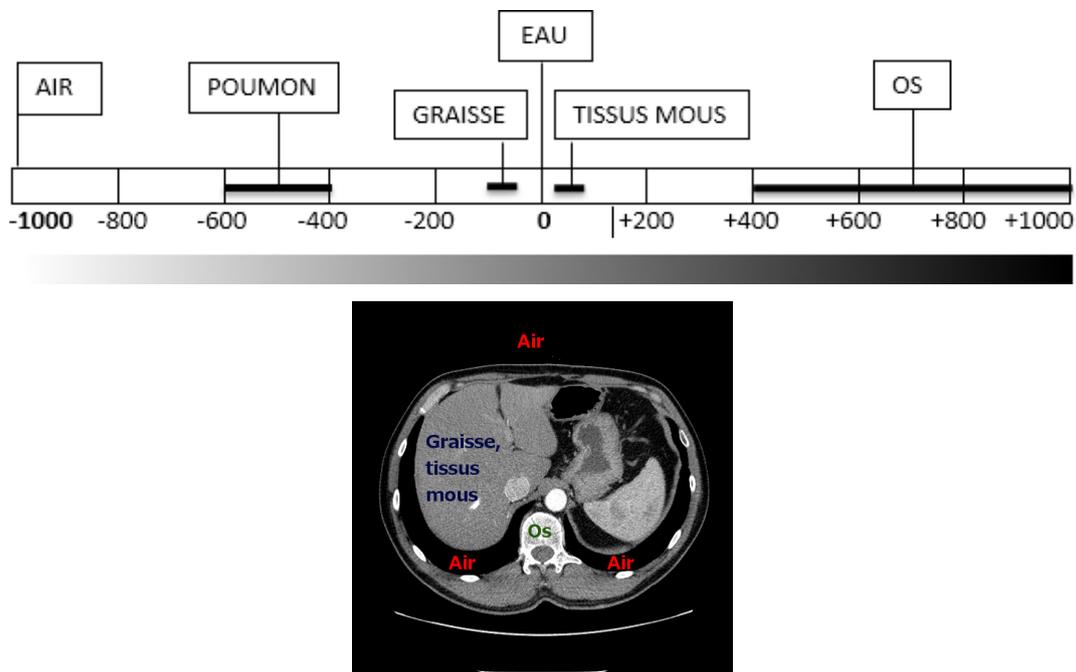


FIGURE 1.3 – Illustration de l'échelle de Hounsfield et de sa signification sur le type de tissu dans les images scanner

2000 niveaux de gris différents, suivant l'échelle de Hounsfield. Celle-ci est élaborée à partir des coefficients d'atténuation du tissu dans chaque élément rapporté à celui de l'eau. Elle indique, en fonction d'un niveau de gris donné, une certaine densité et donc un type de tissu spécifique : les os sont blancs, l'air est noir, la graisse et les tissus mous sont dans des nuances de gris, comme cela est illustré dans la Figure 1.3. Des problèmes de complexité se posent dès lors que les images sont aussi volumineuses.

Dans le but de réduire cet intervalle immense d'intensités, le radiologue réalise un fenêtrage : il choisit une gamme significative de valeurs de pixels. Les paramètres sont tout d'abord choisis en fonction de la zone anatomique, puis éventuellement ajustés par exemple lorsqu'une anomalie est détectée et que le radiologue souhaite en examiner les détails le plus précisément possible. La Figure 1.4 présente deux images scanner, l'une de l'abdomen et l'autre du thorax, visualisées selon deux fenêtrages différents.

Afin d'améliorer le contraste des images acquises, et donc augmenter l'efficacité du diagnostic, l'injection d'un produit de contraste est largement répandue. Une première série d'images est tout d'abord acquise, il s'agit de la phase de pré-injection ou de non-réhaussement. Ensuite, le patient reçoit l'injection, et 3 séries d'images sont acquises à des temps différents. La première, juste après l'injection, est appelée la phase artérielle. La seconde, quelques dizaines de secondes après, la phase portale. Et la dernière, appelée phase tardive, est acquise quelques minutes après. La diffusion dans le temps du produit réhausse les vaisseaux et les lésions. Les radiologues utilisent systématiquement les

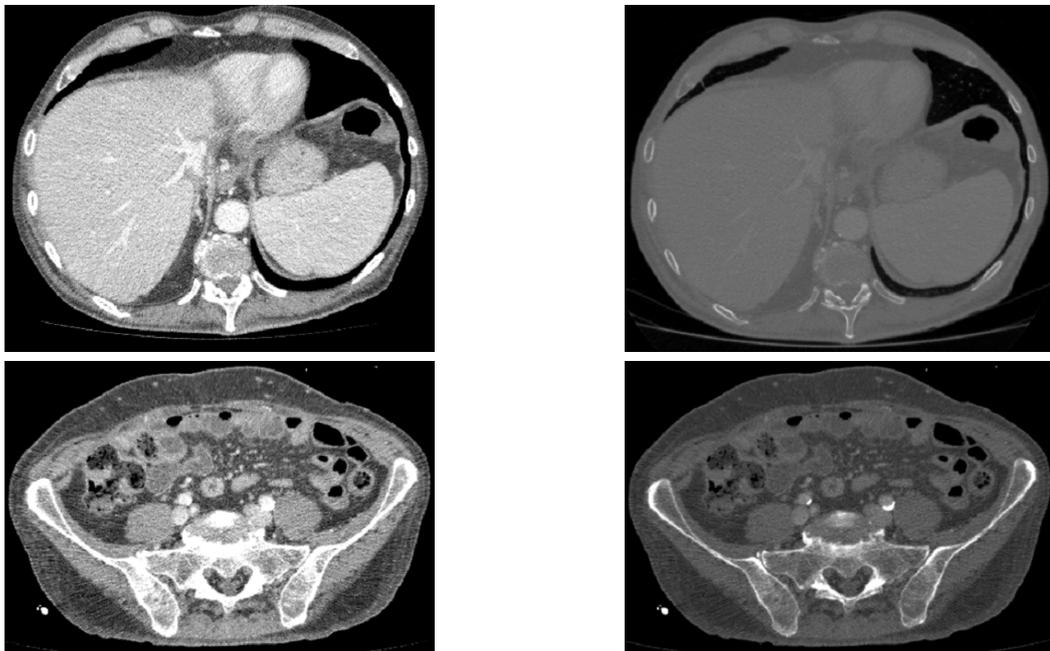


FIGURE 1.4 – Image scanner de l'abdomen (haut) et du pelvis (bas) selon deux fenêtrages

informations essentielles données par cette diffusion spatiale et temporelle au cours des phases. En effet, le réhaussement du contraste varie d'une phase à une autre, et une lésion qui ne se distingue pas du foie sain dans une phase pourra être "révélée" dans une autre. Le réhaussement peut être homogène ou très hétérogène, sur toute la lésion ou périphérique, en fonction des phases et du type de lésion.

Un exemple illustratif est proposé dans la Figures 1.5, tandis que les stratégies d'acquisition en fonction du contexte clinique, extraites d'informations écrites dans [BM09], sont présentées dans le tableau 1.1.

Plusieurs approches sont envisageables pour répondre aux problématiques cliniques relatives aux lésions du foie sur des images scanner. Il est intéressant de détecter dans une image du foie une structure anormale. Il est tout aussi important de pouvoir différencier les types de lésions, et c'est ce sur quoi porte notre travail. Notre volonté est non seulement de pouvoir distinguer des tumeurs malignes des nodules bénins, mais également d'aller plus loin en construisant un système capable de reconnaître différentes classes de lésions, couvrant la majorité des cas pouvant se présenter.

Pour pouvoir caractériser les nodules, deux étapes sont nécessaires. La première est

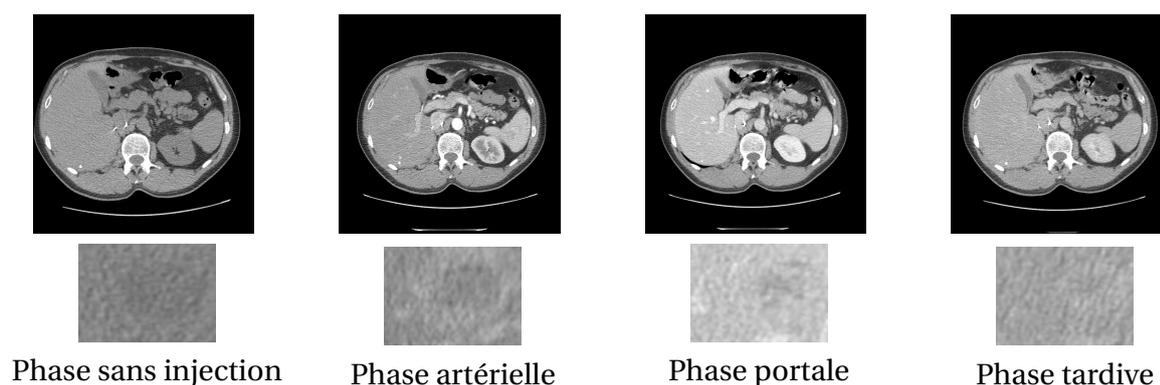


FIGURE 1.5 – Illustration de la tomographie multiphasique : abdomen (haut) et lésion hépatique (abcès, en bas) acquis sur différentes phases

SCÉNARIO	CONTEXTE CLINIQUE
Monophasique (portale)	Pas de suspicion d'une condition hépatique pathologique spécifique
Duophasique (portale, tardive)	Scénario pathologique dont la cause primaire est en dehors du foie (suspicion de métastases hépatiques hypervasculaires)
Triphasique (pré-injection, artérielle, portale)	Cirrhose, CHC, adénome ou Hyperplasie Nodulaire Focale connue ou suspectée

TABLE 1.1 – Scénarios d'acquisition scanner du foie en fonction du contexte clinique

une étape d'extraction de descripteurs visuels : il s'agit d'extraire informatiquement des images les caractéristiques d'une lésion, notamment des informations sur sa texture comme nous le verrons par la suite. Ensuite, c'est sur ces éléments que s'appuie un algorithme de classification supervisée afin de différencier les types de lésions. Comme cela a été évoqué, il est essentiel de s'appuyer sur des mécanismes d'apprentissage pour pouvoir tenter de reproduire le processus diagnostique du radiologue. Cette thèse étant réalisée dans le domaine du traitement d'images, le choix a été fait de se concentrer sur les descripteurs, plutôt que sur la classification. Dans ce travail, l'effort sera donc fourni dans le but d'optimiser les descripteurs existants, et d'en proposer de nouveaux. La classification s'appuiera sur des algorithmes reconnus.

Ce projet est au coeur d'une collaboration entre IMAIOS, une jeune start-up proposant divers supports multimédia de formation à destination des professionnels de la santé⁵ et l'équipe ICAR du Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier (LIRMM). L'objectif du projet global de Recherche et Développement chez Imaios est de développer un outil permettant par exemple à un médecin généraliste ou un étudiant, de sélectionner la partie des images qu'il souhaite analyser, soit dans une seule image, soit dans une série d'images, de l'envoyer en ligne sur le site d'Imaios et de récupérer des informations sur cette zone, par exemple la pathologie et des informations cliniques associées. Le champ est réduit dans ce manuscrit aux lésions focales du foie sur des acquisitions scanner.

1.4 Plan du manuscrit

L'imagerie médicale connaît actuellement une formidable explosion du nombre et de la qualité des techniques qui y sont associées. L'ensemble des disciplines du secteur de la santé pourrait bénéficier des améliorations qui ne manqueront pas d'en découler. Un état de l'art sera réalisé dans une première partie, avant de détailler nos contributions.

Les travaux existants seront abordés dans cette première section en balayant les différents domaines à la croisée du diagnostic assisté par ordinateur. Les Chapitres 2 et 3, techniques, détailleront les méthodes permettant d'extraire des descripteurs visuels des images, ainsi que les algorithmes de classification et les protocoles d'évaluation de celle-ci. Ils se restreindront au contexte d'application qui nous intéresse ici, à savoir les lésions du foie sur image scanner. Les descripteurs visuels sont des mesures extraites des images, elles représentent leurs niveaux de gris, leur forme ou leur texture, et ce sont les représentations de la texture qui dominent. Les algorithmes de classification d'apprentissage supervisé seront présentés après une brève introduction historique aux prémices de la classification. Les méthodes employées afin d'évaluer leurs performances seront également décrites. Une synthèse des études portant sur des images scanner de lésions hépatiques sera ensuite

5. www.imaios.com

été effectuée dans le Chapitre 4. Les travaux applicatifs portant sur des acquisitions scanner monophasiques, puis multiphasiques, seront explorés : les différents objectifs, bases de données, méthodes (descripteurs et classificateurs), résultats. Enfin, une synthèse sera présentée en fin de partie afin de conclure sur les travaux existants.

Dans la section des contributions sera tout d'abord présenté notre système complet de classification dans le Chapitre 5 : la présentation des données, la construction de la base de lésions, le premier ensemble de descripteurs choisis parmi ceux de la littérature, et la méthode de classification. Les résultats obtenus seront analysés. Puis nous introduirons dans le Chapitre 6 des examens multiphasiques dans le but de prendre en compte le réhaussement temporel des lésions au cours du temps, ce qui, nous le montrerons, permet d'améliorer les résultats de classification sur notre base de données. L'emploi du multiphase sera de plus validé par une expérience faisant intervenir deux experts radiologues. De nouveaux descripteurs seront ensuite proposés dans le Chapitre 7, notamment des mesures adaptées au contexte multiphasique. Un descripteur de texture plus rapide et plus performant remplacera les classiques mesures statistiques d'Haralick. Notre principale contribution, les profils temporels et les cartes de densité des lésions seront détaillés et les résultats obtenus analysés et comparés aux précédents. Enfin, notre outil sera appliqué dans le chapitre 8 à diverses caractérisation des lésions, en plus de la classification multi-classe initiale, et un ensemble de descripteurs plus performant sera extrait de l'ensemble initial.

Nous terminerons en présentant les perspectives de nos travaux dans le Chapitre 9.

Première partie

État de l'art

Descripteurs visuels

Préambule

Dans ce chapitre sur les descripteurs visuels, nous allons tout d'abord introduire les notions essentielles, avant d'évoquer les différentes familles de caractéristiques, déterminées en fonction de leur objet : niveaux de gris, texture et forme, et de conclure par une synthèse des attributs qui auront été détaillés. Nous nous restreignons aux éléments présents dans les travaux portant sur l'imagerie scanner du foie. Comme nous le verrons, c'est principalement des descripteurs de texture qui sont calculés. Quant à la forme, elle nous intéresse peu puisque dans notre propre système de classification les lésions sont sélectionnées manuellement par le spécialiste, les descripteurs de ce type ne seront pas détaillés ici.

Contents

2.1	Présentation générale	14
2.2	Caractérisation des niveaux de gris ou des couleurs	14
2.3	Caractérisation de la texture	16
2.4	Caractérisation de la forme	27
2.5	Conclusion	29

2.1 Présentation générale

Les descripteurs visuels sont des mesures exprimant le contenu d'une image, ses caractéristiques. Les concevoir est une tâche ambitieuse, car comme le dit l'adage, une image vaut mille mots. Le système de visualisation humain est complexe. Il est possible de s'en approcher, par exemple en essayant de caractériser la forme d'une structure, ou de s'en dissocier, en menant une approche purement statistique.

Ces mesures décrivent donc les couleurs ou niveaux de gris d'une image, ses textures ou ses formes, ce que nous allons développer dans chaque section ci-dessous. Elles sont calculées sur l'image entière, sur l'image divisée en petits blocs de pixels de taille égale, ou bien sur des régions d'intérêt (RI). Bien entendu, le choix d'un ensemble de descripteurs se fera en fonction de l'application, ils se doivent d'être à la fois :

pertinents : permettant de bien caractériser les images d'un domaine donné (ils seront sûrement différents pour des photographies de forêts et pour des bandes dessinées)

répondant à des contraintes spécifiques : un temps de calcul court pour un outil prévu pour fonctionner en temps réel, ainsi qu'une certaine robustesse par rapport au bruit ou à certaines transformations (translation, rotation, changement d'échelle).

Les descripteurs peuvent être utilisés pour atteindre différents objectifs :

- recherche par le contenu : recherche d'une image similaire à une image de requête dans une base de données d'images
- reconnaissance : déterminer si l'image sur laquelle sont calculés les descripteurs correspond à un modèle
- classification : déterminer la classe représentative la plus proche de l'image
- alignement et recalage : transformer une image de plusieurs façons afin de correspondre à une autre image, en partie ou en totalité
- approximation et simplification : construire une image sur moins d'éléments qui est toujours similaire à l'originale
- détection et/ou segmentation : détecter un objet particulier dans une image et/ou le segmenter

Les différentes valeurs sont classiquement concaténées afin de former un vecteur de grande dimension qui sera inséré dans un système de classification ou de recherche par le contenu.

2.2 Caractérisation des niveaux de gris ou des couleurs

Afin de décrire au mieux les couleurs ou niveaux de gris d'une image, l'histogramme est l'outil idéal : c'est l'estimation de leur distribution de probabilité.

Il peut être employé comme descripteur en tant que tel, mais il est courant d'en extraire des statistiques. Classiquement, les mesures suivantes sont calculées, dont les for-

mules sont exprimées pour n intervalles, dans le cas d'une image en niveaux de gris, et p_i le nombre de pixels présents dans l'intervalle i de l'histogramme.

- **moyenne** : intensité moyenne des pixels de l'image correspondant au moment d'ordre 1 :

$$\mu = \frac{1}{n} * \sum_{i=1}^n p_i \quad (2.1)$$

- **écart-type** : mesure de la dispersion de la variable autour de la moyenne, il correspond au moment d'ordre 2. Son carré, la variance, nous indique que la densité de probabilité est concentrée en un point unique et la variable aléatoire connue avec certitude lorsqu'elle est nulle :

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (p_i - \mu)^2 \quad (2.2)$$

- **coefficient de dissymétrie (skewness)** : mesure de l'asymétrie de la distribution par rapport à la moyenne. Un histogramme obtiendra une valeur nulle s'il est parfaitement symétrique, une valeur positive si la distribution est décalée à gauche de la moyenne (et donc une queue de distribution étalée vers la droite), et une valeur négative dans le cas d'une distribution décalée à droite de la moyenne. Cette mesure notée S correspond au moment d'ordre 3 de la variable centrée réduite sur le cube de l'écart-type :

$$S = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{p_i - \mu}{\sigma} \right)^3 \quad (2.3)$$

- **coefficient d'aplatissement (kurtosis, du grec courbe, arrondi, bossu)** : il s'agit d'une valeur que nous nommons K , représentant l'écrasement de la distribution. Il correspond au rapport entre le moment d'ordre 4 de la variable centrée réduite et le carré de la variance. Il mesure, hors effet de dispersion (exprimé par l'écart type), la disposition des masses de probabilité autour de leur centre, et permet de savoir si elles se regroupent de façon proche ou lointaine du centre de probabilité. Lorsqu'il est positif, la distribution est "pointue", dans le cas contraire elle est "écrasée" :

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \frac{(p_i - \mu)^4}{\sigma^2} - 3 \frac{(n-1)^2}{(n-2)(n-3)} \quad (2.4)$$

Deux remarques sur les formules précédentes sont à faire. Premièrement, p_i est souvent remplacé par $p_i * p(p_i)$ avec $p(p_i)$ la probabilité de la variable p à l'intervalle i en considérant l'histogramme comme une variable aléatoire discrète. Deuxièmement, les mesures d'ordre supérieur à deux présentent un grand intérêt pour la séparation de variables aléatoires indépendantes non gaussiennes.

Un exemple pratique est illustré Figure 2.1. L'histogramme présente deux ensembles : le plus important représente l'air (le fond, noir), le second les tissus mous (gris clair à gris foncé). Le blanc (tissus osseux) est très peu représenté. Intuitivement, les mesures extraites sur cet histogramme sont peu représentatives, puisque moyennées entre ces deux amas. Ce problème est atténué lorsque les images étudiées sont restreintes à des zones d'intérêt.

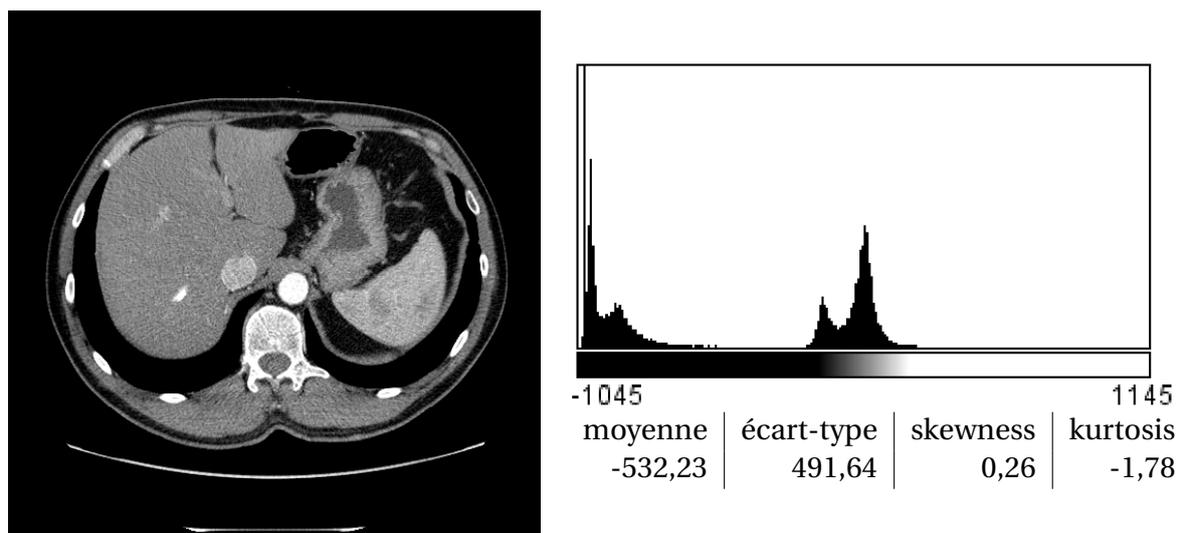


FIGURE 2.1 – Exemple d'un histogramme et des statistiques extraites de celui-ci sur une image DICOM scanner abdominale

L'histogramme et les statistiques extraites de celui-ci sont présents dans presque tous les ensembles d'attributs visuels décrits dans la littérature, comme dans [JP12]. Ils sont en effet très simples à implémenter, rapides à calculer, et permettent une première analyse des images, avant de réaliser des études plus fines de texture et/ou de forme.

2.3 Caractérisation de la texture

Nous allons décrire ici les descripteurs de texture que nous pouvons trouver dans des systèmes de classification ou de recherche par le contenu. Ils ont une grande importance dans le travail sur les images médicales. Bien entendu, l'ensemble qui suit n'est pas exhaustif, l'objectif étant de proposer au lecteur une explication sur les attributs qui seront mentionnés par la suite.

Tout d'abord nous allons définir le concept de texture, avant de passer en revue les descripteurs de texture basé sur les statistiques, puis ceux basés sur des modèles avant de terminer avec ceux issus du traitement du signal.

2.3.1 Définition de la texture

Malgré de nombreuses recherches dans ce domaine, le concept de texture n'est toujours pas clairement établi. Coggins a regroupé dans [Cog82] un certain nombre de définitions diverses.

Le terme texture provient du latin *textura* qui définit la disposition et l'ordre des fils sur une toile. Sa première définition est restée puisqu'on emploie toujours le terme de texture dans le secteur textile, afin de désigner l'opération de tissage et la surface du vêtement.

Au niveau général, la texture est une propriété captée par le sens du toucher (douceur, rugosité...). Dans le domaine de la peinture, la texture est définie en quatre types : texture réelle (combinaison entre l'aspect et le toucher de la peinture), texture simulée (par exemple créée dans le but de ressembler à autre chose que de la peinture), texture abstraite (qui ne représente pas l'objet mais son concept transcrit en termes de texture) et texture inventée. En musique, la texture est la forme selon laquelle les différentes voix (vocales ou instrumentales) s'associent. Des textures musicales courantes sont la monodie (une seule ligne mélodique) et l'homophonie (les voix mélodiques suivent le même rythme mais sur des notes différentes).

En informatique, plusieurs définitions se complètent. Une texture synthétique est constituée très simplement par la répétition d'un motif. Une texture naturelle est forcément plus complexe, on peut la définir comme l'arrangement spatial des couleurs ou intensités dans une image ou une région de l'image. C'est en quelque sorte une illusion, représentant la texture physique réelle qui elle se définit également en termes de toucher. Enfin, une texture peut nommer une image servant à recouvrir la surface d'un objet virtuel, notamment un objet en 3 dimensions.

2.3.2 Descripteurs fondés sur les statistiques

Ces mesures, fondées sur les théories statistiques, restent cependant aisées à appréhender. Nous allons décrire le principe des matrices de relation d'intensité, ainsi que de la mesure d'auto-corrélation.

Matrices de relations d'intensité

Le principe est de construire une matrice exprimant une certaine relation entre les intensités des pixels de l'image. Des mesures statistiques sont ensuite calculées sur ces matrices plutôt que sur l'image d'origine. Celle-ci est en effet volumineuse et une haute corrélation est présente entre ses valeurs.

A chaque fois, une relation de voisinage est définie à l'aide d'un couple (angle, distance), ce qui induit une sensibilité à la rotation. Afin de pallier à cet écueil, elle est classiquement calculée sur un ensemble d'angles afin d'obtenir une invariance à la rotation : di-

rection horizontale (0°), verticale (90°), ainsi que selon les axes des deux diagonales (45° et 135°). Il est possible de ne considérer que la moyenne de ces quatre matrices.

La plus connue de ces matrices est la matrice dite de co-occurrence, mais il en existe d'autres. n est le nombre de niveaux de gris de l'image dans ce qui suit.

- **Matrice de co-occurrence** Elle représente la distribution de valeurs de pixels de l'image co-occurentes selon le couple (angle, distance). Sur une image en niveaux de gris, elle est appelée Grey-Level Co-occurrence Matrix (GLCM), ou Space Grey Level Dependence Matrix (SGLDM). Des expériences sur la perception des textures par les êtres humains indiquent que les probabilités de second ordre de ce qui est décrit par la matrice de co-occurrence jouent un rôle important dans la discrimination des textures [HLH11].

La matrice M , de taille $n * n$, est remplie avec le nombre de pixels correspondant à chaque information de "voisinage". Les lignes sont les valeurs des pixels de référence, et les colonnes celles des pixels "voisins". La valeur $M(i, j)$ de la matrice est donc le nombre de pixels d'intensité i ayant comme "voisin" un pixel d'intensité j . La matrice de co-occurrence est dans la plupart des cas rendue symétrique. La GLCM est concentrée autour de la diagonale, la plupart des pixels ayant une valeur proche de ses voisins.

La matrice de co-occurrence est souvent calculée sur des fenêtres, et non sur l'image entière. Elle est en règle générale très volumineuse, parfois même plus que l'image d'origine, et dispersée. C'est pourquoi elle n'est guère employée comme mesure elle-même. Des statistiques en sont extraites, les 14 valeurs communément calculées sont appelées les descripteurs d'Haralick, du nom de leur concepteur [Har79]. Communément, seul un groupe de 9 descripteurs est extrait, car la corrélation au sein du groupe de 14 descripteurs est importante.

Un exemple de la matrice de co-occurrence d'une image scanner abdominale transformée en 256 niveaux de gris, et de quelques statistiques qui en sont extraites, est présenté dans la Figure 2.2.

Dans les formules qui suivent, la valeur $P_{i,j}$ est la probabilité qui se trouve dans la case (i, j) de la matrice de co-occurrence, c'est donc la probabilité qu'une combinaison (pixel de valeur i , voisin de valeur j) soit présente dans l'image).

- Statistiques descriptives sur la GLCM
 - Moyenne :

$$\mu = \sum_{i,j=0}^{N-1} i * P_{i,j} \quad (2.5)$$

dans le cas d'une matrice de co-occurrence rendue symétrique, les moyennes obtenues sur le i et le j sont identiques

- Variance :

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{i,j} (i - \mu)^2 \quad (2.6)$$

c'est une mesure de la dispersion des valeurs autour de la moyenne. On peut remplacer la variance par l'écart-type, qui est la racine carrée de la variance élevée au carré.

- Corrélation :

$$\sum_{i,j=0}^{N-1} P_{i,j} \frac{(i - \mu)^2}{\sigma^2} \quad (2.7)$$

il s'agit d'une mesure des dépendances linéaires des niveaux de gris entre les pixels voisins

- Mesures de contraste : attribuent des poids en fonction de la distance à la diagonale de la GLCM

- Contraste (ou sum of squares variance) :

$$\sum_{i,j=0}^{N-1} P_{i,j} (i - j)^2 \quad (2.8)$$

lorsque i et j sont égaux (on est donc sur une cellule de la diagonale de la GLCM), les pixels sont identiques à leur voisin et la valeur obtenue est 0. Si i et j diffèrent de 1, le contraste est faible et le poids attribué est 1. Le poids augmente de façon exponentielle avec la différence $i - j$ (et donc en s'éloignant de la diagonale de la GLCM).

- Dissimilarité :

$$\sum_{i,j=0}^{N-1} P_{i,j} |i - j| \quad (2.9)$$

les poids augmentent de façon linéaire lorsque l'on s'éloigne de la diagonale de la GLCM

- Homogénéité (ou moment inverse de la différence) :

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (2.10)$$

les poids diminuent de façon exponentielle lorsque l'on s'éloigne de la diagonale de la GLCM

- Mesures d'ordre : deux images peuvent avoir des mesures de contraste semblables, mais avec un ordonnancement différent des valeurs de pixels, c'est pourquoi les mesures d'ordre évaluent la régularité des valeurs de pixels dans la fenêtre. Elles attribuent des poids en fonction du nombre d'occurrences d'une paire.

- Second moment angulaire :

$$\sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (2.11)$$

la racine carrée de cette mesure est appelée énergie ou uniformité. Les probabilités elles-mêmes sont utilisées en tant que poids, ce qui est logique puisqu'elles sont déjà une mesure d'apparition d'une occurrence. La valeur du second moment angulaire est élevée lorsque la fenêtre est très régulière.

- Probabilité Maximale :

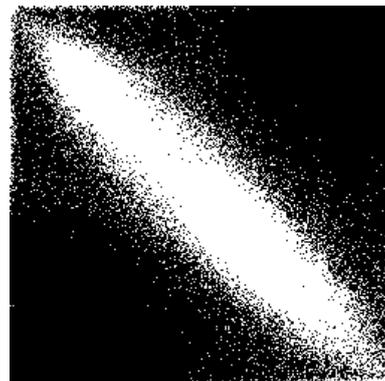
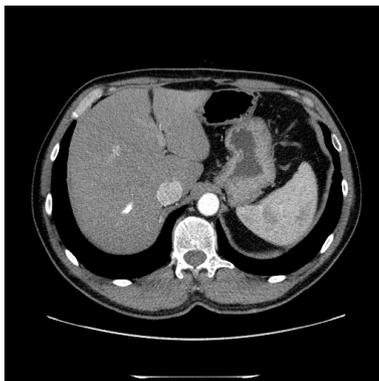
$$\max(P_{i,j}) \quad (2.12)$$

cette mesure est tout simplement la probabilité maximale trouvée dans la fenêtre. Une valeur élevée indique qu'une paire de pixels domine les autres au sein de la fenêtre.

- Entropie :

$$\sum_{i,j=0}^{N-1} P_{i,j} (-\log_2 P_{i,j}) \quad (2.13)$$

ce concept est apparu en thermodynamique, l'entropie réfère à la quantité de chaos. Lorsque les valeurs considérées sont équiprobables, l'entropie est égale à 0,5. De façon pratique, lorsqu'une des probabilités utilisées est nulle (ce qui reviendrait à calculer un logarithme népérien de 0, soit une opération impossible), on ignore l'étape correspondante de calcul dans la somme en passant directement à la suivante.



moyenne	corrélacion	contraste	homogénéité	énergie	entropie
3.99	0.99	172.30	0.61	0.27	0.88

FIGURE 2.2 – Exemple d'une matrice de co-occurrence et des statistiques extraites de celle-ci sur une image DICOM scanner abdominale

– Run-Length Matrix

Ici, ce sont les répétitions des intensités des pixels dans l'image qui sont étudiées. Une séquence (run) est définie comme un set de pixels consécutifs, colinéaires ayant la même intensité. La longueur de la séquence est tout simplement le nombre de pixels qui y est contenu.

La matrice $M(i,j)$ est définie en spécifiant la direction puis en comptant l'occurrence des séquences pour chaque niveau de gris et longueur dans cette direction. La matrice est de dimension $i * j$, avec i le nombre de niveaux de gris de l'image, et j la longueur maximale d'une séquence.

Galloway a proposé l'utilisation d'une matrice de ce type dans le but d'extraire 7 descripteurs de texture [Gal75]. Les statistiques calculées sur cette matrice expriment des caractéristiques visuelles de la texture. Ces descripteurs sont des mesures des séquences courtes/longues, des séquences courtes/longues d'intensité faible/forte, des mesures de non-uniformité en termes d'intensité et de longueur de séquence, et le ratio entre le nombre de séquences et le nombre de pixels de l'image. D'autres ont ensuite proposé des descripteurs plus performants. Dans tous les cas, ils sont tous issus d'un raisonnement intuitif sur les propriétés de la distribution run-length, sans réel fondements théoriques.

Ces méthodes aboutissent à des résultats plutôt satisfaisants, et ce dans différents domaines car elles sont largement répandues, notamment la matrice de co-occurrence. Cependant, leur plus gros inconvénient est leur lourdeur, tant en termes de temps de calcul que de place en mémoire. En effet, une matrice de co-occurrence calculée sur une petite image de taille 16 x 16 pixels qui contiendrait 256 niveaux de gris a une taille de 256 x 256, dans le cas d'une image médicale de 16 bits la matrice est donc énorme, même pour une lésion.

Mesures d'auto-corrélation ou auto-covariance

La fonction d'auto-corrélation permet de valider la régularité d'une texture répétitive, ainsi que sa granularité. Elle correspond à l'auto-covariance divisée par la variance de l'image et est reliée au spectre de la transformation de Fourier, que nous verrons plus loin. Si le grain d'une texture est épais, sa fonction d'auto-corrélation descendra doucement, si le grain est fin, elle baissera rapidement, enfin si la texture est régulière, cette fonction contiendra des extrema locaux. [GMM04] présente un exemple d'application dans le domaine médical. Formellement, cette fonction AC est définie comme ci-dessous sur l'image en deux dimensions $I(x,y)$:

$$AC(x,y) = \frac{\sum_{u=0}^N \sum_{v=0}^N I(u,v) * I(u+x,v+y)}{\sum_{u=0}^N \sum_{v=0}^N I^2(u,v)} \quad (2.14)$$

2.3.3 Descripteurs fondés sur des modèles

Le principe est de tenter de modéliser l'image, généralement par une approche statistique. Nous allons présenter deux méthodes permettant d'extraire des descripteurs selon cette voie.

Champs de Markov

Les champs de Markov (en anglais, Markov Random Fields ou MRF) modélisent par une approche statistique le phénomène de dépendance au sein des pixels. Cette méthode a été proposée dans les années 1920 sous le nom de modèle d'Ising en mécanique ([Isi24]) et a été introduite dans le traitement d'images dans les années 1970 [HS78]. La texture est ici considérée comme l'échantillon d'un processus stochastique défini par un ensemble de paramètres.

Cette approche consiste à considérer les valeurs des pixels comme les réalisations d'une variable aléatoire, et le champ lui-même comme sa fonction de densité de probabilité. L'idée principale est que des pixels voisins ont habituellement une intensité proche dans une image, alors que les valeurs des pixels sont indépendantes des pixels qui sont au-delà de leur voisinage proche. L'image est donc vue comme un échantillon d'un processus aléatoire, dont la corrélation entre les pixels est proportionnel à leur séparation géométrique.

La structure du voisinage au sein des pixels de l'image qui est illustrée dans la figure 2.3 définit l'ordre du champ de Markov. Par exemple, un champ de Markov du premier ordre considère les 4 plus proches voisins d'un pixel.

5	4	3	4	5
4	2	1	2	4
3	1	p	1	3
4	2	1	2	4
5	4	3	4	5

FIGURE 2.3 – *Voisinage considéré dans un champ de Markov (les chiffres dans les cases indiquent l'ordre du voisinage par rapport au pixel central)*

Dans le but d'appliquer les champs de Markov sur une image, la fonction de probabilité jointe des pixels doit être calculée. Elle définit la probabilité des événements définis en termes de l'axe à la fois des abscisses et de celui des ordonnées. Le souci qui se pose est que le calcul de cette fonction nécessite une mémoire et un temps de calcul importants. Pour donner un ordre d'idée, sur une image en niveaux de gris de 8 bits qui mesure 256

x 256 pixels, 256^n configurations devront être estimées. Deux alternatives communes à ce calcul sont les champs de Gibbs, ainsi que les champs de Markov gaussiens.

Dans le premier cas, le théorème d'Hammersley-Clifford établit l'équivalence entre un champ de Markov et une distribution de Gibbs [HC71].

Dans le second cas, le champ, au lieu d'être la vraie fonction de probabilité calculée sur les pixels de l'image, est une Gaussienne. Les paramètres du champ Gaussien peuvent être calculés à l'aide d'une estimation des moindres carrés afin de se rapprocher des données au plus près. Les mesures servant de descripteurs visuels seront les paramètres de champ, ainsi que sa moyenne et sa variance.

Fractales

Les analyses par fractales ou multi-fractales sont de plus en plus courantes ces dernières années. Le terme vient d'un article de Mandelbrot des années 1970. Une fractale est un objet restant identique à toutes les échelles. Pour une application traitant les textures, il faut donc admettre l'hypothèse qu'une texture est une fractale pour un certain niveau d'observation. Le principe est d'analyser les changements dans les propriétés d'une image lorsque la résolution varie. Une propriété importante des fractales est l'aire de leur surface. Les changements observés dans l'aire de la surface des niveaux de gris d'une image à différentes résolutions forment une signature fractale de la texture. Pour comparer deux textures, il suffit d'analyser cette signature.

2.3.4 Descripteurs fondés sur une représentation fréquentielle

Le travail avec des transformations et des filtres consiste en l'application de fonctions mathématiques sur l'image. C'est sur cette représentation résultat que sont ensuite extraits les descripteurs visuels de l'image.

Transformation de Fourier

La transformée de Fourier permet de passer d'une représentation spatiale d'une image à sa représentation dans le domaine fréquentiel.

La transformation de Fourier de base est à une dimension, mais elle se généralise en deux dimensions pour travailler sur des contenus visuels. Les images sont considérées comme des fonctions à deux variables représentant l'intensité en fonction de l'abscisse et de l'ordonnée. Pour toute fonction continue f , avec x, y les coordonnées spatiales et ν_x, ν_y les coordonnées spectrales du signal, nous avons :

– Transformation 1D :

$$F(\nu) = \int_{-\infty}^{+\infty} f(x) e^{-i2\pi x \nu} dx \quad (2.15)$$

– Transformation 2D :

$$F(\nu_x, \nu_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-i2\pi(\nu_x x + \nu_y y)} d\nu_x d\nu_y \quad (2.16)$$

Le résultat de la transformée de Fourier est un ensemble de nombres complexes. Il est donc possible de travailler au choix sur la représentation de la phase ou de l'amplitude.

En général, c'est le module de la transformée, appelé spectre de Fourier, qui est étudié. En effet, il permet de rendre compte de la distribution énergétique de l'image : la périodicité et l'orientation des motifs de l'image sont respectés, ce qui a un avantage évident lorsque l'on cherche à définir une texture.

Dans l'espace transformé, les basses fréquences représentent les grandes surfaces homogènes ou floues, tandis que les hautes fréquences représentent les contours ou le bruit.

Un exemple de transformée de Fourier est présenté dans la Figure 2.4.

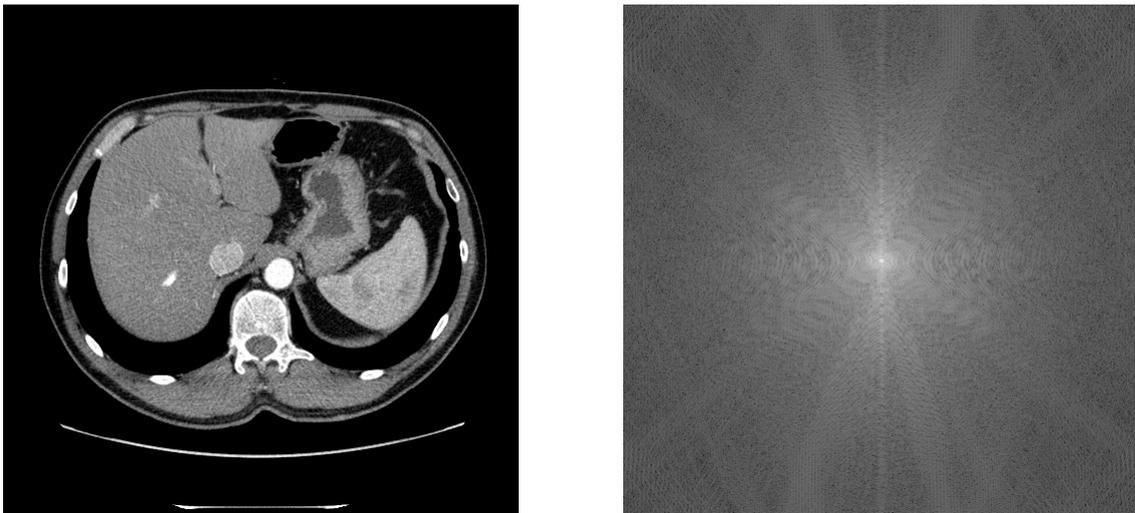


FIGURE 2.4 – La transformée de Fourier (à droite) d'une image médicale d'un scanner abdominal (à gauche)

Filtres

De nombreux filtres peuvent être utilisés. Law a proposé en 1980 [Law80] des mesures "d'énergie de texture", qui ont depuis été utilisées dans des applications diverses et variées. Les descripteurs de Law sont extraits après un ensemble de filtrages en 3 étapes. Tout d'abord, 25 noyaux de convolution sont appliqués sur l'image. Au départ, nous avons 5 noyaux à 1 dimension, chacun d'eux étant capable de détecter une caractéristique précise de l'image, comme une vague. Ces noyaux sont convolués avec chacun d'entre eux

afin de former 25 noyaux à 2 dimensions cette fois. Ce sont ces noyaux 2D qui seront appliqués à l'image d'origine, aboutissant à 25 images filtrées. Dans un second temps, une mesure d'énergie de la texture est calculée sur chaque pixel de ces images convoluées à l'aide d'une opération de fenêtrage. Chacun des pixels des 25 images est remplacé par la valeur de cette mesure d'énergie calculée sur le voisinage local du pixel. Ces nouvelles images sont appelées Texture Energy Measure (TEM) images. Tous les noyaux de convolution ont une moyenne nulle, à l'exception d'un seul, dont l'image résultat permet de normaliser les images TEM. Enfin, les descripteurs sont combinés en additionnant les TEM images qui avaient été générées par des noyaux de convolution transposés. On obtient 14 descripteurs de texture pour chaque pixel. La Figure 2.5 illustre la filtration effectuée par cette méthode.

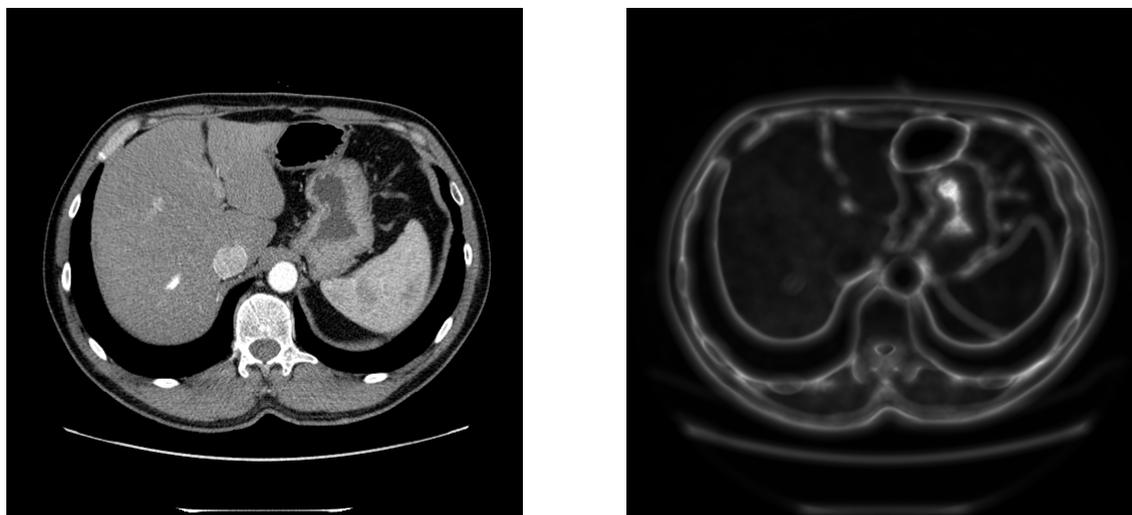


FIGURE 2.5 – L'une des images TEM obtenues à l'aide de la méthode de Law (à droite) d'une image médicale d'un scanner abdominal (à gauche)

Ondelettes

Les ondelettes sont des oscillations qui "résonnent" (l'amplitude, tout d'abord nulle, augmente puis finit par redevenir nulle) en réaction à un signal, qui peut être de toute sorte, notamment audio ou visuel. Classiquement, on les convolue avec ce signal afin d'extraire des informations exploitables. Elles sont employées de façon groupée, chaque ondelette étant à même de détecter quelque chose de particulier dans le signal. En général, toutes les ondelettes du groupe dérivent de la même ondelette "mère", dont l'échelle et l'orientation sont modifiées.

Les ondelettes sont fréquemment utilisées, et ce, dans des domaines d'application très divers, notamment sur des images médicales [QLC⁺12]. En tant que descripteurs visuels,

différentes sortes d'ondelettes peuvent être utilisées : classiques, de Gabor, curvelets ou ridgelets (des extensions du concept de base).

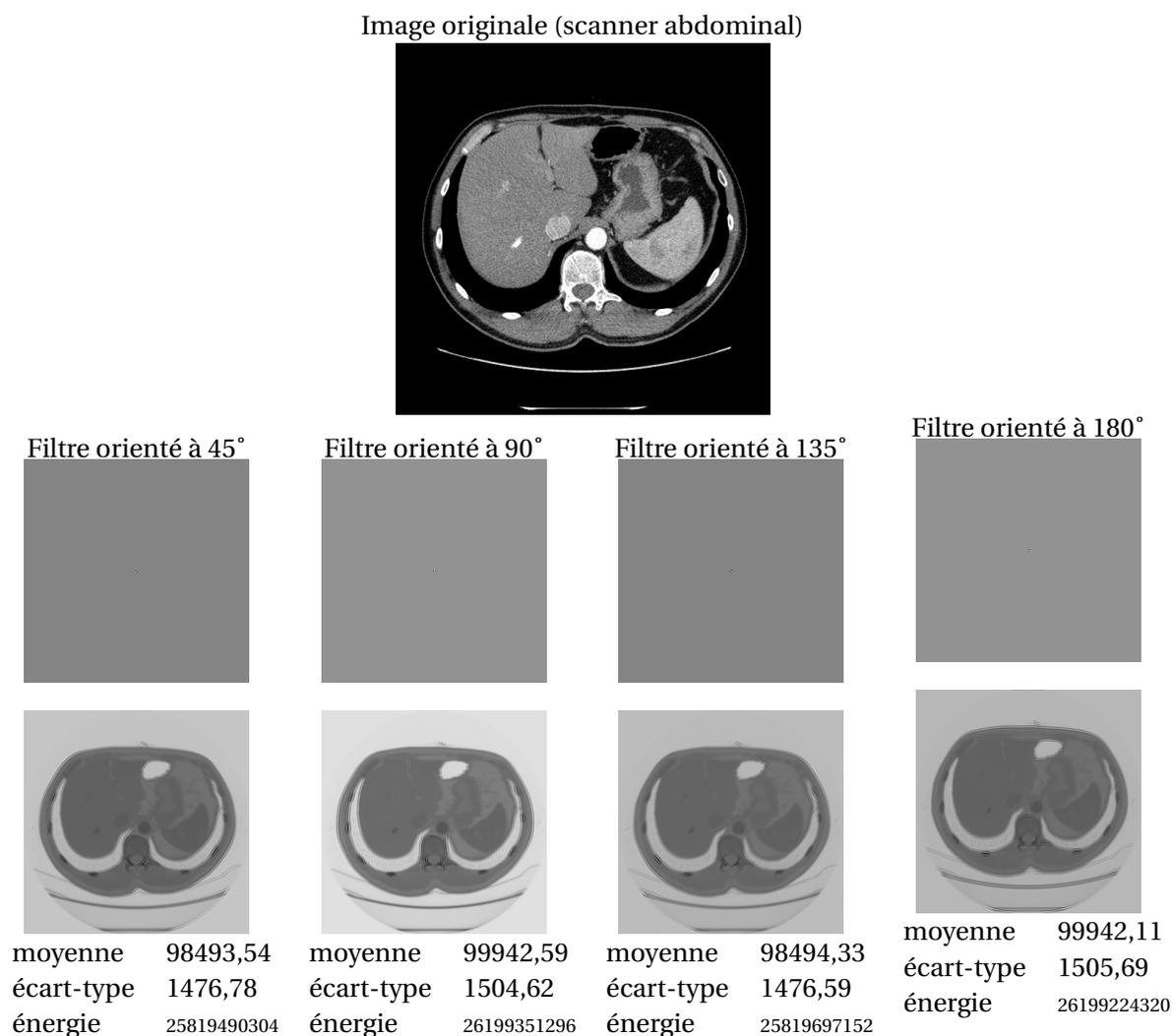


FIGURE 2.6 – Ondelettes de Gabor (facteur d'échelle de 2 pour 4 orientations) : pour chaque orientation sont présentés le filtre, l'image filtrée puis les descripteurs statistiques extraits

Les ondelettes de Gabor consistent en un groupe de filtres de Gabor de différentes fréquences et directions. Un filtre de Gabor est une fonction Gaussienne modulée par une sinusoïde complexe dans le domaine spatial. Dans le domaine fréquentiel, c'est tout simplement une gaussienne. On peut extraire des valeurs statistiques de moyenne, écart-type et énergie qui serviront de descripteurs. La Figure 2.6 présente un exemple d'ondelettes de Gabor.

2.3.5 Problématiques liées à la texture

L'emploi de descripteurs de texture pose un certain nombre de questions, notamment quant à leur invariance à des transformations diverses. Un grand nombre de publications témoigne de l'intérêt porté à sujet, à la fois théorique [HRDA] et relatif à l'application sur des images médicales. L'article [GLZ⁺13] par exemple s'intéresse aux contraintes sur les données (notamment sur la base d'apprentissage) permettant de rendre possible l'invariance à la rotation de descripteurs de texture. Des techniques ont été proposées pour rendre invariants la plupart des descripteurs décrits précédemment appliqués sur des images médicales : invariance au changement d'échelle des descripteurs d'ondelettes ou de fractales [HUVW13], invariance à la rotation de mesures issues de la transformation de Fourier [MKY13]...

2.3.6 Synthèse

Des approches radicalement différentes permettent donc d'appréhender la texture d'une image. Le choix de l'une ou de l'autre dépendra de l'application et surtout des contraintes, notamment en termes de temps de traitement, car certaines de ces méthodes nécessitent des calculs complexes. La mise en oeuvre pratique de ces propositions n'est également pas toujours aisée, et si un groupe de descripteurs est employé, mieux vaut regarder les corrélations éventuelles entre les différentes mesures. Dans l'idéal, il faudra également s'intéresser à leur invariance. Ce sont cependant des outils puissants.

2.4 Caractérisation de la forme

La forme est nettement moins employée dans ce domaine applicatif que la texture, ce qui est intuitif lorsque l'on regarde des exemples visuels de lésions hépatiques, c'est pourquoi cette partie sera moins fournie et détaillée que les précédentes.

Il existe tout d'abord des mesures géométriques simples, par exemple de dimensions (rayon, axes, longueur des côtés selon la forme de la zone d'intérêt), ou d'aire.

Dans le cas d'ellipses, l'excentricité et la compacité peuvent également être calculées. L'excentricité est une mesure d'aplatissement de l'ellipse variant entre 0 lorsque l'ellipse est un cercle, et 1 lorsque l'ellipse est une parabole. La compacité est la comparaison à un cercle ayant le même périmètre que l'ellipse, dans le cas d'un cercle la valeur de la compacité est donc 1.

L'approche statistique des moments, permettant de caractériser les contours d'un objet, peut être déclinée de nombreuses façons. Il s'agit de la décomposition de la fonction représentant les contours de l'objet étudié en une suite infinie de moments (l'équivalence de ces deux représentations a été énoncée par Hu).

Le nombre de coefficients conservé est volontairement limité afin de ne travailler qu'avec les mesures les plus significatives. Les moments sont calculés à plusieurs ordres

différents (moments d'ordre 1, 2, 3...). Selon l'objectif à atteindre, différentes bases sont possibles, elles influent sur la robustesse des descripteurs à la translation et la rotation : moments géométriques, complexes, ou orthogonaux (dits de Legendre [ZMHP11] et de Zernike [Zer34]).

Type	Approche	Descripteurs	Paramètres
Niveaux de gris	Histogramme	Mesures statistiques	Nombre d'intervalles
Texture	Statistiques	Statistiques sur GLCM ou Run-length)	Voisinage (orientation, distance)
		Auto-corrélation / Auto-covariance	-
	Modèles	Champs de Markov (Informations du modèle et statistiques)	Voisinage
		Fractales	-
	Traitement du signal	Fourier	- représentation (phase ou amplitude)
		Law (valeurs calculées après filtrage)	-
Ondelettes		fonction mère, échelle, orientation	
Forme	Mesures géométriques	Aire, compacité, excentricité	-
	Moments	Mesures effectuées sur les moments géométriques, complexes, orthogonaux etc...	Fonction et ordre

TABLE 2.1 – Synthèse des descripteurs visuels couramment extraits d'images médicales

2.5 Conclusion

Le Tableau 2.1 présente les différents descripteurs évoqués dans ce chapitre. Des approches très variées de descripteurs sont envisageables, et la littérature sur le sujet est prodigieuse. Bien entendu, c'est l'application visée qui détermine en premier lieu les caractéristiques qui devront être employées. Classiquement, dans le cas d'images médicales du foie, l'ensemble de descripteurs de base est composé des statistiques sur l'histogramme ainsi que des descripteurs d'Haralick sur la matrice de co-occurrence.

Les niveaux de gris sont des informations utiles, mais seuls ils n'arrivent pas à introduire suffisamment d'information pour réaliser une différenciation. La forme des lésions est extrêmement variable et un vrai problème survient pour distinguer la lésion du fond lorsqu'elles ne sont pas homogènes.

Les descripteurs extraits des images de lésions hépatiques représentent généralement des informations sur la texture, nous verrons dans les contributions comment ils seront utilisés et combinés. Cependant, ils vont généralement de pair avec une complexité importante de calcul, ce qui peut être corrigé par une réflexion sur leur construction.

Nous verrons comment ces descripteurs sont employés dans les outils sur l'imagerie hépatique dans le chapitre 3.

Classification supervisée

Préambule

Dans ce chapitre sur la classification, nous allons tout d'abord introduire les notions essentielles du domaine avant d'évoquer l'apprentissage supervisé qui nous intéresse tout particulièrement ici, et de se pencher pour terminer sur l'évaluation des performances.

Contents

3.1	Présentation générale	32
3.2	Principaux algorithmes de classification supervisée	33
3.3	Evaluation des performances	39
3.4	Conclusion	42

3.1 Présentation générale

La différenciation des lésions à partir de descripteurs visuels est la seconde étape de l'outil que nous souhaitons concevoir. Nous l'avons évoqué dans l'introduction, le processus conduisant à un diagnostic est complexe et nécessite un apprentissage.

L'objectif d'une classification est d'identifier les classes auxquelles appartiennent des objets à partir d'éléments descriptifs (attributs, caractéristiques...). La classification fait partie du problème plus général de reconnaissance de motif ("pattern recognition"), qui assigne une valeur de sortie à une valeur d'entrée donnée. Elle implique également la notion d'apprentissage. Les deux sont issus de l'intelligence artificielle.

Les domaines d'application de la classification, de la reconnaissance de motif et de l'intelligence artificielle sont bien évidemment très variés. On peut dire que l'ensemble des domaines scientifiques fait désormais appel à ces techniques d'une façon ou d'une autre.

L'intelligence artificielle est la capacité d'une machine, d'un logiciel, à raisonner et à prendre des décisions, en fonction de son environnement mais aussi de ses décisions passées en vue de maximiser ses chances de succès à répondre à un problème donné. C'est donc un ensemble de disciplines qui cherchent à modéliser techniquement certaines manifestations de l'intelligence naturelle. C'est un sujet qui passionne l'être humain de par les nombreux questionnements philosophiques qu'il induit : y a-t-il des limites à l'intelligence artificielle et une machine peut-elle avoir une conscience ? et bien d'autres.

Il n'est donc pas étonnant d'en trouver la trace dès les mythes de la Grèce Antique. Au cours des siècles des tentatives d'automates diverses et variées furent réalisées. Plus récemment, dès le 19ème siècle ce sujet apparaît régulièrement dans la littérature. A partir des années 1950, c'est un sujet fédérateur de projets. Mais c'est depuis les années 1990 que la recherche dans ce domaine explose littéralement, et s'introduit partout : dans le data mining, la logistique, le diagnostic médical, et à peu près tous les domaines scientifiques et industriels.

Les travaux de Fisher sur la classification binaire (deux classes) ouvrirent la voie dans les années 1930 [Fis36]. Ils lui ont permis de définir ce qu'on appelle le coefficient linéaire de Fisher, qui représente la séparation entre deux distributions comme étant le ratio de la variance inter-classes sur la variance intra-classe.

Il existe deux grandes familles de classificateurs : supervisés et non supervisés. Dans le cas d'une classification supervisée, les classes sont connues à l'avance, et des échantillons de ces classes sont disponibles pour effectuer un apprentissage. Dans le cas d'une classification non supervisée, l'objectif est d'obtenir des séparations dans les données définissant des classes, dont on ne connaît à l'origine ni le nombre ni les caractéristiques. On se concentre ici sur la classification supervisée, adaptée au problème qui nous intéresse : notre système doit "apprendre" à reconnaître les différentes classes de lésions à partir de nodules identifiés qui lui seront présentés.

La classification se fait sur les attributs des observations. Les observations sont les objets à classer. Les attributs sont les valeurs, les caractéristiques extraites des observations

en amont de la classification. Ils peuvent être logiques, symboliques ou numériques. Bien entendu, si le nombre d'attributs et/ou d'observations est trop faible, la classification ne pourra pas être efficace. Dans notre cas, les attributs correspondent aux mesures numériques des descripteurs visuels extraits de nos images de lésion du foie.

La classification met en jeu deux phases. La première est celle de l'apprentissage : l'algorithme va tenter de séparer les données en leur assignant une classe en fonction des valeurs des attributs. La base de données d'apprentissage doit être suffisamment étoffée pour permettre au classificateur de distinguer les classes. La deuxième phase est celle de la validation ou du test. Il s'agit d'entrer les valeurs des attributs d'une nouvelle donnée, l'algorithme retournant une étiquette de classe.

Les classificateurs peuvent se différencier en faibles et forts. Une méthode faible sera capable de différencier les classes mieux que le hasard, tandis qu'une méthode forte sera fortement corrélée avec la "vraie" classification.

L'objectif à ne pas perdre de vue est d'obtenir un subtil équilibre entre un apprentissage qui serait incomplet sur les données, et un sur-apprentissage. Ce dernier se produit lorsque la classification s'est trop adaptée aux données présentes en perdant de vue un objectif plus global. Ce problème se traduit dans les faits par un résultat excellent sur les données d'apprentissage, mais faible sur les données de test. La classification est souvent dépendante des paramètres des algorithmes, pour lesquels il faut fixer une valeur ou une fonction.

3.2 Principaux algorithmes de classification supervisée

Les méthodes de classification existantes sont bien entendu très nombreuses et variées. Seules des techniques utilisées dans le cadre de la caractérisation des lésions hépatiques sont présentées dans cette partie.

3.2.1 Arbres de décision

Le principe en est extrêmement simple. L'apprentissage construit un arbre dont chaque noeud est une règle sur les données, effectuant un partitionnement récursif sur ces dernières. L'idéal serait de stopper la croissance de l'arbre au bon moment, mais c'est un problème difficile. Dans l'état actuel des choses, il est commun de plutôt ajouter une seconde étape d'élagage. Les branches peu représentatives étant supprimées selon un critère déterminé. Lors de la classification, l'algorithme parcourt l'arbre jusqu'à la feuille correspondant aux valeurs données par chaque noeud.

Les arbres de décision sont aisément compréhensibles pour nous humains, ce qui est sans doute en partie à l'origine de leur succès. Cette propriété est illustrée par l'arbre de décision donné en exemple dans la Figure 3.1. De plus, ils ne requièrent pas d'étape de normalisation des données et sont capables de fonctionner lorsque certaines valeurs sont

manquantes. Les algorithmes nommés ID3 [Qui86] et C4.5 [Qui93], tous deux développés par Quilan sont les pionniers de cette famille. Une implémentation libre du second est disponible : J48 [KH02].

Les arbres sont reliés par la théorie aux règles discriminantes et aux tables de décision. En effet, un système de règles acycliques définies sur la région de l'espace des attributs est équivalent à une table de décision définie sur cette même région. De plus, un arbre de classification sur cette région est équivalent à une table de décision non ambiguë sur cette région.

Les tables de décision tentent d'un point de vue logique de déterminer un ensemble de conditions sur les entrées déterminant le comportement en sortie. On définit exhaustivement l'étiquette de sortie pour chaque combinaison des valeurs d'entrées. Les conditions les plus discriminantes sont posées en premier. Au sein d'une combinaison, chaque condition peut être vraie, fausse ou ne pas influencer sur le résultat final.

Un état de l'art détaillé peut être consulté pour plus de détails [NY10].

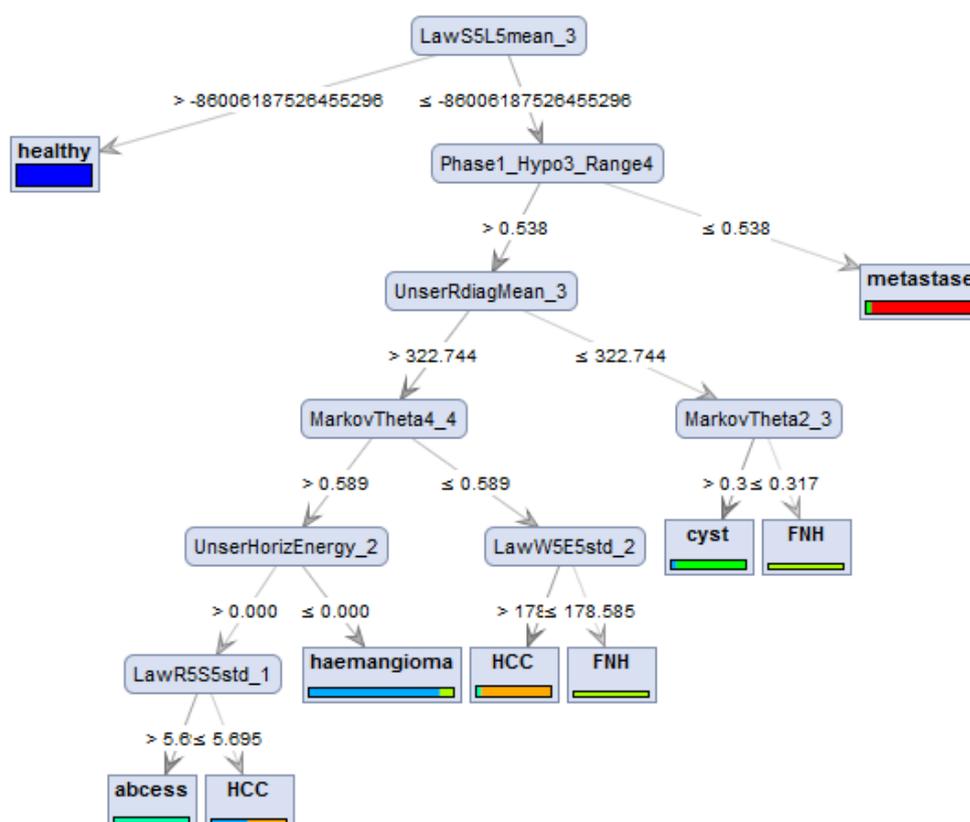


FIGURE 3.1 – Illustration du principe de l'arbre de décision Vue réalisée à partir de RapidMiner Community Edition

3.2.2 Réseaux de neurones

Ces méthodes s'inspirent du fonctionnement du réseau neuronal du cerveau. Ce sont des algorithmes puissants, capables de résoudre des problèmes non linéaires. Une fonction de décision assigne à chaque neurone une valeur de sortie en fonction de sa valeur d'entrée. Les neurones sont disposés en couches et reliés entre eux : un neurone est connecté en entrée à tous les neurones de la couche précédente, et en sortie à tous ceux de la couche suivante. Ce principe a été proposé fin 1800 ([Bai73, Jam90]).

La phase d'apprentissage consiste à optimiser les poids reliant les neurones, en minimisant une erreur de classification en fonction d'un paramètre à régler, le taux d'apprentissage. Ce dernier doit être un compromis entre l'erreur sur l'apprentissage et le sur-apprentissage.

Les réseaux de neurones sont robustes à des attributs bruités car l'ajustement des poids va leur assigner un poids faible à la fin de la période d'apprentissage. Cependant, ils sont relativement lents.

Le lecteur pourra se référer à cet état de l'art généraliste [EPRH02] des applications des réseaux de neurones dans le traitement des images pour plus de détails.

3.2.3 Support Vector Machine

L'approche des machines à vecteurs de support (Support Vector Machines ou SVM en anglais) consiste à se placer dans un espace de plus haute dimension (qui peut même être infini) que celui d'origine contenant les attributs, en espérant y découvrir une séparation linéaire dans les données qui n'apparaissait pas dans l'espace originel. Une illustration de cette technique est présentée Figure 3.2. Le but est bien sûr d'établir une marge maximale autour de la séparation des données.

Cette marge est obtenue à l'aide d'une fonction noyau. Au lieu d'effectuer un produit scalaire dans l'espace de grande dimension, elle permet de ne réaliser qu'une simple évaluation d'une fonction en un point dans l'espace d'origine. Cette dernière est bien moins coûteuse, et la transformation à appliquer pour le changement d'espace n'a pas besoin d'être connue. La fonction noyau peut être par exemple linéaire, polynomiale ou gaussienne.

Ces algorithmes ont été développés dans les années 1990 à partir de la théorie statistique de l'apprentissage Vapnik-Chervonenkis [Vap79, VC64]. L'intérêt de ces techniques est leur capacité à classer des données qui ne sont pas linéairement séparables.

Des données sont linéairement séparables s'il est possible de tracer sur le graphe une séparation entre les deux classes, consistant en une ligne lorsque l'espace contient deux dimensions, et en un hyperplan lorsque la dimension est supérieure. Dans le cadre d'une classification binaire, si les données ne respectent pas cette contrainte, le SVM est adaptable à l'aide d'une "marge douce". Cette approche permet de traiter le fait que certains

points se retrouvent du mauvais côté de la frontière de la marge tout en appartenant à l'autre classe.

A l'origine, le fonctionnement du SVM est prévu pour des classifications binaires exclusivement, cependant des adaptations ont été effectuées afin de discriminer des données multi-classes. Trois stratégies coexistent : le multi-classe pur et dur bien sûr, mais également le un contre tous (en anglais, One Against All) et le un contre un (en anglais One Against One ou pairwise). Dans les deux derniers cas, le SVM ne tente pas directement de distinguer toutes les classes, il réalise plusieurs discriminations avant de combiner les résultats obtenus en un résultat multi-classe final.

De nombreuses implémentations existent à l'heure actuelle, comme par exemple libSVM¹, qui est librement distribué, ou Sequential Minimal Organization (SMO [Pla98a]).

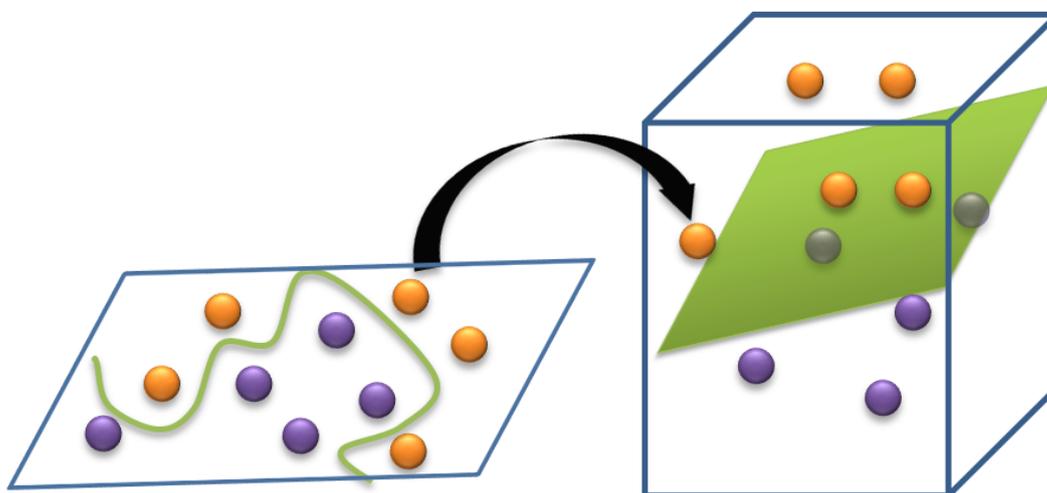


FIGURE 3.2 – Illustration du principe du SVM

3.2.4 Sac de mots ("bag of words")

Cette technique de "bag of words" (BoW) était à l'origine employée dans l'étude du langage d'un document. Il s'agissait de construire un dictionnaire de mots, sans se préoccuper de la grammaire ou de la syntaxe. Une application courante en est la détection de messages électroniques indésirables.

Transposée à des images, cette méthode s'appuie non plus sur des mots classiques, mais sur des mots visuels. Ces mots sont tout simplement des descripteurs visuels, tels que nous les avons décrits dans le Chapitre 2. Chaque image est découpée en petites zones

1. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

appelées patches, sur lesquelles sont extraits les descripteurs. Chaque patch est donc représenté par un ensemble de valeurs, dont le nombre est identique pour tous les patches.

Un apprentissage par partitionnement est réalisé afin d'associer le groupe de mesures extraites de chaque patch à un unique "mot". L'ensemble des mots forme le dictionnaire. L'image complète est représentée par un histogramme sur ces mots. On peut donc définir cette technique comme la représentation d'une distribution basée sur des descripteurs visuels indépendants.

L'inconvénient des BoW est l'abstraction totale qui est faite des relations de dépendance spatiale entre les pixels, qui ont de façon intuitive une grande importance dans les représentations visuelles.

3.2.5 Synthèse des méthodes présentées

Bien entendu, la liste d'algorithmes de classification qui ont été présentés précédemment ne se veut pas exhaustive, elle est présente pour nous permettre d'appréhender les méthodes couramment utilisées dans le domaine applicatif qui nous intéresse. Il existe également d'autres familles de méthodes, comme celles autour de la théorie de Bayes ou de la régression. Nous allons terminer par le Tableau 3.1 qui résume les informations essentielles sur les classificateurs évoqués dans cette partie.

Famille de méthodes	Exs. d'algos.	Paramètres nécessaires	Temps de calcul	Convergence
Arbres	C4.5, J48	si élagage, paramètres d'élagage	Lent	Non garantie
Réseaux de neurones	"Multilayer Perceptron"	Poids initiaux	Lent	
Support Vector Machines	libSVM, SMO	Fonction noyau, paramètres de cette fonction noyau	Rapide	Oui (ou approximation si optimisation du temps de calcul)
Bag of Words		Taille du dictionnaire	Rapide	Non

TABLE 3.1 – Tableau comparatif de quelques familles d'algorithmes de classification

3.2.6 Apprentissage par ensembles

Dans le but d'améliorer les performances de classification, la méthode par ensembles, aussi appelée agrégation de classificateurs ou meta-classification, est envisageable. Ces algorithmes apportent une réponse unique à un problème de classification en combinant plusieurs algorithmes et/ou résultats.

En effet, le théorème du "no free lunch" (littéralement : "aucun repas n'est gratuit") [Wol96] nous certifie qu'aucun algorithme d'apprentissage n'est supérieur aux autres pour tous les problèmes. Il y a également des arguments statistiques favorables à cette méthodologie : minimisation du bruit et fiabilité plus importante des valeurs moyennées.

Des approches variées sont envisageables et peuvent se combiner entre elles :

- Différents algorithmes d'apprentissage
- Différents paramètres des algorithmes
- Différentes attributs : attributs différents ou sélection de sous-ensembles d'attributs
- Différents jeux de données : échantillonnage aléatoire des données (bagging) ou selon les données mal classées (boosting)

La combinaison permettant d'obtenir un résultat à partir des multiples classifications peut être réalisée de façon parallèle, par exemple à l'aide d'un système de votes ou de sélection, ou en série, avec des classifications en cascade.

Deux exemples de telles méthodes sont les Forêts d'Arbres Décisionnels (Random Forest en anglais, proposés dans le début des années 2000 [Bre01]), et les Probabilistic Boosting Tree [Tu05].

Une Forêt d'Arbres Décisionnels construit pendant l'apprentissage un certain nombre d'arbres de décision qui travailleront chacun sur un sous-ensemble d'attributs différent. La classe retenue est le mode des classes individuellement choisies par les arbres, soit la classe qui apparaît le plus grand nombre de fois.

La méthode par Probabilistic Boosting Tree construit un arbre de décision dont chaque nœud est la combinaison d'un certain nombre de classifieurs faibles et dont le résultat est un classifieur fort. Elle fait appel à une stratégie de division et conquête.

Il faut cependant être prudent en mettant en œuvre une telle approche. En effet, rien ne garantit que la méthode par ensembles produira de meilleurs résultats qu'une classification individuelle.

3.3 Evaluation des performances

Réaliser une classification est une chose, l'évaluer en est une autre. Il est impératif de pouvoir quantifier les performances d'un système de classification, afin par exemple de comparer plusieurs méthodes entre elles.

3.3.1 Techniques d'évaluation

Plusieurs approches sont envisageables pour évaluer les performances d'un système de classification. Le lecteur pourra se référer à [Koh95] pour plus de détails.

Resubstitution (Hold-Out) : Afin d'évaluer les performances d'un système de classification dans le cas de bases contenant un grand nombre d'observations, la méthode de la resubstitution est largement plébiscitée. La base de données est séparée de façon aléatoire en trois sous-ensembles :

- l'ensemble d'apprentissage afin de construire le modèle prédictif
- l'ensemble de validation afin de vérifier les performances du modèle et d'ajuster au besoin les paramètres éventuels de l'algorithme de classification
- l'ensemble de test afin d'obtenir une idée des performances futures du modèle. Si le modèle est beaucoup plus proche de l'ensemble d'apprentissage que de celui de test, dans ce cas un problème de sur-apprentissage peut être soupçonné.

Validation croisée : Afin d'évaluer les performances d'un système de classification dans le cas de bases de petite taille comme c'est hélas le cas dans le domaine qui nous intéresse, la méthode la plus répandue est celle de la validation croisée ("cross-validation").

Le principe en est extrêmement simple : la base de données est divisée en deux sous-ensembles, l'un servant à l'apprentissage du classifieur et l'autre permettant de tester ce dernier. Il s'agit d'une validation dite par set de test. Seule, elle ne donne qu'une vague idée des performances du système.

La classification peut en outre être effectuée sur plusieurs partitionnements différents, on la nomme alors validation croisée sur k partitionnements. Cette méthode introduit un biais optimiste dans les résultats.

Afin d'obtenir une expérimentation exhaustive et limiter le biais, on peut réaliser cette dernière sur des partitionnements particuliers, selon une méthode appelée Leave-One-Out (LOO), littéralement "laissons-en un de côté". Pour n éléments dans la base, le set d'apprentissage est constitué de $n - 1$ éléments, l'observation restante étant l'unique élément servant au test. La classification est réitérée n fois, toutes les observations de la base constituant donc à tour de rôle l'élément de test.

Bootstrap : C'est une technique statistique proposée à la fin des années 1970 par Bradley Efron [Efr79] permettant à la fois d'évaluer la stabilité d'un modèle prédictif et d'en améliorer la robustesse. Dans la pratique, la distribution de probabilité des données

est de forme inconnue. On tente de l'approcher, mais si l'échantillon à disposition n'est pas de taille suffisante cette approximation peut être de mauvaise qualité et l'apprentissage biaisé. Le but de cette technique est de pallier à cet écueil et d'obtenir une approximation plus précise de la vraie loi statistique sur les données sans avoir recours à de nouvelles observations.

Le principe est de ré-échantillonner un grand nombre de fois les observations à l'aide de la méthode de Monte-Carlo. Les variables sont affectées à une distribution de probabilités, puis les statistiques d'apparition sont estimées en réalisant un grand nombre de tirages aléatoires. Celle-ci permet de calculer des quantités inconnues en utilisant une suite de nombres au hasard avec une convergence plus rapide vers la solution qu'une exploration systématique. Le bootstrap fournit une distribution et pas seulement un nombre, sur laquelle il est donc possible de calculer un écart-type et un intervalle de confiance pour la moyenne.

L'avantage de cette technique par rapport à la resubstitution ou la validation croisée qui donnent des mesures optimistes, est de quantifier l'optimisme et le corriger.

3.3.2 Mesures d'évaluation

Une fois la resubstitution ou la validation croisée effectuée, des mesures sont calculées afin d'évaluer les résultats de classification. Nous en décrivons quelques-unes, les plus courantes, ci-dessous.

Précision, taux d'erreur : la précision (en anglais : "accuracy") est le pourcentage d'observations classifiées correctement, le taux d'erreur celui d'observations classifiées incorrectement.

$$\text{Précision} = \frac{\text{nombre d'observations correctement classifiées}}{\text{nombre total d'observations}} \quad (3.1)$$

$$\text{Taux d'erreur} = \frac{\text{nombre d'observations incorrectement classifiées}}{\text{nombre total d'observations}} \quad (3.2)$$

$$\text{Taux d'erreur} = 1 - \text{Précision} \quad (3.3)$$

Ces deux mesures sont liées par la relation 3.3. Elles réclament cependant une distribution de classe relativement uniforme. En cas de classification incorrecte, un poids unique est appliqué.

Matrice de confusion : elle permet d'obtenir une vue d'ensemble des résultats de la classification et de savoir pour chaque classe les étiquettes qui ont été assignées à ses observations (dans le cas d'une classification binaire : vrais positifs, vrais négatifs, faux positifs et faux négatifs, présentés dans le Tableau 3.2).

		cible	
		classe1	classe2
modèle	classe 1	vrais positifs A	vrais négatifs B
	classe 2	faux positifs C	faux négatifs D

TABLE 3.2 – Matrice de confusion dans le cas d'une classification binaire

Elle permet en outre, de calculer différentes mesures : la mesure de précision évoquée auparavant, les valeurs prédictives positive et négative, la spécificité. Leurs équations sont présentées ci-dessous, avec A,B,C,D les mesures présentées dans le Tableau 3.2 :

$$\text{sensibilité} = \frac{a}{a + c} \quad (3.4)$$

$$\text{spécificité} = \frac{d}{b + d} \quad (3.5)$$

$$\text{précision totale} = \frac{a + d}{a + b + c + d} \quad (3.6)$$

$$\text{valeur prédictive positive} = \frac{a}{a + b} \quad (3.7)$$

$$\text{valeur prédictive négative} = \frac{d}{c + d} \quad (3.8)$$

La valeur prédictive d'une classification binaire est la probabilité qu'une condition soit présente en fonction du résultat de ce test. La valeur prédictive positive est donc la probabilité que la condition soit présente lorsque le test est positif, et la valeur prédictive négative celle que la condition soit absente lorsque le test est négatif.

La sensibilité d'un test mesure sa capacité à donner un résultat positif lorsque l'hypothèse de départ est vérifiée. Elle s'oppose à la spécificité, qui mesure celle à donner un résultat négatif lorsque l'hypothèse de départ n'est pas vérifiée.

La précision totale (déjà évoquée auparavant) est le ratio entre le nombre d'éléments correctement identifiés sur le nombre total d'éléments.

Risque Bayésien : Il permet d'introduire des coûts inégaux pour les décisions. Par exemple, dans le cas d'un dépistage de cancer, le coût de diagnostiquer un cancer qui n'existe pas est plus petit que celui d'en manquer un. Cette méthode implique cependant de connaître le modèle statistique des données, ce qui introduit un biais optimiste dans les résultats lorsque la base de données est petite et sert à la fois pour l'apprentissage et la validation. Dans la pratique, deux mesures de substitution sont utilisées : le risque attendu et le risque structurel.

Courbes ROC (Receiver Operating Characteristics) : Il s'agit d'un système de coordonnées très répandu dans le domaine médical, dans lequel l'axe des abscisses représente le taux de faux positifs et l'axe des ordonnées celui des vrais positifs. L'aire sous la courbe sert de mesure d'évaluation. Un avantage non négligeable de cette technique est la possibilité de comparer deux résultats de classification. En effet, si une courbe domine l'autre (c'est-à-dire si elle est en tout point au-dessus et à gauche de l'autre courbe sur le graphe) c'est que l'algorithme de classification de la courbe dominante obtient de meilleurs résultats sur les données. Cette méthode d'évaluation a également l'énorme avantage d'être insensible aux effectifs de classe déséquilibrés, problème qui est posé ci-après.

3.3.3 Le cas des effectifs de classe déséquilibrés

Une classification conduite sur deux classes dont l'une contient 95% des observations, l'autre les 5% restants, et qui assigne l'étiquette de la classe majoritaire à tous les points obtient une mesure de précision totale de 95%. À première vue, ce score paraît excellent, il ne représente cependant rien. Il faut donc imaginer d'autres mesures d'évaluation lorsque les effectifs des classes sont déséquilibrés.

Dans le cas d'une classification binaire, le F-score ainsi que les courbes ROC peuvent être mesurés. Dans le cas d'une classification multi-classe, le Balanced Error Rate ou rappel moyen peut être calculé. Il correspond à la moyenne des erreurs sur chaque classe. Il a été proposé dans [RC05] et a été utilisé sur les Challenges Interspeech entre 2009 et 2011 :

$$\text{BET} = 1 - k * \sum_i \frac{A_{ii}}{\sum_j A_{ij}} \quad (3.9)$$

3.4 Conclusion

Ce chapitre a permis d'introduire les notions de base de la classification, de balayer les différentes familles d'algorithmes et de décrire les méthodes permettant leur évaluation. La technique à choisir est fonction du problème et des données, il n'y a pas d'algorithme surpassant tous les autres. Afin de ne plus effectuer de choix manuel, les processus de méta-classification sont de plus en plus couramment employés.

L'objectif de notre projet de caractérisation de lésions hépatiques étant plutôt de nous intéresser à l'extraction de descripteurs visuels, nous nous appuierons donc sur les méthodes existantes de classification, notamment les machines à vecteurs de support, que nous venons de présenter. La matrice de confusion sera notre méthode d'évaluation, que nous réaliserons après une validation croisée "Leave One Out". Ceci est tout à fait adapté à la petite taille de notre base de données.

Caractérisation des lésions hépatiques dans des images scanner

Préambule

Nous allons nous intéresser dans ce chapitre aux travaux de recherche portant sur la caractérisation des lésions hépatiques, dans le cadre d'une classification, mais également de détection ou de recherche par le contenu (en anglais : Content-Based Image Retrieval ou CBIR). Ces trois objectifs sont en effet atteints en s'appuyant de manière similaire sur des descripteurs visuels extraits des images. Nous nous limiterons à la modalité d'acquisition d'images scanner, car l'essentiel des expérimentations autour du foie ont été menées sur celle-ci.

La plupart des études sont menées sur des acquisitions monophasiques, que nous verrons dans un premier temps dans la section 4.1, avant de passer en revue les rares travaux sur du multiphase en section 4.2.

Contents

4.1	Caractérisation des lésions hépatiques dans des acquisitions monophasiques	44
4.2	Caractérisation des lésions hépatiques sur acquisitions multiphasiques	48
4.3	Conclusion	54

Comme évoqué dans l'Introduction, les travaux sur les lésions du foie peuvent être réalisés dans différents cadres. La détection permet d'extraire les lésions présentes sur une image de coupe du foie, la recherche par le contenu de retourner les images d'une base de données qui sont les plus proches de l'image requête, et la classification de différencier les différents types de nodules.

4.1 Caractérisation des lésions hépatiques dans des acquisitions monophasiques

Cette section présente les travaux réalisés sur des examens scanner du foie monophasiques. Certaines équipes se concentrent sur une phase injectée (artérielle ou portale) car elle permet de prendre en compte le réhaussement lié à l'injection du produit de contraste. D'autres font le choix opposé en travaillant sur des images non réhaussées, ce qui réduit les coûts et les risques pour le patient. Enfin, la phase dont sont extraites leurs images n'est pas systématiquement mentionnée.

Nous allons tout d'abord analyser les objectifs, les données, les méthodes d'extraction de descripteurs et de classification, les résultats obtenus, avant de réaliser une synthèse et mettre en avant les forces et faiblesses de ces travaux. Au total, huit études de classification vont être présentées, dont certaines travaillent également sur un objectif de détection et/ou de segmentation. Le tableau ci-dessous présente ces références.

4.1.1 Objectifs

Classification

Nous étudierons huit expérimentations autour de la classification. Les classes présentes dans ces travaux sont le foie sain, les kystes, les angiomes, les CHC et les métastases.

Souvent, la classification est réduite à la distinction entre tumeurs bénignes et malignes, qui a un intérêt clinique évident (Chen [CCC⁺], Huang [HCS06], Kumar [KM10, KMR13]). La classe des lésions bénignes est représentée uniquement par les angiomes, celle des lésions malignes par les CHC seuls dans [KM10], et par les CHC et les métastases dans [HCS06]. Une équipe de Stanford présente une classification tri-classe kystes, angiomes, métastases ([ARC12, BGD⁺04, SPRG13]).

Un des autres objectifs est de distinguer le foie sain : Wang [WZL⁺09] tente de différencier le foie sain des angiomes et des HCC, Mougiakakou [MVNN07] y ajoutant les kystes en quatrième classe.

Référence	Objectifs	Phase	Base de données	Zone d'intérêt	Autre traitement	Descripteurs	Classifieur	Evaluation
Adcock et al. [ARC12]	Classification tri-classe	non décrite	108 lésions : 45 kystes, 45 métastases, 18 angiomes	manuelle (lésion et bordure)		Distance de Bottleneck entre la lésion et toutes les autres selon la méthode des codes barres	SVM	LOO
Bilello et al. [BGD ⁺ 04]	Classification tri-classe	portale	56 slices du foie de 51 patients : 22 kystes, 11 angiomes, 22 métastases, 1 double kyste		selon 2 axes : segmentation de la lésion (basée sur la forme) ou détection de celle-ci	Filtration par des Gaussiennes (paramètres du filtre et valeur moyenne du résultat)	SVM	LOO
Chen et al. [CCC ⁺]	Classification binaire	non décrite			Segmentation du foie	GLCM	Réseau de neurones	
Huang et al. [HCS06]	Classification binaire bénin / malin	sans injection	164 tumeurs : 80 malignes (HCC et métastases) + 84 bénignes (angiomes)	cercle défini au sein de la lésion	égalisation d'histogramme	texture : coefficients normalisés d'autocovariance 2D	SVM (noyau radial)	k-fold validation croisée
Kumar et al. [KM10, KMR13]	Classification binaire bénin / malin	non décrite	200 zones de taille fixe (40*40 pixels) positionnées au centre des lésions : 100 bénignes (angiomes) et 100 malignes (CHC)	Segmentation du foie et des lésions		Statistiques sur l'histogramme et la GLCM, curvelets et statistiques de la GLCM calculée sur le résultat des curvelets	Réseau de neurones	Séparation en 2 sets de taille identique pour l'apprentissage et le test
Mougiakakou et al. [MVNN07]	Classification 4-classes	sans injection	147 images : 76 tissus sains, 19 kystes, 28 angiomes et 24 CHC	manuelle		Statistiques sur l'histogramme et la GLCM, descripteurs de Law, fractales, différences de niveaux de gris	Réseaux de neurones avec une étape de sélection, ainsi qu'un ensemble de 5 classifieurs dont le résultat final est établi à l'aide d'un vote	Bootstrap
Sadafari et al. [SPRG13]	Classification tri-classe ou détection	non décrite	73 zones manuelles de 73 patients : 25 kystes, 24 métastases et 24 angiomes			Bag-of-words (PCA sur les zones d'intérêt découpées en petits carrés de taille égale)	SVM	LOO
Wang et al. [WZL ⁺ 09]	Classification tri-classe	portale		90 zones d'intérêt carrées de taille unique 32*32		Statistiques sur l'histogramme et sur différentes matrices proches de la GLCM (SGLD, GLRLM, GLDM)	SVM	Validation croisée

TABLE 4.1 – Synthèse des éléments caractéristiques des systèmes de classification des lésions hépatiques existants basés sur des examens monophasique

De façon générale, la classification des lésions du foie sur des examens monophasiques reconnaît uniquement 2 ou 3 classes (sauf Mougiakakou dont la méthode s'appuie sur 4 classes [MVNN07]).

4.1.2 Données

Intéressons-nous tout d'abord à la phase sur laquelle les études sont menées. Deux travaillent sur les images non injectées ([HCS06, MVNN07]), trois expérimentations sont effectuées sur la phase portale ([NBR⁺10], [BGD⁺04], [WZL⁺09]), les équipes restantes ne précisant pas sur quelle phase leurs examens scanner ont été acquis.

Les bases de données contiennent un nombre variable de lésions : pour la classification, entre 56 et 200, pour une moyenne de 120 lésions environ.

Les zones d'intérêt sont de façon quasi systématique définies manuellement par un radiologue, sauf lorsqu'elles sont extraites par une segmentation préalable ([BGD⁺04], [KMR13]). Chen ne précise ni le nombre ni la méthode d'extraction des zones d'intérêt [CCC⁺]. Leur forme est variable : certaines sont des carrés de taille unique, d'autres sont rondes, enfin pour certaines la forme n'est pas décrite (elles pourraient être des polygones ou des ellipses).

Les images sont classiquement pré-traitées à l'aide d'une égalisation d'histogrammes afin d'en augmenter le contraste.

4.1.3 Méthodes

Dans certains cas, la classification est précédée par une segmentation. Chen segmente le foie [CCC⁺], Kumar le foie puis les lésions [KM10, KMR13], et [BGD⁺04] les lésions uniquement. Dans d'autres, les auteurs travaillent selon deux axes distincts : la détection en plus de la classification.

Les descripteurs extraits des lésions sont principalement des descripteurs de texture. Tout d'abord des techniques très employées comme les statistiques sur la matrice de co-occurrence ou ses consœurs (4 cas sur 8), ou Law. Sont également présentes des méthodes moins répandues : les coefficients d'auto-covariance, les fractales, les curvelets, les filtres de Gaussiennes, la méthode des code-barres et les bag-of-words. Dans 3 cas sur 8, des descripteurs sur les niveaux de gris sont calculés, à savoir les statistiques sur l'histogramme des niveaux de gris. Les valeurs des caractéristiques extraites peuvent être normalisées ou réduites (par une analyse en composantes principales par exemple).

Seuls deux types d'algorithmes de classification sont représentés : le SVM dans 5 cas sur 8, et les réseaux de neurones dans les cas restants.

Concernant l'évaluation de la classification, elle est généralement réalisée à l'aide d'une validation croisée (5 cas sur 8), voire d'une validation croisée avec LOO (3 cas

sur 8). Le bootstrap est employé dans les articles de Mougiakakou [MVNN07], et Kumar sépare la base de données en deux sets de taille identique pour l'apprentissage et le test [KM10, KMR13]. Chen ne précise pas quelle méthode d'évaluation est employée dans ses travaux [CCC⁺].

4.1.4 Résultats

Différents résultats de classification sont proposés : des matrices de confusion, différentes mesures telles la précision, la sensibilité, la spécificité, le F-score, les courbes ROC. Seule la précision totale étant commune à tous les articles, c'est celle-ci que nous allons évoquer.

Huang obtient au mieux 81.7% avec égalisation d'histogrammes, et 66.50% sans, pour une classification bénign / malin [HCS06]. Il compare les différentes mesures obtenues en faisant varier les paramètres du SVM.

Kumar [KM10, KMR13] obtient entre 79 et 94% selon les groupes de caractéristiques utilisées autour des curvelets (le meilleur étant les mesures extraites de la matrice de co-occurrence des coefficients de curvelets). Est également proposée une comparaison entre les résultats obtenus avec les curvelets et ondelettes (respectivement 94.33 contre 88.6%).

Adock [ARC12] compare les résultats obtenus avec une filtration 1D et 2D, en fonction du type et de l'aire de la lésion. La filtration 2D permet d'améliorer la qualité de la classification par rapport à une filtration 1D. Moins il y a de classes à différencier (4, 3, 2), meilleurs sont les résultats. La taille semble avoir un rôle à jouer : la classification est meilleure lorsque les lésions sont petites (mais ce résultat est à tempérer car ce sous-ensemble de lésions contient davantage de kystes et de métastases proportionnellement aux angiomes par rapport à l'ensemble des lésions de toutes tailles, kystes et métastases qui sont plus aisément reconnus par le système).

Mougiakakou [MVNN07] teste de façon systématique plusieurs architectures, plusieurs groupes de descripteurs, plusieurs méthodes de sélection, plusieurs systèmes de vote permettant de combiner les résultats, et plusieurs ensembles de classifieurs.

4.1.5 Synthèse des travaux

Les systèmes basés sur des images monophasiques de lésions du foie sont variés. Ils s'appuient sur des descripteurs, des algorithmes de classification et des méthodes d'évaluation généralement connus et éprouvés. Les forces de ces propositions sont la liaison avec la clinique dans certains objectifs (comme une distinction sain / pathologique ou bien bénign / malin), et l'automatisation de certains logiciels (avec une détection ou une segmentation avant la classification).

Leurs bases de données sont de petite taille, en comparaison avec celles qu'on peut trouver dans d'autres applications du domaine médical, en raison des difficultés évoquées de construire une banque d'images. Une intervention manuelle (comme la définition des zones d'intérêt) est la plupart du temps nécessaire au cours du processus. Enfin, il paraît ardu, voire impossible de comparer réellement les différentes propositions, par manque de grande base de données accessible publiquement, et par le manque de détails fournis (sur les données ou les méthodes).

4.2 Caractérisation des lésions hépatiques sur acquisitions multiphasiques

Quelques rares équipes se sont penchées sur l'amélioration éventuelle pouvant résulter d'une étude menée sur des examens multiphasés. Afin de se rendre compte des différents schémas de réhaussement du liquide de contraste au cours des phases en fonction des lésions (mais aussi des difficultés inhérentes à la variabilité des lésions au sein d'une même famille), de nombreux travaux médicaux existent sur le sujet. Nous pouvons cependant citer l'article de Nino-Murcia [NMOJ⁺00] expliquant simplement cet aspect clinique d'un point de vue plus informatique. Nous allons présenter les travaux sur le multiphase avant de réaliser une synthèse de leurs propositions.

4.2.1 Un système de Diagnostic Associé par Ordinateur entièrement automatique

Ishiguro et al. décrivent dans [IMMS05] un système de classification complet basé sur les changements temporels des niveaux de gris des pixels (représentant les unités de Hounsfield) entre les phases. Une interface graphique permet de l'exploiter.

Le but est de détecter et reconnaître des vaisseaux sanguins, des kystes, des angiomes, des métastases et des lésions CHC. La base de données contient 20 lésions, dont 11 sont sur les quatre phases et 9 sur deux phases seulement (sans injection et phase portale), soit un total de 58 images. Il n'y a pas d'indication quant au nombre de patients dont sont issus les examens, ni quant à la taille des lésions.

Dans un premier temps, le foie est segmenté dans les images, selon une méthode qui n'est pas décrite et pourrait être manuelle. Les foies segmentés sont ensuite filtrés et alignés entre les phases. Dans un second temps, des segments d'images sont sélectionnés dans les différentes phases selon un critère de voisinage avant d'être fusionnées sur les phases 2 par 2 ("segment mapping"). Dans un troisième temps, les descripteurs visuels suivants sont extraits :

- taille, position, circularité, aplatissement, irrégularité

- valeur relative des intensités entre les phases
- entropie
- nombre de segments de chaque phase

Enfin, la classification est effectuée selon des règles discriminantes pondérées.

Un exemple visuel du résultat de classification obtenu sur 3 lésions est présenté dans cette référence, sans précision sur les scores. Les auteurs évoquent dans la discussion une précision de classification variant selon la taille et le type de lésions, sans donner plus de détails.

Le système proposé est complexe et semble vraiment complet en vue d'un processus de CAD entièrement automatisé (sauf la segmentation mais une référence à des techniques de segmentation automatique du foie est citée). Cependant, l'absence de toute évaluation de la segmentation et de la classification ne permet pas de se faire une idée de la qualité du système proposé.

4.2.2 Une méthode de classification en 3 classes

Duda et al. ont pour objectif dans [DKBW06] de distinguer trois types de tissu : le foie sain, le HCC et le cholangiocarcinoma.

Leur base de données contient 165 lésions acquises sur 3 phases (avant injection, phase artérielle et phase portale) soit un total de 495 images. Les images scanner ont été acquises sur 22 patients selon un protocole défini et unique. Les zones d'intérêt, circulaires, ont été manuellement définies sur les images en évitant les vaisseaux. Elles ont un rayon entre 30 et 70 pixels et ne se chevauchent pas.

Les descripteurs extraits, représentant la texture, sont issus de 4 méthodes classiques : des statistiques de premier ordre, l'entropie de Law, les caractéristiques extraites de la matrice de co-occurrence et de la matrice Run-Length.

La classification est réalisée à l'aide de deux algorithmes : SVM et Dipolar Decision Trees (DDT).

Plusieurs classifications sont effectuées et leurs résultats comparés à l'aide d'une mesure de précision. Pour chaque algorithme de classification sont testés chaque groupe de descripteurs séparément, ainsi que la totalité des descripteurs regroupés, le tout sur chaque phase distincte ainsi que par les trois phases associées.

Voici une synthèse de leurs résultats pour chaque type de classification choisie.

- Expérimentations sur les descripteurs : les meilleures précisions sont observées lorsque la classification s'appuie sur tous les descripteurs
- Expérimentations sur les phases : les meilleures précisions sont obtenues lorsque la classification s'appuie sur toutes les phases
- Expérimentations sur les algorithmes de classification : les meilleurs résultats proviennent de l'algorithme DDT

La meilleure précision, correspondant à 99.53% de tissus correctement étiquetés, est obtenue avec l'algorithme DDT sur tous les descripteurs et toutes les phases.

L'analyse proposée est intéressante de par la diversité des axes d'expérimentation. La seule faiblesse de cette étude est le manque de variabilité "naturelle" dans la base de données d'images. Il y a en effet un nombre faible de patients (22 patients pour 165 lésions soit environ 7 lésions provenant du même patient). Si l'on prend plusieurs lésions de même type sur un patient, il existe une probabilité forte pour que leur dynamique de réhaussement soit proche. D'autre part, un protocole strict a été respecté dans l'acquisition des données. La base de données contient donc moins de variabilité intra-classes qu'une base plus réaliste en termes de clinique. Ces deux éléments introduisent un biais permettant de qualifier la précision de classification obtenue comme étant optimiste. Cette base n'est cependant pas atypique dans le domaine. En effet, les données portant sur les lésions du foie n'étant pas dans des bases publiques comme cela est le cas pour les mammographies ou les scanners de poumon dans la lutte contre le cancer, constituer une base d'images est un défi.

4.2.3 Une méthode de classification en 4 classes utilisant des descripteurs temporels

Une équipe de l'Université de Shanghai Jiao Tong a également conçu un système de classification de lésions du foie multiphasiques, décrit dans [YSW09]. Ils ajoutent à leur outil une étape de sélection des meilleurs descripteurs visuels dans [SSW⁺09], et réalisent une optimisation globale dans [SS11].

Première approche

Leur base de données contient 131 zones d'intérêt acquises sur 4 phases : 64 sont du foie sain, 14 sont des kystes, 26 sont des CHC et 27 sont des angiomes.

La lésion est segmentée manuellement puis une zone d'intérêt carrée de 16 x 16 pixels est extraite. Les descripteurs extraits de ces zones d'intérêt sont :

- Statistiques sur l'histogramme des niveaux de gris
- Quelques-uns des descripteurs d'Haralick sur la matrice de co-occurrence
- Descripteurs temporels issus d'études médicales sur l'IRM : ici S représente la valeur moyenne des pixels de la zone d'intérêt sur la phase i et T le temps écoulé sur la phase i, pour i allant de 0 à 3 (respectivement avant injection, artérielle, portale, tardive) :
 - Intensité Relative du Signal :

$$RSI(i) = \frac{S(i)}{S(0)} \quad (4.1)$$

pour i de 1 à 3

- Tendance d'évolution de l'intensité :

$$ICT(i) = \frac{S(i+1) - S(i)}{T(i+1) - T(i)} \quad (4.2)$$

pour i de 0 à 3

- Ratio du réhaussement du signal :

$$SER = \frac{S(1) - S(0)}{S(3) - S(0)} \quad (4.3)$$

Le SER et le RSI sont notamment employés sur des images IRM dans l'objectif de mesurer la malignité de certaines tumeurs, ou les détecter, à l'aide de seuils définis sur ces valeurs [ACN⁺11, YTK⁺12].

La classification est effectuée à l'aide d'un SVM après conversion du problème multi-classe en 3 classifications binaires après la mise à l'échelle des valeurs dans l'intervalle [-1, +1]. Tout d'abord une première classification différencie le sain du non sain, puis si l'élément est étiqueté comme n'étant pas sain une deuxième classification prend le relais et distingue les kystes des autres lésions, et enfin une troisième classification sépare les angiomes des lésions cancéreuses si la deuxième classification n'avait pas désigné un kyste.

Différentes expérimentations sont présentées en résultats, sur lesquels sont estimés des courbes ROC et leur aire sous la courbe, ainsi qu'une mesure de précision sur laquelle nous revenons ci-dessous :

- Expérimentations sur les descripteurs : chaque ensemble de caractéristiques est testée séparément
- Expérimentations sur les phases : chaque phase est testée séparément
- Expérimentations sur les classes : sain vs non sain, kyste vs autre, angiome vs carcinome.
- Expérimentations sur la sélection : ces expérimentations sont détaillées dans un autre article décrit ci-après

Sélection d'un sous-ensemble optimal de descripteurs

Leur base de données a été étoffée puisque le nombre de zones d'intérêt acquises sur 4 phases est passé de 131 à 179 : 84 sont du foie sain, 22 sont des kystes, 32 sont des CHC et 41 sont des angiomes. La méthode utilisée avant la classification est en revanche identique, tout comme l'ensemble de descripteurs calculés. La technique proposée est une sélection des descripteurs en deux étapes, le but étant d'améliorer la précision de la classification en y introduisant un sous-ensemble de descripteurs optimal.

Pour la première étape, 5 algorithmes différents sélectionnent chacun un sous-ensemble de descripteurs sur le critère de la distance de Manhattan, aboutissant

donc à 5 sous-ensembles qui sont fusionnés dans un même vecteur. Ce vecteur sera l'ensemble de descripteurs introduit dans le classifieur SVM.

Dans la seconde étape, l'algorithme SVM est combiné à un algorithme génétique et une fonction de "fitness" dont les critères pour l'obtention d'un sous-ensemble sont un taux de précision important associé à un faible nombre de descripteurs.

En plus des expérimentations présentées dans le premier article mais réalisées sur la nouvelle base de données, des expériences ont été menées sur les ensembles de descripteurs : totalité des descripteurs, sous-ensembles produits par la première et seconde méthode. La précision de classification en validation croisée k-partitions est estimée afin d'évaluer les performances.

Dans ce qui suit, un "cas" est un couple (type de classification binaire, phase). La précision est meilleure sur la phase portale dans 7 cas sur 9, sur la phase artérielle dans les 2 cas restants. Concernant le premier sous-ensemble sélectionné, contenant 24 descripteurs, la précision est supérieure dans 10 cas sur 12. La seconde étape réduit l'ensemble à 10 descripteurs et obtient une meilleure précision que la non-sélection et la sélection opérée par l'étape une dans tous les cas. Les meilleures précisions obtenues sont 98.29% pour la classification sain / pathologique en phase portale, 98.96% pour la classification kyste / autre en phase artérielle, et 96.43% pour la classification angiome vs carcinome en phase portale.

Seconde approche

La méthode générale est conservée. La base de données a été réduite à 68 échantillons : 11 de tissu sain, 11 kystes, 26 tissus cancéreux et 20 angiomes.

Tout d'abord, la foie est segmenté automatiquement par un algorithme amélioré de montée des eaux. Ce dernier, basé sur la morphologie et un traitement non linéaire, offre une moindre sensibilité au bruit et à la sur-segmentation. Ensuite, les images sont alignées entre les phases à l'aide de descripteurs Scale-Invariant Feature Transform (SIFT). Cette étape permet également de réduire les erreurs de segmentation. Les régions d'intérêt sont détectées et extraites de façon automatique par une transformation en ondelettes de Gabor. En effet, ces dernières reconnaissent les régions à haute fréquence, ce qui est effectivement le cas des lésions. Au niveau des descripteurs extraits, seules les mesures temporelles ont été conservées. Les nouvelles mesures introduites sont extraites sur l'histogramme des niveaux de gris d'une image obtenue à l'aide de la technique du scattergram. Cette dernière combine les informations de deux images issues de deux phases temporelles (dont une est toujours la phase de pré-contraste) en une seule image. De nouveau, la classification est faite de façon hiérarchique : sain / pathologique, puis kyste / non kyste, puis angiome / cancer, et . Cette fois cependant, l'algorithme de classification est un réseau de neurones – Expérimentations sur les phases : les phases sont testées deux à deux : pré-contraste et artérielle, et veineuse, et portale

- Expérimentations sur les classes : sain versus non sain, kyste versus autre, angiome versus carcinome.
- Expérimentations sur les ensembles de descripteurs : statistiques sur le scattergram, statistiques sur le scattergram réduit, descripteurs temporeux et scattergram, descripteurs temporeux et scattergram réduit, totalité des descripteurs réduit

Discussion

L'approche de ces articles est globale et les expériences ont été menées de façon quasiment exhaustive, ce qui rend ces travaux très intéressants. Tout d'abord les phases sont traitées séparément puis deux à deux, ensuite chaque ensemble de descripteurs est testé avant qu'une sélection ne soit effectuée sur ceux-ci. Enfin au niveau des classes l'approche retenue est de réaliser 3 classifications binaires afin de distinguer au mieux 4 classes.

Les résultats présentés montrent que la distinction entre kystes et les autres pathologies est excellente, suivie de près par celle entre foie sain et non sain. Différencier l'angiome du carcinome est une tâche plus ardue. On peut regretter que les 4 phases aient été toujours testées séparément et jamais regroupées dans la classification, ce qui aurait permis de voir éventuellement une amélioration introduite par le multiphase dans les résultats. Il est difficile de tirer une conclusion quant à la phase permettant d'obtenir le meilleur score de classification sur une distinction binaire spécifique, car elle varie selon les expérimentations. Enfin, les descripteurs temporels, bien que provenant de travaux sur l'IRM, obtiennent d'excellents résultats de classification tout en étant très simples à implémenter et rapides à calculer.

4.2.4 Synthèse des travaux

Les systèmes basés sur des images multiphasiques de lésions du foie sont peu nombreux. Leurs bases de données sont, ici aussi, de petite taille. Le nombre de classes reste cependant faible (3 et 4 classes).

Tout comme les travaux portant sur des images hépatiques monophasiques, ils s'appuient sur des descripteurs, des algorithmes de classification et des méthodes d'évaluation principalement connus et éprouvés. On peut cependant se réjouir de l'introduction de valeurs étudiant la dimension temporelle dans les travaux de Ye.

Les équipes de Duda et de la School of Software ont une approche très rigoureuse dans leurs expérimentations. Un grand nombre de tests sont présentés avec leurs résultats.

4.3 Conclusion

Nous avons passé en revue dans ce chapitre différents travaux portant sur les lésions hépatiques des images scanner.

Peu s'inscrivent dans un contexte multiphasique, ce que pourtant nous jugeons indispensable et vers quoi nous orienterons nos propres recherches. L'évolution temporelle des lésions étant un élément important de la réflexion clinique, il semble en effet naturel que les informations pouvant en être extraites améliorent la classification des nodules. Nous proposerons dans les contributions qui suivent une approche tenant compte de cet élément temporel.

Dans les trois approches présentées, il est à noter que les bases de données sont plutôt de petite taille, et que le nombre de classes est plutôt limité, nous souhaitons donc dans notre propre étude élargir les types de lésions reconnus, à défaut d'avoir pu obtenir une grande base de lésions.

Deuxième partie

Contributions

Mise en place des données et développement d'une première méthode de caractérisation sur des examens monophasiques

Préambule

Ce chapitre présente les différents aspects du système que nous proposons. La construction de la base de données sera tout d'abord détaillée, puis les méthodes d'extraction de descripteurs visuels, de classification et d'évaluation. Cet ensemble sera notre système de base, pour lequel nous présenterons des résultats de référence, et que nous validerons et tenterons d'optimiser dans les chapitres suivants de cette partie.

Contents

5.1	Introduction	58
5.2	Description des données	58
5.3	Description de la méthode	64
5.4	Analyse des résultats	66
5.5	Conclusion	68

5.1 Introduction

Nous avons construit un premier outil complet permettant de caractériser les lésions hépatiques : il effectue un pré-traitement sur les images, extrait les descripteurs visuels, lance la classification puis l'évalue.

La construction de la base de données a été construite en étroite collaboration avec un radiologue. Nous avons souhaité un ensemble réaliste, proche de ce sur quoi les médecins travaillent dans leur routine quotidienne. La présentation des différentes classes de lésions, d'un point de vue à la fois clinique et visuel, sera effectuée pour une meilleure compréhension des contraintes de l'application.

Dans cette version préliminaire, les descripteurs visuels et la méthode de classification sont des classiques de l'état de l'art. Un premier problème va immédiatement se poser : celui du temps de calcul. Nous verrons qu'il est dû à l'extraction d'un groupe de descripteurs en particulier. Après une analyse des possibilités, une solution sera mise en place.

Nous serons ensuite fin prêts pour des expérimentations plus poussées. Les résultats présentés dans ce chapitre serviront de référence et d'éléments de comparaison pour ces futurs travaux.

5.2 Description des données

Nous allons détailler ici la construction de notre base de données de lésions du foie. Tous les choix dans le cadre de cette construction ont été effectués par un radiologue expérimenté, ou en collaboration avec celui-ci.

5.2.1 Définition clinique des classes

Les lésions sont également appelées nodules dans ce qui suit. Un nodule est défini par une formation anormale, généralement de forme arrondie, qui se développe à la surface d'un tissu ou dans un organe.

Nous l'avons évoqué dans l'introduction, l'imagerie hépatique a une grande importance à l'heure actuelle. Notre radiologue a choisi de représenter un plus grand nombre de classes que les travaux existants, afin de couvrir environ 95 pour cent des possibilités pouvant se présenter lorsqu'un patient passe un scanner du foie. Le foie sain est également représenté, il est en effet important qu'un système informatisé puisse distinguer des tissus sains des tissus anormaux.

Le tableau 5.1 présente les différentes classes représentées dans notre base.

Dans ce qui suit, ces classes seront nommées comme des "lésions" pour plus de simplicité, même si techniquement parlant le tissu sain n'en fait pas partie.

famille	classe	informations cliniques
tissus sains	sain	il s'agit du "parenchyme hépatique", c'est-à-dire les tissus sains du foie exclusivement sans lésion ni vaisseau
lésions bénignes	adénome	l'adénome se trouve le plus souvent chez des femmes jeunes sous contraception orale
	angiome	l'angiome est une tumeur bénigne qui se développe à partir des vaisseaux sanguins. En règle générale, le patient, typiquement un adulte entre 30 et 50 ans, ne présente aucun symptôme, elle est donc découverte le plus souvent de façon fortuite. Dans de nombreux cas, plusieurs angiomes sont présents. La plupart du temps, l'angiome est petit et ne nécessite aucun traitement ni surveillance, par contre s'il est imposant, une intervention peut être nécessaire pour le retirer.
	Hyperplasie Nodulaire du Foie (HNF)	il s'agit d'une tumeur bénigne et rare, de taille très variable. Dans certains cas, notamment lorsque sa taille est suffisamment grande, on observe une cicatrice centrale. C'est une lésion qui n'évolue que rarement, et qui ne nécessite aucun traitement ni surveillance, sauf s'il est de grande taille (car il peut entraîner une hémorragie)
	kyste	il s'agit d'une cavité remplie de liquide. Les kystes sont très fréquents, ils ne dégénèrent jamais, ne se traitent pas, il n'est même pas besoin de les surveiller.
lésions malignes	Carcinome Hépatocellulaire (CHC)	il s'agit du cancer primitif du foie (il se développe en premier lieu dans le tissu hépatique)
	métastase	les métastases sont des lésions cancéreuses qui ont "migré" depuis le foyer primitif de la maladie, il peut par exemple s'agir de métastases induites par un cancer colorectal ou du sein.

TABLE 5.1 – Présentation des classes de notre base de données d'un point de vue clinique

5.2.2 Description visuelle des classes

Bien entendu, ces classes se caractérisent par des aspects visuels différents, et notamment des cinétiques de réhaussement particulières. Le tableau 5.2 montre l'aspect des lésions sélectionnées au cours des différentes phases de l'examen scanner. Il permet de se rendre compte de l'utilité potentielle des examens multiphasés comparés à des images acquises sur une phase unique.

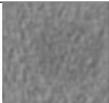
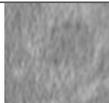
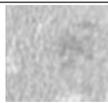
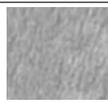
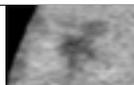
	1 : phase pré-injectée	2 : phase artérielle	3 : phase portale	4 : phase tardive
Abcès				
Adénome				
Kyste				
HNF				
Angiome				
CHC				
Métastase				

TABLE 5.2 – Aspect visuel des lésions par type et par phase

Les caractéristiques courantes des classes de lésions à chaque phase sont quant à elles détaillées dans le tableau 5.3. Les radiologues se fondent en partie sur ces informations pour établir un diagnostic, mais bien entendu, une grande variabilité est observable. Les différentes informations présentées sont : les caractéristiques de contours, l'homogénéité, la présence éventuelle d'une cicatrice centrale, et la densité par rapport au foie sain. La densité des tissus par rapport au parenchyme hépatique est un élément important du diagnostic, les tissus sont dits hypodenses s'ils sont moins denses que le foie sain (ce qui se traduit visuellement par un niveau de gris plus sombre que le foie), isodenses s'ils se confondent avec lui et hyperdenses s'ils sont plus denses.

classe	contours	homogénéité	densité	divers
tissus sains	Homogène avec un réhaussement très léger en phases artérielles et portales.			
adénome	Nets, capsule parfois visible	Hétérogène au temps artériel et/ou portal, homogène au temps tardif,	Isodense (phase 1), réhaussement centripète en "mottes" périphériques (phase 2), isodense (phases 3 et 4)	Pas de cicatrice
angiome	Nets, capsule parfois visible	Hétérogène	Hyperdense avec un réhaussement périphérique (phase 2), Hyperdense selon un réhaussement centripète (phase 3), Hyperdense selon un réhaussement homogène (phase 4)	Cicatrice centrale
HNF	Mal limités	Homogène (sauf zone étoile)	Hyperdense selon un réhaussement hétérogène (phase 2), Plus ou moins isodense sauf la cicatrice (phase 3), Isodense parfois hypodense (phase 4)	Petite cicatrice centrale
Kyste biliaire simple	Pas de paroi	Homogène	Densité équivalente à celle de l'eau, sans réhaussement au cours des phases	Pas de cicatrice, taille très variable (quelques mm à plusieurs cm)
CHC	Nets	Hétérogène	Hypodense voire isodense (phase 1), Hyperdense selon un réhaussement diffus hétérogène (phase 2), plus ou moins isodense (phase 3), hypodense (phase 4)	Pas de cicatrice centrale, peut être de grande taille
métastase	Contours irréguliers et mal délimités	Homogène (hétérogène si nécrose centrale)	Paroi épaisse et réhaussée, si nécrose zone correspondante non réhaussée (phases 3 et 4)	Pas de cicatrice centrale, souvent multiples

TABLE 5.3 – Présentation des caractéristiques des lésions au cours des phases temporelles

5.2.3 Sélection et récupération des images médicales

Il s'agit d'une étude rétrospective de cas cliniques réels, ce qui implique que les patients n'ont pas été irradiés pour les besoins de nos recherches, et qu'aucun protocole spécifique n'a été suivi. Nous retrouvons donc des examens "de routine" qui introduisent une certaine variabilité dans les données mais sont d'un autre côté beaucoup plus réalistes. Les progrès dans l'imagerie scanner étant constants, les examens sélectionnés n'ont pas été acquis avant 2010 pour ne pas introduire une trop grande différence de qualité ou méthode.

Le radiologue a tout d'abord sélectionné un ensemble de cas susceptibles d'entrer dans l'étude. Puis il a passé en revue ces cas et retiré les dossiers pour lesquels le diagnostic n'était pas clairement établi (par manque de compte-rendus de travaux complémentaires biologiques ou d'imagerie). Enfin, il a éliminé dans les dossiers restants ceux qui présentaient un profil trop atypique, tout en conservant tout de même une certaine variabilité. La plupart des examens comprennent 4 phases, mais certains n'en contiennent que 3.

Plusieurs lésions d'une même classe peuvent provenir d'un patient, mais dans le cadre de la construction de notre base de données 2 (voir ci-après), ce nombre est limité à 3. Cette restriction complique la construction de la base par rapport à certains travaux existants, et freine son extension, mais permet d'obtenir une base plus réaliste. En effet, les schémas de réhaussement de plusieurs lésions de même type sur le même patient devraient être très proches, ce qui devrait introduire un biais dans les résultats de classification.

5.2.4 Régions d'intérêt

Les images sont au format DICOM, ce qui implique que le nombre de niveaux de gris est très large. Un pré-traitement est donc effectué afin de seuiller les niveaux de gris au-delà de l'intervalle $[-2048, +2047]$. Quelques petits essais nous ont permis de conclure que cela n'influe en rien sur les résultats de classification. En effet, les pixels de valeur basse correspondent à du noir donc à l'air, ceux de valeur haute à du blanc donc à du tissu osseux, ce qui laisse intacts les niveaux de gris centraux qui nous intéressent ici. Les images ont une taille unique de 512 x 512 pixels.

Le rectangle englobant la lésion a été délimité à la main. Il est de même taille sur les 4 phases, mais sa position peut varier légèrement. En effet, il est difficile pour le patient de rester complètement immobile pendant tout l'examen, surtout au niveau de l'abdomen que la respiration modifie en permanence.

Afin de se rapprocher au plus près de la lésion et d'éliminer les tissus proches mais n'en faisant pas partie, nous sommes partis du constat que la plupart des nodules avaient une forme arrondie ou elliptique. Notre zone d'intérêt est donc l'ellipse in-

cluse dans la zone d'intérêt rectangulaire définie précédemment. Elle est extraite de façon automatique. Les zones d'intérêt sont illustrées dans la figure 5.1.



FIGURE 5.1 – Exemple de nos zones d'intérêt : rectangulaire manuelle (à gauche), elliptique extraite (à droite)

5.2.5 Présentation des bases de données utilisées

Dans les travaux qui suivent, deux bases de données différentes vont être employées, respectivement nommées "Base 1" et "Base 2". La seconde est une amélioration de la première, selon cinq axes principaux. La Base 1 est tout de même présentée ici, puisqu'une expérimentation difficilement réalisable une seconde fois sur la nouvelle base (car faisant intervenir deux médecins radiologues, voir le Chapitre 6) a été conduite sur celle-ci.

La Base 1 ne contient pas de foie sain, ce qui a été corrigé dans la Base 2 car, cliniquement parlant, il est important de le distinguer des tissus pathologiques. Le parenchyme hépatique correspond dans tous les cas à une zone rectangulaire de 32x32 pixels. De plus, plusieurs lésions peuvent provenir du même patient dans la Base 1, alors que ce nombre est limité à trois dans notre Base 2. Les zones d'intérêt manuelles ont été redéfinies afin d'être plus précises et surtout d'avoir la même taille sur les quatre phases des acquisitions. Certaines lésions (notamment des kystes, généralement petits) ont été écartés car nous avons considéré dans la Base 2 une taille minimale requise de 17 * 17 pixels. Enfin, les adénomes, rares et donc plutôt difficiles à trouver, ont été écartés de notre Base 2 car après réexamen des acquisitions (au moment de la construction de la seconde base) le radiologue avec lequel nous travaillons les a qualifiés d'atypiques. Cette classe, nous le verrons, était très bien différenciée des autres lors des expériences de classification sur la Base 1, mais ne correspondait hélas pas à une généralité clinique. Il ne nous a pas été possible d'acquérir d'autres examens scanner d'adénomes pour les insérer dans la Base 2.

Le Tableau 5.4 présente les effectifs des classes au sein des deux bases de données, ainsi que le total des lésions et le nombre de patients différents sur lesquelles elles ont été extraites. Elles contiennent toutes deux une centaine de lésions. Les 6 abcès ont été conservés entre les deux bases. Le nombre d'angiomes, HNF, kystes, métastases, est diminué dans la Base 2 (respectivement de 13 à 12, de 8 à 6, et de 38 à 16) pour les raisons évoquées ci-dessus. Enfin, la classe des CHC a été complétée, passant de 8 lésions dans la Base 1 à 12 dans la Base 2.

Les expérimentations seront conduites sur les examens monophasiques ou multiphasiques. Selon le cas, les bases seront donc appelées "Base 1 monophasique", "Base 2 monophasique", "Base 1 multiphasique" et "Base 2 multiphasique".

TYPE - BASE	Base 1	Base 2
Abcès	6	6
Adénome	10	0
Angiome	13	12
CHC	8	12
Foie sain	0	37
HFN	8	6
Kyste	25	11
Métastase	38	16
TOTAL	107	100
PATIENTS DIFFERENTS	33	37

TABLE 5.4 – Composition de nos bases de données

5.3 Description de la méthode

Nous allons présenter ici la méthode générale suivie ainsi que les aspects techniques de programmation, avant de détailler les descripteurs visuels extraits des images hépatiques et l'algorithme de classification.

5.3.1 Principe

Le dispositif de notre système de caractérisation est très simple, il est présenté dans la Figure 5.2. Tout d'abord, des descripteurs visuels sont extraits sur les zones d'intérêt pré-traitées des lésions. Les différentes mesures sont concaténées en un immense vecteur qui est injecté dans un algorithme de classification. Afin d'évaluer la qualité de cette classification, nous travaillons sur un modèle de Leave One Out (voir chapitre 3).

Notre ensemble de descripteurs de base contient :

- 4 mesures statistiques sur l'histogramme des niveaux de gris (moyenne, écart-type, kurtosis, skewness)
- 42 valeurs de Law : 3 statistiques (moyenne, écart-type, entropie) calculées sur les 14 images d'énergie obtenues après application de différents filtres sur les images
- 4 attributs des champs de Markov (moyenne de la gaussienne, theta)
- 36 descripteurs d'Haralick (les 9 couramment utilisés sur 4 directions : horizontale, verticale, et les deux diagonales)

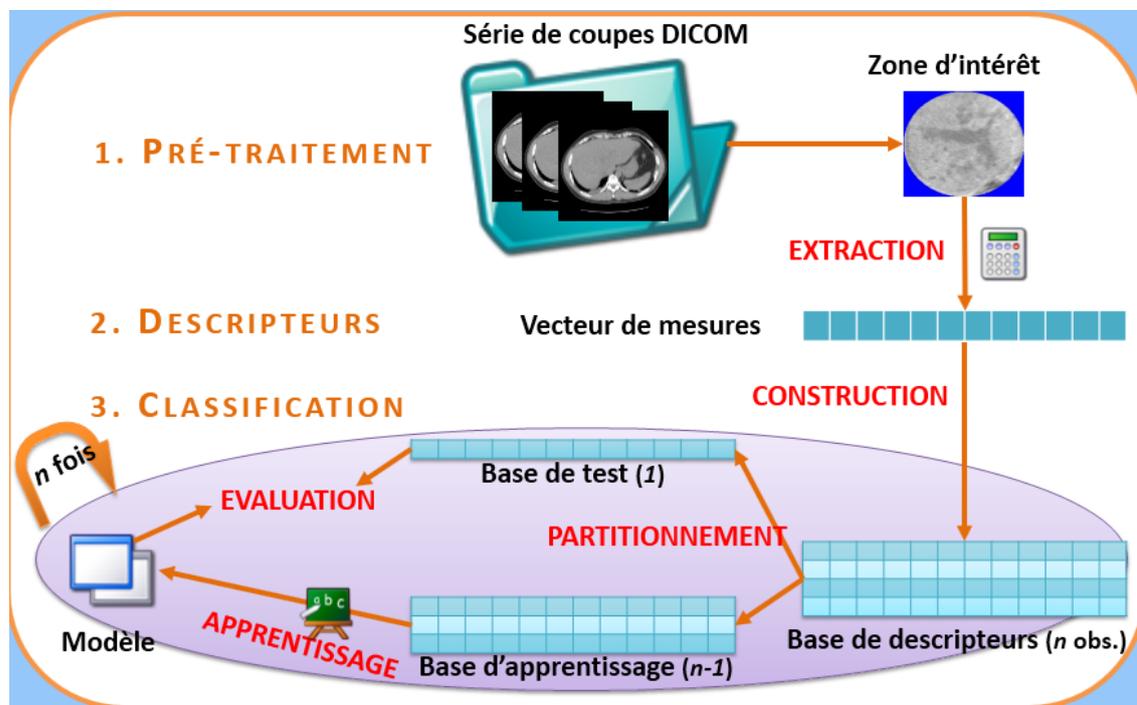


FIGURE 5.2 – Présentation du système complet de caractérisation des lésions hépatiques

Tous ces descripteurs ont été décrits dans le chapitre 2. Ils sont répandus et éprouvés, ce qui nous semble idéal pour une première approche. Nous appellerons ces descripteurs "Ensemble 1". Il est à noter que le temps de calcul nécessaire à l'extraction de cet ensemble de mesures est très important, ce qui pose problème dans un cadre industriel et sur lequel nous travaillerons dans le chapitre 7.

Parmi les algorithmes de classification décrits dans le chapitre 1.3, nous avons décidé d'employer un SVM, principalement pour les deux raisons suivantes. Tout d'abord, les machines à vecteur de support sont en règle générale plus rapides que les arbres ou les réseaux de neurones. De plus, l'implémentation choisie, Support Method Optimization (SMO) [Pla98b]), est fournie gratuitement dans le logiciel open source Weka¹. Le noyau du SVM est polynomial d'exposant 1.

1. Université de Waikato www.cs.waikato.ac.nz

5.3.2 Programmation

Notre programme est écrit en C++, et nous utilisons la librairie Insight Segmentation and Registration Toolkit (ITK²). Elle propose en effet de nombreuses implémentations de techniques de manipulation des images, tout particulièrement médicales, et des algorithmes divers et variés de traitement (filtrage, segmentation, recalage etc.). La société Kitware³ a développé cet ensemble d'outils qui reste sous licence libre. Nous avons adapté certaines classes (comme par exemple celle du type d'images) pour correspondre à nos besoins.

Le logiciel open source Weka est appelé par le programme principal afin d'effectuer la classification. Afin de contrôler l'évaluation croisée Leave One Out (voir Chapitre 3), des scripts écrits en PERL par nos soins extraient les résultats des expériences de classification. Tout le développement est réalisé sous le système d'exploitation libre Linux (distribution Ubuntu).



5.4 Analyse des résultats

5.4.1 Présentation

Nous allons détailler les résultats de classification obtenus à l'aide des méthodes décrites ci-dessus sur notre Base 1 monophasique, et sur les zones d'intérêt elliptiques, avec l'Ensemble 1 de descripteurs.

La matrice de confusion des résultats de classification multi-classe est présentée dans le Tableau 5.5. Pour une explication sur ces mesures choisies dans le but d'évaluer la qualité de la classification, le lecteur pourra se référer au Chapitre 3.

2. www.itk.org

3. www.kitware.com

classe réelle \ étiquetée	Ab	Ad	Ky	HN	An	CH	Mé
Abcès	2	0	1	1	1	1	0
Adénome	0	10	0	0	0	0	0
Kyste	0	1	23	0	0	0	1
HNF	0	0	0	0	2	0	4
Angiome	3	1	0	2	2	0	1
CHC	1	2	0	0	0	3	7
Métastase	4	2	3	1	2	2	24
Instances correctement classées : 59,81%							

TABLE 5.5 – Matrice de confusion et taux de réussite global de classification obtenus sur notre Base 1 sur l'Ensemble 1 de descripteurs

5.4.2 Discussion

Globalement, le score de classification multi-classe, inférieur à 60% de réussite, n'est pas satisfaisant. Des disparités sont à noter cependant entre les différentes classes.

Les classes kyste et adénome sont très bien reconnues (respectivement 23 lésions sur 25 reconnues, et 10 lésions sur 10). Les kystes sont en effet insensibles à l'évolution du produit de contraste entre les phases, et sont hypodenses en permanence.

Les métastases sont ensuite reconnues dans environ 63% des cas (24 lésions sur 38). Cependant, le vecteur de confusion de ces nodules est très étalé : les métastases sont confondues avec toutes les autres classes.

Enfin, les abcès, les angiomes et les CHC sont moins bien étiquetés que par le hasard (entre 22 et 33% de réussite seulement). La pire classe est celle des HNF, puisqu'aucun des nodules de cette classe n'a été correctement étiqueté au cours de l'expérience. Les angiomes, CHC et HNF sont des lésions qualifiées d'hypervasculaires, en raison d'un schéma de réhaussement particulier : ils sont beaucoup plus denses en phases artérielles et/ou portales avant de revenir à leur densité normale. Quant aux abcès, ils ne sont pas hypervasculaires mais ont également un schéma de réhaussement particulier. Sur la phase portale, on peut les confondre avec d'autres lésions.

En n'extrayant des descripteurs que sur la phase portale, cette évolution n'est pas prise en compte. C'est pourquoi, dès le chapitre suivant, nous allons introduire un système basé sur des examens scanner multiphasiques.

5.5 Conclusion

Ce chapitre présente les différentes caractéristiques du premier système proposé.

Les images sur lesquelles portent ces travaux ont été décrites, leur lien avec la pratique médicale effectuée, et la construction de nos bases de données détaillée. Ce sont des régions d'intérêt elliptiques, au format d'images médicales DICOM dont la dynamique des niveaux de gris est très large, et nos bases de données contiennent 8 classes de lésions différentes.

Le principe général de notre outil a ensuite été évoqué : une présentation des étapes, un premier ensemble de descripteurs sélectionnés à partir de la littérature existante, la méthode de classification choisie (un algorithme SVM) et son évaluation grâce à une validation croisée "Leave One Out".

Enfin, les premiers résultats obtenus ont été analysés. Ils sont insatisfaisants, et cette méthode présente certaines limitations. Nous allons l'améliorer selon trois axes. Tout d'abord, prendre en compte la dynamique temporelle de réhaussement des lésions en travaillant sur des examens multiphasiques dans le Chapitre 6. Puis nous allons analyser l'ensemble des descripteurs existants et l'étendre dans le Chapitre 7, en étudiant notamment les nouvelles informations introduites par les examens multiphasiques et en résolvant la problématique du temps de calcul important que prend ce premier outil pour extraire les descripteurs. Enfin, nous allons appliquer cet ensemble de descripteurs à d'autres objectifs de caractérisation des lésions hépatiques.

Examens scanner multiphasiques

Préambule

Dans ce chapitre, les examens multiphasiques sont introduits dans nos expérimentations, en lieu et place des examens monophasiques classiquement présents dans les travaux existants sur les nodules hépatiques. Un comparatif des données et méthodes avec les outils multiphasiques que nous avons détaillés dans le Chapitre 4 sera effectué. Nous validerons cette approche par une comparaison entre les résultats de classification obtenus sur des examens monophasiques et multiphasiques, puis par une expertise de deux radiologues.

Contents

6.1	Introduction et comparaison avec les systèmes existants	70
6.2	Amélioration des résultats par la prise en compte du multiphase . . .	72
6.3	Comparaison avec une classification par des experts	74
6.4	Problème de la normalisation des intensités	77
6.5	Conclusion	83

6.1 Introduction et comparaison avec les systèmes existants

Le lecteur peut se référer à l'introduction pour plus de précisions cliniques concernant les acquisitions scanner multiphasiques, le principe étant rappelé dans la Figure 6.1. Dans ce chapitre, nous souhaitons aller dans le même sens que le cheminement du radiologue vers le diagnostic. Pour cela, il se sert des informations visuelles évoluant au cours du temps (est-ce que la lésion se réhausse ? de façon homogène ou en mottes périphériques ? etc.). C'est pour cette raison qu'il nous semble indispensable que notre outil extraie des informations des images provenant des différentes phases de l'examen scanner. Les résultats de classification que nous avons obtenu dans le chapitre précédent sur nos bases de données semblent également pointer vers cette direction.

Notre outil évolue donc pour prendre en compte de 3 à 4 images par lésion, correspondant aux acquisitions sur les différentes phases. Afin de ne pas avoir un temps de calcul trop important, un parallélisme sur les cœurs du processeur a été mis en place. Pour chaque lésion, les calculs sur les images de ses différentes phases sont réalisées en parallèle. La Figure 6.2 présente le nouveau dispositif. Tous les descripteurs, calculés sur toutes les phases, sont concaténés pour former un unique vecteur par lésion. Ce vecteur sera ensuite entré, comme cela été le cas, dans le système de classification - évaluation.

Nous avons comparé les principales caractéristiques de notre système avec les travaux de la School of Software de Shanghai [YSW09, SSW⁺09, SS11] et Duda et al. [DKBW06], tous deux décrits dans le Chapitre 4 évoquant les études existantes sur les lésions du foie. Le troisième outil présenté contenant trop peu de détails et aucun résultat quantitatif, nous l'avons écarté de ce comparatif. Les données ainsi que les méthodes sont détaillées dans le Tableau 6.1.

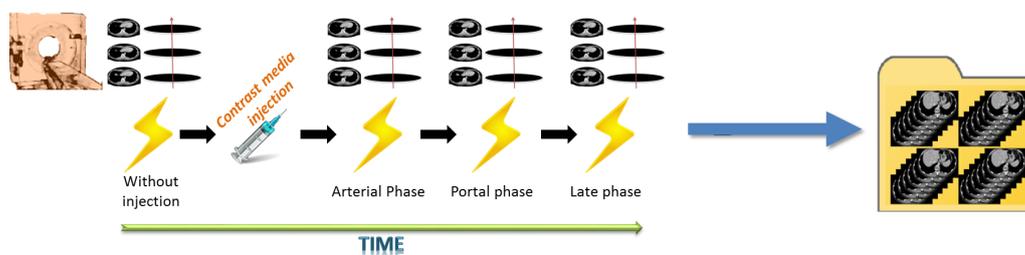


FIGURE 6.1 – *Principe des acquisitions tomographiques multiphasiques*

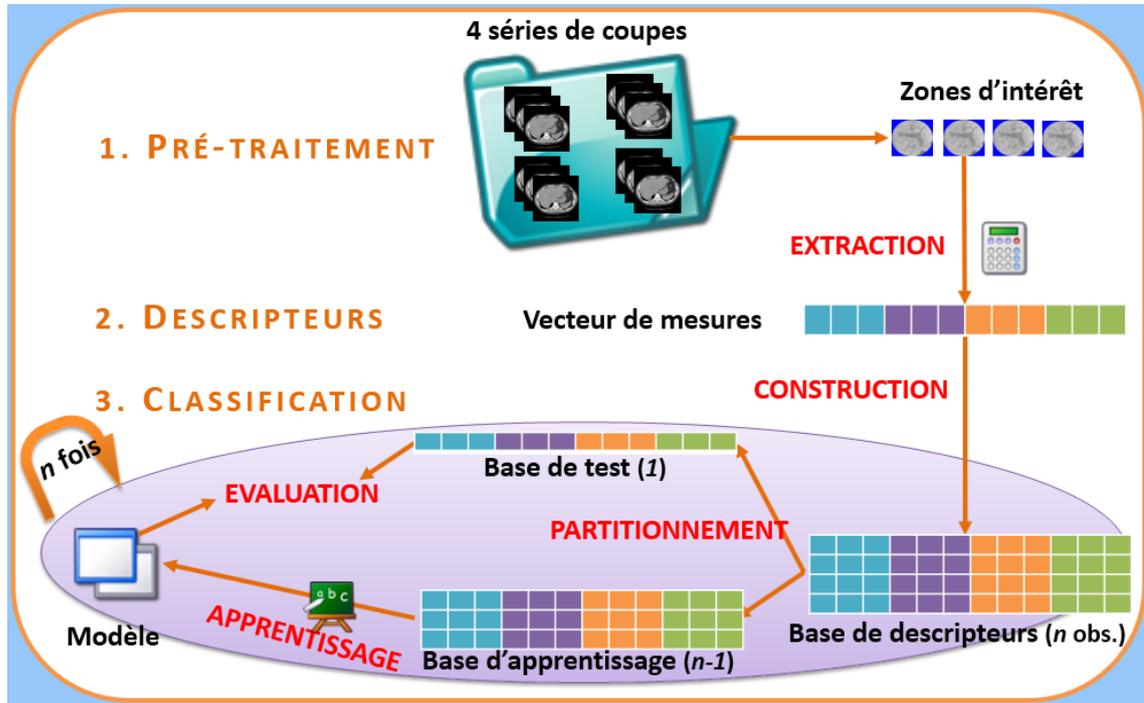


FIGURE 6.2 – Présentation du système complet de caractérisation des lésions hépatiques sur acquisitions multiphasiques

Malgré un nombre de lésions inférieur, notre base est étoffée d'un plus grand nombre de classes différentes. De plus, tout comme dans notre base, les zones d'intérêt ont été délimitées manuellement dans les autres travaux, mais ils n'englobent pas la lésion entière contrairement à nos zones d'intérêt. D'après les radiologues, sur certaines classes comme les angiomes, l'observation du réhaussement des tissus en périphérie de la lésion est essentielle pour leur reconnaissance, d'où notre choix.

Concernant les méthodes, les descripteurs visuels ainsi que les algorithmes de classification sont classiques. L'intéressant travail de différenciation séquentielle réalisé par la School of Software [YSW09,SSW⁺09,SS11] est cependant à noter, car il permet de répondre aux 3 questions cliniques essentielles : est-ce qu'il s'agit d'un tissu sain ou pathologique ? s'il est pathologique, est-il bénin (kyste) ou non, et si non de quelle type de lésion s'agit-il (CHC ou angiome) ?

Caractéristiques	[YSW09, SSW ⁺ 09, SS11]	[DKBW06]	NOTRE TRAVAIL (Base 1 / 2)
Nombre de lésions	131	165	107 / 100
Phases	4	3 (pas de phase tardive)	3 ou 4 phases selon les examens
Classes	CHC / kyste / angiome / foie sain	CHC / foie sain / cholangio-carcinoma	<ul style="list-style-type: none"> - abcès / CHC / kyste / angiome / HNF / métastase - Base 1 : adénome - Base 2 : foie sain
Zones d'intérêt	manuelle, carrée de taille 16x16 pixels	manuelle, carrée d'un rayon de 30 à 70 pixels	ellipse automatiquement extraite du cadre rectangulaire manuellement délimité
Descripteurs	<ul style="list-style-type: none"> - Statistiques sur l'histogramme - Statistiques sur la matrice de co-occurrence - Descripteurs temporels issus de l'IRM 	<ul style="list-style-type: none"> - Statistiques sur l'histogramme - Statistiques sur la matrice de co-occurrence - Mesure d'entropie de Law - Statistiques sur la matrice Run-Length 	<ul style="list-style-type: none"> - Statistiques sur l'histogramme - Champs de Markov Gaussiens - Mesures de Law - Statistiques sur la matrice de co-occurrence
Classification	SVM	SVM	<ul style="list-style-type: none"> - SVM - Arbre de décision dipolaire
Objectif	3 différenciations binomiales séquentielles : <ul style="list-style-type: none"> - sain versus pathologique - kyste versus non-kyste - CHC versus angiome 	Distinguer les 3 classes	Distinguer les 7 classes

TABLE 6.1 – Comparaison de notre système avec les travaux de la littérature

6.2 Amélioration des résultats par la prise en compte du multiphase

6.2.1 Analyse des résultats

Nous avons évalué la qualité de la classification des lésions sur la Base 1 avec l'Ensemble 1 de descripteurs. Les résultats obtenus sont regroupés dans le Tableau 6.2.

Ces derniers montrent que l'introduction d'informations multiphasiques permet une amélioration significative de la performance globale de classification multi-classe : on passe de 59,81% à 68,22% de lésions reconnues, soit une augmentation de près de 10%.

Cette augmentation n'est cependant pas homogène parmi les classes, seules certaines en bénéficient : les angiomes (de 2 à 7 lésions reconnues sur 9 sur la Base 1) et les CHC (de 3 à 6 lésions sur 13 sur la Base 1). Ceci s'explique par le fait que ces nodules sont "hypervasculaires", ce qui signifie qu'ils se réhaussent de façon importante en phases artérielle et portale. Par contre, pour les autres, et ce même pour les HNF qui sont également hypervasculaires, l'influence paraît nulle. Il est enfin à noter que les examens multiphasiques ne permettent pas une meilleure distinction sur les HNF, malgré le fait qu'ils soient hypervasculaires, et les abcès, qui ne sont pas hypervasculaires mais ont un réhaussement temporel particulier.

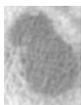
Les abcès par exemple présentent une variabilité importante, comme cela est illustré dans la Figure 6.3. Pour différencier ces deux classes des autres types de nodules, l'introduction de nouveaux descripteurs semble donc indispensable.

classe réelle \ étiquetée	Examens monophasiques							Examens multiphasiques						
	Ab	Ad	Ky	HN	An	CH	Mé	Ab	Ad	Ky	HN	An	CH	Mé
Abcès	2	0	1	1	1	1	0	1	0	0	1	0	0	3
Adénome	0	10	0	0	0	0	0	0	9	0	0	0	0	1
Kyste	0	1	23	0	0	0	1	0	0	24	0	0	0	1
HNF	0	0	0	0	2	0	4	0	0	0	1	0	1	4
Angiome	3	1	0	2	2	0	1	0	1	1	0	7	0	0
CHC	1	2	0	0	0	3	7	0	0	1	0	1	6	5
Métastase	4	2	3	1	2	2	24	3	1	1	3	0	5	25
Instances correctement classées	59,81%							68,22%						

TABLE 6.2 – Matrices de confusion et précisions totales de classification obtenue suite à une classification des lésions de notre Base 1 monophasique et multiphasique sur l'Ensemble 1 de descripteurs



La lésion est plus dense (sombre) que le parenchyme hépatique mais elle est hétérogène.



Le nodule a l'aspect d'un haricot très foncé, avec une séparation nette du tissu hépatique sain gris, et présente des zones blanches sur les contours de la lésion (dûes à l'inflammation).



La lésion est hétérogène, avec des zones hypodenses et d'autres hyperdenses par rapport au foie sain, mais ces dernières étant diffuses et non localisées sur les contours.

FIGURE 6.3 – *Illustration de la variabilité intrinsèque des abcès : 3 abcès de notre base de données acquis sur la même phase (portale)*

6.3 Comparaison avec une classification par des experts

A titre expérimental, nous avons souhaité soumettre deux radiologues ayant tous deux une dizaine d'années d'expérience dans la lecture d'examens scanner abdominaux à un test "à l'aveugle". Leur objectif était d'attribuer une classe à chaque lésion contenue dans notre Base 1 monophasique. Cependant, ils n'avaient accès qu'aux informations qui sont entrées dans notre outil informatique. Ils n'avaient donc à leur disposition aucun environnement. Ni clinique, car aucune information sur le patient, ses symptômes, les autres examens pratiqués n'était disponible. Ni visuel, en effet seule la zone d'intérêt 2D sur les 4 phases scanner était visible, sans le foie autour et sans pouvoir se déplacer entre les différentes coupes de l'examen.

Bien évidemment, cette expérience n'est pas réaliste d'un point de vue pratique clinique, cependant elle remplit deux objectifs. Tout d'abord, avoir une idée des contraintes de notre thématique de travail : les deux radiologues expérimentés, avec aussi peu d'informations à leur disposition, vont-ils pouvoir établir le type de lésions de manière fiable ? Deuxièmement, savoir si notre outil est efficace en comparaison avec ces tests à l'aveugle.

Deux lectures leur ont été demandées, l'une sur des examens monophasiques, l'autre sur des examens multiphasiques, à 2 mois d'intervalles afin qu'aucun effet mémoire ne vienne fausser les résultats. Toujours dans cette idée, les lésions n'étaient pas présentées dans le même ordre et leur système de numérotation était différent.

Ces résultats sont présentés dans le Tableau 6.3. Le but n'étant bien évidemment pas de réaliser une étude clinique entre les avis convergents ou divergents des experts, comme cela peut se voir dans des articles portant sur des thèmes médicaux, nous avons considéré qu'une lésion était "reconnue" lorsque les deux experts l'avaient correctement étiquetée.

Tout d'abord, les résultats obtenus sur les tests des radiologues montrent une amélioration lorsque les examens sont multiphasiques, par rapport à des examens mo-

CLASSE	#	Succès		Coeff. de Dice
		Experts	Tool	
Abcess	6	1	1	0
Adenoma	10	0	9	0
Cyst	25	25	24	0.958
FNH	6	1	1	0
Haemangioma	9	7	7	0.667
HCC	13	0	6	0
Metastasis	38	17	25	0.468
TOTAL	107	51	73	0.618

TABLE 6.3 – Comparaison des scores obtenus par les experts radiologues et notre outil, ainsi que le coefficient de Dice entre les deux

nophasiques. Cette observation va bien évidemment dans le sens de l'aspect clinique de la problématique.

Nous allons maintenant discuter des résultats classe par classe, principalement sur les examens multiphasiques.

Abcès

Un seul des 6 abcès de la base est reconnu correctement, à la fois par les experts et par notre outil, mais ce n'est pas le même dans les deux cas (le coefficient de Dice est nul). Les abcès sont difficiles à identifier en dehors du contexte clinique, ce qui explique le résultat des radiologues. Pour notre outil, nous l'avons dit, la variabilité intrinsèque de cette classe est un problème.

Adénomes

Seul 1 des 10 adénomes de la base a été identifié par les radiologues, mais ils sont très bien reconnus par notre système de caractérisation (9 sur 10). Les adénomes sont censés être aisés à reconnaître, mais dans un certain contexte clinique habituel, et de plus il semble que les adénomes présents dans notre base de données ne soient pas typiques de leur genre (remarque qui rejoint celle qui a été faite lors de la construction de la Base 2, d'où l'absence d'adénomes dans celle-ci).

Kystes

Ces lésions très répandues, et pour lesquelles l'aspect visuel ne varie pas au cours de l'évolution temporelle sur les phases, sont très bien identifiées, à la fois par les

radiologiques (100% de réussite) et par notre outil (de 92% à 96% de réussite), sur les acquisitions monophasiques et multiphasiques.

HNF

Les résultats pour cette classe sont catastrophiques, 1 HNF sur 6 seulement a été correctement identifié, à la fois pour les radiologues et par notre outil, avec un coefficient de Dice nul, ce qui signifie que la lésion reconnue par les praticiens n'est pas la même que celle reconnue par notre système. La caractéristique typique de ce type de lésion est une cicatrice centrale, présente chez 60 à 70% des patients, cependant celle-ci n'est distinguable qu'à partir d'une certaine taille de nodule, et elle se confond parfois avec le reste des tissus la lésion.

Angiomes

Les radiologues et notre outil obtiennent un score identique sur les acquisitions multiphasiques, nécessaires à la reconnaissance de ces lésions hypervasculaires : 7 lésions sur 9 sont reconnues. Cependant, seulement 4 de ces lésions identifiées sont communes.

CHC

Les résultats sont mauvais. Ces nodules sont habituellement détectés dans un contexte clinique suspect (cirrhose, surpoids) et avec l'aide d'autres analyses. Sans ces informations extérieures, les praticiens n'arrivent pas du tout à les reconnaître (aucun des 13 nodules n'a été identifié correctement). Notre outil est ici plus performant que les experts, qui n'ont pas pour habitude de détecter ces lésions en dehors du climat évoqué, puisque 6 des 13 CHC ont été identifiés sur les acquisitions multiphasiques.

Métastases

Le vecteur de confusion est très étalé, en effet l'aspect visuel des métastases est variable et elles peuvent ressembler à toutes les autres classes de lésions, ce qui fait de leur reconnaissance une tâche difficile. Les experts en ont reconnus à peu près la moitié (49%). De façon surprenante cependant, notre outil obtient un bon score sur cette classe (79% de reconnaissance). Cela est peut-être justement dû à la variabilité que nous avons évoqué, associée à l'effectif important de cette classe. Notre outil a peut-être tendance à mettre dans la classe des métastases toutes les lésions qu'il ne peut pas étiqueter par une autre, ce qui conduirait à un biais dans les résultats. Cette hypothèse est illustrée par le vecteur de confusion des lésions HNF de notre outil (4

lésions HNF sur 6 sont identifiées comme des métastases, à la fois sur les examens monophasiques et multiphasiques).

La première observation de cette expérimentation radiologues contre outil est que les résultats obtenus par notre outil sont plus qu'encourageants. La seconde est que pour certaines classes, d'autres descripteurs sont à extraire des images afin de permettre leur caractérisation (abcès, HNF et CHC).

6.4 Problème de la normalisation des intensités

Ces nouveaux échanges avec les radiologues nous ont permis, en plus de construire une seconde base de données, de réfléchir à une normalisation des intensités des lésions. Dans les travaux qui suivent, nous travaillerons désormais sur la Base 2.

Les unités Hounsfield sont censées représenter de façon exacte la densité des tissus. Cependant, une grande variabilité est observable entre les différents examens et patients. La figure 6.4 présente les courbes des moyennes de niveaux de gris en unités Hounsfield de chaque lésion au cours des phases de notre Base 2 multiphasique, regroupées par classes de lésions.

Le premier constat est que toutes les classes ont une dynamique globale assez homogène, y compris le foie sain. Les tissus se réhaussent entre les phases 1 et 2, puis 2 et 3, et enfin retournent vers leur état originel à partir de la phase 3 sans pour autant l'atteindre. Le kyste est nettement marqué de par ses intensités basses, mais les autres classes, y compris le foie sain, semblent être à des niveaux d'intensité proches. Une variation entre 20 et 60 Unités Hounsfield est présente, même entre les minima et maxima des moyennes de niveaux de gris du parenchyme hépatique des différents examens de notre base de données Base 2. En effet, le réhaussement dépend de la dose de produit de contraste et de la diffusion particulière qui se produit. L'absorption du produit de contraste varie énormément d'un individu à l'autre et en fonction des conditions d'injection ([Bae10]). Certaines fois, l'artériel ressemble plus à du portal ou inversement. Cette remarque de la variabilité des unités Hounsfield alors qu'on a une impression visuelle de proche s'applique également aux kystes. Globalement, les dynamiques sont similaires sauf chez les abcès de notre base dont la variabilité des schémas de réhaussement est très importante.

Nous avons souhaité atténuer les effets de cette variabilité et pouvoir effectuer une comparaison de la densité d'une lésion par rapport au parenchyme hépatique réel du patient au moment de l'acquisition. A cet effet, nous avons proposé une normalisation par le foie sain, dont le principe et les effets sont décrits ci-après.

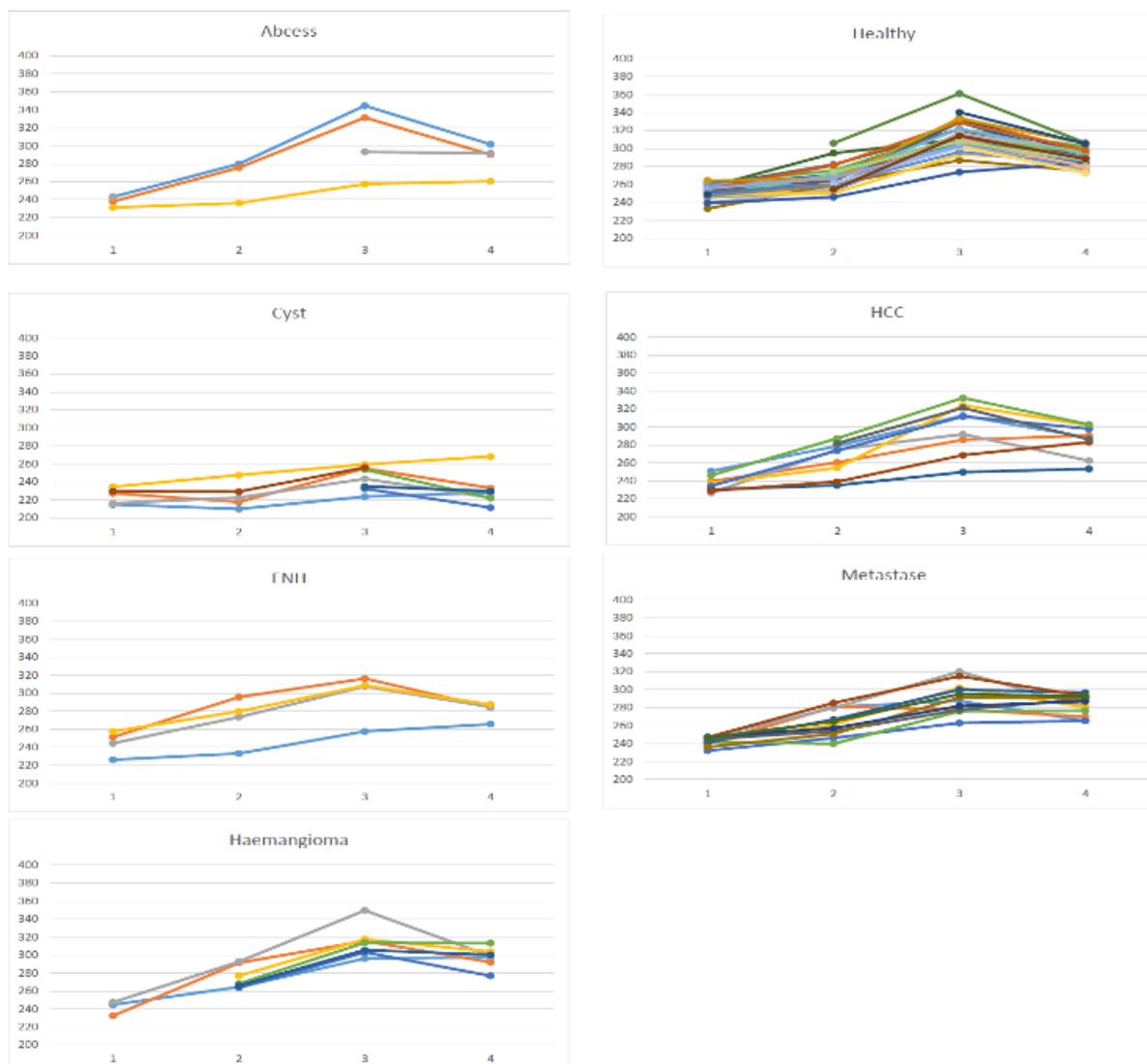


FIGURE 6.4 – Courbes des moyennes de niveaux de gris de chaque lésion de notre Base 2 sur les différentes phases regroupées par type de lésion (abscisse : phase, ordonnée : intensité)

6.4.1 Principe

Les radiologues se référant à la densité des tissus par rapport au parenchyme hépatique sain, nous avons proposé la normalisation de nos images de lésions par le foie sain.

Pour chaque lésion monophasique de la Base 2, une zone de foie sain a été sélectionnée. La normalisation des images de lésions s'effectue par la soustraction de la moyenne des niveaux de gris de la zone de foie sain. Ainsi, on a une valeur de référence commune sur tous les examens de la base, et les informations visuelles extraites le seront en fonction de cette référence "foie sain".

Afin d'illustrer cette transformation, un examen multiphasique de notre Base 2 est présenté dans la figure 6.5, à la fois dans sa version d'origine et dans sa version normalisée.

6.4.2 Influence sur la dynamique temporelle des intensités

La figure 6.6 présente l'évolution des moyennes de niveaux de gris normalisés au cours des phases des lésions de notre Base 2 multiphasique, regroupées par classes de lésions. Comme cela a été expliqué ci-dessus, la moyenne des niveaux de gris du foie sain est nulle pour chaque lésion et sur chaque phase. La dynamique de réhaussement est affectée par cette normalisation. Après consultation auprès d'un radiologue, les schémas après normalisation sont plus proches de la théorie cli-

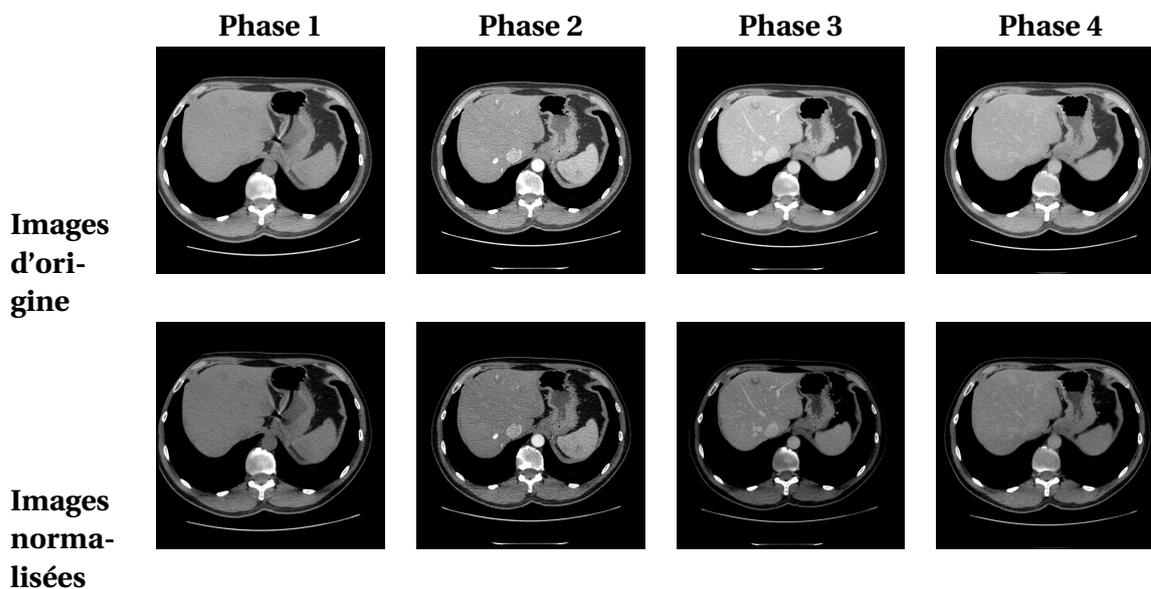


FIGURE 6.5 – Images d'origine et normalisées d'une coupe scanner du foie scanner

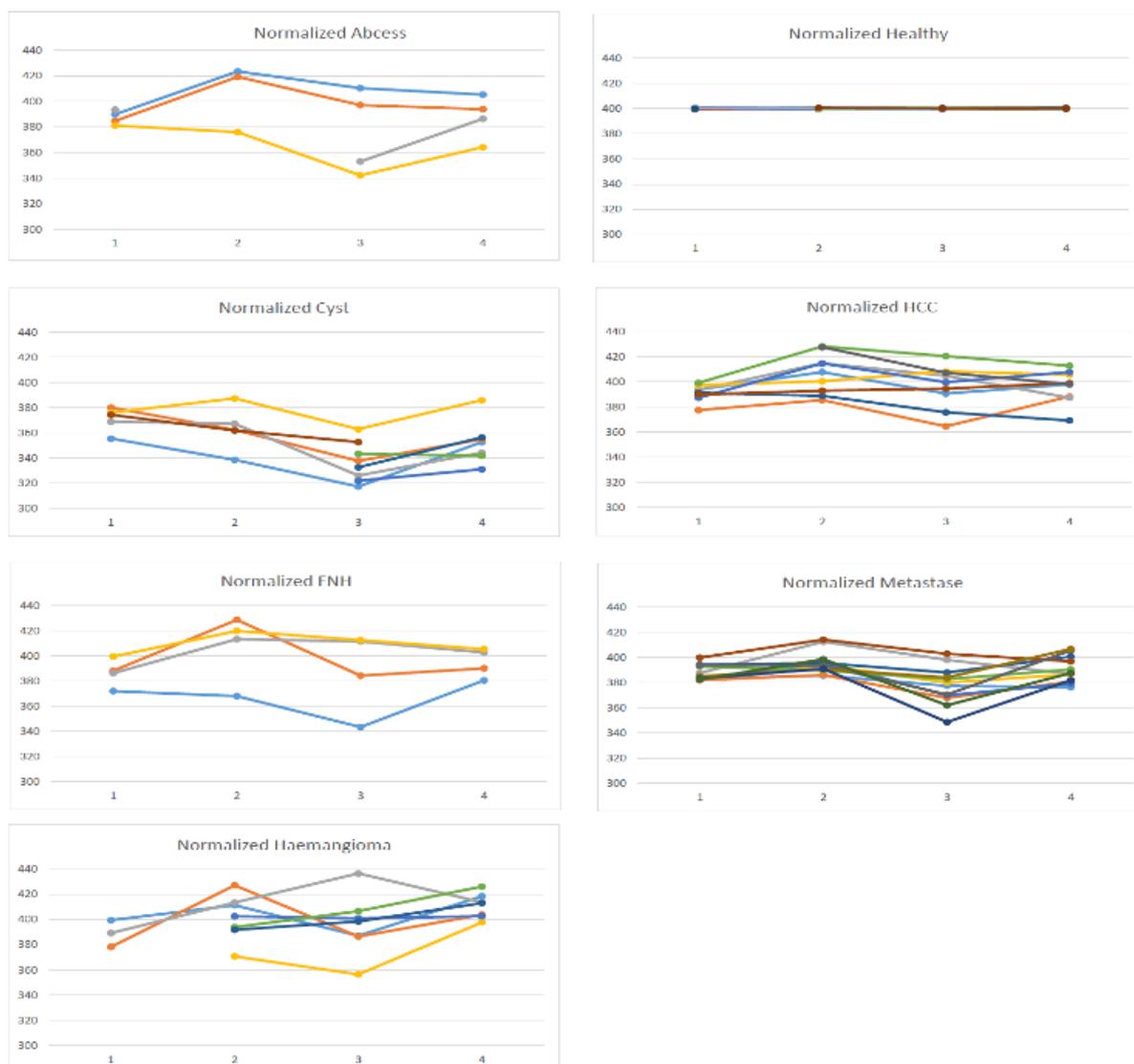


FIGURE 6.6 – Courbes des moyennes de niveaux de gris normalisés de chaque lésion de notre Base 2 sur les différentes phases regroupées par type de lésion (abscisse : phase, ordonnée : intensité)

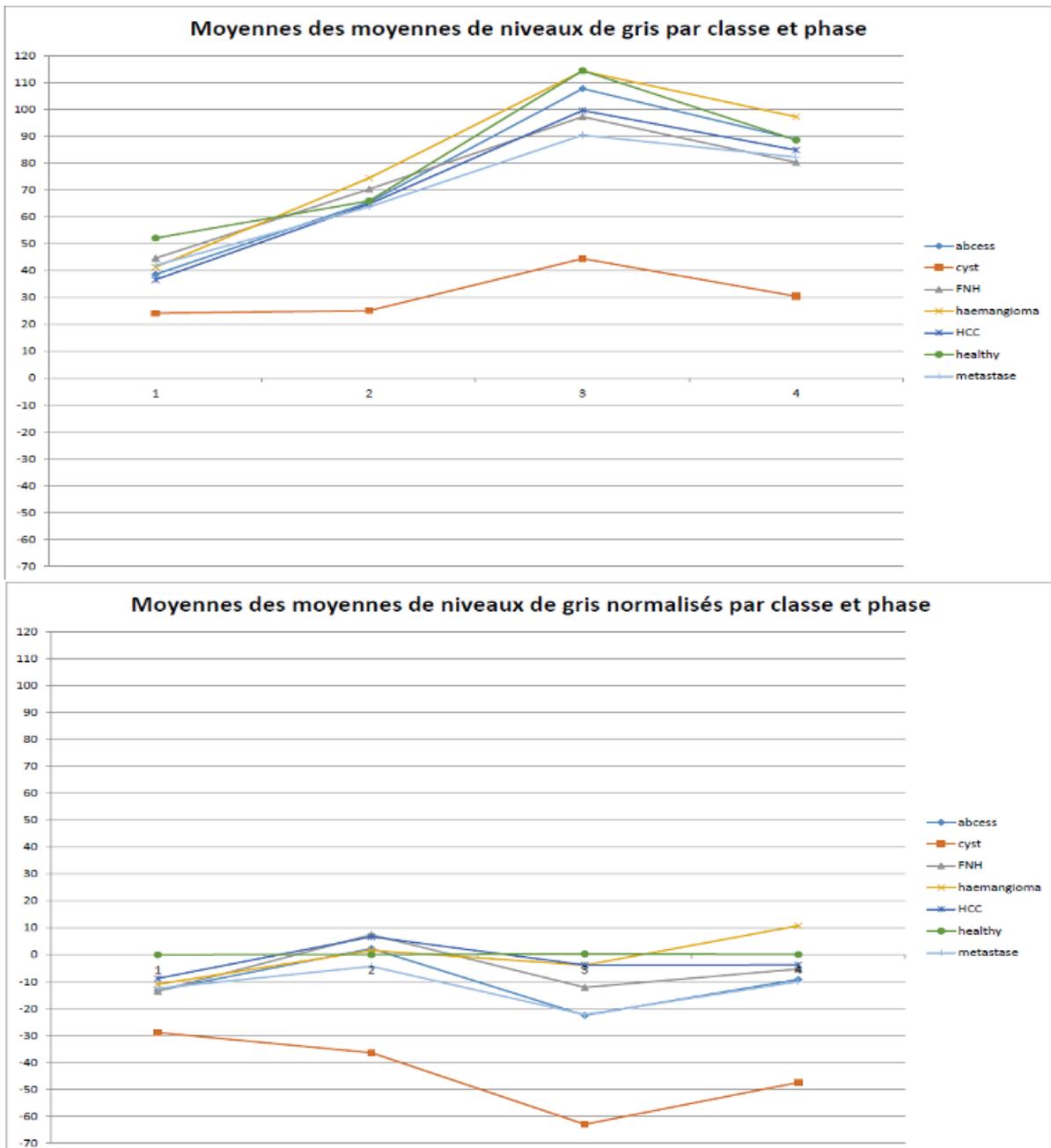


FIGURE 6.7 – Courbes des moyennes de niveaux de gris de chaque classe de lésion de notre Base 2 sur les différentes phases : non normalisées (en haut) et normalisées (en bas). En abscisse : la phase, en ordonnée : l'intensité

nique, ce qui nous l'espérons permettra l'extraction de descripteurs plus pertinents. Il y a moins de variations sur ces courbes normalisées. Globalement, les intensités montent avant de descendre et d'entamer une seconde montée entre les phases temporelles. Le kyste est à nouveau distinct des autres classes, à la fois par sa dynamique car l'intensité moyenne descend entre les phases 1 et 2, et par la valeur de l'intensité elle-même, qui est beaucoup plus basse que les autres types de nodules comme cela était le cas sur les images non normalisées. Une petite différence de dynamique sur le CHC est également à observer, entre les phases 3 et 4 la courbe de moyenne des intensités reste stable.

Du fait de cette visualisation, on s'aperçoit cependant de l'impossibilité de caractériser les différentes classes de lésions sur les seules moyennes de leurs niveaux de gris, puisqu'entre les classes une grande similarité peut être observée. Cette problématique est illustrée par la moyenne effectuée au sein de chaque classe dans la figure 6.7. Prenons le cas d'un angiome. Le réhaussement se caractérisera par des "mottes" périphériques (donc plus claires) mais un centre de la lésion moins dense. La moyenne des niveaux de gris ne permet donc pas de traduire cette information. Même lorsque les niveaux de gris sont normalisés, les centres des classes et les dynamiques sont trop proches pour pouvoir différencier les lésions sur ces critères. Des descripteurs spécifiques devront donc être envisagés par la suite ; ils le seront dans le chapitre suivant.

6.4.3 Influence sur la classification

L'analyse des dynamiques temporelles des niveaux de gris des lésions au cours des phases a montré que la normalisation par le foie sain telle que présentée précédemment permettait d'obtenir des courbes plus proches de ce que les radiologues considéraient en théorie pour les différents types de nodules. L'expérimentation qui suit va permettre de montrer que la normalisation a une influence positive sur les résultats de classification des lésions. La caractérisation est effectuée sur la Base 2 multiphasique normalisée avec l'Ensemble 1 de descripteurs. Les résultats sont regroupés dans le Tableau 6.4.

Globalement, le pourcentage d'instances correctement classées augmente de 6% grâce à la normalisation. En se penchant sur la matrice de confusion, il apparaît que ce sont les CHC et métastases qui en bénéficient (respectivement +33,3% et +12,5% d'augmentation), les autres scores restant stables. La normalisation n'a pas d'influence sur les classes qui posaient problème auparavant, à savoir les HNF et les abcès dont les taux de reconnaissance sont toujours très faibles.

classe réelle \ éti- quetée	Examens d'origine							Examens normalisés						
	Sa	An	Ab	Ky	HN	CH	Mé	Sa	An	Ab	Ky	HN	CH	Mé
Foie sain	36	0	0	0	0	0	1	36	0	1	0	0	0	0
Angiome	0	10	0	0	0	1	1	0	10	1	0	0	1	0
Abcès	0	2	1	0	0	1	2	0	3	1	0	0	0	2
Kyste	0	0	0	9	2	0	0	0	0	0	9	2	0	0
HNF	0	2	0	2	0	1	1	0	2	0	2	0	2	0
CHC	0	2	0	0	0	5	5	0	1	0	0	1	9	1
Métastase	0	0	2	0	0	5	9	0	0	2	0	0	3	11
Instances correc- tement classées	70%							76%						

TABLE 6.4 – Matrices de confusion et précisions totales de classification obtenue suite à une classification des lésions de notre Base 1 monophasique et multiphasique sur l'Ensemble 1 de descripteurs

6.5 Conclusion

Nous avons introduit dans ce chapitre des examens scanner multiphasiques en lieu et place des examens monophasiques sur lesquels nous travaillions précédemment. Les résultats montrent que travailler sur des examens scanner multiphasiques permet d'augmenter les performances de caractérisation des lésions hépatiques, ce qui correspond d'ailleurs à se rapprocher de l'aspect clinique de notre problématique, afin de travailler sur l'ensemble des informations disponibles au lieu de se restreindre aux indications extraites sur une phase unique. Le chapitre suivant présente nos travaux sur cette thématique, en détaillant de nouveaux descripteurs visuels, notamment des mesures prenant en compte l'évolution des intensités de la lésion au cours des phases temporelles.

Enfin, nous avons présenté une normalisation des images par le foie sain, qui influence de façon positive la classification multi-classe.

Intégration de nouveaux descripteurs visuels

Préambule

Après avoir construit un ensemble de descripteurs classiques de la littérature puis les avoir appliqués sur des acquisitions scanner multiphasiques, nous présentons un nouvel ensemble de descripteurs de texture plus performant, ainsi que nos différentes propositions de descripteurs spécifiques à l'application clinique de caractérisation des lésions focales du foie. Une analyse de chacun d'entre eux est effectuée, avant la présentation des résultats de classification obtenus sur notre base de données à l'aide de ces descripteurs.

Contents

7.1	Introduction	86
7.2	Un descripteur visuel plus rapide pour la texture	86
7.3	Des descripteurs visuels adaptés au multi-phase	91
7.4	Conclusion	99

7.1 Introduction

Ce chapitre présente plusieurs nouveaux descripteurs, selon différents axes : la diminution du temps de calcul de l'étape d'extraction des mesures de descripteurs, et la prise en compte des informations temporelles au cours des phases scanner. Leur principe sera décrit, et leurs performances de classification analysées.

Dans tout ce chapitre, nous travaillerons sur la Base 2 multiphasique normalisée à l'aide de la méthode décrite précédemment dans le Chapitre 6.

7.2 Un descripteur visuel plus rapide pour la texture

Cette section présente les différentes problématiques autour du calcul de la matrice de co-occurrence, ainsi que les solutions existantes pour lesquelles un comparatif sera effectué. Nous verrons que pour notre base de données les histogrammes d'Unser, sur lesquels sont extraits des descripteurs statistiques similaires à ceux proposés par Haralick sur la matrice de co-occurrence, sont idéals, tant en termes de temps de calcul qu'en termes d'influence sur les résultats de classification.

7.2.1 Temps de calcul

Pour l'ensemble des lésions de notre Base 2, le temps de calcul des descripteurs est de 120 minutes avec la parallélisation que nous avons mise en place dans le chapitre précédent.

Malgré le fait que l'étape de classification ne prenne que 0,05 seconde, il ne serait donc pas possible d'extraire les descripteurs puis de réaliser la classification en temps réel.

En regardant plus finement le temps de calcul, il apparaît que 99,2% est consacré aux seuls descripteurs d'Haralick.

Nous allons cerner les facteurs qui influencent le temps de calcul de la matrice de co-occurrence et/ou des descripteurs qui en sont extraits, avant de faire un bref état de l'art des solutions qui peuvent être mises en place, et enfin notre choix pour résoudre cette problématique, à savoir l'emploi d'histogrammes d'Unser en lieu et place de la matrice de co-occurrence.

7.2.2 Matrice de co-occurrence

Les méthodes d'extraction de caractéristiques basées sur la matrice de co-occurrence sont puissantes et sont employées à grande échelle depuis une quarantaine d'années. Des expériences sur la perception par les humains de la texture semblent justifier l'emploi de tels types de mesures. La matrice de co-occurrence

peut être utilisée telle quelle comme descripteur, mais souvent des mesures statistiques en sont extraites, comme les valeurs proposées par Haralick.

Les principaux inconvénients de cette méthode sont :

- le temps de calcul de la matrice de co-occurrence
- la place en mémoire requise pour stocker la matrice de co-occurrence si besoin est. En effet, elle peut être plus grande que l'image d'origine sur laquelle elle a été extraite. Pour une image de 256 niveaux de gris, cette matrice contiendra 256x256 valeurs, et ce quelle que soit la taille de l'image, même 16x16 pixels. Sur les images médicales, cette place est donc importante.
- le temps de calcul des descripteurs sur une matrice de taille importante (donc extraite sur une image dont l'intervalle de niveaux de gris est large). Sur les images médicales, ce problème se pose également.

7.2.3 Alternatives

Afin de pallier à cette lenteur, de nombreuses propositions ont été réalisées ces trente dernières années. Ces méthodes présentent deux approches distinctes, soit en travaillant sur le processus de calcul soit en tentant d'obtenir une approximation de la matrice.

Optimisation du temps de calcul de la matrice de co-occurrence et/ou des descripteurs

Cet ensemble de techniques n'a aucune incidence sur les valeurs des descripteurs extraits.

- Optimisation du calcul des descripteurs en exploitant la sparsité de la matrice de co-occurrence : elle ne contient en effet que peu de valeurs, regroupées autour de sa diagonale. Il est donc très intéressant de ne stocker que les valeurs non nulles de la matrice, et ainsi de ne faire les calculs que sur ces quelques valeurs au lieu de parcourir toute la matrice. Il existe plusieurs possibilités. L'une étant de stocker une liste d'indices de colonnes par ligne, chaque indice correspondant à une valeur non nulle de cette ligne de la matrice ([CJ96]). L'autre étant de stocker la matrice dans une table de hachage ([CZ01]).
- Optimisation du calcul des descripteurs : en observant les différents parcours de la matrice de co-occurrence nécessaires à l'extraction des descripteurs, nous avons mis en place dans notre logiciel une méthode très simple permettant de ne réaliser que trois boucles de parcours de la matrice de co-occurrence pour tous les calculer. Il s'agit de regrouper les descripteurs en fonction du parcours qui est nécessaire pour leur calcul.
- Parallélisme pour le calcul de la matrice et des descripteurs : les performances techniques des processeurs actuels permettent la mise en place simplifiée de trai-

tements parallèles [BTK05], voire de traitements parallèles sur les processeurs graphiques (Graphics Processing Units ou GPU) [HGMM08].

Approximation de la matrice de co-occurrence

Du fait de l'approximation, il y a une perte d'information. Ces techniques ont une influence sur les valeurs des descripteurs extraits, et donc sur la qualité des traitements (classification, segmentation etc.).

- Quantification des niveaux de gris : cette méthode définit des intervalles dans lesquels les niveaux de gris d'origine sont regroupés (par exemple pour passer de 512 à 256 niveaux de gris, ils sont regroupés deux par deux). Elle peut être uniforme (tous les intervalles sont de taille égale) ou adaptative (les plages de niveaux de gris les plus représentés dans l'image sont plus détaillées avec des pas de quantification plus fins).
- Réduction de la taille de l'image d'origine
- Histogrammes d'Unser : [Uns86] ils vont être développés ci-après

Histogrammes d'Unser

Approche théorique de la matrice de co-occurrence et des histogrammes d'Unser

La matrice de co-occurrence est l'estimation des fonctions de probabilité jointes de deux pixels de l'image dans une position relative donnée. Haralick insiste sur le fait que "texture and tone bear an inextricable relationship to one another" et que "one property can dominate the other at times". La propriété dominante d'un patch de petite taille et montrant peu de variations d'une image est le ton. Au contraire, si des variations importantes sont présentes, l'information dominante est cette fois textuelle. Lorsqu'il propose la matrice de co-occurrence, il s'appuie donc sur l'hypothèse que l'information de texture d'une image I est contenu dans la relation globale ou moyenne qu'entretiennent les tonalités de gris de l'image I entre elles.

L'idée principale des histogrammes d'Unser est de remplacer la matrice de co-occurrence par des estimations des fonctions de probabilité de premier ordre le long de ses deux principaux axes : ce sont les histogrammes de somme et de différence.

Dans les deux cas, les matrices ou histogrammes sont extraits selon 4 directions différentes (horizontale, verticale, et selon les deux diagonales) afin d'être robuste à la rotation. Ils peuvent être entrés tels quels comme descripteurs dans un système de classification, mais généralement plusieurs mesures statistiques sont calculées, classiquement 9 : la moyenne, la variance, l'énergie, la corrélation, l'entropie, le contraste, l'homogénéité, le cluster shade et le cluster prominence. Au total, 36 mesures sont donc extraites dans les deux cas.

Calcul de la matrice de co-occurrence et des histogrammes d'Unser

Afin d'avoir une idée de la façon dont les histogrammes d'Unser et la matrice de co-occurrence sont calculées, un exemple très simple de caractérisation de texture est présenté dans la Figure 7.1. Les valeurs des pixels vont de 0 à 6, le voisinage choisi est d'une distance horizontale d'un pixel vers la droite.

L'ordonnée de la matrice de co-occurrence correspond à la valeur du pixel courant, tandis que son abscisse représente la valeur du pixel voisin. L'exemple montre ce qui arrive pour les 3 pixels de valeur 1 et qui ont pour voisin un pixel de valeur 2. La case [1][2] de la matrice de co-occurrence contient un 3. Habituellement, on rend cette matrice symétrique, ce qui implique que la valeur est la même pour une configuration pixel - voisin et voisin - pixel, c'est pour cela que la case [2][1] contient également un 3.

Unser propose la construction de deux tables spatiales pour caractériser la somme et la différence. La table des sommes est très simplement remplie à l'aide de la somme de la valeur d'un pixel et de son voisin, et la même méthode est appliquée pour la table des différences. Ensuite les histogrammes de somme et de différence sont extraits de ces tables. En pratique, les histogrammes peuvent être directement calculés sur l'image de texture.

7.2.4 Expérimentations

Le Tableau 7.1 présente les résultats obtenus à l'aide des techniques d'approximation. Nous savons déjà que ces techniques sont beaucoup plus rapides, à la fois parce que le temps de calcul de la matrice elle-même est plus faible, mais également parce que l'extraction de statistiques prend moins de temps sur de plus petites matrices. Il nous reste maintenant à déterminer dans quelles proportions, et si leur utilisation est pertinente en termes de qualité de classification. Pour cela, plusieurs expériences ont été effectuées. En plus des temps de calcul, la précision de classification est donnée afin d'évaluer l'influence de l'approximation sur la qualité de la classification des lésions.

Les observations pouvant être faites sur ces résultats sont :

- la quantification permet de diminuer le temps de calcul par 9, mais à peu d'influence sur les performances de classification
- les histogrammes d'Unser permettent de diminuer le temps de calcul par 2308, et ont une influence positive sur les performances de classification (globalement +8%)

Ces résultats nous permettent de valider le remplacement de la matrice de co-occurrence par les histogrammes d'Unser.

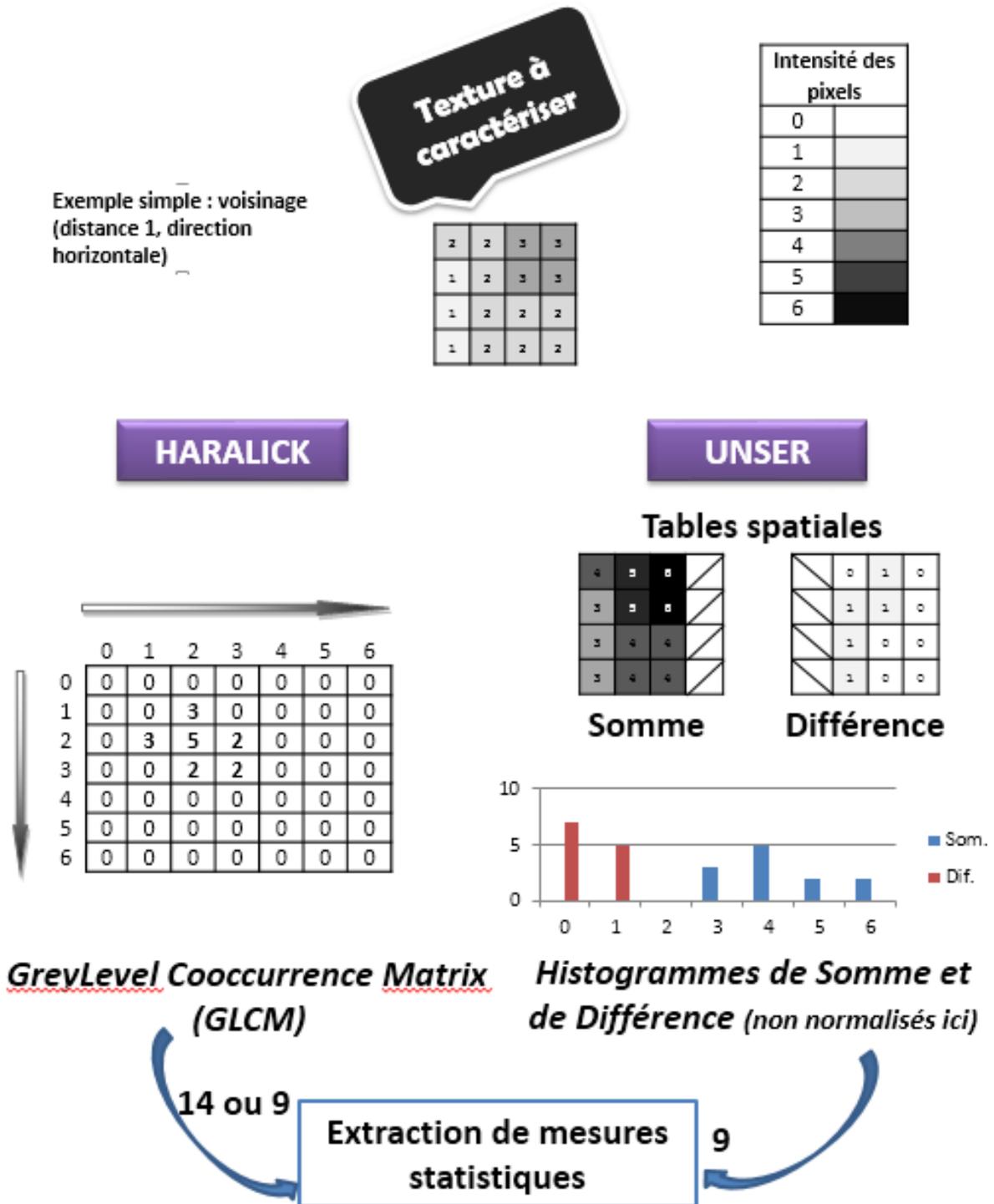


FIGURE 7.1 – Présentation des méthodes d’Haralick et Unser sur un exemple pratique

Méthode	Temps de calcul (avec parallélisation)	Vrais positifs
Classique	6693s	56%
Quantification (801 -> 400 niveaux de gris)	754s	57%
Histogrammes d'Unser	2,90s	64%

TABLE 7.1 – Comparaison des approximations de la matrice de co-occurrence : temps de calcul parallélisé et scores de classification obtenus sur l'ensemble de notre Base 2 multiphasique normalisée

7.2.5 Synthèse autour de la matrice de co-occurrence

Nous avons abordé les problématiques autour de la méthode classique d'extraction de descripteurs visuels sur la matrice de co-occurrence. Différentes méthodes permettant de réduire le temps de calcul de la matrice et des descripteurs ont été présentées et évaluées. Nos résultats démontrent que l'approximation de la matrice de co-occurrence par les histogrammes dits d'Unser donne d'excellents résultats sur notre base de données 2 multiphasique. En effet, le temps de calcul des histogrammes et des statistiques extraites est considérablement réduit (de 111 minutes à 3 secondes), et de plus les résultats de classification sont supérieurs de 8% à la méthode classique.

Le remplacement des mesures d'Haralick par celles extraites sur les histogrammes d'Unser dans l'ensemble des descripteurs "Ensemble 1" conduit à la dénomination "Ensemble 2", qui sera désormais l'ensemble de descripteurs standard.

Après avoir résolu cette problématique de temps de calcul, nous allons maintenant présenter des descripteurs spécifiquement conçus par rapport aux informations supplémentaires apportées par les examens multiphasiques.

7.3 Des descripteurs visuels adaptés au multi-phase

Les expérimentations menées précédemment, ainsi que les observations des radiologues, nous conduisent naturellement à l'exploitation des variations temporelles des niveaux de gris au cours des différentes phases de l'acquisition scanner.

Ces descripteurs n'apparaissent pas dans les travaux sur les lésions hépatiques présentés dans le Chapitre 4, en dehors de trois mesures. Nous allons tout d'abord expérimenter ces dernières, avant de détailler nos propositions, basées sur ce que nous appellerons les profils temporels des lésions, puis sur leurs carte de densité.

7.3.1 Descripteurs temporels issus des travaux existants

Ye et al. [YSW09] ont présenté trois mesures évaluant l'évolution des intensités des lésions hépatiques au cours du temps, qui ont été décrites dans le Chapitre 4. Ces valeurs sont issues du monde médical et habituellement calculées sur des images provenant d'acquisitions IRM.

Nous avons testé deux de ces mesures, la troisième n'étant pas accessible car il faut y intégrer le temps écoulé entre les phases, et nous n'avons pas ces informations à disposition. En effet, pour chaque examen, ce temps est différent : même s'il y a des temps plus ou moins standards, en fonction de la façon dont se déroule l'examen et du patient, mais également de ce que le radiologue veut visualiser, une certaine variabilité est observée. Dans l'article précité, un protocole unique et précis a été suivi pour toutes les acquisitions contenues dans leur base de données, ce qui n'est pas notre cas puisque nos examens sont issus de la routine clinique.

Ces mesures sont l'Intensité Relative du Signal (RSI) calculée entre chaque phase, soit 3 valeurs, et le ratio de réhaussement du signal (le lecteur pourra se référer à la Section 4.2.3 pour les équations et plus de détails).

La classification multi-classe conduite avec ce groupe de mesures uniquement aboutit à un résultat faible : 37% d'instances correctement classées, mais cette mesure seule ne permet pas de rendre compte du problème rencontré. La matrice de confusion, non reproduite, montre en effet que toutes les lésions ont été étiquetées par une classe unique (celle du foie sain, probablement car celle dont l'effectif est le plus important). L'algorithme de classification ne parvient donc pas à distinguer les classes de lésions sur la base de ces seules informations. Nous verrons plus loin dans ce chapitre si ces mesures temporelles, ajoutées à notre ensemble de descripteurs, permet d'améliorer la qualité de la classification ou non.

7.3.2 Profils temporels

Présentation

Nous avons tenté de caractériser la variation de la moyenne des niveaux de gris des pixels composant la lésion à l'aide d'un ensemble de symboles très simples : —, -, -, 0, +, ++, +++. Le 0 correspond à une stabilité entre les deux phases, les trois symboles contenant des "-" correspondent à une baisse de la moyenne des niveaux de gris, et les trois symboles contenant des "+" à une hausse de celle-ci. Bien entendu, il faut fixer un seuil afin de déterminer si une variation est petite ("+" ou "-"), moyenne ("++" ou "--"), ou importante ("+++ ou "—").

Le groupe de 3 descripteurs représente donc la variation de la moyenne des intensités entre chaque phase (phase 1 vers 2, 2 vers 3 et 3 vers 4).

Les seuils sont fixés de façon empirique, après plusieurs essais. Une variation est considérée petite lorsqu'elle est supérieure ou égale à 5 unités Hounsfield, moyenne lorsqu'elle est supérieure ou égale à 15, et importante lorsqu'elle est supérieure ou égale à 30.

Visualisation

La Figure 7.2 présente les profils temporels de chaque lésion de notre Base 2, regroupés par classe de nodule. Le foie sain n'est pas représenté car les variations sont nulles partout, sauf pour une image.

Une variabilité intra-classe importante est observée sur ces courbes. Chaque HNF de notre Base 2 a un profil temporel différent. Même les kystes, censés être relativement stables entre les phases, présentent des variations d'intensité et plusieurs profils différents. Les métastases semblent avoir trois profils principaux, de même dynamique mais translatée. Il est impossible de définir un profil unique par classe sur la base de ces courbes.

Analyse des résultats

Après avoir visualisé ces profils, il semble difficile de différencier les lésions sur la base de ces seules informations. Cependant, ils ne représentent que 3 valeurs nominales, c'est donc un tout petit ensemble de descripteurs. Nous avons tenté de réaliser une classification multi-classe pour confirmer cette intuition. Les résultats complets ne sont pas présentés : en effet, la matrice de confusion montre que quasiment toutes les lésions sont étiquetées comme saines, et le taux d'instances correctement classées est de seulement 28%. L'algorithme de classification est perdu entre les classes. Même les kystes, qui sont censés rester stables au cours du temps et bien différenciés des autres lésions car toujours hypodenses, ne sont pas correctement reconnus ici.

La variabilité intra-classe des données est trop importante pour que ce descripteur basé sur les moyennes des intensités fonctionne, ce qui était intuitif lors de leur analyse, et qui s'est révélé confirmé par les expérimentations sur nos données.

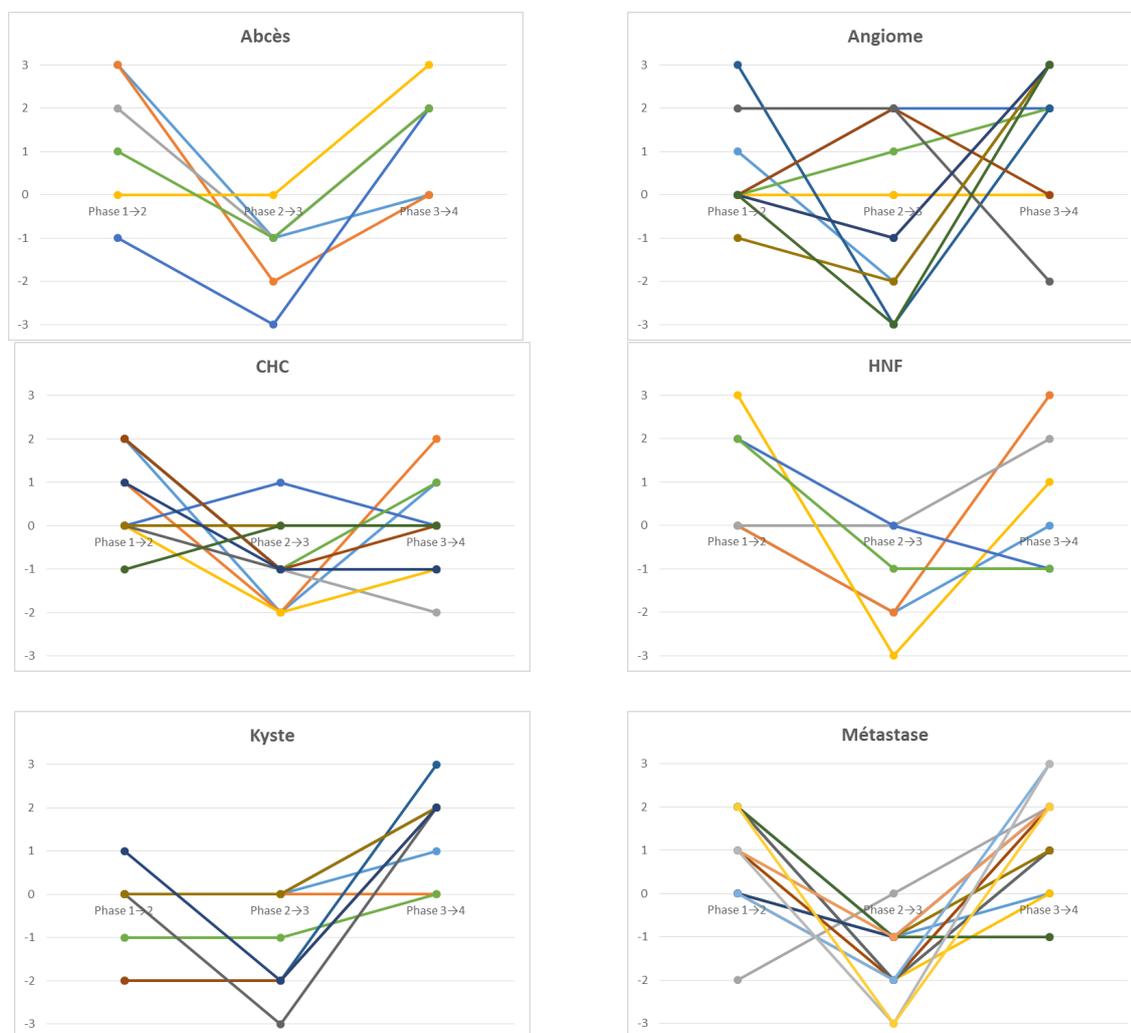


FIGURE 7.2 – Profils temporels des lésions par classe (abscisse : changement de phase, ordonnée : valeurs arbitraires représentant le changement d'intensité, de -3 (forte baisse) à +3 (forte hausse) en partant par 0 (stabilité))

7.3.3 Cartes de densité

Présentation

L'expérience précédente nous a montré que la moyenne des niveaux de gris, même évoluant au cours des phases, ne permet pas une bonne caractérisation des différentes classes de lésion.

La densité par rapport au foie sain est un élément important du processus clinique aboutissant à un diagnostic. Nous avons tenté de nous en approcher informatiquement parlant.

La caractérisation des lésions présentée ici est une carte représentant pour chaque pixel de la lésion sa densité par rapport au foie sain. Cette dernière est donc quadricolore, les couleurs n'ayant d'autre but qu'une visualisation immédiate des zones, leur valeur précise de pixel n'ayant pas une importance en soi.

- bleu : valeurs en dehors de l'intervalle des niveaux de gris considéré (après seuillage) ou à l'extérieur de la zone d'intérêt elliptique
- noir : tissus hypodenses (plus sombres que le foie sain)
- gris : tissus isodenses (proches du foie sain)
- blanc : tissus hyperdenses .

Les extrema de l'intervalle d'isodensité ont été déterminés en visualisant les effets de plusieurs seuils sur du foie sain, l'objectif étant de lisser le bruit afin d'obtenir une zone de foie sain homogène tout en conservant les variations qui seront intéressantes pour distinguer les lésions. Un exemple de l'effet de ces seuils sur les couleurs d'une zone de parenchyme hépatique est donné dans la figure 7.3. Le choix s'est porté après quelques essais sur l'intervalle [-20; +20].

	Phase 1	Phase 2	Phase 3	Phase 4
Intervalle [-10; +10]				
Intervalle [-20; +20]				
Intervalle [-25; +25]				
Intervalle [-30; +30]				

FIGURE 7.3 – Effets du seuil d'isodensité sur la carte de densité du parenchyme hépatique d'un examen multiphasique de notre base

Il est donc possible à partir de là d'étudier, non pas l'évolution de la moyenne des ni-

veaux de gris d'une lésion entre les phases, mais l'évolution de ses zones de densité, ce qui se rapproche plus de la pratique médicale.

Quelques exemples de ces cartes sont présentées dans la Figure 7.4.

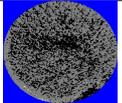
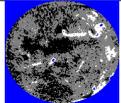
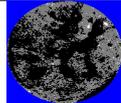
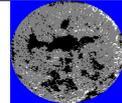
Lésion	Commentaires	Phase 1	Phase 2	Phase 3	Phase 4
Kyste	hypodense				
HNF	bi-zone : hypodense et isodense				
CHC	hétérogène sur toutes les phases				

FIGURE 7.4 – Exemples de cartes de densité sur notre Base 2

Descripteurs extraits des cartes de densité

Comme cela a été évoqué précédemment, ces cartes permettent de quantifier non pas l'évolution de la moyenne des niveaux de gris d'une lésion entre les phases, mais celle de la répartition de ses zones de densité. La Figure 7.5 présente donc les camemberts de la répartition moyenne entre les trois zones (iso, hypo et hyper, la zone hors intervalle ou hors ellipse n'étant absolument pas pertinente ici) de chaque classe de lésion sur toutes les phases.

On constate pour les kystes qu'ils sont reconnaissables sur toutes les phases. Quant aux abcès, il semble possible de les différencier sur les phases 3 et 4. Les HCC et les métastases paraissent se distinguer sur la phase 2. Enfin, une fois qu'il ne reste plus que le choix entre les angiomes et les FNH, ce choix semble pouvoir se faire sur la phase 4. Ces observations restent cependant à confirmer : l'algorithme de classification va-t-il pouvoir s'y appuyer ?

Analyse des résultats de classification

Les Tableaux 7.2 et 7.3 présentent les résultats de classification obtenus avec pour seul groupe de descripteurs les pourcentages des zones de densité sur les quatre phases. Sur chacune des phases, 3 mesures uniquement sont employées, soit 12 au total.

Les résultats montrent qu'avec cette méthode très simple aboutissant à 12 valeurs caractéristiques seulement, le résultat de caractérisation multi-classe global est très satisfaisant : 78% des lésions sont bien reconnues, contre 76% avec l'Ensemble 1 de descripteurs.

En analysant plus finement les résultats en fonction des classes de lésions, il apparaît cependant que pour les types de nodules posant problème précédemment (abcès



FIGURE 7.5 – Répartition des zones de densité par rapport au foie sain moyennées par classe de lésion et phase (gris clair : hyperdensité, gris moyen : isodensité, gris foncé : hypodensité)

et HNF), ces descripteurs ne parviennent pas à mieux les différencier. Pour le foie sain, les CHC et les métastases, une petite amélioration est observable. Par rapport à l'Ensemble 1 de descripteurs, la précision globale baisse cependant légèrement (de 0,75 à 0,71), et le nombre de faux positifs augmente (0,03 à 0,04).

Ces résultats sont excellents car, à partir d'un ensemble de descripteurs à l'effectif très réduit mais construits à partir d'observations cliniques, la qualité de la classification multi-classe est meilleure qu'avec le groupement important de descripteurs issus de la littérature testé précédemment.

classe réelle \ étiquetée	Sa	An	Ab	Ky	HN	CH	Mé
Foie sain	37	0	0	0	0	0	0
Angiome	0	9	0	2	0	1	0
Abcès	0	1	0	1	0	4	0
Kyste	0	0	0	9	0	1	1
HNF	0	2	0	2	0	2	0
CHC	0	1	1	0	0	10	0
Métastase	3	0	0	0	0	0	13
Instances correctement classifiées	78%						

TABLE 7.2 – Matrice de confusion obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les descripteurs de cartes de densité

	Abcès	Angiome	CHC	HNF	Kyste	Métastase	Sain	MOYENNE
Valeur prédictive positive	1.00	0.75	0.00	0.82	0.00	0.83	0.81	0.78
Valeur prédictive négative	0.05	0.04	0.01	0.06	0.00	0.09	0.01	0.04
Précision	0.93	0.69	0.00	0.64	0.00	0.56	0.93	0.71
F-score	0.96	0.72	0.00	0.72	0.00	0.67	0.87	0.74
Aire sous la courbe ROC	0.98	0.83	0.75	0.94	0.64	0.90	0.95	0.91

TABLE 7.3 – Mesures de la qualité de la qualification obtenues suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les descripteurs de cartes de densité

7.4 Conclusion

Dans ce chapitre nous avons présenté les différents descripteurs que nous proposons pour la caractérisation des lésions hépatiques.

Un premier ensemble de mesures de texture (histogrammes Unser), plus générales, remplace avantageusement les descripteurs classiques extraits de la matrice de co-occurrence, que ce soit en termes de temps ou de performances de classification.

Plusieurs pistes ont ensuite été explorées, basées sur le multiphase et les informations qu'il apporte. L'étude des variations temporelles des moyennes des niveaux de gris ayant donné des résultats insatisfaisants, les descripteurs proposés en second caractérisent la répartition des zones de densité des pixels des lésions par rapport au parenchyme hépatique sain et donnent de très bons résultats. Ils permettent de n'extraire en tout et pour tout que 12 mesures des images : ce sont les pourcentages de chaque type de densité par rapport au foie sain, et ce sur les 4 phases. La précision de caractérisation multi-classe de 0,71 est à comparer avec celle de l'ensemble complet de descripteurs contenant 312 valeurs qui est à 0,76. De plus, ils sont simples à mettre en place et leur temps de calcul est ridiculement faible.

Ces descripteurs étudiant les variations temporelles ou la densité par rapport à un fond peuvent se généraliser à d'autres applications dont les données sont composées de plusieurs images comme ici au cours du temps. On pense bien entendu au domaine médical : les acquisitions scanner et IRM multiphasiques. Les résultats obtenus à l'aide de nos cartes de densité sont plus que satisfaisants.

À partir de l'ensemble de descripteurs que nous avons proposé, le chapitre suivant va explorer des méthodes de pré-traitement des données, de sélection des attributs, ainsi qu'expérimenter plusieurs objectifs de caractérisation différents.

Expérimentations autour la caractérisation de lésions hépatiques

Préambule

Après avoir testé et analysé des descripteurs de la littérature et d'autres proposés par nos soins, nous allons ici exploiter différentes méthodes de traitement des données et expérimenter des approches de caractérisation des lésions hépatiques diverses. Nous allons donc présenter des techniques de normalisation, remplacement des valeurs manquantes et sélection d'un sous-ensemble d'attributs. Puis la reconnaissance d'une classe contre toutes les autres et la différenciation entre deux classes uniquement seront effectuées. La clinique ne sera pas oubliée, puisqu'il s'agit ici d'un thème lié à la pratique médicale : les reconnaissances sain / pathologique / bénin / malin seront proposées.

Contents

8.1	Introduction	102
8.2	Pré-traitement des données	102
8.3	Sélection d'un sous-ensemble d'attributs	103
8.4	Introduction d'une classe "inconnue"	106
8.5	Conclusion	109

8.1 Introduction

Une fois l'ensemble de descripteurs construit, il est désormais possible d'utiliser un certain nombre de techniques de traitement des données dans le but d'améliorer le résultat de classification, et également de caractériser les lésions selon plusieurs objectifs. En plus de la classification multi-classe qui a déjà été expérimentée dans les chapitres précédents, les distinctions une classe contre une classe, ainsi qu'une classe contre toutes les autres, seront réalisées. Ils seront tous décrits avant leurs résultats présentés et analysés.

Tous les travaux qui suivent ont été effectués sur la Base 2 multiphasique.

8.2 Pré-traitement des données

Dans le but d'améliorer les résultats de classification, deux étapes ont été ajoutées dans notre système de caractérisation. La première consiste au remplacement des valeurs manquantes, la seconde à la normalisation des valeurs. Pour cette expérimentation, c'est l'outil RapidMiner (Community Edition) ¹ qui a été employé.

Le remplacement des valeurs manquantes consiste à insérer la valeur moyenne de l'attribut en lieu et place de chacune. Elles sont dûes à l'absence d'une phase dans l'examen, comme cela est parfois le cas dans notre base de données (c'est donc l'ensemble des caractéristiques de la phase en question qui manque dans ce cas). Jusqu'ici, nous ne les avons tout simplement ignorées.

La normalisation effectuée est une normalisation statistique, parfois appelée standardisation. L'objectif est d'obtenir des données dont la distribution est normale, avec une moyenne nulle et une variance égale à 1. Les valeurs sont tout d'abord diminuées de la moyenne puis divisées par l'écart-type. Cette transformation permet une comparaison plus aisée des variables.

L'influence de ces deux manipulations des données sur la qualité de la classification est présentée dans le tableau 8.1 une évaluation Leave One Out. L'algorithme retenu est un arbre de décision dont le critère de séparation est le ratio de gain. En effet, il n'est pas possible d'entrer des données dont certaines valeurs sont manquantes dans les implémentations SVM disponibles dans RapidMiner. L'ensemble de descripteurs comporte toutes les valeurs précédemment employées. Les histogrammes, variables d'Unser, de Markov, de Law, les profils temporels, les mesures issues de l'IRM et les cartes de densité sont donc présentes pour un total de 571 valeurs. La Base 2 multiphasique est ici normalisée par le foie sain, tout comme dans les autres expérimentations.

1. www.rapidminer.com

Ces résultats montrent tout d'abord que la normalisation n'influence pas en soi la qualité de la classification, au contraire du remplacement des valeurs manquantes, qui permet un gain immédiat de précision de l'ordre de 9% dans le cas étudié.

Pré-traitement	Aucun	Rempl. val. manquantes	Normalisation	Rempl. et Norm.
Précision	0.60	0.69	0.60	0.69

TABLE 8.1 – Comparaison de la précision de classification obtenue sur la Base 2 multiphasique avec divers pré-traitements des données

8.3 Sélection d'un sous-ensemble d'attributs

8.3.1 Proposition d'un sous-ensemble de groupes de descripteurs

Nous allons tout d'abord analyser la précision de classification multi-classe sur notre Base 2 multiphasique normalisée à l'aide de chaque groupe de descripteurs, à la fois de l'Ensemble 1 (pour lequel on a remplacé les mesures d'Haralick par Unser selon les expérimentations menées dans la Section 7.2), et de ceux adaptés au multiphase et proposés dans la Section 7.3. Les résultats sont présentés dans le Tableau 8.2.

L'ensemble de descripteurs, nommé Set 3, contient le Set 2 additionné des mesures extraites des cartes de densité.

Dans ce qui suit, les groupes seront nommés ainsi : Histogramme pour les statistiques extraites sur l'histogramme des niveaux de gris, Markov pour celles extraites sur les GMRF, Law pour celles extraites après la méthode de filtrage de Law, Unser pour les statistiques extraites sur les histogrammes d'Unser, "Sélection" pour le sous-ensemble de descripteurs déterminé par l'algorithme de sélection évoqué, "Meilleurs groupes" et "Meilleurs individuels" pour les sous-ensembles respectifs de groupes et de descripteurs déterminés par seuillage sur les précisions de classification qu'ils obtiennent.

Globalement, la meilleure précision est obtenue par la combinaison de tous les descripteurs (0,76), suivie par le groupe de mesures extraits des cartes de densité (0,71). L'ensemble des tissus sains sont reconnus par Law et la combinaison de tous les descripteurs, suivis par les statistiques sur les histogrammes d'Unser et les cartes de densité (avec une précision respective de 0,95 et 0,93). Les angiomes sont le mieux reconnus par la combinaison de toutes les mesures (précision de 0,89). Les cartes de densité offrent une précision de 0,70 sur ce type de lésion, tandis que les histogrammes, les statistiques d'Unser et les mesures de Law obtiennent 0,67. Les abcès obtiennent malheureusement une précision toujours nulle, aucun groupe de descripteurs ou la combinaison de tous ne semblant en mesure de les distinguer parmi les autres nodules. Les kystes sont bien caractérisés par les histogrammes d'Unser (précision de

0,82), les cartes de densité et la combinaison de tous les descripteurs obtenant une précision quant à eux de 0,75. De façon étonnante cependant, car leur hypodensité stable les distingue bien de toutes les autres lésions, leurs scores sont moins bons que pour les métastases. Les HNF sont associés à une précision nulle pour chaque groupe de descripteurs, mais la combinaison de tous permet d'augmenter légèrement cette mesure à 0,14. Les CHC n'obtiennent une précision au-dessus de 0,5 qu'avec les mesures sur les cartes de densité. Les métastases ne sont correctement caractérisées que par les descripteurs issus des cartes de densité, ainsi que la combinaison de toutes les valeurs (précisions respectives de 0,93 et 0,94). Les autres groupes de descripteurs obtiennent une précision faible entre 0,2 et 0,4.

Nous avons tenté par des expérimentations exhaustives de construire un sous-ensemble des groupes de descripteurs obtenant une meilleure précision de classification que la combinaison de tous. Nous sommes parvenus à une précision globale équivalente, en enlevant les statistiques sur les histogrammes et les mesures temporelles issues des IRM. Nous appellerons ce nouvel ensemble de mesures "Ensemble 2". Les précisions globales et sur chaque classe sont présentées dans le Tableau 8.3. Il est à noter que l'ajout de 280 valeurs aux descripteurs extraits des cartes de densité augmente la précision globale de 6% seulement.

Les résultats obtenus par les deux ensembles sur le foie sain, les angiomes, les HNF, les kystes et les métastases sont équivalents. Par contre, l'Ensemble 2 permet une meilleure caractérisation des abcès (précision de 0,17 contre 0,00), tandis que l'Ensemble 1 différencie mieux les CHC (précision de 0,33 contre 0,25).

Cette sélection manuelle d'un sous-ensemble de 292 attributs nous fait prendre conscience du besoin d'opérer une réelle sélection sur le nombre élevé d'attributs entrés dans notre algorithme de classification. C'est ce que nous allons proposer ci-après.

Descripteurs	Sain	Angiome	Abcès	Kyste	HNF	CHC	Métastase	Global
Histogramme	0.47	0.67	0.00	0.60	0.00	0.00	0.40	0.38
Unser	0.83	0.67	0.00	0.75	0.00	0.42	0.53	0.61
Markov	0.37	0.00	0.00	0.00	0.00	0.00	0.20	0.17
Law	1.00	0.67	0.00	0.64	0.00	0.00	0.33	0.57
IRM	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.14
Profils	0.32	0.14	0.00	0.00	0.00	0.00	0.00	0.36
Cartes densité	0.93	0.70	0.00	0.64	0.00	0.56	0.93	0.71
Tous	1.00	0.89	0.00	0.75	0.14	0.33	0.94	0.76

TABLE 8.2 – Précision de classification obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les différents groupes de descripteurs présentés

Descripteurs	Sain	Angiome	Abcès	Kyste	HNF	CHC	Métastase	Global
Ensemble 2	1.00	0.88	0.17	0.75	0.13	0.25	0.94	0.77
Tous	1.00	0.89	0.00	0.75	0.14	0.33	0.94	0.76

TABLE 8.3 – Précision de classification obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les différents groupes de descripteurs

Base	Base 2	Base 2 décorrélée
Nombre d'attributs	571	211
Temps de calcul	16s	8s
Précision	0.76	0.78

TABLE 8.4 – Comparaison de la précision de classification obtenue sur la Base 2 multiphasique en fonction du nombre d'attributs

8.3.2 Sélection d'un sous-ensemble d'attributs à partir de la matrice de corrélation

La matrice de corrélation regroupe les coefficients de corrélation des différentes variables présentes dans l'ensemble d'attributs. Étant donné le nombre d'attributs en jeu ici, cette matrice est gigantesque, c'est pourquoi elle n'est pas présentée ici. En la parcourant, on constate une forte corrélation entre un certain nombre d'attributs. Cette corrélation conduit à une redondance dans les données. Il serait donc possible de supprimer certaines mesures, sans pour autant perdre globalement beaucoup d'information.

Les résultats obtenus par une telle suppression, conduite manuellement, sont présentés dans le Tableau 8.4. Le seuil retenu est une corrélation de 0,95. Le SVM utilisé possède un produit scalaire comme noyau, les données sont normalisées et les valeurs manquantes remplacées par la moyenne de l'attribut.

Il apparaît tout d'abord que le nombre d'attributs a grandement baissé, passant de 571 à 211 mesures, et ce pour une performance de classification très légèrement supérieure (+2% sur la précision). Le second gain est celui du temps de calcul qui diminue de moitié (de 16 à 8 secondes sur un processeur Intel core i7).

8.3.3 Synthèse autour de la sélection

Les deux expérimentations présentées ici, visant à opérer une sélection d'attributs, montrent qu'il est possible d'obtenir une qualité de classification équivalente, voire légèrement supérieure, tout en ne travaillant que sur un sous-ensemble d'attributs dont le nombre a drastiquement baissé.

8.4 Introduction d'une classe "inconnue"

L'expérimentation présentée ici consiste à ne plus obliger l'algorithme à choisir entre les classes présentes par l'introduction d'une nouvelle classe. Si l'instance courante apparaît éloignée de toutes les classes connues par rapport au modèle construit lors de l'apprentissage, celle-ci sera étiquetée "inconnue". L'objectif est d'identifier les lésions qui posent problème.

Afin de réaliser cette expérience, on décide de corriger les étiquettes obtenues par l'algorithme de classification lorsque l'indice de confiance de la classe est trop faible ou qu'il ne permet pas faire un choix entre deux classes. L'algorithme de classification est l'arbre de décision décrit précédemment, car sur les résultats obtenus sur notre Base 2 à l'aide du SVM l'indice de confiance est toujours élevé et ne permet pas de placer des instances dans une classe à part.

Voici les changements apportés par l'ajout de la classe "inconnue".

- **Abcès** : 4 abcès sur les 6 sont correctement reconnus, le cinquième étant identifié comme un CHC. Le dernier est étiqueté comme du foie sain par le premier passage, puis placé dans la classe inconnue lors du second.
- **Kyste** : 5 kystes sur 11 sont correctement reconnus, le sixième étant identifié comme une métastase. Les 5 kystes restants sont étiquetés comme du foie sain par le premier passage, puis placés dans la classe inconnue lors du second.
- **HNF** : 4 HNF sur 6 sont correctement reconnus. Les 2 HNF restants sont étiquetés comme du foie sain par le premier passage, puis placés dans la classe inconnue lors du second.
- **Angiome** : 3 angiomes sur 12 sont correctement reconnus. Les 9 angiomes restants sont placés dans la classe inconnue lors du second passage. Ils étaient étiquetés comme du foie sain dans 8 cas et comme un CHC dans le dernier cas.
- **CHC** : 10 lésions de type CHC sur 12 sont correctement reconnues. Les 2 restants sont placés dans la classe inconnue lors du second passage. Ils étaient respectivement étiquetés comme du foie sain et comme un CHC.
- **Foie sain et métastases** : ces deux classes sont correctement reconnues dans 100% des cas.

D'après ces résultats, 18 lésions sur 100 sont désormais classées "inconnues", ce qui est relativement élevé. Les informations détaillées permettent de conclure que beaucoup de lésions étaient placées dans la classe "foie sain" lorsque la lésion était à une distance élevée par rapport aux classes existantes dans le modèle généré. Ceci s'explique probablement par le fait que la classe "foie sain" contient le plus grand nombre d'instances (37 sur un total de 100). Seules trois lésions initialement étiquetées comme une métastase et deux CHCs se voient ensuite affectées à la classe "inconnue". Enfin, une seule lésion, correctement étiquetée au départ, se retrouve déplacée dans la classe "inconnue".

Cette expérimentation permet de réaliser qu'un certain nombre de lésions sont mal étiquetées par l'algorithme car trop distantes du modèle de classes construit lors de l'apprentissage, et de ce fait placées dans la classe dominante (le foie sain). La construction du modèle reste donc encore à améliorer pour notre base de données.

8.4.1 Base de données 3 sans abcès ni HNF

Etant donné les problèmes rencontrés pour reconnaître les deux classes que sont les abcès et les HNF, nous avons réalisé une classification multi-classe sur les six autres types de lésions uniquement. Les précisions de classification sont présentées dans le Tableau 8.5.

Toutes les classes, sauf les angiomes dont la précision diminue légèrement, bénéficient de cette suppression, notamment les CHC pour lesquels la précision de classification augmente de façon spectaculaire de 0,25 à 0,71. La précision totale est supérieure de 13% sur cette nouvelle base.

Classification	Sain	Angiome	Kyste	CHC	Métastase	Global
6 classes	1.00	0.82	0.92	0.71	1	0.90
8 classes	1.00	0.88	0.75	0.25	0.94	0.77

TABLE 8.5 – Comparaison des résultats obtenus par l'Ensemble 2 de descripteurs sur la Base 2 multiphasique normalisée et la Base 2 sans abcès ni HNF

8.4.2 Reconnaissance d'une classe

Il s'agit d'une différenciation binaire : l'objectif est de déterminer si une lésion est d'un certain type (par exemple un angiome) ou non (le "non" contenant toutes les autres classes de lésions). Cette reconnaissance est effectuée pour chacune des classes de lésions.

Les résultats proposés dans le Tableau 8.6 montrent les précisions de classification obtenues pour chaque type de nodule. Les résultats sur le parenchyme hépatique, les abcès, les HNF et les métastases sont équivalents à ceux obtenus en multi-classe. Les angiomes sont moins bien reconnus par rapport à toutes les autres lésions qu'en caractérisation multi-classe (baisse de précision de 0,88 à 0,75). Par contre, les kystes sont mieux reconnus (hausse de précision de 0,67 à 0,75).

Classe	Précision		
	cette classe	classe "autre"	globale
Abcès	0,17	0,95	0,90
Angiome	0,75	0,99	0,96
CHC	0,42	0,92	0,86
HNF	0,14	0,95	0,90
Kyste	0,67	0,95	0,91
Métastase	0,94	1,00	0,99
Foie sain	1,00	0,98	0,99

TABLE 8.6 – Précisions de classification obtenues suite à une classification "une classe contre toutes les autres" sur notre Base 2 multiphasique normalisée

8.4.3 Caractérisation clinique sain - pathologique - cancéreux

Cette expérimentation vise à se rapprocher de deux objectifs cliniques. Le premier est de différencier les tissus sains et pathologiques. Lorsqu'il s'agit d'une lésion et non du parenchyme hépatique, le but est de déterminer si le nodule est cancéreux ou non. Les CHC et les métastases sont étiquetés comme malins du fait de leur nature cancéreuse, les kystes, angiomes, abcès et le HNF comme pathologiques.

Les précisions de classification obtenues sont de :

- sain : 1,00
- cancéreux : 0,77
- non cancéreux : 0,78
- global : 0,86

Ces scores sont plutôt satisfaisants, même s'ils restent à améliorer pour une application réellement clinique.

Les valeurs importantes du point de vue clinique sont les faux positifs, c'est-à-dire les lésions qui sont étiquetées comme du parenchyme hépatique sain, ainsi que les lésions malignes reconnues comme bénignes. En effet, dans la pratique médicale, mieux vaut effectuer des tests supplémentaires si suspicion il y a, que passer à côté d'une lésion, voire d'une lésion cancéreuse. Il y a seulement 7 faux positifs sur toutes les instances, qui nous le rappelons, sont au nombre de 100. L'ensemble de ces faux positifs sont des lésions cancéreuses étiquetées comme pathologiques non cancéreuses.

8.5 Conclusion

Nous avons dans ce chapitre proposé des traitements des données permettant d'améliorer nos résultats de classification, avant de décrire deux sous-ensembles d'attributs, l'un obtenu manuellement, le second automatiquement. Enfin, différents objectifs de classification ont été explorés.

La normalisation et le remplacement des valeurs manquantes permettent d'améliorer la précision de classification.

Deux méthodes de sélection ont ensuite été détaillées. Suite à une analyse de tous les groupes de descripteurs, nous avons proposé un premier sous-ensemble d'attributs, contenant 292 valeurs, et constitué des caractéristiques suivantes : histogrammes d'Unser, champs de Markov, descripteurs de Law, profils temporels et cartes de densité. Ce sous-ensemble permet d'obtenir une précision de classification multi-classe équivalente à l'ensemble de départ. Une méthode complètement automatique nous a ensuite permis d'extraire 211 attributs permettant d'obtenir un score de classification légèrement plus élevé qu'avec notre ensemble de départ.

Enfin, d'autres caractérisations que la classification multi-classe réalisée jusqu'ici ont été expérimentées : reconnaissance d'une classe, classification sain / pathologique / cancéreux en lien avec l'aspect clinique de notre problématique.

La distinction binaire est beaucoup plus efficace, comme on pouvait s'y attendre, que celle multi-classe, que ce soit pour distinguer un type de lésion d'un autre, que pour différencier un nodule de tous les autres. Les scores sont excellents, et les abcès et les HNF sont enfin reconnus de façon satisfaisante dans ces expérimentations.

De plus, nous avons testé la classification multi-classe sur une base diminuée des abcès et des HNF. En effet, ce sont deux classes très mal reconnues malgré nos efforts. L'amélioration des résultats de classification est sensible.

Conclusion et Perspectives

Préambule

Nous présentons ici une synthèse de ce manuscrit, en décrivant tout d'abord nos contributions et les difficultés rencontrées au cours de la thèse, puis d'exposer les limitations de ces travaux, pour lesquelles nous présentons des perspectives.

Contents

9.1	Résumé de nos contributions	112
9.2	Travaux futurs	114

9.1 Résumé de nos contributions

L'évolution des techniques d'acquisition des imageries médicales et leur importance de plus en plus grande dans la prise en charge du patient (diagnostic, préparation d'intervention, suivi etc...), font émerger de nouveaux besoins autour du traitement informatique des images. La reconnaissance du type de lésions hépatiques est un grand enjeu, notamment car le cancer du foie, létal et très répandu, est souvent diagnostiqué trop tard pour sauver le patient. C'est dans ce cadre qu'est né le projet de recherche de ce manuscrit, fruit d'une collaboration entre la société IMAIOS, spécialisée dans l'imagerie médicale, et le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM).

Nous avons présenté nos travaux portant sur les descripteurs visuels extraits sur des images scanner abdominales multiphasiques. **Dans le cadre de cette thèse a été construit un système complet et automatique permettant, à partir d'images de lésions au format médical DICOM, d'extraire des descripteurs visuels de divers nodules hépatiques puis de les différencier à l'aide de ces derniers.** Il est à noter que contrairement aux travaux de la littérature présentés dans la première partie de ce manuscrit, **nos contributions portent sur des lésions multiphasiques et d'un nombre de classes différentes élevé (7 au total).**

Les contributions décrites s'articulent autour de divers axes, que nous présentons ci-dessous : normalisation des niveaux de gris des images, descripteurs visuels (s'appuyant notamment sur les informations temporelles ou de densité des tissus), applications diverses autour du même objectif.

Normalisation des densités : Une procédure de normalisation par le foie sain a été mise au point afin de réduire la variabilité sur l'échelle des niveaux de gris entre les examens médicaux. Son influence positive sur les dynamiques de réhaussement des lésions au cours des phases temporelles a été validée par un radiologue.

Descripteurs visuels : Divers caractéristiques ont été proposées.

- Notre principale contribution, les cartes de densité des tissus par rapport au foie sain, permettent d'obtenir un très petit ensemble de mesures numériques (12 seulement) rapides à calculer, et efficaces pour différencier les lésions.
- Le remplacement d'un ensemble de mesures statistiques de texture, nommées descripteurs d'Haralick [Har79], mais nécessitant des ressources et un temps de calcul important, par une approximation proposée par Unser [Uns86] a été réalisé avec succès, puisque en plus d'éliminer les problèmes de complexité évoqués, les résultats de classification s'en trouvent améliorés.

- Une analyse des descripteurs, classiques de la littérature comme de ceux que nous avons proposé, a permis d’aboutir à un ensemble de descripteurs plus petits pour une performance supérieure.

Traitements des données : La normalisation statistiques et le remplissage des valeurs manquantes ont permis d’améliorer nos résultats de classification. Deux approches différentes ont été expérimentées afin de sélectionner un sous-ensemble d’attributs pour lequel le résultat de classification obtenu est équivalent voire supérieur à l’ensemble de départ.

Caractérisation des lésions hépatiques : A partir des descripteurs évoqués ci-dessus, diverses applications formant un ensemble assez complet ont été proposées : classification multi-classe, différenciation d’une classe par rapport à toutes les autres, différenciation sain / pathologique non cancéreux / cancéreux.

Cette thèse portant sur un thème assez pratique et en lien avec un autre domaine que l’informatique, les difficultés inhérentes à ce type de projet ont été rencontrées. Tout d’abord la construction d’une base de données, en collaboration avec l’Hôpital Lapeyronie de Montpellier, qui a pris beaucoup de temps et qui reste encore à étoffer. Cette base en elle-même n’est pas idéale du point de vue traitement d’images, mais elle s’en rapproche du point de vue de la pratique clinique. Les lésions ne sont pas toutes artificiellement typiques, mais comportent une part de variabilité, et l’ensemble proposé est le plus proche possible de ce que rencontrent les radiologues dans leur routine clinique. De par ce choix elle comporte de plus un nombre de classes important, qui introduit une difficulté supplémentaire dans la caractérisation. Ensuite, le dialogue entre un informaticien et des radiologues n’est pas aisé puisque le langage parlé n’est pas le même. Les informations cliniques essentielles pour pouvoir réaliser ces travaux ne sont pas intuitivement compréhensibles et les articles et autres documents médicaux sur le sujet sont souvent complexes pour un néophyte. Il a fallu prendre le temps d’établir des récapitulatifs de différents éléments (les caractéristiques visuelles des lésions, leur dynamique de réhaussement etc...) Enfin, les examens scanner, du fait du positionnement du patient ou de la façon dont le produit de contraste s’est diffusé, et les lésions hépatiques présentent une variabilité très importante (forme, texture, niveaux de gris, densité par rapport au foie sain), comme nous avons pu l’illustrer dans ce manuscrit.

Nos travaux présentent certaines limitations. Tout d’abord, la base de données, trop petite, induit une difficulté supplémentaire pour la caractérisation qui a tendance à se perdre, faute d’éléments suffisants dans chaque classe. Malgré les efforts déployés, certaines classes de lésions ne sont pas reconnues de façon satisfaisante (HNF et abcès, qui contenaient peu d’éléments et dont les caractéristiques sont très variables d’une lésion à une autre).

9.2 Travaux futurs

Plusieurs perspectives nous semblent intéressantes à mettre en place dans les travaux futurs. Tout d'abord, un agrandissement de la base de données semble indispensable, notamment sur certaines classes de lésions dont l'effectif est actuellement faible. Nous allons maintenant détailler quelques axes possibles de développement.

Descripteurs visuels s'appuyant sur les informations cliniques : se rapprocher de la pratique clinique, comme nous l'avons fait avec les cartes de densité, permettra sans aucun doute de proposer des descripteurs visuels plus pertinents, notamment pour différencier les deux classes qui posaient problème dans ce manuscrit. Par exemple, à partir d'une certaine taille une cicatrice centrale est visible sur les HNF, caractéristique de cette classe (voir Figure 9.1. La détecter pourrait aider à la différenciation de cette lésion. Les cartes de densité elles-mêmes pourraient permettre l'extraction d'autres mesures que celles de pourcentages que nous avons proposées dans ce manuscrit. Une autre piste serait de toutes les redimensionner à la même taille et de se servir de la distance entre ces images comme descripteur (à l'aide d'une mesure de comparaison directe comme celle d'information mutuelle).

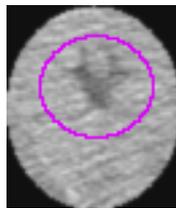


FIGURE 9.1 – Cicatrice centrale d'une lésion HNF

Descripteurs exploitant la temporalité des données : Nous avons tout juste commencé à explorer les informations apportées par les examens multiphasiques. Il serait notamment intéressant de développer des "descripteurs de texture temporels", permettant de rendre compte de l'évolution de la texture au cours des phases de l'examen scanner.

Descripteurs extraits sur des images radiales : Étant donné que les nodules ont des formes approchant des cercles ou des ellipses, une étude approfondie des "images radiales" (obtenues par une transformée en coordonnées polaires) devrait permettre l'extraction de descripteurs améliorant la caractérisation des lésions. Quelques exemples de ces images radiales sont présentées dans la Figure 9.2. Une piste à explorer serait par exemple l'étude de l'homogénéité de zones cibles de la lésion que sont son contour (plus ou moins bien délimité) et son centre.

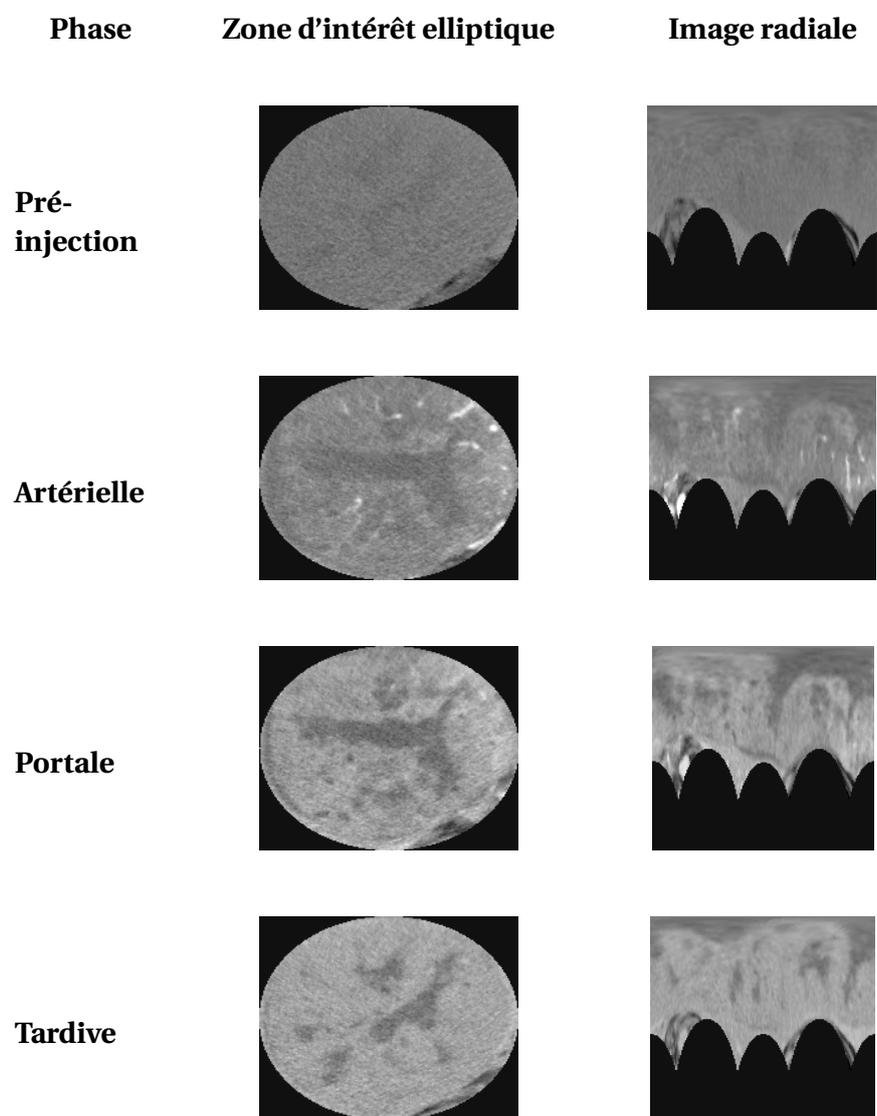


FIGURE 9.2 – Illustration de la transformée en coordonnées polaires sur une lésion

Apprentissage actif : L'apprentissage actif consiste à ajouter l'humain dans la boucle. En donnant son avis, ce dernier permet au système de s'améliorer petit à petit.

Multi-échelle : le passage au multi-échelle en incluant une zone de foie autour de la lésion permettrait probablement une amélioration intéressante des résultats de classification. Cette approche introduit le problème de la segmentation du foie afin d'éviter l'air, les artères et autres éléments indésirables, mais certains algorithmes de segmentation du foie existent déjà (notamment suite au challenge MICCAI "Segmentation of the Liver 2007"¹) et semblent donner des résultats plutôt satisfaisants.

Caractérisation : Au cours de cette thèse, la méthode de classification elle-même n'a pas été analysée, réaliser une étude comparative des résultats obtenus avec différents algorithmes, et l'emploi de méthodes de méta-classification, permettrait sans aucun doute de les améliorer.

Sémantique : Certains articles étudiant la classification ou la recherche par le contenu proposent, en plus des descripteurs visuels, des descripteurs sémantiques, c'est-à-dire des mots-clés ou des expressions décrivant le contenu de l'image. Il semblerait que la mise au point de systèmes hybrides utilisant en parallèle les deux types de descripteurs soit une voie à explorer pour la discipline, comme cela est par exemple décrit par [STS⁺11].

Applications annexes : Enfin, il serait intéressant d'appliquer nos cartes de densité à d'autres lésions que celles du foie (poumons, sein etc...), à d'autres modalités d'images (échographie, IRM etc...), et, pourquoi pas, à d'autres applications extérieures au domaine médical.

1. <http://www.sliver07.org>



Listes des publications

Conférences internationales avec actes

A. Quatrehomme, I. Millet, D. Hoa, G. Subsol et W. Puech : Assessment of an Automatic System Classifying Hepatic Lesions on Multi-Phase CT Scan Images, *EUSIPCO 21st European Signal Processing Conference*, 2013.

A. Quatrehomme, I. Millet, D. Hoa, G. Subsol et W. Puech : Assessing the Classification of Liver Focal Lesions by Using Multi-phase Computer Tomography scans, *Third MICCAI International Workshop on Medical Content-based Retrieval for Clinical Decision Support*, Octobre 2012, Lecture Notes in Computer Science 7723, pages 80-91, Springer, 2013.

A. Quatrehomme, D. Hoa, G. Subsol et W. Puech : Content-based Computer Tomography Image Retrieval on a Whole-body Anatomical Reference Set : Methods and Preliminary Results, *2nd International Workshop on Medical Image Analysis and Description for Diagnosis Systems - MIAD 2011*, Rome (Italy), 2011.

A. Quatrehomme, D. Hoa, G. Subsol et W. Puech : Review of Features Used in Recent Content-Based Radiology Image Retrieval Systems, *Proceedings of the third International Workshop on Image Analysis*, Nîmes (France), 2010.



Bibliographie

- [ACN⁺11] V.A. Arasu, R.C. Chen, D.N. Newitt, C.B. Chang, H. Tso, N.M. Hylton, and B.N. Joe. Can signal enhancement ratio (SER) reduce the number of recommended biopsies without affecting cancer yield in occult MRI-detected lesions? *Acad Radiology*, 18(6) :716 – 721, 2011. Cité page 51.
- [ARC12] A. Adcock, D. Rubin, and G. Carlsson. Classification of Hepatic Lesions using the Matching Metric. *CoRR*, abs/1210.0866, 2012. Cité pages 44, 45, and 47.
- [Bae10] KT. Bae. Intravenous contrast medium administration and scan timing at CT : considerations and approaches. *Radiology*, 256 (1) :32–61, 2010. Cité page 77.
- [Bai73] A. Bain. *Mind and body. The theories of their relation*. D. Appleton and company, 1873. Cité page 35.
- [BGD⁺04] M. Bilello, S. Burak Gokturk, T. Desser, S. Napel, R. Brooke Jeffrey Jr., and C. F. Beaulieu. Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous-phase CT. *Medical Physics*, 31 (2584), 2004. Cité pages 44, 45, and 46.
- [BL05] A. Baert and R. Lencioni. *Focal Liver Lesions*. Medical Radiology / Diagnostic Imaging. 2005. Cité page 5.
- [BM09] D. T. Boll and E. M. Merkle. Diffuse Liver Disease : Strategies for Hepatic CT and MR Imaging. *RadioGraphics*, 29 :1591–1614, 2009. Cité page 8.

- [Bre01] L. Breiman. Random Forests. 45(1) :5–32, 2001. Cité page 38.
- [BTK05] A. Bouridane, M. Tahir, and F. Kurugollu. An FPGA based coprocessor for GLCM and Haralick texture features and their application in prostate cancer classification. 43 (2) :205 – 215, 2005. Cité page 88.
- [CCC⁺] E. L. Chen, P. C. Chung, C. L. Chen, H. Tsai, and C. I. Chang. An automatic diagnostic system for CT liver image classification. *IEEE Trans Biomed Engineering*. Cité pages 44, 45, 46, and 47.
- [CJ96] D. Clausi and M. Jernigan. A fast method to determine co-occurrence texture features using a linked list implementation. 36 :506 – 509, 1996. Cité page 87.
- [Cog82] J. Coggins. *A Framework for Texture Analysis Based on Spatial Filtering*. PhD thesis, University of Michigan, 1982. Cité page 17.
- [CZ01] D. Clausi and Y. Zhao. Rapid determination of co-occurrence texture features. In *Geoscience and Remote Sensing Symposium*, volume 4, pages 1880 – 1882, 2001. Cité page 87.
- [DKBW06] D. Duda, M. Kretowski, and J. Bezy-Wendling. Texture Characterization for Hepatic Tumor Recognition in Multiphase CT. *Biocybernetics and Biomedical Engineering*, 26 (4) :15–24, 2006. Cité pages 49, 70, and 72.
- [Doi07] K. Doi. Computer-aided diagnosis in medical imaging : Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4-5) :198–211, 2007. Cité page 4.
- [Efr79] B. Efron. Bootstrap Methods : Another Look at the Jackknife. 7 (1) :1 – 26, 1979. Cité page 39.
- [EPRH02] M. Egmont-Petersen, D. De Ridder, and H. Handels. Image processing with neural networks - a review. 35 :2279 – 2301, 2002. Cité page 35.
- [Fis36] R. Fisher. The Use of Multiple Measurements in Taxonomic Problems. 7 (2) :179 – 188, 1936. Cité page 32.
- [Gal75] M. M. Galloway. Texture analysis using gray level run lengths. In *Comput. Vis. Graph.*, volume 4, page 172–179, 1975. Cité page 21.
- [GLZ⁺13] Z. Guo, Q. Li, L. Zhang, J. You, D. Zhang, and W. Liu. Is local dominant orientation necessary for the classification of rotation invariant texture? *Neurocomputing*, 116 :182 – 191, 2013. Cité page 27.
- [GMM04] T. Glatard, J. Montagnat, and I. Magnin. Texture based medical image indexing and retrieval : application to cardiac imaging. In *MIR '04, Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pages 135 – 142, 2004. Cité page 21.

- [Har79] R. M. Haralick. Statistical and Structural Approaches to Texture. In *Proceedings of the IEEE*, volume 67 n° 5, 1979. Cité pages 18 and 112.
- [HC71] J. M. Hammersley and J. M. Clifford. Markov fields on finite graphs and lattices. 1971. Cité page 23.
- [HCS06] Y. L. Huang, J. H. Chen, and W. C. Shen. Diagnosis of hepatic tumors with texture analysis in nonenhanced computed tomography images. 2006. Cité pages 44, 45, 46, and 47.
- [HGMM08] N. Harder, M. Gippa, G. Marcus, and R. Munner. Haralicks texture features using Graphics Processing Units. In *Proceedings of the World Congress on Engineering*, volume 1, 2008. Cité page 88.
- [HLH11] L. Hallum, M. Landy, and D. Heeger. Human primary visual cortex (V1) is selective for second-order spatial frequency. *Journal of Neurophysiology*, 105(5) :2121–2131, 2011. Cité page 18.
- [HRDA] A. Hillion, C. Roux, I. Donescu, and O. Avaro. Generalized second order invariance in texture modelling. Cité page 27.
- [HS78] M. Hassner and J. Sklansky. Markov random fields models of digitized image texture. In *Fourth International Joint Conference on Pattern Recognition*, pages 538 – 540, 1978. Cité page 22.
- [HUVW13] S. Hegenbart, A. Uhl, A. Vécsei, and G. Wimmer. Scale invariant texture descriptors for classifying celiac disease. *Medical Image Analysis*, 17(4) :458 – 474, 2013. Cité page 27.
- [IMMS05] M. Ishiguro, I. Murase, N. Moriyama, and R. Sekiguchi. A Classification Method of Liver Tumors based on Temporal Change of Hounsfield Unit in CT Images. In *Proceedings of SPIE 5747 Medical Imaging 2005 : Image Processing*, volume 5747, pages 822 – 830, 2005. Cité page 48.
- [Isi24] E. Ising. *Contribution to the Theory of Ferromagnetism*. PhD thesis, 1924. Cité page 22.
- [Jam90] W. James. *The Principles of Psychology*. Henry Holt and Co, 1890. Cité page 35.
- [JP12] A. Jadhav and R. Pawar. Color histogram based medical image retrieval system. In *International Conference on Electrical Engineering and Computer Science Engineering (ICEECS)*, 2012. Cité page 16.
- [KH02] K. Karimi and H.J. Hamilton. TimeSleuth : A Tool for Discovering Causal and Temporal Rules. In *ICTAI*, 2002. Cité page 34.

- [KM10] S. S. Kumar and Dr R. S. Moni. Diagnosis of Liver Tumor from CT Images using Curvelet Transform. *International Journal on Computer Science and Engineering*, 02 (04) :1173–1178, 2010. Cité pages 44, 45, 46, and 47.
- [KMR13] S. S. Kumar, Dr R. S. Moni, and J. Rajeesh. An automatic computer-aided diagnosis system for liver tumours on computed tomography images. *Computers and Electrical Engineering*, 39 :1516–1526, 2013. Cité pages 44, 45, 46, and 47.
- [Koh95] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, volume 2 (12), page 1137–1143, 1995. Cité page 39.
- [Law80] K. I. Laws. *Textured Image Segmentation*. PhD thesis, University of Southern California, 1980. Cité page 24.
- [MKY13] R. Maani, S. Kalra, and Y.H. Yang. Noise robust rotation invariant features for texture classification. *Pattern Recognition*, 46(8) :2103 – 2116, 2013. Cité page 27.
- [MVNN07] S. Mougiakakou, I. Valavanis, A. Nikita, and K. Nikita. Differential diagnosis of CT focal liver lesions using texture features, feature selection and ensemble driven classifiers. *Artificial Intelligence in Medicine*, 41 :25–37, 2007. Cité pages 44, 45, 46, and 47.
- [NBR⁺10] S. Napel, C. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma, and D.L. Rubin. Automated retrieval of CT images of liver lesions on the basis of image similarity : method and preliminary results. *Radiology*, 256 (1) :243–252, 2010. Cité page 46.
- [NMOJ⁺00] M. Nino-Murcia, E.W. Olcott, R.B. Junior Jeffrey, R.L. Lamm, C.F Beaulieu, and K.A. Jain. Focal liver lesions : pattern-based classification scheme for enhancement at arterial phase CT. *Radiology*, 215 :746 – 751, 2000. Cité page 48.
- [NY10] X. Niuniu and L. Yuxun. Review of decision trees. In *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, volume 5, pages 105 – 109, 2010. Cité page 34.
- [Pla98a] J. Platt. Fast Training of Support Vector Machines Using Sequential Minimal Optimization, 1998. Cité page 36.
- [Pla98b] John Platt. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998. Cité page 65.

- [QLC⁺12] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and Christian Roux. Fast wavelet-based image characterization for highly adaptive image retrieval. *IEEE transactions on image processing*, 4 (21) :1613 – 1623, 2012. Cité page 25.
- [Qui86] R. Quinlan. Induction of Decision Trees. 1 :81–106, 1986. Cité page 34.
- [Qui93] R. Quilan. C4.5 : Programs for Machine Learning. 1993. Cité page 34.
- [Rö95] W. Röntgen. On a new kind of ray : A preliminary communication. *Würzburg Physical and Medical Society*, 1895. Cité page 2.
- [RC05] I. Read and S. Cox. Automatic pitch accent prediction for text-to-speech synthesis. In *Interspeech 2007*, pages 482–485, 2005. Cité page 42.
- [SPRG13] M. Safdari, R. Pasari, D. Rubin, and H. Greenspan. Image patch-based method for automated classification and detection of focal liver lesions on CT. In *Medical Imaging 2013 : Computer-Aided Diagnosis*, volume 8670, 2013. Cité pages 44 and 45.
- [SS11] S. Su and Y. Sun. Key techniques research in computer-aided hepatic lesion diagnosis system based on multi-phase CT images. In *Image and Signal Processing (CISP), 2011 4th International Congress on*, volume 4, pages 1921–1927, 2011. Cité pages 50, 70, 71, and 72.
- [SSW⁺09] S. W. Shuqin, Y. Sun, Q. Weng, J. Ye, L. Gu, L. Qian, and J. Xu. Improvement of Feature Selection in Multi-phase CT Images of Hepatic Lesions. In *Proceedings of the 2009 International Symposium on Bioelectronics and Bioinformatics*, pages 72–75, 2009. Cité pages 50, 70, 71, and 72.
- [STS⁺11] S. Seifert, M. Thoma, F. Stegmaier, M. Hammon, M. Kramer, M. Huber, H.P. Kriegel, A. Cavallaro, and D. Comaniciu. Combined semantic and similarity search in medical image databases . In *Medical Imaging 2011 : Advanced PACS-based Imaging Informatics and Therapeutic Applications*, volume 7967, 2011. Cité page 116.
- [TB13] A. Thomas and A. Banerjee. *The History of Radiology*. Oxford Medical Histories. Oxford University Press, 2013. Cité page 2.
- [Tu05] Z. Tu. Probabilistic boosting-tree : learning discriminative models for classification, recognition, and clustering. In *IEEE ICCV*, volume 2, pages 1589 – 1596, 2005. Cité page 38.
- [Uns86] M. Unser. Sum and Difference Histograms for Texture Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(1) :118 – 125, 1986. Cité pages 88 and 112.

- [Vap79] V. Vapnik. Estimation of Dependences Based on Empirical Data [in Russian]. 1979. Cité page 35.
- [VC64] V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*. 1964. Cité page 35.
- [Wol96] D. Wolpert. The Lack of A Priori Distinctions between Learning Algorithms, 1996. Cité page 38.
- [WZL⁺09] L. Wang, Z. Zhang, J. Liu, B. Jiang, X. Duan, Q. Xie, D. Hu, and Z. Li. Classification of Hepatic Tissues from CT Images Based on Texture Features and Multiclass Support Vector Machines. 5552 :374–381, 2009. Cité pages 44, 45, and 46.
- [YSW09] J. Ye, Y. Sun, and S. Wang. Multi-Phase CT Image Based Hepatic Lesion Diagnosis by SVM. In *2nd International Conference on Biomedical Engineering and Informatics*, pages 1–5, 2009. Cité pages 50, 70, 71, 72, and 92.
- [YTK⁺12] S. Yamazoe, T. Takahara, K. Shimizu K, K. Ouchi, T. Mogami, J. Harada, and K. Fukuda. Diffusion-weighted imaging with relative signal intensity statistical thresholding for delineating prostate cancer tumors. *Magnetic Resonance in Medical Science*, 11(1) :1–8, 2012. Cité page 51.
- [Zer34] F. Zernike. Beugungstheorie des Schneidenverfahrens und Seiner Verbesserten Form, der Phasenkontrastmethode. In *Physica I*, 1934. Cité page 28.
- [ZMHP11] Y. Zhang, B.J. Matuszewsk, A. Histace, and F. Precioso. Statistical Shape Model of Legendre Moments with Active Contour Evolution for Shape Detection and Segmentation. In *Computer Analysis of Images and Patterns*, volume 6854 of *Lecture Notes in Computer Science*, pages 51–58. 2011. Cité page 28.



Table des figures

1.1	<i>L'une des premières images radiologiques réalisée sur une main humaine [Rö95]</i>	2
1.2	<i>Echographie (gauche), scanner (milieu) et IRM (droite, couleurs artificielles) de l'abdomen</i>	6
1.3	<i>Illustration de l'échelle de Hounsfield et de sa signification sur le type de tissu dans les images scanner</i>	6
1.4	<i>Image scanner de l'abdomen (haut) et du pelvis (bas) selon deux fenêtrages</i>	7
1.5	<i>Illustration de la tomographie multiphasique : abdomen (haut) et lésion hépatique (abcès, en bas) acquis sur différentes phases</i>	8
2.1	<i>Exemple d'un histogramme et des statistiques extraites de celui-ci sur une image DICOM scanner abdominale</i>	16
2.2	<i>Exemple d'une matrice de co-occurrence et des statistiques extraites de celle-ci sur une image DICOM scanner abdominale</i>	20
2.3	<i>Voisinage considéré dans un champ de Markov (les chiffres dans les cases indiquent l'ordre du voisinage par rapport au pixel central)</i>	22
2.4	<i>La transformée de Fourier (à droite) d'une image médicale d'un scanner abdominal (à gauche)</i>	24
2.5	<i>L'une des images TEM obtenues à l'aide de la méthode de Law (à droite) d'une image médicale d'un scanner abdominal (à gauche)</i>	25
2.6	<i>Ondelettes de Gabor (facteur d'échelle de 2 pour 4 orientations) : pour chaque orientation sont présentés le filtre, l'image filtrée puis les descripteurs statistiques extraits</i>	26
3.1	<i>Illustration du principe de l'arbre de décision Vue réalisée à partir de RapidMiner Community Edition</i>	34

3.2	<i>Illustration du principe du SVM</i>	36
5.1	<i>Exemple de nos zones d'intérêt : rectangulaire manuelle (à gauche), elliptique extraite (à droite)</i>	63
5.2	<i>Présentation du système complet de caractérisation des lésions hépatiques</i>	65
6.1	<i>Principe des acquisitions tomographiques multiphasiques</i>	70
6.2	<i>Présentation du système complet de caractérisation des lésions hépatiques sur acquisitions multiphasiques</i>	71
6.3	<i>Illustration de la variabilité intrinsèque des abcès : 3 abcès de notre base de données acquis sur la même phase (portale)</i>	74
6.4	<i>Courbes des moyennes de niveaux de gris de chaque lésion de notre Base 2 sur les différentes phases regroupées par type de lésion (abscisse : phase, ordonnée : intensité)</i>	78
6.5	<i>Images d'origine et normalisées d'une coupe scanner du foie scanner</i> . .	79
6.6	<i>Courbes des moyennes de niveaux de gris normalisés de chaque lésion de notre Base 2 sur les différentes phases regroupées par type de lésion (abscisse : phase, ordonnée : intensité)</i>	80
6.7	<i>Courbes des moyennes de niveaux de gris de chaque classe de lésion de notre Base 2 sur les différentes phases : non normalisées (en haut) et normalisées (en bas). En abscisse : la phase, en ordonnée : l'intensité</i>	81
7.1	<i>Présentation des méthodes d'Haralick et Unser sur un exemple pratique</i> .	90
7.2	<i>Profils temporels des lésions par classe (abscisse : changement de phase, ordonnée : valeurs arbitraires représentant le changement d'intensité, de -3 (forte baisse) à +3 (forte hausse) en partant par 0 (stabilité))</i>	94
7.3	<i>Effets du seuil d'isodensité sur la carte de densité du parenchyme hépatique d'un examen multiphasique de notre base</i>	95
7.4	<i>Exemples de cartes de densité sur notre Base 2</i>	96
7.5	<i>Répartition des zones de densité par rapport au foie sain moyennées par classe de lésion et phase (gris clair : hyperdensité, gris moyen : isodensité, gris foncé : hypodensité)</i>	97
9.1	<i>Cicatrice centrale d'une lésion HNF</i>	114
9.2	<i>Illustration de la transformée en coordonnées polaires sur une lésion</i> . .	115



Liste des tableaux

1.1	<i>Scénarios d’acquisition scanner du foie en fonction du contexte clinique</i>	8
2.1	<i>Synthèse des descripteurs visuels couramment extraits d’images médicales</i>	28
3.1	<i>Tableau comparatif de quelques familles d’algorithmes de classification</i>	37
3.2	<i>Matrice de confusion dans le cas d’une classification binaire</i>	41
4.1	<i>Synthèse des éléments caractéristiques des systèmes de classification des lésions hépatiques existants basés sur des examens monophasique</i>	45
5.1	<i>Présentation des classes de notre base de données d’un point de vue clinique</i>	59
5.2	<i>Aspect visuel des lésions par type et par phase</i>	60
5.3	<i>Présentation des caractéristiques des lésions au cours des phases temporelles</i>	61
5.4	<i>Composition de nos bases de données</i>	64
5.5	<i>Matrice de confusion et taux de réussite global de classification obtenus sur notre Base 1 sur l’Ensemble 1 de descripteurs</i>	67
6.1	<i>Comparaison de notre système avec les travaux de la littérature</i>	72
6.2	<i>Matrices de confusion et précisions totales de classification obtenue suite à une classification des lésions de notre Base 1 monophasique et multiphasique sur l’Ensemble 1 de descripteurs</i>	73
6.3	<i>Comparaison des scores obtenus par les experts radiologues et notre outil, ainsi que le coefficient de Dice entre les deux</i>	75

6.4	<i>Matrices de confusion et précisions totales de classification obtenue suite à une classification des lésions de notre Base 1 monophasique et multiphasique sur l'Ensemble 1 de descripteurs</i>	83
7.1	<i>Comparaison des approximations de la matrice de co-occurrence : temps de calcul parallélisé et scores de classification obtenus sur l'ensemble de notre Base 2 multiphasique normalisée</i>	91
7.2	<i>Matrice de confusion obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les descripteurs de cartes de densité</i>	98
7.3	<i>Mesures de la qualité de la qualification obtenues suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les descripteurs de cartes de densité</i>	98
8.1	<i>Comparaison de la précision de classification obtenue sur la Base 2 multiphasique avec divers pré-traitements des données</i>	103
8.2	<i>Précision de classification obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les différents groupes de descripteurs présentés</i>	104
8.3	<i>Précision de classification obtenue suite à une classification multi-classe sur notre Base 2 multiphasique normalisée avec les différents groupes de descripteurs</i>	105
8.4	<i>Comparaison de la précision de classification obtenue sur la Base 2 multiphasique en fonction du nombre d'attributs</i>	105
8.5	<i>Comparaison des résultats obtenus par l'Ensemble 2 de descripteurs sur la Base 2 multiphasique normalisée et la Base 2 sans abcès ni HNF</i>	107
8.6	<i>Précisions de classification obtenues suite à une classification "une classe contre toutes les autres" sur notre Base 2 multiphasique normalisée</i> . . .	108

Abstract

Medical imaging has taken benefits from recent advances and is becoming more and more important in the patient healthcare process. New needs arise, which are related to image processing. In particular, liver lesion classification is a big issue, as liver cancer is wide-spread and leads to death, most of the time because of the diagnosis which is made too late. In this context, a research project in collaboration between IMAIOS company and the Laboratory of Informatics, Robotics and Micro-electronics of Montpellier (LIRMM) has been setup, and the results are described in this report.

This thesis presents an automated system that first extracts visual features from CT scan images of lesions, then differentiate them based on these features. The contributions are: intensity normalization using healthy liver values, analysis and experimentations based on new visual features, which use temporal information or tissue density, characterisation of the lesions in different clinical classes. This work has been performed on multi-phase Computed Tomography acquisitions.

Keywords: *Medical Image Processing, Computer Aided Diagnosis, Liver focal lesions, Multi-Phase Computer Tomography, Classification*

Résumé

L'évolution des techniques d'acquisition des imageries médicales et leur importance de plus en plus grande dans la prise en charge du patient (diagnostic, préparation d'intervention, suivi etc...), font émerger de nouveaux besoins autour du traitement informatique des images. La reconnaissance du type de lésions hépatiques est un grand enjeu, notamment car le cancer du foie, léthal et très répandu, est souvent diagnostiqué trop tard pour sauver le patient. C'est dans ce cadre qu'est né le projet de recherche décrit dans ce manuscrit, fruit d'une collaboration entre la société IMAIOS et le Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM).

Cette thèse présente un système automatique permettant, à partir d'images de lésions au format médical DICOM, d'extraire des descripteurs visuels de divers nodules hépatiques puis de les différencier à l'aide de ces derniers. Les contributions décrites s'articulent autour de divers axes : normalisation des niveaux de gris des images de lésions par rapport au foie sain, proposition, analyse et tests de descripteurs visuels, s'appuyant notamment sur les informations temporelles ou de densité des tissus, caractérisations diverses des différents types de lésions grâce à ces descripteurs. Les données sur lesquelles ces travaux ont été effectués sont des examens scanner multiphasiques.

Mots clefs : *Imagerie Médicale, Diagnostic Assisté par Ordinateur, Lésions hépatiques focales, Tomographie Multi-Phase, Classification*
