

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab Blida

N° D'ordre :



Faculté des sciences

Département d'informatique

Mémoire Présenté par :

Bensaâda Lillia

En vue d'obtenir le diplôme de master

Domaine : Mathématique et informatique

Filière : Informatique
Spécialité : Informatique
Option : Ingénierie de logiciel

**Sujet : Lecture morphométrique de pathologies mammographiques
pour l'aide à l'interprétation**

Soutenu le :

M. Cherif zahar
Mme. Arkam
Mme. Reguieg F.Z
Mme. Benblidia Nadjia

Président
Examineur
Promotrice
Co-promotrice

Promotion
2014 / 2015



Remerciements

Pour ma part, voilà déjà cinq ans que je fréquente l'université Saad Dahleb de Blida 1 et, le département d'informatique en particulier. S'il me fallait alors nommer toutes les personnes qui, durant ce laps de temps, m'ont témoigné leur intérêt, leur amitié sur les plans intellectuel, professionnel ou humain, la liste en serait d'une longueur monotone. De plus, ma mémoire pouvant me faire défaut, je risquerais, ô drame, d'en oublier. C'est pourquoi, afin de ne pas me mettre dans une situation aussi compromettante, je ne citerai nommément personne.

Avant toute chose, je tiens à prosterner remerciant Allah le tout puissant et miséricordieux pour m'avoir donné, la bonne santé et, l'accomplissement de ce projet.

Ma grande gratitude, mes profonds respects, mes sincères remerciements et reconnaissance à ma promotrice Mlle **Reguieg**, pour son encadrement, sa disponibilité, sa patience et ses judicieux conseils, ainsi qu'à ma Co-promotrice Mlle **Benblidia** pour sa disponibilité et son aide.

Je tiens à remercier également:

- les membres de mon jury qui ont tous accepté avec gentillesse de lire, ma prose et de commenter mon travail.
- A tous les enseignants qui, par leurs enseignements, leurs encouragements et leur aide, ont contribué à ma formation.
- A tous mes partenaires universitaires.
- Enfin, à mes proches, pour l'amour et le soutien qu'ils m'ont témoigné, lors de la rédaction de ce mémoire. Qu'ils me pardonnent si je les ai un peu délaissés, durant cette période.

*"L'imagination est plus importante que le savoir."
A. Einstein*

Dédicaces

À

*mes parents,
avec tout mon amour.*

*Nyssia,
parce que c'est Nyssia !*

*Ma promotrice,
pour son indéfectible soutien.*

Ma famille et mes amis.

Lillia



Résumé

Le traitement des images médicales est une discipline riche et variée, mais dans laquelle les nombreuses méthodes existantes sont délicates, à appliquer aux problèmes réels. Les travaux de ce mémoire présentent, un système de détection de pathologies en mammographie, s'articulant autour de quatre axes. L'application des opérateurs de la morphologie mathématique pour le prétraitement des images et, l'utilisation d'algorithmes de segmentation par, la coopération entre le clustering flou (Fuzzy C-Means) et les modèles déformables, est réalisée par le second axe, afin d'extraire les régions d'intérêt. La troisième étape est la caractérisation des structures extraites, par des paramètres de formes et de texture. La quatrième étape utilise, cette description pour la classification des masses et des cas sains, par les supports à vecteur machine (SVM), en vue d'une aide à l'interprétation.

Mots clés: image mammographique, modèles déformables, morphologie mathématique, segmentation coopérative, texture et morphologie, SVM, aide à l'interprétation.

Abstract

The medical image is a rich and varied discipline, but in which the many existing methods are difficult to apply to real problems. The work of this thesis present a pathology detection system for mammography, articulated around four axes. The application of mathematical morphology operators for the image preprocessing and the use of segmentation algorithms, the cooperation between the fuzzy clustering (Fuzzy C-Means) and deformable models is performed by the second axis, to extract the regions of interest. The third axe is the characterization of structures extracted by parameters of shape and texture. The fourth axe uses this description for the classification of masses and healthy cases, support vector machine (SVM), for an aid to interpretation.

Keywords: mammographic image, deformable models, mathematical morphology cooperative segmentation, texture and morphology, svm, aid to interpretation.

المخلص: المعالجة للصور الطبية هو الانضباط غني ومتنوع، ولكن فيه العديد من الأساليب الحالية يصعب تطبيقها على المشاكل الحقيقية. عمل هذه الأطروحة يقدم نظام الكشف عن أمراض لتصوير الثدي، التي تتمحور حول أربعة محاور. تطبيق مشغلي مورفولوجيا الرياضية للصورة تجهيزها واستخدام خوارزميات تجزئة، ويتم تنفيذ التعاون بين المجموعات غامض (ف-سي-سائل) ونماذج تشوه يتم تنفيذها بواسطة المحور الثاني، لاستخراج المناطق المثيرة للاهتمام. الخطوة الثالثة هي توصيف هياكل استخراجها من قبل المعلومات من الشكل والملمس. أما الخطوة الرابعة تستخدم لتصنيف الكتل السرطانية وحالات صحية، وذلك بواسطة رفوف ناقلات الجهاز (س-ف-م)، لمساعدة التفسير.

كلمات المفتاحية: صورة بالأشعة، نماذج تشوه، الرياضية مورفولوجيا، تجزئة تعاونية، الملمس ومورفولوجيا، المساعدات التفسير، س-ف-م.

TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE	1
CHAPITRE 1 CONTEXTE MAMMOGRAPHIQUE	3
1.1 Introduction.....	3
1.2 Anatomie du sein.....	4
1.3 Cancer du sein.....	5
1.3.1 Types de Cancer	5
1.3.2 Les symptômes du cancer du sein	6
1.4 Les facteurs de risque du cancer du sein.....	8
1.4.1 Les facteurs de risque non modifiables	8
1.4.2 Les facteurs de risque modifiables en mammographie	9
1.5 Le contexte mammographique.....	10
1.5.1 Mammographie analogique.....	11
1.5.2 Mammographie numérique	12
1.5.3 Les incidences en mammographie.....	13
1.6 Pathologies mammaires	14
1.6.1 Kystes.....	14
1.6.2 Les calcifications.....	14
1.6.3 Les masses.....	16
1.8 Conclusion.....	19
CHAPITRE 2 PRETRAITEMENT ET SEGMENTATION EN IMAGERIE MAMMOGRAPHIQUE.....	21
2.1 Introduction.....	21
2.1 Système d'aide à la détection automatique	22
2.3 Pretraitement en imagerie mammographique	22
2.4 Introduction à la morphologie mathématique	23
2.4.1 Opérateurs de base et transformations premières de la morphologie mathématique.....	23

2.5	Segmentation en imagerie mammographique.....	28
2.5.1	Les modèles déformables.....	29
2.5.2	La segmentation par approche mixte	35
2.5.3	Segmentation par Classification.....	36
2.6	Parcours succinct de quelques travaux sur la detection des pathologies mammaires....	39
2.7	Méthodologie Adoptée.....	40
2.8	Conclusion.....	41

**CHAPITRE 3 CARACTERISATION ET CLASSIFICATION PAR LES
SUPPORTS A VECTEUR MACHINE D’IMAGES
MAMMOGRAPHIQUES.....44**

3.1	Introduction.....	44
3.2	Caracterisation texturale.....	44
3.2.1	Types de texture.....	46
3.2.2	Matrice de concurrence.....	49
3.2	Analyse de la forme	48
3.3.1	La compacité.....	48
3.3.2	La convexité.....	48
3.3.3	L’excentricité	48
3.3.4	La rectangulativité.....	49
3.3.5	La surface.....	49
3.4	Introduction à la classification	50
3.4.1	Principe de la classification non-supervisé.....	50
3.4.2	Principe de la classification supervisée	50
3.5	Les machines à vecteur support (SVM).....	53
3.5.1	Principe de fonctionnement	54
3.5.2	Classification des SVM multi-classes.....	55
3.6	Conclusion.....	57

CHAPITRE 4	APPLICATION ET RESULTATS SUR DES IMAGES	
MAMMOGRAPHIQUES.....		59
4.1	Introduction.....	59
4.2	Présentation de l'environnement du travail.....	59
4.2.1	Matériel utilisé.....	59
4.2.2	Langage utilisé.....	59
4.2.3	Présentation du 'gui'.....	61
4.3	Présentation du systeme general ' <i>mammoflousvm</i> '.....	62
4.4	Résultats et interpretation.....	65
4.4.1	Prétraitement.....	65
4.4.2	La segmentation.....	68
4.4.3	La caractérisation par les paramètres de formes et texturaux.....	71
4.4.4	La classification.....	77
4.4	Conclusion.....	79
CONCLUSION GÉNÉRALE		80
BIBLIOGRAPHIE		82

LISTE DES FIGURES

Figure1. 1 : Structure anatomique du sein [4].....	4
Figure1. 2 : Principe de la mammographie de dépistage [9].....	10
Figure1. 3 : Mammographie numérique(MN) et les clichés associés [9]	13
Figure1. 4 : Exemples d'incidences en mammographie [8].....	14
Figure1. 5 : Types des calcifications [10]	16
Figure1. 6 : Masses mammographiques [7]	17
figure1. 7 : Catégories de masses [7].....	17
Figure1. 8 : Différentes densités mammaires [7]	22
Figure2. 1 : Système de détection automatique de pathologies mammographiques [13]	22
Figure2. 2 : Érosion d'une image mammographique.....	24
Figure2. 3 : Dilatation d'une image mammographique	25
Figure2. 4 : Ouverture d'une image mammographique par un élément structurant circulaire	25
Figure2. 5 : Fermeture d'une image mammographique par un élément structurant circulaire	26
Figure2. 6 : a) organigramme de prétraitement des pathologies mammographiques	27
Figure2. 7 : Exemple simple d'application des contours actifs sur une image mammographique [20].....	30
Figure2. 8 : Fonction d'ensembles de niveaux [22].....	32
Figure2. 9 : Algorithme des FCM [28]	39
Figure3. 1 : Exemple d'une texture périodique [40]	45
Figure3. 2 : Exemple d'une texture aléatoire [40].....	46
Figure3. 3 : Boîte englobante	49
Figure3. 4 : (a) Recherche de l'hyperplan optimal ; (b) Hyperplan optimal, marge et vecteurs supports [51].....	53
Figure3. 5 : Exemples de quatre types de problèmes de discrimination binaire où il s'agit de séparer les points bleus des croix rouges. La frontière de décision est représentée en noir. ...	55

Figure4. 1 : L'icône du logiciel Matlab.....	60
Figure4. 2 : Fenêtre principale du guide [55].....	62
Figure4. 3 : Système d'aide à la décision mammographique.....	63
Figure4. 4 : Image mammographiques de la base MIAS [54]	64
Figure4. 5 : Prétraitement d'une image saine (disque de taille= 80).....	65
Figure4. 6 : Prétraitement d'une image saine (disque de taille=100).....	65
Figure4. 7 : Prétraitement d'une image saine (disque de taille =60).....	66
Figure4. 8 : Prétraitement d'une masse bénigne (disque de taille=110).....	66
Figure4. 9 : Prétraitement d'une masse bénigne (disque de taille=80).....	66
Figure4. 10 : Prétraitement d'une masse bénigne (disque de taille=100).....	67
Figure4. 11 : Prétraitement d'une masse maligne (disque de taille=80).....	67
Figure4. 12 : Prétraitement d'une masse maligne (disque de taille=120).....	67
Figure4. 13 : Prétraitement d'une masse maligne (disque de taille=80).....	68
Figure4. 14 : Résultats de la segmentation sur les masses malignes.....	69
Figure4. 15 : Résultats de la segmentation sur les masses bénignes.....	70
Figure4. 16 : Résultats de la segmentation sur les images saines	71
Figure4. 17 : Influence des caractéristiques sur des masses malignes.....	75
Figure4. 18 : Influence des caractéristiques sur des masses bénignes	76
Figure4. 19 : Influence des caractéristiques sur des cas sains.....	76

LISTE DES TABLEAUX

Tableau4. 1 : Paramètres de formes et texturaux adoptés.....	71
Tableau4. 2 : Tableau des paramètres de masses malignes.....	72
Tableau4. 3 : Paramètres des masses bénignes.....	73
Tableau4. 4 : Paramètres des cas sains.....	74

Introduction générale

Le cancer du sein est un problème de santé publique important. Ce type de cancer est le plus fréquent et le plus meurtrier chez les femmes. À ce jour, la méthode de prévention s'étant montrée la plus efficace, est le dépistage au moyen de mammographies. Celles-ci sont obtenues à partir de mammographes à rayon X. Ces appareils sont munis de deux plaques qui compressent le sein. Une source de rayons X, éclaire ce dernier et l'image est formée sur une pellicule sensible au rayonnement.

Le cancer du sein est une néoplasie touchant les cellules mammaires, qui commencent à proliférer de manière incontrôlée. C'est dans ce contexte, que l'imagerie médicale joue un rôle important aux niveaux du diagnostic et de la thérapie tout type de cancer et, plus particulièrement pour le cancer du sein. En raison de la délicatesse du cancer de sein, un grand nombre de groupes de recherche travaille, sur le problème d'aide au diagnostic. Depuis ces dernières décennies, les nouveaux travaux ne cessent d'être publiés et, le besoin de la recherche dans ce domaine, ne cesse de s'accroître.

Le travail que nous allons présenter dans ce mémoire, a suscité l'intérêt du laboratoire de traitement du signal et de l'image LATSI (laboratoire du traitement du signal et de l'image de l'université de Blida 1) ainsi que celui du laboratoire de recherche pour le développement des systèmes informatisés LRDSI (département d'Informatique de l'Université de Blida 1) depuis plus d'une dizaine d'années, pour la mise au point d'un système de détection des lésions anormales du sein, à partir des clichés mammographies numérisés.

Le cadre applicatif proposé, dans ce mémoire, est celui de l'analyse des masses mammographiques et des cas sains. C'est ce que nous allons tenter de savoir, avec l'application que nous développerons, où son but est de savoir s'il y a pathologie ou pas.

Cet objectif est justifié par le fait que bien que, la mammographie constitue la principale modalité d'investigation pour le dépistage de lésions cancéreuses; elle représente une modalité complexe à interpréter, à cause de la variété de la densité des tissus (Gras-glandulaire, dense glandulaire, grassex), de la grande diversité existante dans les zones de tumeur en termes de type, de forme, de contours, etc. Ainsi, face à la complexité de l'interprétation des mammographies, les systèmes d'aide au diagnostic sont devenus essentiels. De tels systèmes ne se limitent pas à améliorer la visualisation de l'image, ils jouent plutôt le rôle d'un « second lecteur » de l'image afin d'aider le radiologue dans sa décision de diagnostic.

L'objectif de ce projet, se focalise en premier, sur l'extraction des régions d'intérêt à partir d'images qui proviennent de l'examen utilisant un mammographe, l'appareil dédié au dépistage du cancer du sein, pour établir un premier diagnostic et, ainsi confirmé la présence d'une pathologie mammaire, suivant un prétraitement suivant les opérateurs de la morphologie mathématique et une segmentation coopérative par les modèles déformables et les méthodes de clustering flou, un traitement selon les opérations de la morphologie mathématique, pour extraire les masses.

La seconde phase du système, est consacrée à la caractérisation des différentes régions d'intérêt, suivant des paramètres de formes et texturaux adaptés aux données mammographiques, en vue d'une classification par les machines à support de vecteurs (SVM); des lésions mammaires, pour l'aide à l'interprétation.

De nombreuses méthodes ont été développées pour répondre aux différents problèmes d'interprétation et d'extraction de l'information, à partir d'images médicales. Afin de mener à bien, notre prélude à la recherche, nous structurons notre mémoire selon quatre chapitres:

- **Le chapitre 1**, Introduit le contexte mammographique, présente les notions essentielles liées à l'anatomie et aux pathologies du sein.
- **Le chapitre 2**, parcourt les principales techniques de prétraitement données par les opérateurs de la morphologie mathématique ainsi que les approches de segmentation étudiées par, les contours actifs et les techniques de clustering flou. La dernière étape du chapitre, décrit la méthodologie adoptée pour l'extraction des masses mammographiques et des cas sains.
- **Le chapitre 3**, présente l'analyse texturale et morphologique et, spécifie les paramètres choisis pour la caractérisation des lésions mammographiques suivie d'une classification par les machines à support de vecteurs (SVM).
- **Le chapitre 4**: est consacré à l'application du système développé sur des images mammographiques réelles.

Nous achèverons notre manuscrit par une conclusion générale.



CHAPITRE 1

CONTEXTE MAMMOGRAPHIQUE

Chapitre 1

Contexte mammographique

1.1 Introduction

Le cancer du sein est le cancer, le plus fréquent dans le monde. Il constitue la première cause de décès chez les femmes. L'Agence Internationale de l'Organisation Mondiale de la Santé pour la recherche sur le cancer [1] a déclaré que chaque année, plus de 150.000 femmes meurent du cancer du sein et un million de nouveaux cas, en sont détectés. Les facteurs tels que le vieillissement de la population, les nouveaux modes de vie, l'obésité due à l'alimentation riche en graisse, la consommation régulière d'alcool (plus courante chez les femmes depuis quelques décennies) ainsi que le traitement hormonal substitutif, contribuent à l'augmentation de l'incidence de ce cancer.

En Algérie, le docteur Nezzal Lahcène [2] du service de la médecine préventive et d'épidémiologie du Chu de Constantine a déclaré que pour la wilaya de Constantine, la projection pour l'année 2014 a induit une incidence de 39,38% nouveaux cas de cancer pour les femmes âgées de plus de 15 ans. Ce qui se traduit, par des chiffres compris entre 162 et 247 nouveaux cas de cancer du sein chez des femmes de moins de 45 ans. Chaque année, ce nombre augmentera de 7%.

Comme avec toute forme de cancer, le dépistage et le diagnostic précoce du cancer du sein est l'un des facteurs les plus importants, contribuant à la guérison de la maladie. Le type d'imagerie actuellement utilisée pour cette tâche, est la mammographie. Dans ce contexte, la quantité de clichés que le radiologue doit interpréter en un temps limité, est importante. De plus, la complexité de la tâche de détection est particulièrement élevée, en mammographie. Ces raisons motivent l'utilisation d'outils d'aide à la détection par ordinateur, afin d'améliorer les performances du lecteur, en minimisant en particulier le nombre de signes radiologiques associés à des lésions malignes qui, pourraient ne pas être détectées [2].

Notre objectif principal est la détection des masses par analyse morphologique et texturale de l'image mammographique. Dans ce chapitre, nous présentons un aperçu général sur l'imagerie mammographique, ses modalités quelques notions sur l'anatomie du sein, pour aborder enfin les pathologies mammaires et, situer ainsi, le cadre de notre application.

1.2 Anatomie du sein [3][4]

Le sein est composé d'une **glande mammaire** (figure 1.1) de fibres de soutien (ligaments de Cooper) et de graisse (tissu adipeux); le tout est recouvert par la peau. La quantité de chacune de ses composantes, peut varier d'une femme à l'autre. Le sein est situé par-dessus le muscle pectoral. On trouve également dans le sein des nerfs, des vaisseaux sanguins et lymphatiques. La **glande mammaire** est divisée en 15 à 20 sections qu'on appelle **lobes**, composés de **lobules**. Ceux-ci sont reliés à des **canaux**, qui se rendent sous le mamelon (situé au centre du sein). On peut également, observer des chaînes de ganglions lymphatiques qui filtrent les microbes et protègent le corps, contre l'infection et la maladie. Le cancer du sein, peut se développer tant au niveau d'un canal galactophore, que d'un lobule et il peut également, se retrouver au niveau des ganglions lymphatiques.

Le sein est une masse de tissus glandulaire, graisseux et conjonctif. Il est constitué de:

- lobules – glandes qui produisent le lait ;
- canaux – tubes qui transportent le lait des lobules au mamelon ;
- tissus graisseux et conjonctif – ils entourent et protègent les canaux et les lobules et donnent sa forme au sein ;
- aréole – surface ronde, rosée ou brunâtre qui entoure le mamelon et qui contient les petites glandes sudoripares qui sécrètent la sueur comme lubrifiant lors de l'allaitement ;
- mamelon – région située au centre de l'aréole d'où le lait est expulsé.

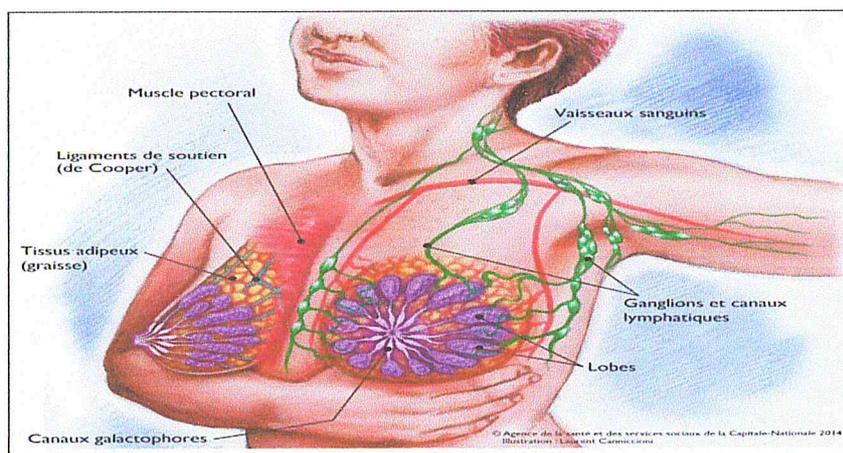


Figure 1. 1 : Structure anatomique du sein [4]

1.3 Cancer du sein [5][6]

Le cancer est une pathologie au cours de laquelle les cellules^{*}, se reproduisent indéfiniment et anarchiquement, en dehors des lois d'équilibre des tissus.

La cellule cancéreuse a pour propriété de détruire les cellules saines qui l'entourent et, de se transporter vers d'autres organes, en passant par les vaisseaux sanguins ou les vaisseaux lymphatiques (métastases).

1.3.1 Types de Cancer

Il existe différents types de cancer du sein. On peut diviser les cancers du sein en deux grandes catégories : Cancer Non invasif et Invasif.

a. Cancer non invasif

C'est le type le plus fréquent, du cancer du sein non-invasif chez la femme. C'est une forme de cancer très précoce. Elle se développe à l'intérieur des canaux de lactation du sein. Les cellules anormales ne se dispersent pas, à l'extérieur des canaux de lactation. On diagnostique beaucoup plus fréquemment ce type de cancer depuis l'utilisation plus répandue de la mammographie. Dans la majorité des cas, il est alors non-palpable. Ce type de cancer ne se dissémine pas aux ganglions ou à d'autres organes. Il n'a pas le potentiel de faire des métastases! Il est avantageux de le diagnostiquer à ce stade, car il est alors guérissable dans plus de 98 % des cas. Par conséquent, on peut être très optimiste.

b. Cancer Invasif

Cette catégorie de cancer est aussi appelée, cancer infiltrant. Ce terme signifie que les cellules cancéreuses, traversent la paroi du canal de lactation. Le cancer est souvent palpable et a la possibilité, de se propager aux ganglions de l'aisselle. Dans la plupart des cas, le cancer invasif est encore bien guérissable; cependant, si on néglige de le traiter, il pourrait se propager dans diverses parties du corps (métastases).

Il existe plusieurs types de cancer du sein invasif, à savoir le carcinome lobulaire infiltrant et le carcinome canalaire infiltrant. Ce dernier type, peut-être sans particularité aucune, ou encore démontrer des caractéristiques particulières, qui sont souvent associées à un meilleur pronostic.

* Une cellule est l'élément fondamental constituant les organes d'un être vivant.

1. 3.2 Les symptômes du cancer du sein

Les symptômes du cancer du sein, peuvent résulter d'un changement anormal de la glande mammaire. On cite en particulier, l'apparition d'un nodule ou d'un écoulement mammaire, ou un changement particulier de la peau du sein.

a. Nodule

Un nodule palpé (bosse) dans le sein n'est pas nécessairement un cancer, surtout chez les jeunes femmes. Des nodules peuvent apparaître avant les menstruations et disparaître après celles-ci. Environ 90 % des nodules observés chez les femmes de 20 à 50 ans, ne sont pas des cancers. Les fibroadénomes et les kystes, sont les nodules bénins les plus fréquents. La dysplasie mammaire ou maladie fibro-kystique du sein, est une cause fréquente de seins granuleux ou présentant des nodules dans les 2 seins.

Les femmes de plus de 50 ans, celles ayant déjà fait un cancer du sein et celles avec une histoire familiale importante; ont plus de risque, qu'un nodule palpé soit cancéreux. Un nodule est inquiétant, lorsqu'il est ferme, peu mobile et à contours irréguliers et si des changements de la peau l'accompagnent.

b. Douleur

Seulement 5 % des cancers du sein s'accompagnent de douleurs. Plus de 66% des femmes ressentent des douleurs aux seins d'intensité variables. La douleur peut être associée au syndrome prémenstruel et être, variable d'un mois à l'autre. La grossesse, la prise d'hormones à la ménopause, peuvent être des causes de douleurs aux seins. La douleur peut aussi, être causée par des problèmes au niveau des muscles, des articulations et des nerfs. L'examen physique des seins et de la cage thoracique, peut permettre de préciser, son origine.

c. Écoulement

En général, lorsque les deux seins coulent, il ne s'agit pas d'un cancer du sein. L'écoulement de lait appelé « galactorrhée » est habituellement observé au niveau des deux seins et, met en cause plusieurs canaux. Il peut persister plusieurs mois, après l'allaitement. Certains médicaments, peuvent être responsables d'une galactorrhée. Un traumatisme ou une chirurgie au thorax, peuvent aussi déclencher une galactorrhée. Plus rarement, une tumeur bénigne de l'hypophyse (petite glande située dans le cerveau) peut être responsable d'une galactorrhée persistante.

d. Changements de la peau



Inversion ou rétraction du mamelon

L'inversion du mamelon est présente chez environ 3 % des femmes, sans qu'il y ait de maladies associées. Dans 87 % des cas, cette inversion apparaîtra sur les deux seins. L'inversion lente, du mamelon sur plusieurs années, est plus souvent due à une galactophorite.



Voissure (bombement du sein)

La voissure d'un sein peut avoir diverses causes. Un volumineux kyste, pourrait entraîner ce genre de déformation tout comme une prothèse mammaire déplacée.



Rétraction cutanée

La rétraction cutanée, associée au cancer du sein entraîne une déformation du sein et elle peut être, ou non accompagnée de peau d'orange comme le dessin ci-contre. Des changements cicatriciels, un kyste inflammatoire du sein ou un kyste infecté de la peau du sein; peuvent à l'occasion donner un peu de rétraction au niveau de la peau.



Peau d'orange

Il s'agit d'un épaissement de la peau du sein causé par une obstruction des canaux lymphatiques du sein. Le gonflement de la peau qui en résulte, accentue les pores de la peau, lui donnant l'aspect d'une pelure d'orange. On observe la peau d'orange (voir dessin précédent) lorsqu'un cancer obstrue les canaux lymphatiques. La peau d'orange s'observe aussi suite à des traitements de radiothérapie du sein bien qu'il n'y ait plus de cancer dans le sein. On peut aussi observer de la peau d'orange lorsqu'il a une infection du sein comme une lymphangite, une mastite ou un abcès.



Ulcération du mamelon

Il peut être d'origine traumatique, cependant la persistance d'une ulcération après deux semaines tout comme la présence d'un eczéma persistant du mamelon ou de l'aréole doit faire penser à une maladie de Paget du mamelon ou à un cancer ulcéré.

1.4 Les facteurs de risque du cancer du sein [6]

Un facteur de risque est quelque chose, comme un comportement, une substance ou un état, qui accroît le risque d'apparition d'un cancer. La plupart des cancers, sont attribuables à de nombreux facteurs de risque, mais il arrive parfois que le cancer du sein se développe chez des femmes, qui ne présentent aucun des facteurs de risque décrits ci-dessous.

La plupart des cancers du sein, affectent la femme. La principale cause du développement de ce cancer chez la femme est le fait, que leurs cellules mammaires sont exposées aux hormones femelles, l'œstrogène et la progestérone. Ces hormones, en particulier l'œstrogène, sont associées au cancer du sein. Elles favorisent, la croissance de certains de ces cancers.

Le risque d'avoir ce type de cancer augmente, avec l'âge. Le cancer du sein est surtout observé chez les femmes âgées de 50 à 69 ans. Les causes étant multiples et inconnues, mais les travaux scientifiques ont pu identifier, certains facteurs de risques pour développer un cancer du sein.

1.4.1 Les facteurs de risque non modifiables

- Le sexe et l'âge: le cancer du sein est plus courant chez les femmes que les hommes car, les cellules des canaux mammaires des femmes, sont plus développées que les cellules des hommes. En effet, les cellules des seins de femmes sont continuellement exposées à l'hormone femelle œstrogène, hormone essentielle à la bonne croissance, au bon développement et à la vie reproductive normale de la femme, mais qui constitue également un facteur de risque du cancer du sein. Le risque de cancer du sein augmente significativement avec l'âge, mais les femmes de tout âge, peuvent présenter un cancer du sein.
- Les antécédents personnels du cancer: le fait d'avoir un antécédent personnel du cancer, est considéré comme un facteur de risque de récurrence. Le cancer du sein peut réapparaître tôt après que le premier cancer ait été diagnostiqué et traité ou, peut

survenir après de nombreuses années. Le cancer peut réapparaître dans le même sein du cancer d'origine, dans l'autre sein ou dans une autre partie du corps.

- Les antécédents familiaux du cancer et les facteurs génétiques: de nombreuses personnes croient que si le cancer est fréquent dans leur famille, ceci constitue la cause la plus courante du cancer du sein. Les recherches indiquent que seulement 5 à 10 pour cent des cancers du sein sont, de nature héréditaire.
- La menstruation précoce et la ménopause tardive : l'âge à laquelle une femme commence et termine ses menstruations, peut influencer sur son risque de présenter un cancer du sein. Si les premières menstruations ont commencé tôt « avant 12 ans » ou si la ménopause s'est manifestée, à un âge tardif « après 55 ans ».
- La densité des seins: le sein est constitué de différents types de tissus: le tissu glandulaire, le grasseux et le conjonctif. La densité des seins décrit la quantité relative, des différents tissus dans le sein. Les seins denses présentent moins de tissu grasseux et plus de tissu conjonctif et, glandulaire. Le tissu mammaire devient généralement moins dense avec l'âge, bien que certaines femmes aient toujours des seins denses sans égard à leur âge. La recherche indique que les femmes qui ont des seins denses présentent un risque accru du cancer du sein.
- Les affections des seins : il existe un certain nombre d'affections bénignes (non cancéreuses) du sein. Certaines de ces affections sont associées, à un risque accru du cancer du sein. Une affection proliférative (les cellules de seins grossissent et se reproduisent plus rapidement, que la normale) est toujours bénigne, mais elle indique qu'il y a une augmentation modérée de risque, d'avoir un cancer du sein.

1.4.2 Les facteurs de risque modifiables

- Le poids corporel : la recherche montre que les femmes qui ont un surplus du poids, peuvent présenter un risque plus élevé du cancer du sein. Ce risque augmente si la prise du poids, se fait après la ménopause.
- La consommation d'alcool et le tabagisme: la consommation d'alcool et Le tabagisme sont des facteurs de risque majeurs de maladies chroniques, incluant le cancer.

- L'exposition aux hormones synthétiques: les hormones synthétiques «œstrogènes et progestérone», peuvent également accroître le risque pour une femme, de présenter un cancer du sein.
- L'exposition aux rayonnements dans le cadre médical: les recherches ont prouvé que l'exposition à des niveaux élevés de rayonnement, est liée à la formation de différents types du cancer, incluant le cancer du sein.

1.5 Le contexte mammographique [7][8][9]

Une mammographie (ou mastographie) est un examen radiologique du sein. Il permet d'obtenir des images de l'intérieur du sein, à l'aide des rayons X pour, détecter ainsi, certaines anomalies. La quantité de rayons X délivrée lors d'une mammographie, est très faible. Il n'y a donc pas de risque de développer un cancer du sein à cause des rayons X reçus, même si des mammographies de dépistage sont répétées. Une mammographie est pratiquée dans deux circonstances: dans le cadre d'un dépistage ou d'un diagnostic.

- **Dépistage**: les femmes se soumettent de façon systématique à une mammographie, en vue du dépistage précoce du cancer du sein. On cherche des signes du développement d'un cancer du sein, même en l'absence de symptômes. La figure 1.2 indique un des résultats possibles.

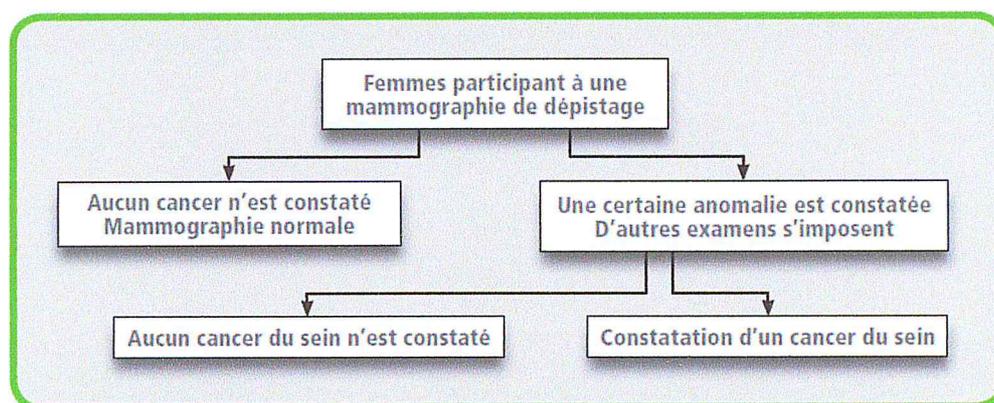


Figure 1.2: Principe de la mammographie de dépistage [9]

- **Diagnostic**: il se pratique généralement pour vérifier la présence d'un cancer du sein, après avoir constaté une grosseur ou tout autre signe ou symptôme comme de la douleur, l'épaississement de la peau, un écoulement au mamelon ou la modification de

la taille, ou de la forme d'un sein. Il peut aussi servir d'examen complémentaire, si une mammographie de dépistage révèle une anomalie.

1.5.1 Mammographie analogique

La mammographie analogique utilise comme détecteur le couple écran-film. Le film photographique fut le premier détecteur à être utilisé en radiographie, dès la découverte des rayons X. Il est sensible à la lumière et aux rayons X, dans une moindre mesure. Il contient une émulsion contenant des cristaux de d'halogénure d'argent (souvent de bromure d'argent). Ces cristaux, soumis aux photons, se dissocient en ions par effet photolytique, créant ainsi une image latente. C'est un système d'imagerie indirect. L'image latente est transformée en image réelle après plusieurs étapes se déroulant dans l'obscurité ou sous une lumière inactinique. La révélation est réalisée en plongeant le film dans une solution basique, qui réduit les ions argent positifs en argent métallique. La fixation de l'image est obtenue, en plongeant le film dans une solution acide permettant de stopper ces réactions de réduction. Après lavage et rinçage du film pour éliminer les différents réactifs, les zones du film les plus irradiées contiennent l'argent métallique et sont les plus opaques à la lumière. Les zones non-irradiées du film sont transparentes et apparaissent blanches si, on le place sur un négatoscope. Avec l'arrivée des nouveaux détecteurs, cette habitude a été conservée. Ainsi, en radiographie, les images sont présentées de façon à ce que les zones les plus exposées soient noires et les zones les moins exposées soient blanches. Pour améliorer la sensibilité du film aux rayonnements très pénétrants que sont les rayons X ou gamma, il est couplé à des écrans renforçateurs, disposés de part et d'autre du film. Ils sont constitués de sels fluorescents qui convertissent les rayons X en photons lumineux. Le couple écran film est disposé à l'abri de la lumière, dans une cassette qui est placée derrière l'objet à radiographier. Le couple écran-film possède une courbe sensitométrique d'allure sigmoïde ce qui oblige à exposer ce détecteur à une quantité précise de photons (latitude d'exposition) pour obtenir un contraste satisfaisant.

Le couple-écran film, seul détecteur analogique, est resté longtemps une référence en radiographie du fait de son excellente résolution spatiale et de sa bonne sensibilité. Néanmoins, il est de moins en moins utilisé, au profit des systèmes de détection numériques qui permettent de délivrer des doses moins importantes au patient tout en conservant une qualité d'image suffisante à un diagnostic. La mammographie conventionnelle (ou analogique, ou « classique ») produit une image imprimée sur un film argentique (un cliché).

Les limites du couple écran-film sont]:

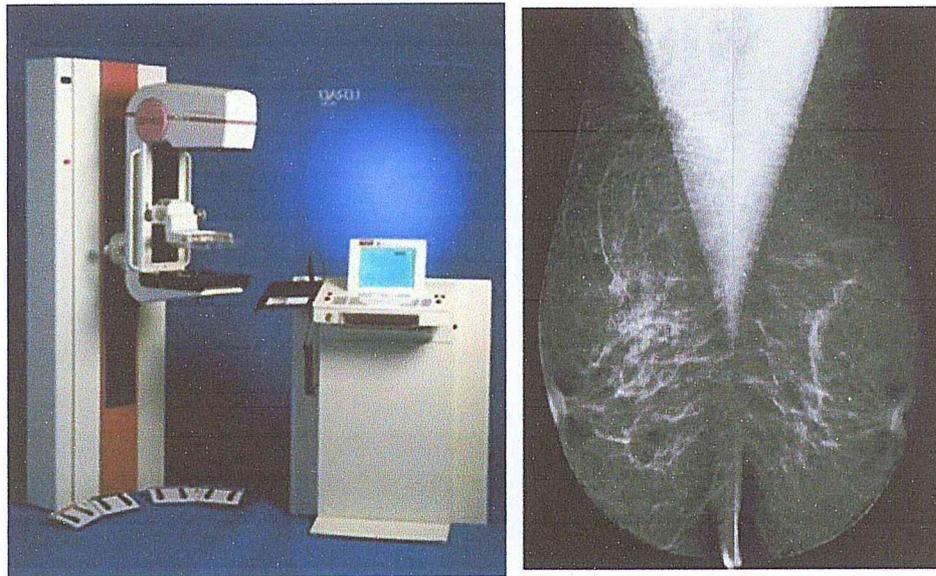
- une faible latitude d'exposition;
- la granularité du film, rend le film plus ou moins sensible, en variant la résolution spatiale. Plus le grain est fin, moins l'émulsion est sensible. Inversement, plus le grain est grand, plus l'émulsion est sensible mais ceci s'accompagne, d'une perte de résolution spatiale;
- le temps de développement, les difficultés de stockage et de transmission constituent également une limitation.

1.5.2 Mammographie numérique

En mammographie numérique, on remplace le capteur analogique (couple écran-film) par un capteur numérique. La chaîne de la mammographie numérique, est plus complexe que la chaîne analogique. Elle intègre des éléments électroniques susceptibles, de se détériorer au cours du temps. L'image finale est le résultat d'une acquisition et de divers traitements, avec sa visualisation sur un écran d'ordinateur, ou sur un film imprimé par reprographe laser.

Il y a trois types de détecteurs numériques (les systèmes à balayage, les écrans radio luminescents à mémoire et les capteurs numériques plein champ) qui utilisent un mode indirect de conversion des rayons X en charges électriques avec; une première étape de conversion des photons X en photons lumineux ou un mode direct de conversion des rayons X en charges électriques. Avec ces détecteurs, l'image numérique obtenue est différente, de l'image analogique parce que son contraste, son niveau de bruit et son contenu fréquentiel, sont différents ce qui peut augmenter l'influence du bruit anatomique. De plus, ils peuvent être modifiés après l'acquisition de l'exposition. Pour cette raison, elle doit être lue sur des consoles de diagnostic qui pour des raisons pratiques, doivent avoir au minimum 2 écrans.

La détection radiologique des lésions du sein exige des images de grande précision, qui peuvent être obtenues seulement, avec le mammographe (figure 1.3). Le mammographe possède entre autres un dispositif de compression du sein, qui permet à la fois d'obtenir des images de haute résolution et de diminuer les doses de rayonnement utilisées. Le mammographe dispose aussi des accessoires nécessaires, pour réaliser les incidences complémentaires.



a : Le mammographe

b : clichés associés

Figure 1.3 : Mammographie numérique(MN) et les clichés associés [9]

1.5.3 Les incidences en mammographie

Etant donné la complexité de l'anatomie du sein, la mammographie est généralement prise sous différentes directions, appelées incidences. Les incidences (figure 1.4) les plus fréquemment utilisées sont l'incidence de face appelée aussi Cranio-Caudale (CC), l'incidence oblique externe nommée Médio Latérale Oblique(MLO) et l'incidence de profil. Dans le cas de l'incidence de face, le porte cassette (le détecteur) ainsi que le tube radiogène sont horizontaux. Cette incidence explore, bien la région centrale et la région interne du sein. Dans le cas de l'incidence oblique externe, le détecteur est incliné de 40 à 60° en fonction de la morphologie de la patiente de telle façon que, l'orientation des fibres du grand pectoral, soit parallèle à la cassette. L'incidence oblique médio-latérale prend en partie, le muscle pectoral et permet dans ces conditions, d'explorer la quasi-totalité de la glande mammaire sur une seule incidence. Dans le cas de l'incidence de profil, le porte-cassette est vertical et placé entre les deux seins (profil interne). Cette incidence, pareillement à celle de face, explore la région centrale et interne du sein.

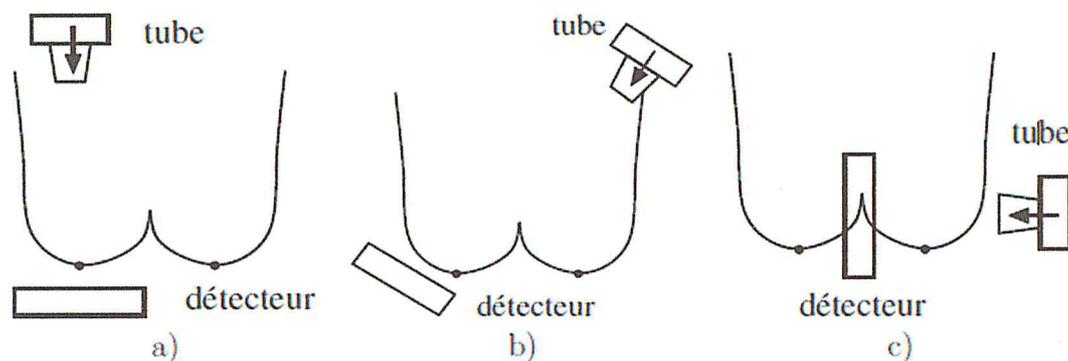


Figure 1.4 : Exemples d'incidences en mammographie [8]
 a) Cranio Caudale (CC), b) Médiolatérale Oblique (MLO) et c) Profil interne.

1.6 Pathologies mammaires [8][9] [10]

Nous limitons la description des pathologies mammaires, uniquement aux kystes, masses et aux calcifications.

1.6.1 Kystes

Si l'un des canaux galactophoriques est obstrué et, qu'un segment du canal ou un lobule n'a plus de sortie, cette cavité peut se remplir avec un liquide. Ce remplissage peut exercer une pression qui gonfle la membrane du canal galactophorique; la cavité devient alors sphérique ou ovale. Selon le type de liquide qui s'est accumulé dans le kyste, l'atténuation de cette lésion est plus élevée que le tissu environnant, (kyste huileux).

Les kystes se traduisent presque toujours, sur les clichés de mammographie, par des opacités circonscrites avec des bords nets. Parfois les cellules du tissu conjonctif qui entourent l'épithélium, sont le siège d'un dépôt d'oxalate ou de phosphate de calcium. Dans ce cas, on parle de kystes calcifiés. Comme l'atténuation du calcium est beaucoup plus élevée, que celle des autres tissus, les kystes calcifiés sont généralement faciles à détecter.

1.6.2. Les calcifications

Les calcifications mammaires sont des dépôts de calcium, qui se forment dans le tissu du sein. Elles n'ont aucun lien avec la quantité de calcium, absorbée au cours de l'alimentation ou obtenue, par l'intermédiaire de compléments alimentaires.

Les calcifications mammaires sont assez courantes et, la plupart ne sont pas associées au cancer. Afin de s'en assurer, le radiologue étudie leur taille, leur forme et leur disposition à l'aide, d'une mammographie sur laquelle, elles apparaissent souvent sous forme de petits

points blancs. Certaines de leurs caractéristiques, comme une forme irrégulière ou certains regroupements, peuvent être suspectes.

Les calcifications se distinguent par leurs caractéristiques spécifiques; parmi lesquelles, on trouve:

- **la taille** des microcalcifications : les grosses microcalcifications sont souvent bénignes et ont un plus fort contraste que les calcifications malignes;
- **la forme** : les microcalcifications rondes ou ovales, ont plus de chances d'être bénignes que les microcalcifications irrégulières, en "pierres écrasées", ou linéaires ou branchées, qui ont plus de risque d'être malignes. Les calcifications trop petites pour être identifiées même sur des mammographies agrandies, dites pulvérulentes ont un risque intermédiaire;
- **l'orientation**: les amas malins, ont tendance à avoir une forme triangulaire ou losangique orientée vers le mamelon;
- **le nombre** : le nombre de microcalcifications est également un paramètre important à prendre en considération. Le cliché centré, agrandi est important pour préciser ce nombre. En effet, certaines microcalcifications, ne sont pas visibles en dehors du cliché centré et agrandi. En pratique courante, il est difficile d'évoquer un foyer s'il n'a pas au moins 4 ou 5 microcalcifications groupées. Le foyer de microcalcifications est d'autant plus suspect, que ces microcalcifications sont nombreuses et concentrées.
- **Le polymorphisme** : si les microcalcifications dans un amas ont des formes variées, l'amas a plus de risque d'être malin.

Ces caractéristiques sont déduites des classifications des microcalcifications. La classification morphologique de Le Gal, a le mérite d'être claire et largement diffusée. Celle-ci inclut 5 types de microcalcifications (figure 1.5).

- Type 1 Microcalcifications annulaires, arciformes ou polyédriques. Risque de cancer du sein quasi nul.
- Type 2 Microcalcifications rondes et de tailles variables. Risque de carcinome évalué à 22%.
- Type 3 Microcalcifications poussiéreuses pulvérulentes. Risque de cancer estimé à 36%.
- Type 4 Microcalcifications irrégulières associées à un risque de cancer estimé à 56 %.
- Type 5 Microcalcifications vermiculaires ou branchées. Risque de carcinome estimé à 91 %.

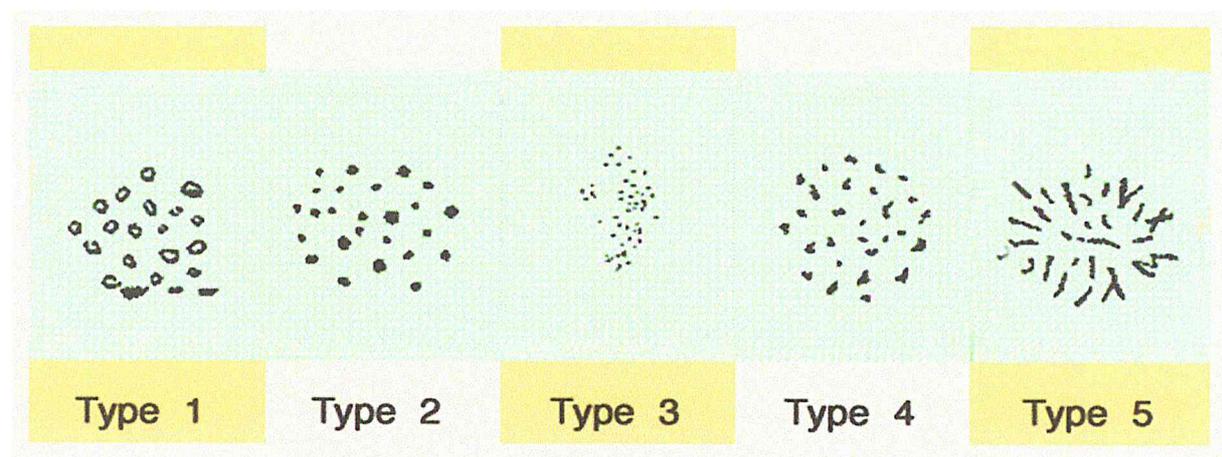


Figure 1. 5 : Types des calcifications [10]

✓ Distribution

La distribution des microcalcifications, est un critère fondamental. Elle présente leur répartition dans le sein et joue, un rôle important dans la prise de décision de la malignité. Les différentes distributions possibles des microcalcifications sont détaillées dans ce qui suit:

- Microcalcifications diffuses/éparses: dans ce cas, les microcalcifications sont distribuées de façon aléatoire dans l'ensemble du sein.
- Micro calcifications groupées, en amas ou en cluster: ces termes sont utilisés lorsque de multiples microcalcifications (au moins cinq) occupent un petit volume tissulaire.
- Distribution linéaire: les microcalcifications sont disposées les unes derrière les autres sous forme d'une ligne. Il s'agit généralement de dépôts calciques dans un galactophore.
- Distribution segmentaire: elle suggère des dépôts calciques dans des canaux galactophores ainsi que leurs branches; ce qui évoque la possibilité d'un cancer mammaire étendu.

1.6.3 Les masses

Une masse, est définie comme une lésion occupant dans l'espace une position, qui peut être vue, sous deux incidences différentes. Néanmoins, la masse peut être difficilement perceptible en mammographie, car elle est similaire aux tissus avoisinants. Pour cette raison, sa détection est considérée comme une tâche difficile. La caractérisation d'une masse est basée sur sa forme (ronde, ovale, lobulée, irrégulière), sur ses contours (circonscrits, micro-lobulés, masqués, indistincts, spiculés) et sur sa densité (haute, moyenne, faible, grasseuse).

Etant plus denses que le tissu mammaire adjacent, les masses apparaissent blanches sur les mammogrammes (figure 1.6).

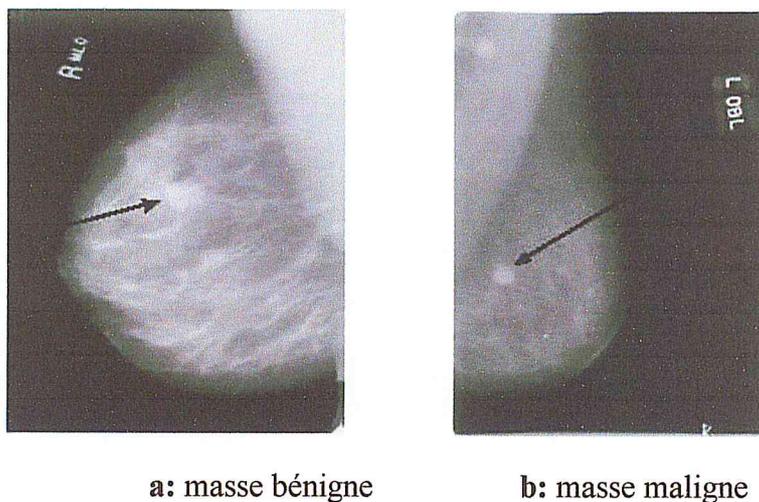


Figure 1.6 : Masses mammographiques [7]

Les masses sont caractérisées par: leur taille, leur forme, leurs contours et leur densité.

- ✓ **La taille :** elle est variable de quelques millimètres, à plusieurs centimètres. Cependant la taille ne prédit pas le caractère malin, sauf sur des clichés successifs lorsque l'on voit la taille augmenter régulièrement.
- ✓ **La forme :** selon la description du BIRADS [11], les masses mammaires peuvent avoir la forme ronde, ovale, lobulée ou irrégulière.
 - **Ronde :** il s'agit de masse sphérique, circulaire ou globuleuse.
 - **Ovale :** elle présente une forme elliptique (ou en forme d'œuf).
 - **Lobulée :** la forme de la masse présente une légère ondulation.
 - **Irrégulière :** cette appellation est réservée aux masses dont la forme est aléatoire et ne peut être, caractérisée par les termes cités ci-dessus.

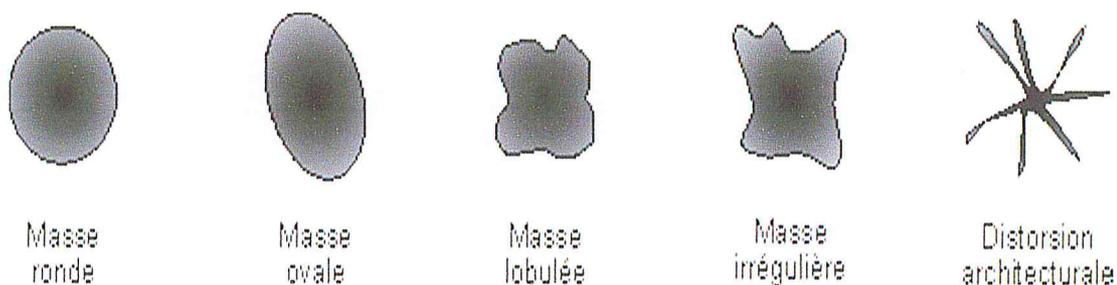


Figure 1.7 : Catégories de masses [7]

✓ Le contour

Le contour des masses mammaires est soit circonscrit, soit micro-lobulé, soit masqué soit indistinct, soit spiculé. On détaille dans ce qui suit ces différentes notions.

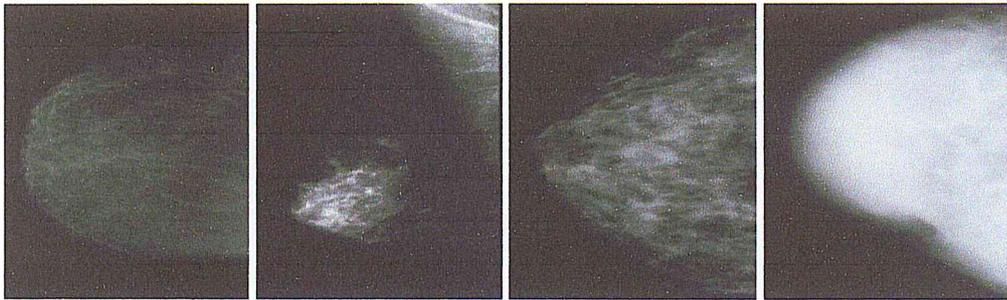
- Circonscrit: il s'agit d'une transition brusque entre la lésion et le tissu environnant. Le contour est alors net et bien défini. Pour qu'une masse soit qualifiée de circonscrite, il faut qu'au moins 75% de son contour soit nettement délimité.
- Micro-lobulé: dans ce cas, de courtes dentelures du contour créent de petites ondulations.
- Masqué: un contour masqué est un contour qui est caché par le tissu normal adjacent. Ce terme est employé pour caractériser une masse circonscrite dont une partie du contour est cachée.
- Indistinct: dans ce cas, le contour est mal défini. Ce caractère indistinct, peut correspondre à une infiltration.
- Spiculé: la masse est caractérisée par des lignes radiaires en prenant naissance, sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.

✓ La densité

L'aspect du sein normal, est très variable d'une femme à l'autre. Le facteur le plus remarquable est la grande variabilité de la densité radiologique de l'aire mammaire. Wolfe [22] est le premier qui a établi une relation entre la densité du tissu mammaire (figure 1.8) et le risque de développer, un cancer.

Afin de standardiser les comptes rendus mammographiques, la classification BIRADS de l'ACR définit 4 classes de la composition du sein.

- Stade 1: le sein est presque entièrement graisseux et homogène, radio transparent et facile à lire (moins de 25 % de la glande mammaire).
- Stade 2: il y a des opacités fibroglandulaires dispersées. Le sein est graisseux et hétérogène (approximativement 25 à 50 % de la glande mammaire).
- Stade 3: le tissu mammaire est dense et hétérogène (approximativement 51 à 75 % de la glande mammaire).
- Stade 4: le tissu mammaire est extrêmement dense et homogène. La mammographie est alors difficile à interpréter puisque la densité peut masquer une lésion (plus de 75 % de la glande mammaire).



Radio-transparent

Faible densité

Densité moyenne

Forte densité

Figure 1.8: Différentes densités mammaires [7]

1.8 Conclusion

Dans ce chapitre, nous avons introduit quelques notions générales concernant, l'imagerie mammaire. Nous avons ensuite parcouru les principales pathologies qui peuvent atteindre les tissus du sein, en passant par l'examen de base qu'est la mammographie. En imagerie mammographique, un système d'aide au diagnostic désigne tout système permettant d'apporter une aide au radiologue lors de son interprétation d'images des systèmes d'aide au diagnostic. L'objectif de tels systèmes consiste en l'amélioration des résultats de détection en attirant l'attention des experts sur des zones suspectes. Cette aide est devenue indispensable et peut être considérée comme une " seconde lecture " de mammographie.

Le chapitre suivant, introduit les techniques de prétraitement et de segmentation de l'image, adoptées dans le cas de notre étude.



CHAPITRE 2

*PRÉTRAITEMENT ET SEGMENTATION
EN IMAGERIE MAMMOGRAPHIQUE*

Chapitre 2 Prétraitement et segmentation en imagerie mammographique

2.1 Introduction

Dans la vision par ordinateur, on distingue les traitements de bas niveau et les traitements de haut niveau. Cette distinction est liée, au contenu sémantique des entités traitées et extraites de l'image. Les traitements de bas niveau opèrent, sur les grandeurs calculées à partir des valeurs attachées, à chaque point de l'image. A l'opposé, les traitements de haut niveau, s'appliquent à des entités de nature symbolique associées, à une représentation de la réalité extraite de l'image.

La segmentation est l'opération fondamentale de l'analyse d'image [12]. Elle consiste à partitionner l'image en un ensemble de régions connexes. Ces régions sont exploitées, pour extraire des paramètres de forme, de position et de, taille afin de réaliser l'étape de reconnaissance des entités caractérisées. Cependant, l'image médicale nécessite, une opération de prétraitement en amont de la segmentation, pour améliorer sa qualité.

Dans ce qui suit, nous allons étudier les techniques de prétraitement adoptées dans le cadre de notre initiation à la recherche, pour le rehaussement du contraste de certaines structures pathologiques, dans l'image mammographique, afin de faciliter la tâche de la segmentation. Nous parcourons aussi, les approches de la segmentation, qui se basent, sur les méthodes par clustering flou et les modèles par les ensembles de niveaux.

2.2 Système d'aide à la détection automatique [13]

Un système de détection automatique dédié à l'analyse d'images mammographiques (figure 2.1), suit trois phases essentielles, avant d'atteindre la prise de décision. Les étapes de détection automatique de pathologies mammographiques, se résument selon:

- un prétraitement, qui sert à améliorer la qualité de l'image avant toute manipulation.
- Une segmentation, qui permet de détecter les lésions étudiées.
- Une description, qui a pour but de caractériser les lésions à travers des attributs morphologiques et texturaux.
- Une classification qui permet la prise de décision en utilisant un classifieur adéquat.

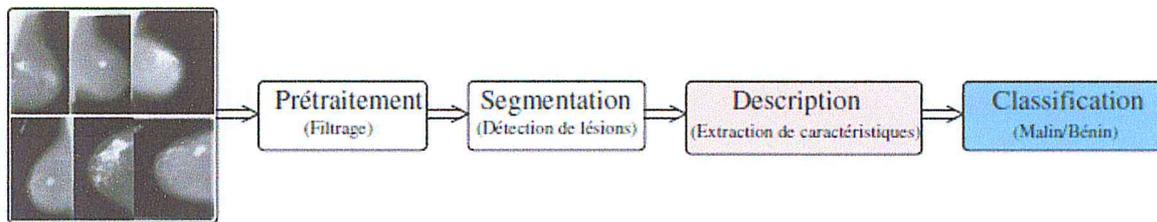


Figure 2.1: Système de détection automatique de pathologies mammographiques [13]

2.3 Prétraitement en imagerie mammographique

Les structures que l'on recherche, n'étant pas toujours facilement discernables, une étape de prétraitement destinée à les mettre en évidence, peut faciliter leur extraction. Dans le cas de l'image mammographique, il est souvent difficile de détecter certaines lésions, du fait de la variabilité des tissus mammaires. Dans ce cadre, on a recours à un prétraitement spécifique du cliché mammographique, pour rehausser son contraste.

Les algorithmes utilisés sont basés sur une succession de techniques mathématiques dont la principale fonction, est d'améliorer la qualité de l'image.

Le problème majeur des algorithmes de rehaussement du contraste réside dans le fait que certaines régions peuvent ne pas être rehaussées convenablement alors que d'autres peuvent être l'objet d'un excès de rehaussement. Un manque de rehaussement du contraste peut causer des faux négatifs (FN). En effet, plusieurs détails concernant la lésion peuvent être négligés. Dans ce cas, certaines lésions peuvent ne pas être détectées et par la suite non diagnostiquées.

Ce qui ne répond pas à l'objectif principal de la détection précoce d'un cancer. Un excès de renforcement du contraste peut causer des faux positifs (FP). Dans ce cas, plusieurs détails inexistantes réellement peuvent s'ajouter à la lésion. D'où, certaines régions normales du tissu mammaire peuvent être considérées comme lésions, ce qui va engendrer des biopsies inutiles.

Une approche couramment utilisée pour l'amélioration du contraste est la modification globale de l'histogramme [14]. Cette méthode consiste à réaffecter les valeurs d'intensité des pixels afin de rendre la nouvelle répartition des intensités plus uniforme. Ceci peut être réalisé par l'égalisation d'histogramme ou par l'étirement d'histogramme. Toutefois, ces transformations ont l'inconvénient de faire ressortir le bruit dans les images déjà bruitées.

Dans ce contexte, une alternative proposée en tant que transformation locale pour l'amélioration du contraste est basée, sur les opérateurs de la morphologie mathématique.

2.4 Introduction à la morphologie mathématique [15][16]

La morphologie mathématique est une science de la forme et de la structure. Le principe de base de cette méthode est de comparer un objet qu'on analyse à un autre objet de référence, connu par sa taille et sa forme, appelé élément structurant.

La morphologie mathématique offre un grand nombre d'outils très puissants de traitement et d'analyse d'image. On appelle morphologie mathématique *ensembliste*, quand le traitement se porte sur des images binaires et, quand l'utilisation se fait sur des images en niveaux de gris, on fait référence à la morphologie mathématique *fonctionnelle*.

Les opérations de base, sont l'érosion et la dilatation. Celles-ci, sont à la base des autres opérateurs plus complexes tels que l'ouverture, la fermeture et, le chapeau haut de forme.

2.4.1 Opérateurs de base et transformations premières de la morphologie mathématique

a. Élément structurant

Un élément structurant B est un ensemble qui présente, une forme géométrique (carrée, circulaire, linéique...) et une taille spécifique.

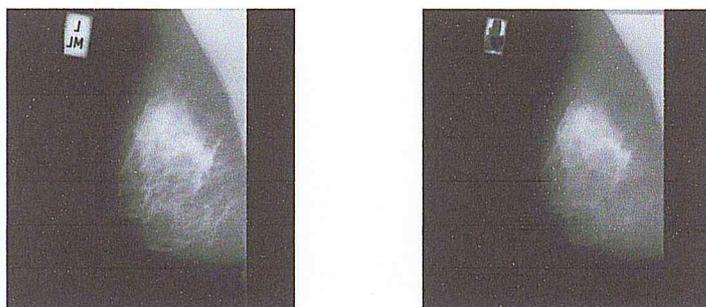
b. Erosion

Soit X l'ensemble à éroder, et B un élément structurant de géométrie simple, par exemple un cercle. L'érosion consiste à faire déplacer B de telle sorte que son centre occupe toutes les positions x de l'espace. Pour chaque position, on pose la question : est-ce que B est complètement inclus dans X ?

L'ensemble des positions x correspondant à une réponse positive, forme le nouvel ensemble Y , appelé érodé de X par B , noté $X \ominus B$ (figure 2.2).

Cet ensemble satisfait l'équation suivante :

$$X \ominus B = \{x \text{ dans } X : B \subset X\} \quad 2.1$$



a) Image mammographique initiale b) Image mammographique érodée

Figure 2.2: Érosion d'une image mammographique

Le traitement obtenu par l'érosion indique les points suivants:

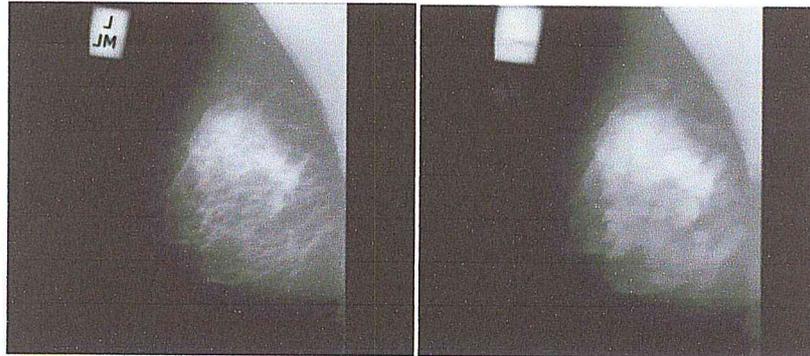
- Les objets de taille inférieure à celle de l'élément structurant vont disparaître.
- Les autres seront "amputés" d'une partie correspondant à la taille de l'élément structurant.
- S'il existe des trous dans les objets, c'est à dire des "morceaux" de fond à l'intérieur des objets, ils seront accentués.
- Les objets reliés entre eux par un élément de taille inférieure, à celle de l'élément structurant, vont être séparés.

c. Dilatation

L'opération de dilatation se définit de manière analogue à l'érosion. En prenant le même élément structurant B , on pose pour chaque point x la question "Bx touche-t-il l'ensemble X ?", c'est à dire, y a-t-il une intersection non vide entre Bx et X ?

L'ensemble des points de l'image (Figure 2.3) correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X, noté comme suit :

$$y = x \oplus B = \{ \text{dans } 1 : B \cap X \neq \emptyset \} \quad 2.2$$



a) Image mammographique initiale b) Image mammographique dilatée

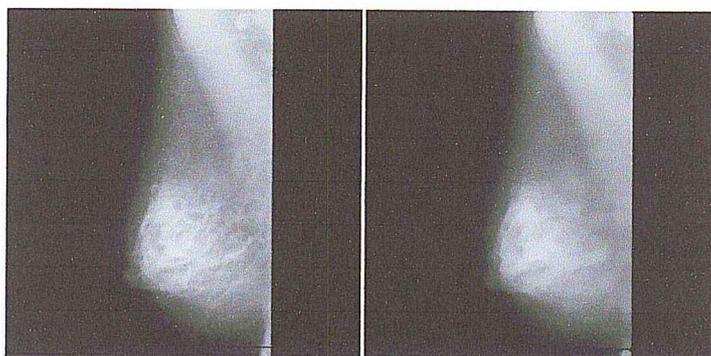
Figure 2.3: Dilatation d'une image mammographique

d. Ouverture

Puisque l'érosion et la dilatation sont des transformations itératives, il est possible d'effectuer sur un ensemble X une érosion, puis de dilater l'ensemble érodé par un même élément structurant B. Le résultat est nommé l'ouvert de X par B, noté comme suit :

$$OB(X) = (X \ominus B) \oplus B \quad 2.3$$

Après une ouverture, on ne retrouve pas l'ensemble de départ; L'ensemble ouvert est plus régulier et moins riche en détails que l'ensemble X initial (figure 2.4). La transformation par ouverture adoucit les contours, coupe les isthmes étroits, supprime les petites îles et les caps étroits.



a) Image mammographique initiale b) Image mammographique avec ouverture

Figure 2.4 : Ouverture d'une image mammographique par un élément structurant circulaire

L'opération d'ouverture de l'image selon la figure ci-dessus, permet:

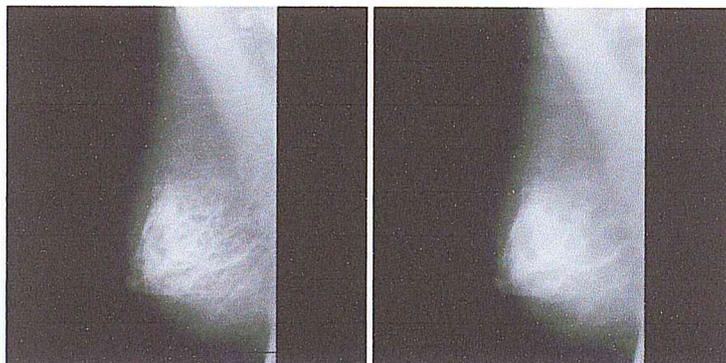
- de lisser les formes,
- d'éliminer les composantes connexes plus petites que B,
- de conserver souvent la taille et la forme.

e. Fermeture

La fermeture est l'opération "inverse" de l'ouverture, c'est à dire que l'on applique tout d'abord une dilatation puis une érosion (toujours en gardant le même élément structurant). Le résultat est nommé le fermé de X par B, noté comme suit :

$$F^B(X) = (X \oplus B) \ominus B \quad 2.4$$

Un ensemble fermé, est également moins riche en détails que l'ensemble initial (figure 2.5). La transformation par fermeture, bouche les canaux étroits, supprime les petits lacs et les golfes étroits.



a) Image mammographique initiale b) Image mammographique avec fermeture

Figure 2.5 : Fermeture d'une image mammographique par un élément structurant circulaire

L'opération de fermeture suivant le résultat de la figure ci-dessus permet de:

- boucher les trous plus petits que B,
- conserver la taille et la forme des régions traitées.

f. Chapeau haut de forme et chapeau bas de forme

On appelle *transformation chapeau haut-de-forme* (ou *top hat*) la différence entre l'image originale et l'ouvert de cette image, ou entre le fermé de l'image et l'image originale. Le premier type de chapeau haut-de-forme est appelé *chapeau haut-de-forme blanc*, car il permet

de détecter ce que l'ouverture a fait disparaître, c'est-à-dire les pics ou parties claires de l'image originale. Le second type du chapeau haut-de-forme, dual du premier, est appelé *chapeau haut-de-forme noir*, car il détecte les vallées, ou parties sombres d'une image.

Le chapeau haut de forme noir est représenté par l'équation suivante :

$$WTH(x) = f(x) - f_{Bn}(x) \quad 2.5$$

Le chapeau haut de forme blanc :

$$BTH(x) = f_{Bn}(x) - f(x) \quad 2.6$$

Où B est l'élément structurant, n représente la taille de l'élément structurant.

Dans notre cas, nous adoptons, un élément structurant de la forme d'un disque, pour rehausser les masses. L'organigramme de la figure 2.6, illustre les étapes que nous avons élaborées, afin de rehausser le contraste des masses mammographiques.

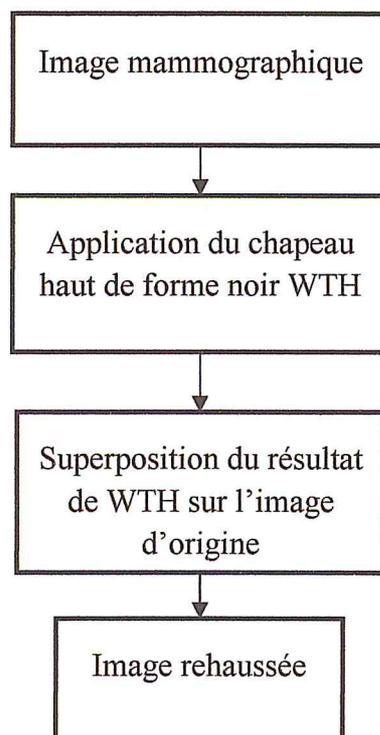


Figure 2.6: a) organigramme de prétraitement des pathologies mammographiques

2.5 Segmentation en imagerie mammographique

La segmentation est un traitement de bas niveau, qui consiste à créer une partition de l'image en sous-ensembles appelés régions. La segmentation est alors obtenue par extraction des composantes connexes des pixels appartenant à la même classe. Ainsi une région sera constituée de pixels, d'une même classe. Mais, il peut aussi y avoir dans l'image plusieurs régions correspondant à une même classe [12].

➤ De façon plus précise, on peut définir la segmentation comme étant une partition d'une image I en n ensembles R_i appelés régions tels que :

1. $\bigcup_{i=1}^n R_i = 1$
2. $R_i \cap R_j = \emptyset$
3. Les sous-ensembles $R_i ; i=1, \dots, n$ sont connexes.
4. Il existe un prédicat P tel que: $P(R_i) = \text{vrai}, \forall i = 1, \dots, n$
5. $P(R_i \cup R_j) = \text{Faux}, \forall (i, j); i \neq j$ et R_i, R_j sont contiguës

Où P désigne un prédicat défini sur l'ensemble, des points de R_i et \emptyset un ensemble vide.

La segmentation est une étape importante pour l'extraction des informations qualitatives de l'image. Elle fournit une description de haut niveau, chaque région est connectée à ses voisines dans un graphe et chaque région porte une étiquette donnant des informations qualitatives comme sa taille, sa couleur, sa forme, son orientation. L'image se réduit donc à un graphe de nœuds étiquetés, qui contient presque toute l'information utile au système.

Fondamentalement, la segmentation est un processus qui consiste à découper une image en régions connexes présentant une homogénéité selon un certain critère. L'union de ces régions doit redonner l'image initiale [12]. Dès lors, la segmentation sera valide si les conditions suivantes sont vérifiées:

- tous les composants de l'image doivent être classés, autrement dit chaque pixel doit appartenir à une classe;
- les régions doivent être homogènes;
- les différentes régions doivent être séparées (disjointes).

L'objectif de la détection de contours, est la détermination des points de l'image séparant des zones de caractéristiques différentes. Chaque méthode de détection de contours, est justifiée

de manière plus ou moins explicite, par la définition du type de transitions recherchées [14]. De manière duale, le but de la segmentation en régions, est de partitionner l'image en ensembles de points connexes possédant des propriétés d'homogénéité. Bien qu'il existe une multitude d'algorithmes de segmentation selon le domaine et les contraintes étudiées, il en découle quatre grandes approches [12] qui peuvent se distinguer par:

1. La segmentation fondée sur les régions.
2. La segmentation fondée sur les contours.
3. La segmentation fondée sur la classification ou le seuillage des pixels.
4. La segmentation fondée sur la coopération entre les trois premières segmentations.

La segmentation est invoquée dans la plupart des applications médicales telles que le diagnostic, la localisation des pathologies, la visualisation des structures anatomiques et la chirurgie assistée par ordinateur, etc. Elle prouve également son intérêt dans le domaine de la biométrie pour l'identification des individus où elle est appelée à mettre en valeur, les caractéristiques de la signature biométrique utilisées. De nombreuses méthodes ont été publiées dans la littérature [55]. Cependant, le choix d'une technique reste un problème posé et dépend fortement de l'application envisagée ainsi que du type d'image à traiter. Nous décrivons ici les principales approches, utilisées dans le contexte de notre projet, les modèles déformables et les méthodes par classification.

2.5.1 Les modèles déformables

Un modèle déformable [17] est une courbe ou surface, fermée ou non, qui évolue depuis une position initiale, vers les contours de l'objet à segmenter en minimisant une fonction d'énergie qui se compose de deux termes d'énergie:

une énergie interne qui permet de mesurer l'adéquation avec la forme et une énergie externe, qui mesure l'adéquation avec l'apparence.

Ce modèle consiste à définir un contour fermé et à le faire évoluer, vers l'objet d'intérêt. Les déformations de cette courbe sont déduites de la dérivation d'une fonctionnelle à optimiser, caractérisant l'objet. Les travaux sur les contours actifs se divisent en deux variantes, qui sont des méthodes uniquement basées sur des informations contour, ou bien celles intégrant des caractéristiques sur les régions et donc des informations plus globales sur les objets à segmenter d'une image.

Les algorithmes de segmentation fondés sur les modèles déformables ont l'avantage, par rapport aux méthodes dérivatives de fournir des contours fermés. Parmi ces méthodes, on trouve les contours actifs et les ensembles de niveaux.

1. Les contours actifs (Snakes) [17][18][19]

Le modèle général des contours actifs, se présente sous la forme d'une courbe dont l'initialisation est située à proximité du contour recherché. L'évolution du modèle s'effectue ainsi, selon un processus itératif de déformation contrôlé par un test de convergence. Cette évolution est réalisée sous l'action de deux forces: une force interne issue de la courbe elle-même permettant de gérer la régularité du contour actif et; une force externe calculée à partir des données de l'image.

La segmentation d'images à base de contours actifs constitue l'une des méthodes les plus utilisées en segmentation itérative d'images dans, de nombreux domaines d'application tels que:

- La reconnaissance de formes;
- La simulation;
- Le suivi de scènes;
- La localisation d'une forme précise dans une image complexe;
- La segmentation d'images optiques ou médicales.

Le principe de cette méthode, pour la segmentation de la région d'intérêt, consiste à initialiser un contour grossièrement autour de l'objet à détecter et, à le faire déplacer par un processus d'itération afin d'épouser au mieux, les contours de la région recherchée. Le processus d'itération est appliqué de façon à conserver certaines caractéristiques; telles que la courbure, la répartition des points, ou d'autres contraintes liées à la disposition des points. Ceci est explicité par un exemple simple, donné par la figure 2.7.



Figure 2.7: Exemple simple d'application des contours actifs sur une image mammographique [20]

La segmentation par cette approche, a montré des avantages remarquables en termes de structure (courbe fermée), de qualité des contours obtenus et, de rapidité de convergence de l'algorithme. Lorsque l'initialisation est proche du contour de la région d'intérêt, et comme toute méthode de détection de contours, elle a ses limites.

H. Cheng et al [20], ont obtenu des résultats intéressants pour le cas d'images mammographiques, dont le contraste est faible. Néanmoins, ces méthodes présentent des inconvénients:

- il s'agit de méthodes semi-automatiques, où le contour initial des courbes doit être fixé par l'utilisateur;
- si le contour actif est initialisé loin de la position finale attendue, la convergence n'est pas garantie et, le temps de calcul devient assez important.

2. Les ensembles de niveaux (levels sets) [21][22]

La méthode des ensembles de niveaux est une méthode de simulation numérique utilisée pour l'évolution des courbes et; des surfaces dans les domaines discrets. C'est une représentation variationnelle contours qui évoluent dans le temps, mais différemment des contours actifs. En effet, les ensembles de niveaux peuvent changer de topologie si les contours l'imposent. Par exemple, changer de connexité (un contour simple peut par exemple donner naissance à plusieurs contours fils qui vont évoluer séparément).

Pour permettre son évolution ultérieure, la fonction ψ dépend du temps:

$$\psi : D \times \mathbb{R}^2 \rightarrow \mathbb{R} \quad (x, t) \rightarrow \psi(x, t)$$

A un instant donné t , la surface Γ est définie comme l'ensemble des points x vérifiant la condition :

$$\Gamma = \{x \mid \psi(x, t) = 0\} \quad 2.7$$

Un choix habituel pour l'initialisation de la fonction $\psi(x)$ est la distance euclidienne signée, qui sépare, x du contour Γ . Le signe dépend de la position du point x à l'intérieur ou, à l'extérieur de Γ . Sur la figure 2.10, la région R_{in} est l'ensemble des points pour lesquels ψ est positive :

$$R_{in} = \{x \mid \psi(x, t) \geq 0\} \quad 2.8$$

La fonction des ensembles de niveaux peut-être initialisée, à partir de la distance euclidienne comme suit :

$$\psi \begin{cases} 0 & \text{si } x \in \Gamma \\ \inf_{y \in D} \|x - y\| & \text{si } x \in R_{in} \\ \inf_{y \in D} \|x - y\| & \text{si } x \text{ n'appartient pas à } R_{in} \end{cases} \quad 2.9$$

Les variables qui vont être modifiées par la méthode d'évolution, sont les niveaux de la fonction ψ en chaque point de D .

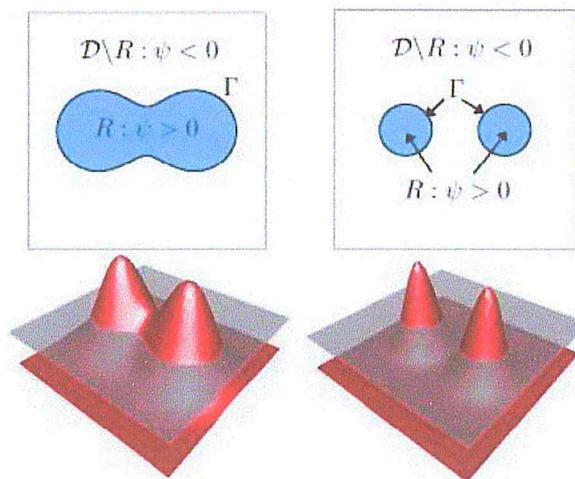


Figure 2.8 : Fonction d'ensembles de niveaux [22]

➤ **Modèle de Chan et Vese [23]**

Soit I , l'image à segmenter en deux régions. Cette image est définie dans un domaine spatial telle que $I = \{I(x, y), (x, y) \in \Omega\}$ et soit $\tau(t)$, la courbe d'évolution que définit le contour des objets. La courbe initiale nécessaire au déroulement de l'algorithme est souvent un cercle de rayon quelconque et centré en un point quelconque de l'image. La courbe $\tau(t)$ divise le domaine Ω en deux régions, Ω_0 et Ω_1 qui sont respectivement les régions intérieure et extérieure à $\tau(t)$. Cette courbe $\tau(t)$ est représentée implicitement par la courbe de niveau zéro d'une fonction auxiliaire $\Phi(x, y, t)$ appelé fonction level set définie telle que :

$$\begin{cases} \Phi(x, y, t) > 0 & \text{pour } (x, y) \in \Omega_1 \\ \Phi(x, y, t) = 0 & \text{pour } (x, y) \in \tau \\ \Phi(x, y, t) < 0 & \text{pour } (x, y) \in \Omega_0 \end{cases} \quad 2.10$$

La fonction $\Phi(x, y, t)$ est habituellement prise comme la distance signée du point (x, y) à l'interface $\tau(t)$.

A la fonction level set, est associée une fonction Heaviside $H(\Phi)$ qui prend la valeur 1 pour $\Phi \geq 0$ et vaut 0 ailleurs. Pour chaque région Ω_k ($k = 0$ ou 1) en liaison avec cette fonction Heaviside, une fonction d'appartenance g_k est définie telle que $g_0 = 1 - H$ et $g_1 = H$. Chaque région ou plus exactement chaque classe Ω_k est caractérisée par une fonction lisse I_k qui est l'intensité moyenne de la région exprimée par l'expression suivante :

$$I_k = \frac{\int_{\Omega_k} I(x, y) dx dy}{\int_{\Omega_k} dx dy} = \frac{\int_{\Omega} I(x, y) g_k(x, y) dx dy}{\int_{\Omega} g_k(x, y) dx dy}, \text{ avec } k \in \{0, 1\} \quad 2.11$$

Dans le cas d'une image de taille $M * N$ pixels, l'expression numérique de l'intensité moyenne de la région Φ_k est :

$$I_k = \frac{\sum_{i=1}^M \sum_{j=1}^N I(i, j) g_k(i, j)}{\sum_{i=1}^M \sum_{j=1}^N g_k(i, j)}, \text{ avec } k \in \{0, 1\} \quad 2.12$$

La segmentation par le modèle de Chan et Vese, consiste à partitionner l'image en régions d'intensités homogènes en faisant évoluer la courbe $\tau(t)$, donc la fonction $\Phi(x, y, t)$, jusqu'à ce qu'il ait coïncidence avec les frontières des régions.

L'évolution de la courbe $\tau(t)$ est réalisée par la minimisation de la fonction d'énergie exprimée par:

$$E(\Phi, I_k) = \sum_{k=0}^1 \lambda_k [I(x, y) - I_k]^2 g_k(x, y) dx dy + \mu \int_{\Omega} |\nabla H(\Phi)| dx dy \quad 2.13$$

Le second terme de cette expression est la longueur de la courbe τ pondérée par le paramètre positif μ fixé, par l'utilisateur. Les λ_k sont des constantes positives souvent prises à 1.

Pour solutionner le problème de minimisation, on utilise l'équation correspondante d'Euler-Lagrange. Cela conduit à une équation d'évolution, qui s'exprime directement à l'aide de la fonction de level set Φ :

$$\frac{\partial \Phi}{\partial t} = \delta_{\varepsilon}(\Phi) \left[\mu \operatorname{div} \left(\frac{\nabla \Phi}{|\nabla \Phi|} \right) - \sum_{k=0}^1 \lambda_k (I - I_k)^2 \frac{\partial g_k}{\partial H} \right] \quad 2.14$$

δ_ϵ est une version régularisée de la fonction de Dirac, qui est la dérivée de la fonction de Heaviside normalisée. Ces deux fonctions sont données par les expressions suivantes:

$$H_\epsilon(\Phi) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \left(\frac{\Phi}{\epsilon} \right) \right) \quad 2.15$$

$$\delta_\epsilon(\Phi) = H'_\epsilon(\Phi) = \frac{1}{\pi\epsilon} \left(\frac{1}{1 + \left(\frac{\Phi}{\epsilon} \right)^2} \right) \quad 2.16$$

Lorsque les constantes λ_k sont prises à 1, l'expression (2.14) devient :

$$\Delta\Phi = \Delta t \cdot \delta_\epsilon(\Phi) \left[\mu \operatorname{div} \left(\frac{\nabla\Phi}{|\nabla\Phi|} \right) + (I - I_0)^2 - (I - I_1)^2 \right] \quad 2.17$$

L'équation (2.13) est résolue en procédant, par une méthode numérique itérative. A chaque itération les valeurs de I_k et g_k sont mises à jour jusqu'à atteindre le nombre d'itérations fixé par l'utilisateur. La courbe τ finale obtenue, représente le contour des objets contenus dans l'image.

A la fin de la segmentation, l'image résultante est désigné par:

$$U(x, y) = \sum_k I_k g_k(x, y) \quad 2.18$$

✓ Les avantages et les inconvénients des ensembles de niveaux

La méthode des ensembles de niveaux présente de nombreux avantages [28]

❖ les avantages

1. Adaptation au changement de topologie: plusieurs objets peuvent être segmentés à la fois.
2. La formulation n'est pas limitée aux courbes en deux dimensions. Elle peut être facilement étendue, aux dimensions plus élevées.
3. La prise en compte des changements de topologie automatique.
4. Les propriétés géométriques de la courbe, telles que la courbure ou le vecteur normal qui sont importantes pour les contours actifs, peuvent être simplement déterminés.

❖ Les inconvénients

1. Le temps de calcul est important.
2. L'implémentation implique plusieurs problèmes.

3. Le résultat dépend de l'initialisation.

2.5.2 La segmentation par approche mixte

La dualité entre les régions et les contours peut être exprimée de quatre manières différentes [25].

- ❖ Les régions sont situées à l'intérieur des contours. Par conséquent, il n'y a pas de points contours à l'intérieur d'une région ;
- ❖ Un point de contours est situé sur ou du moins à proximité (distance à définir) d'une frontière de région;
- ❖ Une frontière de région est fermée par nature, un contour doit l'être aussi
- ❖ Un contour doit être situé sur la totalité de la frontière commune entre deux régions.

Cette dualité a donné naissance à la coopération dans la segmentation d'image. Selon la manière de faire coopérer deux processus de segmentation région et contour, on peut distinguer trois approches différentes: La coopération séquentielle, la coopération par fusion de résultats et la coopération mutuelle.

- ❖ La coopération séquentielle : une des techniques de segmentation (région ou contour) est exécutée, son résultat est exploité par la deuxième technique pour améliorer les critères ou les paramètres de la segmentation.
- ❖ La coopération des résultats : les deux types de segmentation sont exécutés indépendamment. La coopération se fait sur les résultats obtenus afin d'atteindre une meilleure segmentation.
- ❖ La coopération mutuelle : les deux types de segmentation coopèrent mutuellement au cours de leur processus d'exécution.

2.5.3 Segmentation par Classification

1. Les méthodes de segmentation par partitionnement [26][27]

La classification par partitionnement est un processus itératif, chargé de partitionner l'ensemble des pixels de l'image N , en C classes. Dans le processus de classification, les éléments sont affectés à la classe la plus proche. Les représentations des classes sont

régulièrement mises à jour, selon un critère de minimisation. Ces méthodes de classification nécessitent trois prés-requis:

- chaque objet (pixel) doit être représenté par un vecteur d'attributs,
- un critère pour classifier les objets dans des classes est fixé, une distance entre éléments est définie,
- un critère qui représente la stratégie de fusion ou de dissociation des classes, est donné.

Parmi les méthodes par partitionnement les plus connues, on trouve la méthode des k-moyennes (k-means) et l'algorithme des k plus proches voisines (KPPV).

a. KPPV

C'est une approche très simple et directe. Elle ne nécessite pas d'apprentissage mais simplement le stockage des données d'apprentissage. Son principe est le suivant:

- une donnée de classe inconnue, est comparée à toutes les données stockées. On choisit pour la nouvelle donnée la classe majoritaire parmi ses K plus proches voisins (Elle peut donc être lourde pour des grandes bases de données) au sens d'une distance choisie.

Afin de trouver les K plus proches d'une donnée à classer, on peut choisir la distance euclidienne.

Soient deux données représentées par deux vecteurs x_i et x_j , la distance entre ces deux données est donnée par:

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad 2.19$$

b. K-means

L'algorithme des centres mobiles (k-means) est la technique de classification la plus simple. Cette technique utilise comme critère d'évaluation d'une partition, la distance quadratique moyenne. Son principe considère un ensemble de points que nous essayons, de partitionner en k classes. Nous posons:

- $x_j^{(i)}$: le pixel j appartenant à la classe i,
- y_i : le centroïde de la classe i,
- L_i : le nombre de points de la classe i.

- $D(x_j^{(i)}, y_i)$: la distance entre $x_j^{(i)}$ et y_i ; dans le cas de l'algorithme k-means, c'est la distance Euclidienne.
- Di : la distance totale de la classe i : $Di = \sum_{j=1}^{Li} d(x_j^{(i)}, y_i)$
- D est la distorsion globale pour l'ensemble des vecteurs.

$$D = \sum_{i=1}^k Di \quad 2.20$$

La classification optimale, est celle qui permet de minimiser la distorsion totale D donnée par l'équation précédente. La procédure d'optimisation, doit tenir compte des hypothèses suivantes:

- pour un ensemble donné de centroïdes, la classification qui minimise Di est celle, pour laquelle chaque pixel est affecté, à la classe dont le centroïde est le plus proche.
- Pour chaque classe i , on considère le pixel y_i qui minimise la distance totale de classe Di .

L'algorithme de k-means est le suivant :

1. Choisir aléatoirement k centres de classes.
 2. Affecter les individus aux classes dont la distance est minimale.
 3. Calculer les nouveaux centres de gravité des classes.
 4. Affecter chaque individu à la nouvelle classe.
 5. Tant que le critère d'arrêt n'est pas vérifié, aller à l'étape 3.
- L'algorithme s'arrête, lorsque les centres de gravité, sont fixes.

2. Les méthodes de segmentation par classification floue [26][28]

Une classe est un groupe d'individus présentant des similitudes communes. D'une part, les similitudes sont plus au moins fortes entre les individus d'une même classe. D'autre part, un même individu peut présenter des similitudes avec les individus d'autres classes. Le degré d'appartenance d'un individu n'est pas localisé à une classe déterminée mais se trouve distribuée sur plusieurs classes. Nous retrouvons là le concept même de l'ensemble flou, dans ce formalisme, un élément peut appartenir plus au moins fortement à plusieurs ensembles flous. Parmi les méthodes de classification floues, l'algorithme fuzzy C-means (FCM) est le plus utilisé dans la segmentation d'image. Durant les itérations, le degré d'appartenance flou est mis à jour comme une fonction de distance par rapport aux classes.

➤ **Méthode de Fuzzy C-Means (FCM)**

Fuzzy C-Means est un algorithme de classification non supervisée floue. Il introduit la notion d'ensemble flou, dans la définition de classes: chaque point dans l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leurs centres de gravité. Il utilise un critère de minimisation des distances intra-classe et de maximisation des distances inter-classes, mais en donnant un certain degré d'appartenance à chaque classe, pour chaque pixel. Cet algorithme nécessite une connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective. Ainsi, il permet d'obtenir une partition floue de l'image, en donnant à chaque pixel un degré d'appartenance à une classe donnée. Le cluster auquel est associé un pixel est celui dont le degré d'appartenance sera le plus élevé.

Les principales étapes de l'algorithme Fuzzy C-means sont:

1. La fixation arbitraire d'une matrice d'appartenance.
2. Le calcul des centroïdes des classes.
3. Le réajustement de la matrice d'appartenance suivant la position des centroïdes.
4. Calcul du critère de minimisation et retour à l'étape 2 s'il y a non convergence de critère.

Ces étapes sont regroupées dans l'algorithme de la figure 2.8.

Entrer k classes, centre v de classes, m degré flou des données, partition initiale.

Sortie : partition finale.

Fixer arbitrairement un nombre de classes K

Initialiser la matrice de degré d'appartenance $U_{K \times N}[u_{ij}]$. Cette matrice représente la partition floue des données et doit vérifier la condition de normalisation.

$$\sum_{j=1}^K u_{ij} = 1; j \in \{1, \dots, N\}$$

Répéter

Calculer les centres de classes

$$v_j = \frac{\sum_{k=1}^N x_k u_{ik}^m}{\sum_{k=1}^N u_{ik}^m}, i = 1, \dots, K$$

Mettre à jour la matrice d'appartenance

$$\mu_{ji} = \frac{1}{\sum_{j=1}^K \left(\frac{1}{d_j^2(x_i, v_j)} \right) \left(\frac{1}{m-1} \right)}, i = 1, \dots, K \text{ et } j = 1, \dots, N$$

Jusqu'à convergence

Figure 2.9 : Algorithme des FCM [28]

2.6 Parcours succinct de quelques travaux sur la détection des pathologies mammaires

Les approches menées pour la détection des pathologies mammaires sont diverses et variées. On retrouve des adaptations de toutes les méthodes du traitement d'images mettant en œuvre, l'analyse multirésolution [29], la segmentation par croissance de régions adaptative [30][31], la détection par la morphologie mathématique mise en exergue par la ligne de partage des eaux dans les travaux de Vachier [32], pour la détection des masses, ainsi que l'extraction des structures mammaires, basée sur les ensembles de niveaux dans la recherche de Boujelben et al [33], etc.

Les équipes d'imagerie pour l'aide à la décision des laboratoires LRDSI et LATSI, s'intéressent à la détection de tumeurs mammaires, issues des images mammographiques et

cytologiques et; consacre ses travaux, à l'analyse topologique et texturale pour la réalisation d'un système d'aide au diagnostic, suivant une classification connexionniste, par les arbres de décision ainsi que par les SVM et les algorithmes génétiques en hybridation avec les réseaux de neurones [34][35][36][37][38][39]. L'objectif de ce travail, est de réaliser, une classification via les multiSVM (support à vecteur machine multi classes) multicouche, pour la discrimination des signes malins et bénins mammaires.

2.7 Méthodologie adoptée

L'objectif de la segmentation d'image mammographique est d'extraire les différentes structures d'intérêt d'une manière précise, sans altération de l'information. Les deux modèles FCM et Chan et Vese peuvent être appliqués, au cas de notre étude, pour une meilleure performance des résultats. Bing Nan Li et al [40], ont proposé la coopération des ensembles de niveaux avec les FCM, pour l'automatisation de la segmentation d'une image médicale. L'approche est scindée en deux étapes : la première consiste en la classification floue spatiale et la seconde, utilise les résultats pour initialiser la segmentation de levels set. Les régions d'intérêt sont obtenus, via l'évolution du modèle déformable, grâce à la minimisation de la fonctionnelle d'énergie.

L'algorithme de "Fuzzy levels set" automatise l'initialisation de la segmentation de level set, en utilisant la classification floue spatiale. Il emploie un FCM avec les restrictions spatiales pour déterminer les régions d'intérêt d'une image médicale.

Bénéficiant de l'initialisation flexible comme dans l'équation (2.14), la fonction des levels set améliorée, peut être adaptée à des résultats de FCM directement pour l'évolution.

Supposons que la composante d'intérêt dans un des résultats de la FCM est:

$$R_K: \{r_K = \mu_{nk}, n = x \times N_y + y\}.$$

Il est convenable alors, de lancer la fonction d'ensemble de niveau 2.21 comme suivant l'équation initiale 2.21.

$$\phi_0(x, y) = -4\varepsilon (0.5 - B_k) \quad 2.21$$

où ε est un paramètre de régulation de la fonction Dirac. La fonction Dirac est alors définie par l'équation 2.22).

$$\delta_\varepsilon(x) = \begin{cases} 0, & |x| > \varepsilon \\ \frac{1}{2\varepsilon} [1 + \cos \frac{\pi x}{\varepsilon}], & |x| \leq \varepsilon \end{cases} \quad 2.22$$

B_k est une image binaire obtenue à partir de l'équation 2.23.

$$B_k = R_k \geq b_0 \quad 2.23$$

Où $b_0 \in (0,1)$ est un seuil réglable. Grâce à la classification floue spatiale, B_k peut se rapprocher de la composante d'intérêt, qui peut être facilement ajustée par b_0 .

L'amélioration proposée permet plusieurs avantages. L'évolution des levels set est adaptée à la distance, de l'objet authentique. À l'approche de l'objet, la fonction dynamique, ralentit automatiquement l'évolution, pour épouser la forme de l'objet à détecter, autour de paramètres d'itération et de régulation appropriés.

2.8 Conclusion

Ce chapitre a été consacré aux techniques de prétraitement et de segmentation adoptées, dans le cadre de notre projet. Nous avons opté pour la coopération entre les FCM et les modèles des ensembles de niveaux, pour automatiser la segmentation des masses et des cas normaux mammaires. L'approche étudiée, utilise le résultat du clustering comme la fonction initiale du modèle déformable. L'avantage de l'algorithme coopératif, permet de se rapprocher précisément des contours vrais positifs, suivant des paramètres de contrôle appropriés.

Le chapitre suivant, est consacré à la caractérisation et la classification par les SVM des images mammographiques.

CHAPITRE 3

CARACTERISATION ET

CLASSIFICATION PAR LES

SUPPORTS A VECTEUR MACHINE

D'IMAGES MAMMOGRAPHIQUES

Chapitre 3

Caractérisation et classification par les supports à vecteur machine d'images mammographiques

3.1 Introduction

L'objectif de ce chapitre, est de faire une analyse morphologique et texturale via des attributs choisis qui, permettent de caractériser les différentes pathologies étudiées. Ces caractéristiques regroupées autour de vecteurs représentant la signature de chaque image, nous permettront d'établir une classification par les supports à vecteur machine (SVM), des masses et des cas sains des clichés mammographiques. L'analyse de l'image à travers ses attributs (niveaux de gris ou texture,...) consiste souvent, à extraire un certain nombre de propriétés caractéristiques et, à les exprimer sous forme paramétrique.

3.2 Caractérisation texturale [40][41] [42]

L'analyse des textures joue un rôle important, dans l'interprétation automatique des scènes. Un tel problème, est souvent rencontré dans divers domaines tels que l'imagerie médicale, le contrôle de qualité, la télédétection, etc. En effet, l'utilisation seule des niveaux de gris, ne semble pas être en mesure, de conduire à une bonne analyse de certaines images. Beaucoup d'images possèdent, des zones texturées non homogènes au sens de niveaux de gris, d'où la nécessité d'introduire l'information texturale. L'étude des propriétés d'une image, se fait alors par le traitement numérique de celle-ci via l'information texturale. Cette dernière montre que la distribution spatiale et l'organisation des niveaux de gris, d'une image, comportent beaucoup d'informations, qui méritent d'être formaliser par des paramètres quantitatifs, servant à l'identification de la texture. Plusieurs méthodes de caractérisation de la texture des images, ont été ainsi proposées dans la littérature et, qu'on peut regrouper essentiellement en trois catégories:

1) Méthodes statistiques

- Matrice de cooccurrence;
- Longueurs de plage.

2) Méthodes à base de modèle

- Modèles Fractales ;
- Modèles AR (Autoregressive Models).

3) Méthodes fréquentielles)

- Fourier ;
- Gabor ;
- Ondelettes.

Nous cherchons à avoir une représentation statistique des différentes textures des pathologies mammographiques, en utilisant une méthode statistique du deuxième ordre, appelée *Matrice de cooccurrence*. Les matrices de cooccurrence, seront complètement détaillées car elles serviront de référence pour la deuxième partie de ce mémoire, concernant la construction de paramètres destinés, à la discrimination de texture des différentes régions d'intérêts.

3.2.1 Types de texture

a. Texture périodique

Une texture périodique, est une texture de primitives arrangées d'une façon particulière, rangée. Le motif de base se répète alors, de manière régulière. On parle alors, de texture ordonnée (figure 3.1).

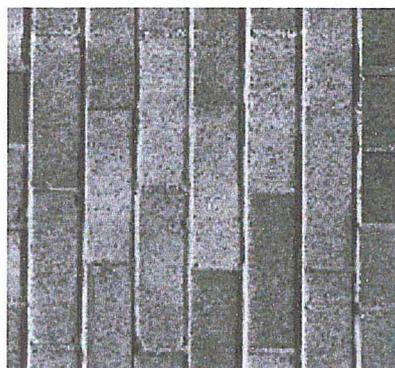


Figure3. 1: Exemple d'une texture périodique [40]

b. Texture aléatoire

Une texture aléatoire semble totalement désordonnée. Il est alors impossible d'isoler un motif de base (cf. Figure 3.2).

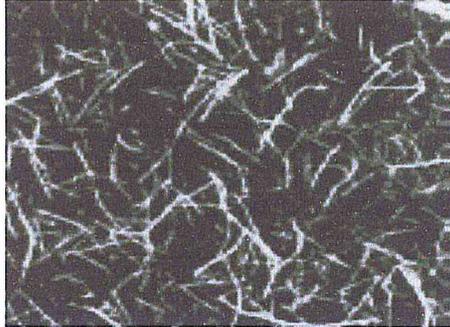


Figure 3. 2: Exemple d'une texture aléatoire [40]

3.2.2 Matrices de concurrence [40] [42]

Du fait de leur richesse en information de texture, les matrices de cooccurrences sont devenues les plus connues et, les plus utilisées pour extraire ces caractéristiques de textures. Elles estiment des propriétés des images relatives, à des statistiques de second ordre. Une matrice de cooccurrence, mesure la probabilité d'apparition des paires de valeurs de pixels situés à une certaine distance, dans l'image. Elle est basée sur le calcul de la probabilité $P(i, j, \delta, \theta)$ qui représente le nombre de fois où un pixel de niveau de couleur i , apparaît à une distance relative δ , d'un pixel de niveau de couleur j et selon une orientation θ donnée.

Les directions angulaires θ classiquement utilisées, sont 0, 45, 90 et 135 degrés.

Ceci conduit à une grande complexité, calculatoire. Plusieurs études ont été menées pour déterminer une distance ou une orientation optimale en concluant, qu'en pratique, une distance courte donne généralement de bons résultats [42].

Afin de limiter le nombre de calculs, on prend généralement comme valeurs une orientation de 0° et une distance de, 1.

Les caractéristiques extraites à partir de ces matrices, contiennent des informations notamment sur l'homogénéité, les dépendances linéaires entre les niveaux de gris, le contraste et la complexité de l'image.

Nous définissons les 4 paramètres considérés, comme les plus pertinents, dans le cas de notre travail.

1. Variance

La variance (équation 3.1), mesure l'hétérogénéité de la texture. Elle augmente lorsque les niveaux de gris, différent de leur moyenne μ . La variance est indépendante du contraste.

$$VAR = \sum_i \sum_j ((i - \mu)^2 P_{ij}(\delta, \theta)) \quad 3.1$$

2. Entropie

Ce paramètre (équation 3.2) mesure le désordre dans l'image. Contrairement à l'énergie, l'entropie atteint de fortes valeurs lorsque la texture, est complètement aléatoire (sans structure apparente). Elle est fortement corrélée (par l'inverse) à l'énergie.

$$ENT = - \sum_i \sum_j (\log P_{ij}(\delta, \theta) P_{ij}(\delta, \theta)) \quad 3.2$$

3. Contraste

La valeur en est d'autant plus élevée, que la texture présente un fort contraste. Ce paramètre (équation 3.3) est fortement non corrélé, à l'énergie [2].

$$CST = \sum_i \sum_j ((i - j)^2 P_{ij}(\delta, \theta)) \quad 3.3$$

4. Corrélacion

La corrélation (équation 3.4) mesure la dépendance linéaire (relativement à (δ, θ)) des niveaux de gris de l'image. La corrélation n'est corrélée ni à l'énergie, ni à l'entropie.

$$COR = \sum_i \sum_j \frac{(i - \mu)(j - \mu) P_{ij}(\delta, \theta)}{\sigma^2} \quad 3.4$$

3.3 Analyse de la forme [43][44]45]

Dans la littérature, des dizaines de paramètres de forme, permettent de caractériser la morphologie des particules. Une des difficultés de l'analyse morphologique, réside dans le choix du paramètre de forme le plus adapté à la nature des particules.

Les caractéristiques que nous avons choisies de calculer, ce sont celles qui nous paraissent les plus pertinentes, pour modéliser numériquement les régions d'intérêt mammaires. Elles permettent ainsi, d'identifier les zones cancéreuses, en se basant sur les descripteurs morphologiques.

Dans notre approche on va utiliser cinq paramètres basés sur la compacité, la convexité, l'excentricité, la rectangularité et la surface.

3.3.1 La compacité

La compacité (équation 3.5) est une mesure de la complexité du contour vis à vis de l'aire A.

$$\text{Comapcité} = \frac{4\pi A}{P^2} \quad 3.5$$

Où P, représente le périmètre de la région d'intérêt.

3.3.2 La convexité

La convexité (équation 3.6) est la différence relative entre un objet et une coque convexe la contenant. On la calcule à partir du rapport, au périmètre de la coque convexe o de l'objet et le périmètre de l'objet.

$$\frac{P_{ch}}{P_{img}} \quad 3.6$$

Où P_{ch} est le périmètre de la coque convexe de l'objet, et P_{img} est le périmètre de l'objet

3.3.3 L'excentricité

L'excentricité (équation 3.7) définit comment l'objet, est distribué dans l'espace tout au long de son axe. Elle représente, le rapport entre la longueur du grand axe et celle, du petit axe de l'objet. Cette mesure peut être calculée par la méthode des axes principaux ou bien, par la méthode du rectangle minimum englobant.

$$\text{excentricité} = \frac{(\mu_{02} - \mu_{20})^2 + 4\mu_{11}}{A} \quad 3.7$$

où A est l'aire du secteur, et les moments centraux peuvent être calculés par :

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \bar{x})^p (y - \bar{y})^q \quad 3.8$$

Avec $p + q > 1$

3.3.4 La rectangularité

Il s'agit d'un descripteur de forme intitulé rectangularité ou boîte englobante. Il est souvent utilisé, pour décrire le taux de rectangularité et d'allongement d'une région [43]. On désigne par R la région étudiée et B_E la boîte englobante qui est définie par le plus petit rectangle contenant la lésion étudiée. La mesure de rectangularité notée *Rectangularité* peut alors se définir comme suit :

$$\text{Rectangularité} = \frac{\text{aire}(R)}{\text{aire}(B_E)} \quad 3.9$$

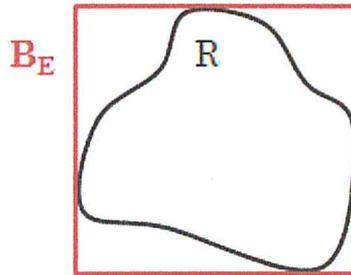


Figure3. 3 : Boîte englobante

3.3.5 La surface

Il s'agit de l'aire en pixels, qui est calculée à partir du nombre de pixels contenus dans une lésion.

Les paramètres formes fournissent, des informations importantes pour distinguer la masse et la non-masse.

3.4 Introduction à la classification

Les méthodes de classification ont pour but d'identifier les classes, auxquelles appartiennent des objets à partir de certains traits descriptifs ou caractéristiques.

Elles s'appliquent à un grand nombre d'activités humaines et, conviennent en particulier au problème de la prise de décision automatisée.

La procédure de classification sera extraite, automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification. Dans ce cas, il s'agit en effet, d'extraire une règle générale à partir des données observées. La procédure générée, devra classer correctement les exemples de l'échantillon et, avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions. Le terme de classification peut désigner deux approches distinctes: la classification supervisée et la classification non-supervisée.

Les méthodes non supervisées ont pour but de constituer des groupes d'exemples (ou des groupes d'attributs) en fonction des données observées, sans connaissance a priori. En revanche les méthodes supervisées utilisent la connaissance a priori sur l'appartenance d'un exemple à une classe pour construire un système de reconnaissance de ces classes [46].

3.4.1 Principe de la classification non-supervisé

Le but de la classification non-supervisée (ou du clustering) est de créer un ensemble de clusters, regroupant des données de mêmes caractéristiques. Le regroupement de données est basé, sur la notion de distance par rapport à des centroïdes, c'est à- dire les centres de masse de chaque classe. Les classes sont a priori inconnues et sont créées selon la ressemblance des données. Les principales méthodes de la classification non-supervisée sont les algorithmes, de classification hiérarchique et les méthodes de regroupement [46].

3.4.2 Principe de la classification supervisée

Dans le cadre de la classification supervisée, les classes sont connues et, l'on dispose d'exemples de chaque classe. L'objectif de la classification supervisée, est d'apprendre à l'aide d'un ensemble d'entraînement, une procédure de classification qui permet de prédire l'appartenance d'un nouvel exemple, à une classe. Les systèmes d'apprentissage permettant d'obtenir une telle procédure, peuvent être basés sur des hypothèses probabilistes (classifieur

naïf de Bayes), sur des notions de proximité (plus proches voisins) ou sur des recherches dans des espaces d'hypothèses (arbres de décisions, . . .) ou suivant les réseaux de neurones ou les SVM. Nous décrivons dans la suite quelques méthodes de classification supervisée [46].

1. Classifieur bayésien naïf [46]

Le classifieur bayésien naïf est une méthode qui trouve son fondement théorique dans le théorème de Bayes. Il permet les inférences probabilistes et repose sur l'hypothèse que les solutions recherchées, peuvent être trouvées à partir de distributions de probabilité dans les données et, dans les hypothèses. Cette méthode permet de déterminer la classification d'un exemple quelconque spécifiée en termes d'attributs en supposant que, les attributs $\{x_1, \dots, x_n\}$ de l'espace d'entrée X sont indépendants les uns des autres. On peut voir les probabilités qui sont estimées sur l'ensemble d'échantillons X .

Points positifs : simple et facile à programmer, souvent efficace.

Point négatif : très sensible à la présence d'attributs corrélés.

2. k-PPV [46]

L'algorithme des k plus proches voisins (noté k -PPV), est une méthode basée sur la notion de proximité (voisinage) entre les exemples et le raisonnement, à partir de cas similaires pour prendre une décision. Autrement dit des entrées x_i semblables devraient avoir des valeurs y_i semblables. Le principe est le suivant: on note x un nouvel exemple décrit, par un vecteur de p attributs. On trouve alors, parmi l'ensemble d'exemples d'apprentissage, les k plus proches voisins de x et on associe à x la classe majoritaire parmi ses k voisins, lui ressemblant le plus dans la, base d'apprentissage. Cette méthode dépend donc des trois éléments suivants:

1. Le nombre de voisins retenus.
2. La mesure de distance entre exemple.
3. La combinaison des classes.

Points positif: Elle ne pose aucune hypothèse sur la forme des classes à apprendre. La méthode est simple, puisqu'il n'y a pas besoin d'apprentissage d'un modèle de classification et son pouvoir prédictif, est souvent bon.

Point négatif : La performance de cette méthode diminue lorsque la dimension augmente, puisque pour chaque nouvelle classification, il est nécessaire de calculer toutes les distances de x à chacun des exemples d'apprentissage. De plus, la performance dépend fortement de k , le nombre de voisins choisi et il est nécessaire d'avoir un grand nombre d'observations pour obtenir une bonne précision des résultats.

3. Les réseaux de neurones [48]

Le principe consiste à apprendre à classifier correctement des données à partir d'un jeu d'exemples déjà classifiés, c'est-à-dire l'apprentissage par l'expérience. Un réseau de neurones est constitué d'un graphe pondéré orienté dont les nœuds symbolisent les neurones. Ces neurones possèdent une fonction d'activation qui permet d'influencer les autres neurones du réseau, les fonctions, les plus souvent utilisées, sont la fonction signe ou la fonction sigmoïde. Les neurones peuvent être organisés de différentes manières, c'est ce qui définit l'architecture et le modèle du réseau. L'architecture la plus courante est celle dite du perceptron multicouche.

Points positifs : excellentes performances, minimise la fréquence d'erreur.

Point négatif : difficulté de mise en œuvre.

4. Arbres de décision [46]

Les arbres de décision sont une des méthodes les plus connues en classification. Le principe des arbres de décision, est de réaliser la classification d'un exemple par une suite de tests sur les attributs qui le décrivent. Concrètement, dans la représentation graphique d'un arbre,

1. Un nœud interne correspond à un test sur la valeur d'un attribut.
2. Une branche part d'un nœud et correspond à une ou plusieurs valeurs de ce test.
3. Une feuille est un nœud d'où ne part aucune branche et correspond à une classe.

Une règle de décision (de la forme si . . . alors . . .) est créée pour chaque chemin partant de la racine de l'arbre et, parcourant les tests (en faisant des conjonctions) jusqu'à la feuille qui est l'étiquette de la classe. Les arbres de décision sont particulièrement appréciés, car ils permettent une compréhension aisée, mais lors d'une tâche de classification relativement complexe, l'utilisateur n'est plus capable d'explorer efficacement les résultats obtenus sous forme textuelle.

5. Bagging

Cette méthode a été proposée par L. Breiman [49]. Le principe du bagging est d'entraîner un algorithme d'apprentissage, sur plusieurs bases d'apprentissage obtenues par tirage avec remise (ou bootstrap) d'exemples d'apprentissage, dans l'échantillon d'apprentissage S . Pour chaque tirage b , une hypothèse h_b est obtenue. L'hypothèse finale est basée, sur les moyennes des hypothèses obtenues.

Le point positif : cette méthode est facile à mettre en place et, s'adapte à tout type de classifieur.

Les points négatifs : elle est coûteuse en temps de calcul. Une amélioration de la qualité de la classification, se fait au détriment de l'interprétation.

Nous présentons dans ce qui suit, classification supervisée adoptée dans notre étude, par les machines à support vecteur (SVM).

3.5 Les machines à vecteur support (SVM) [50][51][52][53]

Les machines à vecteur support (SVM) ont reçu une attention considérable, ces dernières années et un grand nombre d'applications sont décrites, dans la littérature [50].

L'objectif des SVM est de maximiser la marge de séparation, entre les classes. Les SVM sont des algorithmes d'apprentissage faisant l'hypothèse que les éléments sont linéairement séparables, s'ils sont préalablement projetés dans un espace de dimensions adéquat.

Le principe des approches SVM est de trouver l'hyperplan optimal, parmi l'ensemble des hyperplans possibles (figure (3.4)) permettant de classifier correctement les données (i.e. les données des classes de label +1 et -1 se trouvent de chaque côté de l'hyperplan) mais également tel que sa distance aux vecteurs (exemple d'apprentissage) les plus proches soit maximale (i.e. le plus loin possible de tous les exemples). Les vecteurs les plus proches sont alors dénommés "vecteurs supports" et la distance constitue la marge optimale.

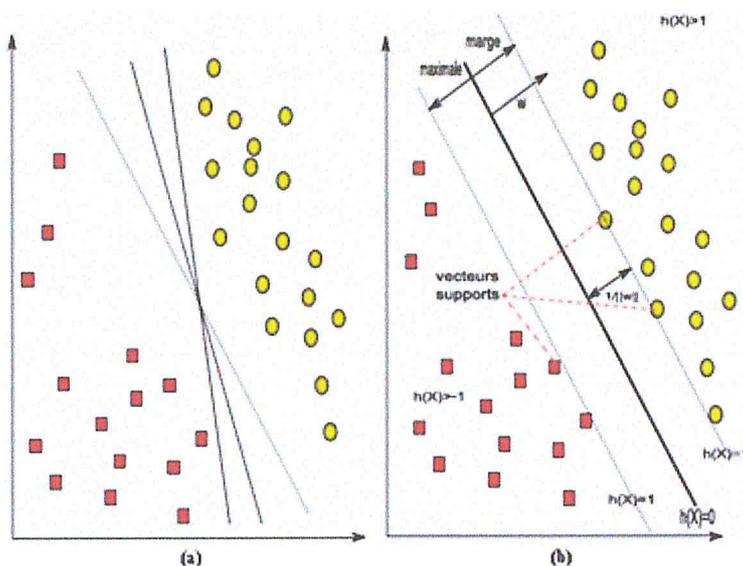


Figure3. 4: (a) Recherche de l'hyperplan optimal ; (b) Hyperplan optimal, marge et vecteurs supports [51].

3.5.1 Principe de fonctionnement [50][51]

Le principe de base des SVM, consiste de ramener le problème de la discrimination à celui, linéaire, de la recherche d'un hyperplan optimal. Deux idées ou astuces permettent d'atteindre cet objectif:

- La première consiste à définir l'hyperplan comme solution d'un problème d'optimisation, sous contraintes dont la fonction objectif ne s'exprime qu'à l'aide de produits scalaires, entre vecteurs et dans lequel le nombre de contraintes "actives" ou vecteurs supports contrôle, la complexité du modèle.
- Le passage à la recherche de surfaces séparatrices, non linéaires est obtenu par l'introduction d'une fonction noyau (kernel) dans le produit scalaire induisant, implicitement une transformation non linéaire des données vers un espace intermédiaire, de plus grande dimension. D'où l'appellation couramment rencontrée de machine à noyau ou kernel machine. Sur le plan théorique, la fonction noyau définit un espace hilbertien, dit auto-reproduisant et isométrique par la transformation non linéaire de l'espace initial et dans lequel, est résolu le problème linéaire (figure 3.5).

Cet algorithme devient largement, utilisé dans de nombreux types d'applications et s'avère un concurrent sérieux des algorithmes les plus performants (agrégation de modèles). L'introduction de noyaux, spécifiquement adaptés à une problématique donnée, lui confère une grande flexibilité pour s'adapter à des situations très diverses (reconnaissance de formes, de caractères, détection, diagnostic...). À noter que, sur le plan algorithmique, ces algorithmes sont plus pénalisés par le nombre d'observations, c'est-à-dire le nombre de vecteurs supports potentiels, que par le nombre de variables. Néanmoins, des versions performantes, permettent de prendre en compte des bases de données volumineuses, dans des temps de calcul acceptables.

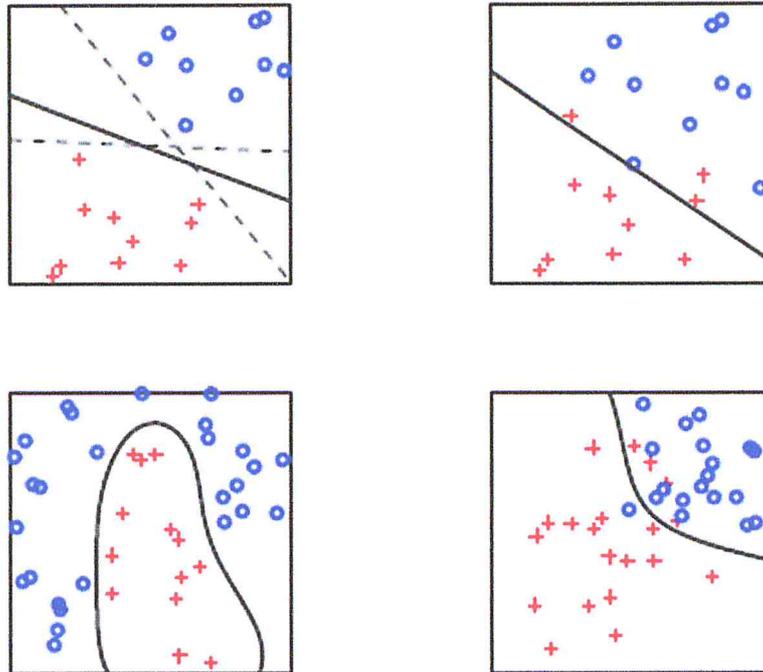


Figure 3. 5: Exemples de quatre types de problèmes de discrimination binaire où il s'agit de séparer les points bleus des croix rouges. La frontière de décision est représentée en noir.

3.5.2 Classification des SVM multi-classes [50][51][53]

L'adaptation des SVM bi classe au cas multiclass, peut se faire de trois façons différentes. Le choix va dépendre de la taille du problème.

1. L'approche un contre tous consiste, à entraîner un SVM bi-classe en utilisant les éléments d'une classe contre tous les autres. Il s'agit de résoudre les problèmes de l'ordre de, c SVM chacun de taille n .
2. L'approche un contre un : consiste à entraîner $\frac{c(c-1)}{2}$ SVM sur chacun des couples de classes, puis à décider de la classe gagnante soit par un vote majoritaire soit, en traitant les résultats grâce à l'estimation de probabilités a posteriori. Le nombre de classifieurs SVM à entraîner, peut être réduit en utilisant un codage astucieux pour les classes, à travers un code correcteur d'erreur ou un graphe direct acyclique.
3. L'approche globale, consiste à traiter le problème en une seule fois. Cela peut se faire en posant formellement le problème, par exemple si l'on note $f_\ell(x) - b_\ell$ la fonction de discrimination associée à la classe ℓ se réduit à la relation 3.10.

$$\left\{ \begin{array}{l} \min_{f \in \mathcal{H}, b, \xi \in \mathbb{R}^n} \frac{1}{2} \sum_{\ell=1}^c \|f_{\ell}\|_{\mathcal{H}}^2 + \frac{c}{d} \sum_{i=1}^n \sum_{\ell=1, \ell \neq y_i}^c \xi_{i\ell} \\ \text{avec } y_i (f y_i(x_i) + b y_i - f_{\ell}(x_i) - b_{\ell}) \geq 1 - \xi_{i\ell} \\ \xi_{i\ell} \geq 0 \text{ pour } i = 1, \dots, n; \ell = 1, \dots, c; \ell \neq y_i \end{array} \right. \quad 3.10$$

Ce problème est un programme quadratique, de même nature que celui des SVM et de taille $n \times c$. L'estimateur associé, est non consistant mais donne de bons résultats en pratique.

1. Trouver l'hyperplan

On peut montrer que pour un hyperplan, canonique $w \cdot x + b = 0$, la marge M est donnée par l'expression 3.11.

$$M = \frac{1}{\|w\|} \quad 3.1$$

Où $\|w\| = \sqrt{w_1^2 + \dots + w_n^2}$. On voit donc que plus la norme $\|w\|$ est petite, plus la marge de l'hyperplan canonique correspondant, est grande. Ainsi, afin de trouver l'hyperplan qui sépare le mieux les données, il faut trouver celui qui respecte les conditions d'un hyperplan canonique et pour lequel, la norme $\|w\|$ est minimale.

La recherche du meilleur hyperplan peut donc s'écrire, sous la forme du problème d'optimisation suivant:

$$\text{Minimiser } \|w\| \text{ sujet à } y_i (w \cdot x_i + b) \geq 1, i = 1, \dots, l \quad 3.12$$

Les contraintes $y_i(w \cdot x_i + b) \geq 1, i = 1, \dots, l$ assurent d'une part que l'hyperplan sépare les données correctement et d'autre part, qu'il est canonique. En effet, $y_i(w \cdot x_i + b) > 0$ si et seulement si $(w \cdot x_i + b) = y_i$, donc si et seulement si x_i est du bon côté de l'hyperplan. Ainsi, l'hyperplan doit correctement séparer les données. Ensuite, on peut montrer qu'imposer $y_i(w \cdot x_i + b) \geq 1$ assure que pour toutes les données qui ne sont pas sur la marge, $|w \cdot x_i + b| > 1$ et que $|w \cdot x_i + b| = 1$ pour les données sur la marge, donc que l'hyperplan est canonique.

Nous avons ainsi formulé un problème d'optimisation dont la solution optimale, est l'hyperplan canonique séparant les données avec la plus vaste marge possible. Cependant, il

est possible de formuler un problème équivalent, mais avec une fonction objectif plus simple. En effet, comme la relation 3.13 ;

$$\| w \| = \sqrt{w \cdot w} \quad 3.2$$

minimiser $\| w \|$ est équivalent à minimiser $w \cdot w$. Évidemment, minimiser $w \cdot w$ est équivalent à minimiser $\frac{1}{2} w \cdot w$ (cette petite modification permet d'éviter d'avoir une constante dans la représentation duale du problème. Par conséquent, afin de trouver l'hyperplan canonique qui sépare les données avec la plus grande marge possible, il suffit de résoudre le problème d'optimisation suivant :

$$\text{minimiser } \frac{1}{2} w \cdot w \text{ Sujet à } y_i(w \cdot x_i + b) \geq 1, i = 1, \dots, l \quad 3.3$$

Une propriété très intéressante de ce problème est que la relation 3.15, est une fonction strictement convexe. Ceci assure qu'il n'y a pas de minimum relatif et qu'il n'existe qu'une unique solution optimale.

$$f(w) = w \cdot w = w_1^2 + \dots + w_n^2 \quad 3.15$$

Dans notre projet, nous avons opté pour la méthode "un contre un" qui consiste, à construire trois classifieurs binaires, en confrontant chacune des classes. En phase de test, l'échantillon à classer, est analysé par chaque classifieur et, un vote majoritaire permet de déterminer sa classe.

3.6 Conclusion

Dans ce chapitre, on a abordé les descripteurs morphologiques, suivant les différents paramètres qui nous informent sur la compacité, la convexité, excentricité, la surface et la rectangularité, afin d'identifier le changement de la forme.

Une description texturale, a également été prise en compte suivant l'entropie, la variance, le contraste, la corrélation et l'entropie, afin d'analyser la granularité mammographique.

Ces descriptifs morphologiques et texturaux, seront intégrés dans les multi SVM pour la classification des masses mammographiques.

Le dernier chapitre, est consacré à la mise en œuvre du système de détection de masses et de cas sains, dont l'application, est faite sur des cas réels d'images mammographiques.



CHAPITRE 4

**APPLICATION ET RÉSULTATS SUR DES
IMAGES MAMMOGRAPHIQUES**

sur des images mammographiques

4.1 Introduction

L'objectif de ce chapitre, est de présenter l'application développée, les traitements illustrés par des images mammographiques dans le but de l'extraction et de la classification des pathologies mammaires issues de la base de données MIAS (Mammographic Image Analysis Society) [54]. Ces images de type MLO (incidence latérale oblique) ont une résolution spatiale de 1024 x 1024 pixels. Notre système détecte les masses et les cas sains de la base MIAS suivant les machines à support de vecteurs (Multi-SVM), en vue d'une aide à la décision.

4.2 Présentation de l'environnement du travail

Notre système est développé, sous environnement Windows 7, à l'aide de l'outil de programmation MATLAB de la version R2014b.

4.2.1 Matériel utilisé

L'implantation de nos codes, est réalisée grâce à un PC portable, ayant un double processeur Intel Core I5 de fréquence de 2.50 Ghz et d'une capacité de mémoire vive « RAM de 8 Go » sous système d'exploitation 64 bits.

4.2.2 Langage utilisé

Notre système est développé sous environnement Windows, à l'aide de l'outil de programmation MATLAB (abréviation de " **Matrix Laboratory**") de la version R2014b [55], qui est un environnement informatique conçu pour le calcul matriciel. L'élément de base est une matrice, dont la dimension n'a pas à être fixée. Matlab est un outil puissant qui permet la résolution de nombreux problèmes, en beaucoup moins de temps. On peut se rendre compte de la puissance du logiciel en lançant la commande *demo*. Les compléments d'information peuvent être facilement obtenus, grâce à l'aide en ligne. Matlab (figure 4.1) s'élargit sur la

bibliothèque des fonctions mathématiques, l'environnement graphique, ainsi que sur une interface de développement.

On le trouve dans les applications de :

- Calcul numérique,
- développement d'algorithmes et la programmation
- modélisation et simulation,
- analyse et visualisation de données,
- création de graphiques scientifiques,
- création d'application avec interfaces utilisateurs.
- l'interfaçage avec d'autres langages (Fortran, C++).
- le développement d'interface utilisateur (GUI).
- L'utilisation des boîtes à outils (Toolboxes) spécialisées.

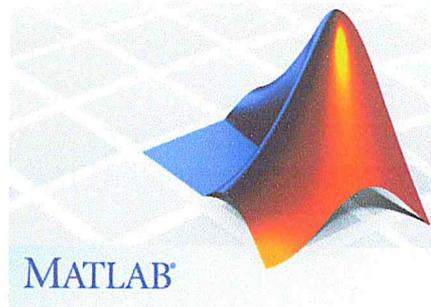


Figure4. 1 : L'icône du logiciel Matlab

S'il est parfaitement adapté à l'automatique et au traitement du signal, sa facilité d'emploi avec des nombres complexes et ses possibilités d'affichages graphiques, en font un outil intéressant pour plusieurs types d'applications. De plus, des " toolboxes " (boîtes à outils) sont disponibles dans de nombreux domaines du traitement du signal, de la commande des systèmes, des réseaux de neurones, de la logique floue, des ondelettes, de la simulation, etc.

➤ Première utilisation de Matlab

Pour découvrir matlab d'une façon efficace, il faut utiliser son aide en ligne. On présente ci-dessous, quelques commandes intéressantes:

- **Help** : « help » donne la liste des aides générales possibles;
- **Helpwin**: ouvre une fenêtre et donne accès, à une aide détaillée;
- **Info** : fournit des informations sur la version de Matlab;
- **Lookfor<mot clé>** : Recherche du mot-clé indiqué dans toutes les entrées d'aide.
- **Help+ nom de commande** : donne l'utilisation de cette commande;
- **Démo** : lance une démo générale de matlab;
- **Help demos** : donne une liste des démos existantes.

Quelques commandes importantes, liées à l'espace de travail :

- **Clear**: supprime la variable de l'espace de travail;
- **Clc**: efface l'écran son toucher les variables.
- **CTRL + C** : (touches du clavier) permet d'interrompre un programme (ie : récupérer la main).
- **Exit/quit**: permet de sortir de matlab;
- **Who/whos**: donne le nom de variable présente dans l'espace de travail;
- **Look for+commande**: donne une Liste de commandes, qui ont un rapport avec le

nom de commande.

- **mkdir** : réalise la création d'un répertoire;
- **pwd** : indique le répertoire courant;
- **cd** : permet de changer de répertoire;
- **path** : visualise ou change le chemin de recherche du répertoire de MATLAB.

4.2.3 Présentation du 'gui'

Les interfaces graphiques (ou interfaces homme-machine) sont appelées GUI (pour Graphical User Interface) sous MATLAB. Elles permettent à l'utilisateur d'interagir avec un programme informatique, grâce à différents objets graphiques (boutons, menus, cases à cocher...). Ces objets sont généralement actionnés à l'aide de la souris ou du clavier. Malgré le fait que les interfaces graphiques semblent secondaires, par rapport au développement du cœur d'une application, elles doivent néanmoins être conçues et développées avec soin et rigueur. Leur

efficacité et leur ergonomie sont essentielles dans l'acceptation et l'utilisation de ces outils par les utilisateurs finaux.

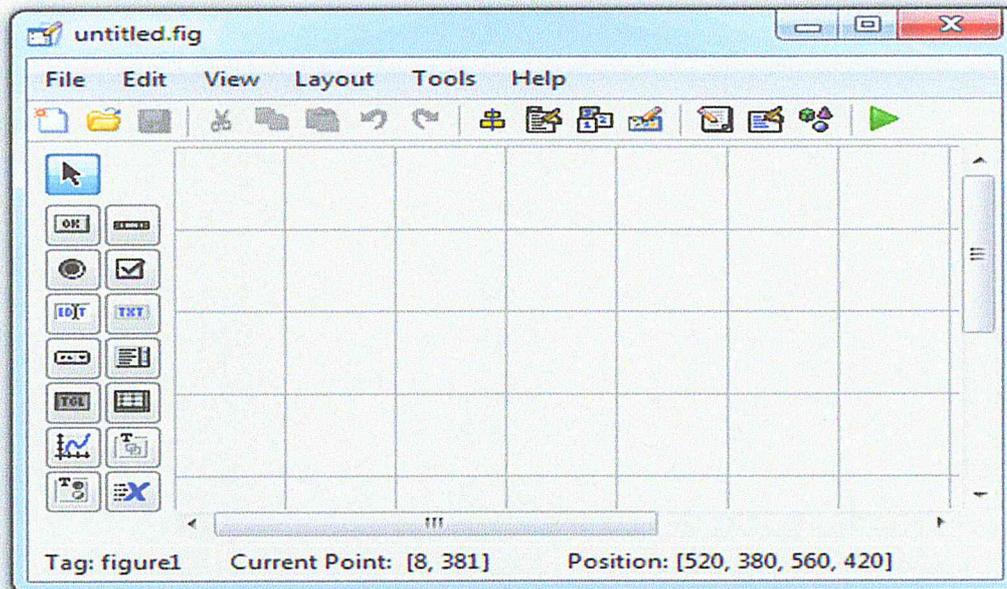


Figure4. 2: Fenêtre principale du guide [55]

4.3 Présentation du système général 'MammoFlouSVM'

Le système 'MammoFlouSVM', a pour objectif, de classifier les pathologies mammographiques moyennant le classifieur des Multi-SVM. L'identification de ces zones s'est faite, par les étapes de la figure 4.3.

1. Rehaussement du contraste par, chapeau haut de forme, afin de faciliter la détection des masses sur le mammographe.
2. Segmentation par le modèle hybride du clustering flou et de Chan- Vese, afin de détecter les régions d'intérêt.
3. Caractérisation texturale et morphologique des masses et des cas sains.
4. Classification par les multi-SVM, des masses et des cas sains.

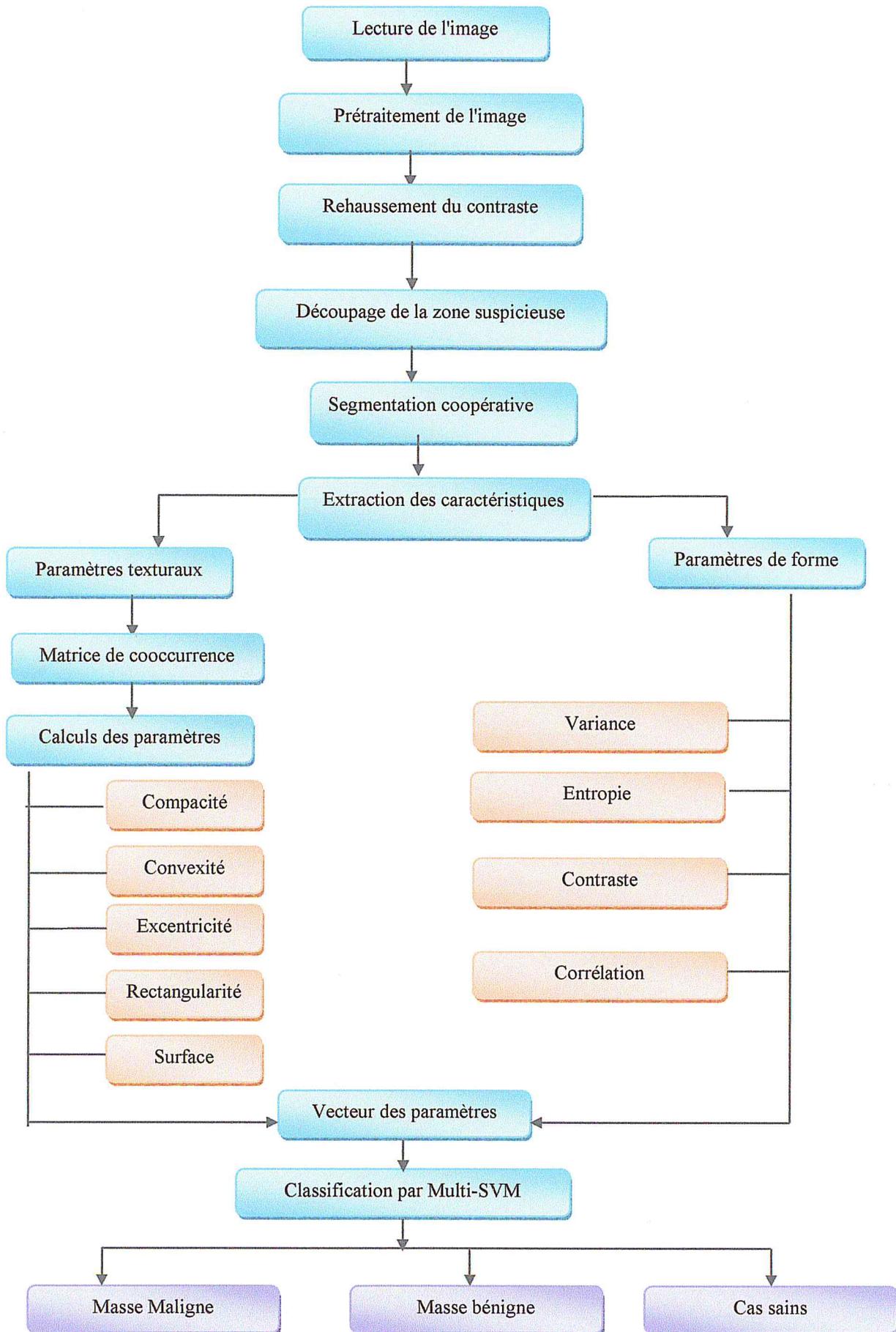


Figure4.3 : Système d'aide à la décision mammographique

✓ Images étudiées

Les images mammographiques utilisées, dans le cas de cette application, proviennent de la base d'images MIAS, qui font l'objet de notre étude. Celle-ci, contient 322 images, dont 200 représentent des cas sains et 122, sont réparties entre les masses et les calcifications. Nous nous sommes intéressées à l'analyse des masses et des cas sains.

Nous avons considéré 239 images dans notre étude. Nous présentons dans la figure 4.4 quelques exemples de la base MIAS.

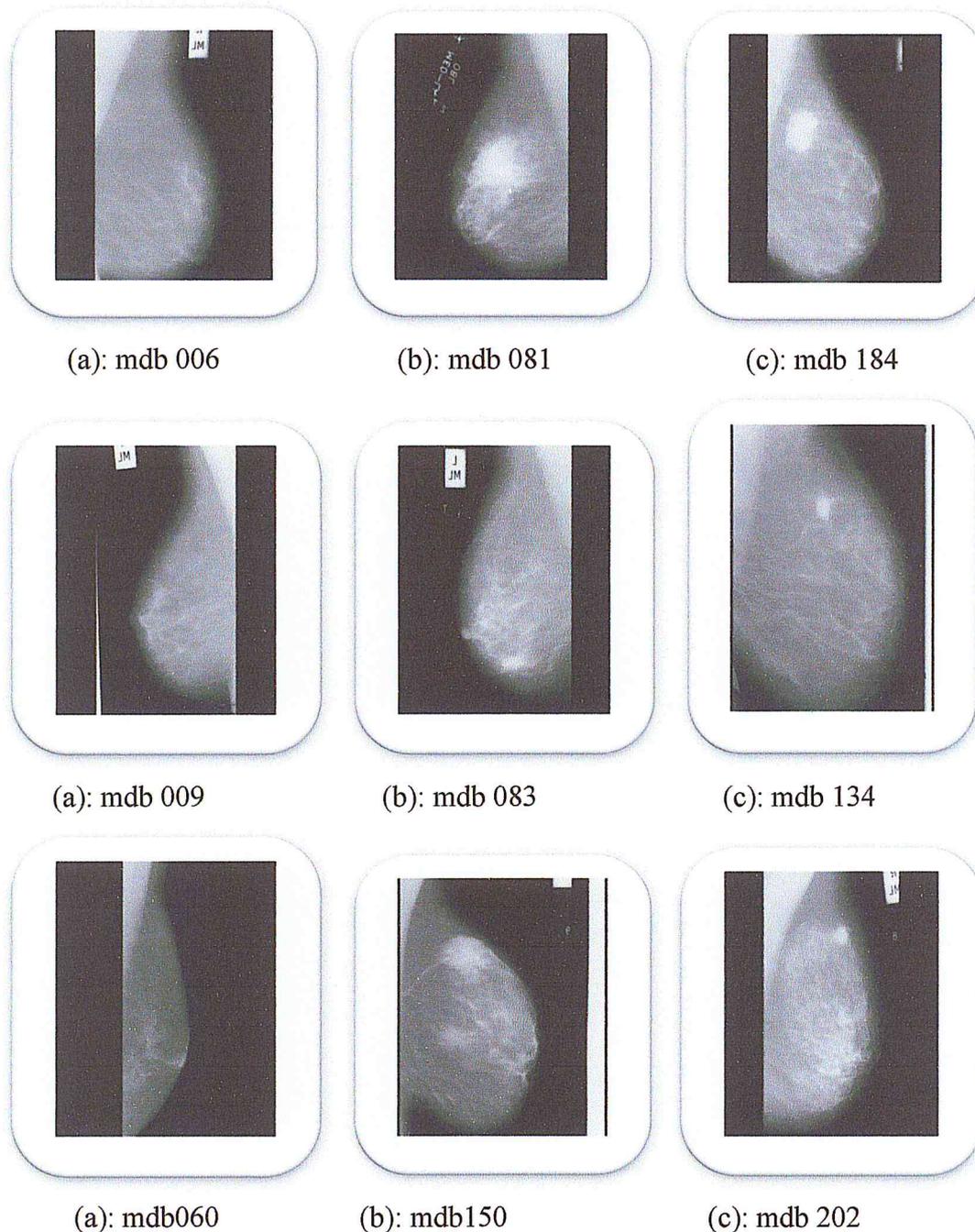


Figure4. 4 : Image mammographiques de la base MIAS [54]

a): cas sain. b): masse bénigne. c): masse maligne

4.4 Résultats et interprétation

L'application 'MammoFlouSVM', effectue l'analyse des masses et des cas sains, en vue d'une classification par les multi-SVM. Les résultats des différentes étapes, se succèdent selon l'organisation suivante.

4.4.1 Prétraitement

Notre prétraitement, consiste en premier lieu à appliquer sur l'image originale une transformation du chapeau haut de forme, pour faire apparaître les structures claires de l'image. Une superposition de ce résultat sur l'image originale, d'améliorer sa qualité visuelle. Les figures suivantes illustrent les étapes de rehaussement du contraste des images mammographiques.

L'opération du chapeau haut de forme, est basée sur un élément structurant de forme d'un disque, avec une taille qui varie entre 60 et 120 selon l'image considérée.

❖ Cas sains

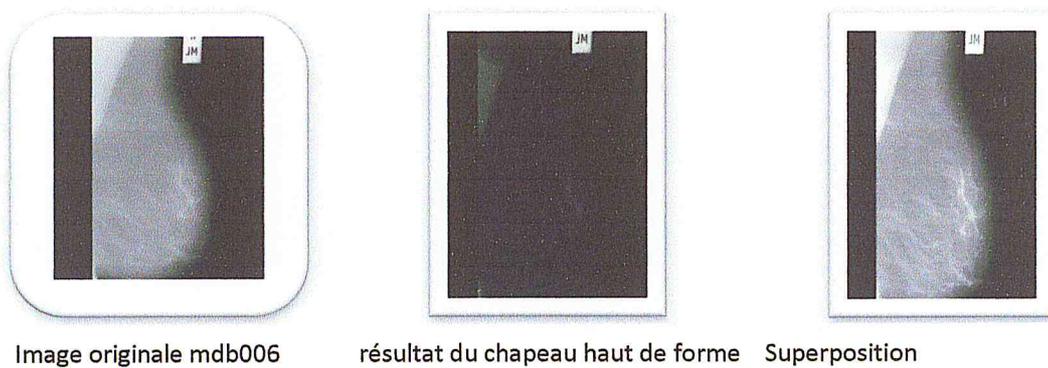


Figure4. 5 : Prétraitement d'une image saine (disque de taille= 80)

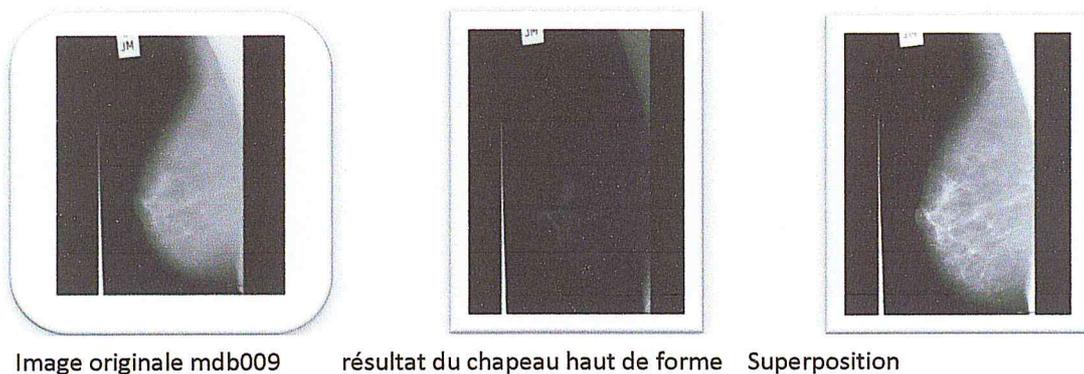


Figure4. 6 : Prétraitement d'une image saine (disque de taille=100)



Figure4. 7 : Prétraitement d'une image saine (disque de taille =60)

❖ Cas de masses bénignes



Figure4. 8 : Prétraitement d'une masse bénigne (disque de taille=110)

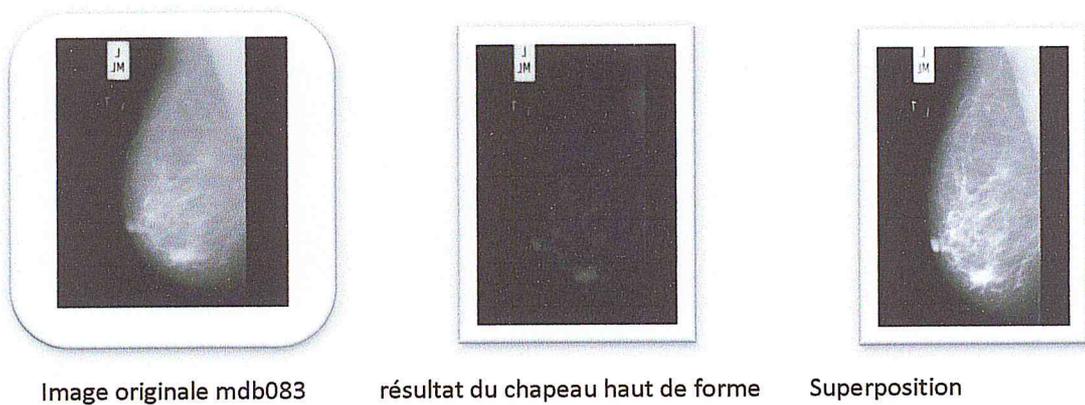


Figure4. 9 : Prétraitement d'une masse bénigne (disque de taille=80)



Image originale mdb150 résultat du chapeau haut de forme Superposition

Figure4. 10 : Prétraitement d'une masse bénigne (disque de taille=100)

❖ Cas de masses malignes



Image originale md184 résultat du chapeau haut de forme Superposition

Figure4. 11 : Prétraitement d'une masse maligne (disque de taille=80)

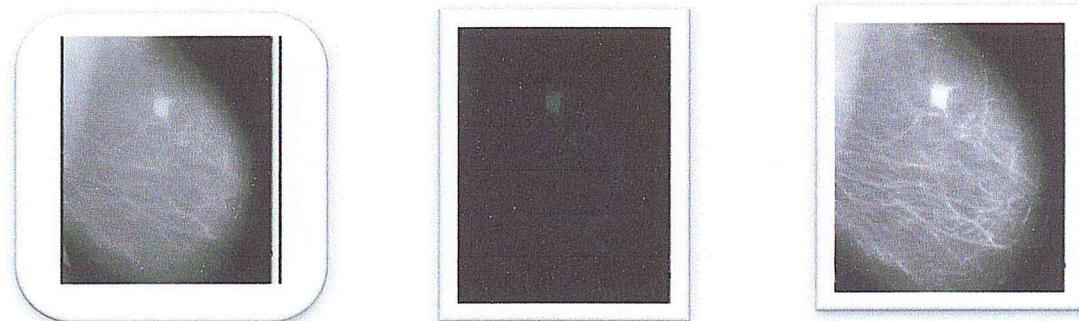


Image originale mdb134 résultat du chapeau haut de forme Superposition

Figure4. 12 : Prétraitement d'une masse maligne (disque de taille=120)

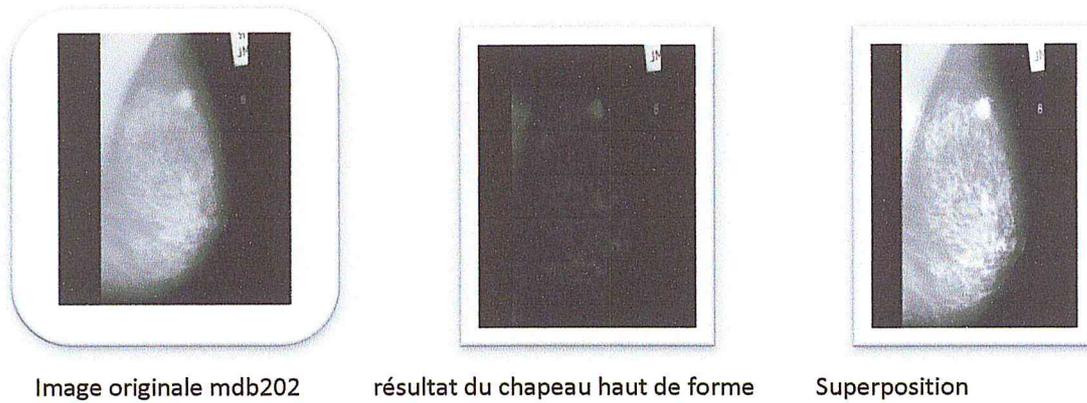


Figure4. 13 : Prétraitement d'une masse maligne (disque de taille=80)

4.4.2 La segmentation

La détection des structures tumorales s'est faite, par une technique de segmentation coopérative moyennant, les contours actifs et la segmentation par clustering. Pour les cas sains, nous avons pris en considération toute la glande mammaire.

➤ Cas des masses

Dans notre application, nous avons opté pour une approche coopérative, utilisant une segmentation par une technique de clustering flou pour délimiter la région de la lésion, suivie des levels set, pour affiner le contour de la masse.

Le choix du nombre d'itérations a été pris en compte judicieusement, puisque lors de nos tests, nous avons remarqué que la taille de la masse (telle que expertisée par un médecin spécialiste), influe sur le nombre d'itérations considéré, plus elle est grande, plus le nombre d'itérations est consistant et vice versa.

Au départ on sélectionne le nombre d'itérations voulu, en prenant en compte le cluster de la zone suspectieuse pour l'initialisation du level set. L'évolution se fait par l'algorithme décrit dans le chapitre précédent, jusqu'à la convergence (nombre d'itérations atteint).

En appliquant le programme de segmentation sur la lésion suspectieuse, on obtient ainsi la région segmentée (figures 4.14, 4.15 et 4.16).

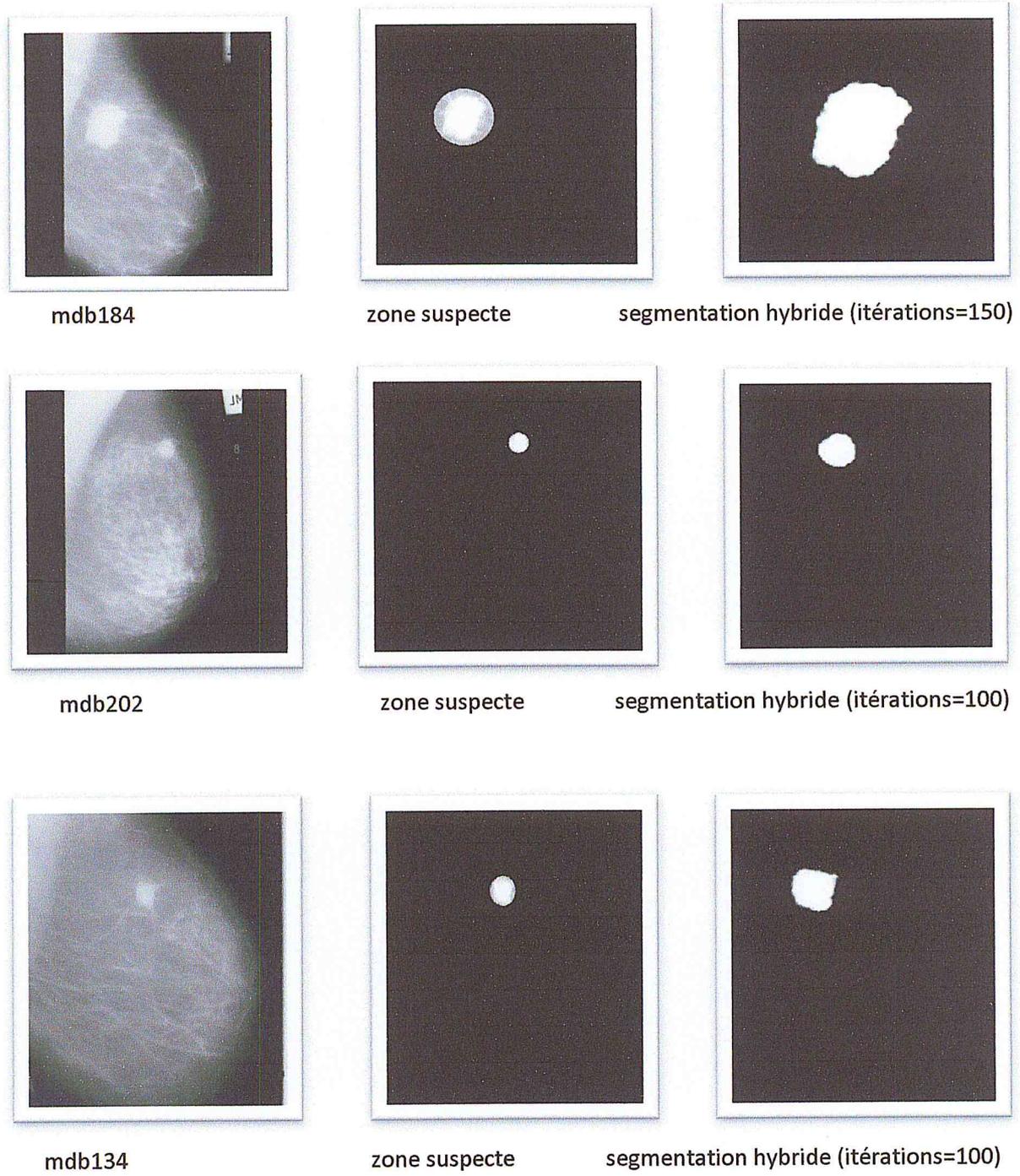
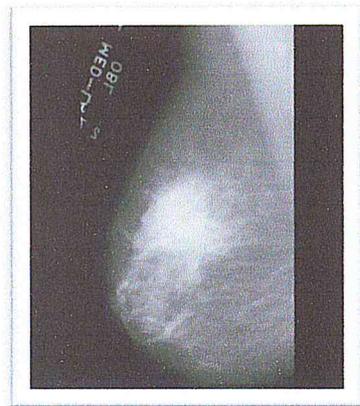
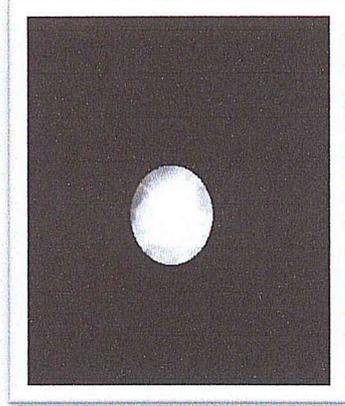


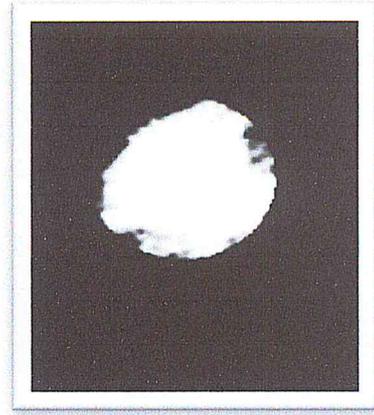
Figure4. 14 : Résultats de la segmentation sur les masses malignes



mdb080



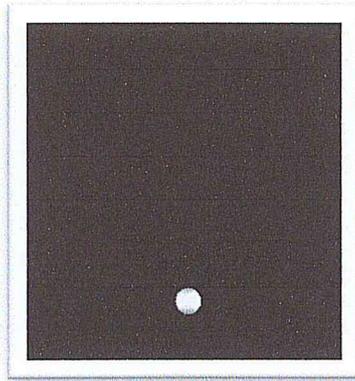
zone suspecte



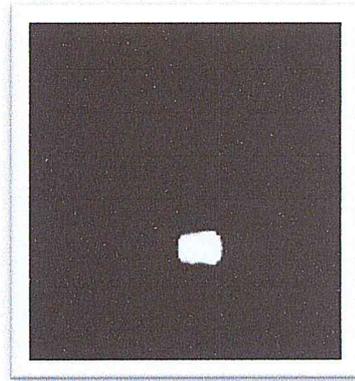
segmentation hybride (itérations=100)



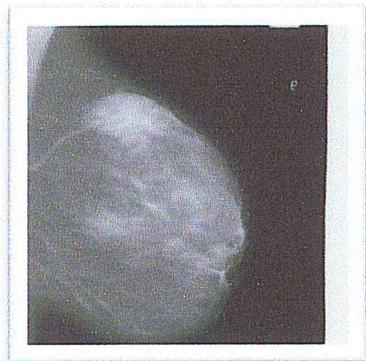
mdb083



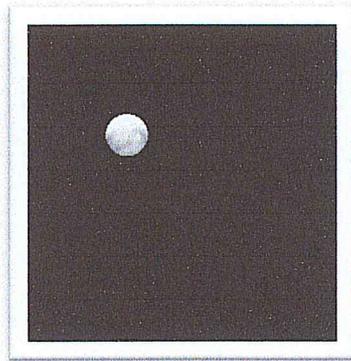
zone suspecte



segmentation hybride(itérations=100)



mdb150

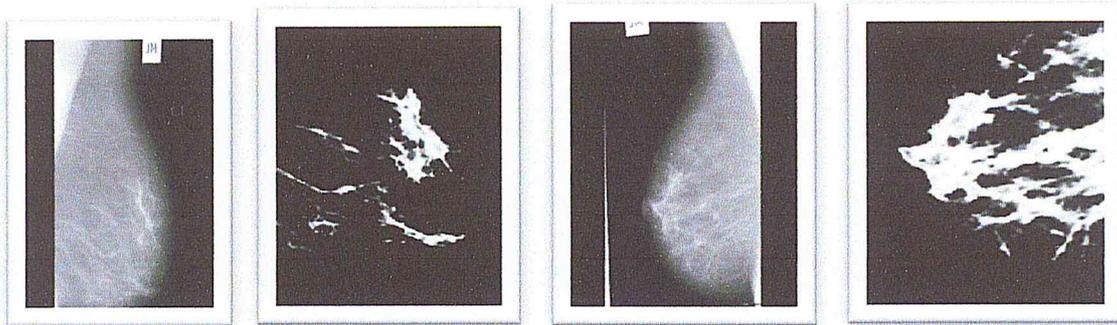


zone suspecte



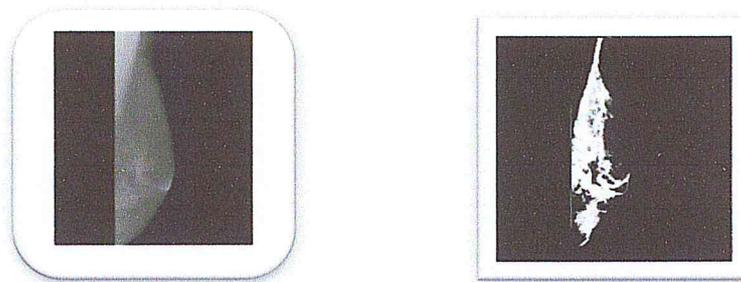
segmentation hybride (itérations=120)

Figure4. 15 : Résultats de la segmentation sur les masses bénignes



Segmentation de mdb006 (disque=80 et itérations=100)

Segmentation de mdb009 (disque=100 et itérations=100)



Segmentation de mdb060 (disque=80 et itérations=100)

Figure4. 16: Résultats de la segmentation sur les images saines

4.4.3 La caractérisation par les paramètres de formes et texturaux

Cette étape, permet l'extraction des caractéristiques texturaux et de formes, pour les régions segmentées. Les paramètres adoptés sont données par le tableau 4.2.

Paramètres de formes	Paramètres texturaux
Variance	Compacité
Entropie	Convexité
Contraste	Excentricité
Corrélation	Rectangularité
	Surface

Tableau4. 1: Paramètres de formes et texturaux adoptés

Notre base est répartie sur trois classes, suivant des masses malignes, et bénignes, ainsi que, des cas sains. Cette base a été analysée en explorant quatre paramètres de texture (variance, entropie, contraste et, corrélation) et cinq attributs de forme (la compacité, la convexité, l'excentricité, la rectangularité et la surface).

L'objectif, est de réaliser une base de connaissances via, cette description.

Les tableaux suivants, représentent l'évaluation des paramètres utilisés pour la caractérisation des images mammographiques.

Image	Variance	Entropie	Contraste	Corrélation	Compacité	Excentricité	Convexité	Rectangularité	Surface
mdb115	1.4767e+19	0.1161	0.2959	0.9850	0.4152	0.72959	0.18078	0.42021	300993
mdb155	2.5135e+18	0.4980	0.2424	0.9895	0.37182	0.75724	0.018341	0.34423	1.9215e+05
mdb075	2.5087e+18	0.9544	0.2409	0.9896	0.47832	0.79225	0.20557	0.34104	1.9003e+05
mdb023	2.4857e+18	0.2730	0.2674	0.9885	0.47304	0.44073	0.30842	0.34084	1.9006e+05
mdb028	2.5182e+18	0.4489	0.2742	0.9881	0.44557	0.44741	0.30975	0.34085	1.9088e+05
mdb270	2.5279e+18	0.7281	0.2477	0.9893	0.41821	0.68398	0.15495	0.34104	1.9098e+05
mdb178	2.5169e+18	0.1161	0.2735	0.9882	0.43169	0.67904	0.20667	0.34086	1.9104e+05
mdb184	2.5620e+18	0.1161	0.2764	0.9880	0.32433	0.85021	0.018998	0.34239	1.9318e+05
mdb202	2.5216e+18	0.9544	0.2449	0.9894	0.46381	0.43596	0.30893	0.34104	1.9038e+05
mdb134	2.5004e+18	0.1161	0.2709	0.9883	0.4509	0.57942	0.30918	0.34085	1.9053e+05

Tableau4. 2 : Tableau des paramètres de masses malignes

Image	Variance	Entropie	Contraste	Corrélation	Compacité	Excentricité	Convexité	Rectangularité	Surface
mdb132	2.2302e+18	0.9284	0.2392	0.9899	0.089777	0.53035	0.13856	0.16798	1.2808e+05
mdb080	2.5086e+18	0.9993	0.2404	0.9896	0.48616	0.41979	0.20553	0.34104	1.8999e+05
mdb017	2.5257e+18	0.4980	0.2453	0.9894	0.4574	0.34013	0.20912	0.34673	1.933e+05
mdb002	2.5363e+18	0.4489	0.2478	0.9893	0.41418	0.72754	0.20701	0.34104	1.9136e+05
mdb104	2.4267e+18	0.4489	0.3678	0.9838	0.4634	0.4536	0.31774	0.35458	1.8906e+05
mdb083	2.5204e+18	0.9745	0.2445	0.9894	0.46255	0.63828	0.3089	0.34104	1.9036e+05
mdb167	2.5088e+18	0.6962	0.2398	0.9897	0.48558	0.76335	0.30826	0.34104	1.8997e+05
mdb150	2.5022e+18	0.2006	0.2721	0.9882	0.4293	0.5979	0.12383	0.34085	1.9078e+05
mdb127	2.5083e+18	0.8351	0.2428	0.9895	0.42978	0.77359	0.15439	0.34104	1.9028e+05

Tableau4. 3 : Tableau des Paramètres des masses bénignes

Image	Mdb313	mdb060	Mdb009	mdb006	Mdb084	Mdb250	Mdb201 p	Mdb201	Mdb040	mdb003
Variance	1.5890e+ 18	1.8305e+ 18	1.6954e+ 18	1.9695e+ 18	5.372e+1 6	7.3804e+ 16	1.6295e+ 18	2.8184e+ 18	1.3437e+ 17	6.3049e+ 18
Entropie	0	0	0	0	0	0	0	0	0	0
Contraste	0.3909	0.18882	0.4440	0.3097	0.19337	0.38024	0.3046	0.3347	0.20105	0.1890
Corrélation	0.9837	0.99211	0.9816	0.9869	0.99167	0.98245	0.9872	0.9850	0.99168	0.9921
Compacité	0.49531	0.17439	0.66425	0.27136	0.27134	0.5527	0.57607	0.13439	0.3762	0.12909
Excentricité	0.56366	0.48426	0.53447	0.66987	0.76218	0.47404	0.67291	0.729	0.50128	0.72475
Convexité	0.03139	0.034319	0.070793	0.012832	0.03269	0.086721	0.051995	0.01428	0.029563	0.031838
Rectangularité	0.54108	0.19796	0.57925	0.41356	0.33415	0.41846	0.62352	0.3772	0.30239	0.31743
Surface	1.1655e+ 05	1.4804e+ 05	75182	82837	73356	85825	99901	79316	88804	1.2995e+ 05

Tableau4. 4 : Tableau des Paramètres des cas sains

L'extraction de paramètres de forme texturaux des régions étiquetées, nous permet de constituer des vecteurs caractéristiques de chacune des formes étudiées, en vue d'une classification par les Multi-SVM, pour une prise de décision.

La mesure de ces descripteurs, nous permet de réaliser des graphes statistiques, afin de voir leur influence sur les différentes masses étudiées.

Nous remarquons que dans le cas d'images saines, les paramètres descriptifs, sont du même ordre, alors que pour les cas pathologiques, ils varient suivant le type de tissu mammaire et la lésion.

Nous montrons à cet effet (figures 4.17 ; 4.18, la variation de l'entropie, de la corrélation, de l'excentricité, de la compacité et de la rectangularité, suivant les masses bénignes et malignes ainsi que les cas sains.

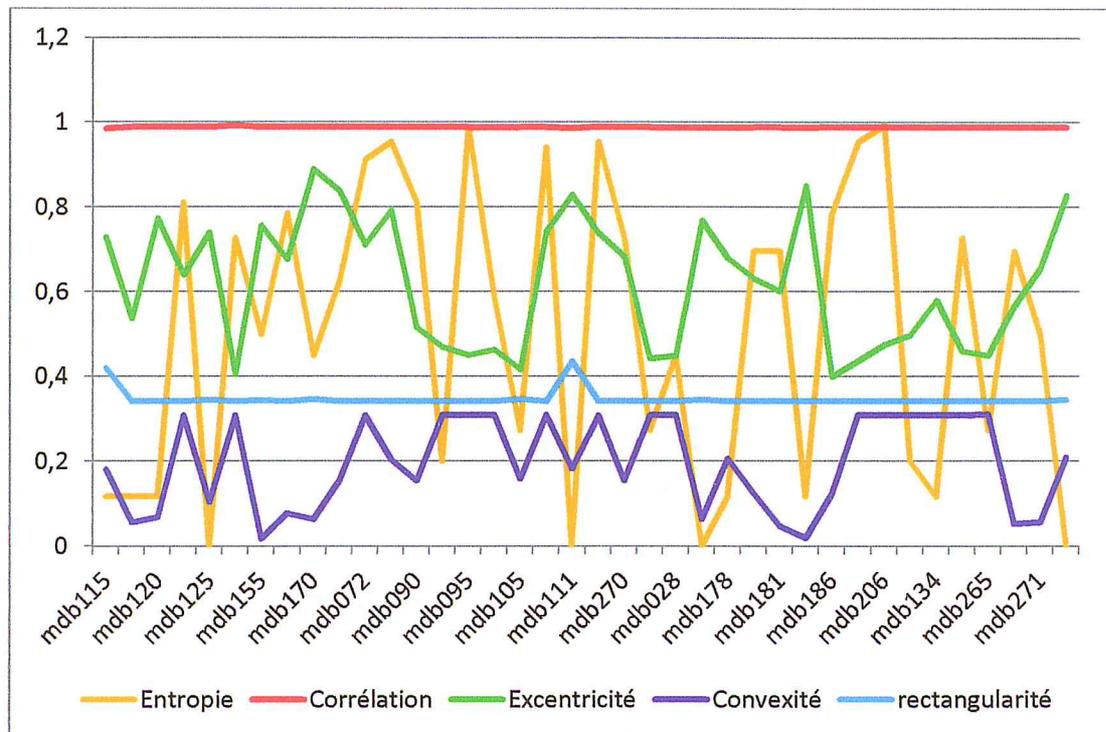


Figure4. 17 : Influence des caractéristiques sur des masses malignes.

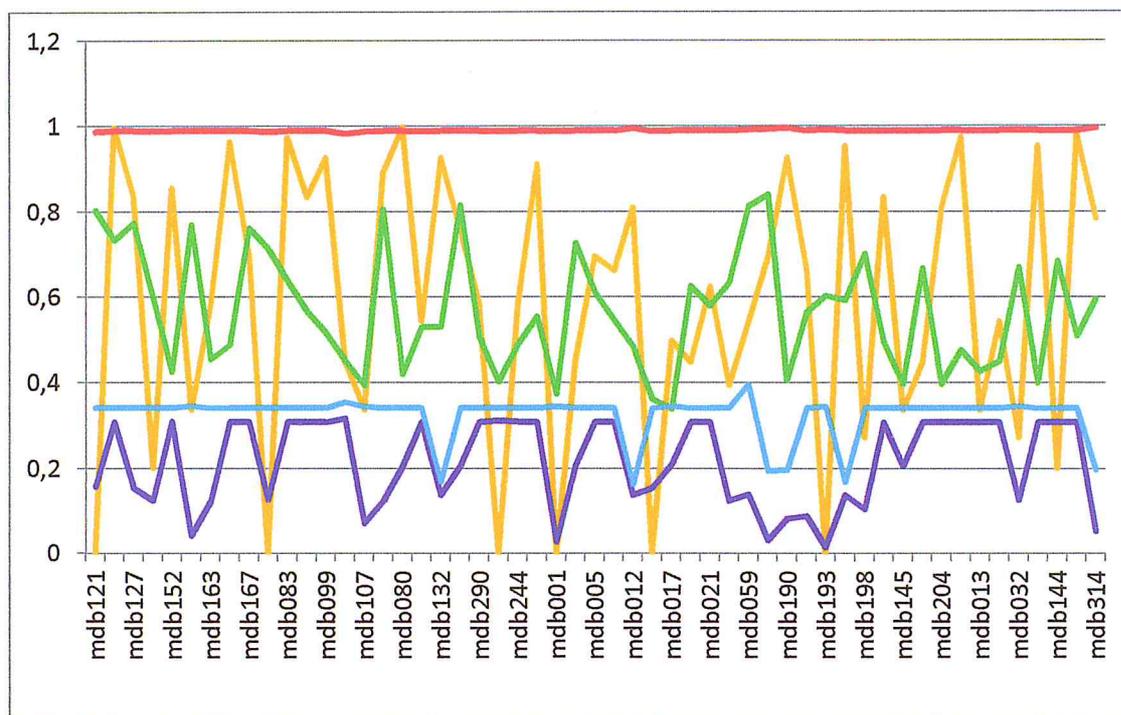


Figure4. 18 : Influence des caractéristiques sur des masses bénignes

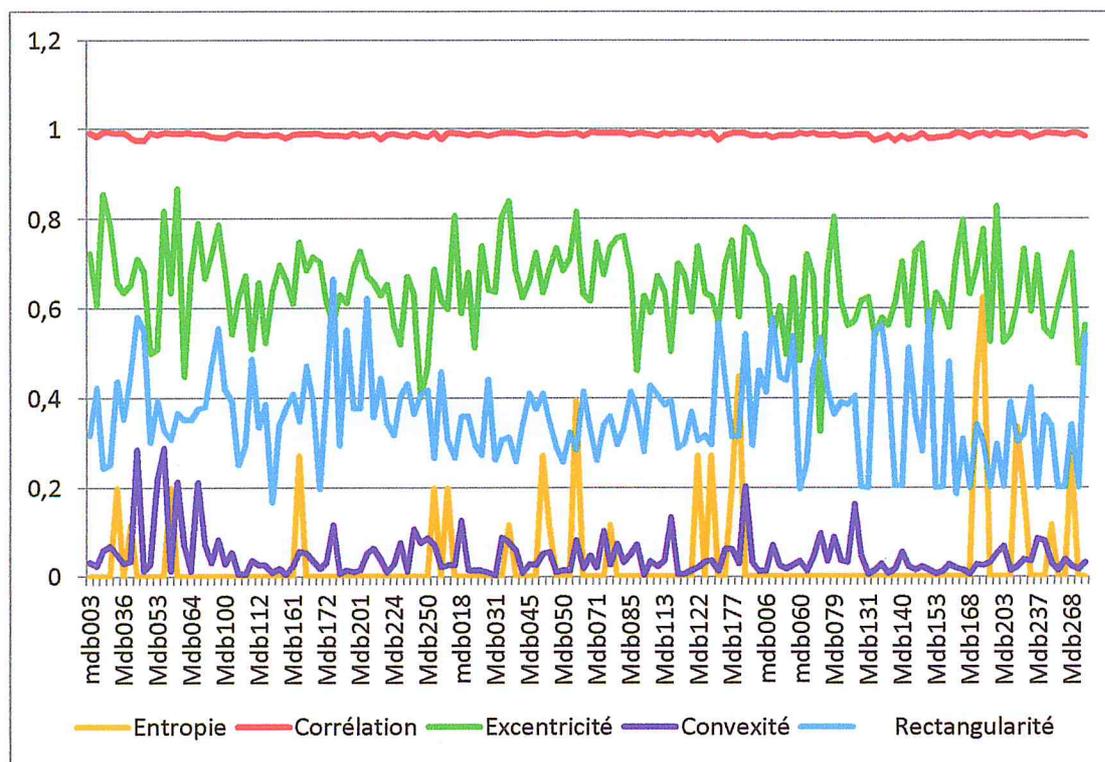


Figure4. 19 : Influence des caractéristiques sur des cas sains

4.4.4 La classification

La classification effectuée par le classifieur des SVM, permet l'identification du tissu normal et anormal, en utilisant des descripteurs de texture et de forme. L'apprentissage réalisé est en mode supervisé, car les images utilisées sont déjà diagnostiquées par des experts. Les résultats obtenus sont très encourageants, dans la mesure où une amélioration dans l'apprentissage peut réduire, le nombre de faux positifs.

1. La base considérée

Dans notre étude, nous avons sélectionné les 2/3 de la base (c'est-à-dire 159 images représentant des masses) pour l'apprentissage et 1/3 pour la phase d'exploitation (autour de 80 images représentant les masses). L'apprentissage, se fait via la stratégie "un contre un", suivant un kernel gaussien, pour la séparation des classes. L'avantage des SVM, est la rapidité de l'entraînement et l'optimisation des données.

La stratégie un contre un utilise le modèle SVM, pour chaque paire de classes. Le nombre total de modèles utilisés pour un problème de m classes, est $m(m-1)/2$ modèles. Chaque modèle réalise une classification binaire et, sépare les données en deux classes. Il considère les données de la première classe comme des éléments positifs et les données de la deuxième classe, comme des éléments négatifs. Ce système de classification utilise le vote majoritaire, comme un critère de décision. Pour classer une nouvelle observation on utilise, une mesure de vote calculée à partir de la sortie de chaque modèle et l'observation est affectée à la classe qui obtient le plus grand nombre, de votes. Il est possible que deux classes obtiennent le même nombre de votes.

1. Intégration de trois SVM binaires

Les SVM étant des classificateurs binaires, c'est-à-dire, un problème à deux classes. La résolution d'un problème multi-classes est réalisée en le transformant en une combinaison de problèmes binaires.

2. Phase d'apprentissage

L'apprentissage vise à découvrir les relations et motifs cachés dans des collections des données brutes. Un système d'apprentissage prend en entrée un ensemble d'exemples dont il cherche une définition. Dans notre cas, on cherche à apprendre le système à classifier les masses et les cas sains à partir des attributs extraits, fournis à l'entrée des SVM, selon la stratégie un contre un. Ces attributs sont présentés sous forme de vecteurs, qui représentent les signaux diffusés par le sein. La sortie du modèle est sous forme d'un vecteur, où on peut qualifier la tumeur de bénigne si c'est -1, maligne si c'est 1, sinon si c'est 0, il s'agit d'un cas sain. Ce qui fait correspondre trois classes à savoir (1 : masse maligne, -1 : masse bénigne, 0 : cas sain).

L'apprentissage a produit les résultats où:

- La matrice des poids correspond aux valeurs : [0.850299401197, 0.317365269461, 0.832335329341].
- Codes des classes : [-1;0;1] ;
- Les probabilités a posteriori pour chaque classe correspondent aux valeurs suivantes: [0.167664670658683, 0.682634730538922, 0.149700598802395];
- Taux d'erreur: 0.011976047904192.

3. Résultats de test

Une fois la phase d'apprentissage achevée, nous avons testé les performances du SVM pour déterminer le maximum a posteriori d'appartenance à chaque classe. Nous avons remarqué que le taux de classification obtenu est de 99% pour la phase de test.

4.5 Conclusion

La détection des pathologies mammaires est une tâche ardente à la variabilité des tissus, d'une patiente à une autre. A cet effet, le prétraitement des images mammographiques, a été d'une grande utilité, pour leur segmentation. La coopération, des techniques des modèles déformables et du clustering flou, a produit des résultats performants, quant à la détection des masses, suivant des paramètres de détection adaptés à chaque cas. Nous arrivons ainsi, à mieux distinguer les différentes formes. Les level sets permettent de gérer automatiquement le changement de topologie de la courbe en cours d'évolution, ce qui a engendré des résultats intéressants, comme le montrent les figures ci-dessus.

Dans ce travail, l'approche des multi-SVM a été utilisée pour la classification des masses et des images saines mammographiques, suivant la stratégie "un contre un". Les résultats de simulation sur les images de la base MIAS, sont satisfaisants suivant le taux de classification (99%).

Cette étude a confirmé la performance des SVM, dans le domaine de la reconnaissance des pathologies mammaires, suivant un choix judicieux des attributs de caractérisation.

Conclusion générale

La détection automatique du cancer du sein a été étudiée, pendant plus de deux décennies, grâce aux techniques de la vision par ordinateur. Cette étude a été menée, dans le but de la détection de masses et de cas sains en mammographie numérique.

Dans ce projet, nous avons étudié et développé, un système de détection et de classification de structures pathologiques mammographiques établies par les masses malignes et bénignes et les cas sains, en vue d'une aide à la décision. Les traitements ont été effectués sur les images mammographiques issues de la base de données MIAS.

Dans un premier lieu, les images mammographiques originales, ont été préparées en utilisant les fonctions de la morphologie mathématique en particulier, celle du chapeau haut de forme pour faire un rehaussement de contraste afin d'améliorer la qualité de l'image et, faciliter la phase de segmentation.

La détection des zones d'intérêt s'est faite par une technique de segmentation coopérative, qui a prouvé sa robustesse, lorsque nous choisissons les paramètres convenables, à savoir le nombre d'itérations pour les contours actifs. L'initialisation est effectuée, par l'approche du clustering flou, région et le raffinement par l'approche du modèle déformable. Les résultats engendrés avec une variation de paramètres, sont satisfaisants, puisque l'application coopérative sur des images mammographiques, a détecté la forme des lésions convenablement.

Une caractérisation texturale et morphologique des masses et des cas sains, a été abordé en second lieu. Les facteurs extraits, fournissent des mesures quantitatives de la granularité des masses et, les facteurs morphologiques, permettent une description de la géométrie.

En dernier lieu, la phase de classification des régions d'intérêts segmentées est réalisée suivant les machines à vecteurs de support SVM, pour une discrimination des masses et des cas sains.

Les résultats que nous avons obtenus sont acceptables et, permettent d'espérer à un aboutissement, en étudiant d'autres pathologies mammaires. Nous pouvons extraire d'autres types de caractéristiques texturales et morphologiques de l'image, en vue de caractériser davantage de pathologies.

Notre système, a pu classifier les masses de type malin ou bénin et les images saines, avec un taux de reconnaissance de 99%.

Notre système constitue un outil d'aide à la décision, qui peut donner aux radiologues et aux experts une seconde opinion dans la détection des anomalies. L'entreprise d'une étude de classification par un autre type de classifieur, tel que les arbres de décision, serait envisageable pour une comparaison avec les résultats des machines à support de vecteurs. Bien sûr, d'autres options peuvent être explorées, car comme nous le savons, le monde de la recherche n'est pas prêt de s'arrêter.

Bibliographie

- [1] L'OMS s'inquiète de la hausse des cancers dans le monde, <http://www.lemonde.fr/sante/article/2013/12/12/hausse-du-nombre-de-cancers-le-constat-alarmiste-de-l-oms.html>, décembre 2013, date d'accès, janvier 2015.
- [2] M. Bouziani, le cancer du sein en augmentation <http://www.algerie360.com/algerie/cancer-du-sein-le-nombre-de-cas-en-augmentation-novembre-2014>, date d'accès Janvier 2015.
- [3] Anatomie et physiologie du sein, Société canadienne du cancer, <http://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/anatomy-and-physiology/?region=qc>, date d'accès, septembre 2014.
- [4] les seins et les glandes mammaires (notions avancées), [http://bv.alloprof.qc.ca/science-et-technologie/l'univers-vivant/le-maintien-de-la-vie/les-tissus,-les-organes-et-les-systemes-biologiques/les-systemes-biologiques/le-systeme-reproducteur-feminin/l'anatomie-du-systeme-reproducteur-de-la-femme/les-seins-et-les-glandes-mammaires-\(notions-avancees\).aspx](http://bv.alloprof.qc.ca/science-et-technologie/l'univers-vivant/le-maintien-de-la-vie/les-tissus,-les-organes-et-les-systemes-biologiques/les-systemes-biologiques/le-systeme-reproducteur-feminin/l'anatomie-du-systeme-reproducteur-de-la-femme/les-seins-et-les-glandes-mammaires-(notions-avancees).aspx), 1996, date d'accès, septembre 2014.
- [5] Cancer du sein, institut National du Cancer, France, <http://www.e-cancer.fr/cancerinfo/les-cancers/cancer-du-sein/le-sein>, 2010, date d'accès, septembre 2014.
- [6] A. Chevallier, « Tumeurs du sein », Cours sur l'anatomie pathologique, Faculté de Médecine, Université de Nice, <http://anapath.unice.fr/sitelocal/disciplines/niveaudiscipline/anapath/lecon14/leconimprim.pdf> 2000 , date d'accès, septembre 2014.

- [7] P. Kestener, *Analyse multifractale 2D et 3D à l'aide de la transformation en ondelettes: application en mammographie et en turbulence développée*, Thèse de doctorat, Université de Bordeaux, 2003.
- [8] « Comprendre la mammographie », fédération Nationale des Centres de Lutte contre le cancer,
http://www.centreleonberard.fr/Portals/0/Documents/parcours_de_soins/examens/sein_mammographie.pdf, Septembre 2002, date d'accès, septembre 2014.
- [9] A. Le Treut, M. H. Dilhuydy, *Mammography : a guide to interpretation*, Ed Mosby, 1991.
- [10] J.F. HERON, les calcifications mammaires cancérologie générale, Faculté de Médecine de Caen France,
http://www.oncoprof.net/Generale2000/g04_Diagnostic/Mammographie/g04_mm10.html, décembre 2011, date d'accès, novembre 2014.
- [11] Classification en six catégories des images mammographiques en fonction du degré de suspicion de leur caractère pathologique, correspondance avec le système BIRADS de l'*American College of Radiology*(ACR), ANAES / Service des recommandations et références professionnelles / Février 2002.
- [12] J. P. Cocqueret & S. Philipp, *Analyse d'images: filtrage et segmentation*. Masson, 1995.
- [13] A. Rick, *Représentation de La Variabilité dans Le Traitement d'Images Flou*, Thèse de doctorat, université Paris VI, 1999.
- [14] G. Kom , A. Tiedeu , M. Kom, C. Nguemgne, J. Gonsu , *Détection automatique des opacités dans les mammographies par la méthode de minimisation de la somme de l'inertie*, ITBM-RBM, Elsevier, pp 347–356, 2005.
- [15] S. Beucher, *Segmentation d'images et morphologie mathématique*, thèse de doctorat, Ecole Nat. Sup. des Mines de Paris, 1990.
- [16] M.Coster et J.L.Cherman, « *Traitement d'image par morphologie mathématique application au domaine médical et industriel* », Thèse de doctorat, université B. Pascal, 1987.

- [17] J.-L. ROSE, Croissance de régions variationnelles et contraintes géométriques tridimensionnelles pour la segmentation d'image, Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2008.
- [18] J. J. Rousselle, les contours actifs, méthode de segmentation: application à l'imagerie médicale, thèse de Doctorat, Université F. Rabelais de Tours, juillet 2003.
- [19] J.M. Rendon Mancha, " régions actives morphologiques: application à la vision par ordinateur ", thèse de Doctorat, Université R. Descartes, Paris V, 2002.
- [21] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai and H. Du: Approaches for automated detection and classification of masses in mammograms, Elsevier, vol. 39, no.4, pp. 646-668, 2006.
- [22] P. Vigneaux: 'Méthodes des level set pour des problèmes d'interfaces micro fluidiques', thèse de doctorat, Université Bordeaux, 2007.
- [23] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image. Process.* vol. 10, no. 2, pp. 266–277, Feb. 2001.
- [24] R. Crandall, Image Segmentation Using the Chan-Vese Algorithm, ECE532 project, http://math.arizona.edu/~rcrandall/ECE532_ProjectPaper.pdf, 2009, date d'accès, janvier 2015.
- [25] M. Meliani, segmentation d'Image par coopération régions-contours, mémoire de Magister en informatique, Ecole Doctorale STIC, Alger, 2010.
- [26] F. Chamroukhi, " Méthodes de classification pour la segmentation d'image ", cours de Master, Université de Toulon - Var. <http://chamroukhi.univ-tln.fr/courses/Chamroukhi-cours-image-seg.pdf>, 2012, accès Janvier 2015.
- [27] LAKHDARI Rachida, La détection des microcalcifications dans les images mammographiques, mémoire de magister, Université A. B. Belkaid de Tlemcen, 2011.
- [28] J.C. Atine, Méthodes d'apprentissage flou : application à la segmentation d'images biologiques, thèse de doctorat, Institut National des Sciences Appliquées de Toulouse, 2011.
- [29] S. Liu , E.J. Delp, Multiresolution detection of stellate lesions in mammograms, *IEEE transactions on medical imaging*, pp 10-112, 1997.

- [30] L. Shen, R. M. Rangayyan, J.E L.Desautels, Detection and classification of mammographic calcifications, International Journal of Pattern Recognition and Artificial Intelligence, Vol 7, N^o 6, pp 1403-1415, June 1993.
- [31] F.Z Saïdi, réalisation d'algorithmes de segmentation: application sur des images du cancer du sein. Master en Électronique Biomédicale, Université Aboubekr Belkaid-Tlemcen. Juin 2012.
- [32] C.VACHIER, extraction de caractéristiques, segmentation d'image et morphologie mathématique, thèse de doctorat, école des mines de Paris, 1995.
- [33] A. Boujelben, H. Tmar, J.Mnif, M. Abid, Automatic application level set approach in detection calcifications in mammographic image, International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 4, pp 1-14, 2011.
- [34] F. Boudierba, A. Y. Brahim, analyse topologique d'images mammographiques, mémoire d'ingénieur, Département d'électronique, Université S. Dahlab de Blida, Septembre 2009.
- [35] K. F. Cissé, caractérisation texturale des images mammographiques, mathématique, mémoire de Master, Département d'électronique, Université S. Dahlab de Blida, Juillet 2011.
- [36] S.Haoulia, Détection morphologique et texturale de pathologies mammographiques pour l'aide à la décision au diagnostic, mémoire de Master, Département d'électronique, Université de Blida 1, Juillet 2013.
- [37] N.Rendja, H. Belaïd, détection morphométrique d'opacités et de microcalcifications mammographiques pour l'aide à la décision, mémoire de Master, Département d'électronique, Université de Blida 1, Juin 2014.
- [38] A.S .Aïdi, Analyse connexionniste de pathologies mammographiques pour l'aide à la décision, mémoire de Master, Département d'électronique, Université de Blida 1, Juin 2014.
- [39] M. Djebari, A. Kerkache, Détection et classification connexionniste de structures pathologiques mammographiques, mémoire de Master, Département d'électronique, Université de Blida 1, Juillet 2013.
- [40] B. N. Li et al. Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation. Computers in Biology and Medicine 41 (2011) pp: 1–10.

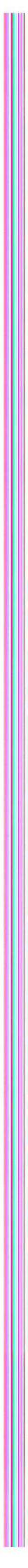
- [41] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification", *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-3, No.6, November 1973, pp.610-621.
- [42] S. Herlidou, caractérisation tissulaire en IRM par l'analyse de texture : étude du tissu musculaire et de tumeurs intracrâniennes, Thèse de doctorat, Université de Rennes, 1999.
- [43] A. M. Khuzi, R. B. Beng, Z. W. Wan and N. A. Mengsc: Identification of masses in digital mammogram using gray level co-occurrence matrices, *Biomedical Imaging and Intervention Journal*, vol. 5, no. 3, pp. e17, 2009.
- [44] K.Hassan, Exploitation des grandes bases de données d'image sur la grille de calcul, Mémoire de DEA, Université Lumière Lyon 2, France, Juillet 2003.
- [45] I. Cheikhrouhou, Description et classification des masses mammaires pour le diagnostic du cancer du sein, thèse de doctorat, Université d'Evry-Val d'Essonne, 2012.
- [46] F.Z.Regueieg, N.Benblidia, M.Guerti, Morphometric appreciation of mammary breast calcifications around multiresolution analysis by inertia and frequeniel factors, ACIT2005, Amman, Jordanie, Décembre 2005.
- [47] C. Bouveyron, modélisation et classification des données de grande dimension application à l'analyse d'images, thèse de Doctorat, université J. Fourier, 2006.
- [48] G. Dreyfus, M. Samuelides, J.-M. Martinez, M. B. Gordon, F. Badran, S. Thiria, L. Hérault, réseaux de neurones (méthodologies et applications), Eyrolles, 2004
- [49] L. Breimann, J. Friedman, R. Olsen, and C. Stone. Classification and regression trees. California: Wadworth International, 1984.
- [50] Machines à vecteurs supports, <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-svm.pdf>, date d'accès, Février 2015.
- [51] O. Kadri, L. H. Mouss, M. D. Mouss, A. Abdelhadi.Reconnaissance des Formes par SVM pour le Diagnostic du Système de Pasteurisation d'une Usine de Lait.Laboratoire d'Automatique et Productique Université de Batna,janvier 2013.

[52] L. O. Martins et al. Detection of Masses in Digital Mammograms using K-means and Support Vector Machine. *Electronic Letters on Computer Vision and Image Analysis* 8(2):39-50, 2009.

[53] F. Eddaoudi. Aide à la détection des pathologies du sein par les techniques de segmentation et de codage de textures des images mammographiques. Thèse de doctorat. Université de Rabat. Juillet 2012.

[54] Medical Image/Video Engineering Group, MIAS: Mammographic Image Analysis Society, <http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>.

[55] R. C. Gonzalez, *Digital Image processing using Matlab*, Pearson publication, 2005.



1

2