

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

وزارة التعليم العالي و البحث العلمي

BADJI MOKHTAR-ANNABA UNIVERSITY
UNIVERSITY BADJI MOKHTAR-ANNABA



جامعة باجي مختار - عنابة

Faculté des Sciences de l'Ingéniorat
Département d'Informatique

Année : 2016/2017

THESE

Présentée en vue de l'obtention du diplôme de

Doctorat en Sciences

Intitulée :

Extraction des connaissances pour la segmentation d'images mammographiques

Option
Informatique

Par
Chokri FERKOUS

Devant le Jury

Pr. Yamina Mohamed Ben Ali	Professeur	Université d'Annaba	Président
Pr. Hayet-Farida Merouani	Professeur	Université d'Annaba	Directrice de Thèse
Pr. Hamid Seridi	Professeur	Université de Guelma	Examineur
Pr. Nadjia Benblidia	Professeur	Université de Blida	Examineur
Dr. Nabiha Azizi	Maitre de conférences classe A à L'université de Annaba		Examineur
Dr. Smain Mazouzi	Maitre de conférences classe A à L'université de Skikda		Examineur

Résumé

Le cancer du sein est le cancer le plus fréquent chez la femme. Il représente plus du tiers de l'ensemble des nouveaux cas de cancer chez la femme. La mammographie est actuellement la modalité d'imagerie la plus efficace pour la détection et le diagnostic du cancer du sein. Des études rétrospectives montrent que, dans les dépistages actuels du cancer du sein, environ 30 pour cent des cas de cancer du sein sont manqués par les radiologues. Avec les progrès des techniques de traitement et d'analyse d'images, il est envisagé que les radiologistes auront la possibilité de diminuer cette marge d'erreur et donc d'améliorer leur diagnostic.

Le but de ce travail est de proposer un système d'aide au diagnostic pour différencier entre les quatre classes BI-RADS dans les mammographies. Ce système s'inspire de l'approche du médecin lors d'un examen radiologique comme il a été convenu dans le système d'aide à la rédaction des comptes rendu BI-RADS (Breast imaging reporting and data system), où les masses sont décrites par leurs formes, leurs contours et leurs densités. La segmentation des masses dans notre approche est manuelle, car il est supposé que la détection est déjà faite. Lorsque la région suspecte est disponible, 22 caractéristiques de formes, de textures et de contours ont été automatiquement calculées à partir de la zone détectée ; seul l'âge du patient a été utilisé comme caractéristique humaine dans ce travail. La classification est finalement effectuée à l'aide d'un Perceptron Multicouches (PMC) selon deux scénarios distincts ; le premier schéma consiste à classer les masses pour différencier entre les quatre catégories BI-RADS (2, 3, 4 et 5). Dans le second, nous classons les anomalies sur les deux classes (bénignes ou malignes). L'approche proposée a été évaluée sur 480 masses mammographiques extraites de la base de données numériques pour le dépistage de mammographie (DDSM), et les résultats obtenus sont encourageants.

Mots-clefs BI-RADS, mammographie, extraction de caractéristiques, classification.

ملخص

إنّ سرطان الثدي هو السرطان الأكثر انتشاراً بين النساء في جميع أنحاء العالم. التصوير الشعاعي للثدي يعتبر حالياً طريقة التصوير الأكثر فعالية للكشف وتشخيص سرطان الثدي. وتظهر الدراسات حالياً أنه خلال الكشف عن سرطان الثدي، ما يقرب من 10 في المئة من حالات سرطان الثدي عن طريق الأشعة تصنف بالخطأ. لكن مع التقدم المسجل في ميدان تقنيات تحليل ومعالجة الصور، فمن المتوقع أن تكون هناك فرصة لأطباء الأشعة لتقليل هذا الهامش من الخطأ، وبالتالي تحسين التشخيص.

إن الهدف من هذا العمل هو اقتراح نظام للتشخيص بمساعدة الحاسوب للتمييز بين أربع فئات BI-RADS في تصوير الرقمي للثدي بالأشعة السينية. فكرة هذا النظام مستوحاة من منهجية الطبيب أثناء الفحص الإشعاعي كما تم الاتفاق عليه في النظام المساعد لكتابة التقارير الخاصة بصور الماموجرام BI-RADS (Breast imaging reporting and data system) الذي يتبعه أطباء الأشعة، حيث توصف الكتل المتواجدة على الصور بشكلها ونوعية حدودها وكثافتها. تجزئة الكتل وتحديد حدودها في هذا العمل تتم بطريقة يدوية من طرف خبير مختص في الأشعة لأننا نفترض أن هته الكتل تم الكشف عنها مسبقاً. وبمجرد توفر حدود هته الكتلة، يمكننا إجراء عملية إستخراج الميزات الخاصة بها. 22 خاصية يتم حسابها بصفة أوتوماتيكية تعبر عن شكل ونوعية حدود الكتلة وخصائص كثافتها. في هذا العمل تستخدم ميزة يدوية واحدة فقط، وهي المتمثلة في عمر المريض. أما فيما يخص تصنيف وتحديد نوعية هته الكتل تتم باستخدام المستقبالات المتعددة الطبقات (Perceptron Multicouches) وفقاً لخطين منفصلتين. الأولى تستعمل للتمييز بين فئات الكتل الأربعة BI-RADS 2, BI-RADS 3, BI-RADS 4 و BI-RADS 5. أما في الخطوة الثانية نصنف هته الكتل على فئتين (حميدة وخبيثة) وقد تم تقييم المنهج المقترح على 480 كتلة مستخرجة من قاعدة البيانات الرقمية للتصوير الإشعاعي للثدي (DDSM)، والنتائج التي تم الحصول عليها جد مشجعة.

الكلمات المفتاحية: BI-RADS، تصوير الثدي بالأشعة السينية، إستخراج الخصائص، التصنيف .

Abstract

Breast cancer is the most diagnosed cancer in women around the world. Mammography is currently the most effective imaging modality for the detection and diagnosis of breast cancer. Retrospective studies show that in current breast cancer screening, approximately 10 percent of breast cancer cases are missed by radiologists. With the advances in image processing and analysis techniques, it is envisaged that radiologists will be able to reduce this margin of error and thus improve their diagnosis.

The goal of this work is to propose a computer-aided diagnosis system to differentiate between four Bi-RADS classes in digitized mammograms. This system is inspired by the approach of the doctor during the radiologic examination as it was agreed in BI-RADS (Breast imaging reporting and data system), where masses are described by their form, their boundary and their density. The segmentation of masses in our approach is manual because it is supposed that the detection is already made. When the segmented region is available, the features extraction process can be carried out. 22 visual characteristics are automatically computed from shape, edge and textural properties; only one human feature is used in this work, which is the patient's age. Classification is finally done using a Multi-Layer Perceptron (MLP) according to two separate schemes; the first one consists to classify masses to distinguish between the four BI-RADS classes (2, 3, 4 and 5). In the second one we classify abnormalities on two Classes (Benign and malign). The proposed approach has been evaluated on 480 mammographic masses extracted from the Digital Database for Screening Mammography (DDSM), and the obtained results are encouraging.

Keywords BI-RADS, mammogram, features extraction, classification.

REMERCIEMENTS

Au terme de ce travail, je voudrai exprimer ma profonde gratitude envers Dieu tout puissant qui, grâce à son aide, j'ai pu finir ce travail.

Je tiens à exprimer mes remerciements les plus sincères à Madame **Merouani Hayet-Farida**, Professeur à l'université Badji Mokhtar-Annaba, ma directrice de recherche, pour m'avoir conseillée, encouragée et soutenue tout au long de ces années de thèse avec patience et disponibilité, et pour la confiance qu'elle m'a accordée.

Je tiens à remercier Madame **Yamina Mohamed Ben Ali**, Professeur à l'université de Badji Mokhtar - annaba, pour avoir accepté de présider le jury de cette thèse.

Un grand merci également aux membres du jury de soutenance qui m'ont fait l'honneur de bien vouloir évaluer mon travail. Je suis particulièrement reconnaissant aux examinateurs **Seridi Hamid**, Professeur à l'Université 08 mai 1945 de Guelma, **Benblidia Nadjia**, Professeur à L'Université Saâd Dahlab de Blida, **Azizi Nabiha**, Maitre de conférences classe A à l'Université Badji Mokhtar-Annaba et **Mazouzi Smain**, Maitre de conférences classe A à l'Université 20 Aout 1955 de Skikda.

J'exprime une réelle reconnaissance aussi à Mr. **Beledjhem Nedjmeddine**, radiologue à l'hôpital Ibn Rochd d'Annaba (ALGERIE), Pour sa disponibilité, ses conseils et ses remarques qui m'ont guidé tout au long de ma recherche. Ce qui m'a amenée à utiliser la norme BI-RADS dans mon travail pour la classification des anomalies mammaires

FERKOUS Chokri

DEDICACE

*A la mémoire de mon père,
A ma chère mère, pour son amour, son affection et ses prières,
A mon épouse "Hadjer", pour sa patience et son aide,
A ma fille "Nada",
A mes frères et mes sœurs qui m'ont tant soutenu et encouragé,
Et à tous mes collègues et mes amis.*

FERKOUS Chokri

Table des matières

RESUME	I
ملخص	II
ABSTRACT	III
REMERCIEMENTS	IV
DEDICACE	V
TABLE DES MATIERES	1
LISTE DES FIGURES	5
LISTE DES TABLEAUX	7
ABREVIATIONS ET ACRONYMES	8
INTRODUCTION GENERALE	10
CHAPITRE I: RECONNAISSANCE DE FORMES & VISION ARTIFICIELLE	15
I.1. INTRODUCTION	16
I.2. LA RECONNAISSANCE DE FORMES	16
I.3. PROCESSUS DE LA RDF	17
I.3.1. <i>Acquisition de données</i>	17
I.3.2. <i>Prétraitement de données</i>	17
I.3.3. <i>Extraction et sélection de caractéristiques</i>	19
I.3.4. <i>Apprentissage et classification</i>	20
I.3.4.1. Méthodes supervisées	20
I.3.4.1.1. Méthodes basées sur le concept de similarité	21
I.3.4.1.2. k plus proches voisins (kppv).....	21
I.3.4.1.3. Réseaux de neurones	23
I.3.4.1.4. Arbres de décision.....	27
I.3.4.1.5. Séparateurs à vaste marge (SVM)	29
I.3.4.1.6. La classification Bayésienne	30
I.3.4.2. Méthodes non supervisées	31
I.3.4.2.1. Méthodes basées sur le concept de similarité	31
I.3.4.2.2. Méthode K – moyennes	31
I.3.4.2.3. Cartes de Kohonen (Self-Organizing Maps).....	33

I.3.4.3.	Autres types d'apprentissage.....	33
I.4.	LA VISION PAR ORDINATEUR.....	34
I.4.1.	<i>Applications industrielles de la vision artificielle</i>	34
I.4.2.	<i>Analyse d'image</i>	36
I.4.2.1.	Analyse de bas niveau.....	36
I.4.2.2.	Analyse de haut niveau.....	37
I.5.	CONCLUSION.....	37
CHAPITRE II:	MAMMOGRAPHIE ET DEPISTAGE DU CANCER DU SEIN	39
II.1.	INTRODUCTION.....	40
II.2.	LA GLANDE MAMMAIRE.....	40
II.3.	LA MAMMOGRAPHIE.....	41
II.3.1.	<i>Les incidences de base</i>	42
II.3.1.1.	L'incidence de face ou incidence cranio-caudale (CC).....	42
II.3.1.2.	L'incidence oblique externe.....	42
II.3.2.	<i>La qualité de la mammographie</i>	43
II.3.2.1.	Caractéristique de positionnement.....	43
II.3.2.2.	Caractéristiques de la mammographie.....	44
II.3.3.	<i>Limite de la mammographie</i>	45
II.3.3.1.	L'échographie.....	46
II.3.3.2.	Imagerie par Résonance Magnétique (IRM).....	47
II.4.	LES OPACITES DU SEIN.....	47
II.5.	DEPISTAGE DU CANCER DU SEIN.....	49
II.5.1.	<i>Les densités asymétriques</i>	49
II.5.2.	<i>Les masses</i>	50
II.5.2.1.	Les masses régulières.....	51
II.5.2.2.	Les masses irrégulières.....	52
II.5.3.	<i>Les calcifications</i>	53
II.6.	LA CLASSIFICATION DES ANOMALIES MAMMOGRAPHIQUES.....	54
II.6.1.	<i>La classification de LeGal</i>	55
II.6.2.	<i>La classification de Lanyi</i>	55
II.6.3.	<i>La classification de BI-RADS</i>	56
II.6.3.1.	BI-RADS 0.....	56
II.6.3.2.	BI-RADS 1.....	56
II.6.3.3.	BI-RADS 2.....	56
II.6.3.4.	BI-RADS 3.....	57
II.6.3.5.	BI-RADS 4.....	58
II.6.3.6.	BI-RADS 5.....	59
II.6.3.7.	BI-RADS 6.....	60
II.7.	CONCLUSION.....	61

CHAPITRE III: EXTRACTION DE CARACTERISTIQUES MAMMAIRES	62
III.1. INTRODUCTION.....	63
III.2. LES CARACTERISTIQUES DES MASSES MAMMOGRAPHIQUES.....	63
<i>III.2.1. Les descripteurs de texture en mammographie</i>	<i>63</i>
III.2.1.1. Les statistiques du premier ordre.....	64
III.2.1.2. La matrice de co-occurrence.....	65
III.2.1.3. La méthode Run difference matrix (RDM)	70
III.2.1.4. La transformée de Fourier	73
III.2.1.5. L'analyse fractale	73
<i>III.2.2. Les descripteurs de forme en mammographie</i>	<i>74</i>
III.2.2.1. Les descripteurs géométriques.....	75
III.2.2.2. Les descripteurs spécifiques	80
<i>III.2.3. Les descripteurs de contours</i>	<i>83</i>
1. Le coefficient d'aplatissement (Kurtosis)	86
2. L'entropie	86
3. L'index de la probabilité maximale	86
<i>III.2.4. Autres caractéristiques.....</i>	<i>87</i>
III.3. CONCLUSION.....	87
CHAPITRE IV: LES ÉLÉMENTS D'UN SYSTÈME DE DÉPISTAGE AUTOMATIQUE DU CANCER DU SEIN (DACS).....	89
IV.1. INTRODUCTION.....	90
IV.2. LES SYSTEMES DE DEPISTAGE AUTOMATIQUE DU CANCER DU SEIN	90
<i>IV.2.1. Les avantages des systèmes DACS.....</i>	<i>91</i>
<i>IV.2.2. L'histoire des systèmes DACS.....</i>	<i>94</i>
IV.3. LES TECHNIQUES UTILISEES PAR LES SYSTEMES DACS	95
IV.4. PROCESSUS DE DEPISTAGE AUTOMATIQUE DU CANCER DU SEIN	99
<i>IV.4.1. Prétraitement d'images.....</i>	<i>100</i>
<i>IV.4.2. Segmentation</i>	<i>104</i>
IV.4.2.1. Les méthodes de seuillage.....	105
IV.4.2.2. Les méthodes basées régions.....	106
IV.4.2.3. Les méthodes basées contours.....	107
<i>IV.4.3. Extraction de caractéristiques</i>	<i>108</i>
<i>IV.4.4. La sélection des caractéristiques</i>	<i>109</i>
<i>IV.4.5. La classification</i>	<i>111</i>
<i>IV.4.6. Évaluation de la performance.....</i>	<i>112</i>
IV.5. CONCLUSION.....	116
CHAPITRE V: LA CLASSIFICATION AUTOMATIQUE DES MASSES MAMMOGRAPHIQUES ..	118
V.1. INTRODUCTION.....	119

Table des matières

V.2.	PROCESSUS DE CLASSIFICATION AUTOMATIQUE DES MASSES	119
V.2.1.	<i>La sélection de la région d'intérêt (ROI)</i>	121
V.2.2.	<i>La Segmentation des masses mammaires</i>	122
V.2.3.	<i>L'extraction de caractéristiques des masses mammographiques</i>	122
V.2.3.1.	Extraction de caractéristiques de forme	122
V.2.3.2.	L'extraction de caractéristiques de contours.....	126
V.2.3.3.	L'extraction de caractéristiques de la densité des masses.....	128
V.2.3.3.1.	L'extraction de caractéristiques texturale.....	128
V.2.3.3.2.	Extraction de caractéristiques de l'intensité	130
V.2.3.3.3.	Les caractéristiques de texture et de densité utilisées	131
V.2.3.3.4.	Autres caractéristiques	132
V.2.3.4.	La classification.....	132
V.3.	BASE DE DONNEES DE MAMMOGRAPHIES.....	133
V.3.1.	<i>Une description des fichiers « .JPEG »</i>	134
V.3.2.	<i>Description de fichier « .ics ».....</i>	134
V.3.3.	<i>Description du fichier overlay.....</i>	135
V.3.4.	<i>La description des fichiers « .16_PGM ».....</i>	136
V.4.	RESULTATS ET DISCUSSIONS.....	137
V.4.1.	<i>Les images de test et d'apprentissage</i>	137
V.4.2.	<i>1er scénario (La classification BI-RADS)</i>	138
V.4.3.	<i>2^{ème} scénario (masses malignes Vs bénignes)</i>	142
V.4.4.	<i>Comparaison avec les méthodes existantes.....</i>	142
V.5.	CONCLUSION.....	145
	CONCLUSION GENERALE	146
	RÉFÉRENCES BIBLIOGRAPHIQUES.....	150
	LISTE DES PUBLICATIONS.....	164

Liste des Figures

FIGURE I.1: LES ETAPES PRINCIPALES D'UNE APPROCHE DE RECONNAISSANCE DE FORME TYPIQUE [HARTERT 2010].	18
FIGURE I.2: EXEMPLE DE CLASSIFICATION EN UTILISANT LA METHODE KPPV AVEC ($k=1$ ET $k=6$)	22
FIGURE I.3: REPRESENTATION D'UN NEURONE ARTIFICIEL [TILMANNE 2009].	24
FIGURE I.4: SCHEMA D'UN PERCEPTRON MULTICOUCHES (PMC) [TILMANNE 2009].	26
FIGURE I.5: UN EXEMPLE D'ARBRE DE DECISION POUR LA CLASSIFICATION DE 100 INDIVIDUS PRIS DE LA BASE IRIS. [RAKOTOMALALA 2005].	27
FIGURE I.6: LA REPRESENTATION D'UN CLASSIFIEUR DE TYPE SVM.	29
FIGURE I.7: QUELQUES APPLICATIONS INDUSTRIELLES DE LA VISION PAR ORDINATEUR.	35
FIGURE II.1: SCHEMA ANATOMIQUE DU SEIN [DRAKE 2015].	41
FIGURE II.2: A : RESULTAT DE L'INCIDENCE DE FACE, B : RESULTAT DE L'INCIDENCE OBLIQUE.	44
FIGURE II.3: EXEMPLES D'IMAGES ECHOGRAPHIQUES DU SEIN.	46
FIGURE II.4: EXEMPLE D'IRM MAMMAIRE BI-SEIN.	47
FIGURE II.5: UNE SURDENSITE SUR LE CLICHE MAMMOGRAPHIE.	48
FIGURE II.6: UNE ASYMETRIE DE DENSITE OBSERVEE SEULEMENT SUR LA MAMMOGRAPHIE GAUCHE.	50
FIGURE II.7: LES FORMES DES MASSES.	50
FIGURE II.8: LES TYPES DES CONTOURS DES MASSES.	51
FIGURE II.9: MASSE REGULIERE RONDE DANS LA MAMMOGRAPHIE (INCIDENCE LATERALE-OBLIQUE)	51
FIGURE II.10: IMAGE DE MASSES A CONTOURS REGULIERS.	52
FIGURE II.11: IMAGES DE DISTORSION ARCHITECTURALE	53
FIGURE II.12: IMAGES DE MASSES AVEC DES LIMITES EN SPICULES.	53
FIGURE II.13: LES TYPES DE CALCIFICATIONS.	54
FIGURE II.14: ASPECT TYPIQUE DE FIBROADENOME PARTIELLEMENT CALCIFIE : BI-RADS 2	57
FIGURE II.15: MICROCALCIFICATIONS BENIGNES : LAIT CALCIQUE : BI-RADS 2 [LEVY 2005].	57
FIGURE II.16: MASSES BIEN CIRCONSCRITES DISCRETEMENT LOBULEES, SANS MICRO LOBULATION, NON CALCIFIEES, NON TYPIQUEMENT LIQUIDIENNES EN ECHOGRAPHIE : BIRADS 3 [LEVY 2005].	58
FIGURE II.17: MICRO CALCIFICATIONS GRANULAIRES IRREGULIERES PEU NOMBREUSES : BIRADS 4	59
FIGURE II.18: IMAGES SPICULEE SANS CENTRE DENSE : CENTRE PROLIFERATIF D'ASCHOFF [LEVY 2005].	59
FIGURE II.19: MASSE MAL CIRCONSCRITE AUX CONTOURS FLOUS ET IRREGULIERS : BI-RADS 5	60
FIGURE II.20: MICRO CALCIFICATIONS VERMICULAIRES, ARBORESCENTES, NOMBREUSES ET GROUPEES : BI-RADS 5 [LEVY 2005].	60
FIGURE III.1: LES VOISINS D'UNE DISTANCE $D = 2$ DU PIXEL 'x' SELON LES 4 DIRECTIONS.	66
FIGURE III.2: EXEMPLE D'UNE IMAGE.	66
FIGURE III.3: LA MATRICE DE COOCCURRENCE $M^p_{i,j}$ DE L'IMAGE (FIGURE III.2).	67
FIGURE III.4: LE DEPLACEMENT 'r'.	70
FIGURE III.5: RUN DIFFERENCE MATRIX (RDM).	71
FIGURE III.6: LA MATRICE RDM DE L'IMAGE (FIGURE III.2).	71

Liste des Figures

FIGURE III.7: RELATION ENTRE LA COMPLEXITE DE LA FORME ET LE TYPE DE CONTOURS DE LA MASSE ET MALIGNITE : A)MASSE CIRCONSCRITE CIRCULAIRE, B) CIRCONSCRITE OVALE, C) MACRO LOBULEE, D) MICRO LOBULEE, E)IRREGULIERE, F) SPECULE. [BRUCE 1999]	74
FIGURE III.8: CALCUL DE LA MESURE DE CIRCULARITE [CHEIKHROUHOU 2007].	76
FIGURE III.9: CALCUL DE LA MESURE DE RECTANGULARITE [CHEIKHROUHOU 2007].....	77
FIGURE III.10: LA COQUE CONVEXE ET LE CONTOUR D'UNE MASSE MAMMAIRE [BEHNAM 2010].	78
FIGURE III.11: UN EXEMPLE DE CARACTERISTIQUES ÉQUATION III-44 ÉQUATION III-48 A L' AIDE D'UNE MASSE EN ROTATION A DES ANGLES DIFFERENTS [BEHNAM 2010].	80
FIGURE III.12: EXEMPLE DE DEUX ONDELETTES I_1 ET I_2	83
FIGURE III.13: LA REPRESENTATION DES ONDELETTES SUR UNE MASSE MAMMOGRAPHIQUE.....	84
FIGURE III.14: MODELE SCHEMATIQUE DU CALCUL DES PROBABILITES DE CONTOUR DE CHAQUE ONDELETTE.	85
FIGURE IV.1: SCHEMA GENERAL POUR LA DETECTION ET LE DIAGNOSTIC AUTOMATIQUE DU CANCER DU SEIN.....	100
FIGURE IV.2: LES ARTEFACTS DES MAMMOGRAPHIES.	102
FIGURE IV.3: LA DETECTION DE CONTOURS DE LA GLANDE MAMMAIRE EN UTILISANT LES CONTOURS ACTIFS [FERKOUS 2010b].	102
FIGURE IV.4: EXEMPLES D'EXTRACTION DE LA GLANDE MAMMAIRE DES MAMMOGRAPHIES MDB053, MDB172 ET MDB194 DE LA BASE DE DONNEES MIAS [FERKOUS 2010b].	103
FIGURE IV.5: RESULTAT D'UNE MAUVAISE DETECTION DE LA MAMMOGRAPHIE "MDB152" DE LA BASE DE DONNEES MIAS [FERKOUS 2010].....	103
FIGURE IV.6: RESULTAT DE SEPARATION DU MUSCLE PECTORAL DE LA MAMMOGRAPHIE MDB020 MDB004 DE LA BASE DE DONNEES MIAS[FERKOUS 2010].....	104
FIGURE IV.7: LE RESULTAT DE SEGMENTATION DE DEUX MASSES CIRCONSCRITE ET SPICULEE [LI 2008].	108
FIGURE IV.8: LA COURBE ROC — FFP VS. FVP.....	114
FIGURE IV.9: LES COURBES ROC	115
FIGURE V.1: LES ETAPES DE LA METHODOLOGIE PROPOSEE.....	121
FIGURE V.2: EXEMPLE DE SELECTION DES POINTS V_i	123
FIGURE V.3: LES FORMES DE MASSES (A) RONDE (B) LOBULE (C) IRREGULIERE.....	124
FIGURE V.4: LA PRESENTATION D'UNE FORME D'ONDE I AYANT UN CENTRE Q_i ET LONGUEUR L_i	126
FIGURE V.5: UNE MASSE AVEC MICRO-CALCIFICATIONS CLASSEE BI-RADS 5 [ROBIN 2016].	128
FIGURE V.6: LA LISTE DES FICHIERS DU REPERTOIRE B-3024-1.	134
FIGURE V.7: LE FICHIER B-3024-1.ICS.	135
FIGURE V.8: CONTENU DU FICHIER B_3024_1.RIGHT_CC.OVERLAY.....	136
FIGURE V.9 : LA REPRESENTATION DES VALEURS DE DEPLACEMENT DE LA CHAINE CODEE.	136
FIGURE V.10: LA VISUALISATION D'UN FICHIER ".16_PGM".	137
FIGURE V.11: DIFFERENTES FORMES DES MASSES MAMMAIRES.	140
FIGURE V.12: DIFFERENTES TYPES DE CONTOURS DES MASSES MAMMAIRES.	140
FIGURE V.13: DIFFERENTES DENSITES DES MASSES MAMMAIRES.	141

Liste des tableaux

TABLEAU II-1: LES CLASSES DES MICRO-CALCIFICATIONS SELON LEGAL.....	55
TABLEAU III-1: INTERPRETATION BI-RADS POUR LES VALEURS DE DENSITE DU SEIN.....	87
TABLEAU IV-1 : RELATION ENTRE VP, VN, FP ET FN DE LA MATRICE DE CONFUSION.....	112
TABLEAU IV-2: MESURES DU PERFORMANCE D'UNE CLASSIFICATION BINAIRE.....	113
TABLEAU V-1: LES CARACTERISTIQUES DE FORME.....	124
TABLEAU V-2: RESULTATS DES ENERGIES DE CONTINUTE, DE COURBURE ET DE L'IRREGULARITE DES MASSES DE LA FIGURE V.3.....	124
TABLEAU V-3: LES AVANTAGES ET LES INCONVENIENTS DES DESCRIPTEURS GEOMETRIQUES.....	125
TABLEAU V-4: LES CARACTERISTIQUES DE CONTOURS UTILISEES DANS NOTRE SYSTEME.....	127
TABLEAU V-5: LES CARACTERISTIQUES DE DENSITE.....	131
TABLEAU V-6: AUTRES CARACTERISTIQUES UTILISEES.....	132
TABLEAU V-7: LA REPARTITION DES MASSES UTILISEES DANS CETTE ETUDE SELON LEUR FORME, LEUR TYPE DE CONTOURS ET LEUR CLASSE BI-RADS.....	138
TABLEAU V-8: LA CATEGORISATION BI-RADS.....	139
TABLEAU V-9: LA MATRICE DE CONFUSION DU PREMIER SCENARIO (LA CLASSIFICATION BI-RADS).....	141
TABLEAU V-10: LA MATRICE DE CONFUSION DU DEUXIEME SCENARIO (CLASSIFICATION MALIGNE/ BENIGNE).....	142
TABLEAU V-11: COMPARAISON AVEC D'AUTRES METHODES EXISTANTES.....	144

Abréviations et Acronymes

AD	L'ANALYSE DISCRIMINANTE
ADL	ANALYSE DISCRIMINANTE LINEAIRE
ALDGD	ACCENTUATION DE LONGUE DISTANCE POUR UNE GRANDE DIFFERENCE
ANOVA	L'ANALYSE DE LA VARIANCE
BFG	BANCS DE FILTRES DE GABOR
BI-RADS	SYSTEME DE COMPTES-RENDUS ET DE DONNEES D'IMAGERIE MAMMAIRE (BREAST IMAGING REPORTING AND DATA SYSTEM)
CC	CRANIO CAUDAL
DA	L'ANALYSE DISCRIMINANTE
DACS	DEPISTAGE AUTOMATIQUE DU CANCER DU SEIN
DBCM	METHODE DIFFERENTIELLE DE COMPTAGE DE BOITES (DIFFERENTIAL BOX COUNTING METHOD)
DDIFM	DISTRIBUTION DE LA DIFFERENCE MOYENNE
DDISM	DISTRIBUTION DE LA DISTANCE MOYENNE
DDSM	LA BASE DE DONNEES NUMERIQUE POUR LA MAMMOGRAPHIE DE DEPISTAGE
DEAO	DETECTION ASSISTEE PAR ORDINATEUR
DxAO	DIAGNOSTIC ASSISTE PAR ORDINATEUR
EGD	ÉCART DE GRANDE DIFFERENCE
ELM	MACHINE D'APPRENTISSAGE EXTREME
ENTR	ENTROPIE
FDA	L'ADMINISTRATION DES ALIMENTS ET DE DROGUES DES ÉTATS-UNIS (FOOD AND DRUG ADMINISTRATION)
GLCM	MATRICE DE COOCCURRENCE DES NIVEAUX DE GRIS
IRMA	BASE DE DONNEES MAMMOGRAPHIQUE (IMAGE RETRIEVAL IN MEDICAL APPLICATIONS)
KPPV	K-PLUS PROCHES VOISINS
KURT	KURTOSIS
LDA	ALLOCATION DE DIRICHLET LATENTE (LATENT DIRICHLET ALLOCATION)
MCC	MICRO CALCIFICATION
MLO	MEDIO-LATERALE OBLIQUE
MTRL	LES MODELES TERNAIRES RADIAL LOCAUX

NF	NOMBRE DE FORMES D'ONDE
OEP	OPTIMISATION D'ESSAIM DE PARTICULES
PC	PROBABILITE DE CONTOURS
PMC	PERCEPTRON MULTICOUCHE
PSA	ANALYSE SPECTRALE DE PUISSANCE (POWER SPECTRAL ANALYSIS)
RDM	RUN DIFFERENCE MATRIX
RDNG	REPARTITION DE LA DIFFERENCE DE NIVEAU DE GRIS
ROI	REGION D'INTERET
ROIms	REGIONS D'INTERETS AVEC MASSES
SGLDM	METHODE SPATIALE DE DEPENDANCE DES NIVEAUX DE GRIS (SPATIAL GRAY LEVEL DEPENDENCE METHOD)
SMD	LE MOMENT D'ORDRE 2 DE LA DDIFM
SMG	LE MOMENT D'ORDRE 2 DE DISTRIBUTION DE LA DIFFERENCE DE NIVEAUX DE GRIS
SVM	SEPARATEURS A VASTE MARGE

INTRODUCTION GENERALE

L'augmentation de la puissance des calculateurs dans les dernières décennies a permis le développement du traitement d'images numérique, ce traitement qui nécessite un très grand nombre d'opérations mathématiques. Les puissances des processeurs actuels permettent d'effectuer du traitement complexe en temps réel. Les logiciels utilisés pour traiter les images numériques sont une preuve à cette révolution, dans le domaine du traitement d'images, on distingue deux grands axes principaux : la synthèse d'images et l'analyse d'images, ces deux axes sont étroitement liés. Par exemple, la reconstruction 3D en imagerie médicale utilise les techniques de synthèse pour reconstituer des formes à partir des données fournies par les coupes des scanners ou des balayages ultrasoniques.

L'analyse d'images est le domaine qui nous préoccupe ici particulièrement, elle présente un intérêt dans les domaines d'extraction automatique de caractéristiques de l'image.

Le domaine qui nous intéresse est celui de l'imagerie médicale qui constitue aujourd'hui un outil extraordinaire pour l'aide au diagnostic pour les médecins. Les modalités d'acquisition d'images sont nombreuses et la quantité d'images manipulées s'accroît dans le domaine médical. La radiographie par rayon X constitue la part la plus importante du marché médical, son principe consiste à impressionner sur un film radiographique les différences de densité d'un organe. Le film sera plus ou moins noirci selon l'organe radiographié. L'os apparaît blanc, les tissus mous seront dans différents niveaux de gris et l'air sera noir.

La mammographie est un examen radiographique du sein, effectué chez la femme pour le dépistage du cancer du sein. Toutefois, la mammographie constitue la modalité principale pour le diagnostic du cancer du sein. L'image mammographique possède également l'avantage d'être très pratique et simple d'utilisation. L'analyse des images mammographique permet de fournir au médecin des informations quantitatives et fiables, ces informations complètent l'aspect visuel et qualitatif de l'examen [Verma 2001]. Les avantages potentiels de la mammographie numérique sont :

1. Possibilité de traitement d'image

Actuellement, les images mammographiques sont présentées au format numérique. (Un contraste, une densité et une amélioration de contours peuvent être obtenus).

2. Possibilité d'utilisation des systèmes de dépistage automatiques du cancer du sein (DACS)

Plusieurs outils de détection et de diagnostic assisté par ordinateur sont en cours de développement. Dans un proche avenir, ces outils permettront un dépistage mammographique rentable, dans lequel une deuxième lecture en utilisant les systèmes DACS sera recommandée. Le cancer du sein est l'un des grandes préoccupations de la santé qui a commencé à réclamer de l'importance dans la recherche médicale en raison de sa prévalence élevée et les taux de détection depuis les dernières décennies.

Selon les statistiques de l'institut national de la santé publique Algérienne (INSP) [Hamdi 2014] [Hamdi 2015], le nombre de décès annuels dus à cette maladie est évalué à 3500. En outre, on recense 10910 nouveaux cas de maladie par an, L'incidence a été multipliée par 7 en 28 ans. La prévalence a grimpé de 9.3 pour 100.000 habitants en 1986 à 65.2 pour 100.000 habitants actuellement.

En raison des différents types de fatigue et les facteurs humains, le taux d'échec a été élevé. Il a été observé que les radiologues manquaient environ 30% de l'ensemble des lésions cancéreuses [Nagappan 2008]. Aussi, le pourcentage global de cancer du sein détecté par nombre des mammaires biopsies réalisées sur la base de mammographie se situe entre 10 et 50% [Säbel 1996].

La prévention du cancer du sein jusqu'à présent n'est pas une solution fiable. Cela nécessite qu'une détection précoce soit obligatoire afin de réduire ou d'empêcher la perte de vie due au cancer du sein.

Dans le passé, la variabilité de la terminologie utilisée dans les rapports mammographiques a souvent mènes à des interprétations erronées. BI-RADS (Breast Imaging-Reporting And Data System) de l'ACR (American College of Radiology) est une procédure de normalisation créé pour améliorer la clarté et de réduire la variabilité de la terminologie utilisée dans les rapports mammographique [Obenauer 2005].

Lors d'un examen mammographique, le radiologue ne fait que résoudre un problème de vision ; c'est à partir de ce point et de récentes techniques d'analyse et prétraitement d'image, que l'idée de réaliser un système d'aide au diagnostic devient réalisable. La recherche menée dans ce domaine actuellement est encore à un stade peu avance. Pour l'instant, seul le problème de la

classification des micro-calcifications (MCC) a prouvé sa faisabilité. Les résultats plus qu'encourageants obtenus dans ce domaine ont permis d'espérer un aboutissement aux autres problèmes plus complexes comme celui des masses.

Le diagnostic assisté par ordinateur (DxAO) est devenu l'un des sujets de recherche les plus importants en radiologie; un système DxAO typique pour la classification mammographique doit comprendre quatre grandes étapes : le prétraitement, la segmentation, l'extraction de caractéristiques et la classification.

Parmi les étapes ci-dessus, c'est l'étape d'extraction de caractéristiques qui nous intéresse le plus dans notre travail en raison de son importance et de l'influence sur le résultat de la classification.

Dans les rapports mammographiques, les radiologues caractérisent les masses mammaires par leurs densités, leurs formes et leurs types de contours pour les classer dans l'une des classes BI-RADS (2, 3, 4 et 5); ceci est justifié par le fait que les radiologues ne peuvent pas confirmer la malignité d'une anomalie quelle que soit ses caractéristiques sans faire une étude de suivi ou une biopsie.

Les DxAO ne remplaceront jamais les experts, mais, doivent être complémentaire à eux, et les radiologues peuvent utiliser la sortie du DxAO comme une deuxième opinion pour prendre les décisions finales [Doi 2007], cela signifie que le jugement d'un DxAO devrait être du même type et doit utiliser le même lexique utilisé par les experts. C'est pourquoi un système DxAO mammographique doit classer les masses à l'aide des classes BI-RADS, de la même manière que les experts en radiologie. A cet effet, nous avons développé un système DxAO inspiré de l'approche du médecin lors de l'examen radiologique tel qu'il a été convenu dans BI-RADS, ce qui permet de classer une masse mammographique en se basant sur un ensemble de caractéristiques quantifiées sur les quatre classes BI-RADS mentionnées précédemment. Les contributions de notre travail sont les suivantes :

- En plus de la classification des masses comme malignes ou bénignes, notre système DxAO peut distinguer entre les quatre classes de BI-RADS (2, 3, 4 et 5), qui est la même procédure pratiquée par les radiologues. C'est pourquoi; nous pouvons dire que notre approche est la première technique qui catégorise les masses mammographiques en se basant sur la classification BI-RADS de l'ACR grâce à une étape d'extraction automatique des caractéristiques.
- Ce travail met l'accent sur toutes les caractéristiques utilisées par les radiologues durant un examen mammographique incluant la densité, la forme et les

descripteurs de contours, en plus l'âge du patient et la taille de la masse, comme il a été convenu dans le lexique BI-RADS.

- Les masses utilisées à l'étape expérimentale sont sélectionnés par un radiologue expérimenté afin de toucher toutes les catégories de densités, de formes et types de contours.

Cette thèse est organisée de manière méthodique, en cinq chapitres comme suit :

Dans le 1^{er} chapitre, un état de l'art sur la reconnaissance des formes et la vision artificielle est présenté, avec un accent particulier sur les techniques d'apprentissage automatique, à savoir, les approches supervisées et non-supervisées ; ainsi que les avantages et les inconvénients de chaque méthode.

Dans le 2^{ème} chapitre, nous présentons le domaine de diagnostic des anomalies mammaires et de la mammographie numérique ; à savoir, les informations générales sur la structure de la glande mammaire, les types de tumeurs du sein ; où nous étudions les différents types d'anomalies selon leurs formes, leurs contours et leurs contrastes; ce chapitre nous permet de mieux cerner le problème de diagnostic abordé. Ensuite, des questions de fond concernant les différentes modalités d'imagerie telles que l'échographie, la mammographie numérique et l'imagerie par résonance magnétique (IRM). Vers la fin de ce chapitre, l'analyse et l'interprétation des mammographies numériques à l'aide du lexique BI-RADS est présentée.

Le 3^{ème} chapitre expose en premier lieu les descripteurs de texture les plus fréquents telles que les statistiques de premier ordre, la matrice de co-occurrence, la méthode « Run difference matrix (RDM) », la transformée de Fourier et l'analyse fractale. Ensuite, nous présentons les différents descripteurs de forme et de contours. L'étude de ces descripteurs permet d'identifier les avantages et les inconvénients de chaque méthode.

Dans le 4^{ème} chapitre, nous présentons premièrement l'analyse de la littérature des systèmes de détection et de diagnostic automatique (DACS) des mammographies numériques. Puis, nous exposons les concepts fondamentaux du traitement numérique utilisées dans les applications de mammographie en mettant l'accent sur les techniques de filtrages, de segmentation et l'extraction de caractéristiques.

Dans le cinquième et dernier chapitre nous présentons la conception de notre travail, ainsi que l'application testée sur la base d'image mammographique DDSM (The Digital Database for Screening Mammography). Cette application d'aide au diagnostic appliquée sur les images mammographiques, permet l'extraction de caractéristiques de formes, de contours et de densités des masses mammographiques, la classification est effectuée à l'aide d'un Perceptron

Multicouches (PMC) selon deux scénarios distincts ; le premier consiste à catégoriser les masses pour différencier entre les quatre classes BI-RADS (2, 3, 4 et 5). Dans le second, nous classons les anomalies sur les deux classes (bénignes et malignes). L'approche proposée a été évaluée sur 480 masses mammographiques extraites de la base de données numérique pour le dépistage des mammographies (DDSM).

Enfin, la conclusion générale présente une synthèse de ce travail et les perspectives qui en découlent.

Chapitre

1

**RECONNAISSANCE DE
FORMES & VISION
ARTIFICIELLE**

I.1. Introduction

L'être humain reçoit en permanence des informations très diverses et très complexes sur le monde qui l'entoure par l'intermédiaire de ses cinq sens. Au sein du cerveau humain, il a été décrit de nombreuses voies qui forment un système chargé de traiter les informations pour la description des objets (reconnaissance des visages, reconnaissance de caractères, reconnaissance vocale, etc.). Malgré l'abondance de ces informations, le cerveau est capable de restituer chaque objet observé et de lui attribuer une représentation cohérente appelée "description humaine".

La reconnaissance de formes (RDF) constitue l'outil principal de l'intelligence artificiel, en abordant des problèmes dont les humains résolvent à leur insu. L'objectif de la RDF est de développer des machines qui reconstruisent et interprètent un environnement en se basant sur des mesures énergétiques [Besag 1991].

La vision artificielle est sans doute l'un des principaux champs d'application de la Reconnaissance des Formes. C'est une science qui repose essentiellement sur l'intelligence artificielle, les théories mathématique et le traitement de signal. Lorsque le cerveau humain voit un texte, il peut distinguer les caractères. Cette opération est simple pour le cerveau humain mais elle est très difficile à traduire par des algorithmes pour qu'elle soit compréhensible par la machine.

Dans la première partie de ce chapitre, l'intérêt général de la reconnaissance des formes est illustré. Par la suite, les différentes étapes d'un système de reconnaissance de forme générique seront présentées de manière synthétique. Nous détaillerons plus particulièrement les méthodes de classification accompagnées de leurs avantages et inconvénients. Ces méthodes peuvent être réparties en trois catégories : la classification supervisée, dont différentes approches sont présentées à la section I.3.4.1, la classification non supervisée (plus connue sous le nom de « clustering », en anglais), qui fait l'objet de la section I.3.4.2 et finalement les approches semi-supervisées. Dans la deuxième partie, nous abordons la notion de la vision artificielle qui consiste à tenter d'imiter la vision humaine par le truchement des techniques de l'intelligence artificielle.

I.2. La reconnaissance de formes

La RDF est l'ensemble de techniques et méthodes visant à identifier des motifs à partir des données brutes afin de prendre une décision dépendante de la catégorie attribuée à ce motif, elle se considère comme une branche de l'intelligence artificielle qui fait largement appel aux

techniques d'apprentissage automatique et aux statistiques. Les formes à reconnaître peuvent être de natures très variées. Il peut s'agir de contenu visuel (caractère, visage, empreinte digitale, images médicales, rayon X, IRM ou images satellitaires ...) ou sonore (reconnaissance de parole ou de locuteur) ou bien d'autres. La RDF s'intéresse à la conception et à la réalisation de systèmes (matériels ou logiciels) capables de percevoir, et dans une certaine mesure, d'interpréter des signaux captés dans le monde physique [Dubuisson 1990].

La RDF est un des nombreux aspects de l'intelligence artificielle. A partir d'un ensemble de données ou d'informations apprises, elle offre la possibilité d'interpréter toute nouvelle observation (ou forme). Les observations déjà connues sont regroupées en classes, constituant des prototypes auxquels la nouvelle observation est comparée pour être identifiée. Les algorithmes utilisés permettent donc de classer des observations dont les propriétés ont varié par rapport à une observation type [Zouari 2002].

I.3. Processus de la RDF

Les informations issues du monde réel (via le capteur) et fournies au système de RDF sont généralement trop volumineuses et peu pertinentes. Le processus de RDF consiste à une réduction progressive et sélective de l'information [Lepage 2004], la *Figure 1.1* représente les étapes de traitement d'un processus de RDF typique.

I.3.1. Acquisition de données

Le monde physique qui nous entoure est considéré comme un espace analogique de dimension n appelé l'espace de formes F . C'est celui qui est présenté dans sa forme la plus primaire, c'est-à-dire dont nous devons déterminer les caractéristiques les plus apparentes avant l'étape du codage. L'opération de codage consiste en une conversion numérique du monde physique continu vers un monde discret.

I.3.2. Prétraitement de données

La phase de prétraitement permet de sélectionner les informations nécessaires à l'application à partir de l'espace de représentation de la forme F [Barrat 2010]. Cette sélection passe par la suppression du bruit dû aux conditions d'acquisition et de la qualité de la forme, par la normalisation des données ainsi par la suppression des redondances.

Le prétraitement de données concerne le filtrage et la transformation des informations qui permettent d'obtenir des données plus adaptées à l'extraction de caractéristiques informatives. Cette étape peut être exécutée à l'aide de plusieurs méthodes [Trouilhet 1994] telles que :

l'analyse de Fourier standard, les méthodes Temps-Fréquence (exemple : filtre de Gabor), l'analyse des ondelettes, etc. Lorsque les données sont obtenues, des caractéristiques structurelles peuvent être obtenues en utilisant une technique de segmentation.

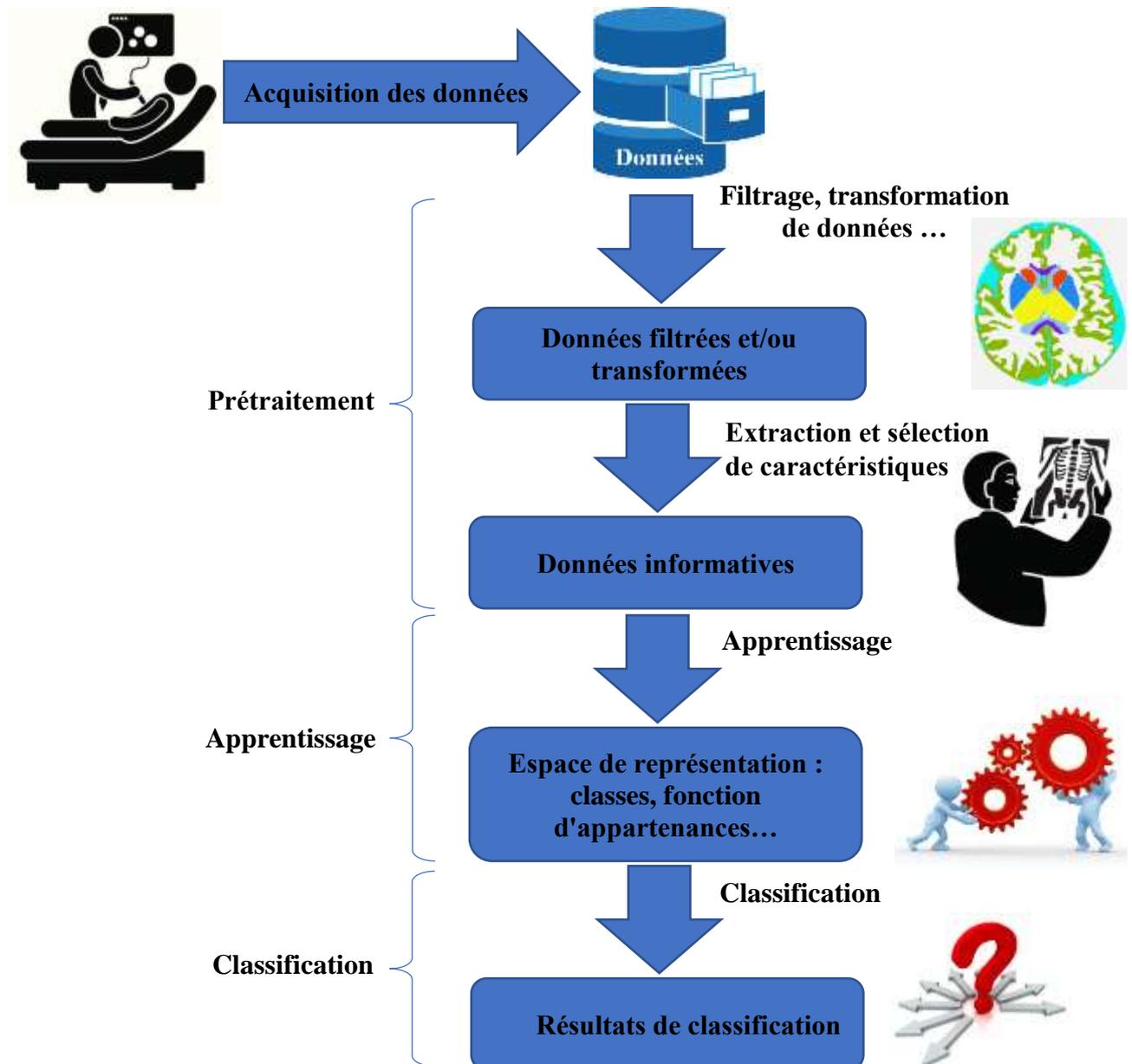


Figure I.1: Les étapes principales d'une approche de reconnaissance de forme typique [Hartert 2010].

Les techniques de segmentation peuvent utiliser différents types de primitives [Colomer 2002] qui sont connectées en maintenant la structure de chaque échantillon. En conservant la connexion entre les différentes parties d'un signal, une technique de segmentation permet d'obtenir une signature, qui est une représentation simplifiée et caractéristique de chaque signal

I.3.3. Extraction et sélection de caractéristiques

La première étape de l'analyse des données traitées permet d'obtenir des caractéristiques statistiques (des valeurs scalaires) ou structurelles. Les caractéristiques statistiques fournissent des informations spatiales telles que le nombre de pics présents dans un signal, l'écart-type des données, la moyenne quadratique, la valeur maximale ou minimale, le coefficient d'aplatissement (Kurtosis), etc. Lors de la deuxième phase, la sélection des caractéristiques les plus informatifs sera effectué. Généralement un grand nombre de caractéristiques est calculé à partir des données recueillies par les capteurs du système. C'est pourquoi, il est nécessaire d'extraire et/ou de sélectionner les caractéristiques les plus utiles pour représenter la forme [Benzecri 1979], c'est-à-dire ceux qui permettront de discriminer au maximum les classes de formes. Les méthodes de sélection des paramètres choisissent le sous-ensemble de paramètres les plus informatifs. Les méthodes d'extraction des paramètres créent un sous-ensemble de nouveaux paramètres par combinaison des paramètres existants. La réalisation de la phase de sélection de caractéristiques, se fait par l'utilisation des techniques d'analyse des données [Benzecri 1979] telles que l'Analyse en Composantes Principales (ACP), les algorithmes génétiques, la mesure discriminante de Fisher, etc. Plus les caractéristiques du système permettent de bien discriminer les formes analysées, plus les résultats de classification sont considérés bons. L'ensemble des caractéristiques trouvés par ces méthodes représente les paramètres qui permettent de caractériser chaque forme. Lorsque les données issues de l'observation du fonctionnement d'un système sont représentées par des paramètres statistiques, elles sont transformées en formes, c'est-à-dire en points, dans l'espace de représentation. Les groupes de formes similaires sont appelés classes. Si les caractéristiques de forme sont bien déterminées, les classes sont bien discriminées et elles sont situées dans différentes régions de l'espace de représentation. Chaque classe est associée à un mode de fonctionnement (normal ou défaillant). Ces formes, avec leurs assignements à une classe, constituent l'ensemble d'apprentissage. Elles sont représentées par des caractéristiques ou attributs, c'est à dire des points, dans l'espace de représentation.

L'objectif de l'extraction et de la sélection de caractéristiques est d'identifier les caractéristiques importantes pour la discrimination entre classes. Après avoir choisi le meilleur ensemble de caractéristiques, il s'agit de réduire la dimensionnalité de l'ensemble des caractéristiques en trouvant un nouvel ensemble, plus petit que l'ensemble original, qui néanmoins, contient la plupart de l'information.

- ✓ Dans les systèmes de reconnaissance des caractères, les caractéristiques utilisables peuvent venir des squelettes ou des contours des segments, elles peuvent également venir de la densité des pixels, des moments, des lieux caractéristiques ou des transformées mathématiques.
- ✓ Dans les applications liées à l'analyse de texture et de densité des niveaux de gris telles que l'imagerie médicale et l'analyse des scènes, les caractéristiques utilisables peuvent venir de la matrice de cooccurrence des niveaux de gris (GLCM) ou de la méthode Run difference matrix (RDM), des descripteurs de Fourier, aussi bien que de diverses primitives structurelles.
- ✓ Dans l'analyse et la reconnaissance de formes d'ondes telles que les signaux électroencéphalographique, électrocardiographique, sismique ou audible aussi bien que les images ayant une forme de courbes, les caractéristiques utilisables généralement peuvent venir du spectre de puissance, de fonctions d'approximation, et de plusieurs types de segments de traits structurels.

I.3.4. Apprentissage et classification

La classification par apprentissage automatique est introduite dans ses objectifs et ses principes. En s'appuyant sur les approches qui ont montré leurs efficacités (Les arbres de décision, Les réseaux de neurones artificiels, les séparateurs à vaste marge, k-moyennes, k-plus proches voisins, etc.), aussi, les grands principes de l'apprentissage automatique (supervisées, non-supervisées et semi-supervisées) pour la classification automatique sont présentés dans cette section.

I.3.4.1. Méthodes supervisées

Dans le cadre de la reconnaissance de formes, une méthode de classification supervisée, dite aussi discrimination est la tâche qui consiste à discriminer des données, de façon supervisée (c'est à dire avec l'aide préalable d'un expert) ; les méthodes de classification supervisée tentent d'apprendre, à partir d'objets étiquetés (pour lesquelles la classe est connue), constituant un échantillon d'apprentissage, une fonction de classification. Cette fonction permettra d'associer une valeur de classe à chaque objet non étiqueté.

Généralement, on passe par une première étape dite d'apprentissage où il s'agit d'apprendre une règle de classification à partir de données annotées (étiquetées) par l'expert et donc pour lesquelles les classes de sortie sont connues, afin de prédire les classes de nouvelles données,

pour lesquelles (on suppose que) les données sont inconnues. Parmi les méthodes d'apprentissages supervisés :

I.3.4.1.1. Méthodes basées sur le concept de similarité

L'un des moyens le plus simple de faire de la classification est de définir une fonction de distance entre les vecteurs caractéristiques, et d'affecter chaque image d'entrée inconnue à la classe dont le barycentre est le plus proche de l'image requête, selon la fonction de distance définie [Barrat 2010].

Les notions de distance et de similarité sont proches, dans le sens où elles permettent toutes deux d'identifier si deux objets sont différents ou non. Cependant, à la différence des mesures de similarité, les mesures de distance accordent une valeur maximale à deux objets complètement différents et minimale (tend vers zéro) à deux objets identiques. De plus, une mesure D peut être considérée comme une distance sur un ensemble E (où E est un \mathbb{R} -espace vectoriel), si et seulement si D est une application de $E \times E$ dans \mathbb{R}^+ telle que :

- $D(a, b) = 0 \Leftrightarrow a = b$ (réflexivité),
- $\forall a, b \in E, D(a, b) = D(b, a)$ (symétrie),
- $\forall a, b, c \in E, D(a, b) + D(b, c) \geq D(a, c)$ (inégalité triangulaire)

Il existe plusieurs mesures de similarité et de distances dans la littérature (Euclidienne, Minkowski, Mahalanobis ...). Certaines sont des distances, c'est-à-dire des mesures qui ont les propriétés de non-négativité, réflexivité, symétrie.

La distance classique la plus utilisée est la distance de Minkowski (ou L_p -norme) qui se base sur la distance euclidienne, Il existe cependant des mesures spécifiques aux histogrammes: L'intersection d'histogrammes [Swain 1991] et La distance entre histogrammes cumulés [Stricker 1995].

L'inconvénient des mesures telles que la distance Euclidienne ou l'intersection d'histogrammes est qu'elles comparent les composantes des vecteurs une par une, sans prendre en compte les autres composantes.

I.3.4.1.2. k plus proches voisins (kppv)

La méthode de classification k plus proche voisins (KPPV) [Cover 1967] est la méthode non paramétrique la plus simple. La méthode des KPPV est une méthode de classification géométrique très utilisée en reconnaissance de formes, en raison de sa simplicité et de sa robustesse. Les caractéristiques sont exploitées dans un espace métrique de représentation, généralement \mathbb{R}^n muni de la distance Euclidienne.

Elle permet de reconnaître les formes lorsque la distribution des classes n'est pas convexe [Taconet 2006]. L'algorithme de base est facile à écrire et à programmer. En contrepartie, elle nécessite un grand volume de données d'apprentissage, ce qui veut dire un grand nombre de points d'apprentissage à stocker et à examiner, autrement dit, des ressources mémoire élevées et un temps d'exécution important dans une implémentation brute de l'algorithme.

Depuis plusieurs décennies, les chercheurs ont essayé d'améliorer la méthode, d'une part en réduisant le volume des données nécessaires et en optimisant les algorithmes pour réduire le temps de traitement. Keller et al. [Keller 1985] proposent une méthode de lissage flou des étiquettes, afin de filtrer les incertitudes. Une vote majoritaire proposé par Dudani [Dudani 1976], voisinage par classe de Hattori et al. [Hattori 1999], paramétrique de Arif [Arif 2005]. La *Figure I.2* représente une illustration d'un problème de classification à 2 classes (A et B) avec un algorithme *KPPV* où $k = 1$ puis $k = 6$. Un ensemble de points sont répartis dans le plan. Les données étiquetées de classe A (respectivement de classe B) sont représentées par les points noirs (respectivement par les points blancs). Une observation non étiquetées (individu à classer), est représentée par le forme étoile x .

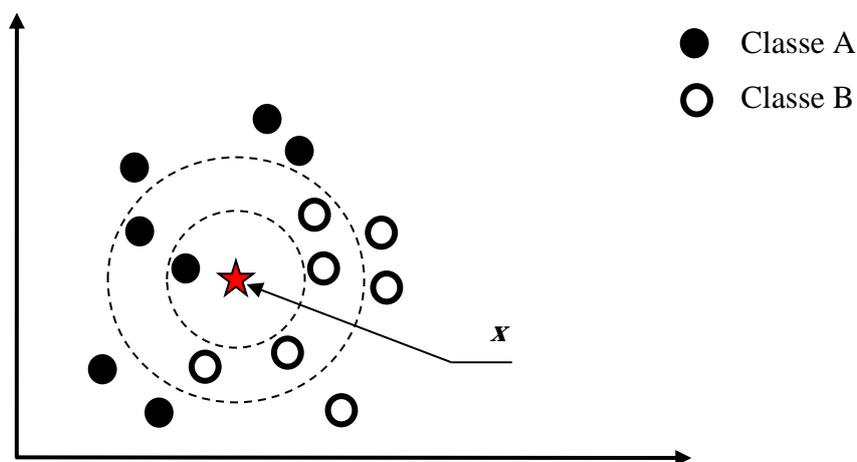


Figure I.2: Exemple de classification en utilisant la méthode *kppv* avec ($k=1$ et $k=6$)

Afin de classer l'individu x dans un voisinage de $k = 1$ point, on recherche le plus proche voisin de x . Le petit cercle noir entoure le point à classer et son plus proche voisin. Le plus proche voisin de x est un point noir, x sera donc affecté à la classe A.

Considérons maintenant le même problème, mais avec un voisinage de $k = 6$ points. Afin de classer l'individu x , on recherche les 6 points les plus proches de x . le grand cercle noir qui entoure l'individu à classer x et ses six plus proches voisins. Parmi les 6 points les plus proches

de x , il y en a 4 points blancs et 2 points noirs. Un vote majoritaire est effectué et l'individu x sera donc affecté à la classe des points blancs (Classe B).

L'avantage principal de la méthode KPPV est sa simplicité et le fait qu'elle ne nécessite pas d'apprentissage. L'introduction de nouveaux individus permet d'améliorer la qualité de classification sans nécessiter la reconstruction d'un modèle. C'est une différence majeure avec des méthodes telles que les réseaux de neurones et les arbres de décision. Cette méthode est également utilisable même en présence de caractéristiques de grande dimension.

L'inconvénient majeur de cette méthode est qu'elle est lourde car chaque nouvel individu à classer est comparé (sur la base de ses caractéristiques) à tous les individus stockés (sauf si un schéma d'indexation a été appliqué, auquel cas la comparaison se fera uniquement avec les individus d'une région). D'autre part, la performance des KPPV est très dépendante de la mesure de distance appliquée et de la valeur k utilisée.

I.3.4.1.3. Réseaux de neurones

Un réseau de neurones (Artificial Neural Network) est un modèle de calcul dont la conception est schématiquement inspirée du fonctionnement de vrais neurones. Les réseaux de neurones sont généralement optimisés par des méthodes d'apprentissage de type statistique, si bien qu'ils sont placés d'une part dans la famille des applications statistiques, qu'ils enrichissent avec un ensemble de paradigmes permettant de générer de vastes espaces fonctionnels, souples et partiellement structurés, et d'autre part dans la famille des méthodes de l'intelligence artificielle qu'ils enrichissent en permettant de prendre des décisions s'appuyant d'avantage sur la perception que sur le raisonnement logique formel.

Ce sont les deux neurologues *Warren McCulloch* et *Walter Pitts* [McCulloch 1943] qui ont mené les premiers travaux sur les réseaux de neurones. Ils constituèrent un modèle simplifié de neurone biologique communément appelé neurone formel. Ils montrèrent également théoriquement que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes.

La fonction des réseaux de neurones formels à l'instar du modèle vrai est de résoudre divers problèmes. A la différence des méthodes traditionnelles de résolution informatique, on ne doit pas construire un programme pas à pas en fonction de la compréhension de celui-ci. Les paramètres les plus importants de ce modèle sont les coefficients synaptiques. Ce sont eux qui construisent le modèle de résolution en fonction des informations données au réseau. Il faut donc trouver un mécanisme, qui permet de les calculer à partir des grandeurs acquises du problème, c'est le principe fondamental de l'apprentissage. Dans un modèle de réseau de

neurones formels, apprendre, c'est d'abord calculer les valeurs des coefficients synaptiques en fonction des exemples disponibles. La représentation graphique d'un réseau de neurones artificiel est donnée par la Figure I.3.

Le neurone calcule la somme de ses entrées puis cette valeur passe à travers la fonction d'activation pour produire sa sortie. La fonction d'activation (ou fonction de seuillage) sert à introduire une non linéarité dans le fonctionnement du neurone.

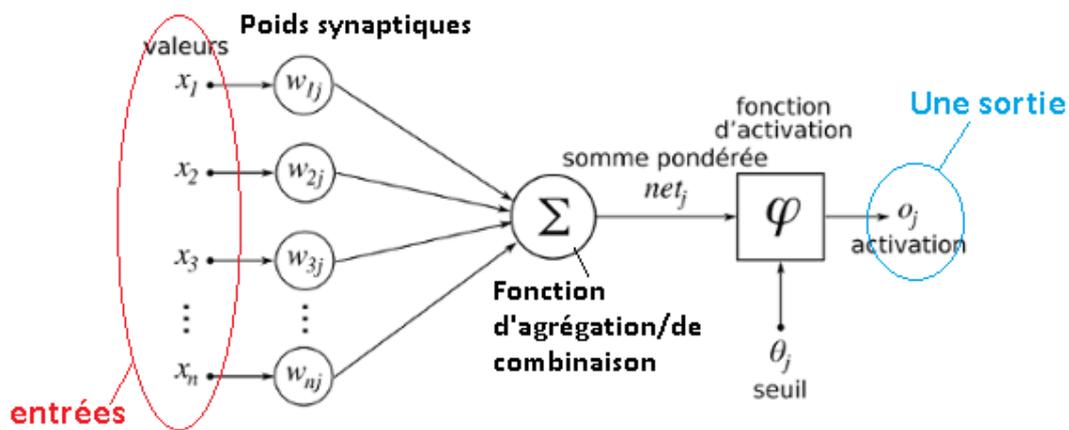


Figure I.3: Représentation d'un neurone artificiel [Tilmanne 2009].

Les fonctions de seuillage présentent généralement trois intervalles :

1. en dessous du seuil, le neurone est non-actif (souvent dans ce cas, sa sortie vaut 0 ou 1),
2. au voisinage du seuil, une phase de transition,
3. au-dessus du seuil, le neurone est actif (souvent dans ce cas, sa sortie vaut 1).

En général, un réseau de neurone est composé d'une succession de couches dont chacune prend ses entrées à partir des sorties de la couche précédente. Chaque couche i est composée de N_i neurones, prenant leurs entrées sur les N_{i-1} neurones de la couche précédente. A chaque synapse est associé un poids synaptique, de sorte que les N_{i-1} sont multipliés par ce poids, puis additionnés par les neurones de niveau i , ce qui est équivalent à multiplier le vecteur d'entrée par une matrice de transformation. Mettre l'une derrière l'autre, les différentes couches d'un réseau de neurones, reviendrait à mettre en cascade plusieurs matrices de transformation et pourrait se ramener à une seule matrice, produit des autres, s'il n'y avait à chaque couche, la fonction de sortie qui introduit une non linéarité à chaque étape. Ceci montre l'importance du choix judicieux d'une bonne fonction de sortie : un réseau de neurones dont les sorties seraient linéaires n'aurait aucun intérêt.

Plusieurs types de réseaux de neurones ont été reportés dans la littérature, notamment le perceptron proposé par Rosenblatt [Rosenblatt 1958], les cartes autoorganisatrices de Kohonen [Kohonen 1989] et les réseaux basés sur le modèle de Hopfield [Hopfield 1982], etc.

Les réseaux de neurones fonctionnent en répartissant les valeurs des variables dans des automates (les neurones). Ces unités sont chargées de combiner entre elles leurs informations pour déterminer la valeur du paramètre de discrimination. C'est de la connexion de ces unités entre elles qu'émerge la capacité de discrimination du RNA. Chaque neurone reçoit des informations numériques en provenance de neurones voisins ; à chacune de ces valeurs est associée un poids représentatif de la force de la connexion. Chaque neurone effectue localement un calcul dont le résultat est transmis ensuite aux neurones avals.

La famille de réseau majoritairement employé est le perceptron multi-couches (PMC). À lui seul ce type de réseau recouvre plus de 95 % des applications scientifiques et industrielles.

Le PMC est un modèle de réseau à propagation par couche (Figure I.4). Les neurones y sont organisés en couches successives : une couche d'entrée, une couche de sortie et entre les deux une ou plusieurs couches intermédiaires, appelées aussi couches cachées. Il n'existe pas de connexion entre les neurones d'une même couche, en revanche tout neurone d'une couche est connecté à tous les neurones de la couche suivante. La « couche » d'entrée n'est pas une réelle couche de neurones car elle se contente de coder les variables d'observation. La couche de sortie code la variable de discrimination. Les valeurs d'activité des neurones sont propagées dans le réseau, de l'entrée vers la sortie, sans retour arrière. La présence d'une couche cachée permet de modéliser des relations non linéaires entre les entrées et la sortie. En théorie une seule couche cachée suffit, mais le fait de disposer d'une seconde couche cachée permet de modéliser plus facilement une fonction de discrimination non continue. En pratique, la plupart des problèmes sont résolus avec un ou deux niveaux, trois au maximum.

L'objectif général d'un RNA est de trouver la configuration des poids de connexion entre neurones pour qu'il associe à chaque configuration d'entrée, une réponse adéquate. L'utilisation d'un RNA se fait en deux temps. Tout d'abord une phase d'apprentissage qui est chargée d'établir des valeurs pour chacune des connexions du réseau, puis une phase d'utilisation proprement dite, où l'on présente au réseau une entrée et où il nous indique en retour « sa » sortie calculée.

Dans le cas du PMC, on utilise un apprentissage supervisé. Les valeurs des poids de connexion sont créées tout d'abord au hasard et le système cherche par itérations successives à obtenir une modélisation des données. A chaque étape, une entrée est présentée au réseau, il propage ces

valeurs vers les neurones de sortie. Cette sortie calculée est comparée avec la réponse attendue et le système modifie les poids en conséquence. Cette altération des connexions est obtenue par l'algorithme de rétropropagation du gradient d'erreur. Ce calcul est chargé de rétropropager dans le réseau les erreurs constatées sur les sorties. En théorie, on ne peut jamais être sûr que cet algorithme finisse par déterminer un ensemble de poids convenable pour tous les couples d'entrées-sorties. En pratique, on ne construit pas un seul RNA, mais plusieurs modèles en jouant sur les paramètres de cet algorithme, et en cherchant à obtenir un modèle qui colle au mieux aux données.

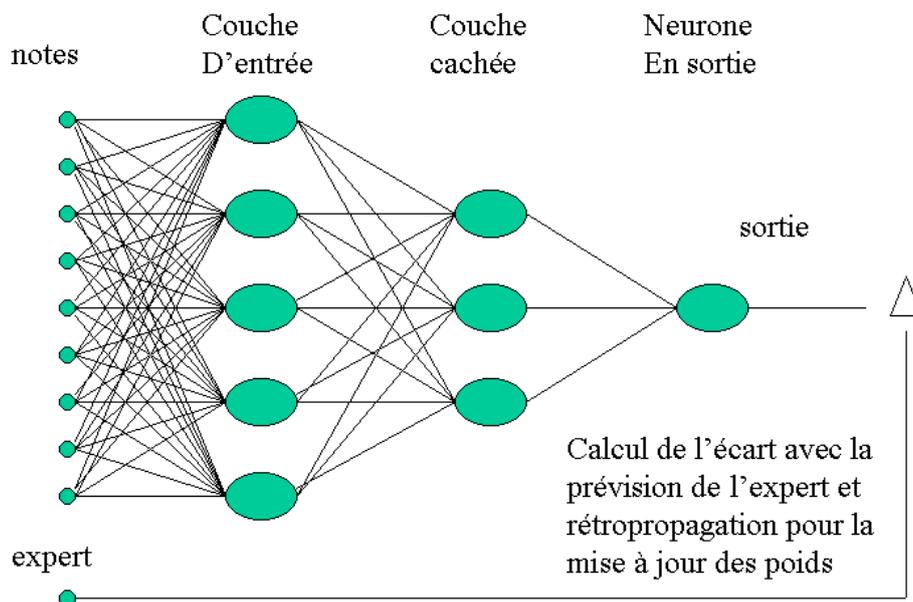


Figure I.4: Schéma d'un perceptron multicouche (PMC) [Tilmanne 2009].

Grâce à leur capacité de classification et de généralisation, les réseaux de neurones sont généralement utilisés dans des problèmes de nature statistique, tels que la classification automatique, reconnaissance de caractère, approximation d'une fonction inconnue ... etc.

Les réseaux de neurones sont célèbres pour leurs bonnes performances, en termes de taux et de vitesse de classification. Ils ont aussi l'avantage de fonctionner en présence de données manquantes ou fausses. Ces avantages font la popularité des approches neuronales en reconnaissance de formes [Barrat 2010].

L'inconvénient majeur des réseaux de neurones est l'absence de méthode systématique permettant de définir la meilleure architecture du réseau et le nombre de neurones à placer pour chaque couche cachée, il est difficile de choisir l'architecture et de régler les paramètres d'apprentissage. En effet, le résultat de l'apprentissage est un réseau défini par une fonction

d'activation et un très grand nombre de poids à valeurs réelles. Ces poids sont difficilement interprétables. De plus les réseaux de neurones ont un côté « boîte noire » : les résultats de la classification sont difficilement interprétables.

I.3.4.1.4. Arbres de décision

Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre). C'est en particulier le cas pour l'aide au diagnostic médical où le médecin doit pouvoir interpréter les raisons du diagnostic. Les arbres de décision répondent à cette contrainte car ils représentent graphiquement un ensemble de règles et sont aisément interprétables. Pour les arbres de grande taille, la procédure globale peut être difficile à appréhender, cependant, la classification d'un élément particulier est toujours compréhensible.

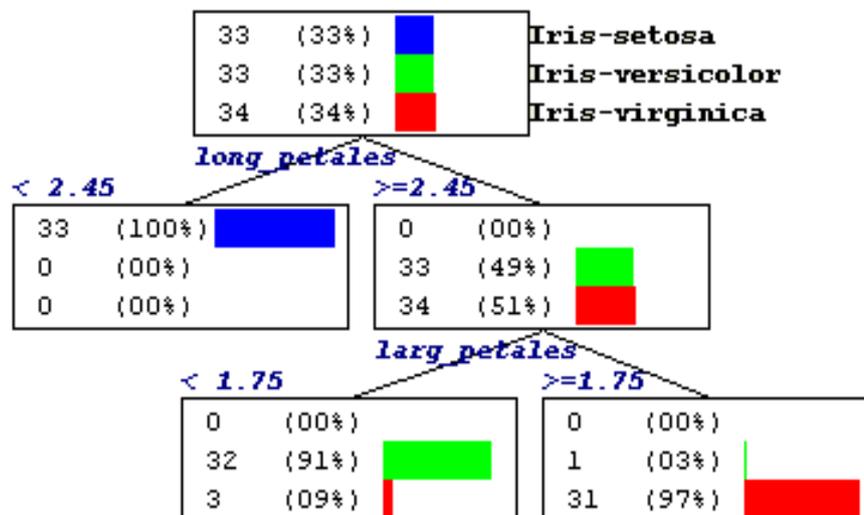


Figure I.5: Un exemple d'arbre de décision pour la classification de 100 individus pris de la base IRIS. [Rakotomalala 2005].

Les premiers algorithmes de classification par arbres de décision sont anciens. Les travaux les plus marquants sont :

- ✓ CHAID (CHi-squared Automatic Interaction Detector) développée par Kass et al. [Kass 1980].
- ✓ La méthode « itérative Dichotomiser 3 » (ID3) et son successeur (C4.5) réalisées par Quinlan et al. [Quinlan 1986] [Quinlan 1992].
- ✓ Exhaustive CHAID [Biggs 1991].
- ✓ MARS (Multivariate adaptive regression splines) [Friedman 1991]
- ✓ CART (Classification and Regression Tree) [Breiman 1996].

L'idée de l'apprentissage des arbres de décision commence par la division récursive des éléments de la base d'apprentissage par des tests définis à l'aide des caractéristiques jusqu'à ce que l'on obtienne des sous-ensembles d'exemples ne contenant (presque) que des exemples appartenant tous à une même classe.

Les méthodes vont différer par les choix effectués pour ces différents opérateurs, c'est-à-dire sur le choix d'un test (par exemple, utilisation du gain et de la fonction entropie) et le critère d'arrêt (quand arrêter la croissance de l'arbre, soit quand décider si un nœud est terminal). Tous ces algorithmes se distinguent par le ou les critères de segmentation utilisés, par les méthodes d'élagages implémentées, par leur manière de gérer les données manquantes dans les prédicteurs. Dans toutes les méthodes, on trouve les trois opérateurs suivants :

1. **Décider si un nœud est terminal**, c'est-à-dire décider si un nœud doit être étiqueté comme une feuille. Par exemple : tous les exemples sont dans la même classe, il y a moins d'un certain nombre d'erreurs.
2. **Sélectionner un test à associer à un nœud**. Par exemple : aléatoirement, utiliser des critères statistiques.
3. **Affecter une classe à une feuille**. On attribue la classe majoritaire sauf dans le cas où l'on utilise des fonctions coût ou risque.

L'avantage de cette méthode est qu'elle permet à l'utilisateur d'interpréter le résultat de la classification, puisque le processus de classification est représenté graphiquement. De plus, une fois l'arbre construit, le temps de décision est réduit. Enfin, cette approche est non paramétrique. Ces avantages rendent les arbres de décisions adaptés aux problèmes de reconnaissance de formes. D'une autre part, les performances des arbres de décision tendent à se dégrader lorsque le nombre de classes devient trop important. De plus, les variables situées en haut de l'arbre ont beaucoup de poids. En effet, elles influencent grandement le reste de la construction. Ceci peut conduire à changer complètement les prédictions en cas de modification d'une de ces variables, ce qui fait des arbres de décision une méthode d'apprentissage plutôt instable. Enfin, comme pour les réseaux de neurones, l'apprentissage n'est pas incrémental et par conséquent, si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage pour s'adapter à cette évolution.

I.3.4.1.5. Séparateurs à vaste marge (SVM)

Les SVM (Support Vector Machine) souvent traduit par l'appellation de Séparateur à Vaste Marge ou Les machines à vecteurs de support sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination.

L'idée des hyperplans à marge maximale a été explorée dès 1963 par Vladimir Vapnik [Vladimir 1963], et en 1973 par Richard Duda et Peter Hart dans leur livre "*Pattern classification and scene analysis*" [Duda 1973].

Les SVM ont été rapidement adoptés pour leur capacité à travailler avec des données de grandes dimensions, le faible nombre d'hyper paramètres, leurs garanties théoriques, et leurs bons résultats en pratique.

Les SVM ont été appliqués à de très nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finance...).

La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Ce choix est justifié par la théorie de Vapnik-Chervonenkis [Vladimir 1963], qui montre que la frontière de séparation de marge maximale possède la plus petite capacité [Hearst 1998]. Le problème est de trouver cette frontière séparatrice optimale, à partir d'un ensemble d'apprentissage.

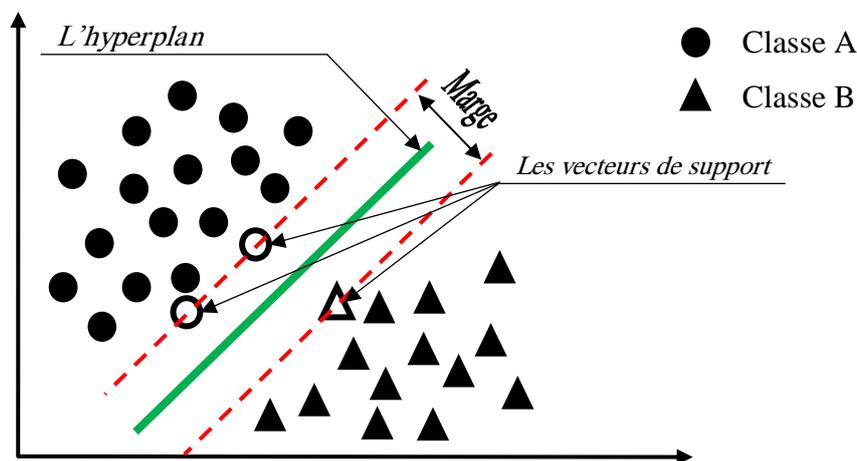


Figure I.6: La représentation d'un Classifieur de type SVM.

*La marge (la distance minimale aux objets d'apprentissage est maximale)
Les objets les plus proches de l'hyperplan s'appellent les vecteurs de support.*

Le principe des SVM est assez simple. Soit l'échantillon d'apprentissage où chaque objet appartient à l'une des deux classes notées A et B et supposées linéairement séparables, l'algorithme recherche la séparation qui maximise la « Marge ». La marge peut être définie

comme la distance euclidienne entre la surface de séparation (Hyperplan) et le point le plus proche de l'ensemble d'apprentissage [Boser 1992]. Une extension de l'algorithme, dite « à marge douce », propose de traiter le cas où les données ne sont pas séparables en acceptant que certains points se situent du mauvais côté de la surface de décision mais en ajoutant une pénalité en fonction de la distance des points à la marge [Cortes 1995].

I.3.4.1.6. La classification Bayésienne

La classification naïve bayésienne est l'une des méthodes les plus simples en apprentissage supervisé basée sur le théorème de Bayes avec une forte indépendance (naïve) des hypothèses [Michalski 1983]. Cette méthode appartient à la famille des classifieurs linéaires, dont le rôle est de catégoriser dans des classes les échantillons d'apprentissage qui ont des propriétés similaires. En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique déterminant l'appartenance d'un individu à une catégorie est indépendante de l'existence d'autres caractéristiques. Par exemple, un animal sera considéré avec un bon degré de confiance comme un chien s'il marche sur quatre pattes et qu'il aboie, en ignorant totalement la possibilité que ces caractéristiques puissent être corrélées.

Notons que $P(C_i)$ est la probabilité a priori d'appartenance à une classe C_i , $P(x)$ est la probabilité d'observer un individu ayant le vecteur de caractéristique x et $P(x/C_i)$ est la probabilité d'observer le vecteur x sachant que la classe est C_i . La règle de Bayes permet alors de calculer la probabilité a posteriori de la classe C_i quand le vecteur de x est observé :

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_j P(x|C_j)P(C_j)} \quad \text{Équation I-1}$$

Dans la pratique, les probabilités de chaque classe $P(C_i)$ et les distributions $P(x/C_i)$ doivent être préalablement calculées à partir d'un échantillon d'apprentissage. Le vecteur x est assigné à la classe C_i si $\forall j \neq i; P(C_i/x) > P(C_j/x)$.

L'avantage des méthodes Bayésiennes est qu'elles sont simples et efficaces, malgré les hypothèses d'indépendance entre les variables, et les hypothèses sur les distributions de probabilités des variables, qui ne sont pas toujours vérifiées dans la pratique. En outre, le raisonnement Bayésien est applicable sur des données bruitées ou manquantes. Il associe des probabilités aux prédictions, ce qui est utile dans les nombreux domaines où les connaissances sont incertaines. Dans ce cas, le résultat de la classification est facilement interprétable.

L'inconvénient de cette méthode est qu'elle nécessite des probabilités dont la détermination requiert plusieurs connaissances a priori ou une grande quantité de données. Elle nécessite aussi un temps de calcul relativement élevé pour déterminer l'hypothèse optimale dans un cas général. De plus un modèle bayésienne probabiliste n'est parfois pas un concept intuitif pour les experts du domaine.

I.3.4.2. Méthodes non supervisées

Contrairement à l'apprentissage supervisé, dans l'apprentissage non-supervisé (clustering) il n'y a pas de prédiction qui explicite les étiquettes. L'utilisation de ce type d'algorithme permet de trouver des structures, des dépendances entre descripteurs.

Les classes (clusters) sont formées par regroupement des objets qui ont certaines caractéristiques en commun.

Le clustering est un outil important pour l'analyse de données. Il vise à trouver les structures intrinsèques des données en les organisant en groupes homogènes et distincts (les clusters). Les objets dans un même cluster doivent être similaires entre eux et différents des objets des autres clusters [Hartert 2010].

I.3.4.2.1. Méthodes basées sur le concept de similarité

Le moyen le plus simple de faire du clustering est de définir une fonction de distance entre les vecteurs caractéristiques. Les partitions sont constituées en minimisant les distances entre les vecteurs d'un même partition (cluster), et en maximisant les distances entre les vecteurs de clusters différents [Hartert 2010].

L'efficacité de ces méthodes dépend de la mesure de distance (ou similarité), utilisée. Il n'est pas évident de trouver une mesure de distance qui permette de donner plus d'importance à un attribut qu'à un autre. Enfin, avec ces méthodes, le nombre de clusters possibles est difficile à contrôler.

I.3.4.2.2. Méthode K – moyennes

L'algorithme des k-moyennes, plus communément appelé k-means, est une méthode couramment utilisée en analyse de données pour la classification automatique de données. Le terme « k-means » a été utilisé pour la première fois par James MacQueen et al. [MacQueen 1965].

Étant donné un ensemble d'individus $I = \{I_1, I_2, \dots, I_n\}$, on cherche à partitionner les n individus en k classes, souvent appelés clusters, $C = \{C_1, C_2, \dots, C_k\}$ ($k \leq n$) en minimisant la distance entre les individus à l'intérieur de chaque classe.

On considère la distance d'un individu à la moyenne des individus de son cluster; la fonction à minimiser est la somme des carrés de ces distances.

$$\text{Min} \sum_{i=1}^k \sum_{P_j \in C_i} \|P_j - \mu_i\|^2 \quad \text{Équation I-2}$$

Où μ_i est la moyenne des individus de la classe C_i .

Dans la méthode des "k-moyennes", le choix des k centres initiaux s'effectue aléatoirement à partir de la population des individus à classifier. La partition des classes est modifiée avec chaque affectation d'un individu i .

Les individus sont géométriquement représentés dans l'espace vectoriel \mathbf{P} muni d'une distance notée d . L'algorithme de la méthode des "k-means" se déroule comme suit :

Étape 1 : On choisit par un tirage aléatoire k individus parmi n individus composant l'ensemble I . Ces k centres notés $\{C_1^0, C_2^0, C_3^0, \dots, C_k^0\}$.

Chaque individu i de l'ensemble I est affecté à une classe. Chacune de ces classes est localisée par son centre. La procédure d'affectation est la suivante :

L'individu i est affecté à la classe m (C_m) si et seulement si

$$d(i, C_m^0) = \inf_{j \in \{1..k\}} \{d(i, C_j^0)\} \quad \text{Équation I-3}$$

Après avoir affecté tous les individus on obtient k classes notées $\{C_1, C_2, C_3, \dots, C_k\}$. de centres respectifs $\{C_1^0, C_2^0, C_3^0, \dots, C_k^0\}$.

Étape 2 : En considérant les k classes obtenues à l'étape 1 on calcule ses centres de gravité.

On obtient donc k nouveaux centres notés $\{C_1^1, C_2^1, C_3^1, \dots, C_k^1\}$.

On utilise la même règle d'affectation qu'à l'étape 1, on obtient k nouvelles classes

$\{C_1, C_2, C_3, \dots, C_k\}$ de centres respectifs $\{C_1^1, C_2^1, C_3^1, \dots, C_k^1\}$.

Étape H : On détermine k nouvelles classes en calculant les centres de gravité des classes obtenues à l'étape (H-1).

La règle d'affectation reste la même qu'à l'étape précédente et on obtient par la suite une nouvelle typologie de l'ensemble I : de centres $\{C_1^H, C_2^H, C_3^H, \dots, C_k^H\}$

L'arrêt de l'algorithme

L'arrêt de l'algorithme de la méthode des "k-means" se fait lorsque deux itérations successives conduisent à une même partition. Ou un nombre maximal d'itération atteint.

La méthode K-means s'avère plutôt rapide et converge souvent vers un optimum local [Filippone 2008]. Par contre, elle est sensible au bruit et aux valeurs aberrantes. Cette méthode est basée sur le concept de similarité, elle dépend largement de la mesure de distance appliquée. De plus, elle requiert que l'on puisse calculer une moyenne des observations pour chaque cluster. En outre, les K-means nécessitent de fixer à l'avance le nombre de classes.

I.3.4.2.3. Cartes de Kohonen (Self-Organizing Maps)

Les cartes de Kohonen [Kohonen 1989] aussi appelées carte auto-organisatrices (Self Organizing Maps : SOM) sont des réseaux à compétition. Ces cartes permettent de représenter sur un petit nombre de dimensions la structure présente dans des données de haute dimensionnalité. C'est une technique de prétraitement fort utile car elle permet de diminuer l'espace de représentation. La carte topologique est composée essentiellement d'une couche compétitive de neurones de sortie. Ces neurones sont alimentés par une couche de neurones d'entrée. Dans ce type de réseau, les neurones sont connectés entre eux selon une notion de voisinage et non selon une notion de couche.

L'avantage principale des Cartes de Kohonen est ses capacités de traiter les données de grande dimension. Par contre, la convergence n'est pas assurée et le temps de convergence est long [Oja 2002].

I.3.4.3. Autres types d'apprentissage

Deux autres types d'apprentissage, à savoir l'apprentissage semi-supervisé et l'apprentissage par renforcement. En effet, l'apprentissage semi-supervisé utilise un ensemble de données étiquetées et non-étiquetés. Il se situe ainsi entre l'apprentissage supervisé qui n'utilise que des données étiquetées et l'apprentissage non-supervisé qui n'utilise que des données non-étiquetées, car il permet de traiter un grand nombre de données sans avoir besoin de toutes les étiquètes [Blum 1998]. Il a été démontré que l'utilisation de données non-étiquetées, en combinaison avec des données étiquetées, permet d'améliorer significativement la qualité de l'apprentissage. L'apprentissage par renforcement consiste à apprendre, à partir d'expériences, ce qu'il convient de faire en différentes situations, de façon à optimiser une récompense quantitative au cours du temps [Khamassi 2005]. Ce type d'apprentissage est utile dans le cas d'apprentissage interactif, comme les techniques de bouclage de pertinence.

I.4. La vision par ordinateur

La vision naturelle permet aux humains de percevoir et de comprendre le monde qui les entoure, la vision par ordinateur vise à reproduire l'effet de la vision humaine en percevant électroniquement l'environnement [Sonka 2014].

Donner à l'ordinateur la possibilité de voir n'est pas une tâche facile, en effet, nous vivons dans un monde à trois dimensions (3D), et quand les ordinateurs tentent d'analyser des objets dans l'espace 3D, les capteurs visuels disponibles comme les caméras de télévision, donnent généralement des images bidimensionnelles (2D), et cette projection en un nombre inférieur de dimensions encourt une énorme perte d'informations. La majorité des techniques de vision par ordinateur se basent sur les résultats issus des méthodes de mathématiques, reconnaissance de formes, intelligence artificielle, psycho-physiologie, informatique, électronique, et d'autres disciplines scientifiques.

La création ou l'utilisation des algorithmes ainsi que les dispositifs d'analyse d'image numérique, est conditionnée par la prise en compte des principes de la perception humaine de l'image. Lorsqu'un humain analyse une image, il exprime l'information contenu au sein de cette dernière en utilisant des descripteurs faciles à percevoir, se sont des descripteurs psychophysiques tels que la couleur, le contraste, les contours, la forme, la texture...etc.

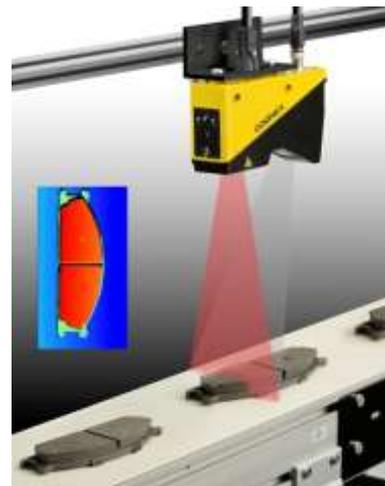
L'analyse de la texture est un domaine important pour la vision par ordinateur. En effet, la majorité des surfaces naturelles présentent une texture et un système performant de vision doit être capable de distinguer la texture de l'environnement pour lequel il a été conçu.

I.4.1. Applications industrielles de la vision artificielle

Les techniques de la vision artificielle sont des disciplines relativement jeunes (~années 60) et qui évoluent rapidement. Elles sont en plein expansion et donnent lieu chaque année à une profusion de travaux, académiques, technologiques et industriels. L'engouement pour ces disciplines s'explique par la multiplication permanente d'applications et d'enjeux industriels dans des domaines aussi variés que : la reconnaissance de caractère Figure I.7-a, l'inspection de pièces automobile Figure I.7-b, les commandes gestuelles Figure I.7-c, la génération des modèles 3D Figure I.7-d, la gestion du trafic routier Figure I.7-e, l'analyse et le diagnostic médicale Figure I.7-f ...etc.



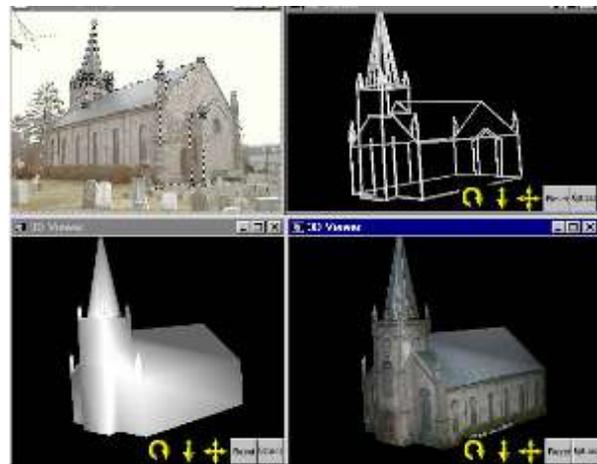
(a) **OCRMax** : logiciel pour la reconnaissance optique de caractères. [Cognex 2017]



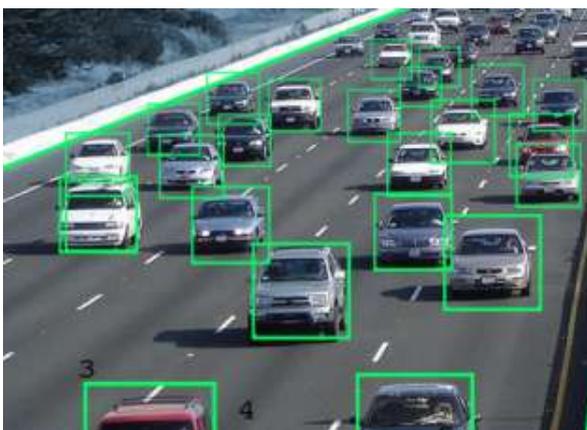
(b) **DS1000 3D** : Système de vision automatisée pour l'inspection de pièces. [DS1000 2017]



(c) **Ground FX** : système de reconnaissance des gestes humains pour jouer ou interagir avec des ordinateurs. [GroundFX 2017]



(d) Le logiciel **PhotoModeler** permet de générer des modèles 3D texturée. [Photomodeler 2017]



(e) **TrafficVision** : logiciel pour la gestion du trafic en temps réel. [Trafficvision 2017]



(f) **Mirada** : logiciel pour l'analyse et le diagnostic d'images médicales. [Mirada 2017]

Figure I.7: Quelques applications industrielles de la vision par ordinateur.

I.4.2. Analyse d'image

Une image numérique est constituée d'une grille rectangulaire d'échantillonnage dont les constituants sont des pixels portant des informations sur l'intensité lumineuse des différents lieux au sein de l'image. De ce fait, Une image numérique est représentée par une matrice bidimensionnelle, dont les éléments sont des nombres naturels correspondant à des niveaux de quantification dans l'échelle de l'intensité lumineuse.

L'analyse d'image de bas niveau et la vision par ordinateur de haut niveau diffèrent dans les données utilisées. En effet, les données de bas niveau sont constituées d'images d'origine représentées par des matrices composées de valeurs d'intensité lumineuse, tandis que les données de haut niveau sont originaires des images aussi, en revanche, seulement les données qui sont pertinentes et qui se rapportent à des objectifs de haut niveau sont extraites, ce qui réduit considérablement la quantité de données.

I.4.2.1. Analyse de bas niveau

Les techniques de bas niveau de la vision par ordinateur représentent la base du traitement numérique de l'image [Sonka 2014], ils utilisent très peu de connaissances sur le contenu des images. Tout d'abord une image d'entrée est capturée par une caméra de télévision en (2D) et numérisée, étant décrite par une fonction d'image $f(x, y)$ dont la valeur est en général l'intensité lumineuse en fonction de deux paramètres (x, y) représentant les coordonnées de l'emplacement dans l'image. Puis l'étape de traitement qui consiste en les opérations effectuées sur les images au plus bas niveau d'abstraction comme l'élimination du bruit au sein de l'image, l'amélioration de certains descripteurs des objets jugés pertinents pour interpréter l'image. L'entrée et la sortie sont des images d'intensité. Ces images iconiques sont généralement de la même nature que les données originales capturées par le capteur, le traitement ne fait pas augmenter les informations du contenu de l'image, bien au contraire, il diminue généralement les informations contenues au sein de l'image. Du point de vue de la théorie de l'information, le meilleur traitement est de ne pas faire un traitement.

La meilleure façon d'éviter l'élaboration d'un traitement est de se focaliser sur une acquisition d'images de haute qualité. Toutefois, le traitement est très utile dans une variété de situations, car il permet de supprimer les informations qui ne sont pas pertinentes pour l'analyse de l'image. Par conséquent, le but du traitement est l'amélioration des données d'image en supprimant les indésirables distorsions et en améliorant certaines caractéristiques importantes de l'image pour l'analyse ultérieure de l'image, bien que les transformations géométriques

d'images comme la rotation et le redimensionnement sont également classés comme des méthodes de traitement.

Sachant que les pixels voisins correspondant à un objet donné dans des images réelles ont essentiellement la même valeur d'intensité lumineuse, de sorte que si un pixel déformé peut être capté à partir de l'image, il peut généralement être restauré comme une valeur moyenne des pixels voisins, ceci est un exemple d'un traitement que nous pourrions effectuer sur une image. La segmentation d'image est la prochaine étape du processus d'analyse d'image, dans laquelle l'ordinateur tente de séparer les objets de l'arrière-plan de l'image. Nous distinguons une segmentation totale et une segmentation partielle : la segmentation totale n'est possible que pour des tâches très simples, un exemple étant la reconnaissance d'objets non jointifs sombres du fond clair. Dans des cas de problèmes plus compliqués, les techniques d'analyse d'image de bas niveau gèrent les tâches de segmentation partielle, dans laquelle seuls les indices qui aideront l'analyse ultérieure de haut niveau sont extraits. La description d'objet dans une image totalement segmentée est également comprise dans le cadre d'analyse de bas niveau d'image.

I.4.2.2. Analyse de haut niveau

L'analyse de haut niveau [Sonka 2014] s'appuie sur des connaissances relatives au contenu de l'image, par exemple, taille de l'objet, sa forme et les relations mutuelles entre les objets dans l'image ; ces données de haut niveau sont généralement exprimées sous une forme symbolique. Les méthodes d'intelligence artificielle sont largement applicables aussi, en effet, la vision par ordinateur de haut niveau tente d'imiter la cognition humaine et la capacité à prendre des décisions en fonction de l'information contenue dans l'image. La vision de haut niveau commence par une certaine forme du modèle formel du monde, puis la réalité perçue sous la forme d'images numériques est comparée au modèle ; l'ordinateur passe en analyse d'image à bas niveau pour trouver des informations nécessaires afin de mettre à jour le modèle.

I.5. Conclusion

La RDF un domaine de recherche actif s'évoluant de plus en plus, où l'analyse de l'image constitue l'outil principal de la vision par ordinateur.

Dans le domaine de la RDF, plus la description effectuée est proche de l'interprétation humaine, plus elle est considérée comme robuste et fidèle à la scène initiale. Les méthodes de description de formes sont variables et dépendent de l'objectif visé (description globale ou locale) et du type de l'objet à analyser (caractère, son, visage, image médicale ...etc.). Dans la littérature, la description d'images est assurée en utilisant la couleur, la texture, les contours et/ou la forme.

Dans le chapitre suivant, nous présentons le domaine de diagnostic des anomalies mammaires et de la mammographie numérique ; à savoir, les informations générales sur la structure de la glande mammaire, les types de tumeurs du sein ; ce qui nous permet de mieux cerner le système de reconnaissance de forme abordé.

Chapitre

2

MAMMOGRAPHIE ET DEPISTAGE DU CANCER DU SEIN

II.1. Introduction

Le cancer du sein constitue la première cause de mortalité chez les femmes âgées de 35 à 50 ans. Selon les chiffres de l'Institut national de la santé publique algérienne [Abid 2009] et le nombre de décès annuels dus à cette maladie est évalué à 3500. En outre, l'Institut recense 7000 nouveaux cas de maladie par an, l'incidence a été multipliée par 5 en 20 ans. La prévalence a grimpé de 10 pour 100000 habitants en 1990 à 50 pour 100000 habitants actuellement. Parfois, les femmes arrivent en consultation à un stade avancé de la maladie, soit quand la médecine ne peut plus faire grand-chose pour les guérir du mal. Donc, plus la maladie est détectée à un stade précoce, plus les chances de guérison sont grandes. Un dépistage systématique des maladies du sein constitue donc une étape importante de la chaîne de traitement des maladies du sein.

La recherche de méthodes de détection automatique des maladies du sein est assez récente. Aujourd'hui, l'analyse des mammographies numérisées semble être une des méthodes les plus encourageantes.

Dans la première partie de ce chapitre nous allons décrire l'anatomie du sein d'une manière générale, et une étude sur la mammographie et les différentes incidences utilisées pour le dépistage du cancer du sein, et dans la deuxième partie nous allons décrire le problème de dépistage du cancer du sein, surtout en ce qui concerne la forme, l'intensité, la taille et les contours des lésions.

II.2. La glande mammaire

C'est une glande annexée à la paroi antérieure du thorax. Elle est constituée de trois entités anatomiques : la peau, la matrice conjonctivo-graisseuse et les tissus adipeux sous-cutanés.

La matrice conjonctivo-graisseuse qui en assure le soutien est perforée en tous sens telle une éponge. A l'intérieur de la matrice conjonctive se développent le système vasculaire, lymphatique et glandulaire. Le système glandulaire, à fonction de lactation, se compose de lobules (qui sécrètent le lait en période de lactation) s'ouvrant sur des canaux galactophoriques (qui drainent le lait vers le mamelon) [Digabel 2008]. Sur le plan profond, on distingue 3 muscles : le grand pectoral sur lequel glisse la glande mammaire, le petit pectoral et le sous-clavier.

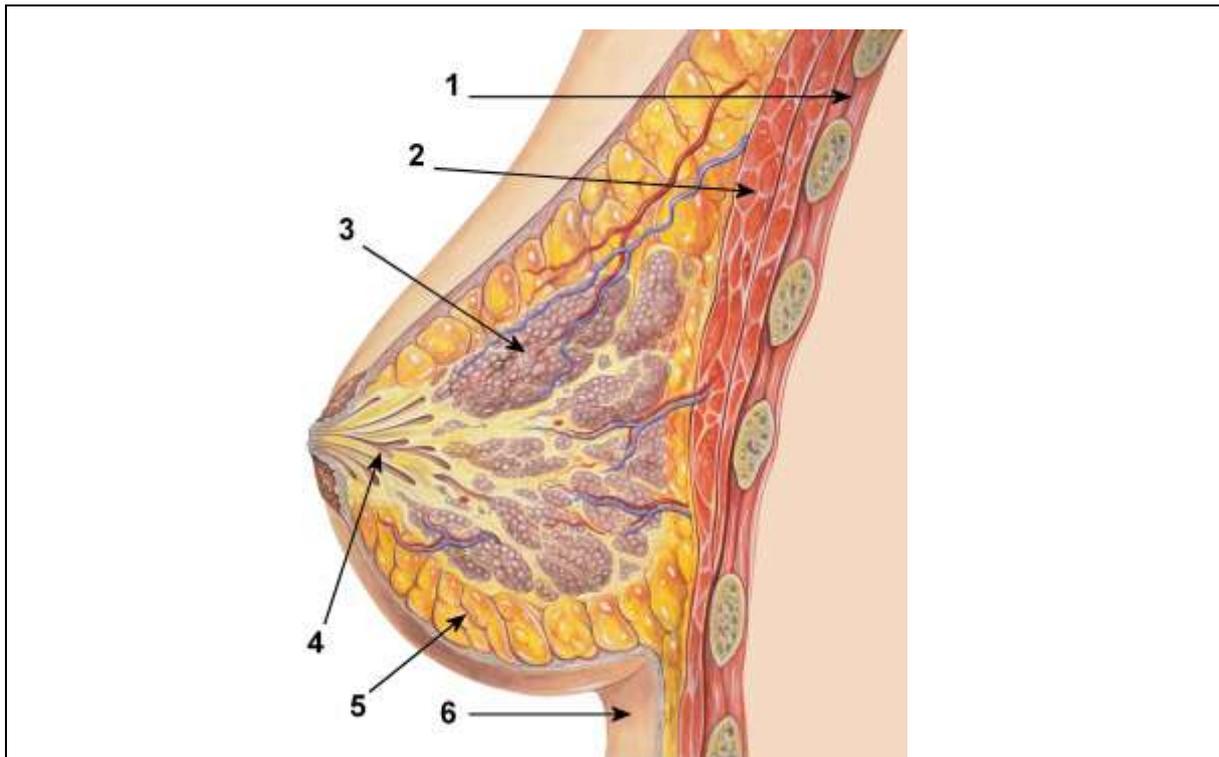


Figure II.1: Schéma anatomique du sein [Drake 2015].

1. Cage thoracique
2. Muscles pectoraux
3. Lobules
4. canaux galactophores
5. Tissu adipeux
6. Peau

II.3. La Mammographie

La mammographie est une technique de radiographie, particulièrement utilisée pour détecter les tissus anormaux qui pourraient être en rapport avec un cancer du sein. Jusqu'à ce jour, seule la technique standard analogique des mammographies est autorisée pour l'utiliser au dépistage du cancer du sein. Les appareillages doivent répondre à un certain nombre de critères pour que les clichés soient d'une bonne qualité. Ils doivent être équipés de format (18×24) et (24×30) pour qu'il couvre une zone d'exploration mammaire suffisante.

La mammographie est réalisée à l'aide d'un appareil spécialement conçu, le scénographe comportant une source de rayons X adaptée à la radiologie des tissus mous et un système de compression du sein. La patiente se présente debout et le manipulateur étale la glande sur une plaque porte film. Puis, il actionne la pédale de compression automatique assurant le mouvement d'un compresseur en plastique qui vient comprimer modérément mais fermement la glande à la limite de l'inconfort sans jamais être douloureux [Heron 2016].

II.3.1. Les incidences de base

Un cliché (*image mammographique*) est tiré selon les deux projections spatiales de face, c'est-à-dire que la source de rayons X est située à la perpendiculaire au-dessus du sein (*incidence cranio-caudale*), et de trois quarts, en oblique (*latérale oblique*).

II.3.1.1. L'incidence de face ou incidence cranio-caudale (CC)

Dans cette incidence la patiente debout face à l'appareil est légèrement tournée, les épaules forment un angle de 30° par rapport au plateau pour adapter celui-ci à la convexité de la cage thoracique. La tête est tournée à l'opposé du sein à radiographier, les bras sont placés le long du corps, les épaules tombantes. La manipulatrice se place du côté du sein à radiographier, ce qui lui permettra de bien contrôler l'étalement de la partie externe de la glande. Elle place la paume de la main sous le sein, bord cubital contre le gril costal de la patiente, elle soulève le sein pour bien dégager le sillon sous-mammaire et affine le réglage en hauteur du plateau. Avec l'autre main, la manipulatrice va former une pince pour bien engager le sein entre le plateau et le palpateur. Elle va ensuite dégager délicatement la main inférieure en tirant vers l'avant et légèrement en dedans. Avec la main supérieure, elle plaque le sein sur le plateau pour éviter tout recul de la glande [Digabel 2008].

Avant la compression, elle vérifie que le mamelon est bien tangentiel, la partie externe du sein bien engagée et que le bord cubital de sa main affleure le bord du centre lumineux. La compression qui doit être rapide, indolore mais efficace est appliquée à mesure que la main supérieure est dégagée en tirant vers l'avant et le dedans.

II.3.1.2. L'incidence oblique externe

Cette incidence, initialement utilisée seule lors de la réalisation du mammothé, permet d'étaler la glande de telle sorte que la visualisation en est optimale avec un tissu glandulaire étalé dans sa totalité et sans superposition.

L'inclinaison varie entre 45° et 60° selon que la femme est bréviligne ou longiligne.

La patiente est placée bras en abduction à 90°, le coude à la hauteur de l'épaule. Le bord supérieur du plateau vient au contact du bord inférieur du bras, le corps de la femme forme un angle de 30° avec le plateau qui doit épouser la convexité du thorax.

La patiente est avancée ou reculée de telle sorte que la limite externe du sein soit dans l'alignement du bord antérieur du plateau porte-cassette.

La manipulatrice saisit le bras et l'épaule de la femme et imprime un mouvement de translation pour engager le creux axillaire dans l'appareil tout en gardant le bras posé sur le plateau. Elle vient se positionner ensuite face à la patiente pour vérifier qu'il n'existe pas de superposition avec le ventre au niveau du sillon sous-mammaire.

Pour bien libérer la partie externe de la glande qui peut être restée bloquée le long du champ du plateau, elle glisse la main sous la partie externe du sein, le bord cubital au plus près du gril costal, elle pose l'autre main sur la partie antérieure et, de la même façon que pour l'incidence de face, elle se sert de la pince formée par ses deux mains pour bien engager le sein entre le plateau et le palpeur. Elle dégage doucement la main inférieure tandis que la main supérieure applique la glande sur le plateau pour éviter tout recul.

Avant d'appliquer la compression, la manipulatrice vérifie que le bord axillaire antérieur est bien engagé. De la même façon que pour l'incidence de face, elle vérifie que le mamelon est bien tangentiel et que le bord cubital de la main, au contact du gril costal, affleure le bord du centreur lumineux. Le sillon sous-mammaire bien engagé dans l'appareil est dégagé de toute superposition. Lors de la compression qui respecte les mêmes règles que pour l'incidence de face, la main antérieure est retirée progressivement tout en tirant vers l'avant et légèrement vers le haut.

II.3.2. La qualité de la mammographie

La qualité de la mammographie s'influe beaucoup sur le processus de diagnostic, un bon cliché permet une bonne lecture de la mammographie, Plusieurs précautions sont donc à prendre afin d'éviter des mauvais clichés.

II.3.2.1. Caractéristique de positionnement

Pour qu'une mammographie soit d'une bonne qualité il ne doit pas contenir de plis, d'artefacts. Le mamelon doit être tangentiel sur les deux incidences. Les constantes correctement choisies privilégient la résolution en contraste. Et pour chaque incidence il y a des critères de qualité [Heid 2007].

Sur l'incidence crano-caudale, le sein doit être bien centré au milieu du film, la partie externe entièrement radiographiée, l'espace graisseux rétro-glandulaire bien visible avec parfois visible également le bord antérieur du muscle pectoral (Figure II.2.A) [Heid 2007].

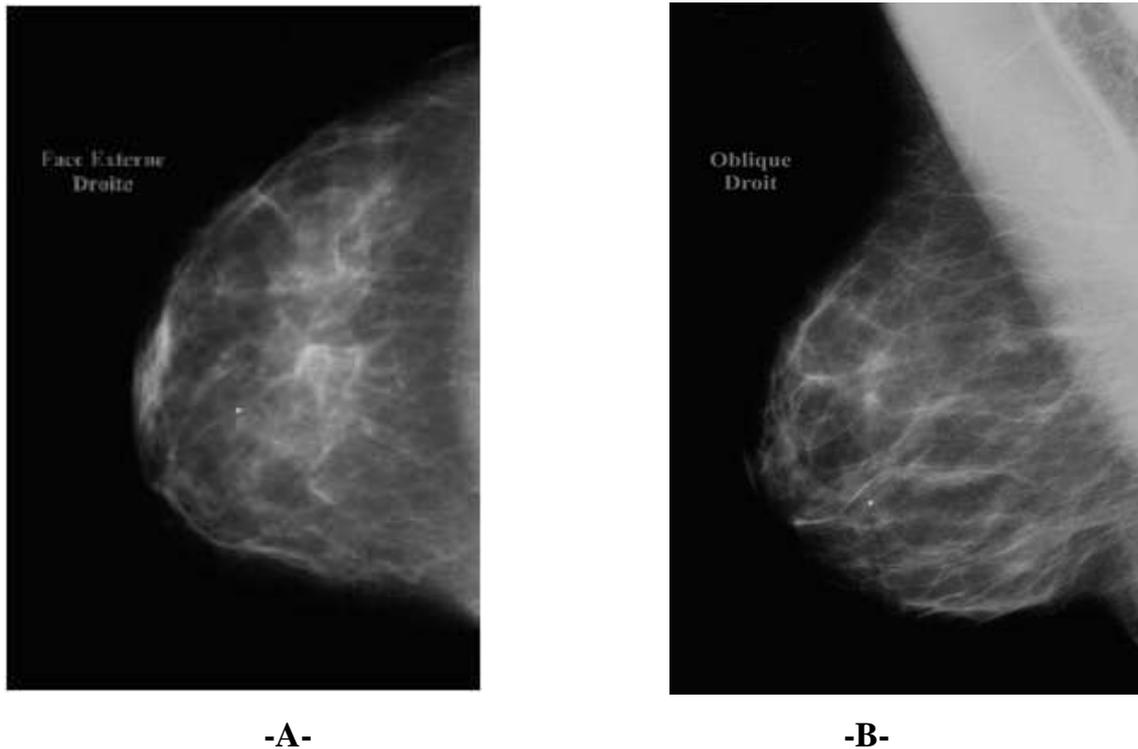


Figure II.2: *A : Résultat de l'incidence de face, B : Résultat de l'incidence oblique.*

Sur l'incidence oblique externe, la glande mammaire sera étalée dans son intégralité. Le sillon sous-mammaire est bien dégagé, sans superposition avec la voussure de l'abdomen. Le muscle pectoral visible sous forme d'une surdensité triangulaire à une pointe inférieure qui se projette sur l'horizontale passant par le mamelon (Figure II.2.B) [Heid 2007].

II.3.2.2. Caractéristiques de la mammographie

La densité mammaire est une entité radiologique appréciant la quantité des structures mammaires radio-opaques ou denses (tissu fibroglandulaire) par rapport au tissu radio-transparent (tissu graisseux). Cette appréciation permet de décrire la capacité de la mammographie à détecter des lésions : plus le sein est dense, plus la sensibilité de la mammographie diminue. De plus, la forte densité est responsable d'un taux plus important de cancers de l'intervalle et de faux positifs. Toutes les études ayant évalué la relation entre densité mammaire et cancer du sein ont montré une augmentation significative du risque de cancer chez les femmes avec des seins denses, qui ne peut s'expliquer par un défaut de détection lors des mammographies antérieures [Tardivon 2008].

De nombreuses classifications ont été proposées pour évaluer la densité mammaire. Celle qui est actuellement utilisée en routine provient du lexique BI-RADS [Lévy 2005] et consiste à apprécier visuellement le pourcentage de tissu radio-opaque sur une incidence mammographique. Quatre types sont ainsi décrits :

- ✓ Type 1 : sein presque entièrement graisseux (< 25 % de glande mammaire) ;
- ✓ Type 2 : présence d'opacités fibroglandulaires éparses (approximativement 25 à 50 % de glande mammaire) ;
- ✓ Type 3 : tissu mammaire dense et hétérogène, ce qui peut compliquer la détection de petites masses (approximativement 51 à 75 % de glande mammaire) ;
- ✓ Type 4 : tissu mammaire très dense, ce qui peut diminuer la sensibilité de la mammographie (> 75 % de glande mammaire).

Cette appréciation de la densité doit apparaître dans le compte-rendu, qui précise également la notion de densité homogène ou hétérogène.

Les mouvements du sein pendant le cliché ne sont pas souhaitables, on distingue les mouvements déclenchés par la douleur de compression, et les mouvements entraînés par un mauvais positionnement, appareil trop haut ou trop bas. Le flou est difficile à reconnaître dans l'absolu, mais une comparaison avec d'autres clichés est souhaitable. Des micro-calcifications peuvent disparaître sur un cliché flou : lorsque l'on reconnaît des calcifications sur une seule incidence il est nécessaire de refaire le cliché où elles n'apparaissent pas [Travers 1995].

Il a été remarqué qu'environ 90 à 95% des seins sont enregistrés sur un film de format 18 × 24. Pour voir sans difficulté tous les seins il faut soit disposer d'un format 24 × 30, soit d'emblée diviser chaque incidence en deux parties sur deux films complémentaires [Travers 1995].

Ces critères permettent aux radiologues de prévenir d'éventuelles dérives et de les corriger avant qu'elles ne soient visibles sur les images radiologiques.

II.3.3. Limite de la mammographie

Après la mammographie, un radiologue analyse les clichés, interroge la patiente et réalise un examen clinique : il examine l'aspect de la peau et du mamelon. Il palpe les seins et recherche de ganglions anormaux. La mammographie ne permet pas toujours de donner d'emblée un diagnostic définitif : elle permet de voir s'il existe une anomalie dans le sein, mais elle ne permet pas de déterminer avec certitude s'il s'agit ou non d'un cancer [David 2002].

Plus les seins sont jeunes, plus ils contiennent de glande et moins ils contiennent de graisse. Or, la glande est mal étudiée par les rayons X et les mammographies sont donc peu adaptées et/ou insuffisantes à l'exploration des seins d'adolescentes ou de femmes jeunes n'ayant pas eu d'enfants.

Egalement, les mammographies sont parfois difficiles à lire chez des femmes ayant des clichés denses. Dans ces cas, la mammographie est souvent complétée par une échographie mammaire.

Pour cela l'analyse de la mammographie reste un examen difficile, nécessitant un appareillage de pointe et une grande habitude de lecture des clichés. D'autres examens complémentaires sont alors nécessaires pour établir le diagnostic.

II.3.3.1. L'échographie

Le principe de l'échographie consiste à appliquer une sonde contre la peau en regard de l'organe à explorer. Cette sonde émet des ultrasons qui traversent les tissus puis lui sont renvoyés sous la forme d'écho. Ce signal, une fois recueilli est analysé par un système informatique qui retransmet en direct une image sur un écran vidéo. La Figure II.3 présente deux exemples d'images échographiques du sein avec a) une lésion maligne et b) une lésion bénigne.

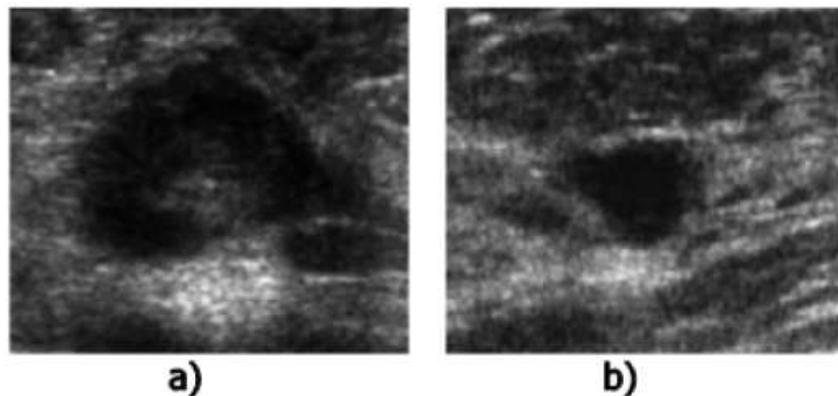


Figure II.3: Exemples d'images échographiques du sein.

La technique d'échographie présente deux avantages majeurs. Elle est d'une part peu coûteuse et d'autre part non invasive. Elle ne présente aucun risque pour la patiente, pour cette raison, elle est généralement utilisée dans le cas où la patiente est enceinte. Toutefois, vu certains inconvénients, l'échographie du sein n'est pas systématique. Elle est utilisée en complément d'une mammographie surtout qu'elle ne révèle que très rarement un cancer non détecté par la mammographie.

En outre, les micro calcifications qui sont de petites tailles (et même les petites masses) sont difficiles à détecter par ultrasons. Il est ainsi difficile de s'assurer que le sein ait été diagnostiqué dans sa totalité à l'issue de ce type d'examen. De plus, les images ultrasonores sont généralement altérées par un bruit spécifique appelé la granularité (reconnu aussi sous le nom speckle) qui est causé essentiellement par les interférences entre les ondes. En conséquence, cette méthode d'imagerie médicale n'est pas généralisée aux campagnes de dépistage. Elle est souvent exploitée comme moyen de repérage lors d'une ponction ou d'une biopsie et aussi comme moyen de repérage préopératoire pour marquer l'emplacement de la lésion.

II.3.3.2. Imagerie par Résonance Magnétique (IRM)

L'Imagerie par Résonance Magnétique est une technique d'imagerie médicale relativement récente (début des années 1980). Cette méthode se base sur l'utilisation d'un aimant (constituant le champ magnétique) et d'ondes de radiofréquences. Son principe consiste à faire vibrer de façon imperceptible les atomes d'hydrogène du corps humain. Placés dans un champ magnétique puissant, tous les atomes d'hydrogène s'orientent vers la même direction. Ils sont alors excités par des ondes radio durant une très courte durée. On dit qu'ils sont mis en résonance. A l'arrêt de cette stimulation, les atomes restituent l'énergie accumulée en produisant un signal. Ce dernier est enregistré et traité sous forme d'image par un système informatique. Nous présentons dans la Figure II.4 l'image IRM des seins gauche et droit.

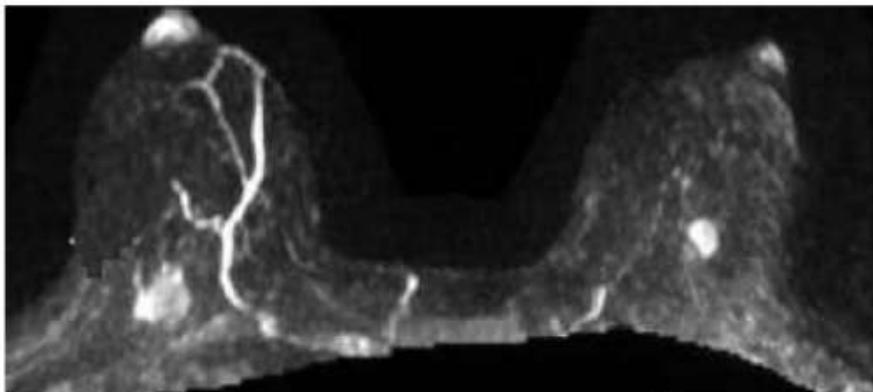


Figure II.4: Exemple d'IRM mammaire bi-sein.

L'IRM permet une étude dans tous les plans de l'espace et l'obtention d'images de grande qualité. De plus, elle fournit une image en haute résolution, ce qui permet une étude approfondie de la pathologie. Cependant, cette technique est très coûteuse et appliquée uniquement aux patients non porteurs de prothèses métalliques. Ainsi, son utilisation est restreinte pour apporter des précisions concernant une anomalie déjà détectée en mammographie ou en échographie.

En tenant compte des conditions générales des différents moyens d'imagerie médicale (limitations de l'examen échographique, coût de l'examen d'IRM), la technique de mammographie semble être la plus adaptée à une utilisation générale dans le cadre de dépistage ou de diagnostic du cancer du sein. Nous détaillons dans ce qui suit le principe de base de ce type d'imagerie médicale.

II.4. Les opacités du sein

Une opacité correspond à une plage de surdensité anormale (sur l'exemple de la Figure II.5, nous avons indiqué l'opacité par une flèche). Les opacités sont considérées comme des

surdensités anormal, car les surdensités normales sont nombreuses sur un cliché mammographique, Une surdensité anormale ne se distingue pas d'une surdensité normale par un critère précis, mais par une combinaison de différentes caractéristiques comme la taille, la densité, les contours, la forme, les texture. Pour un spécialiste c'est l'expérience qui lui permet de différencier entre les surdensités normale et anormale sur une mammographie. La présence d'une surdensité sur plusieurs clichés effectués sur plusieurs incidences implique une forte probabilité que cette surdensité soit anormale [Vachier 1995].

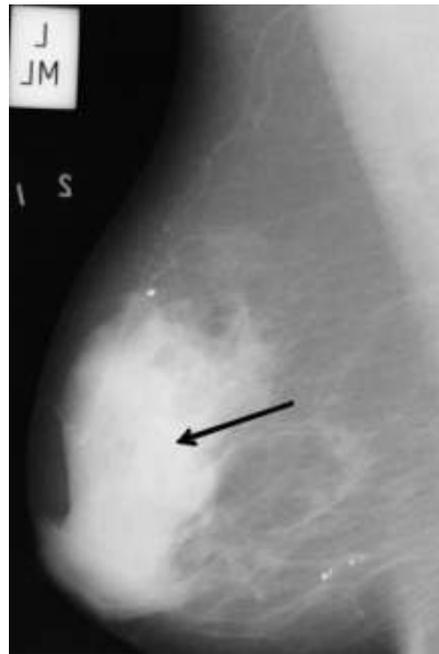


Figure II.5: *une surdensité sur le cliché mammographie.*

On peut dire aussi, qu'une opacité anormale ne correspond pas toujours à une zone cancéreuse, et pour affirmer le degré de malignité d'une opacité plusieurs facteurs sont nécessaires comme la densité de la zone détectée, la taille, la forme, les contours de la surdensité. Par exemple, une opacité de type « Distorsion architecturale » signifie très fortement une tumeur cancéreuse mais une opacité arrondie et homogène est bénigne dans 90% des cas [Vachier 1995].

Les opacités ne constituent pas le seul signe radiologique en mammographie. La présence des microcalcifications, une rupture architecturale ou une désorganisation fibreuse, une dissymétrie de densité entre les deux seins ou encore un épaissement cutané évoquent également un processus pathologique [Vachier 1995].

L'analyse précise d'une opacité ne suffit pas toujours à préjuger de sa nature. Il est nécessaire dans la plupart des cas de comparer le résultat avec celui obtenu par un examen clinique ou par l'analyse des autres signes radiologiques.

II.5. Dépistage du cancer du sein

La mammographie est aujourd'hui reconnue comme la technique de dépistage la plus performante. Elle autorise la détection de lésions et permet donc d'intervenir à un stade très précoce de la maladie. De plus, elle permet également de mettre en évidence les signes secondaires tels que les micro-calcifications. [Vachier 1995] La mammographie constitue également une aide au diagnostic en permettant au radiologue d'évaluer le degré de malignité des lésions détectées et ainsi d'orienter le traitement (acte chirurgical, examens complémentaires échographie ou IRM). L'analyse de la mammographie nécessite des clichés de bonne qualité et une approche d'analyse, en deux étapes [Heron 2016].

Dans la première étape, le but est de trouver une région dans la mammographie qui a une apparence différente par rapport aux autres régions. Dans la deuxième étape le radiologue essaye de distinguer les variations plus ou moins pathologique selon la taille, la forme et les contours de la zone détectée, pour classer l'image comme suspecte [Heron 2016].

On peut classer les anomalies selon trois catégories :

- Les densités asymétriques.
- Les masses.
- Les calcifications.

II.5.1. Les densités asymétriques

Les surdensités asymétriques sont détectées en positionnant les mammographies gauche et droite de la même patiente en miroir, dans les cas normaux on observe une répartition symétrique des densités dans les deux mammographies. Mais dans les cas anormaux on observe une asymétrie c'est-à-dire une surdensité visible sur l'un des deux clichés et invisible sur l'autre, cette asymétrie peut être la conséquence d'une masse, de zones remaniées par une biopsie antérieure ou d'une mauvaise compression.

Une asymétrie focalisée peut être représentée par une forte probabilité, voir une éventuelle masse cancéreuse [Tardivon 2007].

Une densité focale asymétrique se caractérise par sa concentration croissante vers son centre pour former une zone centrale très dense que l'on retrouve sur plusieurs clichés, la Figure II.6 montre une asymétrie sur des mammographies latérale oblique, où on observe une surdensité visible le cliché mammographique droite, et invisible sur la mammographie gauche.

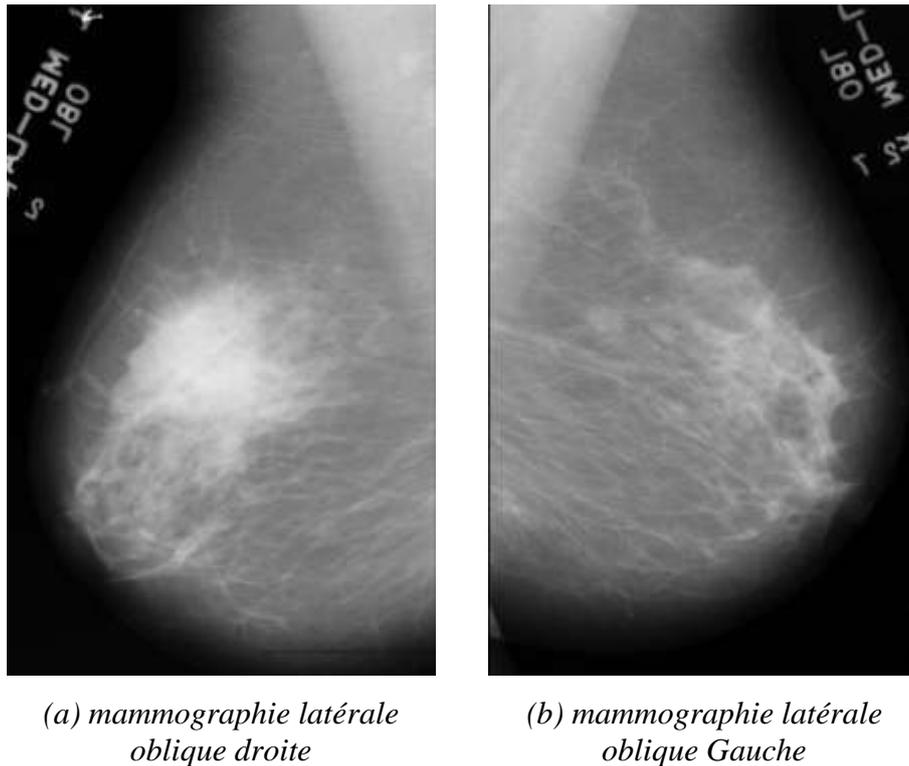


Figure II.6: Une asymétrie de densité observée seulement sur la mammographie gauche.

II.5.2. Les masses

Les masses sont des lésions retrouvées en trois dimensions c'est-à-dire qu'on peut l'observer sur plusieurs incidences. La localisation de la masse est de l'analysée à partir de plusieurs clichés surtout les petites masses qui ne sont pas palpables. La visibilité des contours de la masse évoque la malignité [Tardivon 2007].

La taille de la masse détectée ne donne pas une signification de malignité, sauf sur des clichés successifs lorsque l'on voit la taille augmenter régulièrement. On distingue généralement 5 formes différentes de masses : ronde, ovale, lobulaire, irrégulière ou une distorsion architecturale [Heron 2016].

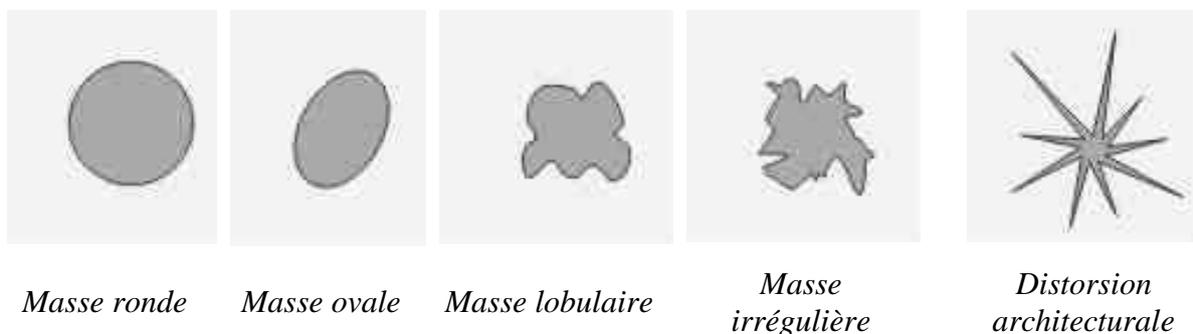


Figure II.7: Les formes des masses.

Les contours de la masse sont très importants à étudier lors de l'étape d'analyse. On distingue généralement 5 types de contours : bien définis, effacés, micro-lobulés, mal définis et avec des

spicules. Les contours effacés correspondent à une superposition de tissu adjacent. Les contours mal définis ou les spicules correspondent à une invasion dans les régions saines [Heron 2016].

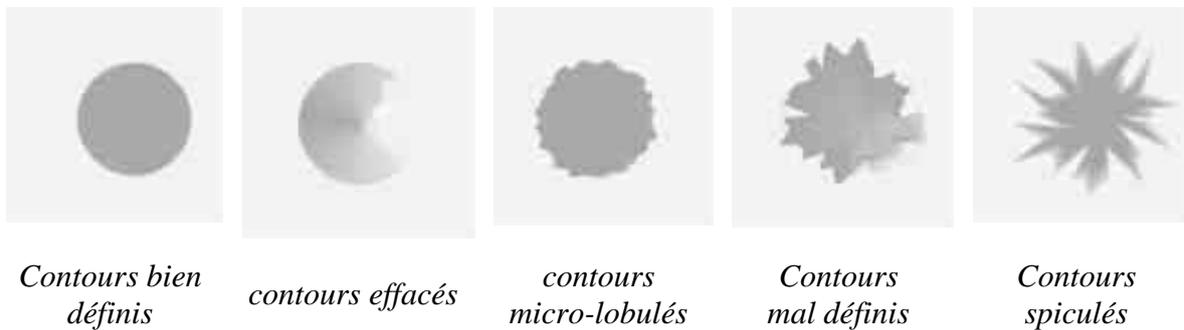


Figure II.8: *Les types des contours des masses.*

Les masses sont en général denses (plus blanches) que le tissu mammaire connexe.

On peut diviser les masses en deux catégories, les masses régulières et les masses irrégulières.

II.5.2.1. Les masses régulières

Ce sont les masses rondes, ovales ou lobulées avec des contours bien définies (Figure II.9) sont a priori des lésions bénignes. Mais cette règle n'est pas absolument vraie, puisque dans certains cas des petits cancers peuvent apparaître. Lorsqu'il y a plusieurs masses dans la même mammographie, dans la plupart des cas, cela signifie qu'il s'agit probablement à des lésions bénignes [Heron 2016].



Figure II.9: *Masse régulière ronde dans la mammographie (incidence latérale-oblique)*

Nous trouvons parfois que les masses rondes ont des contours mal définis en raison de l'importance du tissu adjacent, ces masses peuvent ressembler à une masse cancéreuse aux

contours mal définis. La superposition des tissus avoisinants peut entraîner cet aspect et rendre le diagnostic plus facile surtout si l'image mammographique est très dense [Heron 2016].

Dans la plupart des cas les masses ne sont pas parfaitement rondes ou ovales mais présentent un certain degré de lobulation (Figure II.10). Plus le nombre de lobulations augmente, plus grand est le risque d'avoir une tumeur maligne [Heron 2016].

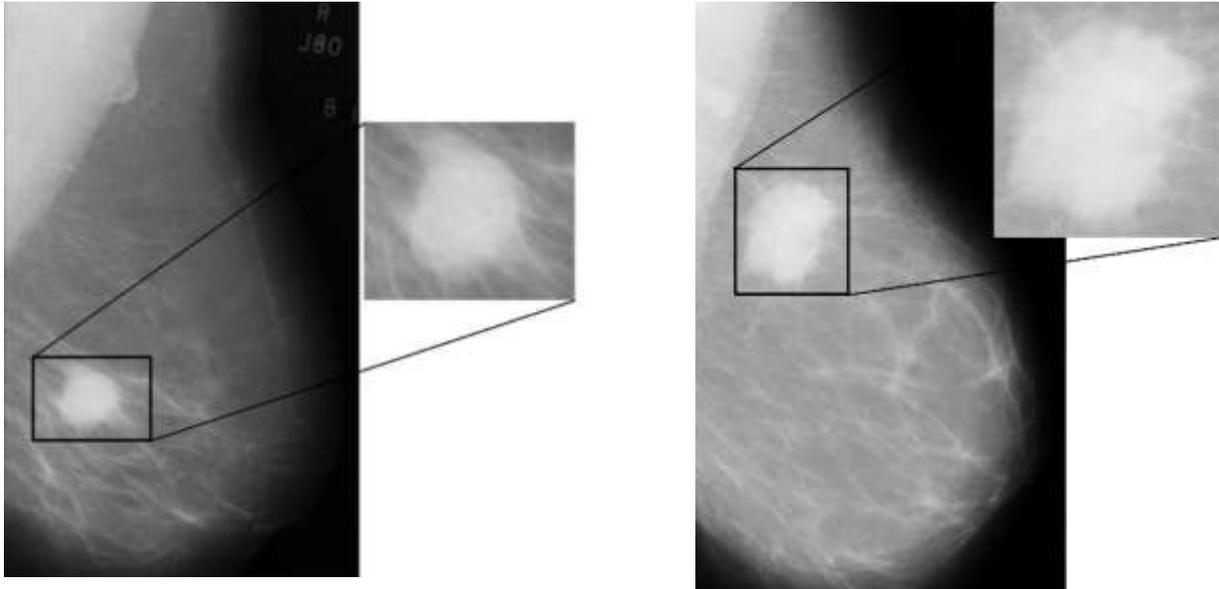


Figure II.10: Image de masses à Contours réguliers.

Lorsque la masse est bien visible et montre un contour mal défini, il faut redouter une lésion maligne.

II.5.2.2. Les masses irrégulières

Les masses irrégulières et les distorsions architecturales sont généralement des lésions malignes, la distorsion architecturale à une forme des rayons convergeant vers un centre avec une rétraction focale. Une distorsion architecturale (Figure II.11) n'est pas une masse en soi, mais une réaction tissulaire traduisant une interruption focale de la distribution normale du tissu mammaire. On peut retrouver une telle distorsion près de la graisse sous-cutanée [Tardivon 2007].

Une telle distorsion permet de localiser un cancer, mais on peut l'observer en cas de cicatrice chirurgicale ou des fois en cas de superposition de tissus denses.

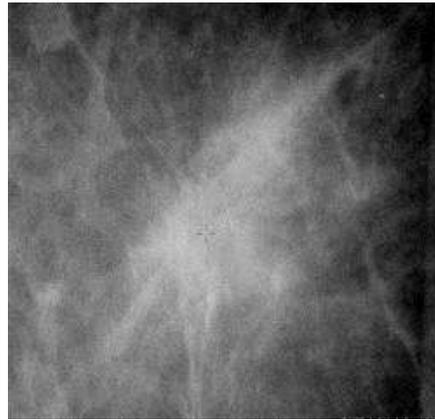


Figure II.11: *Images de distorsion architecturale*

Lorsque les masses ont des contours étoilés ou spiculés (Figure II.12) cela augmente le degré de malignité. Mais aussi les cicatrices chirurgicales, un traumatisme ou une infection peuvent aussi produire des images avec contours spiculés.

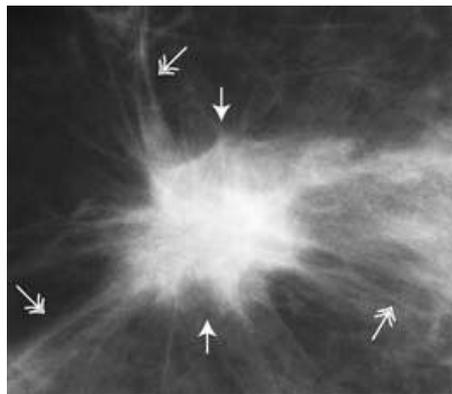


Figure II.12: *Images de masses avec des limites en spicules*

II.5.3. Les calcifications

Les calcifications mammaires sont des dépôts de calcium qui se forment dans le tissu du sein. Les calcifications sont assez courantes et la plupart ne sont pas associées au cancer. Elles sont associées à des images de masse ou de modifications architecturales et peuvent se retrouver dans des pathologies malignes ou bénignes. Les radiologues analysent les calcifications selon leur taille, leur forme, leur nombre et leur distribution. Généralement les calcifications de grande taille régulière, rondes ou ovales sont des tumeurs bénignes, et les calcifications petites, irrégulières, de formes variées sont souvent des tumeurs malignes [Heron 2016]. Elles apparaissent souvent sous forme de petits points blancs. Les calcifications peuvent être détectées :

- Soit à l'intérieur ou au niveau des canalicules mammaires,
- Soit au niveau des lobules,

- Soit au niveau des structures vasculaires,
- Soit au niveau du tissu conjonctif.

Sur le cliché mammographique on peut observer les calcifications lorsque la taille est d'environ 0.2 à 0.3 mm. Les calcifications de moins de 0.5 mm sont souvent des cancers malins, et lorsque la taille des calcifications est de 2 mm ou de plus, il représente souvent des lésions bénignes [Heron 2016], et la présence de plusieurs petites calcifications augmente le degré de malignité. La plupart des radiologues considèrent qu'au-dessus de 4 à 6 micro-calcifications, il faut d'être prudent dans la lecture de la mammographie [Tardivon 2007]. La morphologie de la calcification apporte beaucoup de renseignements et permet de séparer les micro-calcifications bénignes des malignes. Par exemple, les calcifications arrondies ou ovales et d'une taille uniforme, sont probablement bénignes. Au contraire, lorsqu'elles sont irrégulières dans leur taille, et si elles sont représentées par leur hétérogénéité comme « des débris de verre », sont probablement malignes [Heron 2016]. On distingue 5 types de calcifications.

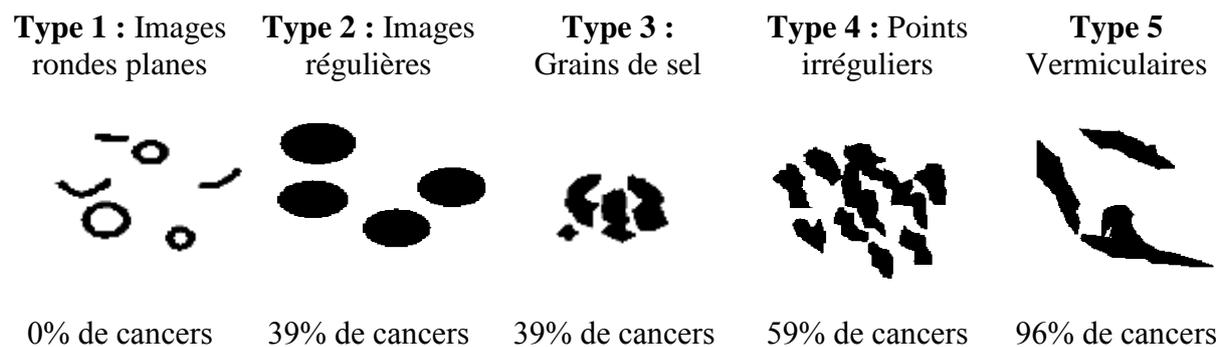


Figure II.13: Les types de calcifications.

La Figure II.13 représente les différents types de calcification et expose le degré de malignité de chaque type.

II.6. La classification des anomalies mammographiques

Il est important d'adopter un lexique standard et une classification commune afin de fournir aux radiologues une description claire et précise des lésions mammaires. Il existe plusieurs classifications comme Le Gal [Gal 1984], de Lanyi [Lanyi 1985] et de BIRADS [Lévy 2005]. La classification que nous avons utilisé dans notre approche est celle de BI-RADS car elle est plus récente par rapport aux autres lexiques et c'est celle qui nous a été recommandé par le radiologue avec lequel nous avons travaillé.

II.6.1. La classification de LeGal

En 1976, Le Gal du Service de Radio diagnostique de l'institut Curie à Paris, a conçu la classification dite de Le Gal [Gal 1984]. Cette classification cherche à évaluer le caractère dangereux ou non des calcifications observées. C'est la classification la plus communément utilisée en France. Elle distingue les micro-calcifications selon leur morphologie, on distingue cinq types de micro-calcifications qui ont une valeur prédictive de malignité croissante comme le montre le Tableau II-1 :

Type 1	Micro-calcifications annulaires, arciformes ou polyédriques. - 0 % de cancer du sein.
Type 2	Micro-calcifications rondes et de tailles variables. - 22 % de cancer du sein.
Type 3	Micro-calcifications poussiéreuses, pulvérulentes. - 36 % de cancer du sein.
Type 4	Micro-calcifications irrégulières - 56 % de cancer.
Type 5	Micro-calcifications vermiculaires ou branchées. - > 90 % de cancer

Tableau II-1: *Les classes des micro-calcifications selon LeGal.*

Cette ancienne manière de classification a l'avantage d'être simple. Toutefois, son défaut principal est qu'elle prend en considération uniquement sur les micro-calcifications et n'intègre pas d'autres paramètres tels que :

- L'étude des masses mammographiques.
- L'étude des distorsions architecturales.
- La disposition des microcalcifications.
- L'étude des signes associées.

II.6.2. La classification de Lanyi

Lanyi [Lanyi 1985] individualise seulement deux types de calcifications : les calcifications rondes lobulaires ou lait calcique, associées à une pathologie bénigne et les calcifications canalaires irrégulières ou en bâtonnets en rapport avec de la nécrose et associées à un cancer.

II.6.3. La classification de BI-RADS

La classification BI-RADS est utilisée par les radiologistes lors de la mammographie, de l'échographie et de l'IRM pour définir les anomalies et permettre de savoir ce qui doit être fait par la suite, soit retour au dépistage, suivi rapproché ou biopsie. BI-RADS [Lévy 2005] comporte sept catégories :

II.6.3.1. BI-RADS 0

L'évaluation est incomplète et nécessite des compléments d'imagerie. Cette catégorie est presque toujours utilisée en situation de dépistage mais rarement en situation diagnostique. Les recommandations sont : comparaison avec clichés antérieurs, clichés complémentaires, échographie etc. C'est une classification d'attente et le radiologue doit indiquer dans quelle mesure la poursuite des investigations.

II.6.3.2. BI-RADS 1

L'examen est négatif : mammographie normale. Les seins sont symétriques et il n'y a pas d'opacité, de distorsion architecturale ou de calcification suspecte. En principe cela ne pose aucun problème si les seins sont radio-transparents. Mais quelle est la certitude en cas de seins denses (type 3 ou 4) ?.

Les solutions possibles résident dans la comparaison avec les mammographies anciennes, l'examen clinique ce qui renvoie à la question plus générale de l'intérêt de l'échographie systématiquement associée à l'exploration des seins denses.

II.6.3.3. BI-RADS 2

Constatations bénignes : Cela correspond pratiquement à une mammographie négative mais le radiologue désire décrire une anomalie dont les caractères radiologiques sont caractéristiques et qui peuvent être étiquetées avec beaucoup de confiance dans l'interprétation. Le risque d'association à un cancer est insignifiant en pratique et il n'y a pas lieu de poursuivre les investigations.

Les anomalies entrant dans cette catégorie sont les suivantes :

- Masses rondes avec macro calcifications (fibroadénome /kystes) (Figure II.14).
- Masses rondes ou ovales à centre clair (ganglion intra mammaire).
- Masse ronde correspondant à un kyste typique en échographie.
- Anomalie de densité grasseuse ou mixte (kyste huileux, lipome, galactocèle).

- Cicatrice connue et calcification sur matériel de suture.
- Macro calcifications sans opacité (adénofibrome, kyste, calcifications vasculaires).
- Micro calcifications annulaires ou arciformes, semi-lunaires (Figure II.15).
- Calcifications cutanées et calcifications punctiformes régulières diffuses.



Figure II.14: Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2 [Lévy 2005].

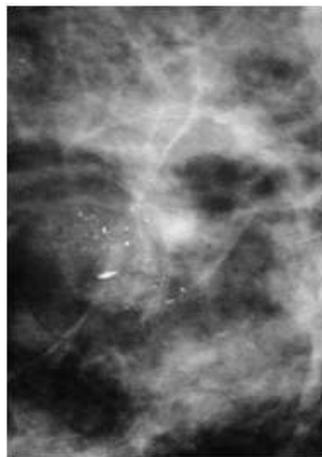


Figure II.15: Microcalcifications bénignes : lait calcique : BI-RADS 2 [Lévy 2005].

II.6.3.4. BI-RADS 3

Anomalie probablement bénigne (> 98 %). Proposition d'une surveillance à court terme. Il s'agit de la catégorie la plus délicate à manier du fait de l'incertitude qu'elle comporte. Il n'y a pas de modification attendue pendant la période de surveillance et les données de la littérature sont en faveur de cette attitude.

La surveillance mammographique est justifiée par le fait que ces lésions probablement bénignes ont effectivement un très faible risque d'être malignes : que la mammographie identifiera, parmi les rares lésions qui se modifient pendant l'intervalle, celles qui sont effectivement malignes après ces cancers seront diagnostiqués tôt dans leur évolution à un stade encore favorable. Ces

lésions doivent être non visibles et ne pas être nouvelles ou en progression par comparaison avec une mammographie antérieure. Il faut préciser dans le compte rendu le rythme suggéré de surveillance.

Au terme d'une surveillance suffisante pendant deux à trois ans, la lésion stable peut être reclassée BI-RADS 2. En cas de modification au cours de la surveillance, la biopsie est la règle. Les données de la littérature sont en faveur de cette attitude puisque le risque de cancer observé dans la littérature est de 0,3 à 1,7 % des lésions classées BI-RADS 3. L'intérêt de l'échographie devant une masse probablement bénigne est évident : quand elle permet de faire le diagnostic immédiat de kyste ou à l'opposé d'une lésion hautement suspecte, elle évite respectivement une surveillance inutile ou une attente de six mois.

Les anomalies susceptibles d'être classées BIRADS 3 sont :

- ✓ Masses ou opacités bien circonscrites rondes ou ovales ou discrètement lobulées, sans micro-lobulation, non calcifiées, non typiquement liquidiennes en échographie (Figure II.16).

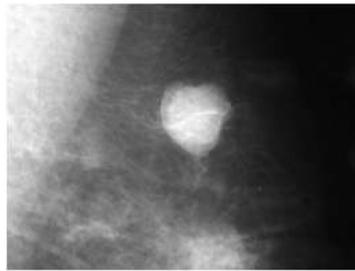


Figure II.16: *Masses bien circonscrites discrètement lobulées, sans micro lobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3 [Lévy 2005].*

- ✓ Asymétries focales de densité à limites concaves et/ou mélangées à de la graisse.
- ✓ Petits amas ronds ou ovales de calcifications amorphes, peu nombreuses, évoquant un début de calcification d'adénofibrome.
- ✓ Une discrète distorsion sans opacité centrale dans la zone d'une biopsie antérieure.
- ✓ Micro calcifications rondes ou punctiformes régulières ou amorphes, peu nombreuses, en petit amas rond isolé.

II.6.3.5. BI-RADS 4

Les anomalies classées BI-RADS 4 sont suspectes, dans ce cas une biopsie doit être envisagée. Ce sont des lésions qui n'ont pas toutes les caractères morphologiques d'un cancer typique mais qui ont une probabilité importante d'être malignes [Bérubé 1998]. Cette probabilité est de 10 à 50 % [Orel 1999]. Une vérification histologique est nécessaire.

Ce sont les :

- ✓ Micro calcifications punctiformes régulières nombreuses et/ou groupées en amas aux contours ni ronds ni ovales.
- ✓ Micro calcifications amorphes, groupées et nombreuses.
- ✓ Micro calcifications irrégulières, polymorphes ou granulaires, peu nombreuses (Figure II.17).

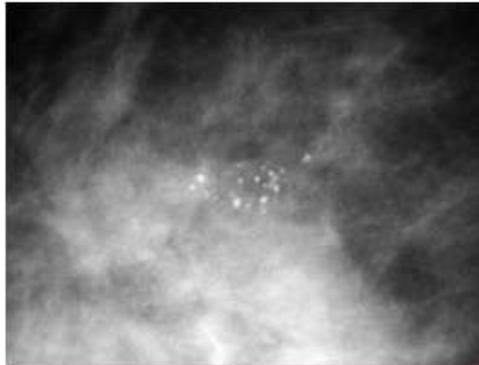


Figure II.17: *Micro calcifications granulaires irrégulières peu nombreuses : BIRADS 4*
[Lévy 2005].

- ✓ Distorsion architecturale sans centre dense (Figure II.18) en dehors d'une cicatrice connue et stable.

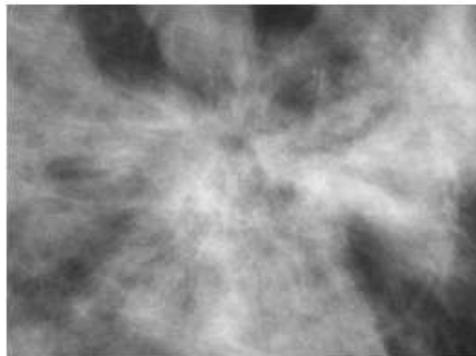


Figure II.18: *Images spiculée sans centre dense : centre prolifératif d'Aschoff* [Lévy 2005].

- ✓ Masses opaques solides (échographie) rondes ou ovales à contours lobulés ou masqués ou ayant augmenté de volume.
- ✓ Asymétries de densité focales, à limites convexes ou évolutives.

II.6.3.6. BI-RADS 5

Les anomalies classées BI-RADS 5 ont une haute probabilité de malignité. Ces lésions ont une très forte probabilité de correspondre à un cancer du sein (> 95 %) [Lévy 2005].

Il s'agit de :

- ✓ Masses spiculées à centre dense.
- ✓ Masses irrégulières à contours indistincts ou spiculés (Figure II.19).

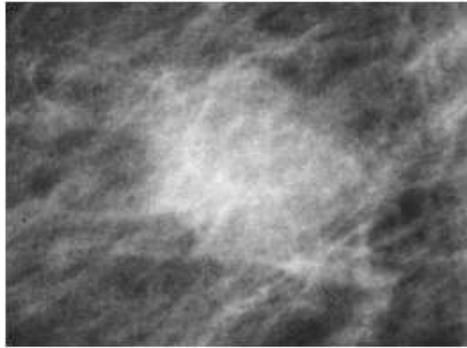


Figure II.19: *Masse mal circonscrite aux contours flous et irréguliers : BI-RADS 5*
[Lévy 2005].

- ✓ Micro calcifications vermiculaires, arborescentes ou micro calcifications irrégulières, polymorphes ou granulaires, nombreuses et groupées (Figure II.20).

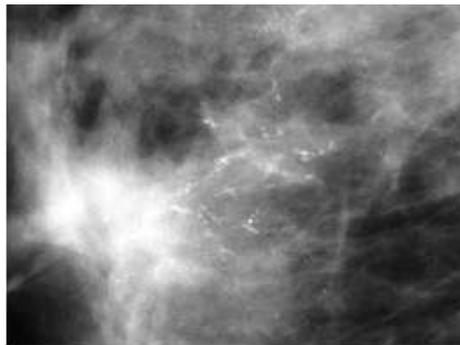


Figure II.20: *Micro calcifications vermiculaires, arborescentes, nombreuses et groupées : BI-RADS 5* [Lévy 2005].

- ✓ Groupements de micro calcifications quelle que soit leur morphologie, dont la topographie est galactophorique.
- ✓ Micro calcifications associées à une distorsion architecturale ou à une opacité.
- ✓ Micro calcifications groupées ayant augmenté en nombre ou micro calcifications dont la morphologie et la distribution sont devenues plus suspectes.

Une biopsie guidée par l'imagerie est utile pour confirmer le diagnostic, d'autant plus que le traitement envisagé comporte la technique de la chimiothérapie.

II.6.3.7. BI-RADS 6

Les mammographies classées BI-RADS 6 Correspond à une anomalie avec malignité prouvée. Une action appropriée (chimiothérapie, hormonothérapie) doit être entreprise.

II.7. Conclusion

Puisque nous ayons qu'une faible connaissance a priori sur la procédure de dépistage du cancer du sein et de la lecture des mammographies, ce chapitre, a été consacré à une description nécessaire sur les images de mammographies, plus particulièrement la forme, la taille et les contours des lésions. Nous avons défini et rappelé tous les outils dont nous aurons besoin pour présenter notre application qui sera décrite dans le chapitre suivant. Il nous permet également de situer le cadre de notre travail. Les outils de l'imagerie médicale permettant le diagnostic du cancer du sein ont été présentés. Une grande attention a été consacrée à l'étude des spécifications des pathologies mammaires à savoir les masses et les micro calcifications. Une telle étude est fortement associée à la présentation des standards adoptés par les radiologues telle que BI-RADS pour classier les lésions mammaires en bénignes-malignes.

Chapitre

3

EXTRACTION DE CARACTERISTIQUES MAMMAIRES

III.1. Introduction

L'extraction de caractéristiques des masses mammographiques est une étape incontournable dans la chaîne de diagnostic automatique des clichés mammographiques. Ces images radiographiques sont présentées en niveau de gris, sachant qu'une masse mammaire se caractérise par sa densité, sa forme et son type de contours. De ce fait, nous présentons dans ce qui suit à étudier les différents descripteurs de texture, de forme et de contours utilisés dans la littérature.

Dans ce chapitre nous dévoilons en premier temps, les différents descripteurs de texture et leur utilité pour caractériser la densité des masses mammographiques. Ensuite, nous présentons les caractéristiques de forme proposées dans la littérature, en commençant par l'étude des descripteurs d'ordre général utilisés dans la majorité des applications de traitement d'images. Ensuite, nous abordons les descripteurs spécifiques couramment appliqués en traitement des masses mammaires. Cette étude permet de distinguer les avantages et les inconvénients des différentes méthodes de caractérisation de la forme des lésions mammaires. A la fin, nous présentons quelques descripteurs de contours ainsi que les caractéristiques humaines utilisées par les radiologues.

III.2. Les caractéristiques des masses mammographiques

Les caractéristiques d'une masse mammographique généralement se divisent en trois grandes catégories : les descripteurs de densité représentés par les données texturales, les descripteurs de forme symbolisés par des fonctions géométrique et les descripteurs de contours caractérisés par des méthodes statistiques. Dans cette section nous présentons un état de l'art de ces trois descripteurs.

III.2.1. Les descripteurs de texture en mammographie

Dans le domaine du traitement de l'image, il n'existe pas de définition conventionnelle de la texture. Les définitions mathématiques proposées peuvent être trop générales et donc imprécises et peuvent aussi être trop détaillées et donc ne couvrent pas tous les cas de texture envisageables.

Haralick et Shanmugam [Haralick 1973] ont décrit la notion de texture comme suit : " Une image texturée est décrite par le nombre et les types de ses primitives tonales ainsi que leurs orientations spatiales. Elle ne peut pas être analysée sans une vue de la primitive tonale de référence. Au fur et à mesure que la résolution augmente, on observe une texture fine puis une texture grossière". Michael Unser [Unser 1984] a défini la texture comme une région d'une

image pour laquelle il est possible de définir une fenêtre de dimensions minimales, telle qu'une observation à travers de celle-ci se traduit par une perception (impression) visuelle identique pour toutes les translations possibles de cette fenêtre à l'intérieur de la région considérée.

Il existe en littérature, plusieurs familles de textures ce qui engendre la proposition de plusieurs descripteurs de texture. En effet, certaines méthodes de description texturale basées sur une analyse fréquentielle regroupent principalement la transformée de Fourier, les filtres de Gabor et la transformée en ondelette. D'autres méthodes basées sur un modèle regroupent les champs de Markov et les mesures fractales. Une autre catégorie de méthodes basée sur une analyse spatiale et statistique des intensités de niveau de gris rassemble les statistiques du premier ordre, les caractéristiques d'auto corrélation et les matrices de co-occurrence. Dans ce qui suit, nous présentons les méthodes d'extraction de caractéristiques les plus utilisées dans le domaine du traitement des images mammographiques.

III.2.1.1. Les statistiques du premier ordre

Les statistiques du premier ordre mesurent la probabilité d'observer un niveau de gris à un emplacement aléatoire d'une image. Elles sont estimées sur les niveaux de gris sans la considération de leur distribution relative. Les statistiques du premier ordre sont calculées à partir de l'histogramme des intensités de l'image noté '*Hist*'. Ce vecteur est obtenu en utilisant la fonction suivante :

$$Hist(i) = \sum p \quad \text{tel que} \quad I(p) = i \quad \text{Équation III-1}$$

Avec '*I(p)*' est l'intensité du pixel '*P*' dans l'image, $i \in \{1, 2, 3, \dots, ng\}$, tel que '*ng*' est le nombre de niveaux de gris.

Parmi les statistiques du premier ordre, nous citons la moyenne (*Moy*), l'écart type (σ_{Hist}) et la variance (*Var*).

- **La moyenne :**

$$Moy = \frac{1}{N} \sum_{i=1}^{ng} i \times Hist(i) \quad \text{Équation III-2}$$

Où '*N*' est le nombre de pixels dans l'image.

- **L'écart-type :**

$$\sigma_{Hist} = \sqrt{\frac{1}{ng} \sum_{i=1}^{ng} (Hist(i) - Moy)^2} \quad \text{Équation III-3}$$

• **La variance :**

$$Var = \frac{1}{N} \sum_{p=1}^N (I(p) - Moy)^2 \quad \text{Équation III-4}$$

L'inconvénient majeur des statistiques du premier ordre basées sur l'histogramme est qu'elles n'exploitent que l'information niveau de gris et ne prennent pas en considération les relations spatiales entre les pixels. La description de texture à base des statistiques d'ordre supérieur (statistiques du second ordre par exemple) s'avère plus efficace.

III.2.1.2. La matrice de co-occurrence

Les méthodes statistiques du second ordre sont les méthodes qui mettent en jeu deux pixels pour le calcul des caractéristiques. La principale méthode du second ordre utilisée pour le traitement des images médicales notamment les images mammographiques est la méthode de matrice de co-occurrence ou méthode de dépendance spatiale des niveaux de gris : SGLD (Spatial Gray Level Dependency) développée par Haralick [Haralick 1973]. C'est une méthode statistique qui consiste à construire des matrices de co-occurrence pour représenter les relations entre les pixels d'une image.

La matrice représente la probabilité conjointe pour deux niveaux de gris i, j soient dans une relation spatiale donnée. Cette relation est définie en termes de la distance et d'angle entre ces deux pixels. L'angle permet d'évaluer la direction de texture et l'application de plusieurs valeurs de distance peut donner une description significative de la taille de la périodicité de la texture. Le choix de la distance et de la combinaison de l'angle, ainsi que le niveau de quantification, est quelque peu arbitraire.

Les matrices de cooccurrences à niveaux de gris (GLCM), introduites par Haralick et al. [Haralick 1973], sont très intéressantes car elles mesurent la distribution des niveaux de gris dans l'image tout en prenant en compte les dépendances spatiales entre les pixels. Cette matrice est une matrice carrée de taille n^2 où n représente le nombre des niveaux de gris de l'image. La plupart des images sont codées sur 256 niveaux de gris par conséquent, la taille des GLCM est considérable. Dans la pratique ces niveaux de gris sont requantifiés (16, 32 ou 64 niveaux de gris) ce qui permet d'avoir un temps de traitement plus restreint [Paulhac 2009], dans notre cas les images sont codées sur 4096 niveaux de gris et nous avons utilisés 64 niveaux de gris pour les GLCM.

Une matrice de cooccurrence mesure la probabilité d'apparition des paires de valeurs de pixels (P_i, P_j) situés à une certaine distance d dans l'image. Elle est basée sur le calcul de la probabilité

$P(i, j, d, \theta)$ qui représente le nombre de fois où un pixel de niveau de gris i apparaît à une distance relative d d'un pixel de niveau de gris j et selon une orientation θ donnée. Afin de ne prendre en compte qu'une information très locale de voisinage, la distance d n'excède pas en générale quelques pixels. Les directions angulaires θ classiquement utilisées sont $0, \pi/4, \pi/2$ et $3\pi/4$. Figure III.1 illustre ce propos.

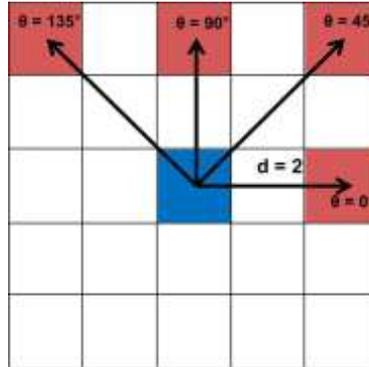


Figure III.1: Les voisins d'une distance $d = 2$ du pixel 'x' selon les 4 directions.

Dans cette méthode, il faut calculer la matrice de co-occurrence pour une région d'intérêt (ROI) carré à quatre angles différents : $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ et à une distance d . En conséquence, les quatre matrices de co-occurrence sont générés pour chaque valeur de distance d , ces matrices sont additionnées pour obtenir une matrice de rotation invariant M^d :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad \text{Équation III-5}$$

Comme les matrices de cooccurrences incluent beaucoup d'informations et sont donc consommatrices en espace mémoire, elles ne sont pas directement utilisées pour caractérisation des textures. Les utilisateurs préfèrent donc calculer à partir de ces matrices quelques descripteurs afin de réduire la quantité d'informations à analysés, tout en conservant la pertinence de ces caractéristiques. Considérons l'image (Figure III.2: *exemple d'une image.*).

1	3	2	5	4	3
2	3	4	2	3	4
3	4	5	6	2	5
7	8	5	6	4	2
3	4	5	2	4	6

Figure III.2: *exemple d'une image.*

Si la direction θ est nulle (direction horizontale) et la distance $d=1$, alors la matrice de co-occurrence $M_{i,j}^1$ est calculée comme suit :

$\begin{matrix} Ng \\ Ng \end{matrix}$	1	2	3	4	5	6	7	8
1	0	0	1	0	0	0	0	0
2	0	0	2	1	2	0	0	0
3	0	1	0	4	0	0	0	0
4	0	2	1	0	2	1	0	0
5	0	1	0	1	0	2	0	0
6	0	1	0	1	0	0	0	0
7	0	0	0	0	0	0	0	1
8	0	0	0	0	1	0	0	0

Figure III.3: La matrice de cooccurrence $M_{i,j}^d$ de l'image (Figure III.2).

Haralick et Shanmugam [Haralick 1973] dans leur article "Textural features for image classification", proposent quatorze descripteurs texturaux extraits à partir des matrices de cooccurrences.

Le calcul de ces descripteurs est basé sur les matrices de cooccurrence. Ces matrices qui mettent effectivement en avant les relations existantes entre les pixels de l'image en faisant intervenir l'aspect local (les niveaux de gris) et l'aspect spatial (d, θ). Ces descripteurs sont les suivants :

1. L'énergie (uniformité)

Ce descripteur mesure l'uniformité de la texture. Il atteint de grandes valeurs lorsque la distribution des niveaux de gris est constante ou de forme périodique. Dans ce dernier cas, lorsque (d, θ) correspond à la période, les valeurs élevées d'énergie sont obtenues pour les matrices M_{θ}^d .

$$Energie^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad \text{Équation III-6}$$

Où : G est le nombre de niveaux de gris utilisé.

2. Le contraste

Le contraste mesure les variations locales des niveaux de gris, lorsque ces variations sont importantes, alors le contraste sera élevé. La valeur en est d'autant plus élevée que la texture présente un fort contraste. Ce paramètre est fortement non corrélé à l'énergie.

$$Contraste^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} ((i-j)^2 \times M_{i,j}^d) \quad \text{Équation III-7}$$

3. La corrélation

La corrélation permet de déterminer si certaines colonnes de la matrice sont égales, c'est-à-dire s'il existe des dépendances linéaires dans l'image. En effet, plus les valeurs sont uniformément distribuées dans la matrice de cooccurrences et plus la corrélation est importante. La corrélation n'est corrélée ni à l'énergie, ni à l'entropie.

$$\text{Corrélation}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad \text{Équation III-8}$$

Où: μ_i^d, σ_i^d : sont la moyenne et l'écart-type de la ligne i de la matrice $M_{i,j}^d$.
 μ_j^d, σ_j^d : sont la moyenne et l'écart-type de la colonne j de la matrice $M_{i,j}^d$.

4. L'homogénéité (Moment différentiel inverse)

L'homogénéité représente le comportement inverse de celui du contraste. En effet, plus la texture possède de régions homogènes et plus le moment différentiel inverse est élevé.

$$\text{Homogénéité}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad \text{Équation III-9}$$

5. La variance

La variance représente l'hétérogénéité de la texture. Elle augmente lorsque les niveaux de gris différent de leur moyenne. Cette mesure est indépendante du contraste. La variance correspond au carré de l'écart-type :

$$\text{Variance}^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu^d)^2 \quad \text{Équation III-10}$$

Où : μ est la valeur moyenne de la matrice de rotation invariante $M_{i,j}^d$.

6. La somme moyenne

$$\text{SM}^d = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad \text{Équation III-11}$$

7. L'entropie de la somme

$$\text{ES}^d = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i + j) M_{i,j}^d \log_{(i+j)M_{i,j}^d} \quad \text{Équation III-12}$$

8. La variance de la somme

$$\text{VS}^d = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [(i + j) - \text{ES}^d]^2 M_{i,j}^d \quad \text{Équation III-13}$$

9. Entropie

L'entropie caractérise la complexité de l'image. Lorsque les valeurs de la matrice de cooccurrences sont égales, l'entropie prend une valeur élevée. Elle permet ainsi de représenter le degré de granulation de l'image. En effet, plus la valeur de l'entropie est grande et plus la granulation est grossière. L'entropie est faible si l'image possède souvent les mêmes couples de pixels. C'est un indicateur de désordre. L'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire (sans structure apparente). Cette mesure est fortement corrélée (par l'inverse) à l'énergie.

$$Entropie^d = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \log_{M_{i,j}^d} \quad \text{Équation III-14}$$

10. Entropie de la différence

$$ED^d = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j) M_{i,j}^d \log_{(i-j)M_{i,j}^d} \quad \text{Équation III-15}$$

11. Variance de la différence

$$VD^d = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [(i-j) - ED^d]^2 M_{i,j}^d \quad \text{Équation III-16}$$

12. Information sur la corrélation

$$Corr1 = \frac{(Entropie^d - HXY1)}{Max(HX, HY)} \quad \text{Équation III-17}$$

Où HX et HY sont respectivement les entropies de M_i^d et M_j^d

$$HXY1 = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \log_{((i \times j)M_{i,j}^d)} \quad \text{Équation III-18}$$

13. Information sur la corrélation 2

$$Corr2 = (1 - \exp^{-2(HXY2 - Entropie^d)})^{1/2} \quad \text{Équation III-19}$$

Où

$$HXY2 = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i \times j) M_{i,j}^d \log_{((i \times j)M_{i,j}^d)} \quad \text{Équation III-20}$$

14. Coefficient de corrélation maximal

$$\text{CorrMax} = (2^{\text{eme Plus grande valeur propre de } Q})^{1/2} \quad \text{Équation III-21}$$

Où :

$$Q(i, j)^d = \frac{\sum_{i=0}^{G-1} (M_{i,k}^d \times M_{j,k}^d)}{((i \times k) M_{i,j}^d)} \quad \text{Équation III-22}$$

Les descripteurs III.11, III.12, III.13, III.15, III.16, III.17, III.19 et III.21 apportent des informations supplémentaires sur la corrélation et les degrés d'homogénéité et de complexité de l'image.

III.2.1.3. La méthode Run difference matrix (RDM)

La méthode Run difference matrix (RDM) est une version étendue et généralisée de la méthode statistique des différences de niveaux de gris [Kim 1991]. Cette méthode est basée sur l'estimation de la fonction de probabilité de densité des différences des niveaux de gris dans l'image [Mohanaiah 2013]. La méthode RDM présente la différence de niveau de gris 'gdif' entre les pixels, lorsque le vecteur de déplacement 'r' entre deux pixels a été donné.

À partir de la Figure III.4, le vecteur de déplacement 'r' est calculé en utilisant la relation suivante :

$$r = [\Delta x, \Delta y] = [x_a, y_b] - [x_n, y_m] \quad \text{Équation III-23}$$

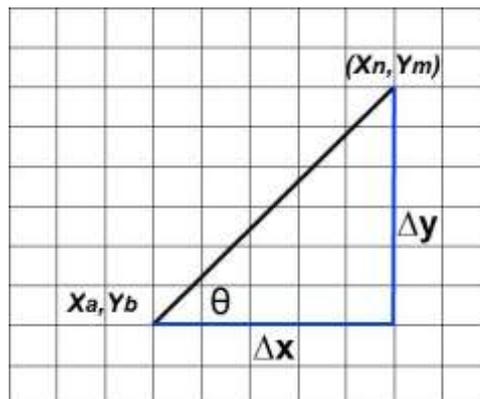


Figure III.4: Le déplacement 'r'.

Soit la région d'intérêt (ROI) contenant un nombre fini de niveaux de gris dans une image 2D, la méthode RDM pourrait être défini comme le nombre total de pair de pixels dans ROI ayant une distance 'r' et la différence de niveau gris 'gdif' avec une direction donnée θ :

$$RDM(r, g_{dif} | \theta) = \# \left\{ \begin{array}{l} P_i(x_i, y_i) : P_j(x_j, y_j) \in ROI^2; \\ \text{Où } |G(x_j, y_j) - G(x_i, y_i)| = g_{dif}, \\ \text{et } r = \sqrt{(x_i, y_i)^2 + P_j(x_j, y_j)^2} \end{array} \right\} / N(r) \quad \text{Équation III-24}$$

Où # désigne le nombre de paires de pixels dans région d'intérêt et $G(x, y)$ est la valeur de niveau gris du pixel (x, y) . La distance 'r' est défini par la fonction suivante :

$$r = \sqrt{\Delta x^2 + \Delta y^2} \quad \text{Équation III-25}$$

L'angle de déplacement θ est défini par :

$$\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right) \quad \text{Équation III-26}$$

Et N est le nombre total des paires de pixels de la région d'intérêt (ROI), il est utilisé pour la normalisation :

$$N(r) = \# \left\{ \begin{array}{l} P_i(x_i, y_i) : P_j(x_j, y_j) \in ROI^2; \\ \text{Où } r = \sqrt{(x_i, y_i)^2 + P_j(x_j, y_j)^2} \end{array} \right\} \quad \text{Équation III-27}$$

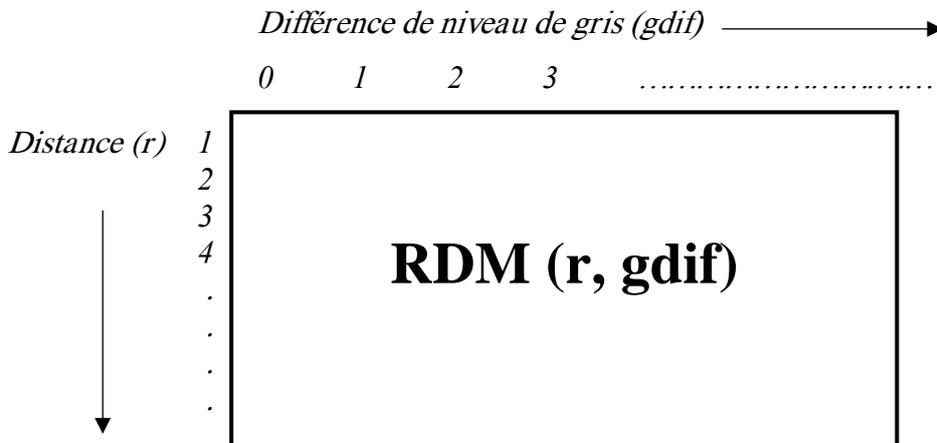


Figure III.5: Run difference matrix (RDM).

Si la direction θ est nulle (direction horizontale), alors la matrice RDM est calculée comme suit :

<i>r</i> \ <i>gdif</i>	0	1	2	3	4	5	6
1	0/25	14/25	6/25	4/25	1/25	0/25	0/25
2	0/20	7/20	10/20	1/20	2/20	0/20	0/20
3	5/15	5/15	1/15	2/15	2/15	2/15	0/15
4	1/10	5/10	1/10	2/10	0/10	0/10	1/10

Figure III.6: La matrice RDM de l'image (Figure III.2).

Il est possible de calculer à partir de la matrice RDM déterminée ci-dessus les trois vecteurs caractéristiques :

La répartition de différence des niveaux grise (*RDNG*)

$$RDNG_{gdif} = \sum_{r=1}^{r_{max}} RDM(r, gdif) \quad \text{Équation III-28}$$

La distribution de différence moyenne (*DDifM*)

$$DDifM_r = \sum_{gdif}^{G-1} RDM(r, gdif) \cdot gdif \quad \text{Équation III-29}$$

Et la distribution de distance moyenne (*DDisM*).

$$DDisM_{gdif} = \sum_{r=1}^{r_{max}} RDM(r, gdif) \cdot r \quad \text{Équation III-30}$$

Où r_{max} est la distance maximale de r , G est le niveau de gris maximal. A partir de ces trois caractéristiques, il est possible d'extraire d'autres caractéristiques telles que :

a) L'écart de grande différence (EGD)

Ce paramètre permet de mesurer la prédominance des grandes différences de niveaux de gris.

$$EGD = \sum_{gdif=0}^{G-1} RDNG(gdif) \cdot \ln(k/gdif) \quad \text{Équation III-31}$$

Où ' k ' est une constante.

b) L'acuité

C'est le paramètre qui mesure le contraste de la zone d'intérêt.

$$Acuité = \sum_{gdif=0}^{G-1} RDNG(gdif) \cdot (gdif)^3 \quad \text{Équation III-32}$$

c) Le Moment d'ordre 2 de la RDNG (SMG)

Ce paramètre mesure la variation des différences de niveau de gris.

$$SMG = \sum_{gdif=0}^{G-1} (RDNG(gdif))^2 \quad \text{Équation III-33}$$

d) Accentuation de longue distance pour une grande différence (ALDGD)

C'est le paramètre qui mesure la proéminence des grandes différences ' $gdif$ ' des niveaux de gris sur une longue distance ' r '.

$$ALDGD = \sum_{gdif=0}^{G-1} DDisM(gdif)(gdif)^2 \quad \text{Équation III-34}$$

e) Le Moment d'ordre 2 de la DDifM (SMD)

$$SMD = \sum_{gdif=0}^{G-1} (DDifM(r))^2 \quad \text{Équation III-35}$$

III.2.1.4. La transformée de Fourier

La transformée de Fourier est l'une des méthodes les plus utilisées en analyse de texture d'une image. A partir de la transformée de Fourier, nous pouvons calculer généralement le spectre de puissance qui est défini comme le carré du module de la transformée de Fourier et qui est, d'après [Petrou 2006] très informatif sur la périodicité d'une texture.

La transformée de Fourier est un outil bien adapté à l'analyse des comportements harmoniques. Cependant elle n'est pas appropriée à l'analyse de formes localisées spatialement. Par ailleurs, la transformée de Fourier est difficilement assimilable par les chercheurs étant donné qu'il est difficile d'exploiter et d'interpréter directement ses valeurs.

III.2.1.5. L'analyse fractale

La mesure fractale est reconnue pour sa grande capacité à caractériser la rugosité. La courbe de Koch représente l'une des premières courbes fractales à avoir été décrite. Elle a été inventée en 1906 par le mathématicien suédois Helge von Koch. Parmi les méthodes d'analyse fractale, la méthode la plus populaire est celle de comptage de boîtes (box counting).

La dimension fractale d'un ensemble A est définie comme suit :

$$D = \log(N) / \log\left(\frac{1}{r}\right) \quad \text{Équation III-36}$$

Telle que ' N ' est le nombre total des copies distinctes similaires a ' A ' et ' $1/r$ ' correspond au facteur d'échelle avec lequel ' A ' est divisé.

Une deuxième méthode basée sur le mouvement Brownien fractionnaire (fractional Brownian motion) est souvent utilisée pour décrire des phénomènes aléatoires. L'avantage de la dimension fractale du mouvement Brownien fractionnaire est son invariance aux transformations linéaires telles que la translation.

L'analyse fractale est généralement utilisée pour générer des textures ainsi que pour les analyser. Cependant, leur utilisation pour la caractérisation texturale est une tâche assez difficile puisqu'elle nécessite de définir de nombreux paramètres souvent complexes à manipuler.

De plus, de telles méthodes sont souvent coûteuses en termes de temps de calcul. Dans le domaine du traitement des images mammographies, particulièrement le cas de classification de lésions de type masse, cette mesure est appliquée sur le contour en tant que descripteur de forme et non en tant que descripteur de texture qui analyse l'intérieur de la forme.

Les descripteurs de texture sont assez variés et sont couramment utilisés dans le domaine de la mammographie. Les descripteurs de texture ont prouvé leur efficacité dans le traitement de ces petites lésions qui se distinguent par leur forme et par leur répartition dans le sein.

Même si une différence de texture a été révélée entre les masses malignes et celle bénignes, d'après la littérature cette différence n'est pas assez importante et consistante pour en tenir compte [Rangayyan 1997]. Par ailleurs, l'étude des caractéristiques pathologiques des masses mammaires montre que la malignité est fortement liée à la forme et au contour de ces lésions. Par conséquent, on s'intéresse, désormais, aux descripteurs de forme appliqués aux masses mammaires.

III.2.2. Les descripteurs de forme en mammographie

Le degré de malignité dans le cas des masses augmente essentiellement avec l'élévation de la complexité du contour. Pour cette raison, il serait plus judicieux d'étudier chaque cas séparément.

Nous résumons dans la Figure III.7, les différentes formes d'une masse qui nous permettent de mieux comprendre la nature des descripteurs de forme les plus adéquats à chercher.

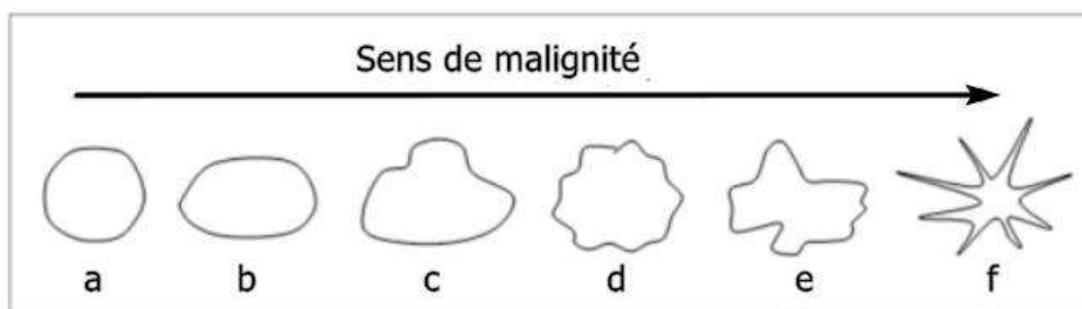


Figure III.7: Relation entre la complexité de la forme et le type de contours de la masse et malignité : a) masse circonscrite circulaire, b) circonscrite ovale, c) macro lobulée, d) micro lobulée, e) irrégulière, f) spéculée. [Bruce 1999]

La forme des masses mammaires peut être modélisée à travers le contour ou bien l'intérieur de l'objet. Les deux approches sont pratiquement équivalentes étant donné que la forme peut être

reconstituée à partir de son intérieur et que l'intérieur peut être reconstitué à partir de son contour. Quel que soit la façon par laquelle la forme a été modélisée, les descripteurs de formes proposés dans la littérature se basent sur des caractéristiques de la forme elle-même. Après plusieurs décennies de recherche dans le traitement d'images, une multitude de descripteurs de forme sont proposés. Nous détaillons dans ce qui suit ces différents types de descripteurs et leurs intérêts.

III.2.2.1. Les descripteurs géométriques

Les descripteurs géométriques permettent de caractériser l'aspect de la forme d'une façon plus ou moins globale et sont généralement indépendants du domaine d'application comme l'aire et le périmètre

a) Le centre de gravité

Cette mesure est étroitement liée à la forme de l'objet, de telle sorte que les coordonnées (x_g, y_g) du centre de gravité sont définies de la manière suivante [Bottigli 2007]:

$$x_g = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{Équation III-37}$$

$$y_g = \frac{1}{N} \sum_{i=1}^N y_i \quad \text{Équation III-38}$$

Où 'N' est le nombre de points dans la région de la forme et (x_i, y_i) sont les points contenus dans cette région.

Si la forme est représentée par le contour de l'objet :

$$x_g = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad \text{Équation III-39}$$

$$y_g = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - y_i x_{i+1}) \quad \text{Équation III-40}$$

Avec 'N' est le nombre de pixels formant les contours de la forme et (x_i, y_i) sont les points englobés par les contours.

b) L'aire

La caractéristique de forme plus triviale qui peut être calculée à partir d'un objet détecté sur une image est l'aire (noté : A). Il peut être définie comme étant le nombre de pixels à l'intérieur de l'anomalie [Bottigli 2007].

c) Le périmètre

Le calcul du périmètre (*noté : P.*) de la masse est aussi couramment utilisé dans le domaine de diagnostic du cancer du sein. Il désigne le nombre de pixels du contour.

d) La circularité

La mesure de circularité représente le degré de ressemblance entre la forme considérée et un cercle. Ce fameux descripteur noté C peut être calculé en dessinant un cercle centré sur le centre de gravité de la région et de même aire.

Soit R : la région de masse et C_{EQ} : le cercle équivalent de même aire que la zone de la masse et ayant comme centre le centre de gravité R .

La mesure de circularité peut alors se définir comme suit [Cheikhrouhou 2007]:

$$C = \frac{\text{aire}(R \cap C_{EQ})}{\text{aire}(R)} \quad \text{Équation III-41}$$

La figure suivante explique d'avantage le principe de calcul de la circularité. Cette mesure est d'une valeur comprise entre 0 et 1. La région est d'autant plus circulaire que sa mesure de circularité est proche de 1.

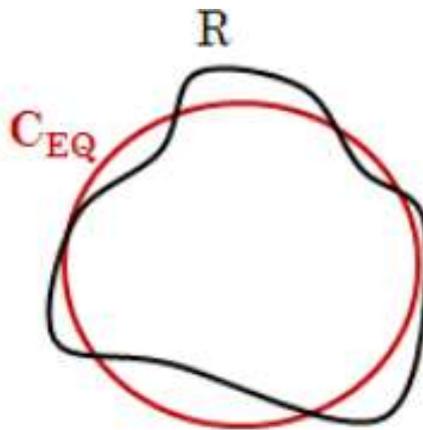


Figure III.8: Calcul de la mesure de circularité [Cheikhrouhou 2007].

En dépit de sa simplicité, ce descripteur est très utilisé dans plusieurs domaines d'analyse de forme. Cependant, son utilité dans le diagnostic du cancer du sein réside dans la nature du caractère pathologique des masses mammaires.

e) La rectangularité

Un autre descripteur de forme intitulé rectangularité ou boîte englobante est souvent utilisé pour décrire le taux de rectangularité et d'allongement d'une région [Yang 2008].

On désigne par R la région étudiée et la boîte englobante (B_E) qui est définie par le plus petit rectangle contenant la lésion étudiée (Figure III.9). La mesure de rectangularité notée $Rect$ peut alors se définir comme suit :

$$Rect = \frac{aire(R)}{aire(B_E)} \quad \text{Équation III-42}$$

Selon cette formule, plus l'objet à décrire est rectangulaire, plus la valeur de rectangularité est proche de 1. Toutefois, cette formule de rectangularité ne tient pas compte de l'orientation de l'objet. En effet, la valeur de rectangularité d'un rectangle parfait et droit est égale à 1 alors que la rectangularité de ce même rectangle mais considéré incliné est différente de 1. Ainsi, ce descripteur est sensible à la rotation.

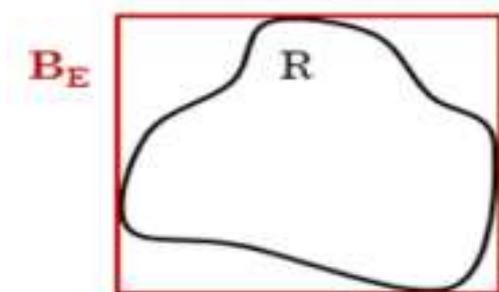


Figure III.9: Calcul de la mesure de rectangularité [Cheikhrouhou 2007].

f) La compacité

Il s'agit d'une mesure de la complexité du contour vis à vis de l'aire [Peura 1997]. Elle est notée par Com et est calculée par la fonction :

$$Com = \frac{P_2}{A} \quad \text{Équation III-43}$$

Où P est le périmètre de l'objet et A est l'aire en pixels.

Cette valeur de compacité permet de distinguer une forme irrégulière d'une forme simple puisqu'elle attribue à la forme irrégulière une valeur plus élevée. En dépit de sa simplicité, ce descripteur présente l'avantage d'être invariant aux transformations géométriques telles que la translation, la rotation et la variation d'échelle.

g) La méthode de coque convexe

Au cours des dernières années, les chercheurs ont utilisé de nombreux types des caractéristiques pour distinguer entre les masses malignes et bénigne. Dans l'étude de Behnam et al. [Behnam 2010], six nouvelles caractéristiques de forme ont été proposées :

1. Zone de différence

Cette zone est définie par la différence entre la zone de coque convexe et la zone de la masse. Pour définir autres caractéristiques, une fonction nommée *variation* (noté : *var*) est tout d'abord défini comme « la projection de la distance entre les pixels les plus éloignés d'une région de masse à tous les angles ».

La coque convexe, ce qui est le plus petit polygone convexe qui contient le contour de la masse et de la région de masse, est représenté sur la Figure III.10. La zone de la coque convexe est le nombre réel de pixels dans la coque convexe de la masse, et la zone de masse est le nombre de pixels dans la région de masse.

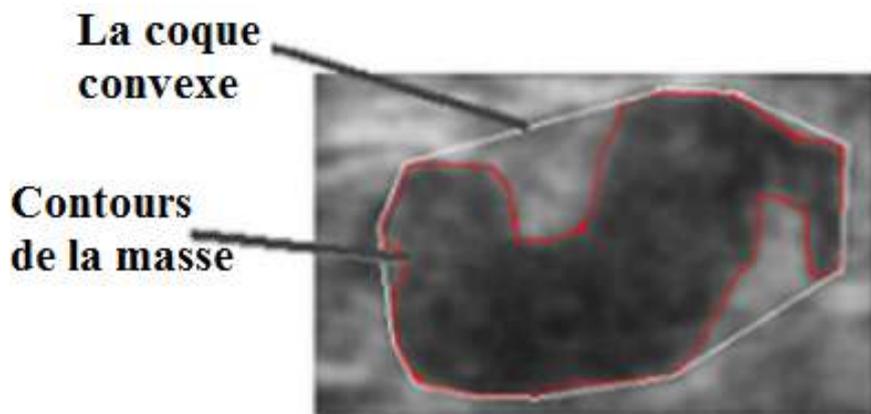


Figure III.10: La coque convexe et le contour d'une masse mammaire [Behnam 2010].

Cinq caractéristiques sont extraites de la fonction *variation* :

2. La Moyenne des variations

La moyenne est la grandeur que devraient avoir toute les variations pour que leur total soit inchangé. C'est un critère de position. Dans la plupart des cas, le total formé par les variations d'une population est la somme de leurs valeurs. La moyenne est alors la moyenne arithmétique.

$$M_v = \frac{1}{L} \sum_{i=1}^L \text{var}(i) \quad \text{Équation III-44}$$

3. La variance des variations

La variance indique de quelle manière les variations '*var(i)*' se dispersent autour de sa moyenne ou son espérance '*M_v*'. Une variance élevée indique que les variations sont très écartées les unes des autres, et vice versa. Elle est nulle lorsque toutes les valeurs sont identiques. Une variance n'est donc jamais négative.

$$VAR_v = \frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^2 \quad \text{Équation III-45}$$

4. L'asymétrie des variations

Le coefficient de dissymétrie (*skewness* en anglais) correspond à une mesure de l'asymétrie de la distribution des variations.

$$SK_v = \frac{\frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^3}{\left(\frac{1}{L} \sum_{i=1}^L (\text{var}(i) - M_v)^2\right)^{3/2}} \quad \text{Équation III-46}$$

5. L'aplatissement des variations

Le coefficient d'aplatissement (*kurtosis* en anglais) correspond à une mesure contrario de la pointicité de distribution des variations. Il mesure la disposition des masses de probabilité des variations autour de leur centre ' M_v ', c'est-à-dire, d'une certaine façon, leur regroupement proche ou loin du centre de probabilité.

$$APP_v = \frac{(1/L) \sum_{i=1}^L (\text{var}(i) - M_v)^4}{\left((1/L) \sum_{i=1}^L (\text{var}(i) - M_v)^2\right)^2} - 3 \quad \text{Équation III-47}$$

6. L'entropie des variations

Le terme entropie caractérise le degré de désorganisation ou de manque d'information d'un système.

$$E_v = - \sum_{i=1}^L \text{var}(i) \log_2 \text{var}(i) \quad \text{Équation III-48}$$

La Figure III.11 illustre la base de caractéristiques (Équation III-44Équation III-48). Pour ces caractéristiques, la sous-image ROI devrait être mise en rotation à différents angles. Rotation avec intervalle de 5° est utilisée. Pour chaque angle des pixels avec des coordonnées minimales Y ('a' dans la figure) et les coordonnées maximales Y ('b' sur la figure) sont trouvé. La projection de la distance entre les pixels sur l'axe Y est calculée et affichée avec une ligne en pointillés dans la figure.

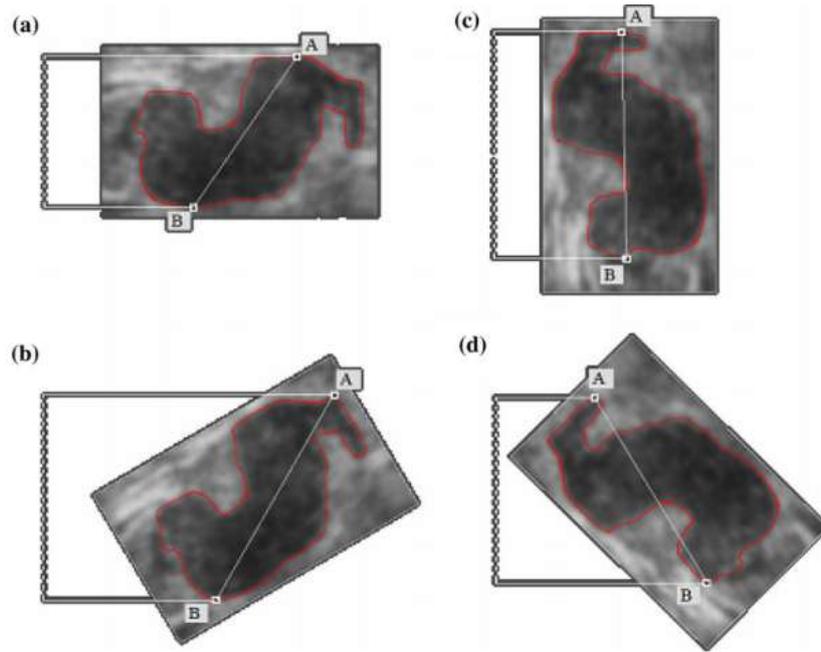


Figure III.11: *Un exemple de caractéristiques Équation III-44 Équation III-48 à l'aide d'une masse en rotation à des angles différents [Behnam 2010].*

III.2.2.2. Les descripteurs spécifiques

Les chercheurs dans le domaine de la reconnaissance de forme se sont penchés sur la proposition de descripteurs dits spécifiques, plus précis et plus informatifs sur les détails de la forme. Ces propositions visent l'obtention de descripteurs robustes capables de révéler l'aspect général d'un objet tout en préservant ses caractéristiques partielles.

a) La courbure

La courbure noté '*Curv*' a été couramment utilisée dans le contexte de l'analyse de formes dans plusieurs domaines. Elle a été reconnue par sa capacité à caractériser les formes d'objets. De nombreuses approches proposées suggèrent minimisant la courbure tout en respectant les contraintes géométriques de directions tangentes calculé sur le contour. Cette méthode est robuste et utilisée largement dans le système de (*DxAO*). En général, la courbure à un point donné '*b*', d'une courbe est définie comme l'inverse du rayon du cercle osculateur dans '*b*'. Le cercle osculateur peut être obtenue comme suit : étant donné deux points '*a*' et '*c*' à proximité de '*b*', nous cherchons le cercle qui traverse les point '*a*', '*b*' et '*c*'. Si ces points sont alignés, le cercle aura un rayon infini et la courbure devient alors nulle.

$$Curv = \frac{1}{R}$$

Équation III-49

Le rayon du cercle osculateur est défini comme suit :

$$R = \frac{A.B.C}{\sqrt{(A+B+C)(A-B+C)(A+B-C)(B-A+C)}} \quad \text{Équation III-50}$$

Avec $A = |ab|$, $B = |bc|$ et $C = |ac|$

Récemment, plusieurs méthodes de calcul de la courbure ont été proposées. Nguyen et al. [Nguyen 2007] ont amélioré l'estimation des cercles osculateurs proposée par Coeurjolly et al. [Coeurjolly 2001] en utilisant des segments flous. Ainsi, l'estimation proposée est plus adaptée aux contours bruités et plus significative dans le cas de points non connexes. Une autre approche proposée par Kerautret et Lachaud [Kerautret 2008] suggère de minimiser la courbure en respectant les contraintes géométriques issues des directions tangentes calculées sur le contours.

b) La longueur radiale normalisée (LRN)

Kilday et al. [Kilday 1993] ont élaboré un ensemble de six caractéristiques pour décrire une forme géométrique en se basant sur la notion de la longueur Radial normalisée (LRN). La longueur radiale (LR) a été définie comme la distance euclidienne entre le centre de gravité de l'objet 'C' et l' $i^{\text{ème}}$ pixel de contours ' P_i '. Ainsi, la longueur radiale normalisée est le résultat de normalisation par la distance maximale trouvée.

$$LRN(i) = \frac{\sqrt{(x(i) - x_g)^2 + (y(i) - y_g)^2}}{Max_{LR}}, i = \{1, 2, \dots, N\} \quad \text{Équation III-51}$$

Avec $(x(i), y(i))$ et (x_g, y_g) les coordonnées de l' $i^{\text{ème}}$ pixel et le centre de gravité respectivement, et 'N' est le périmètre de la masse.

1. La moyenne de LRN (M_{LRN}) : La moyenne de (LRN) est définie par la fonction suivante :

$$M_{LRN} = \frac{1}{N} \sum_{i=1}^N LRN(i) \quad \text{Équation III-52}$$

2. L'écart type de LRN (σ_{LRN}): L'écart type sert à mesurer la dispersion de l'ensemble de $LRN(i)$. Plus il est faible, plus les valeurs sont regroupées autour de la moyenne. Cette mesure décrit l'irrégularité de la forme :

$$\sigma_{LRN} = \sqrt{\frac{1}{N} \sum_{i=1}^N (LRN(i) - M_{LRN})^2} \quad \text{Équation III-53}$$

3. l'Entropie (E) : L'entropie est obtenue à partir de l'histogramme de la longueur radiale. Le périmètre p_k est la probabilité que (LRN) est entre $LRN(i)$ et $LRN(i)+1/N_{bin}$, avec N_{bin} numéro

de l'histogramme normalisée, variant dans l'intervalle [0,1] qui a été divisé en $N_{bins}=100$. La mesure de l'Entropie intègre simultanément la notion de circularité et irrégularité :

$$E = \sum_{k=1}^{100} p_k \log(p_k) \quad \text{Équation III-54}$$

4. Le rapport de surface (A_1) : Le rapport de surface est une mesure du pourcentage de la partie de l'objet hors la région circulaire définie par la moyenne de (LRN):

$$A_1 = \frac{1}{M_{LRN} \cdot N} \sum_{i=1}^N (LRN(i) - M_{LRN}) \quad \text{Équation III-55}$$

5. La rugosité (R) : La rugosité vise à isoler la forme macroscopique de l'objet à partir de la structure fine des bords. Elle fournit des informations sur la moyenne entre les pixels voisins, les contours irréguliers fournissent des valeurs élevées de rugosité.

$$R = \frac{1}{N} \sum_{i=1}^N (LRN(i) - LRN(i+1)) \quad \text{Équation III-56}$$

6. Le taux de croisement en zéro (ZC_1) : Il calcule le nombre de fois que la droite définie par la moyenne de (LRN) intercepte les contours de la lésion, il fournit le degré de spéculation du contour.

c) La longueur Radial normalisée modifiée (LRNM)

À partir des propriétés de la LRN, Chen et al. [Chen 2009] proposent des descripteurs qui montre des performances dans la classification des opacités mammaires. Les nouveaux descripteurs extraits à partir de la (LRNM) sont la différence entre déviation standard (σ_{diff}), l'entropie de la différence entre $LRN(i)$ et LRN_{MAX} nommé E_{diff} , la zone ratio ' A_2 ' et le nombre de passage par zéro ' ZC_2 '

1. Différence des déviations standard (σ_{diff}): σ_{diff} peut estimer le degré d'irrégularité du contour, de sorte que, si le contour devient de plus en plus irrégulier, σ_{diff} atteint les plus grandes valeurs.

$$\sigma_{diff} = |\sigma - \sigma_{ma}| \quad \text{Équation III-57}$$

2. La différence de l'entropie (E_{diff}): Ce descripteur fournit la distribution de la différence entre $LRN(i)$ et Max_{LRN} . P_k est la probabilité que $|LRN(i) - Max_{LRN}|$ est entre $|LRN(i) - Max_{LRN}|$ et $|LRN(i) - Max_{LRN}| + 1/N_{bins}$.

$$E_{diff} = \sum_{k=1}^{100} p_k \log(p_k) \quad \text{Équation III-58}$$

3. Le rapport de surface modifiée (A_2) : La description est obtenue par l'équation suivante :

$$A_2 = \frac{1}{M_{LRN} \cdot N} \sum_{i=1}^N (LRN(i) - MAX_{LRN}) \quad \text{Équation III-59}$$

4. Le taux de croisement en zéro modifié (ZC_2): C'est la mesure du nombre de fois que la courbe $LRN(i)$ intercepte MAX_{LRN} .

III.2.3. Les descripteurs de contours

Le nombre de caractéristiques de contours utilisée dans la littérature pour la catégorisation des masses mammographiques est beaucoup moindre que les descripteurs de forme et de textures, alors que le contour est l'une des caractéristiques les plus importantes pour le diagnostic des anomalies mammaires. Les caractéristiques de contours proposées par Bagheri et al. [Bagheri 2012] sont obtenues à partir d'un ensemble d'ondelettes. Ces caractéristiques ont montré une grande précision pour la classification des masses mammographiques.

Soit Ω la marge de masse segmentée, le point Q_1 est le centre de gravité de la masse, il est calculé par :

$$Q_1 = \frac{1}{|\Omega|} \sum_{i,j \in \Omega} (i, j) \quad \text{Équation III-60}$$

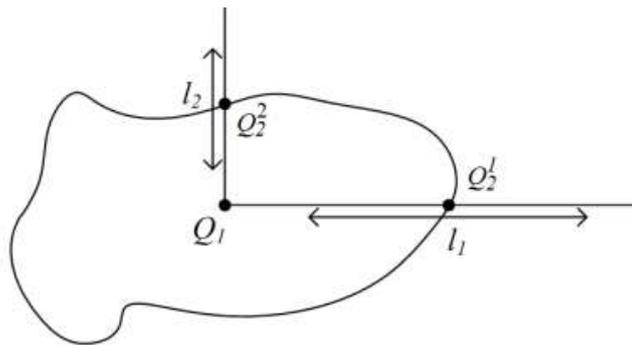


Figure III.12: Exemple de deux ondelettes l_1 et l_2 .

Si Q_2 est le pixel qui est sélectionné en traversant les contours de la masse ; ainsi, l'ondelette l aura la même direction que la ligne Q_1Q_2 . La trace de la marge Ω est parcouru et les ondelettes sont placées tous les 5 pixels selon Bagheri et al. [Bagheri 2012].

La longueur de l'ondelette l est calculée par la fonction suivante :

$$|l| = 2^{(\log_2 r)+1}$$

Équation III-61

Où r est la longueur de $\overline{Q_1 Q_2^i}$. Ainsi, la longueur de l_i est une puissance de deux et appropriée pour la transformation en ondelettes, jusqu'à présent, la longueur et le sens de l'ondelette l_i sont déterminées, les ondelettes seront localisées de telle sorte que leurs éléments centraux soient Q_2^i . La Figure III.12 représente un exemple de deux ondelettes l_1 et l_2 , où Q_1^1 et Q_2^2 correspondent aux pixels centraux sur le contour Ω , un exemple réel de la construction des ondelettes est illustré dans la Figure III.13.

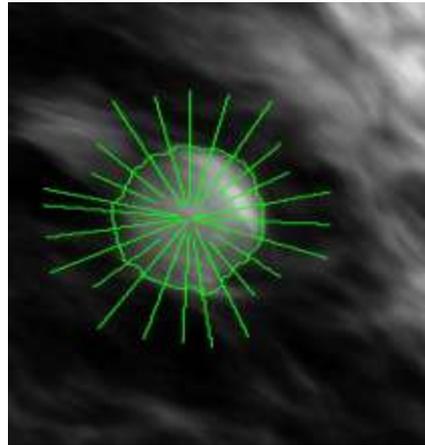


Figure III.13: La représentation des ondelettes sur une masse mammographique
[Bagheri 2012]

Les ondelettes tracées sont la base pour l'analyse de la marge des masses mammographiques, dans la méthode proposée par Bagheri et al. [Bagheri 2012], les auteurs se sont concentré sur l'analyse de ces signaux pour obtenir des informations sur le type de contours. A cet effet, la transformation en ondelettes unidimensionnel a été appliqué à chacune d'entre elles utilisant le filtre de haar [Haar 1910].

Soit n le nombre d'ondelettes et l_i la $i^{\text{ème}}$ ondelette. Après L niveau de décomposition, les coefficients d'ondelettes $\{a_L, d_L, d_{L-1}, \dots, d_1\}$ sont obtenue. L est calculé comme:

$$L = \log_2 |l_i| - 2$$

Équation III-62

Où $|l_i|$ est la longueur de la $i^{\text{ème}}$ ondelette. Il faut seulement utiliser les coefficients de détail d_i , en raison de la texture ambiguë des mammographies d_1 a été écarté. Les valeurs de chaque niveau sont normalisées indépendamment pour être entre 0 et 1.

En utilisant des coefficients d'ondelettes, la distribution de probabilité de contours est calculée pour chaque l_i . Notons que nous ne cherchons pas de contours, mais seulement la distribution de probabilité des contours soient calculées.

Soit D la longueur de d_2 , alors la fonction de probabilité (P) des contours a également D éléments. La probabilité de contours P_i^j ($j^{\text{ième}}$ élément de P_i) est calculée par la fonction :

$$P_i^j = S_j / \sum_{j=1}^D S_j \quad \text{Équation III-63}$$

$$S_j = \sum_{t=2}^L w_t d_t^{u(t,j)} \quad \text{Équation III-64}$$

$$u(t, j) = \left\lfloor \frac{j}{2^{t-2}} \right\rfloor \quad \text{Équation III-65}$$

Où d_t^u est le $u^{\text{ième}}$ élément du $t^{\text{ième}}$ niveau de détails dans la décomposition, w_t est le poids qui est associé à la $t^{\text{ième}}$ niveau de détails, il est calculé par la fonction suivante:

$$w_t = t / \sum_{t=2}^L t \quad \text{Équation III-66}$$

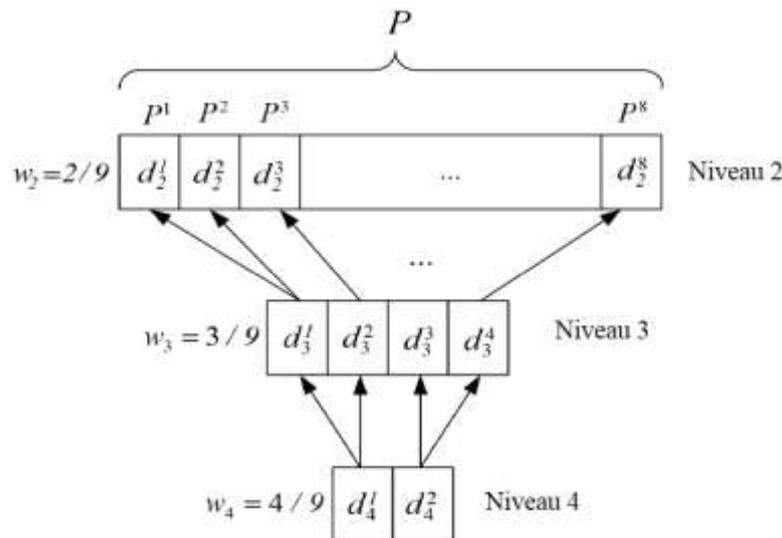


Figure III.14: Modèle schématique du calcul des probabilités de contour de chaque ondelette.

Le plus faible poids w est assigné au deuxième niveau de décomposition et le poids le plus élevé est assigné à $L^{\text{ième}}$ niveau, où $\sum_t w_t = 1$. La raison du choix de ces poids, c'est que l'énergie dans

les niveaux supérieurs est plus élevée que dans des niveaux inférieurs dans la décomposition de chaque ondelette. En outre, l'effet de bruit est aussi plus sensible aux niveaux inférieurs. Pour plus de clarification de l'équation (III.64-III.66) un exemple schématique est illustré dans la Figure III.14, où $L=4$ et $D=8$, chaque P^j est calculé par la multiplication des éléments correspondant par ses poids et après l'additionnés au sens des flèches. Enfin, pour la contrainte de $\sum_j P^j = 1$ les éléments de P sont normalisées à leurs sommes comme il a été formulée par l'équation (III.64).

Pour chaque fonction de distribution de P_i qui a été obtenu à partir de l'ondelette l_i les paramètres suivants sont calculés :

1. Le coefficient d'aplatissement (Kurtosis)

$$k_i = \frac{(1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^4}{((1/D) \sum_{j=1}^D (P_i^j - \bar{P}_i)^2)^2} - 3 \quad \text{Équation III-67}$$

2. L'entropie

$$E_i = - \sum_{j=1}^D P_i^j \log_2 P_i^j \quad \text{Équation III-68}$$

3. L'index de la probabilité maximale

$$R_i = (\text{argmax} P_i^j) - \frac{D}{2} \quad \text{Équation III-69}$$

Le coefficient d'aplatissement (Kurtosis) mesure le nombre de pics sur la fonction de probabilité de contours (P_i). Pour des contours bien définies des valeurs élevées sont obtenues par rapport aux contours mal définis ou effacés. En général, les contours circonscrit ont une transition abrupte le long d'une ondelette. Les contours bien définis possèdent une plus faible entropie. L'index de la probabilité maximale est une autre mesure qui est décalé pour être zéro sur les contours, positif pour l'extérieur et négatif pour l'intérieur de la marge. Cet indice est utilisé pour repérer les variations des endroits les plus probables parmi les contours. Chacun de ces vecteurs ont n éléments (n est le nombre d'ondelettes).

Les caractéristiques suivantes peuvent être obtenues à partir de ces vecteurs.

- ✓ Moyenne et déviation standard de K.
- ✓ Moyenne et déviation standard de E.
- ✓ Variation total de R.

$$\sum_i |R_i|$$

Équation III-70

L'équation de déviation standard est :

$$\sigma = \sqrt{\frac{1}{D} \sum_{i=1}^N (P_i - \mu)^2}$$

Équation III-71

Les caractéristiques Kurtosis et entropie ne dépendent pas de l'exactitude de la segmentation, car elles sont obtenues sur le long des ondelettes et si les contours varient à cause d'une certaine déflation ou inflation, elles restent les mêmes.

III.2.4. Autres caractéristiques

L'âge et la densité sont des caractéristiques humaines. La description de l'interprétation de ces traits humains est basée sur Breast Imaging Reporting and Data System (BI-RADS) [Lévy 2005], qui sont utilisés comme un guide pour la normalisation des rapports mammographiques, le Tableau III-1 illustre l'interprétation BI-RADS pour les valeurs de densité du sein.

Valeur de densité	Interprétation BI-RADS
1	Le sein est presque entièrement de matières grasses.
2	Il y a des densités de fibroglandulaires dispersés qui pourraient masquer une lésion sur la mammographie
3	Le sein est hétérogène dense cela peut diminuer la sensibilité de la mammographie.
4	Le tissu du sein est très dense, ce qui réduit la sensibilité de la mammographie.

Tableau III-1: *Interprétation BI-RADS pour les valeurs de densité du sein.*

III.3. Conclusion

L'étape de l'état de l'art concernant les différents descripteurs appliqués en mammographie est nécessaire. Dans ce contexte, différents descripteurs de texture sont présentés afin d'analyser les méthodes les plus pertinentes. Dans le cas de traitement de lésions de type masses, il est recommandé de s'appuyer sur des descripteurs de forme. Dans ce cadre, différents descripteurs de forme appliqués en mammographie ont été étudiés afin d'analyser les avantages et les inconvénients de chaque proposition. En outre, l'étude des descripteurs de contour a montré que ces derniers ont une très grande importance dans l'interprétation des masse mammaires.

Cette étude détaillée nous permet de tirer profit des points forts des descripteurs existants et de tenir compte des lacunes rencontrées.

Chapitre

4

LES ÉLÉMENTS D'UN SYSTÈME DE DÉPISTAGE AUTOMATIQUE DU CANCER DU SEIN (DACCS)

IV.1. Introduction

Ces dernières années, d'importants efforts ont été faits pour développer des systèmes de dépistage automatique du cancer du sein (DACS) qui peuvent aider les radiologues à détecter et à caractériser les anomalies malignes et bénignes en mammographie numérique. Deux genres de systèmes ont été développés. Le premier type est la détection assistée par ordinateur (DéAO) qui est utilisé pour améliorer la précision des radiologues pour la détection du cancer du sein [Giger 2001] [Vyborny 2000]. Le deuxième type est le diagnostic assisté par ordinateur (DxAO), ce système classe les régions détectées en catégories malignes ou bénignes pour aider les radiologues à reconnaître l'étape suivante (un examen complémentaire, une biopsie ou un suivi à court terme).

Ce chapitre présente et analyse la littérature de référence en traitement informatique des mammographies numériques. Dans la section IV.2, les intérêts des systèmes de détection assistée par ordinateur sont introduits avec la revue de la littérature des techniques informatiques de détection du cancer du sein présentées à la section IV.3. La section IV.4 met en lumière et identifie les principales étapes des systèmes de détection et de diagnostic automatiques, en présentant les concepts fondamentaux du traitement numérique des images et les techniques de segmentation d'images utilisées dans les applications de mammographie numérique.

IV.2. Les systèmes de dépistage automatique du cancer du sein

La plupart des systèmes de dépistage du cancer du sein (DACS) comprennent cinq étapes principales : le prétraitement, la segmentation, l'extraction de caractéristiques, la sélection de caractéristiques et la classification. Le processus de dépistage automatique du cancer du sein débute par une région de lésion ou une région d'intérêt (ROI) de l'anomalie et offre en sortie la probabilité de malignité ou une recommandation à suivre.

La performance d'un système (DxAO) dépend de l'optimisation de la phase de sélection de caractéristiques plus que les méthodes de classification utilisées [Elfarra 2012]. Cependant, l'espace de caractéristiques est très large et complexe en raison de la grande diversité des tissus normaux et de la variété des anomalies. L'utilisation de caractéristiques excessives peut dégrader les performances des algorithmes et augmenter la complexité des classifieurs.

Les systèmes DéAO identifient et marquent les régions suspectes sur les mammographies pour les porter à l'attention des radiologues. Ces systèmes minimisent les erreurs de recherche, de perception et d'interprétation même si les radiologues ne parviennent pas à reconnaître des

anomalies suspectes. Un système DACS est destiné à être utilisée après que le radiologue a terminé l'évaluation des images mammographiques et a pris une décision initiale [Hutt 1996]. À titre d'exemple, si le radiologue identifie une région anormale lors de la lecture initiale d'une mammographie et que cette zone n'est pas marquée par le système DéAO, il est conseillé aux radiologues d'interpréter la mammographie comme positive et de rappeler le patient ; car le système DACS est une adjonction à la mammographie numérique pour diminuer les erreurs de recherche, de détection et d'interprétation, le radiologue prend une décision finale s'il existe une anomalie cliniquement significative et décide si un autre examen de diagnostic supplémentaire est recommandé [Jirari 2005]. L'espoir réside dans le fait que les systèmes DACS amélioreront la sensibilité de la mammographie numérique sans augmenter sensiblement les taux de rappel des patients [Bozek 2008].

IV.2.1. Les avantages des systèmes DACS

Le dépistage humain des mammographies est souvent effectué inconsciemment, sans définition de règles de décisions, ce qui rend la détection et le dépistage assistée par ordinateur une tâche très difficile [Masala 2006].

L'analyse radiologique de la mammographie est faillible, le taux d'échec a été augmenté à cause de la tâche répétitive et fatigante de la détection des anomalies, la mauvaise qualité de l'image, la subtilité de certaines anomalies, l'occlusion des structures anatomiques de l'organe mammaire, la faible prévalence de la maladie et la complexité de la structure mammaire. Ces difficultés peuvent être maîtrisées par des approches telles que « la double lecture », qui fournit une double perception et interprétation. De toute évidence, cette procédure est trop coûteuse, complexe, et prend du temps, en particulier dans les programmes de dépistage avec une grande quantité d'images mammographiques. Le développement des systèmes DACS en tant qu'un deuxième lecteur représente une alternative complémentaire [Mencattini 2010].

Selon Ciato et al. [Ciatto 2003], le système DACS avait presque la même performance de simulation de double lecture conventionnelle. Cependant, Khoo, Taylor et Given-Wilson [Khoo 2005] indiquent que les systèmes DACS augmentent la sensibilité de la lecture unique de 1,3%, alors que la double lecture augmente la sensibilité de 8,2%.

Les ordinateurs sont cohérents et infatigables et ne nécessitent pas des années de pratique pour acquérir l'expérience nécessaire pour analyser les mammographies [Simonetti 1998]. Par conséquent, les systèmes DACS sont plus utiles dans ces situations et dans d'autres circonstances telles que la mammographie de dépistage ; il a été observé que jusqu'à 30% de lésions sont manqué par les radiologues en raison d'un mauvais diagnostic ou d'une mauvaise

interprétation et quand il y a un grand volume d'examens avec une faible incidence de la maladie [Bird 1992], [Masala 2006], dont environ les deux tiers des anomalies sont des lésions évidentes a posteriori.

Les études indiquent que les radiologistes ont un diagnostic faux négatif de 21%. Les systèmes DACS ont le potentiel de réduire ce taux de faux négatifs de 77% [Warren 2000]. Cependant, il y a une certaine controverse sur l'efficacité des systèmes DACS en comparaison avec la performance des radiologues. Les anomalies malignes peuvent être ignorées si les signes pathologiques sont subtils, elles peuvent injustement rejeter par le radiologue comme étant bénignes. Dans ce cas, l'utilisation d'un système DxAO ajouterait du poids à la lésion comme anormale, en réduisant ainsi la possibilité d'erreurs de classification. Warren et al. [Warren 2000] ont observé que plusieurs types d'anomalies malignes très précoces rétrospectivement montrent seulement des subtils changements, mais il est prouvé que les systèmes de DxAO sont suffisamment sensibles pour la détection de tels cas.

De plus, parmi les masses référées à des biopsies chirurgicales, seulement 10 à 20% sont réellement malignes [Simonetti 1998]. Les systèmes DéAO ont en général de bonnes performances pour la détection des micro-calcifications qui peuvent atteindre jusqu'à 99%, le taux de détection des masses mammaires allant de 75 à 89 % [Houssami 2009]. Cependant, les anomalies de type « distorsions architecturales » ne peuvent pas être détectées avec autant de précision [Baker 2003].

Baker et al. [Baker 2003], ont étudiés la sensibilité de deux systèmes DACS commerciaux aux distorsions architecturales ; il a été remarqué que plus de la moitié des cas n'ont pas été détectés. Des améliorations doivent encore être apportées pour augmenter la détection et la classification de ce type d'anomalies.

Les conséquences d'un mauvais diagnostic des lésions bénignes diagnostiquées comme malignes est qu'une biopsie qui implique des coûts et des effets psychologiques tels que l'anxiété des femmes, l'inconfort et le stress. Cependant, le coût et les conséquences d'un cancer manqué sont bien supérieurs à ceux d'une lésion bénigne mal diagnostiquée comme maligne [Rangayyan 2007].

Les systèmes DACS ont besoin d'une image numérique pour l'analyse des mammographies et la caractérisation des anomalies ; dans le cas des clichés mammographiques, une phase de numérisation est nécessaire [Simonetti 1998]. L'utilisation des systèmes DACS sur la base de la mammographie numérique présente des avantages par rapport aux clichés mammographiques qui doivent être numérisés, en plus du temps et de l'argent de la

numérisation et la qualité de l'image qui diminue. Ainsi, avec la mammographie numérique, il a été remarqué la DéAO augmente les détections et taux des faux positifs des systèmes DxAO sont plus élevés pour les systèmes numériques par rapport aux systèmes des clichés mammographiques [Pisano 2005].

Les systèmes DACS par mammographie numérique présentent généralement une valeur prédictive positive élevée et une haute sensibilité en comparaison avec les radiologues. Cependant, ses faux positifs doivent être réduits afin d'augmenter encore plus la valeur prédictive positive [Simonetti 1998].

Au fur et à mesure que les radiologistes prennent la décision finale, certaines fausses alertes des systèmes DACS sont facilement exclues lorsqu'elles sont des calcifications bénignes ou des artefacts d'images. Cependant, l'effet d'un taux élevé de fausses alertes réduira le potentiel des systèmes DACS à maîtriser les erreurs de classification erronée. Les fausses alertes peuvent également dégrader la performance des systèmes, car elles agissent comme perturbateurs, en attirant l'attention des radiologues loin de régions véritablement anormales. Par conséquent, un système DACS puissant nécessite des algorithmes sensibles et spécifiques [Astley 2004].

Selon Freer et Ulissey [Freer 2001], le nombre de cancers détectés a été augmenté de 19,5% après avoir entamé l'utilisation des systèmes DACS, et la proportion de malignités précoces détectées est passée de 73 à 78%. Le taux de sensibilité est passé de 6,5 à 7,7% et la valeur prédictive positive de la biopsie est demeurée inchangée à 38%. Par conséquent, avec cette étude il a été conclu qu'un système DACS peut améliorer la détection des tumeurs malignes au stade précoce sans un effet excessivement néfaste sur le taux de sensibilité ou la valeur prédictive positive de la biopsie. Une autre étude, présentée par Taplin, Rutter et Lehman [Taplin 2006], a indiqué qu'un système DACS augmentait la spécificité de l'interprétation, mais n'affectait pas la sensibilité car les lésions visibles non calcifiées et qui ne sont pas marquées étaient moins susceptibles d'être évaluées comme anormales par les radiologues ; dans la même étude il a été remarqué que la densité mammaire n'a pas affecté les performances des systèmes DACS.

Cependant, des améliorations doivent encore être apportées afin de réduire au minimum la défaillance de ces systèmes en raison de l'importance du dépistage, en raison de la grande variabilité des caractéristiques et de la difficulté de détection des lésions dans les tissus mammaires.

IV.2.2. L'histoire des systèmes DACS

Le premier article qui traite le problème de dépistage du cancer du sein sur la base de la mammographie a été publié en 1967 par Winsberg et al. [Winsberg 1967], cette étude a été basé sur une comparaison bilatérale, qui a été reconnue comme étant utile dans la mammographie de dépistage avec la visualisation de routine d'un grand nombre d'examen généralement normaux. Le concept de détection informatique ou de diagnostic automatique en radiologie a été établi à cette époque. Bien que certains résultats intéressants aient été signalés, ces premières tentatives n'ont pas été couronnées de succès, les ordinateurs n'étant pas suffisamment puissants, les images numériques n'étant pas facilement accessibles et les techniques avancées de traitement et d'analyse d'images n'étant pas disponibles [Dacruz 2011].

En 1980, les améliorations apportées aux techniques de la vision par ordinateur, à la qualité des mammographies et à la numérisation ont commencé à rendre possible la réalisation des systèmes DACS cliniques [Masala 2006], [Simonetti 1998]. Avant cela, le concept était que l'ordinateur remplacerait les radiologues, en examinant la mammographie, qui a été appelé le diagnostic automatique par ordinateur (DxAO). En raison de cette notion, il y a eu quelques critiques dans la phase précoce à la mise en œuvre de logiciels informatiques pour l'aide au diagnostic. A cette époque, le concept de détection et diagnostic assistée par ordinateur apparaîtra [Dacruz 2011].

Des recherches approfondies ont été menées depuis cette date, principalement en ce qui concerne la détection et le diagnostic assisté par ordinateur et l'acceptation des radiologistes ont encouragé le développement de ces techniques. Le DxAO a été introduit en pratique clinique en avril 1995, à l'Université de Chicago, où les mammographies de dépistage sont numérisées et analysées pour des masses et des calcifications par un poste de travail clinique (workstation) [Simonetti 1998].

L'administration des aliments et de drogues des États-Unis (Food and Drug Administration) (FDA) a approuvé le premier système de DxAO en mammographie de dépistage en 1998. En 2001, seulement 130 unités de DxAO étaient en opération clinique aux États-Unis. En 2005, cela a augmenté jusqu'à 1600 [Masala 2006].

Le premier système DéAO approuvé par la FDA a été le logiciel de détection assistée par ordinateur ImageChecker^{MD} de R2 Technology Inc [Hologic 2016], servant à la mammographie numérique est conçu pour offrir le niveau de détection le plus élevé, ainsi que des options de travail souples. Il renferme des outils perfectionnés, comme le repérage des zones d'intérêt, et offre aux radiologistes une plus grande souplesse dans le choix de leurs préférences de lecture.

Ce système incorpore un numériseur pour convertir les clichés mammographies au format numérique, des algorithmes de détection et des alertes apparaissent sur des anomalies suspectes. Il a subi quelques améliorations pour renforcer la détection et pour offrir un examen détaillé sur les régions suspectes, comme la présence d'un seuil pour déterminer si un message d'alerte soit affiché ou non. Le seuil est fixé pour atteindre l'équilibre optimal entre la sensibilité et la spécificité. Le taux de détection des calcifications atteint 98,5%, la sensibilité 0,74 de faux positifs par cas (ensemble de quatre images). La précision de détection des masses était de 85,7% à 1,32 fausses alertes positives par cas [Sampat 2005].

En 2002, deux nouveaux systèmes de DxAO mammographiques ont été approuvés : MammoReader^{MC} de iCad et Second Look^{MC} de CADx. Ils ont un principe similaire à l'Image Checker^{MD}, mais avec des algorithmes différents, et donc répondre différemment aux lésions potentielles [Dacruz 2011].

MammoReader^{MC} a été conçu pour détecter les signes primaires du cancer du sein dans les images mammographique, y compris le regroupement de micro-calcifications, les masses bien définies et mal définies, les lésions spiculées, les distorsions architecturales et les densités asymétriques. La sensibilité globale rapportée était de 89,3% (91,0% dans les cas où les micro-calcifications étaient le seul signe de malignité et 87,4% dans les cas restants où des masses malignes étaient présentes).

Second Look^{MC} détecte principalement les micro-calcifications et masses. La sensibilité du système était de 85% pour le dépistage des cancers détectés.

IV.3. Les techniques utilisées par les systèmes DACS

Les chercheurs ont mis au point des systèmes de détection des mammographies pour détecter les tumeurs du sein [Ganesan 2013], [Wei 2009], [Singh 2011], [Bhateja 2013], [Auephanwiriyakul 2005] et des microcalcifications [Cheng 1998], [Gong 2004]. De nombreuses méthodes de classification ont été déjà développées et appliquées pour caractériser les masses mammographiques comme bénignes ou malignes. Ces méthodes comprennent les ondelettes [Soltanian 2004], [Ferreira 2003], [Bruce 1999], les fractales [Rangayyan 2007], [Guo 2009], les agents [Melouah 2008], les méthodes statistiques combinées aux automates cellulaires [Benmazou 2014], [Petroudi 2003], les méthodes basées sur la vision [Polakowski 1997]. Les méthodes basées sur RNA sont particulièrement populaires [Chou 2004], [Karabatak 2009], [Halkiotis 2007] [Wu 1992] car elles peuvent être mises en œuvre en utilisant le traitement parallèle pour traiter efficacement de grandes quantités d'information, elles peuvent analyser et classer des modèles même lorsque l'information est bruyante, complexe ou

incomplète et adapter leur comportement à une variété de données d'apprentissage. Les réseaux de neurone artificiel ont été utilisés pour détecter et classer les microcalcifications et les masses mammaires sur mammographie.

Yanfeng et al. Dans [Li 2015] ont proposé une méthode pour différencier entre les masses bénignes et malignes à l'aide d'une analyse textural avec des stratégies de sous-échantillonnage multiples. Le classifieur K plus proches voisins (KPPV) est appliqué pour chaque stratégie de sous-échantillonnage et le résultat de la classification est obtenu en effectuant un vote majoritaire sur toutes les stratégies de sous-échantillonnage. Un vote majoritaire a été utilisé pour la décision finale de tous les classifieurs KPPV. L'avantage de cette approche est que les caractéristiques des masses sont extraites sans segmentation; mais, les résultats de cette méthode ne sont pas hautement garantis car la taille de la base des données utilisées est petite; Elle comprend seulement 62 masses malignes et 52 bénignes, et les images mammographiques sont très bien contrastées.

Dans [Muramatsu 2016], Muramatsu et al. Ont proposé un ensemble de caractéristiques extraites à partir des modèles ternaires radial locaux (MTRL). Pour distinguer entre lésions bénignes et malignes. MTRL prend en compte la direction des contours des modèles par rapport au centre des masses. Un autre travail récent sur la classification maligne / bénigne des anomalies est celui de Khan et al. [Khan 2016]. Dans ce travail, les auteurs ont proposé une technique d'optimisation des bancs de filtres de Gabor (BFGs) en se basant sur un algorithme de regroupement incrémental et d'optimisation d'essaim de particules (OEP). Les BFGs sont utilisés pour extraire des micro-modèles de texture locale à multi-échelle et orientation multiple à partir de mammographies. La classification est finalement réalisée en utilisant un classifieur de type séparateur à vaste marge (SVM), l'approche a été évaluée sur un ensemble de 768 masses. Les résultats de Muramatsu et al. [Muramatsu 2016] et Khan et al. [Khan 2016] sont bonnes, mais seules les masses de hautes densités sont prises en considération.

Dans [Dhanalekshmi 2015], deux méthodes ont été proposées par Menon et al. : L'analyse spectrale de puissance (PSA) et la méthode différentielle de comptage de boîtes (DBCM) ; Les deux méthodes ont été appliquées pour le calcul de la dimension fractale des contours de la masse pour différencier entre les lésions bénignes et malignes. Leur document ne se concentre que sur les masses bien définies ; L'étape de classification montre que la méthode PSA donne une meilleure précision que DBCM.

Deux algorithmes de segmentation hybrides ont été proposés par Rahimeh et al. Dans [Rouhi 2016] en se basant sur l'algorithme « ensemble de niveaux ». Après segmentation des régions

d'intérêt, quelques caractéristiques telles que la forme, l'intensité et la texture ont été calculées à partir des masses segmentées. Ensuite, un Algorithme génétique a été appliqué pour sélectionner les meilleures caractéristiques. Enfin, deux classifieurs de type PMC et forêt aléatoire ont été utilisés pour différencier entre les masses bénignes et malignes. Cette méthode a donné de bons résultats, mais seulement 57 des masses de haute densité sont utilisées dans la phase d'évaluation.

Dans [Sidney 2016], les caractéristiques de texture et de forme sont combinées et utilisées par deux classifieurs SVM et ELM pour détecter et classifier les masses mammographiques en se basant sur les régions d'intérêt (ROIs). Les caractéristiques sont extraites en décomposant chaque image à l'aide d'ondelettes multi-résolution. La méthode proposée a été évaluée sur 355 ROI de la base de données IRMA ; la méthode montre des résultats très élevés parce que seuls les tissus gras du sein sont pris en considération.

Elfarra et al. Dans [Elfarra 2012] proposent deux méthodes de sélection de caractéristiques : Algorithme génétique et Stepwise; Ces deux méthodes sélectionnent 14 à partir de 65 descripteurs extraites en utilisant deux méthodes d'analyse statistique textural qui sont la Méthode « *Run Difference Matrix* » (RDM) et la Méthode spatiale de dépendance des niveaux de gris (SGLDM). Cette approche n'est pas totalement automatique ; Quatre caractéristiques humaines sont utilisées dans cette étude, qui sont la densité de la mammographie, la catégorie de forme, le type de contours et l'âge du patient.

Jiang et al. Dans [Jiang 2008] ont proposé un algorithme unique pour segmenter et détecter les contours de la masse ; cette approche combine la méthode de l'entropie maximale avec un modèle de contour actif pour détecter les régions de masse spiculée. Un simple indice de spiculation a été appliqué pour quantifier les niveaux de spiculation de chaque masse. Bien que cet indice quantitatif ait été utilisé pour distinguer entre les masses spiculées et circonscrites.

Ying Wang et al. Dans [Ying 2014] ont proposé une approche pour la classification des masses mammographiques. En se basant sur les caractéristiques latentes des masses, dans la première étape, les régions qui contiennent des masses sont segmentées, ensuite l'allocation de Dirichlet latente (LDA) a été utilisée pour trouver la répartition de thématique cachée des régions segmentées (Régions d'intérêt avec des masses ROIms). Dans la deuxième étape, une extension pyramidale spatiale est introduite pour capturer les modèles de distribution spatiale latente sur les ROIms. Les tracés locaux des contours segmentés sont extraits pour construire un document spécial pour la LDA pour l'exploration des caractéristiques marginales statistiques latentes des masses ; Enfin, tous les thématiques latents ont été combinés, analysés et classifiés à l'aide d'un

classifieur SVM. Rangayyan et al. Dans [Rangayyan 2000] ont proposés des méthodes pour analyser et classifier les masses mammaires. Dans la première étape, les contours des masses ont été dessinés manuellement par un radiologue expert, ce qui permet par la suite d'identifier les principales portions concaves et convexes de la forme ; dans la deuxième étape, les caractéristiques sensibles aux variations locales des contours ont été calculées à l'aide d'un algorithme de modélisation polygonale. Les caractéristiques proposées sont utilisées dans l'étape de classification en utilisant deux schémas d'évaluation ; Le premier consiste à distinguer entre les masses malignes et bénignes, dans le deuxième schéma, les anomalies sont classées soit comme des masses spiculées ou bien circonscrites. Dans les approches [Jiang 2008], [Ying 2014], [Rangayyan 2000], seules les masses spiculées et circonscrites sont pris en considération.

Sampat et al. Dans [Sampat 2005], proposent un algorithme pour classifier les masses en catégories de formes utilisées dans le lexique BI-RADS (rondes, ovales, lobulées ou irrégulières). Pour cette tâche, un ensemble de caractéristiques ont été calculé à l'aide de la transformée Beamlet; un classifieur de type KPPV (k-plus proche voisins) a été utilisé pour différencier entre les types de forme. La méthode a été testée sur un ensemble de 100 images extraites à partir de la base de données DDSM. La précision de classification obtenue atteint 78% pour classifier les masses comme ovales ou rondes et de 72% pour classifier les masses comme lobulées ou rondes.

Un ensemble de caractéristiques représentant la forme géométrique et le type de contours ont été présentées par Vadivel et al. Dans [Vadivel 2013] pour classifier les masses mammographiques en catégories de forme utilisés dans le lexique BI-RADS: rondes, ovales, lobulaires et irrégulières. 17 caractéristiques ont été extraites pour faire la distinction entre les formes régulières et irrégulières. Dans [Surendiran 2010], une méthode de sélection de caractéristiques a été proposée par Surendiran et al. En utilisant l'analyse de la variance par étapes (ANOVA) et l'analyse discriminante (AD) Pour classifier les masses mammaires. Cette approche combine les 17 caractéristiques de contours et de forme pour distinguer entre les masses malignes et les bénignes.

Les résultats des méthodes présentées dans [Surendiran 2010], [Sampat 2005], [Vadivel 2013] étaient bons, bien qu'ils n'aient pas utilisé de caractéristiques texturales. Les résultats de ces études indiquent que la forme et les caractéristiques de contours sont très utiles pour la classification des masses mammographiques.

Grâce à l'étude des travaux connexes, les caractéristiques de texture ont été largement traitées contre les caractéristiques de forme et de contours, bien que les radiologues examinent les masses en fonction de leur densité, leur forme et le type de contours.

Parmi les systèmes DxAO mammographique existants, les techniques appropriées ne sont pas disponibles jusqu'à maintenant parce qu'elles ne répondent pas au besoin des radiologues; Un système approprié doit pratiquer le même lexique utilisé par les radiologues et doit classifier les masses mammographiques dans les classes BI-RADS. La contribution principale de notre étude est le développement d'un nouveau système DxAO capable de distinguer entre les quatre classes BI-RADS (2, 3, 4 et 5), où toutes les caractéristiques utilisées par le radiologue (contours, forme et densité) doivent être quantifiées ; une seule caractéristique humaine a été utilisée dans notre travail. On peut affirmer que ce schéma de classification est différent de tous les travaux antérieurs.

IV.4. Processus de dépistage automatique du cancer du sein

L'objectif des systèmes automatique de dépistage automatique du cancer du sein est de réduire les faux positifs (FP) et d'obtenir une sensibilité élevée pour détecter les cancers que les radiologistes pourraient manquer [Hutt 1996]. Au cours du dépistage du cancer du sein, l'identification incorrecte des lésions malignes comme bénignes chez les patients atteints de cancer du sein, généralement désignée sous le nom de faux positif (FP) [Kocur 1996] [Fogel 1998]. L'utilité clinique des systèmes automatiques dépend du nombre de FP détectés par image, car les radiologistes doivent prendre plus de temps et de soins pour réexaminer les zones mammographiques repérées par les systèmes informatisés [Nagel 1998].

Le traitement automatique des mammographies en utilisant les systèmes DxAO et DéAO en mammographie numérique implique généralement le traitement et l'analyse d'image et la reconnaissance de forme, afin de caractériser les anomalies malignes et bénignes dans les mammographies, la classification (apprentissage automatique) est nécessaire. La Figure IV.1 représente un schéma typique d'un système de dépistage (détection et classification) automatique du cancer du sein sur la base de la mammographie numérique. Les sections suivantes décrivent en détail chacune de ses étapes.

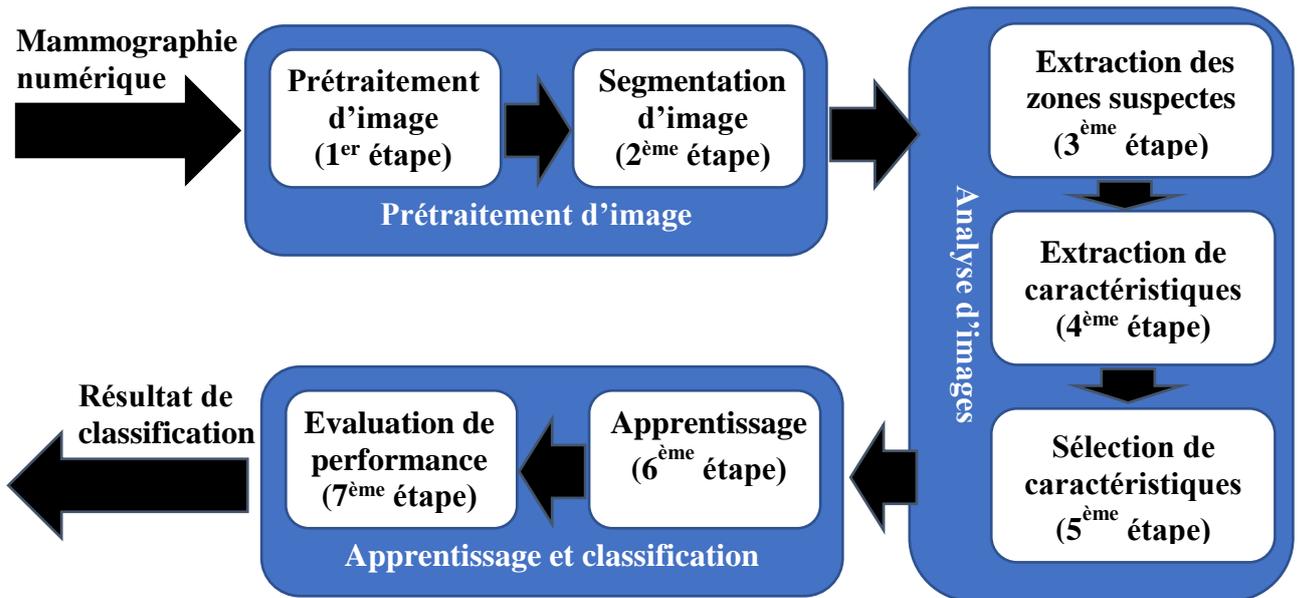


Figure IV.1: schéma général pour la détection et le diagnostic automatique du cancer du sein.

IV.4.1. Prétraitement d'images

Le prétraitement des mammographies numériques implique généralement le bruit et la suppression des artefacts radio-opaques et le réglage de l'image. Le réglage de l'image est habituellement réalisé en effectuant une amélioration du contraste en appliquant des mesures sur histogramme d'image. L'augmentation du contraste est très essentielle dans les mammographies, surtout pour les seins denses [Wang 1998]. Il peut y avoir un contraste entre les anomalies malignes et bénignes, mais il ne peut pas être discernable à l'œil humain. Il est donc difficile de différencier les anomalies malignes et bénignes [Wang 1998]. Les techniques traditionnelles de traitement d'image peuvent à ne pas fonctionner correctement sur les images mammographiques en raison de la grande variation des caractéristiques de densités, de tailles et de formes [Morrow 1992]. Il existe quelques approches pour améliorer les caractéristiques mammographiques, telles que les revues présentées par Thangavel et al. [Thangavel 2005a]. La première technique consiste à supprimer les bruits de fond et de numérisation et l'autre [Thangavel 2005b] utilisée pour augmenter le contraste des zones suspectes.

Les bruits dus aux caractéristiques intrinsèques d'un dispositif d'imagerie ou d'imagerie auront une incidence sur la sensibilité du résultat de la classification. Plusieurs types de filtres ont été rapportés par Qian et al. [Qian 1994] afin de réduire les bruits d'imagerie et de numérisation. Des méthodes telles que le fenêtrage en ligne droite [Chan 1987] et les fenêtres hexagonales [Glatt 1992] ont été introduites en tant que techniques de filtrage non linéaire. Bien que les techniques de filtrage non linéaire se soient avérées plus efficaces pour la suppression du bruit

que les approches linéaires, elles ne montrent pas nécessairement des améliorations significatives dans la préservation des détails de l'image. D'autres techniques incluent le filtrage médian et la préservation des contours. Le filtrage médian s'adapte localement aux images en niveaux de gris en utilisant un critère de seuil empiriquement dérivé [Lai 1989] et [Thangavel 2005b]. D'autres techniques incluent un lissage en préservant des contours, qui recherche un voisinage homogène dans différentes directions d'un pixel donné et la moyenne de ce voisinage. Les artefacts mammographiques sont de petits défauts de continuité de l'émulsion sur les clichés mammographiques, qui ressemblent à des MCC. Les artefacts des mammographies (Figure IV.2) contiennent généralement des étiquettes, de marqueurs et de coins dans la région fond (non-glande mammaire). De tels artefacts sont habituellement radio-opaques dans le sens où ils ne sont pas transparents aux radiations. Ces artefacts sont habituellement nettement définis et plus brillants que les microcalcifications, qui sont généralement présentes dans la région de fond des mammographies. L'un des problèmes liés à la segmentation avérée de la glande mammaire dans les mammographies est dû à l'existence d'artefacts (par exemple une étiquette chevauchant la région du sein), ce qui entraîne souvent une région de fond non uniforme provoquant l'échec de nombreux algorithmes de segmentation. Un algorithme robuste de suppression des artefacts basé sur la morphologie des zones développé par Wirth et al. [Wirth 2007] élimine les artefacts radio-opaques de la région de fond des mammographies.

Dans un travail précédent, nous avons proposé une technique qui permet d'extraire de l'image mammographique la région correspondante à la glande mammaire [Ferkous 2010]. L'intérêt de cette étape est de réduire la fenêtre de travail afin de permettre la recherche d'anomalies de ce limité à une petite zone de la mammographie, ainsi de réduire les temps des traitements lors de la recherche d'une éventuelle anomalie.

La glande mammaire est une région de grande taille présentant sur le cliché mammographique. Notre algorithme de segmentation de la glande mammaire est basé sur l'utilisation de l'algorithme « *Greedy* » des contours actifs et la carte de chanfrein [Ferkous 2010b].

La Figure IV.3 présente les résultats obtenus sur quelques mammographies. L'algorithme n'est sensible ni à la taille de la glande mammaire, ni à la présence ou non d'informations parasites sur le cliché (la présence d'une étiquette blanche par exemple ne modifie pas le comportement de l'algorithme).

Un des grands atouts de cet algorithme est qu'il est totalement automatique. Et que la forme du contour actif initiale est toujours la même (un demi-cercle).

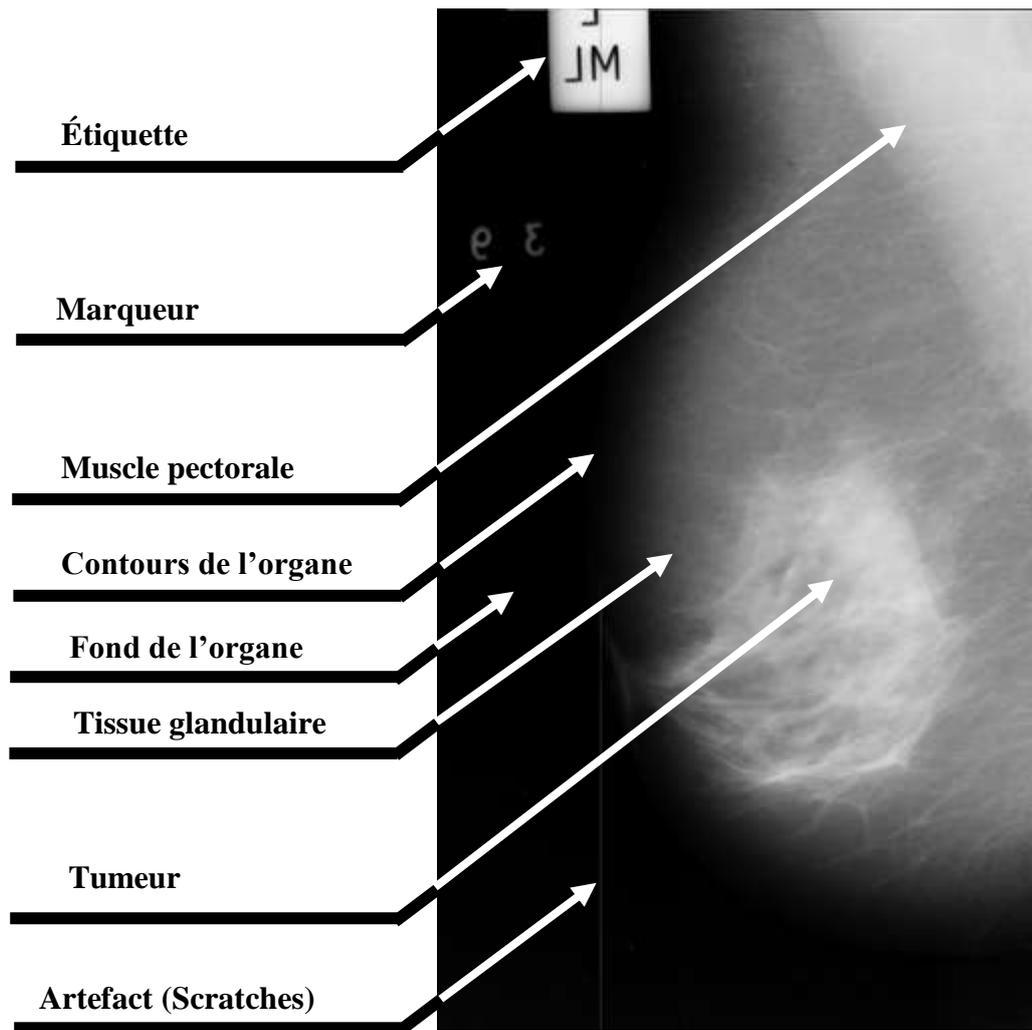


Figure IV.2: Les artefacts des mammographies.

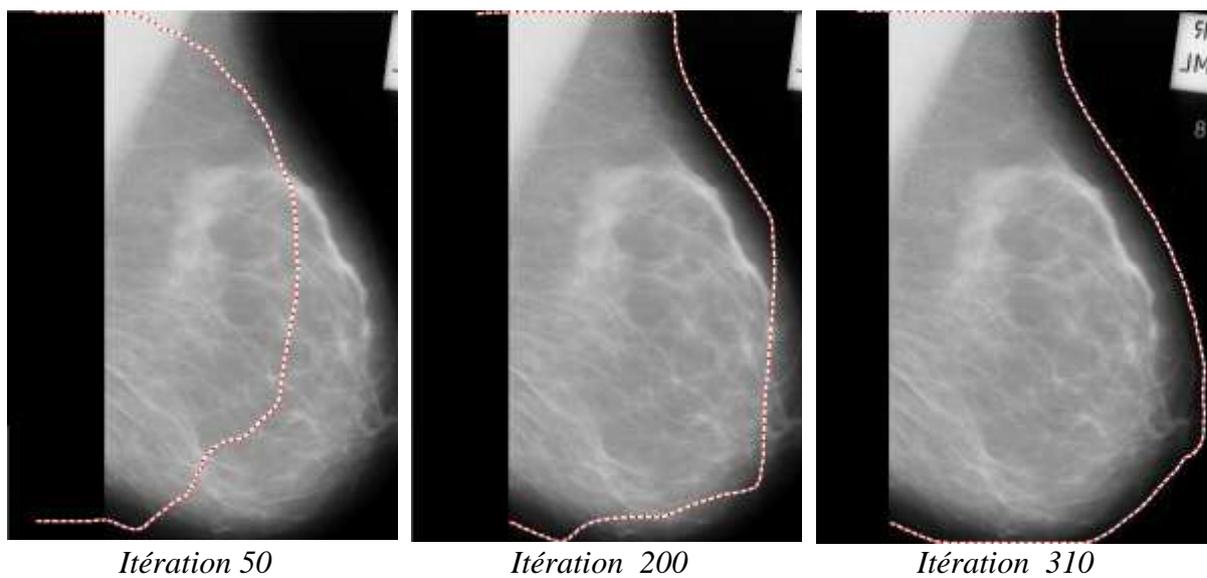


Figure IV.3: La détection de contours de la glande mammaire en utilisant les contours actifs [ferkous 2010b].

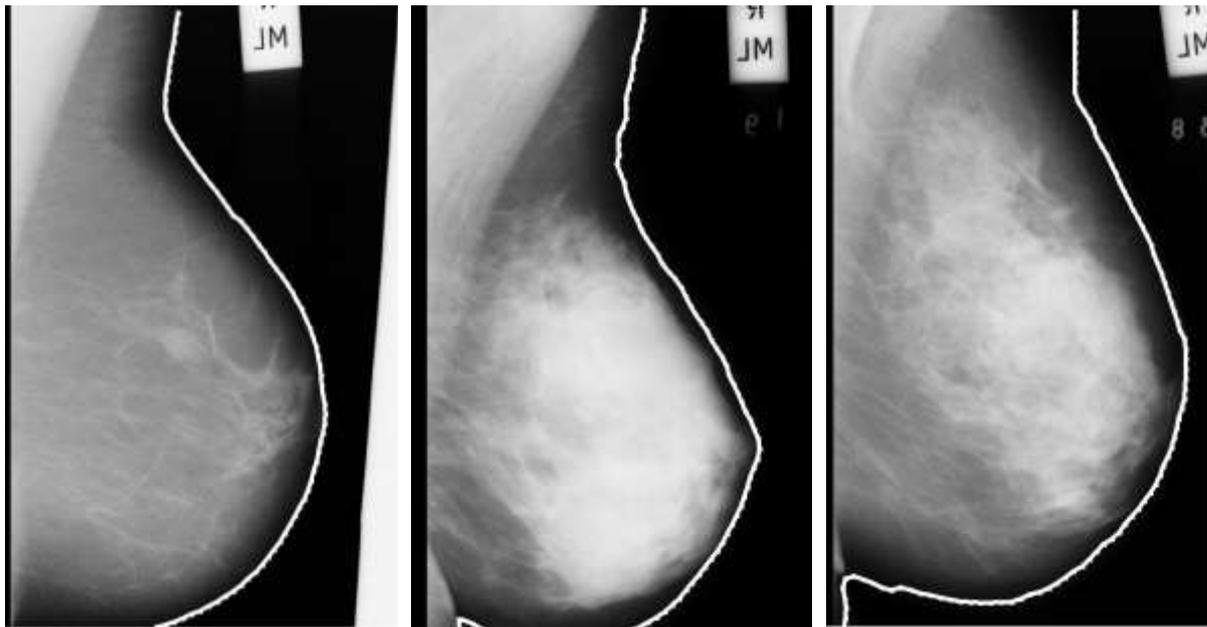


Figure IV.4: Exemples d'extraction de la glande mammaire des mammographies *mdb053*, *mdb172* et *mdb194* de la base de données MIAS [ferkous 2010b].

Cet algorithme a été testé sur la base de données MIAS [Suckling 1994] ayant 322 mammographies, le résultat obtenu est de 99,38% des cas acceptables. La glande mammaire n'est pas séparée que dans 2 mammographies de la base de données, puisque la taille de la glande mammaire est plus grande que l'image mammographique comme le montre la figure suivante :

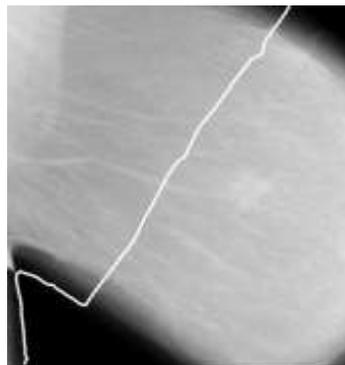


Figure IV.5: Résultat d'une mauvaise détection de la mammographie "*mdb152*" de la base de données MIAS [ferkous 2010].

La présence du muscle pectoral dans les images mammographique influe sur les résultats des méthodes de traitement d'image basées sur l'intensité et peut biaiser les procédures de détection d'anomalies malignes et bénignes [Ganesan 2013b], [Lakshmanan 2015], [Mustra

2015], [Yoon 2016], [Mustra 2013]. Ainsi, lors de l'analyse mammographique, le muscle pectoral doit être supprimé.

Le muscle pectoral représente une région ayant une intensité prédominante dans la plupart des incidences latérale-oblique des mammographies, nous avons remarqué que le muscle pectoral apparaît approximativement à la même densité que les tissus denses de la glande mammaire. Dans [Ferkous 2010] nous avons proposé une technique basée sur la morphologie pour la détection du muscle pectoral de la zone d'intérêt global. La Figure IV.6 présente les résultats de segmentation du muscle pectoral obtenus par notre méthode sur des images de la base de données mammographique MIAS.

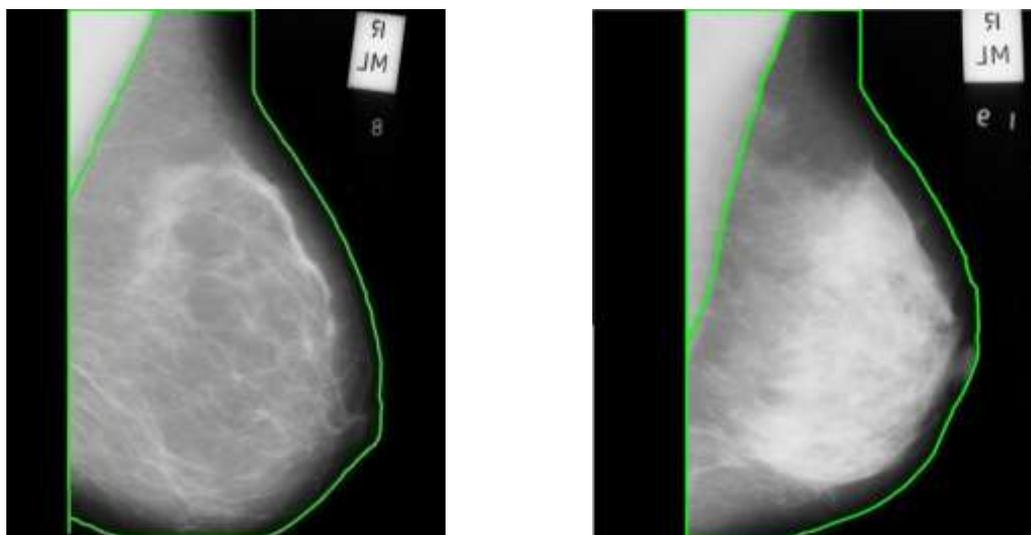


Figure IV.6: *Résultat de séparation du muscle pectoral de la mammographie mdb020 mdb004 de la Base de données MIAS[Ferkous 2010]*

IV.4.2. Segmentation

Après la suppression du bruit et les artefacts, la segmentation des images est utilisée pour identifier les zones suspectes (ROI) dans les images mammographiques. Le but de la segmentation est d'obtenir des ROIs contenant toutes les anomalies mammaires et de localiser les lésions suspectes et les microclassifications. Les anomalies mammographiques telles que les lésions de masse sont extrêmement difficiles à identifier à cause de leurs caractéristiques radiographiques et morphologiques qui ressemblent à celles des tissus mammaires normaux. Les masses mammographiques n'apparaissent pas comme des densités isolées mais comme superposées sur les tissus parenchymateux, puisque la mammographie numérique est une image de projection.

La segmentation est souvent considérée comme l'étape initiale dans un système de diagnostic assisté par ordinateur (DxAO) surtout si on fait abstraction de l'étape de prétraitement qui,

d'après la section précédente, n'est pas indispensable dans le cas de traitement des masses. La phase de segmentation est très importante, puisque les traitements ultérieurs (extraction de caractéristiques et classification) sont fortement liés au résultat de la segmentation. En effet, une bonne détection de contours des lésions mammographiques engendre une description exacte à ses caractéristiques. Ainsi, on peut garantir une classification en minimisant le taux des faux positifs et maximisant le taux des vrais négatifs.

Cette étape consiste à détecter la totalité du sein à partir du fond de l'image ou bien un type d'anomalie bien spécifique comme les micro calcifications et les masses. Il a été démontré que la détection des masses est plus difficile que la détection des micro calcifications. En effet, les masses peuvent être masquées partiellement par le tissu mammaire. Par ailleurs, il y a une grande variabilité dans ces lésions (forme, contour, dimension).

Au cours de la dernière décennie, un certain nombre de techniques de segmentation des mammographies ont été proposées. Ces techniques de segmentation comprennent : la croissance des régions [Mencattini 2008], [Rouhi 2015], Champ aléatoire de Markov [Goubalan 2016], Modélisation de fractale [Shanmugavadivu 2016], [Ashraf 2013], Transformée en ondelettes [Beura 2015], Amélioration du contraste pondéré par densité adaptative [Jenifer 2016], Logique floue [Sridhar 2015], Les opérations morphologiques [Abbas 2015] et les techniques de programmation dynamique [Rosado 2015].

Il existe de nombreuses techniques de segmentation qui ne cessent de se multiplier et de se diversifier. Ces méthodes sont couramment classées en trois catégories : les approches basées pixel, les approches basées contour et les approches basées région. Les approches basées sur les pixels sont généralement fondées sur l'étude des histogrammes de l'image par seuillage, clustering ou clustering flou. Les approches contours abordent la segmentation en tant que recherche de frontières entre les objets (les anomalies) et le fond. Elles consistent à identifier les transitions d'intensité de pixel entre les régions pour définir les bords des anomalies cherchées. Les approches région consistent à partitionner l'image en régions distinctes vérifiant une certaine homogénéité. Ces méthodes permettent d'aboutir directement à une partition de l'image étant donné que chaque pixel est affecté à une région unique. Nous détaillons dans ce qui suit les trois approches citées précédemment.

IV.4.2.1. Les méthodes de seuillage

Le seuillage a pour objectif de segmenter une image en deux ou plusieurs classes. Cette opération consiste à effectuer une partition de l'histogramme en niveaux de gris en utilisant un ou plusieurs seuils. Chaque pic de l'histogramme correspond à une classe. En effet, cette

méthode n'est efficace que si l'histogramme contient réellement des pics séparés. Les méthodes de seuillage ont été largement utilisées pour la segmentation de masses mammaires. Par exemple, les auteurs dans [Matsubara 1998] ont utilisé différentes valeurs de seuils en niveau de gris qui dépendent du type de tissu mammaire et ceci en se basant sur une analyse de l'histogramme. Plus récemment, Mudigonda et al. [Mudigonda 2001] ont utilisé un seuillage multi-niveau pour détecter des contours fermés. L'inconvénient majeur de cette approche est le fait de considérer que les masses ont une densité uniforme par rapport au fond de l'image ce qui n'est pas toujours vérifié.

Cependant, l'inconvénient majeur de ces différentes techniques de seuillage est le choix du seuil ou de l'intervalle de seuillage. En effet, avec un intervalle trop large, on obtient des faux positifs. Dans ce cas, l'image seuillée contient des pixels qui ne font pas partie des objets d'intérêt. Il s'agit généralement de bruit ou de pixels (zones) qui ont un niveau de gris proche de celui des objets recherchés. Avec un intervalle trop étroit, on obtient des faux négatifs. Certains objets d'intérêt ou bien des parties de ces objets n'apparaissent pas dans l'image seuillée

IV.4.2.2. Les méthodes basées régions

Les méthodes basées région regroupent des pixels ayant des propriétés communes, ainsi ces méthodes donnent directement une partition de l'image. Les méthodes basées région les plus populaires sont la méthode de croissance de régions (region growing) et la méthode de division et fusion de régions (split and merge).

La méthode de croissance de régions est basée sur la propagation d'un point germe initial (seed pixel) selon un critère d'homogénéité spécifique et sur la croissance itérative de la région. Cette méthode comporte en général deux étapes, une étape d'initialisation où l'on trouve les germes des régions les plus représentatives de l'image. Ces germes s'accroissent dans la deuxième phase en respectant des contraintes d'homogénéité, de forme géométrique ou de taille. Les algorithmes de croissance de région ont été largement utilisés pour la segmentation des masses. Toutefois, l'inconvénient majeur de cette approche est l'intervention humaine pour la segmentation de chaque objet.

Une alternative des méthodes de croissance de région est la segmentation par division et fusion de régions (split and merge). L'algorithme "division et fusion" a été proposé par Horowitz et Pavlidis en 1974 [Horowitz 1977]. Le processus est décomposé en deux étapes. Dans la première étape, appelée division, on analyse individuellement chaque région de l'image. Si

celle-ci ne vérifie pas le critère d'homogénéité, alors on divise cette région en blocs (le plus généralement en 4 quadrants) et l'on réitère le processus sur chaque sous-région.

Cette approche n'a pas été suffisamment exploitée dans le cadre de la segmentation des masses mammaires. La principale difficulté de cette approche réside dans le parcours de l'ensemble de tous les couples de régions voisines et la fixation du critère d'arrêt de la méthode.

IV.4.2.3. Les méthodes basées contours

Les approches les plus anciennes de détection de contour font intervenir la notion de dérivation afin de détecter les transitions entre les objets et le fond de l'image qui correspondent à des zones de fortes variations d'intensité. Les méthodes dérivatives utilisées se basent sur des opérateurs tels que Roberts, Sobel, Prewitt et Canny. Une fois la norme et la direction du gradient sont calculées en chaque pixel de l'image, ces méthodes extraient des contours d'un seul pixel d'épaisseur en sélectionnant les maxima locaux des normes des gradients. En mammographie, on est souvent confronté à la présence de bruit (d'acquisition), de textures fines ou de frontières pas très nettes, d'où les transitions détectées ne correspondent pas forcément aux contours réels. Il est alors souvent nécessaire d'appliquer un traitement en aval afin d'écarter les transitions dues aux bruits. De plus, les contours extraits sont généralement discontinus et peu précis. Il faut donc, utiliser des techniques de reconstruction de contours par interpolation ou connaître a priori la forme de l'objet recherché afin de connecter les points du contour.

Dans le cas des approches région, plusieurs travaux se basent sur le fait que les intensités sont statistiquement homogènes dans chaque région à segmenter. Or l'inhomogénéité de l'intensité se produit souvent dans les images médicales comme c'est le cas des images mammographiques. Le modèle du contour actif basé région récemment proposée par [Li 2008] est capable de segmenter des images ayant diverses intensités inhomogènes. Par ailleurs, il parvient à fournir un bon résultat de segmentation dans le cas d'objets à contours mal définis ou masqués (ce qui est souvent le cas des masses mammaires). En utilisant le terme de régularisation proposé par [Li 2008], la régularité de la fonction "level set" est intrinsèquement préservée ce qui garantit la précision du calcul et évite les procédures coûteuses de réinitialisation.

La Figure IV.7 présente le résultat de segmentation d'une masse circonscrite (première ligne) et d'une masse spiculée (deuxième ligne) en utilisant l'algorithme proposé par [Li 2008]. Les Figure IV.7.a) montrent la superposition du contour initial sur la région d'intérêt. La Figure

IV.7.b) montrent la propagation du contour et la Figure IV.7.c) présentent la convergence vers le contour de la lésion.

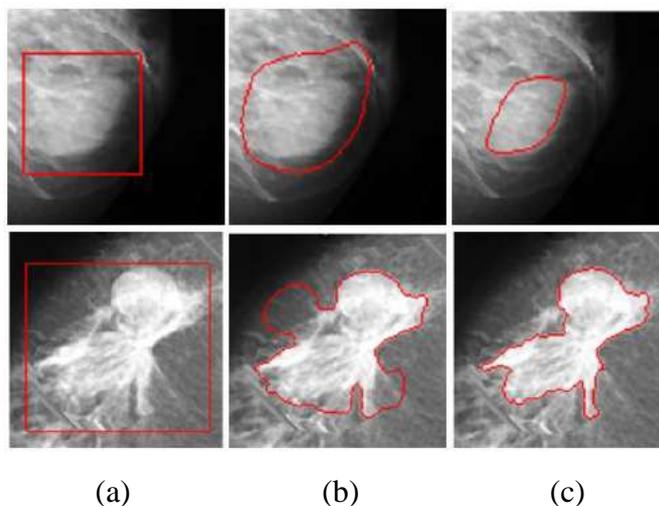


Figure IV.7: Le résultat de segmentation de deux masses circonscrite et spiculée [Li 2008].

IV.4.3. Extraction de caractéristiques

Lors de l'extraction des caractéristiques, les descripteurs sont quantifiés à partir des régions d'intérêts segmentés. Le nombre de caractéristiques sélectionnées pour la classification du cancer du sein rapporté dans la littérature varie selon l'approche employée. Les caractéristiques des différents domaines d'imagerie tels que : la morphologie [Czarnek 2016], [Lima 2015], l'analyse spatial et texturale peuvent être extraites de la mammographie numérique [Zyout 2015]. Les caractéristiques les plus importantes des ROI sont étudiées et analysées pour la classification des tumeurs malignes et bénignes (lésions de masse et calcifications) signalées par les radiologistes sont les suivantes [Veldkamp 2000]:

- i. Polymorphisme vs Monomorphisme: Les calcifications malignes ont tendance à se polymorphiser tandis que les amas bénignes sont principalement caractérisés par des calcifications monomorphes de taille uniforme [Lanyi 2000].
- ii. Taille et contraste : Les calcifications bénignes ont une taille plus grande et un contraste plus dense que les calcifications malignes.
- iii. Irrégularité vs rond et ovale: calcifications linéaires peuvent être une indication d'un carcinome canalaire in situ, car de telles calcifications sont situées dans les conduits glandulaires. Les calcifications bénignes sont la plupart du temps de formes rondes ou ovales et sont souvent situées dans les lobules.
- iv. L'orientation : Les calcifications malignes ont souvent de formes orientées vers le mamelon [Lanyi 2000].

- v. Le nombre : Un MCC avec très peu de calcifications est considéré comme moins suspect. Cinq calcifications ou plus, mesurant moins de 1mm dans un volume d'un centimètre cube, sont considérées comme formant un MCC [Popli 2001].
- vi. L'emplacement : Environ 48 pour cent de tous les cas cancéreux sont situés dans le quadrant supérieur externe de la glande mammaire. Les lésions observées dans ce quadrant sont plus suspectes [Harris 2014].

Au cours de la dernière décennie, un certain nombre de techniques d'extraction de caractéristiques ont été émergé et ont été appliquées à la mammographie numérique pour la détection et la classification des lésions de masse et de MCC. Ces techniques d'extraction de caractéristiques incluent : les caractéristiques de texture [Benmazou 2014], [Zhou 2013], [Torrents 2014], l'analyse de gradient de contours radial [Bhooshan 2014], L'optimisation des banques de filtres Gabor [Khan 2016], les caractéristiques morphologiques [Ferkous 2015] [Sridhar 2015], [Liu 2014], l'analyse par ondelettes [Beura 2015], les caractéristiques de contours des tumeurs [Bagheri 2012], [Dash 2012] et la modélisation neuro-floue [Abubacker 2016].

La littérature étudiée indique que les caractéristiques à base de texture ont été les plus couramment utilisées pour l'identification des lésions de masse et des MCC. Les descripteurs basés sur la texture proposés par Haralick et al. [Haralick 1973] ont été utilisés plus tard par Kramer & Aghdasi [Kramer 1999], Soltanian-Zadeh et al. [Soltanian 2004] et Khuzi et al. [Khuzi 2009] pour augmenter la performance des algorithmes d'apprentissage automatique comme les séparateurs à vaste marge (SVM) et les réseaux de neurones artificiels (RNA).

IV.4.4. La sélection des caractéristiques

Pour la modélisation d'un système d'apprentissage automatique pour la classification d'anomalies mammaires, il est souhaitable d'utiliser un nombre optimal de caractéristiques ; en effet, il devient plus difficile de définir des limites de décision précises dans un espace de représentation très large ; un très grand nombre de caractéristique augmente les besoins de calcul. Cela signifie qu'un sous-ensemble optimal de caractéristiques doit être sélectionné pour réaliser la phase d'apprentissage automatique.

La sélection des caractéristiques pertinentes des anomalies mammaires est une étape très importante pour la réalisation d'un système DxAO efficace. Le succès d'un système de classification dépend principalement des caractéristiques sélectionnées et de l'information fournie pour leur rôle dans le modèle. Certaines des caractéristiques extraites à partir des ROIs dans les images mammographiques ne sont pas significatives lorsqu'elles sont utilisées seules

dans un système DxAO, mais en combinaison avec d'autres caractéristiques elles peuvent être significatives pour la classification. En général, la raison de la sélection des caractéristiques est triple :

- a. Amélioration de la performance de classification du système ;
- b. Classification plus rapide et plus rentable ;
- c. Meilleure compréhension des processus qui génèrent le système DxAO; [Guyon 2003].

Au cours de la dernière décennie, un certain nombre de techniques de sélection de caractéristiques de mammographie ont été émergé et ont été appliquées à la mammographie numérique pour la classification de lésions de masse et des microcalcifications. Ces techniques de sélection des caractéristiques comprennent : Les algorithmes génétiques (AG) [Elfarra 2012], Analyse discriminante linéaire (ADL) [Li 2016], l'analyse de caractéristiques par étapes (Stepwise) [Gierach 2014], L'analyse en composantes principales (ACP) [Stewart 2014] et la modélisation Neuro-génétique [Verma 2001].

Il existe de nombreux avantages notables de la phase de sélection de caractéristiques. Tout d'abord, la sélection des caractéristiques facilite la visualisation et la compréhension des données et réduit les exigences de stockage en réduisant le temps d'apprentissage et en améliorant les performances de la classification. Le pouvoir de discrimination des caractéristiques peut être analysé à travers ce processus. L'objectif est d'éliminer une caractéristique si elle fournit peu ou pas d'information supplémentaire, par la suite la classification se fait en utilisant les caractéristiques restantes [Koller 1996]. De plus, des caractéristiques à haute corrélation peuvent être éliminées pendant ce processus afin de réduire le temps de traitement global sans affecter la précision du classifieur. Seulement quelques caractéristiques peuvent être utiles représentant le vecteur de caractéristiques optimales. Alors que la plupart des caractéristiques éliminées peuvent contenir des informations non pertinentes ou redondantes qui peuvent entraîner une dégradation des performances du classifieur. Les caractéristiques non pertinentes et corrélées sont préjudiciables, car ils peuvent contribuer au bruit et peuvent interagir de manière non-productive avec les algorithmes de classification utilisés [Hsu 2002].

Il existe plusieurs techniques de sélection de caractéristiques qui ont été bien étudiées et publiées. Toutes ces méthodes déterminent la pertinence des caractéristiques générées vers la tâche de classification. Il existe cinq types principaux pour l'évaluation des caractéristiques [Dash 1997] :

1. Distance (mesure de distance euclidienne)

2. Information (entropie, gain d'information, etc.)
3. Dépendance (coefficient de corrélation)
4. Cohérence (biais minimum des caractéristiques)
5. Taux d'erreur des classifieurs (basé sur l'algorithme de classification)

La limite de toutes les méthodes énumérées ci-dessus est qu'elles peuvent conduire à la sélection d'un sous-ensemble redondant certaines caractéristiques. Chen et Lin [Chen 2006] ont conclu que les caractéristiques indépendantes et identiquement distribuées ne sont pas vraiment redondantes. La réduction du bruit et l'utilisation d'un algorithme efficace de séparation des classes peuvent être obtenues en ajoutant des caractéristiques qui sont probablement redondantes. Chen et Lin [Chen 2006] ont indiqué qu'une caractéristique jugée inutile lorsqu'elle est testée seule, peut fournir une amélioration significative de la performance du classifieur quand elle est considérée avec d'autres caractéristiques. En d'autres termes, deux caractéristiques qui sont considérées inutiles par elles-mêmes peuvent être utiles ensemble. Ainsi, la sélection de sous-ensembles de caractéristiques peut fournir de bons résultats de prédiction, par opposition au classement des caractéristiques en fonction de leur pouvoir prédictif individuel.

IV.4.5. La classification

Pendant l'étape de classification, les objets des ROI (anomalies dans le tissu mammaire) sont classés soit bénignes (bénins) soit malignes (malins) sur la base d'un sous-ensemble optimale de caractéristiques de texture, de forme et de contours pour la modélisation d'un système d'apprentissage automatique. Le classifieur est formé sur des anomalies connues (lésions de masse et microcalcifications / MCCs) et combine les caractéristiques choisies et utilise des mesures de confiance pour indiquer qu'un ROI est soit maligne ou bénigne [Duda 2000].

Au cours de la phase d'apprentissage du système de classification, des techniques d'évaluation de performances d'apprentissage et de mémorisation d'un moteur de classification (classifieur) sont identifiées, parmi lesquelles: la validation croisée (VC) et le schéma leave-one-out.

Au cours de la dernière décennie, plusieurs techniques de classification automatique ont été étudiées et utilisées pour la catégorisation des lésions de masse et des MCC dans les mammographies. Ces techniques de classification comprennent: les séparateurs à vaste marge (SVM) [Singh 2014], les réseaux neuronaux artificiels (RNA) [Dheeba 2014], [Tan 2014], LDA [Khan 2016], les Réseaux de neurones convolutifs (CNN) [Arevalo 2016] et le k plus proche voisins [Li 2016]. D'autres techniques incluent la méthode statistique basée sur l'utilisation de

modèles statistiques et le framework de l'analyse bayésienne des images a été développé par Karssemeijer [Karssemeijer 1993].

Les auteurs Yoshida et al. [Yoshida 1996] ont utilisé la transformée en ondelettes squelettique et une technique d'apprentissage supervisée pour la classification des lésions de masse et des microcalcifications. Dheeba et al. [Dheeba 2014] ont proposé une méthode de distinction entre les tumeurs bénignes et malignes en utilisant les RNA. Les auteurs Baharuddin et al. [Baharuddin 2016] ont proposé une approche fondée sur la logique floue pour la différenciation entre anomalies mammaires. Les auteurs Hadjiiski et al. [Hadjiiski 1999] ont appliqué une combinaison hybride de la théorie de la résonance adaptative (TRA) et de l'ADL pour la classification entre anomalies malignes et bénignes.

IV.4.6. Évaluation de la performance

Afin d'évaluer les performances des systèmes de classification des systèmes DxAO (développés à l'aide de techniques d'apprentissage supervisées telles que les RNA et les SVM), la performance de la classification binaire doit être mesurée.

Matrice de confusion		Classé comme	
		Positive	Négative
Les données réelles	Positive	Vrai positive (VP)	Faux positive (FP)
	Négative	Faux négative (FN)	Vrai négative (VN)

Tableau IV-1 : Relation entre VP, VN, FP et FN de la matrice de confusion.

La performance d'un classifieur binaire ne peut pas être décrite par une seule valeur, elle est habituellement mesurée par sa précision (le taux de classification) pendant la phase de test, c'est-à-dire la fraction de points mal-classés sur l'ensemble des cas de la base de test. La performance d'un classifieur binaire peut être mieux décrite en termes de sa sensibilité et de sa spécificité, en mesurant sa performance sur des exemples faux positifs (FP) et faux négatifs (FN) [Veropoulos 2001].

La matrice de confusion du Tableau IV-1 représente la relation entre les différents indices de performance pour la classification binaire.

Pour la classification informatisée des anomalies malignes et bénignes dans les mammographies numériques, les quatre indices de performance (VP, VN, FP et FN) du Tableau IV-2 sont calculés en comparant la production prédite du système d'apprentissage automatique avec les étiquettes réelles déterminées après une biopsie. En utilisant ces quatre mesures de performance, nous pouvons calculer des mesures relatives à la classification binaire.

La mesure de performance	Définition
Vrai positif (VP)	Tumeur marquée comme maligne par une biopsie, qui est classée également comme maligne par la machine.
Faux positif (FP)	Tumeur marquée comme maligne par une biopsie, qui est classée comme bénigne par la machine.
Vrai négatif (VN)	Tumeur marquée comme bénigne par une biopsie, qui est également classée comme bénigne par la machine d'apprentissage.
Faux négatif (FN)	Tumeur marquée comme bénigne par une biopsie, qui est classée comme maligne par la machine d'apprentissage.

Tableau IV-2: Mesures de performance d'une classification binaire.

La *sensitivité* est définie comme le rapport des anomalies malignes qui sont marquées et classées comme maligne, à toutes les tumeurs marquées, elle est donnée par :

$$sensitivité = \frac{Nbr_de_positifs_bien_classés}{Nbr_Anomalies_Classés_Positifs} = \frac{VP}{VP + FN} \quad \text{Équation IV-1}$$

La *spécificité* est définie comme le rapport des anomalies bénignes et également non classées comme bénignes, à toutes les tumeurs non marquées, elle est donnée par :

$$spécificité = \frac{Nbr_de_négatifs_bien_classés}{Nbr_Anomalies_Classés_Négatifs} = \frac{VN}{VN + FP} \quad \text{Équation IV-2}$$

La précision globale est le rapport entre le nombre total d'anomalies correctement classées et la taille de la base de test, elle est donnée par :

$$précision = \frac{Nbr_de_cas_bien_classés}{Nbr_Total_des_Anomalies} = \frac{VP + VN}{VN + VP + FN + FP} \quad \text{Équation IV-3}$$

Dans une courbe ROC (Receiver Operator Characteristics), la *sensitivité* est la portion de masses mammographiques malignes correctement classées par le système d'apprentissage, cette courbe est tracée en fonction de $(1 - spécificité)$, c'est-à-dire la part des tumeurs bénignes qui sont faussement classées par le système d'aide au diagnostic ; pour les différentes valeurs de coupure. L'analyse du ROC est généralement utilisée pour déterminer la valeur d'un seuil

optimale utilisée comme critère de tests des systèmes de diagnostic médical. Il est possible de parvenir à un équilibre optimal entre la *sensitivité* et la *spécificité* qui est nécessaire pour certaines fins. Cela peut être obtenu en changeant la valeur de coupure du système. En outre, si le coût de la non-détection d'une anomalie particulière devient précieux pour la base de test, la valeur seuil peut être modifiée pour atteindre une *sensitivité* très élevée, mais une *spécificité* plus faible [Veropoulos 2001].

Pour visualiser les courbes ROC de la performance d'une classification binaire (Maligne / Bénigne), les paramètres de performance « Fraction des Vrai Positive (FVP) » et « Fraction des Faux Positive (FFP) » peuvent être calculés en utilisant les fonctions suivantes :

$$FVP = \frac{\text{Nbr_de_positifs_bien_classés}}{\text{Nbr_des_Anomalies_Positifs}} \quad \text{Équation IV-4}$$

$$FFP = \frac{\text{Nbr_de_négatifs_bien_classés}}{\text{Nbr_des_Anomalies_négatifs}} \quad \text{Équation IV-5}$$

Lors d'un test de diagnostic médical, la sensibilité (FVP) reflète le pourcentage d'anomalies malignes correctement classés et la spécificité (FFP) représente le pourcentage d'anomalies bénignes correctement classés comme bénigne. Ainsi, les courbes ROC sont des représentations bidimensionnelles le compromis relatif entre la sensibilité (FVP) et la spécificité (FFP) du test de diagnostic médical [Veropoulos 2001], comme le montre la Figure IV.8.

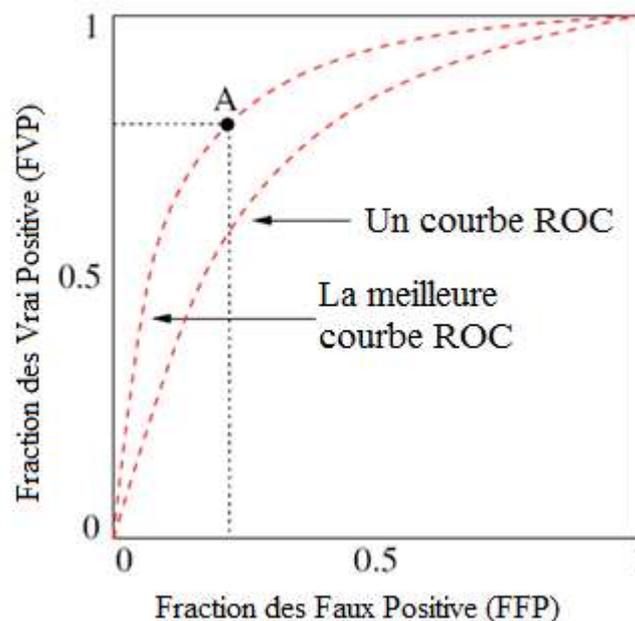


Figure IV.8: La Courbe ROC — FFP vs. FVP

La surface totale sous la courbe (SSC) d'une courbe ROC représentée par MN, qui est une mesure quantitative de la performance d'une classification binaire car elle reflète les performances d'essai du classifieur à tous les niveaux de seuils possibles. Plus l'SSC est grande dans l'intervalle fermé $[0.5, 1]$, meilleure sera la performance de classification [Veropoulos 2001].

Il y a généralement un nombre fini de points sur une courbe ROC dans la plupart d'expérimentation des systèmes de diagnostic médical, ce qui conduit principalement à ne pas trouver une bonne approximation de l'SSC. Ainsi, plus il y a de points, meilleure sera l'estimation de la courbe et la précision du classifieur binaire.

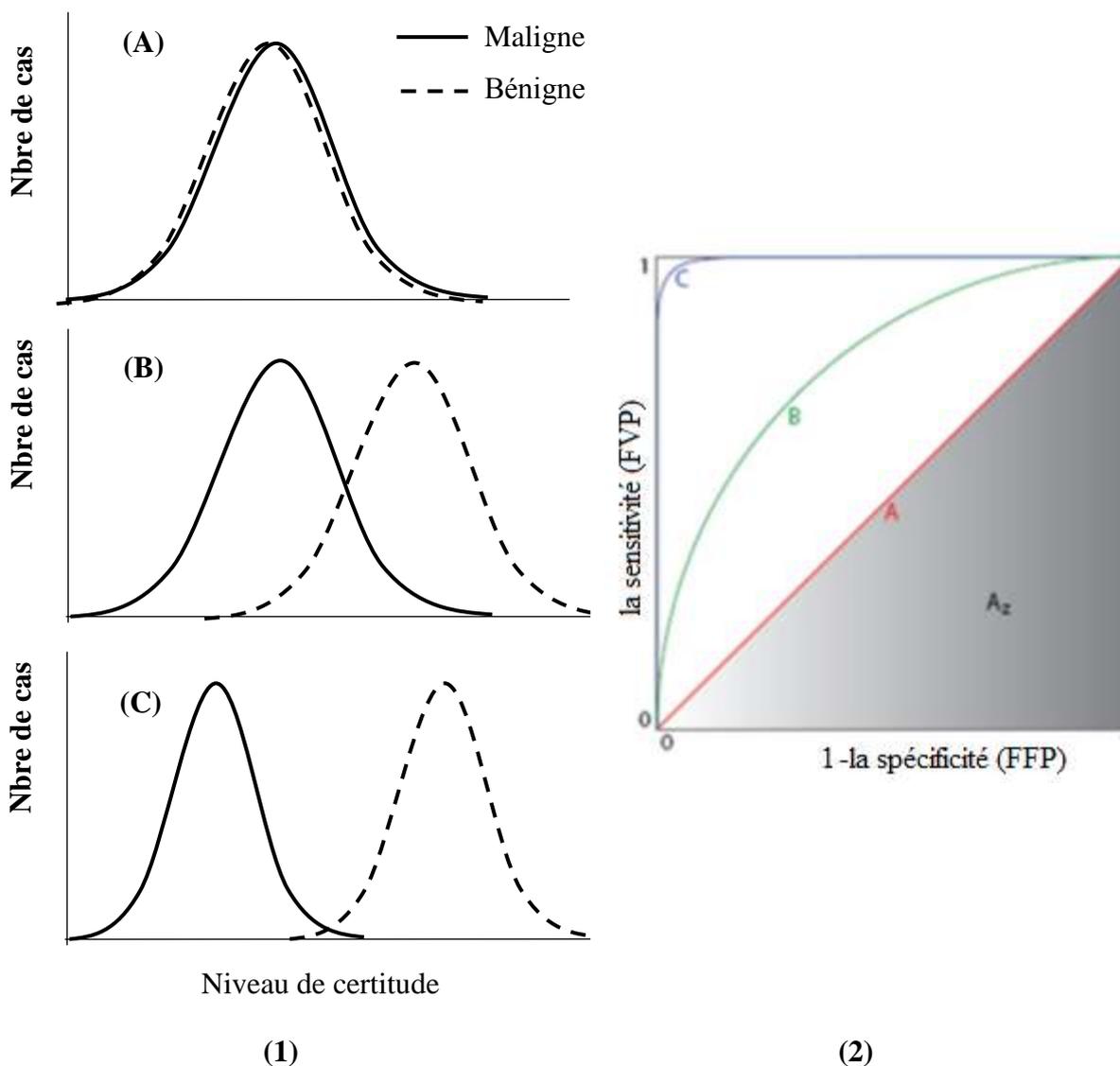


Figure IV.9: Les Courbes ROC

(1) Probabilité qu'une tumeur soit bénigne ou maligne

(2) Les courbes ROC de la Figure IV.9-1 [Veropoulos 2001]

Il existe plusieurs façons de calculer l'SSC d'une courbe ROC. La règle trapézoïdale peut être utilisée pour la calculer, mais elle donne généralement une sous-estimation de la zone. Une autre façon d'obtenir une meilleure approximation de l'SSC consiste à ajuster les données FVP et FFP (respectivement dans les équations (Équation IV-4 et Équation IV-5) dans un modèle binomial utilisant un logiciel d'ajustement de courbes [Veropoulos 2001]. Les distributions d'échantillons de tumeurs bénignes et malignes visualisées à l'aide de courbes ROC sont présentées à la Figure IV.9.

L'axe horizontal représente le niveau de certitude qu'une tumeur est maligne. Lorsque le système a des difficultés à identifier si une tumeur est maligne ou bénigne, les deux distributions se chevauchent, comme le montre la courbe A (voir la Figure IV.9 (a)). L'SSC de la courbe A est 0,5 (voir la Figure IV.9 (b)), est la pire performance que l'on puisse obtenir. La courbe C de la Figure IV.9 (a) présente le plus petit chevauchement entre les parties malignes et bénignes, ce qui fournit une performance quasiment parfaite avec un MN presque égal à 1,0 (voir la Figure IV.9 (b)).

Ces courbes ont été rapidement appliquées dans plusieurs domaines de recherche tels que : la prise de décision médicale, l'apprentissage automatique et la fouille de données.

IV.5. Conclusion

La détection et le diagnostic assisté par ordinateur sont des outils de calcul que les radiologues peuvent l'utiliser, qui vise à améliorer la détection et le diagnostic correct des anomalies mammographiques. Les systèmes DxAO sont des algorithmes computationnels qui caractérisent les lésions par analyse automatique des images. Les outils d'évaluation des DxAO sont basés sur leurs valeurs de faux positifs et négatifs et de vrais positifs et négatifs, et donc sur la sensibilité et la spécificité.

Dans ce chapitre nous avons présenté et discuté la documentation de base sur le traitement informatique des mammographies numériques. Dans la section IV.2, des systèmes de détection assistée par ordinateur ont été introduits avec la revue de la littérature des techniques informatiques de détection du cancer du sein présentées à la section IV.3. La section IV.4 a identifié les principales étapes utilisées par les chercheurs pour le diagnostic automatique du cancer du sein, en mettant l'accent sur les concepts fondamentaux du traitement numérique des images et les techniques de segmentation utilisées dans les applications de mammographie numérique.

Il y a encore une certaine controverse dans ce domaine. Cependant, certaines indications indiquent que cet outil, lorsqu'il est correctement utilisé, il améliore la détection correcte des micro-calcifications et des masses et par conséquent la présence d'une tumeur malignes.

Chapitre

5

LA CLASSIFICATION AUTOMATIQUE DES MASSES MAMMOGRAPHIQUES

V.1. Introduction

L'analyse automatique des images mammographiques est un domaine de recherche en plein essor, c'est un problème qui a été abordé largement dans la littérature puisque les premières recherches dans ce domaine ont plus de vingt ans et vient en parallèle avec le développement des techniques de la mammographies numériques.

Nous présentons dans ce chapitre l'application relative à la classification des masses mammographiques. Ce chapitre conclut et résume les contributions de notre travail, les réalisations et les objectifs de la recherche en ce qui concerne les résultats expérimentaux obtenus sont mis en évidence ainsi que les principales conclusions et l'importance de la recherche. Ce chapitre traite également l'impact et l'importance du système développé pour les radiologues et les hôpitaux pour le dépistage et l'interprétation des mammographies. Les radiologues et les cliniciens bénéficieront du système développé car il les aidera dans leur diagnostic en agissant comme un deuxième lecteur. Cette application fait appel aux notions que nous avons présentées dans cette thèse ainsi qu'à d'autres notions plus classiques en prétraitement d'image. Nous indiquons dans ce chapitre la démarche générale adoptée pour le processus d'extraction de caractéristiques des masses mammaires. Ensuite, nous montrons et discutons les résultats obtenus lors de la procédure de classification.

Ce chapitre est organisé comme suit : dans la section 2, nous présentons les méthodes d'extraction de caractéristiques étudiées et les schémas de classification des masses mammographiques utilisés dans le présent travail ; À la section 3, nous exposons l'organisation de la base mammographique utilisée ; Dans la section 4, nous montrons les résultats expérimentaux et quelques discussions. Enfin, dans la section 5, nous terminons par une conclusion.

V.2. Processus de classification automatique des masses

Lors d'un examen mammographique, le radiologue ne fait que résoudre un problème de vision. C'est à partir de ce point et de récentes techniques d'analyse et de prétraitement d'image, que l'idée de réaliser un système DxAO pour faciliter le travail des radiologues pour mieux diagnostiquer un cliché mammographique. La recherche menée dans ce domaine actuellement est encore à un stade peu avancé. Pour l'instant, seul le problème de classification des micro-calcifications a prouvé sa faisabilité. Les résultats plus qu'encourageants obtenus dans ce domaine ont permis d'espérer un aboutissement aux autres problèmes plus complexes comme celui des masses. Comme nous l'avons dit précédemment, notre but est de proposer un système

de diagnostic assisté par ordinateur basé sur la classification BI-RADS, qui est le processus de classification pratiqué par les radiologues.

La structure du système de classification des masses mammographique que nous proposons s'inspire globalement de l'approche du médecin lors de l'examen radiologique. Notre approche [Ferkous 2017] comprend principalement deux étapes : la première étape consiste à quantifier les caractéristiques employées par les radiologues selon BI-RADS, dans la deuxième étape nous avons classifié les masses mammaires en utilisant deux scénarios, le premier consiste à catégoriser les masses pour la distinction entre les quatre classes BI-RADS (Bi-RADS 2, Bi-RADS 3, Bi-RADS 4 et Bi-RADS 5) ; Dans le deuxième scénario, nous classons les anomalies sur les deux classes (bénignes et malignes).

La difficulté de notre problème réside dans la complexité et la diversité des masses et les différentes caractéristiques perceptibles telle que la graisse, les fibres et les lésions qui rendent les images particulièrement complexes.

Le diagnostic des masses sur la base des mammographies n'est pas une opération mathématique qui permet d'aboutir à un résultat absolument certain. La preuve est que dans certain cas le radiologue a recours à d'autres techniques plus précises comme l'échographie et l'IRM (imagerie par résonance magnétique) pour valider son diagnostic. Ainsi, les mammographies sont utilisées pour détecter d'éventuels signes pathologiques et pour décider de la nécessité d'examens complémentaires plus approfondis, d'un suivi à court terme ou une biopsie.

Le but de notre travail est de réaliser un système d'aide au diagnostic mammographique pour la classification des masses mammographiques. Ce système peut éventuellement donner de fausses alertes aux radiologues mais ne doit pas omettre de signaler la pathologie. Et pour que le système de classification soit pertinent il faut que le nombre de fausses alertes ne doivent pas être trop important.

Dans ce qui suit nous décrivons les méthodes adoptées pour résoudre les étapes principales de notre application, en insistant sur celles qui utilisent les notions présentées dans cette thèse, c'est-à-dire principalement les différentes méthodes d'extraction de caractéristiques, vue dans le troisième chapitre, et en exploitant les connaissances du problème à traité présentées aux chapitres (1, 2 et 3).

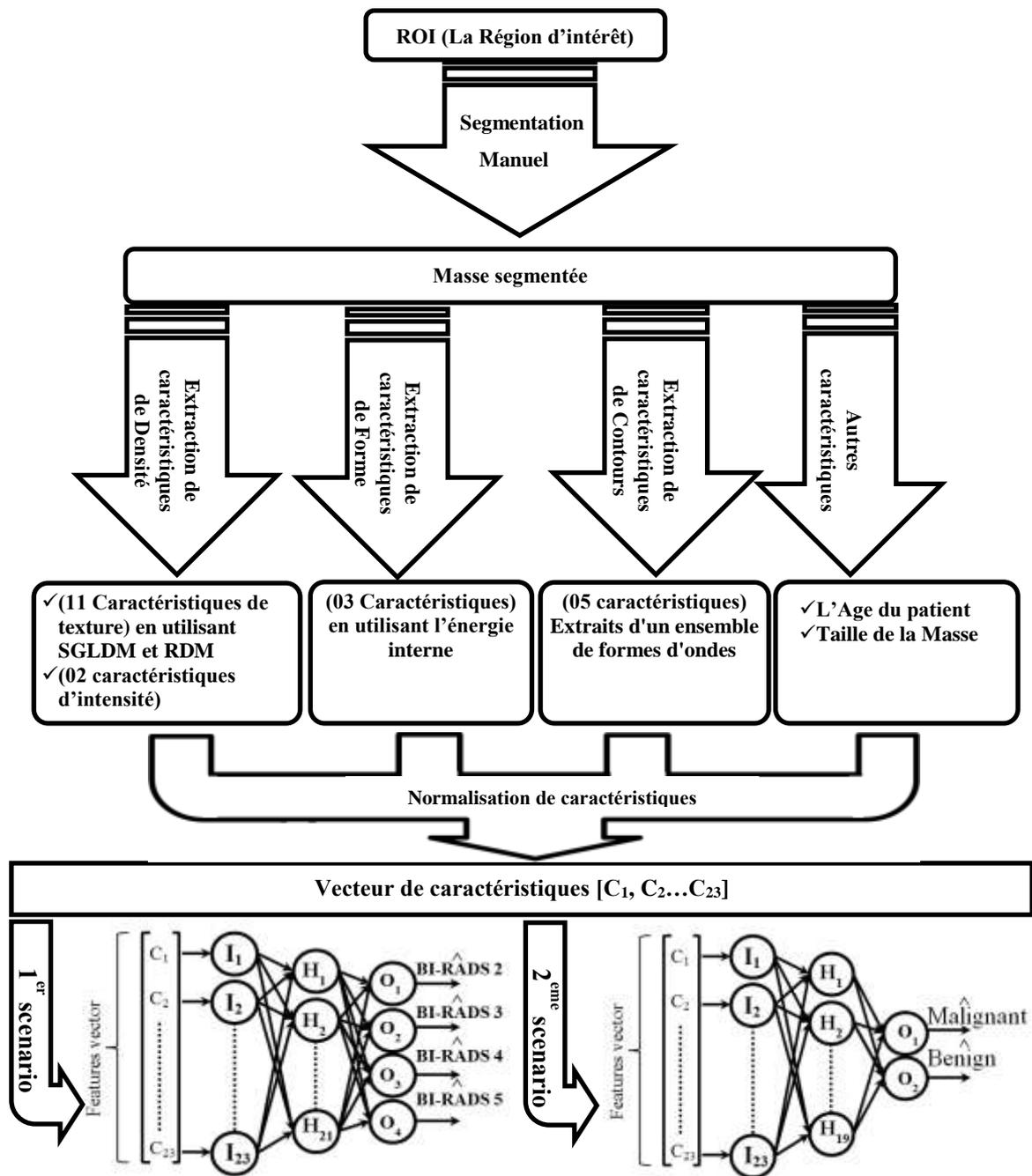


Figure V.1: Les étapes de la méthodologie proposée.

V.2.1. La sélection de la région d'intérêt (ROI)

Les régions d'intérêt (ROI) sont sélectionnées à partir de la base de données numériques pour la mammographie de dépistage (DDSM) par un radiologue expérimenté pour toucher toutes les catégories de masses. Les localisations des anomalies dans la mammographie sont disponibles sous forme de chaînes codées dans le fichier « *. Overlay » (section V.3.3).

V.2.2. La Segmentation des masses mammaires

Dans l'étape de segmentation, les zones suspectées sont séparées du fond. Cette phase est extrêmement importante, car le succès de l'algorithme de classification dépend de cette étape. En raison de la variabilité des degrés de densité et des types de contours utilisés dans notre recherche (Figure V.12, Figure V.13), où la densité peut présenter un faible contraste et les tissus qui entourent et masquent la région suspecte rendent l'étape de segmentation très difficile, ce qui nécessite donc une autre étude entière.

Ici les phases d'extraction de caractéristiques et de classification nous intéressent le plus. Dans notre système, l'étape de segmentation a été faite manuellement par un radiologue expérimenté, parce que nous supposons que la détection de la masse a été déjà faite.

V.2.3. L'extraction de caractéristiques des masses mammographiques

Selon BI-RADS [Lévy 2005], une masse se caractérise par les descripteurs : de sa forme, de ses contours et de sa densité ; De plus, l'âge du patient et la taille de la masse peuvent reclasser une anomalie dans une classe BI-RADS inférieure ou supérieure. La Figure V.1 résume les catégories des caractéristiques utilisées par les radiologues pour la classification des masses mammaires et les méthodes de quantification utilisées dans notre approche.

Le but de cette section est de quantifier les caractéristiques mentionnées précédemment comme suit :

V.2.3.1. Extraction de caractéristiques de forme

Avant d'entamer la phase d'extraction des caractéristiques de forme, une étape de mise à l'échelle est nécessaire, cette phase joue un rôle important dans le processus de calcul des caractéristiques de forme ; en respectant une échelle de référence permet aux masses identiques d'avoir les mêmes "caractéristiques de forme" quelles que soient leurs dimensions.

Lorsque la masse segmentée est mise à l'échelle, l'analyse de la forme peut être effectuée. Cependant, plusieurs caractéristiques de forme ont été proposées dans la littérature. Dans notre travail nous avons utilisé trois caractéristiques : la courbure, la continuité et l'irrégularité extraite de l'énergie interne E_{int} . Cette dernière a été proposée par Kass et al. Dans [Kass 1988] pour l'utiliser dans leur méthode de détection de contours appelée *Snakes*. Dans leur méthode, l'énergie interne gère la cohérence de la courbe; Il maintient la cohésion des points et la rigidité de la courbe.

Mais dans notre travail [Ferkous 2017], nous l'avons utilisée pour décrire la forme de la masse mammographique.

Nous devons calculer l'énergie interne à partir de laquelle les trois caractéristiques de formes seront extraites.

Tout d'abord, un ensemble de points v_i doit être extrait du bord de la masse mammographique pour chaque changement de direction du chemin de contours, la Figure V.2 montre un exemple du processus de sélection.

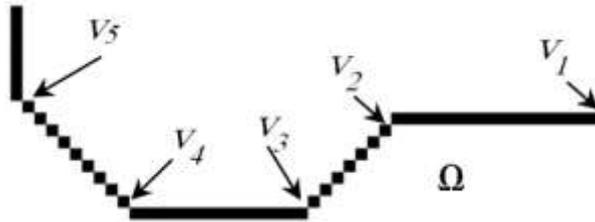


Figure V.2: Exemple de sélection des points v_i .

Les contours de la masse Ω peuvent être représentés par un vecteur constitué de points extraits $[v_1, v_2 \dots v_n]$, maintenant nous pouvons mesurer l'énergie en utilisant la formule suivante:

$$E_{\text{int}}(v) = (\|v'(\Omega)\|^2 + \|v''(\Omega)\|^2) \quad \text{Équation V-1}$$

Les termes v' et v'' sont les dérivées première et seconde de v , le terme du 1er ordre correspond à la tension (comportement comme une membrane mince). Il prend une valeur importante lorsque la courbe se distend. Nous parlerons donc d'énergie de continuité.

Le terme du 2ème ordre correspond à la courbure (comportement comme une section mince). Il prend une valeur importante lorsque la courbe prend de fortes convexités, d'autre part quand elle prend une petite valeur, la courbe tend vers un cercle s'il est fermé ou une ligne si elle est ouverte. Nous parlerons donc d'énergie de courbure.

Pour chaque point v_i , nous discrétisons les termes v' et v'' comme suit:

$$\text{Continuité}(i) = \left\| \frac{dv_i}{d\Omega} \right\|^2 \cong |v_i - v_{i-1}|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 \quad \text{Équation V-2}$$

$$\text{Courbure}(i) = \left\| \frac{d^2v_i}{d\Omega^2} \right\|^2 \cong |v_{i-1} - 2v_i + v_{i+1}|^2 = (x_{i-1} - 2x_i + x_{i+1})^2 + (y_{i-1} - 2y_i + y_{i+1})^2 \quad \text{Équation V-3}$$

Enfin, l'irrégularité n'est rien d'autre que le nombre de points de (v_i). Le Tableau V-1 résume les caractéristiques de forme utilisées dans notre approche.

Les caractéristiques de forme	Valeur	
La Continuité	$c_1 = \sum_{i=1}^n \text{Continuité}(i)$	Équation V-4
La Courbure	$c_2 = \sum_{i=1}^n \text{Courbure}(i)$	Équation V-5
L'irrégularité	$c_3 = \text{Nombre_de_point } s(v_i)$	Équation V-6

Tableau V-1: Les Caractéristiques de Forme.

La première contribution consiste à utiliser les énergies du *Snake* pour décrire la forme de la masse ; une masse circulaire ou ovale seront représentées par des énergies minimales, d'une autre part la masse irrégulière se caractérise par une forme disposant des énergies maximales. Une masse micro-lobulée ou spiculée auront plusieurs points v_i (irrégularité maximale) par rapport aux formes possédant des contours plat.

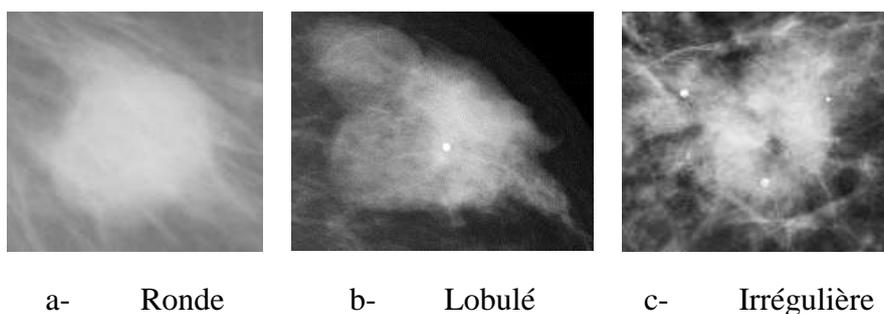


Figure V.3: Les formes de masses (A) Ronde (b) Lobulé (c) irrégulière

Le tableau suivant présente les résultats de continuité, de courbure et de l'irrégularité obtenus sur les formes de la Figure V.3.

	La forme		
	Ronde	Lobulé	Irrégulière
La Courbure	153	1543	2140
La Continuité	3034	3362	3532
L'irrégularité	121	267	284

Tableau V-2: Résultats des énergies de continuité, de courbure et de l'irrégularité des masses de la Figure V.3.

Il est préférable d'utiliser des propriétés invariantes aux transformations géométriques telles que le changement d'échelle. Ceci permet de palier aux différentes transformations que peut subir une image. Cependant, la caractérisation robuste et discriminante des images reste un grand défi en traitement d'images.

Dans ce projet, nous proposons des descripteurs de forme extraites à partir de l'énergie interne du *snake* adaptés au caractère pathologique des masses mammaires. Pour cela, ces descripteurs

doivent respecter les critères proposés par Cheikhrouhou et al. [Cheikhrouhou 2007] :

- ✓ **Invariance** : Une forme peut apparaître de différentes manières et à des échelles différentes, le système visuel humain reconnaîtra presque toujours la même forme. De la même manière, le descripteur de forme doit être invariant par transformations de rotation géométrique.
- ✓ **Unicité** : L’algorithme correspondant au descripteur de forme produit pour chaque forme une description unique ; c’est-à-dire deux formes peuvent produire une même valeur du descripteur, à condition qu’il n’y ait que peu de différences visibles entre ces deux formes.
- ✓ **Conservation de l’information** : Il doit être possible de reconstruire la forme d’origine à partir de sa description. Les descripteurs vérifiant ce critère fournissent une représentation complète de la forme.

Nous présentons dans le Tableau V-3 les avantages et les inconvénients de chaque descripteur géométrique définis dans la section (III.2.2) et la méthode proposée.

<i>Descripteurs</i>	<i>Les avantages</i>	<i>Les inconvénients</i>
<i>A et P</i>	<ul style="list-style-type: none"> ✓ Plus performants quand ils sont associés à d’autres descripteurs 	<ul style="list-style-type: none"> ✓ Insuffisants pour décrire les masses complexes ✓ Non uniques
<i>Com</i>	<ul style="list-style-type: none"> ✓ Proportionnel à la complexité de la forme ✓ Invariant 	<ul style="list-style-type: none"> ✓ Non unique
<i>C</i>	<ul style="list-style-type: none"> ✓ Simple ✓ Distingue les masses circulaires ✓ Invariant 	<ul style="list-style-type: none"> ✓ Insuffisant pour décrire des masses complexes
<i>Rect</i>	<ul style="list-style-type: none"> ✓ Distingue les masses ovales 	<ul style="list-style-type: none"> ✓ Non invariant par rotation ✓ Insuffisant pour décrire des masses complexes
<i>LRN et LRNM</i>	<ul style="list-style-type: none"> ✓ Invariants ✓ Proportionnels à la complexité de la forme 	<ul style="list-style-type: none"> ✓ Dépend de la distance entre le contour et le centre de gravité ✓ Non unique
<i>Curv</i>	<ul style="list-style-type: none"> ✓ Proportionnel à la complexité de la forme 	<ul style="list-style-type: none"> ✓ Non Invariant
Coque convexe	<ul style="list-style-type: none"> ✓ Invariant par rotation ✓ Proportionnel à la complexité de la forme 	<ul style="list-style-type: none"> ✓ Insuffisant pour décrire des masses complexes
L’énergie interne du <i>Snake</i>	<ul style="list-style-type: none"> ✓ Invariant par rotation ✓ Proportionnel à la complexité de la forme ✓ Presque unique ✓ Plus performants quand il est associé à d’autres descripteurs 	<ul style="list-style-type: none"> ✓ Insuffisant pour différencier entre les masses spiculées et micro-lobulées

Tableau V-3: *Les avantages et les inconvénients des descripteurs géométriques.*

V.2.3.2. L'extraction de caractéristiques de contours

L'analyse de contours est une tâche très complexe en raison de la nature difficile des mammographies en termes de netteté et de niveau de contraste. Cependant, le nombre de caractéristiques basé contours existants est beaucoup moins élevé par rapport aux caractéristiques de forme ou de texture. Dans ce travail, les caractéristiques de contours sont extraites à partir d'un ensemble de formes d'onde par l'analyse d'ondelettes utilisée dans [Bagheri 2012], qui peut capter la brusquerité à l'intérieur et à l'extérieur de la masse. La Figure V.4 illustre une présentation de forme d'onde.

Pour les obtenir, les coordonnées du centre de la masse $C(x_c, y_c)$ est d'abord calculée en utilisant les fonctions suivantes:

$$x_c = \sum_{i \in \Omega} i / \text{longueur}(\Omega) \quad \text{et} \quad y_c = \sum_{j \in \Omega} j / \text{longueur}(\Omega) \quad \text{Équation V-7}$$

Tel que Ω symbolise les contours de la masse.

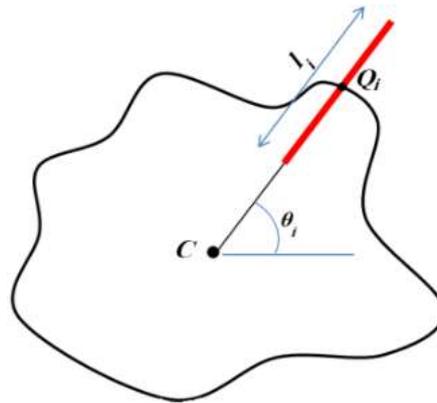


Figure V.4: La présentation d'une forme d'onde i ayant un centre Q_i et longueur l_i .

Si Q est un pixel sélectionné en traversant les contours de la masse, alors la forme d'onde aura la même direction que la ligne CQ . Dans [Bagheri 2012], la longueur de la forme d'onde est calculée relativement au rayon CQ et les centres de la forme d'onde Q_i sont placés tous les 5 pixels au long des contours de la masse Ω .

Dans notre travail, la longueur des formes d'onde l_i a été fixée après une série d'expérimentation.

La deuxième contribution consiste à placer les formes d'ondes tous les $\theta = \pi / 16$ ce qui signifie que le nombre de formes d'ondes NF est fixé à 32 ; l'objectif de fixer le nombre de formes d'onde pour chaque Masse mammographique est de normaliser les caractéristiques de contours, ce qui permet aux masses quelle que soit leurs tailles, d'avoir le même nombre de formes d'ondes. Ainsi, après une série d'expérimentation une longueur des formes d'onde $l = 128$ a été

considérée suffisante pour une meilleure extraction de caractéristiques, cette longueur garantit le non-chevauchement avec le muscle pectoral, les contours de la glande mammaire et les autres masses ; de ce fait, la valeur l est une puissance de deux pour être pertinente pour la transformation d'ondelettes

Un vecteur de probabilité de contours (VPC) est calculé pour chacune de ces formes d'onde. Une description détaillée du calcul de la probabilité de contours se trouve dans la section III.2.3. Après cela, à partir de ces distributions de probabilité (VPC_i) de chaque forme d'onde i , des caractéristiques telles que l'aplatissement ($kurt_i$), L'entropie ($entr_i$) et l'indice de la probabilité maximale (R_i) sont calculés.

$$Kurt_i = \frac{(1/D) \sum_{j=1}^D (VPC_i^j - \overline{VPC_i})^4}{((1/D) \sum_{j=1}^D (VPC_i^j - \overline{VPC_i})^2)^2} - 3 \quad \text{Équation V-8}$$

$$Entr_i = - \sum_{j=1}^D (VPC_i^j \times \log_2(VPC_i^j)) \quad \text{Équation V-9}$$

$$R_i = (IndiceMax VPC_i^j) - \frac{D}{2} \quad \text{Équation V-10}$$

Où D est la longueur du vecteur de probabilité de contours.

La moyenne et l'écart-type de l'aplatissement ($kurt_i$) et l'entropie ($entr_i$), et la variation totale des probabilités maximales représentent les caractéristiques de contours utilisées dans notre système (Voir Tableau V-4).

Les caractéristiques de contours	Valeur	
La moyenne de la Kurtosis	$c_4 = (\sum_{i=1}^{NF} kurt_i) / n$	Équation V-11
L'écart-type de la Kurtosis	$c_5 = \sqrt{(\sum_{i=1}^{NF} (kurt_i - c_4)^2) / n}$	Équation V-12
La moyenne de l'entropie	$c_6 = (\sum_{i=1}^{NF} Entr_i) / n$	Équation V-13
L'écart-type de l'entropie	$c_7 = \sqrt{(\sum_{i=1}^{NF} (Entr_i - c_6)^2) / n}$	Équation V-14
La variation totale	$C_8 = \sum_{i=1}^{NF} R_i $	Équation V-15

Tableau V-4: Les caractéristiques de contours utilisées dans notre système.

V.2.3.3. L'extraction de caractéristiques de la densité des masses

Dans le domaine radiologique, le terme densité est utilisé pour décrire le tissu de la masse mammographique [Lévy 2005]. Les anomalies ayant une haute densité sont trois fois plus susceptibles d'être malignes par rapport à celles ayant une faible densité [Woods 2011] ; une masse graisseuse pouvait être : un kyste huileux, lipome ou galactocèle; cependant que les masses de densité mixte pouvait être ganglion lymphatique intra-mammaire.

La propagation de micro-calcifications sur le tissu de la masse (Figure V.5) recalifié une anomalie dans une classe BI-RADS plus supérieure [Lévy 2005]. C'est pourquoi nous devons étudier la texture et l'intensité des masses mammaires dans cette section.

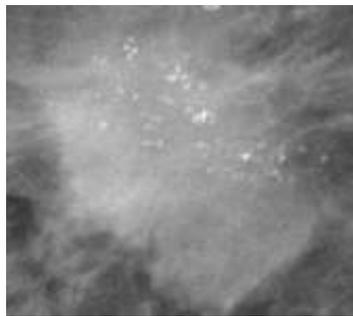


Figure V.5: Une Masse avec micro-calcifications classée BI-RADS 5 [Robin 2016].

V.2.3.3.1. L'extraction de caractéristiques texturale

Parmi les méthodes qui se sont révélées efficaces pour l'extraction des caractéristiques textural des masses mammographiques, Nous distinguons la méthode Run difference matrix (RDM) [Chen 2005] et la méthode spatiale de dépendance des niveaux des gris (SGLDM) [Benmazou 2014] [Maitra 2011]. Nous présentons dans cette section les deux méthodes avant de présenter les caractéristiques de texture utilisées dans notre approche.

a. La méthode spatiale de dépendance des niveaux des gris (SGLDM) :

La matrice $SGLDM M_{i,j}^{\theta,d}$ est formé en calculant le nombre d'occurrences de pixels dont le niveau de gris i est éloigné d'une distance d de tout pixel de niveau de gris j dans une direction définie par l'angle θ . Nous avons calculé huit matrices de cooccurrence statistique $M_{i,j}^{\theta,d}$ à partir des huit directions $\theta = \{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4 \text{ and } 7\pi/8\}$, Et à toutes les distances $d \in \{1, 2, 3, \dots, L/2\}$, où L est la longueur de la région d'intérêt. Pour chaque valeur de distance d ; Les huit matrices de cooccurrence sont additionnées pour obtenir une matrice invariante de rotation M^d :

$$M_{i,j}^d = \sum_{\theta} M_{i,j}^{\theta,d} \quad \text{Équation V-16}$$

Les caractéristiques suivantes sont extraites de la somme des matrices de co-occurrences M^d , pour chaque valeur de distance d :

1. Corrélation:

$$Cor^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d \left[\frac{(i - \mu_i^d)(j - \mu_j^d)}{\sigma_i^d \sigma_j^d} \right] \quad \text{Équation V-17}$$

2. Homogénéité:

$$H^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \left[\frac{M_{i,j}^d}{1 + (i - j)^2} \right] \quad \text{Équation V-18}$$

3. Variance:

$$Var^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d)^2 \quad \text{Équation V-19}$$

4. Energie:

$$E^d = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (M_{i,j}^d)^2 \quad \text{Équation V-20}$$

5. La moyenne des sommes :

$$SM^d = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i + j) \quad \text{Équation V-21}$$

6. Proéminence de Cluster :

$$CP^d = \frac{1}{2} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} M_{i,j}^d (i - \mu_i^d + j - \mu_j^d)^4 \quad \text{Équation V-22}$$

Où:

G est le nombre de niveaux de gris.

μ_i^d, σ_i^d : sont la moyenne et l'écart-type de la ligne i de la matrice $M_{i,j}^d$.

μ_j^d, σ_j^d : sont la moyenne et l'écart-type de la colonne j de la matrice $M_{i,j}^d$.

μ : est la valeur moyenne de la matrice invariante de rotation $M_{i,j}^d$.

b. La méthode Run difference matrix (RDM)

Cette méthode est basée sur l'estimation de la fonction de probabilité de la densité des différences de niveaux de gris dans une image. RDM est la matrice qui se compose de la différence de niveau de gris avec une distance entre les pixels, lorsque le vecteur de déplacement entre deux pixels est donné [Kim 1991]. RDM pourrait être défini comme le

nombre total de pair de pixels dans la région d'intérêt avec une distance r et de différence de niveau de gris g_{dif} avec la direction θ :

$$RDM(r, g_{dif} | \theta) = \# \left\{ \begin{array}{l} P_i(x_i, y_i): P_j(x_j, y_j) \in ROI^2; \\ \text{Où } |G(x_j, y_j) - G(x_i, y_i)| = g_{dif}, \\ \text{et } r = \sqrt{(x_i, y_i)^2 + P_j(x_j, y_j)^2} \end{array} \right\} / N(r) \quad \text{Équation V-23}$$

Où $G(x, y)$ est la valeur de niveau de gris du pixel $P(x, y)$, et $N(r)$ est utilisé pour la normalisation il est défini par le nombre total des pairs de pixels ayant une distance r dans ROI:

$$N(r) = \# \left\{ \begin{array}{l} P_i(x_i, y_i): P_j(x_j, y_j) \in ROI^2; \\ \text{Où } r = \sqrt{(x_i, y_i)^2 + P_j(x_j, y_j)^2} \end{array} \right\} \quad \text{Équation V-24}$$

Elfarrar et al. Dans [Elfarrar 2012] ont proposé d'élargir les matrices RDM en utilisant huit angles différents $\theta = \{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4 \text{ and } 7\pi/8\}$.

En conséquence, les huit matrices RDM sont générées et additionnées pour obtenir une matrice M_{RDM} invariante en rotation.

$$M_{AMD}(r, g_{dif}) = \sum_{\theta} RDM(r, g_{dif} | \theta) \quad \text{Équation V-25}$$

Nous pouvons calculer les deux vecteurs caractéristiques : la répartition de la différence des niveaux de gris (RDNG) et la distribution de la distance moyenne (DDM) à partir de la matrice M_{RDM} obtenue.

$$RDNG(g_{dif}) = \sum_{r=1}^{L/2} M_{RDM}(r, g_{dif}) \quad \text{Équation V-26}$$

$$DDM(g_{dif}) = \sum_{r=1}^{L/2} M_{RDM}(r, g_{dif}) \times r \quad \text{Équation V-27}$$

V.2.3.3.2. Extraction de caractéristiques de l'intensité

Les caractéristiques d'intensité sont déterminées par la moyenne et l'écart type de la luminosité G de masse (équations V.39-V.40).

Où :

$G(P)$ est le niveau de gris du pixel P ,

np est le nombre total de pixels dans la masse.

V.2.3.3.3. Les caractéristiques de texture et de densité utilisées

Dans [Elfarra 2012] Elfarra et al. Proposent une méthode pour identifier les caractéristiques texturales les plus significatives, parmi 65 caractéristiques extraites en utilisant les deux méthodes mentionnées précédemment. Dans ce travail, nous avons choisi d'utiliser les onze descripteurs les plus représentatifs selon [Elfarra 2012] (c9-c19), calculés seulement sur la région segmentée.

Les caractéristiques de densité	Valeur	
Le Moment d'ordre 2 de distribution de la différence des niveaux de gris (RDNG), qui mesure la variation des différences de niveaux de gris.	$c_9 = SMG = \sum_{g_{dif}=0}^{G-1} (DDG(g_{dif}))^2$	Équation V-28
Accentuation de longue distance pour une grande différence, Qui mesure la proéminence de grandes différences des niveaux de gris pour longue distance.	$c_{10} = ALDGD = \sum_{g_{dif}=0}^{G-1} DDM(g_{dif}) \times (g_{dif})^2$	Équation V-29
Écart-type de la variance	$c_{11} = \frac{1}{L-2} \sum_{d=1}^{L/2} (Var^d - \overline{Var})^2$	Équation V-30
La déviation moyenne absolue de l'homogénéité.	$c_{12} = \frac{1}{L-2} \sum_{d=1}^{L/2} H^d - \overline{H} $	Équation V-31
La Moyenne d'homogénéité	$c_{13} = \frac{1}{L-2} \sum_{d=1}^{L/2} H^d$	Équation V-32
La déviation moyenne absolue de la corrélation	$c_{14} = \frac{1}{L-2} \sum_{d=1}^{L/2} Cor^d - \overline{Cor} $	Équation V-33
Minimum du Proéminence de Cluster	$c_{15} = \text{Min}_{d=1}^{L/2} CP^d$	Équation V-34
Moyenne de Proéminence de Cluster	$c_{16} = \frac{1}{L-2} \sum_{d=1}^{L/2} CP^d$	Équation V-35
Minimum de la somme moyenne	$c_{17} = \text{Min}_{d=1}^{L/2} SA^d$	Équation V-36
Minimum de l'énergie	$c_{18} = \text{Min}_{d=1}^{L/2} E^d$	Équation V-37
Maximum de la corrélation	$c_{19} = \text{Max}_{d=1}^{L/2} Cor^d$	Équation V-38
La moyenne de l'intensité	$c_{20} = \sum_{i=1}^{np} G(P_i) / np$	Équation V-39
L'Écart-type de l'Intensité	$c_{21} = \sqrt{\sum_{i=1}^{np} (G(P_i) - c_{20})^2 / np}$	Équation V-40

Tableau V-5: Les caractéristiques de densité.

V.2.3.3.4. Autres caractéristiques

En plus de caractéristiques précédemment cités, deux autres caractéristiques très utiles dans l'étape de classification ont été utilisé dans notre système DxAO : l'âge du patient et la taille de la masse. Ces deux caractéristiques peuvent reclasser une anomalie dans une classe BI-RADS inférieure ou supérieure [Lévy 2005].

Il est facile de calculer la taille réelle de la masse en mm^2 comme le montre (Équation V-41), qui est représenté par la zone englobé par le contour Ω . Où "res" est la résolution utilisée pendant la numérisation.

Dans notre travail, l'âge du patient est la caractéristique humaine unique utilisée.

Autres caractéristiques	Valeur	
Taille de la masse	$c_{22} = (\sum_{p \in \Omega} p) / (res)^2$	Équation V-41
L'âge du patient	$c_{23} = Patient (age)$	Équation V-42

Tableau V-6: *Autres caractéristiques utilisées.*

V.2.3.4. La classification

La classification dans notre approche a été faite en utilisant un PMC (Perceptron Multicouches). PMC est un réseau neuronal artificiel (RNA) avec apprentissage supervisé, à savoir, l'algorithme de rétro-propagation, spécialement dédié aux fins de classification. Il a la capacité de séparer des classes de motifs non linéairement séparables. Notre vecteur de caractéristiques est composé de 23 caractéristiques extraites à l'étape précédente, présentées dans (Tableau V-1, Tableau V-4, Tableau V-5 et Tableau V-6).

En effet, nous avons effectué deux schémas de classification différents, le premier consiste à classier les masses sur les quatre classes BI-RADS (BI-RADS 2, BI-RADS 3, BI-RADS 4 et BI-RADS 5) ce qui signifie que le nombre neurones de la couche de sortie $O = 4$. Le deuxième schéma consiste à classier les anomalies sur deux classes: Bénigne et maligne, dans ce cas, deux neurones seront utilisés dans la couche de sortie.

Dans les deux schémas, le nombre de neurones d'entrée est le même que la longueur du vecteur de caractéristiques calculé dans l'étape précédente ($I = 23$).

Comme mentionné dans [Karsoliya 2012] une couche cachée est suffisante pour résoudre tout problème complexe non linéaire. Pour déterminer le nombre de neurones dans la couche cachée,

nous nous basons sur une heuristique proposée par Berry et al. [Berry 1997] Qui estiment que le nombre de neurones de couche cachée ' H ' doit être inférieur à deux fois le nombre de neurones dans la couche d'entrée. Le nombre de neurones cachés H est calculé pour notre cas en utilisant la règle : $H=O+(0.75 \times I)$ et $H < 2 \times I$.

Cela signifie que : Dans le 1^{er} Scénario $H= 2I$, et Dans le 2^{ème} Scénario le nombre de neurones de la couche cachée $H= 19$.

Toutes les caractéristiques d'entrée prennent des valeurs qui peuvent être très éloignées les unes des autres. Par exemple, la caractéristique intensité peut avoir une valeur maximale de 4096, et la continuité peut avoir un maximum de 1000. Pour éviter cela, nous devons normaliser les valeurs d'entrée en utilisant la formule suivante:

$$f_{Nor} = (f - \text{Min}(f)) / (\text{Max}(f) - \text{Min}(f)) \quad \text{Équation V-43}$$

Où:

f : est la valeur de la caractéristique qui doit être normalisée.

f_{Nor} : La caractéristique normalisée.

$\text{Min}(f)$: la valeur minimal de la caractéristique f .

$\text{Max}(f)$: la valeur maximal de la caractéristique f .

Dans ce travail, nous choisissons «Sigmoidé» comme fonction de transfert, car les valeurs d'entrée de sortie sont normalisées dans l'intervalle $[0,1]$.

V.3. Base de données de mammographies

Pour mesurer les performances, tester et évaluer notre système de classification nous avons besoin d'un référentiel qui s'appelle « Benchmark ». Dans le domaine d'aide au diagnostic du cancer du sein, il existe plusieurs bases d'images : MIAS (Mammographic Image Analysis Society), DDSM (Digital Database for Screening Mammography), AMDI (Indexed Atlas of Digital Mammograms) et WDBC (Wisconsin Diagnosis Breast Cancer). Les mammographies étudiées dans notre recherche proviennent de la base de données DDSM.

Une description de de la base de données (DDSM) de l'université de la Floride du Sud. a été effectuée par « American college of Radiology » dans le lexique de BI-RADS (Breast Imaging Reporting and Data System). La base correspondante contient 2620 dossiers de patientes classés en trois groupes : normaux, bénignes et malignes.

Dans la base de données DDSM, chaque cas est enregistré dans un répertoire distinct. La Figure V.6 montre un exemple de fichiers stockées dans le répertoire "*case3024-1*". Comme vous pouvez le voir, il existe des fichiers "*.ics*", "*.LJPEG*", "*.OVERLAY*" et "*.16_PGM*". Chaque

type de fichier est décrit dans ce qui suit. Veuillez noter que les cas « normaux » ne contiendront aucun fichier « *.overlay » .

```
B-3024-1.ics
B_3024_1.LEFT_CC.LJPEG
B_3024_1.LEFT_MLO.LJPEG
B_3024_1.RIGHT_CC.LJPEG
B_3024_1.RIGHT_MLO.LJPEG
B_3024_1.LEFT_CC.OVERLAY
B_3024_1.LEFT_MLO.OVERLAY
B_3024_1.RIGHT_CC.OVERLAY
B_3024_1.RIGHT_MLO.OVERLAY
TAPE_B_3024_1.COMB.16_PGM
```

Figure V.6: La liste des fichiers du répertoire B-3024-1.

V.3.1. Une description des fichiers « .LJPEG »

Toutes les images de la base de données DDSM ont été stockées dans un format utilisant la compression (Lossless Jpeg). Chaque fichier image est de très grande taille, car les clichés ont été numérisés avec une résolution comprise entre 42 et 100 microns et chaque niveau de gris est codé de 12 à 16 bits.

Nous avons utilisé une application exécutable pour décompresser les mammographies, ce programme est téléchargeable à partir du serveur FTP de de l'Université de Stanford¹. Une fois les images sont décompressées, chaque fichier image contient uniquement les valeurs brutes des pixels. Comme il n'y a pas d'information d'en-tête dans la mammographie, les informations liées à chaque image seront obtenues à partir du fichier ".ics".

V.3.2. Description de fichier « .ics »

Ce fichier ".ics" fournit des informations sur un cas dans son ensemble. En format ASCII, il liste des informations importantes telles que la date de l'étude, l'âge des patients, la date de numérisation des films, le type de numériseur utilisé et une liste des fichiers images. Le fichier ".ics" donne également une cote de densité de tissus mammaires de 1 à 4 selon l'évaluation ACR réalisé par un radiologue expérimenté.

La taille de chaque fichier image, le nombre de bits par pixel, la résolution de numérisation (en microns) et un indicateur sur l'existence ou l'absence d'un fichier «*.Overlay» pour chaque image sont fournies. Comme vous pouvez le voir sur la Figure V.7, les quatre images "LJPEG"

¹ <ftp://figment.csee.usf.edu/pub/DDSM/software/bin/jpeg>

ont des " *Overlay* " et ces fichiers sont répertoriés dans la Figure V.6. Si les lignes de description d'image avaient "*NON-OVERLAY*" au lieu de "*OVERLAY*", les mammographies n'auraient pas de fichiers « *. *Overlay* ».

```
ics_version 1.0
filename B-3024-1
DATE_OF_STUDY 2 7 1995
PATIENT_AGE 42
FILM
FILM_TYPE REGULAR
DENSITY 4
DATE_DIGITIZED 7 22 1997
DIGITIZER LUMISYS
SELECTED
LEFT_CC LINES 4696 PIXELS_PER_LINE 3024 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
LEFT_MLO LINES 4688 PIXELS_PER_LINE 3048 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_CC LINES 4624 PIXELS_PER_LINE 3056 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
RIGHT_MLO LINES 4664 PIXELS_PER_LINE 3120 BITS_PER_PIXEL 12 RESOLUTION 50 OVERLAY
```

Figure V.7: Le fichier B-3024-1.ics.

V.3.3. Description du fichier overlay

Les cas anormaux comportent entre un et quatre fichiers « *.*Overlay* » selon le nombre d'images dans lesquelles le radiologue a marqué des anomalies. Nous pouvons savoir quelles images ont des fichiers overlay en regardant dans le fichier « .ics ». Chaque image qui a "*Overlay*" à la fin de la ligne (et non pas "*Non-Overlay*") aura un fichier "*Overlay*". Chaque fichier "*Overlay*" peut spécifier des anomalies multiples, de sorte que la première ligne du fichier indique le nombre total d'anomalies. Dans le cas d'anomalies multiples, chaque anomalie est ensuite répertoriée l'une après l'autre.

Les informations liées aux anomalies sont : le type de lésion (masses / Calcification), l'évaluation, la subtilité, la pathologie (maligne / Bénigne) et au moins une esquisse codée de contours. Les mots clés qui décrivent le type de lésion sont tirés du lexique BI-RADS de ACR. Le code de classification est une valeur de 1 à 5, et provient également de la norme BI-RADS. La note de subtilité n'est pas de la norme Bi-RADS. Il s'agit d'une valeur de 1 à 5, où 1 signifie «x1», 2 signifie «x2», 3 signifie «x3», 4 signifie «x4» et 5 signifie «x5». Le type de lésion, l'évaluation et la subtilité sont spécifiés par un radiologue expérimenté. De même, les contours pour les régions suspectes sont dérivés des marquages réalisés sur la mammographie par un radiologue expérimenté. Dans certains cas, il y a plus d'un aperçu de la même anomalie. Dans ces situations, le nombre "*TOTAL_OUTLINES*" sera supérieur à un. La Figure V.8 en illustre

un exemple. La première esquisse de contours contiendra tous les autres esquisses de contours, et tous les contours après la première commenceront par le mot «CORE».

```

TOTAL_ABNORMALITIES 1
ABNORMALITY 1
LESION_TYPE CALCIFICATION TYPE PLEOMORPHIC-FINE_LINEAR_BRANCHING
DISTRIBUTION REGIONAL
ASSESSMENT 5
SUBTLETY 4
PATHOLOGY MALIGNANT
TOTAL_OUTLINES 4
BOUNDARY
8 1368 4 4 4 4 4 4 4 4 2 2 2 2 2 2 2 2 ... 0 0 0 0 0 0 0 0 0 0 1 #
    CORE
168 1824 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 ... 1 0 1 1 0 1 1 0 1 1 #
    CORE
384 1848 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 ... 0 0 0 0 0 0 0 0 0 0 0 #
    CORE
368 2192 6 6 6 6 6 6 6 6 0 0 0 0 0 0 0 0 ... 0 0 0 0 0 0 0 0 0 0 0 #
    
```

Figure V.8: Contenu du fichier *B_3024_1.RIGHT_CC.OVERLAY*.

Chaque esquisse de contours est spécifiée comme une chaîne de code. Cette chaîne se trouve sur la ligne après le mot-clé "BOUNDARY" ou "CORE" comme discuté ci-dessus. Les deux premières valeurs de chaque chaîne sont la colonne et la ligne de départ de l'esquisse codée de contours. Après ces deux nombres, chaque nombre représente la direction de déplacement de l'esquisse par un pixel et le caractère "#" indique la fin de la chaîne de code. Les nombres correspondent aux directions sont comme suit :

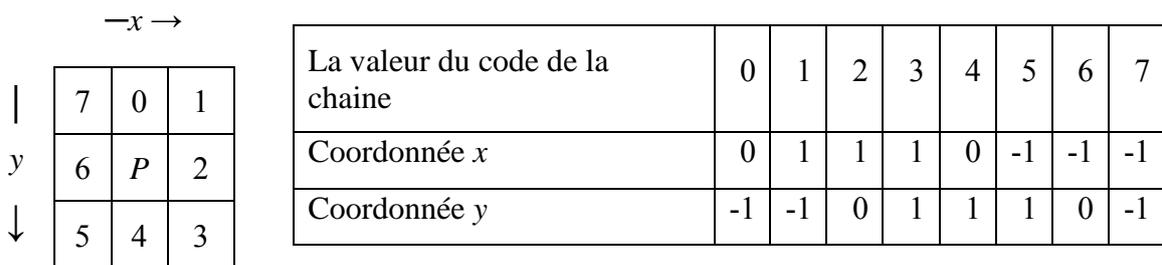


Figure V.9 : la représentation des valeurs de déplacement de la chaîne codée.

V.3.4. La description des fichiers « .16_PGM »

Les fichiers ".16_PGM" sont des images sous-échantillonnées concaténées. Ils sont stockés dans un format PGM (Portable Gray Map) sur 16 bits. Le programme populaire "imageJ" pouvait être utilisé pour visualiser ces images. Un exemple du résultat de l'affichage est présenté sur la Figure V.10. La qualité de l'image affichée peut être médiocre même si le fichier contient

des informations assez bonnes. Le but de ce type de fichiers est simplement de fournir une vue "rapide" sur les quatre images mammographiques LJPEG.

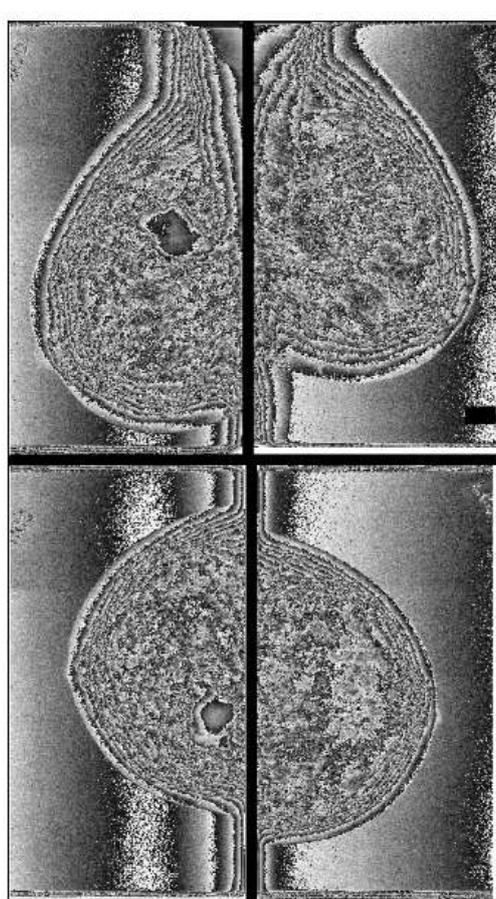


Figure V.10: *La visualisation d'un fichier ".16_PGM".*

V.4. Résultats et discussions

Dans cette section, l'approche proposée est évaluée selon deux scénarios en utilisant plusieurs mesures d'évaluation et les résultats expérimentaux sont présentés et analysés.

V.4.1. Les images de test et d'apprentissage

Les expérimentations ont été faites sur un ensemble de 480 masses choisies à partir de la base de données mammographique (DDSM) [Heath 1998] [Heath 2000]. Dans DDSM, les mammographies ont été scannées en utilisant une résolution entre 0,042 et 0,05 mm comme taille de pixel, avec une taille d'environ 3000×5000 pixels. Chaque anomalie contient des informations sur le type de lésion, l'évaluation Bi-RADS, la pathologie et les régions d'intérêt (ROI) de la masse suspectée. Les anomalies utilisées dans cette étude sont choisies par un radiologue expérimenté pour toucher toutes les catégories de densité, de forme et de type de

contours ; et répartis également sur les quatre classes BI-RADS (2, 3, 4 et 5), comme le montre le Tableau V-7.

Afin d'entraîner et de tester les PMC, nous avons séparé 288 masses pour la phase d'apprentissage et 192 images pour les tests.

Caractéristiques	Classes BI-RADS								Pathologie		Totale
	2		3		4		5		M	B	
	M	B	M	B	M	B	M	B			
Type de contours											
Circonsrite	0	120	7	104	0	0	0	0	7	224	231
Micro-lobulée	0	0	0	0	7	61	0	0	7	61	68
Effacée	0	0	4	5	1	10	0	0	5	15	20
Mal défini	0	0	0	0	3	5	49	6	52	11	63
Spiculé	0	0	0	0	11	22	63	2	74	24	98
Type de forme											
Ronde	0	69	2	43	0	0	0	0	2	112	114
Ovale	0	51	3	55	2	0	0	0	5	106	111
Lobulé	0	0	6	11	8	67	8	5	22	83	105
Irrégulière	0	0	0	0	12	31	104	3	116	34	150
Totale											
	0	120	11	109	22	98	112	8			
		120		120		120		120	145	335	480

Tableau V-7: La répartition des masses utilisées dans cette étude selon leur forme, leur type de contours et leur classe BI-RADS.

V.4.2. 1er scénario (La classification BI-RADS)

La norme BI-RADS (Breast Imaging-Reporting And Data System) a été créée par le Collège Américain de Radiologie (ACR), elle est utilisée par les radiologues en mammographie pour identifier et catégoriser les anomalies mammaires et permet de savoir ce qui devrait être fait par la suite, soit revenir à des tests, une surveillance étroite ou une biopsie. BI-RADS [Lévy 2005] est un système utilisé pour la classification des anomalies mammaires dans un petit nombre de catégories bien définies.

Les sept classes BI-RADS sont représentées dans le tableau suivant :

Catégorie	Signification	Recommandations
BI-RADS 0	L'examen est incomplet	Imagerie supplémentaire est nécessaire (IRM ou échographie) avant qu'une catégorie puisse être assignée.
BI-RADS 1	Négatif (sans cancer)	Continuer la mammographie de dépistage annuelle (pour les femmes ayant plus de 40 ans)
BI-RADS 2	Résultats bénins (100% bénignes)	Continuer la mammographie de dépistage annuelle (pour les femmes ayant plus de 40 ans)
BI-RADS 3	Une anomalie probablement bénigne (> 98% bénigne)	Recevoir un suivi mammographique de 6 mois
BI-RADS 4	Anomalie suspecte nécessitant une biopsie (10 à 50% de malignité)	Peut nécessiter une biopsie
BI-RADS 5	Hautement suspect de malignité (> 98% maligne)	Nécessitant une biopsie
BI-RADS 6	Malignité prouvée par biopsie (100% maligne)	Intervention chirurgicale, chimiothérapie ou hormonothérapie est nécessaire.

Tableau V-8: La catégorisation BI-RADS

Dans la classification BI-RADS, une masse classée BI-RADS 0, BI-RADS 4 ou BI-RADS 5 est considéré comme anomalie positif tandis qu'une masse classée BI-RADS 1, BI-RADS 2 ou BI-RADS 3 est considéré comme anomalie négatif.

Après le dépistage du cancer du sein en se basant sur un cliché mammographique initiale ; les incidences classées BI-RADS 0 signifie que l'analyse est incomplète, il y a une nécessité d'investigations d'imagerie supplémentaires (échographie ou bien IRM). Cette catégorie est importante dans le cadre du dépistage de masse ou en cas de double lecture systématique et qu'il s'avère que la qualité est insuffisante, et lorsqu'il y a besoin de clichés d'agrandissement. Les masses mammaires classées BI-RADS 1 sont négatifs (Bénignes à 100%) et sans inquiétude, tandis que celles classées BI-RADS 6 sont positifs (Malignes à 100%) et la malignité a été prouvée par biopsie ; Par conséquent, un système de diagnostic assisté par ordinateur (DxAO) doit inclure au moins les quatre catégories BI-RADS restantes (BI-RADS 2, BI-RADS 3, BI-RADS 4 et BI-RADS 5).

Selon BI-RADS [Lévy 2005], une masse est caractérisée par sa forme (ronde, ovale, lobulée ou irrégulière) (Figure V.11) ; Son contour (circonscrit, micro-lobulé, masqué, indistinct ou spiculé) (Figure V.12) ; Sa densité (élevée, moyenne, faible, grasseuse ou mixte) (Figure V.13).

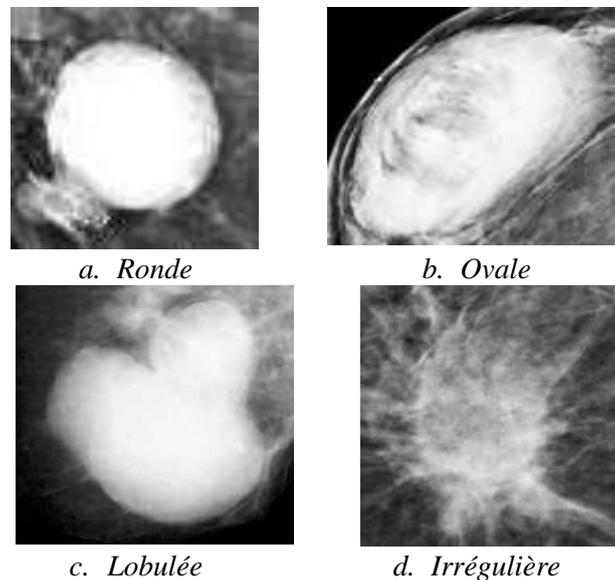


Figure V.11: Différentes formes des masses mammaires.

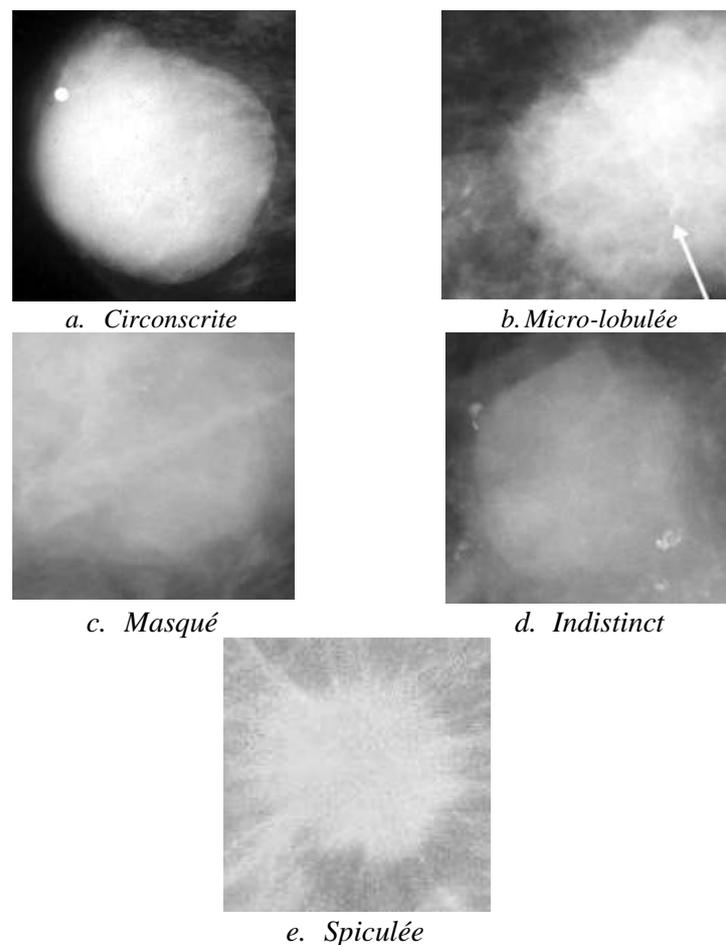


Figure V.12: Différents types de contours des masses mammaires.

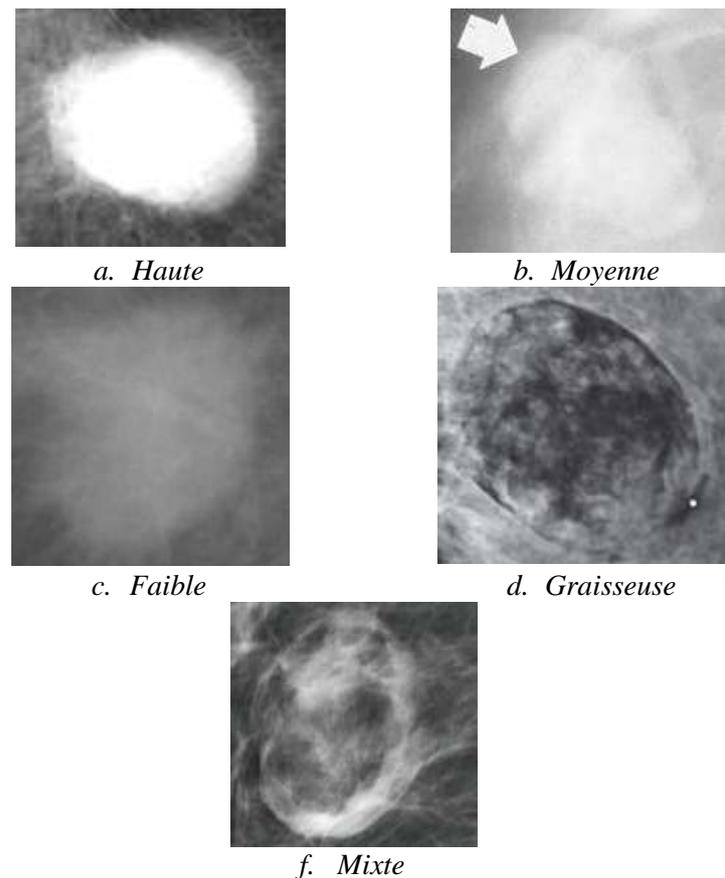


Figure V.13: Différentes densités des masses mammaires.

Dans le premier schéma de classification, le PMC a été entraîné pour apprendre la classe d'appartenance BI-RADS de chaque masse. Cependant, la phase d'apprentissage a duré environ 31.422 secondes avec 34699 époques afin d'atteindre une erreur d'entraînement 0.059. Une fois que l'apprentissage est réalisé, le PMC entraîné est utilisé pour classer les masses de la base de test (qui n'ont pas été vues pendant l'apprentissage), afin d'évaluer sa capacité de généralisation. 161 des 192 masses sont correctement classées dans leurs catégorie BI-RADS; La précision de classification obtenue atteint 83,85%, comme le montre la matrice de confusion suivante (Tableau V-9) :

Classes		Classifié BI-RADS				Taux de classification
		2	3	4	5	
Actual BI-RADS	2	42	4	1	1	87.50%
	3	3	38	5	2	79.17%
	4	1	3	40	4	83.33%
	5	1	2	4	41	85.42%
Bien classifié		161 / 192				83.85%

Tableau V-9: La matrice de confusion du premier scenario (La Classification BI-RADS).

Les plus hauts taux de classification des classes BI-RADS sont obtenues avec BI-RADS 2 et BI-RADS 5, avec 87,50% et 85,42% respectivement, en raison de leurs caractéristiques distinctives; Les masses BI-RADS 2 sont généralement de forme ronde ou ovale (Figure V.11.ab) avec des contours circonscrite (Figure V.12.a) et de densité graisseuse (Figure V.13.d), comme pour BI-RADS 5 les masses ont une forme irrégulière (Figure V.11.d) avec des contours spiculées (Figure V.12.e) et de haute densité (Figure V.13.a). Cependant, les masses BI-RADS 3 ont la plus faible précision en raison de leur chevauchement avec les classes BI-RADS 2 et BI-RADS 4.

V.4.3. 2^{ème} scénario (masses malignes Vs bénignes)

Dans le deuxième schéma, la phase d'entraînement a pour but d'apprendre si chaque masse de la base d'apprentissage sera maligne ou bénigne. Dans ce scénario, le PMC entraîné a été obtenue après 22.391 secondes avec 14843 itérations pour atteindre une erreur d'apprentissage égale à 0,0243.

Le PMC entraîné est ensuite utilisé pour classifier les masses de la base de test (qui n'ont pas été vus pendant la phase d'apprentissage), afin d'évaluer sa capacité de généralisation. Le Tableau V-10 représente la matrice de confusion obtenue à l'étape de test du 2^{ème} scénario, où 48 sur 58 masses malignes sont classées correctement comme malignes tandis que les 10 masses restantes sont classées comme bénignes. 121 de 134 masses bénignes sont bien classées dans leurs catégories, seulement 13 cas ont été classés dans la fausse classe. Le taux de classification atteint 88,02%, avec 82,76% de sensibilité et 90,30% de spécificité.

Classes		Classifiée		Taux de classification
		Bénigne	Maligne	
Classe	Bénigne	121	13	90.30%
	Maligne	10	48	82.76%
Bien Classifiée		169 / 192		88.02%

Tableau V-10: La matrice de confusion du deuxième scénario (Classification Maligne/benigne).

V.4.4. Comparaison avec les méthodes existantes

Afin d'avoir une idée plus claire sur les performances de notre système, il est nécessaire de le comparer avec des méthodes bien connues de la littérature. La comparaison avec les approches

existantes est très importante, mais en même temps c'est une tâche très difficile en raison de plusieurs facteurs. En effet, pour pouvoir comparer deux méthodes de classification, les deux méthodes doivent : traiter les mêmes anomalies, intégrant la même technique de segmentation, ce qui donne les mêmes classes en résultats. En outre, les deux méthodes doivent être évaluées sur la base des mêmes images, avec les mêmes mesures d'évaluation.

Rappelons que tous les travaux de la littérature ne classent pas les masses mammographiques selon la norme BI-RADS ; Dans ce cas nous limitons l'évaluation et la comparaison du premier scénario avec les données réelles. Quant au deuxième schéma, nous avons comparé notre approche à neuf méthodes de la littérature. Les résultats de la comparaison sont résumés dans le Tableau V-11.

Avant d'analyser les résultats obtenus, il faut dire que certaines techniques comparées ne sont pas totalement automatiques ou ont été conçues uniquement pour traiter un type particulier de masses mammographiques. Par exemple, dans [Elfarra 2012] L'âge du patient, le type de contours, la densité de la masse et les caractéristiques de forme sont donnés manuellement au classifieur, ce qui reflète le taux de classification élevé. Tandis que, dans notre approche, seulement l'âge du patient était utilisé comme un caractéristique humain.

La base de masses utilisées dans les systèmes proposés par Jiang et al. [Jiang 2008], Ying et al. [Ying 2014] et Rangayyan et al. [Rangayyan 2000] ne comprennent que des masses circonscrites et spiculées, dans ce cas un très grand nombre de masses ne sont pas prises en considération. Quant à Muramatsu et al. [Muramatsu 2016] Et Khan et al. [Khan 2016] la base d'image utilisée est composée uniquement de masses de haute densité. Rahimeh et al. [Rouhi 2016] considèrent que les masses sont de forte densité par rapport à un tissu de fond; Les algorithmes de segmentation utilisés dans leur travail ne peuvent pas détecter les masses à densité graisseuse ou mixte. Sidney et al. [Sidney 2016] ont obtenu un bon taux de classification en utilisant les caractéristiques de texture et de forme. Mais seuls les tissus graisseux de la glande mammaires sont considérés dans leur expérimentation.

Comme le montre le Tableau V-11, les résultats obtenus par notre méthode sont meilleurs par rapport à ceux obtenus par Rangayyan et al. [Rangayyan 2000], Surendiran et al. [Surendiran 2010], Yanfeng et al. [Li 2015] et Menon et al. [Dhanalekshmi 2015]. Cette classification est bonne compte tenu de la généralité de notre méthode par rapport aux autres techniques et de son applicabilité au traitement de différentes formes de masses, avec multitudes de type de contours et densités.

Technique	Base de données	Scénario de classification	Taux de classification
Rangayyan et al. [Rangayyan 2000]	39 MIAS 15 ALBERTA	Benign and Malign	81.5%
Muramatsu et al. [Muramatsu 2016]	376 NAGOYA	Benign and Malign	90%
Khan et al. [Khan 2016]	768 DDSM	Benign and Malign	93.95%
Sidney et al. [Sidney 2016]	355 IRMA	Normal, Benign and Malign	94.11%
Yanfeng et al. [Li 2015]	114 DDSM	Benign and Malign	85.96%
Surendiran et al. [Surendiran 2010]	300 DDSM	Benign and Malign	87%
Elfarra et al. [Elfarra 2012]	410 DDSM	Benign and Malign	91%
Menon et al. [Dhanalekshmi 2015]	22 MIAS	Benign and Malign	82%
Rahimeh et al. [Rouhi 2016]	57 MIAS	Benign and Malign	90.94%
Our approach [Ferkous 2017]	480 DDSM	BI-RADS 2.3.4.5	83.85%
		Benign and Malign	88.02%

Tableau V-11: *Comparaison avec d'autres méthodes existantes.*

L'avantage le plus important de notre travail par rapport à tous les travaux précédents est qu'un nouveau schéma de classification a été élaboré en se basant sur l'évaluation BI-RADS; Dans lequel toutes les caractéristiques observables sont automatiquement quantifiées; Seulement l'âge du patient a été utilisé comme un caractéristique humain. Ce schéma de classification est le même procéder par les radiologues pour déterminer l'appartenance des anomalies mammaires dans une classe BI-RADS en mammographie. De là, notre système peut aider les radiologues à confirmer ou à vérifier de nouveau.

Les masses utilisées dans l'étape d'expérimentation sont sélectionnées par un radiologue expérimenté pour toucher toutes les catégories de densités, de formes et de types de contours. Quant à l'inconvénient unique de ce travail est que l'étape de segmentation de notre approche est manuelle, ce qui nécessite un radiologue expérimenté pour le faire. Notons qu'une méthode de segmentation efficace permettra au système DxAO de traiter les autres classes BI-RADS (BI-RADS 0, BI-RADS 1 et BI-RADS 6).

V.5. Conclusion

Plusieurs recherches ont largement abordé le problème de classification des masses mammographiques, mais l'exploration de la littérature a révélé qu'aucune étude ou approche n'a été proposée pour permettre la classification des anomalies dans les catégories BI-RADS.

L'objectif de ce travail est de proposer un système DxAO capable de catégoriser une masse mammographique dans l'une des quatre classes BI-RADS (BI-RADS 2, BI-RADS 3, BI-RADS 4 et BI-RADS 5), Qui est la même manière procéder par les radiologues en mammographie. Notre système peut également différencier entre les masses bénignes et malignes. Dans cette étude, nous avons extrait automatiquement 22 caractéristiques. Pour décrire la forme, nous avons utilisé la continuité, la courbure, la taille et l'irrégularité ; En ce qui concerne les caractéristiques de contours, cinq caractéristiques ont été extraites d'un ensemble de formes d'onde par l'analyse d'ondelettes ; Les caractéristiques de densité ont été également calculées en utilisant les méthodes SGLDM et RDM. Une seule caractéristique humaine qui représente l'âge du patient a été utilisée. Notre méthode a été comparée à plusieurs travaux, et les résultats obtenus sont encourageants. Enfin, nous pouvons dire que notre DxAO est le premier système qui gère la classification en respectant la norme BI-RADS.

CONCLUSION GENERALE

Le cancer du sein est la maladie féminine la plus fréquente et le deuxième cancer dans la population après celui de la prostate. En Algérie, le nombre de décès annuels dû à cette maladie est évalué à 3500 selon les statistiques de l'institut national de santé publique Algérienne. En outre, on recense 10910 nouveaux cas de maladie par an. Entre 1986 et 2014, le nombre de cas de cancers du sein a augmenté de 700%. Cette forte évolution peut être s'expliquer par la mise en œuvre spécifique des plans de dépistages, ces programmes qui ont joué un rôle essentiel dans l'orientation des femmes vers les modalités de dépistage et par l'allongement de la durée de vie, mais aussi, selon plusieurs experts, par l'évolution des facteurs de risque environnementaux et comportementaux. Sur le niveau international, selon les statistiques de l'institut national du cancer (INCA) en 2012 le nombre de cancers du sein augmente, près de 1,7 million de cas ont été détectés, soit une augmentation de plus de 20% depuis les dernières estimations en 2008. Ainsi, 6,3 millions de femmes vivaient avec un cancer du sein diagnostiqué dans les 5 dernières années.

Plus un cancer du sein est détecté tôt, plus les chances de guérison sont grandes. Grâce au dépistage précoce, ce niveau de détection permet d'atteindre de très hauts taux de guérison tout en réduisant considérablement l'agressivité des traitements appliqués.

La mammographie est l'examen radiographique qui emploie de faibles doses de radiation pour produire une image du sein sur un cliché. Elle se considère aujourd'hui le meilleur examen pour le dépistage précoce du cancer du sein.

Cependant, cet examen comporte des limites, pouvant être liées à de réelles difficultés d'interprétation ou à des facteurs humains liés au lecteur, comme la fatigue, la distraction ou le niveau d'expérience du radiologue. Les lésions manquées sont souvent liées à une insuffisance de détection du radiologue. Le taux de faux négatifs en première lecture est loin d'être négligeable puisqu'il a été évalué entre 16 et 31 % [Bird 1992]. Afin de réduire ce chiffre, une relecture informatique des mammographies de dépistage est recommandée. Cette double lecture réduit le taux de faux négatifs de 5 à 15 %, en améliorant particulièrement la détection des cancers de petite taille. Les expertises combinées des radiologues et des systèmes de détection et de diagnostic (DACS) atteindraient un taux de réussite élevé. Les résultats montrent que ces systèmes sont vraiment intelligents et que la combinaison des interprétations humaines et de

l'intelligence artificielle se traduira par des diagnostics plus précis et plus cliniquement utiles pour guider les décisions de traitement.

Plusieurs équipes se sont intéressées à des systèmes DACS pour l'aide à la détection et au diagnostic des lésions mammographiques. Ils sont destinés à une lecture automatique qui vise à aider mais pas à se substituer au radiologue. Le diagnostic assisté par ordinateur des clichés mammographiques peut alors servir de deuxième lecture informatique pour le radiologue. Le but de ces systèmes est d'augmenter la sensibilité de la mammographie tout en réduisant le nombre de faux positifs.

Plusieurs recherches ont largement abordé le problème de classification des masses mammographiques, mais l'exploration de la littérature a révélé qu'aucune étude ou approche n'a été proposée pour permettre la classification des anomalies dans les catégories BI-RADS.

L'objectif de ce travail est de proposer un système DxAO capable de catégoriser une masse mammographique dans l'une des quatre classes BI-RADS (BI-RADS 2, BI-RADS 3, BI-RADS 4 et BI-RADS 5), Qui est la même manière procéder par les médecins. Notre système peut également différencier entre les masses bénignes et malignes. Dans cette étude, nous avons extrait automatiquement 22 caractéristiques. Pour décrire la forme, nous avons utilisé la continuité, la courbure, la taille et l'irrégularité ; En ce qui concerne les caractéristiques de contours, cinq descripteurs ont été extraites à partir d'un ensemble de formes d'onde par analyse d'ondelettes ; Les caractéristiques de densité ont été également extraites en appliquant les méthodes d'analyse de texture SGLDM et RDM en plus de la moyenne et de l'écart-type de l'intensité. Une seule caractéristique humaine qui représente l'âge du patient a été utilisée. Selon les scénarios appliqués, des caractéristiques sont ensuite introduites dans un PMC pour sortir soit la classe BI-RADS de la masse d'entrée, soit si elle est maligne ou bénigne. Notre méthode a été comparée à plusieurs travaux, et les résultats obtenus sont encourageants. Enfin, nous pouvons dire que notre méthode est la première technique qui gère la classification comme dans BI-RADS.

En conclusion, cette recherche a montré des résultats encourageants et une performance qui correspond à l'intelligence humaine pour la distinction entre les anomalies malignes et bénignes et entre les quatre classes BI-RADS (2, 3, 4 et 5) dans les mammographies numériques. L'objectif de cette recherche s'est concentré sur l'exactitude et l'efficacité des méthodes de diagnostic assistée par ordinateur utilisées dans le domaine mammographique.

Les résultats expérimentaux présentés au cinquième chapitre de cette thèse mettent en évidence l'importance et les principales contributions de cette recherche :

- ✓ La première contribution consiste à utiliser les coordonnées polaires pour placer les formes d'ondes chaque $\theta = \pi / 16$ au lieu tous les 5 pixels, ce qui signifie que le nombre de formes d'ondes a été fixé à 32 ; l'objectif de fixer le nombre de formes d'onde est de normaliser les caractéristiques de contours, ce qui permet aux masses quelle que soit leurs tailles, d'avoir le même nombre de formes d'ondes. Ainsi, après une série d'expérimentation une longueur des formes d'onde $l = 128$ a été considérée suffisante pour une meilleure extraction de caractéristiques, cette longueur garantit le non-chevauchement avec le muscle pectoral, les contours de la glande mammaire ou bien les autres masses.
- ✓ La deuxième contribution consiste à utiliser les énergies du *Snake* pour décrire la forme de la masse ; une masse circulaire ou ovale seront représentées par des énergies minimales, d'une autre part la masse irrégulière se caractérise par une forme disposant des énergies maximales.
- ✓ Ainsi, un schéma de classification automatique des masses mammographique selon le lexique BI-RADS de l'ACR a été proposé, ce système s'inspire de l'approche du médecin lors d'un examen radiologique. Le système développé dans cette recherche présente quelques avantages notables pour les radiologues. Tout d'abord, ce système aidera les radiologistes cliniques dans le processus d'interprétation mammographique en agissant comme un deuxième lecteur après les radiologues. Deuxièmement, ce système réduira le nombre de faux positifs (PF), ce qui éliminera la nécessité de réaliser des biopsies inutiles et d'économiser des coûts. Enfin, ce système réduira le temps d'examen des patients en inspectant les mammographies et en signalant les résultats en quelques secondes. Le système informatisé de diagnostic du cancer du sein développé dans cette recherche atteint un taux de classification réussie de 83,85% pour différencier entre les quatre classes BI-RADS (2, 3, 4 et 5) et une précision de classification moyenne de 88,02% pour la distinction entre les anomalies malignes et bénignes.
- ✓ Ce travail met l'accent sur toutes les caractéristiques utilisées par les radiologues durant un examen mammographique incluant la densité, la forme et les descripteurs de contours, en plus l'âge du patient et la taille de la masse, comme il a été convenu dans le lexique BI-RADS.
- ✓ Les masses utilisées à l'étape expérimentale sont sélectionnés par un radiologue expérimenté afin de toucher toutes les catégories de densités, de formes et types de contours.

Cependant dans les travaux futurs, il peut être suggéré d'appliquer une technique de prétraitement puissante à l'étape de segmentation applicable à tous les types de masses, ce qui nous permet de traiter les classes BI-RADS restantes (0, 1 et 6).

Il est nécessaire également de chercher d'autres caractéristiques plus efficaces, en se basant sur des méthodes de sélection fiables.

Dans la classification, nous pouvons aussi utiliser d'autres classifieurs, afin d'améliorer le taux de reconnaissance, de maximiser le nombre des vrais positifs (VPs) et de minimiser le nombre des faux négatifs (FNs).

Références bibliographiques

- [Abbas 2015] Abbas, A. H., Kareem, A. A., & Kamil, M. Y. (2015). *breast cancer image segmentation using morphological operations*. international Journal of Electronics and Communication Engineering & Technology, 6 (4), 08-14.
- [Abid 2009] Abid, L. (2009). *Cancer epidemiology in Algeria: Best use of cancer registers*. African Journal of Cancer, 1 (2), 98-103.
- [Abubacker 2016] Abubacker, N. F., Azman, A., Doraisamy, S., & Murad, M. A. A. (2016). *An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification*. Neural Computing and Applications, 1-14. DOI: 10.1007/s00521-016-2290-z.
- [Arevalo 2016] Arevalo, J., González, F. A., Ramos-Pollán, R., Oliveira, J. L., & Lopez, M. A. G. (2016). *Representation learning for mammography mass lesion classification with convolutional neural networks*. Computer methods and programs in biomedicine, 127 (1), 248-257.
- [Arif 2005] Muhammad, Arif. (2005). *Fusion de Données : Ultime Etape de Reconnaissance de Formes, Applications à l'Identification et à l'Authentification* (Thèse pour obtenir le grade de Docteur), l'université de Tours.
- [Ashraf 2013] Ashraf, A. B., Gavenonis, S. C., Daye, D., Mies, C., Rosen, M. A., & Kontos, D. (2013). *A multichannel markov random field framework for tumor segmentation with an application to classification of gene expression-based breast cancer recurrence risk*. IEEE transactions on medical imaging, 32 (4), 637-648.
- [Astley 2004] Astley, S. M. (2004). *Computer-aided detection for screening mammography*. Academic radiology, 11 (10), 1139-1143.
- [Auephanwiriyakul 2005] Auephanwiriyakul, S., Attrapadung, S., Thovutikul, S., & Theera-Umpon, N. (Mai 2005). *Breast abnormality detection in mammograms using fuzzy inference system*. IEEE, The 14th International Conference on Fuzzy Systems, FUZZ. (pp. 155-160). Nevada (USA).
- [Bagheri 2012] Bagheri-Khaligh, A., Zarghami, A., & Manzuri-Shalmani, M. T. (Décembre 2012). *Novel Margin Features for Mammographic Mass Classification*. 11th IEEE International Conference on Machine Learning and Applications (ICMLA), (pp. 139-144). Florida (USA).
- [Baharuddin 2016] Baharuddin, W. N. A., Abdullah, S. N. H. S., Sahran, S., Qasem, A., bin Abdullah, A., Iqbal, R., & Ismail, F. (Mars 2016). *Type 2 Fuzzy Logic for mammogram breast tissue classification*. IEEE, the 2nd International Conference on Industrial Informatics and Computer Systems (CIICS), (pp. 1-6). Sharjah (United Arab Emirates).
- [Baker 2003] Baker, J. A., Rosen, E. L., Lo, J. Y., Gimenez, E. I., Walsh, R., & Soo, M. S. (2003). *Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion*. American Journal of Roentgenology, 181 (4), 1083-1088.
- [Barrat 2010] Barrat, S. (2010). *Modèles graphiques probabilistes pour la reconnaissance de formes*. (Thèse pour obtenir le grade de Docteur), l'université de Nancy 2.

- [Behnam 2010] Behnam, H., Zakeri, F. S., & Ahmadinejad, N. (2010). *Breast mass classification on sonographic images on the basis of shape analysis*. *Journal of Medical Ultrasonics*, 37 (4), 181-186.
- [Benmazou 2014] Benmazou, S., Merouani, H. F., Layachi, S., & Nadjmeddine, B. (2014). *Classification of mammography images based on cellular automata and Haralick parameters*. *Evolving Systems*, 5 (3), 209-216.
- [Benzecri 1979] Benzecri, J. P. (1979). *Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire, addendum et erratum à (BIN. MULT)*. *Les cahiers de l'analyse des données*, 4 (3), 377-378.
- [Berry 1997] Berry, M. J., & Linoff, G. (1997). *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc. ISBN: 978-0-471-17980-1.
- [Bérubé 1998] Bérubé, M., Curpen, B., Ugolini, P., Lalonde, L., & Ouimet-Oliva, D. (1998). *Level of suspicion of a mammographic lesion: use of features defined by BI-RADS lexicon and correlation with large-core breast biopsy*. *Canadian Association of Radiologists journal*, 49 (4), 223-228.
- [Besag 1991] Besag, J., York, J., & Mollié, A. (1991). *Bayesian image restoration, with two applications in spatial statistics*. *Annals of the institute of statistical mathematics*, 43 (1), 1-20.
- [Beura 2015] Beura, S., Majhi, B., & Dash, R. (2015). *Mammogram classification using two-dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer*. *Neurocomputing*, 154 (1), 1-14.
- [Bhateja 2013] Bhateja, V., Urooj, S., Pandey, A., Lay-Ekuakille, A., & Misra, M. (Mars 2013). *Improvement of masses detection in digital mammograms employing non-linear filtering*. *IEEE, International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*. (pp. 406-408). Kerala (India).
- [Bhooshan 2014] Bhooshan, N., Giger, M., Medved, M., Li, H., Wood, A., Yuan, Y., & Newstead, G. (2014). *Potential of computer-aided diagnosis of high spectral and spatial resolution (HiSS) MRI in the classification of breast lesions*. *Journal of Magnetic Resonance Imaging*, 39 (1), 59-67.
- [Biggs 1991] Biggs, D., De Ville, B., & Suen, E. (1991). *A method of choosing multiway partitions for classification and decision trees*. *Journal of Applied Statistics*, 18 (1), 49-62.
- [Bird 1992] Bird, R. E., Wallace, T. W., & Yankaskas, B. C. (1992). *Analysis of cancers missed at screening mammography*. *Radiology*, 184 (3), 613-617.
- [Blum 1998] Blum, A., & Mitchell, T. (juillet 1998). *Combining labeled and unlabeled data with co-training*. *ACM, the 11th annual conference on Computational learning theory* (pp. 92-100). Wisconsin (USA).
- [Boser 1992] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (juillet 1992). *A training algorithm for optimal margin classifiers*. *ACM, Proceedings of the 5th annual workshop on Computational learning theory*. (pp. 144-152). Pittsburgh (USA).
- [Bottigli 2007] Bottigli, U., Cascio, D., Fauci, F., Golosio, B., Magro, R., Masala, G. L., & Stumbo, S. (2007). *Massive Lesions Classification using Features based on Morphological Lesion Differences*. *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, 1 (12), 662-666.

- [Bozek 2008] Bozek, J., Delac, K., & Grgic, M. (Septembre 2008). *Computer-aided detection and diagnosis of breast abnormalities in digital mammography*. IEEE, 50th International Symposium ELMAR. (pp. 45-52). Zadar (Croatia).
- [Breiman 1996] Breiman, L. (1996). *Stacked regressions*. Machine learning, 24 (1), 49-64.
- [Bruce 1999] Bruce, L. M., & Adhami, R. R. (1999). *Classifying mammographic mass shapes using the wavelet transform modulus-maxima method*. IEEE transactions on medical imaging, 18 (12), 1170-1177.
- [Chan 1987] Chan, H. P., Doi, K., Galhotra, S., Vyborny, C. J., MacMahon, H., & Jokich, P. M. (1987). *Image feature analysis and computer-aided diagnosis in digital radiography*. Automated detection of microcalcifications in mammography. Medical physics, 14 (4), 538-548.
- [Cheikhrouhou 2007] Cheikhrouhou, I., Djemal, K., Sellami-Masmoudi, D., Derbel, N., & Maaref, H. (Octobre 2007). *Abnormalities description for breast cancer recognition*. International Conference on E-medical Systems (pp. 199-205). Fez (Maroc).
- [Chen 2005] Chen, W. M., Chang, R. F., Kuo, S. J., Chang, C. S., Moon, W. K., Chen, S. T., & Chen, D. R. (2005). *3-D ultrasound texture classification using run difference matrix*. Ultrasound in medicine & biology, 31 (6), 763-770.
- [Chen 2006] Chen, Y. W., & Lin, C. J. (2006). *Combining SVMs with various feature selection strategies. Feature extraction (Chapitre 12 pp. 315-324)*, Springer Berlin Heidelberg. ISBN: 978-3-540-35488-8.
- [Chen 2009] Chen, C. Y., Chiou, H. J., Chou, Y. H., Chiou, S. Y., Wang, H. K., Chou, S. Y., & Chiang, H. K. (2009). *Computer-aided diagnosis of soft tissue tumors on high-resolution ultrasonography with geometrical and morphological features*. Academic radiology, 16 (5), 618-626.
- [Cheng 1998] Cheng, H. D., Lui, Y. M., & Freimanis, R. I. (1998). *A novel approach to microcalcification detection using fuzzy logic technique*. IEEE transactions on medical imaging, 17 (3), 442-450.
- [Chou 2004] Chou, S. M., Lee, T. S., Shao, Y. E., & Chen, I. F. (2004). *Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines*. Expert Systems with Applications, 27 (1), 133-142.
- [Ciatto 2003] Ciatto, S., Del Turco, M. R., Risso, G., Catarzi, S., Bonardi, R., Viterbo, V., & Navarra, F. (2003). *Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography*. European journal of radiology, 45 (2), 135-138.
- [Coeurjolly 2001] Coeurjolly, D., Miguët, S., & Tougne, L. (Mai 2001). *Discrete curvature based on osculating circle estimation*. 4th International Workshop on Visual Form, (pp. 303-312). Capri (Italy).
- [Cognex 2017] Le système OCRMax de Cognex Corporation, (Massachusetts, United States), <http://www.cognex.com>. (Consulté le 17 Janvier 2017).
- [Colomer 2002] Colomer, J., Melendez, J., & Gamero, F. I. (Juin 2002). *Pattern recognition based on episodes and DTW. Application to diagnosis of a level control system*. 16th International Workshop on Qualitative Reasoning. (pp. 37-43). Sitges (Espagne)
- [Cortes 1995] Cortes, C., & Vapnik, V. (1995). *Soft margin classifiers*. Machine Learning, 20 (1), 273-297.

- [Cover 1967] Cover, T., & Hart, P. (1967). *Nearest neighbor pattern classification*. IEEE transactions on information theory, 13 (1), 21-27.
- [Czarnek 2016] Czarnek, N. M., Clark, K., Peters, K. B., Collins, L. M., & Mazurowski, M. A. (Février 2016). *Radiogenomics of glioblastoma: a pilot multi-institutional study to investigate a relationship between tumor shape features and tumor molecular subtype*. SPIE Medical Imaging (pp. 1-6). California, (USA).
- [Dacruz 2011] Dacruz, C. F. (2011). *Automatic analysis of mammography images: enhancement and segmentation techniques*. (Thèse pour obtenir le grade de Docteur), l'université de porto.
- [Dash 1997] Dash, M., & Liu, H. (1997). *Feature selection for classification*. Intelligent data analysis, 1 (3), 131-156.
- [Dash 2012] Dash, J. K., & Sahoo, L. (Mars 2012). *Wavelet based features of circular scan lines for mammographic mass classification*. IEEE, 1st International Conference on Recent Advances in Information Technology (RAIT), (pp. 58-61). Dhanbad (India).
- [David 2002] David, P., GUER, M. C., LECOANET, C., DEROSNE, R., & DISSAY, E. (2002). *Cancers infracentimétriques du sein: Valeur primordiale de l'échographie*. Le Sein, 12 (3), 213-223.
- [Dhanalekshmi 2015] Dhanalekshmi, P. M., & Phadke, A. C. (2015). *Classification using fractal features of well-defined mammographic masses using power spectral analysis and differential box counting approaches*. Artificial Intelligence and Evolutionary Algorithms in Engineering Systems. (Chapitre 53 pp. 495-501), Springer India. ISBN: 978-81-322-2134-0.
- [Dheeba 2014] Dheeba, J., Singh, N. A., & Selvi, S. T. (2014). *Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach*. Journal of biomedical informatics, 49 (1), 45-52.
- [Digabel 2008] Digabel, C. Quinzin, C & Allioux. (2007). *La mammographie de dépistage . Le dépistage du cancer du sein : un enjeu de santé publique*, 2 (1), 87-98.
- [Doi 2007] Doi, K. (2007). *Computer-aided diagnosis in medical imaging: historical review, current status and future potential*. Computerized medical imaging and graphics, 31 (4), 198-211.
- [Drake 2015] Drake, R. L., Duparc, F., Duparc, J., Mitchell, A. H. G., & Vogl, A. W. (2015). *Gray's Anatomie pour les étudiants*. Elsevier Masson. ISBN : 978-2-294-74095-4.
- [DS1000 2017] Le système DS1000 3D de Cognex Corporation, (Massachusetts, United States), <http://www.cognex.com>. (Consulté le 17 Janvier 2017).
- [Dubuisson 1990] Dubuisson, D. (1990). *Diagnostic et reconnaissance des formes*. Hermes Science Publications. ISBN : 978-2-86601-240-3.
- [Duda 1973] Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. J. Wiley and Sons. ISBN : 978-0471223610.
- [Duda 2000] Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. John Wiley & Sons. ISBN : 978-0-471-05669-0
- [Dudani 1976] Dudani, S. A. (1976). *The distance-weighted k-nearest-neighbor rule*. IEEE Transactions on Systems, Man, and Cybernetics, 6 (4), 325-327.

- [Elfarrar 2012] Elfarrar, B. K., & Abuhaiba, I. S. (2012). *Mammogram Computer Aided Diagnosis*. International Journal of Signal Processing, Image Processing and Pattern Recognition, 5 (4), 1-30.
- [Ferkous 2010] Ferkous, C., & Merouani, H. F. (Novembre 2010). *Détection Des Lésions Dans Les Mammographies Avec Les Contours Actifs*, 7ème Séminaire National en Informatique BISKRA (SNIB'2010), Biskra (Algerie).
- [Ferkous 2010b] Ferkous, C., & Merouani, H. F. (Mars 2010). *Extraction Of The Breast Region In Mammograms Using Active Contours And Chamfer Distance*. 6th International Conference on Electrical Systems and Automatic Control. Hammamet, (Tunisie).
- [Ferkous 2015] Ferkous, C., & Merouani, H. F. (Septembre 2015). *Features Extraction from Mammographic Masses Using Genetic Active Contours*, 9ième édition Atelier des Systèmes Décisionnels (ASD'2015), Tanger (Maroc).
- [Ferkous 2017] Ferkous, C., & Merouani, H. F. (2017). *Mammographic mass classification according to BI-RADS lexicon*. IET Computer Vision. 11 (3), 189-198.
- [Ferreira 2003] Ferreira, C. B. R., & Borges, D. L. (2003). *Analysis of mammogram classification using a wavelet transform decomposition*. Pattern Recognition Letters, 24 (7), 973-982.
- [Filippone 2008] Filippone, M., Camastra, F., Masulli, F., & Rovetta, S. (2008). *A survey of kernel and spectral methods for clustering*. Pattern recognition, 41 (1), 176-190.
- [Fogel 1998] Fogel, D. B., Wasson, E. C., Boughton, E. M., Porto, V. W., & Angeline, P. J. (1998). *Linear and neural models for classifying breast masses*. IEEE transactions on medical imaging, 17 (3), 485-488.
- [Freer 2001] Freer, T. W., & Ulissey, M. J. (2001). *Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center 1*. Radiology, 220 (3), 781-786.
- [Friedman 1991] Friedman, J. H. (1991). *Multivariate adaptive regression splines*. The annals of statistics, 19 (1), 1-67.
- [Gal 1984] Gal, M. L., Chavanne, G., & Pellier, D. (1984). *Valeur diagnostique des microcalcifications groupées découvertes par mammographie*. Bulletin du cancer, 71 (1), 57-64.
- [Ganesan 2013] Ganesan, K., Acharya, U. R., Chua, C. K., Min, L. C., Abraham, K. T., & Ng, K. H. (2013). *Computer-aided breast cancer detection using mammograms: a review*. IEEE Reviews in Biomedical Engineering, 6 (1), 77-98.
- [Ganesan 2013b] Ganesan, K., Acharya, U. R., Chua, K. C., Min, L. C., & Abraham, K. T. (2013). *Pectoral muscle segmentation: a review*. Computer methods and programs in biomedicine, 110 (1), 48-57.
- [Gierach 2014] Gierach, G. L., Li, H., Loud, J. T., Greene, M. H., Chow, C. K., Lan, L., & Mai, P. L. (2014). *Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study*. Breast Cancer Research, 16 (4), 1-16.
- [Giger 2001] Giger, M. L., Karssemeijer, N., & Armato, S. G. (2001). *Guest editorial computer-aided diagnosis in medical imaging*. IEEE Transactions on Medical Imaging, 20 (12), 1205-1208.

- [Glatt 1992] Glatt, L. A., Longbotham, H. G., Arnow, T. L., Shelton, D., & Ravdin, P. (1992). *Application of weighted-majority minimum-range filters in the detection and sizing of tumors in mammograms*. *Proceeding SPIE*, 1652 (1),477-488.
- [Gong 2004] Gong, X., Vedula, A. A., & Glick, S. J. (2004). *Microcalcification detection using cone-beam CT mammography with a flat-panel imager*. *Physics in medicine and biology*, 49 (11), 2183-2195.
- [Goubalan 2016] Goubalan, S. R., Goussard, Y., & Maaref, H. (Septembre 2016). *Unsupervised malignant mammographic breast mass segmentation algorithm based on pickard Markov random field*. *IEEE International Conference on Image Processing (ICIP)*, (pp. 2653-2657). Arizona (USA).
- [GroundFX 2017] le système GroundFX de GestureTek, (Toronto, Canada). <http://www.gesturetek.com/>. (Consulté le 17 Janvier 2017).
- [Guo 2009] Guo, Q., Shao, J., & Ruiz, V. F. (2009). *Characterization and classification of tumor lesions using computerized fractal-based texture analysis and support vector machines in digital mammograms*. *International journal of computer assisted radiology and surgery*, 4 (1), 11-25.
- [Guyon 2003] Guyon, I., & Elisseeff, A. (2003). *An introduction to variable and feature selection*. *Journal of machine learning research*, 3 (3), 1157-1182.
- [Haar 1910] Haar, A. (1910). *Zur theorie der orthogonalen funktionensysteme*. *Mathematische Annalen*, 69 (3), 331-371.
- [Hadjiiski 1999] Hadjiiski, L., Sahiner, B., Chan, H. P., Petrick, N., & Helvie, M. (1999). *Classification of malignant and benign masses based on hybrid ART2LDA approach*. *IEEE transactions on medical imaging*, 18 (12), 1178-1187.
- [Halkiotis 2007] Halkiotis, S., Botsis, T., & Rangoussi, M. (2007). *Automatic detection of clustered microcalcifications in digital mammograms using mathematical morphology and neural networks*. *Signal Processing*, 87 (7), 1559-1568.
- [Hamdi 2014] Hamdi-Cherif, M., Bidoli, E., Birri, S., Mahnane, A., Laouamri, S., Zaidi, Z., & Ayat, A. (2014). *Le cancer à sétif, algérie, 1986–2010*. *Journal Africain du Cancer*, 6 (3), 166-173.
- [Hamdi 2015] Hamdi Cherif, M., Bidoli, E., Birri, S., Mahnane, A., Zaidi, Z., Boukharouba, H., & Bouchaïbi, I. (2015). *Cancer estimation of incidence and survival in Algeria 2014*. *J Cancer Res Ther*, 3 (9), 100-104.
- [Haralick 1973] Haralick, R. M., & Shanmugam, K. (1973). *Textural features for image classification*. *IEEE Transactions on systems, man, and cybernetics*, 3 (6), 610-621.
- [Harris 2014] Harris, J. R., Lippman, M. E., Osborne, C. K., & Morrow, M. (2014). *Diseases of the Breast*. Lippincott Williams & Wilkins. ISBN: 978-1451186277.
- [Hartert 2010] Hartert, L. (2010). *Reconnaissance des formes dans un environnement dynamique appliquée au diagnostic et au suivi des systèmes évolutifs* (Thèse pour obtenir le grade de Docteur) Université de Reims-Champagne Ardenne.
- [Hattori 1999] Hattori, K., & Takahashi, M. (1999). *A new nearest-neighbor rule in the pattern classification problem*. *Pattern recognition*, 32 (3), 425-432.
- [Hearst 1998] Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). *Support vector machines*. *IEEE Intelligent Systems and their Applications*, 13 (4), 18-28.

- [Heath 1998] Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., & Munishkumaran, S. (1998). *Current status of the digital database for screening mammography*. Computational Imaging and Vision, 13 (1), 457-460.
- [Heath 2000] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, W. P. (Juin 2000). *The digital database for screening mammography*. Proceedings of the 5th international workshop on digital mammography. Medical Physics Publishing. 212-218. Toronto (Canada).
- [Heid 2007] Heid, P., & Noël, A. (2007). *Le contrôle qualité en mammographie*. Le dépistage du cancer du sein: un enjeu de santé publique, 2 (1), 99-114.
- [Heron 2016] Heron J.F, *La Mammographie; Cancérologie générale*, www.oncoprof.net. (Consulté le 20 décembre 2016).
- [Hologic 2016] Hologic. *Détection assisté par ordinateur « R2 Image Checker »*. <http://www.christieinmed.com/fr/r2-imagechecker-cad>. (Consulté le 15 décembre 2016).
- [Hopfield 1982] Hopfield, J. J. (1982). *Neural networks and physical systems with emergent collective computational abilities*. Proceedings of the national academy of sciences, 79 (8), 2554-2558.
- [Horowitz 1977] Horowitz, S. L., & Pavlidis, T. (Aout 1977). *Picture segmentation by a directed split-and-merge procedure*. Second International Joint Conference on Pattern Recognition, (pp. 424-433). Copenhagen (Denmark).
- [Houssami 2009] Houssami, N., Given-Wilson, R., & Ciatto, S. (2009). *Early detection of breast cancer: Overview of the evidence on computer-aided detection in mammography screening*. Journal of medical imaging and radiation oncology, 53 (2), 171-176.
- [Hsu 2002] Hsu, C. N., Huang, H. J., & Dietrich, S. (2002). *The ANNIGMA-wrapper approach to fast feature selection for neural nets*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 32 (2), 207-212.
- [Hutt 1996] Hutt, I. (1996). *The computer-aided detection of abnormalities in digital mammograms*. (Thèse pour obtenir le grade de Docteur), Université de Manchester.
- [Jenifer 2016] Jenifer, S., Parasuraman, S., & Kadirvelu, A. (2016). *Contrast enhancement and brightness preserving of digital mammograms using fuzzy clipped contrast-limited adaptive histogram equalization algorithm*. Applied Soft Computing, 42 (1), 167-177.
- [Jiang 2008] Jiang, L., Song, E., Xu, X., Ma, G., & Zheng, B. (2008). *Automated detection of breast mass spiculation levels and evaluation of scheme performance*. Academic radiology, 15 (12), 1534-1544.
- [Jirari 2005] Jirari, M. (Septembre 2005). *A Computer Aided Detection System for Digital Mammograms Based on Radial Basis Functions and Feature Extraction Techniques*. 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 4457-4460. Shanghai (China).
- [Karabatak 2009] Karabatak, M., & Ince, M. C. (2009). *An expert system for detection of breast cancer based on association rules and neural network*. Expert systems with Applications, 36 (2), 3465-3469.
- [Karsoliya 2012] Karsoliya, S. (2012). *Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture*. International Journal of Engineering Trends and Technology, 3 (6), 714-717.

- [Karssemeijer 1993] Karssemeijer, N. (Juillet 1993). *Recognition of clustered microcalcifications using a random field model*. Symposium on Electronic Imaging: Science and Technology IS&T/SPIE's (pp. 776-786). Nijmegen (Netherlands).
- [Kass 1980] Kass, G. V. (1980). *An exploratory technique for investigating large quantities of categorical data*. *Journal of the Royal Statistical Society*, 29 (2), 119-127.
- [Kass 1988] Kass, M., Witkin, A., & Terzopoulos, D. (1988). *Snakes: Active contour models*. *International journal of computer vision*, 1 (4), 321-331.
- [Keller 1985] Keller, J. M., Gray, M. R., & Givens, J. A. (1985). *A fuzzy k-nearest neighbor algorithm*. *IEEE transactions on systems, man, and cybernetics*, 15 (4), 580-585.
- [Kerautret 2008] Kerautret, B., & Lachaud, J. O. (Avril 2008). *Robust estimation of curvature along digital contours with global optimization*. *International conference on Discrete geometry for computer imagery* (pp. 334-345). Lyon (France).
- [Khamassi 2005] Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., & Guillot, A. (2005). *Actor-Critic models of reinforcement learning in the basal ganglia: from natural to artificial rats*. *Adaptive Behavior*, 13 (2), 131-148.
- [Khan 2016] Khan, S., Hussain, M., Aboalsamh, H., Mathkour, H., Bebis, G., & Zakariah, M. (2016). *Optimized Gabor features for mass classification in mammography*. *Applied Soft Computing*, 44 (1), 267-280.
- [Khoo 2005] Khoo, L. A., Taylor, P., & Given-Wilson, R. M. (2005). *Computer-aided Detection in the United Kingdom National Breast Screening Programme: Prospective Study I*. *Radiology*, 237 (2), 444-449.
- [Khuzi 2009] Khuzi, A. M., Besar, R., Zaki, W. W., & Ahmad, N. N. (2009). *Identification of masses in digital mammogram using gray level co-occurrence matrices*. *Biomedical imaging and intervention journal*, 5 (3), 1-13.
- [Kilday 1993] Kilday, J., Palmieri, F., & Fox, M. D. (1993). *Classifying mammographic lesions using computerized image analysis*. *IEEE transactions on medical imaging*, 12 (4), 664-669.
- [Kim 1991] Kim, S. I., Choi, K. C., & Lee, D. S. (Décembre 1991). *Texture classification using run difference matrix*. *IEEE, International Ultrasonics Symposium*, (pp. 1097-1100). Florida (USA).
- [Kocur 1996] Kocur, C. M., Rogers, S. K., Myers, L. R., Burns, T., Kabrisky, M., Hoffmeister, J. W., & Steppe, J. M. (1996). *Using neural networks to select wavelet features for breast cancer diagnosis*. *IEEE engineering in medicine and biology magazine*, 15 (3), 95-102.
- [Kohonen 1989] Kohonen T., (1989). *Self-organization and associative memory*. Springer Berlin Heidelberg. ISBN: 978-3-540-51387-2.
- [Koller 1996] Koller, D., & Sahami, M. (1996). *Toward optimal feature selection*. *Proceedings of the Thirteenth International Conference on Machine Learning (ICML)*. (pp. 284-292). Bari (Italy).
- [Kramer 1999] Kramer, D., & Aghdasi, F. (Septembre 1999). *Texture analysis techniques for the classification of microcalcifications in digitised mammograms*. *IEEE Africa conférence Africon*, 395-400. Cape Town (Afrique du sud).
- [Lai 1989] Lai, S. M., Li, X., & Biscof, W. F. (1989). *On techniques for detecting circumscribed masses in mammograms*. *IEEE Transactions on Medical Imaging*, 8 (4), 377-386.

- [Lakshmanan 2015] Lakshmanan, R., Thomas, V., Jacob, S. M., & Thara, P. (Septembre 2015). *Pectoral Muscle Boundary Detection in Mammograms Using Homogeneous Contours*. IEEE, Fifth International Conference on Advances in Computing and Communications (ICACC) (pp. 354-357). Kochi (India).
- [Lanyi 1985] Lanyi, M. (1985). *Morphologic analysis of microcalcifications*. *Early Breast Cancer* (Chapitre 12, pp. 113-135). Springer Berlin Heidelberg. ISBN : 978-3-642-70194-8.
- [Lanyi 2000] Lanyi, M. (2000). *Differential diagnosis of microcalcifications*. *Radiological Diagnosis of Breast Diseases* (Chapitre 7, pp. 89-136). Springer Berlin Heidelberg. ISBN: 978-3-540-66339-3.
- [Lepage 2004] Lepage, R. (2004). *Reconnaissance d'algues toxiques par vision artificielle et réseau de neurones*. (Thèse pour obtenir le grade de Docteur), l'université du Québec (Canada).
- [Lévy 2005] Lévy, L., Suissa, M., Bokobsa, J., Tristant, H., Chiche, J. F., Martin, B., & Teman, G. (2005). *Présentation de la traduction française du BI-RADS®(Breast Imaging Reporting System and Data System)*. *Gynécologie obstétrique & fertilité*, 33 (5), 338-347.
- [Li 2008] Li, C., Kao, C. Y., Gore, J. C., & Ding, Z. (2008). *Minimization of region-scalable fitting energy for image segmentation*. *IEEE transactions on image processing*, 17 (10), 1940-1949.
- [Li 2015] Li, Y., Chen, H., Rohde, G. K., Yao, C., & Cheng, L. (2015). *Texton analysis for mass classification in mammograms*. *Pattern Recognition Letters*, 52 (1), 87-93.
- [Li 2016] Li, Y., Chen, H., Wei, X., Peng, Y., & Cheng, L. (2016). *Mass classification in mammograms based on two-concentric masks and discriminating texton*. *Pattern Recognition*, 60 (1), 648-656.
- [Lima 2015] Lima, S., Azevedo, W., Cordeiro, F., Silva-Filho, A., & Santos, W. (Aout 2015). *Feature extraction employing fuzzy-morphological decomposition for detection and classification of mass on mammograms*. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS) (pp. 801-804). Milano (Italy).
- [Liu 2014] Liu, X., & Tang, J. (2014). *Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method*. *IEEE Systems Journal*, 8 (3), 910-920.
- [MacQueen 1965] MacQueen, J. (Juin 1965). *Some methods for classification and analysis of multivariate observations*. the fifth Berkeley symposium on mathematical statistics and probability (pp. 281-297). California (USA).
- [Maitra 2011] Maitra, I. K., Nag, S., & Bandyopadhyay, S. K. (2011). *Identification of abnormal masses in digital mammography images*. *International Journal of Computer Graphics*, 2 (1), 17-30.
- [Masala 2006] Masala, G. L. (2006). *Pattern recognition techniques applied to biomedical patterns*. *Int J Biomed Sci*, 1 (1), 47-55.
- [Matsubara 1998] Matsubara, T., Fujita, H., Hara, T., Kasai, S., Otsuka, O., Hatanaka, Y., & Endo, T. (1998). *Development of a new algorithm for detection of mammographic masses*. *Digital Mammography*. 13 (1),139-142.
- [McCulloch 1943] McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5 (4), 115-133.

- [Melouah 2008] Melouah, A., & Merouani, H. F. (Novembre 2008). *Hierarchical segmentation of digital mammography by agents competition*. IEEE, The Third International Conference on Digital Information Management (pp. 442-447). London (Royaume-Uni).
- [Mencattini 2008] Mencattini, A., Rabottino, G., Salmeri, M., Lojacono, R., & Colini, E. (Octobre 2008). *Breast mass segmentation in mammographic images by an effective region growing algorithm*. Springer Berlin Heidelberg, International Conference on Advanced Concepts for Intelligent Vision Systems (pp. 948-957). Juan-les-Pins (France).
- [Mencattini 2010] Mencattini, A., Salmeri, M., Rabottino, G., & Salicone, S. (2010). *Metrological characterization of a CADx system for the classification of breast masses in mammograms*. IEEE transactions on Instrumentation and Measurement, 59 (11), 2792-2799.
- [Michalski 1983] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine learning: An artificial intelligence approach*. Springer-Verlag Berlin Heidelberg. ISBN : 978-3-662-12405-5.
- [Mirada 2017] Le système médicale Mirada réalisé par l'université d'Oxford, (Oxford, UK). <http://www.mirada-medical.com>. (Consulté le 17 Janvier 2017).
- [Mohanaiah 2013] Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). *Image texture feature extraction using GLCM approach*. International Journal of Scientific and Research Publications, 3 (5), 290-294.
- [Morrow 1992] Morrow, W. M., Paranjape, R. B., Rangayyan, R. M., & Desautels, J. E. L. (1992). *Region-based contrast enhancement of mammograms*. IEEE transactions on Medical Imaging, 11 (3), 392-406.
- [Mudigonda 2001] Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. L. (2001). *Detection of breast masses in mammograms by density slicing and texture flow-field analysis*. IEEE Transactions on Medical Imaging, 20 (12), 1215-1227.
- [Muramatsu 2016] Muramatsu, C., Hara, T., Endo, T., & Fujita, H. (2016). *Breast mass classification on mammograms using radial local ternary patterns*. Computers in biology and medicine, 72 (1), 43-53.
- [Mustra 2013] Mustra, M., & Grgic, M. (2013). *Robust automatic breast and pectoral muscle segmentation from scanned mammograms*. Signal processing, 93 (10), 2817-2827.
- [Mustra 2015] Mustra, M., Grgic, M., & Rangayyan, R. M. (2015). *Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms*. Medical & biological engineering & computing, 54 (7), 1003-1024.
- .
- [Nagappan 2008] Nagappan, A., Britto, A. P., Banu, R. W., & Malmurugan, N. (2008). *Novel segmentation technique using wavelet based active contour model for detection of mammographic lesions*. Inf. Technol. J, 7 (3), 490-496.
- [Nagel 1998] Nagel, R. H., Nishikawa, R. M., Papaioannou, J., & Doi, K. (1998). *Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms*. Medical Physics, 25 (8), 1502-1506.
- [Nguyen 2007] Nguyen, T. P., & Debled-Rennesson, I. (Aout 2007). *Curvature estimation in noisy curves*. Springer Berlin Heidelberg, International conference on Computer Analysis of Images and Patterns (pp. 474-481). Vienna (Autriche).

- [Obenauer 2005] Obenauer, S., Hermann, K. P., & Grabbe, E. (2005). *Applications and literature review of the BI-RADS classification*. *European radiology*, 15 (5), 1027-1036.
- [Oja 2002] Oja, M., Kaski, S., & Kohonen, T. (2003). *Bibliography of self-organizing map (SOM) papers: 1998-2001 addendum*. *Neural computing surveys*, 3 (1), 1-156.
- [Orel 1999] Orel, S. G., Kay, N., Reynolds, C., & Sullivan, D. C. (1999). *BI-RADS categorization as a predictor of malignancy*. *Radiology*, 211 (3), 845-850.
- [Paulhac 2009] Paulhac, L. (2009). *Outils et méthodes d'analyse d'images 3D texturées: application à la segmentation des images échographiques* (Thèse pour obtenir le grade de Docteur), l'université Rabelais-Tours.
- [Petrou 2006] Petrou, M., & García Sevilla, P. (2006). *Image processing: dealing with texture*. Wiley. ISBN: 978-0-470-02628-1.
- [Petroudi 2003] Petroudi, S., Kadir, T., & Brady, M. (Septembre 2003). *Automatic classification of mammographic parenchymal patterns: A statistical approach*. IEEE, the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. (pp. 798-801). Cancun (Mexico).
- [Peura 1997] Peura, M., & Iivarinen, J. (Mai 1997). *Efficiency of simple shape descriptors*. the third international workshop on visual form (pp. 443- 451). Capri (Italy).
- [Photomodeler 2017] Le logiciel PhotoModeler de EOS Systems Inc, (Vancouver, Canada) <http://www.photomodeler.com>. (Consulté le 17 Janvier 2017).
- [Pisano 2005] Pisano, E. D., & Yaffe, M. J. (2005). *Digital Mammography*. *Radiology*, 234 (2), 353-362.
- [Polakowski 1997] Polakowski, W. E., Cournoyer, D. A., Rogers, S. K., DeSimio, M. P., Ruck, D. W., Hoffmeister, J. W., & Raines, R. A. (1997). *Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency*. *IEEE transactions on medical imaging*, 16 (6), 811-819.
- [Popli 2001] Popli, M. B. (2001). *Pictorial essay: Mammographic features of breast cancer*. *Indian Journal of Radiology and Imaging*, 11 (4), 175-179.
- [Qian 1994] Qian, W., Clarke, L. P., Kallergi, M., & Clark, R. A. (1994). *Tree-structured nonlinear filters in digital mammography*. *IEEE Transactions on Medical Imaging*, 13 (1), 25-36.
- [Quinlan 1986] Quinlan, J. R. (1986). *Induction of decision trees*. *Machine learning*, 1 (1), 81-106.
- [Quinlan 1992] Quinlan, J. R. (1992). *C4. 5: programs for machine learning*. Morgan Kaufmann Publishers Inc. ISBN :1-55860-238-0
- [Rakotomalala 2005] Rakotomalala, R. (2005). *Arbres de décision*. *Revue Modulad*, 33 (1), 163-187.
- [Rangayyan 1997] Rangayyan, R. M., El-Faramawy, N. M., Desautels, J. L., & Alim, O. A. (1997). *Measures of acutance and shape for classification of breast tumors*. *IEEE Transactions on medical imaging*, 16 (6), 799-810.
- [Rangayyan 2000] Rangayyan, R. M., Mudigonda, N. R., & Desautels, J. E. (2000). *Boundary modelling and shape analysis methods for classification of mammographic masses*. *Medical and Biological Engineering and Computing*, 38 (5), 487-496.

- [Rangayyan 2007] Rangayyan, R. M., & Nguyen, T. M. (2007). *Fractal analysis of contours of breast masses in mammograms*. Journal of Digital Imaging, 20 (3), 223-237.
- [Robin 2016] Robin, S. the Radiology Assistant. ‘*Educational site of the Radiological Society of the Netherlands*’: <http://www.radiologyassistant.nl/en/p4793bfde0ed53/> . (Consulté le 26 Juin 2016).
- [Rosado 2015] Rosado-Toro, J. A., Barr, T., Galons, J. P., Marron, M. T., Stopeck, A., Thomson, C., & Rodríguez, J. J. (2015). *Automated breast segmentation of fat and water MR images using dynamic programming*. Academic radiology, 22 (2), 139-148.
- [Rosenblatt 1958] Rosenblatt F., (1958). *The perceptron : a probabilistic model for information storage and organization in the brain*. Psychological review, 65 (6), 386-408.
- [Rouhi 2015] Rouhi, R., Jafari, M., Kasaei, S., & Keshavarzian, P. (2015). *Benign and malignant breast tumors classification based on region growing and CNN segmentation*. Expert Systems with Applications, 42 (3), 990-1002.
- [Rouhi 2016] Rouhi, R., & Jafari, M. (2016). *Classification of benign and malignant breast tumors based on hybrid level set segmentation*. Expert Systems with Applications, 46 (1), 45-59.
- [Säbel 1996] Säbel, M., & Aichinger, H. (1996). *Recent developments in breast imaging*. Physics in Medicine and Biology, 41 (3), 315.
- [Sampat 2005] Sampat, M. P., Bovik, A. C., & Markey, M. K. (2005). *Classification of mammographic lesions into BI-RADS shape categories using the beamlet transform*. SPIE Proceedings, 5747 (1), 16-25.
- [Shanmugavadivu 2016] Shanmugavadivu, P., Sivakumar, V., & Sudhir, R. (2016). *Fractal dimension-bound spatio-temporal analysis of digital mammograms*. The European Physical Journal Special Topics, 225 (1), 137-146.
- [Sidney 2016] Sidney, M., Abel, G., Wellington, P. (2016). *Detection and classification of masses in mammographic images in a multi-kernel approach*. computer methods and programs in biomedicine, 134 (1), 11-29.
- [Simonetti 1998] Simonetti, G., Cossu, E., Montanaro, M., Caschili, C., & Giuliani, V. (1998). *What’s new in mammography*. European journal of radiology, 27 (2), 234-241.
- [Singh 2011] Singh, N., Mohapatra, A. G., & Kanungo, G. (2011). *Breast cancer mass detection in mammograms using K-means and fuzzy C-means clustering*. International Journal of Computer Applications, 22 (2), 15–21.
- [Singh 2014] Singh, B., Jain, V. K., & Singh, S. (2014). *Mammogram mass classification using support vector machine with texture, shape features and hierarchical centroid method*. Journal of Medical Imaging and Health Informatics, 4 (5), 687-696.
- [Soltanian 2004] Soltanian-Zadeh, H., & Rafiee-Rad, F. (2004). *Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms*. Pattern recognition, 37 (10), 1973-1986.
- [Sonka 2014] Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning Engineering. ISBN: 978-1133593607
- [Sridhar 2015] Sridhar, B., Reddy, K. V. V. S., & Prasad, A. M. (2015). *Mammographic image analysis based on adaptive morphological fuzzy logic CAD system*. International Journal of Biomedical Engineering and Technology, 17 (4), 341-355.

- [Stewart 2014] Stewart, S., Ivy, M. A., & Anslyn, E. V. (2014). *The use of principal component analysis and discriminant analysis in differential sensing routines*. Chemical Society Reviews, 43 (1), 70-84.
- [Stricker 1995] Stricker, M. A., & Orengo, M. (Février 1995). *Similarity of color images*. Symposium on Electronic Imaging: Science & Technology (pp. 381-392). California (USA).
- [Suckling 1994] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., & Taylor, P. (1994). *The mammographic image analysis society digital mammogram database*. Excerpta Medica, 1069 (1), 375-378.
- [Surendiran 2010] Surendiran, B., & Vadivel, A. (2010). *Feature selection using stepwise ANOVA discriminant analysis for mammogram mass classification*. International J. of Recent Trends in Engineering and Technology, 3 (2), 55-57.
- [Swain 1991] Swain, M. J., & Ballard, D. H. (1991). *Color indexing*. International journal of computer vision, 7 (1), 11-32.
- [Taconet 2006] Taconet, B., Zahour, A., Ramdane, S., & Boussellaa, W. (Septembre 2006). *Classification des k-ppv par sous-voisinages emboîtés*. 9ème Colloque International Francophone sur l'Ecrit et le Document (pp. 145-150). Fribourg (Suisse).
- [Tan 2014] Tan, M., Pu, J., & Zheng, B. (Février 2014). *A new mass classification system derived from multiple features and a trained MLP model*. Proc SPIE, Medical Imaging 2014: Computer-Aided Diagnosis (pp. 1-9). California (USA).
- [Taplin 2006] Taplin, S. H., Rutter, C. M., & Lehman, C. D. (2006). *Testing the effect of computer-assisted detection on interpretive performance in screening mammography*. American Journal of Roentgenology, 187 (6), 1475-1482.
- [Tardivon 2007] Tardivon, A. (2007). *Les images détectées*. Le dépistage du cancer du sein : un enjeu de santé publique, 2 (1), 115-129.
- [Tardivon 2008] Tardivon, A. (2008). *Densité mammaire et cancer du sein*. 30^{ème} Journées de la Société Française de Sénologie et de Pathologie Mammaire. La Baule (France).
- [Thangavel 2005a] Thangavel, K., Karnan, M., Sivakumar, R., & Mohideen, A. K. (2005). *Automatic detection of microcalcification in mammograms—A review*. International Journal on Graphics Vision and Image Processing, 5 (5), 31-61.
- [Thangavel 2005b] Thangavel, K., & Karnan, M. (2005). *CAD system for Preprocessing and Enhancement of Digital Mammograms*. International Journal on Graphics Vision and Image Processing, 5 (9), 69-74.
- [Tilmanne 2009] Tilmanne, J., Urbain, J., Kothare, M. V., Wouwer, A. V., & Kothare, S. V. (2009). *Algorithms for sleep–wake identification using actigraphy: a comparative study and new results*. Journal of sleep research, 18 (1), 85-98.
- [Torrents 2014] Torrents-Barrena, J., Puig, D., Ferre, M., Melendez, J., Diez-Presa, L., Arenas, M., & Marti, J. (Juin 2014). *Breast masses identification through pixel-based texture classification*. International Workshop on Digital Mammography (pp. 581-588). Gifu City (Japan).
- [Trafficvision 2017] Le logiciel TrafficVision de Omnibond, (Pendleton, South Carolina). <http://www.trafficvision.com>. (Consulté le 17 Janvier 2017).
- [Travers 1995] Travade, A., Isnard, A., & Gimbergues, H. (1995). *Imagerie de la pathologie mammaire*. Masson. ISBN : 978-2225846472.

- [Trouilhet 1994] Trouilhet, J. F., Coustaux, I., & Guilhot, J. P. (1994). *Présentation d'une méthodologie pour la prise de décision. Application à la reconnaissance automatique de signature acoustique*. Le Journal de Physique IV, 4 (5), 1327--1330.
- [Unser 1984] Unser, M. (1984). *Description statistique de textures*. (Thèse pour obtenir le grade de Docteur), l'université de Lausanne.
- [Vachier 1995] Vachier, C. (1995). *Extraction de caractéristiques, segmentation d'image et morphologie mathématique* (Thèse pour obtenir le grade de Docteur), Ecole Nationale Supérieure des Mines de Paris.
- [Vadivel 2013] Vadivel, A., & Surendiran, B. (2013). *A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories*. Computers in biology and medicine, 43 (4), 259-267.
- [Veldkamp 2000] Veldkamp, W. J., Karssemeijer, N., Otten, J. D., & Hendriks, J. H. (2000). *Automated classification of clustered microcalcifications into malignant and benign types*. Medical Physics, 27 (11), 2600-2608.
- [Verma 2001] Verma, B., & Zakos, J. (2001). *A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques*. IEEE transactions on information technology in biomedicine, 5 (1), 46-54.
- [Veropoulos 2001] Veropoulos, K. (2001). *Machine learning approaches to medical decision making* (Thèse pour obtenir le grade de Docteur), University de Bristol.
- [Vladimir 1963] Vladimir, V., & Lerner, A. (1963). *Pattern recognition using generalized portrait method*. Automation and remote control, 24 (1), 774-780.
- [Vyborny 2000] Vyborny, C. J., Giger, M. L., & Nishikawa, R. M. (2000). *Computer-aided detection and diagnosis of breast cancer*. Radiologic Clinics of North America, 38 (4), 725-740.
- [Wang 1998] Wang, T. C., & Karayiannis, N. B. (1998). *Detection of microcalcifications in digital mammograms using wavelets*. IEEE transactions on medical imaging, 17 (4), 498-509.
- [Warren 2000] Warren Burhenne, L. J., Wood, S. A., D'Orsi, C. J., Feig, S. A., Kopans, D. B., O'Shaughnessy, K. F., & Castellino, R. A. (2000). *Potential contribution of computer-aided detection to the sensitivity of screening mammography I*. Radiology, 215 (2), 554-562.
- [Wei 2009] Wei, J., Chan, H. P., Sahiner, B., Zhou, C., Hadjiiski, L. M., Roubidoux, M. A., & Helvie, M. A. (2009). *Computer-aided detection of breast masses on mammograms: dual system approach with two-view analysis*. Medical physics, 36 (10), 4451-4460.
- [Winsberg 1967] Winsberg, F., Elkin, M., Macy Jr, J., Bordaz, V., & Weymouth, W. (1967). *Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis*. Radiology, 89 (2), 211-215.
- [Wirth 2007] Wirth, M., Nikitenko, D., & Lyon, J. (2005). *Segmentation of the breast region in mammograms using a rule-based fuzzy reasoning algorithm*. International Journal on Graphics, Vision and Image Processing, 5 (2), 45-54.
- [Woods 2011] Woods, R. W., Sisney, G. S., Salkowski, L. R., Shinki, K., Lin, Y., & Burnside, E. S. (2011). *The mammographic density of a mass is a significant predictor of breast cancer*. Radiology, 258 (2), 417-425.
- [Wu 1992] Wu, C. Y., Giger, M. L., Doi, K., Metz, C. E., Nishikawa, R. M., Vyborny, C. J., & Schmidt, R. A. (1992). *Application of artificial neural networks in mammography for the*

diagnosis of breast cancer. Conference of Imaging Technologies and Applications (pp. 19-27). Chicago (USA).

[Yang 2008] Yang, M., Kpalma, K., & Ronsin, J. (2008). *Pattern Recognition Techniques, Technology and Applications* (Chapitre 3, pp. 43-90). Peng-Yeng Yin (Ed.). ISBN : 978-953-7619-24-4.

[Ying 2014] Ying, W., Li, J., & Gao, X. (2014). *Latent feature mining of spatial and marginal characteristics for mammographic mass classification*. *Neurocomputing*, 144 (1), 107-118.

[Yoon 2016] Yoon, W. B., Oh, J. E., Chae, E. Y., Kim, H. H., Lee, S. Y., & Kim, K. G. (2016). *Automatic Detection of Pectoral Muscle Region for Computer-Aided Diagnosis Using MIAS Mammograms*. *BioMed Research International*, 2016 (1), 1-6.

[Yoshida 1996] Yoshida, H., Doi, K., Nishikawa, R. M., Giger, M. L., & Schmidt, R. A. (1996). *An improved computer-assisted diagnostic scheme using wavelet transform for detecting clustered microcalcifications in digital mammograms*. *Academic Radiology*, 3 (8), 621-627.

[Zhou 2013] Zhou, S., Shi, J., Zhu, J., Cai, Y., & Wang, R. (2013). *Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image*. *Biomedical Signal Processing and Control*, 8 (6), 688-696.

[Zouari 2002] Zouari, H., Heutte, L., Lecourtier, Y., & Alimi, A. (Janvier 2002). *Un panorama des méthodes de combinaison de classifieurs en reconnaissance de formes*. 13ème Congrès Francophone de Reconnaissance des Formes et d'Intelligence Artificielle RFIA, (p. 499-508). Angers (France).

[Zyout 2015] Zyout, I., Czajkowska, J., & Grzegorzec, M. (2015). *Multi-scale textural feature extraction and particle swarm optimization based model selection for false positive reduction in mammography*. *Computerized Medical Imaging and Graphics*, 46 (2), 95-107.

Liste des publications

Publications dans des revues internationales

Ferkous, C., & Merouani, H. F. (2017). *Mammographic mass classification according to BI-RADS lexicon*. *IET Computer Vision*. 11 (3), 189-198. DOI : [10.1049/iet-cvi.2016.0244](https://doi.org/10.1049/iet-cvi.2016.0244).

Publications dans des conférences internationales

Ferkous, C., & Merouani, H. F. (Septembre 2015). *Features Extraction from Mammographic Masses Using Genetic Active Contours*, 9ième édition Atelier des Systèmes Décisionnels (ASD'2015), Tanger, Maroc.

Ferkous, C., & Merouani, H. F. (Mars 2010). *Extraction Of The Breast Region In Mammograms Using Active Contours And Chamfer Distance*. 6th International Conference on Electrical Systems and Automatic Control. Hammamet, Tunisie.

Publications dans des conférences nationales

Ferkous, C., Merouani, H. F., & Kefali, A. R (Octobre 2016). L'extraction de caractéristiques pour la classification des masses mammographiques. 5ème journée des sciences et technologies de l'information et de la communication (JSTIC'2016). Guelma, Algerie.

Ferkous, C., & Merouani, H. F. (Novembre 2010). Détection Des Lésions Dans Les Mammographies Avec Les Contours Actifs'', 7ème Séminaire National en Informatique BISKRA (SNIB'2010), Biskra, Algerie.